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

матриц, с помощью которых эти матрицы приводятся к диагональному виду. Если известны собственное значение λ симметрической трехдиагональной матрицы порядка m и соответствующий ему собственный вектор, то алгоритм исчерпывания позволяет с помощью подобных ортогональных преобразований вращения привести эту матрицу к клеточно-диагональной форме, в которой одна клетка опять является симметрической трехдиагональной матрицей порядка $(m-1)$, а вторая имеет порядок 1 и просто совпадает с λ. Циклическое повторение процесса исчерпывания дает возможность привести исходную матрицу к диагональной форме. Этот алгоритм хорошо известен и широко освещен в литературе (см., например, [1]). Вариант для сингулярного разложения двухдиагональной матрицы описан в [2].

При практическом применении известных реализаций алгоритмов исчерпывания иногда обнаруживается численная их неустойчивость, состоящая в том, что окончательная форма преобразованной матрицы отличается от желаемой. В работе выяснены причины этой неустойчивости и приведены способы их устранения, благодаря чему разработаны численно устойчивые варианты алгоритмов исчерпывания трехдиагональных симметрических и двухдиагональных матриц с гарантированной оценкой точности.

Продемонстрируем проблемы, возникающие при реализации алгоритмов исчерпывания, на примере симметрической трехдиагональной матрицы

$$A=\begin{bmatrix} d_1 & b_2 & & & 0 \\ b_2 & d_2 & b_3 & & \\ & \ddots & \ddots & \ddots & \\ & & b_{m-1} & d_{m-1} & b_m \\ 0 & & & b_m & d_m \end{bmatrix},$$

не имеющей нулевых элементов на побочных диагоналях, $b_i \neq 0$. Предположим, что известны собственное значение λ этой матрицы и отвечающий ему собственный вектор $u=(u_1, u_2, \ldots, u_m)^T$, так что имеют место равенства

$$\begin{gathered} (d_1-\lambda)u_1+b_2u_2=0; \\ b_iu_{i-1}+(d_i-\lambda)u_i+b_{i+1}u_{i+1}=0, \quad i=2, \ldots, m-1; \\ b_mu_{m-1}+(d_m-\lambda)u_m=0. \end{gathered} \tag{1}$$

Традиционно алгоритм исчерпывания состоит в следующем. Положим $u^{(1)}=u$ и подберем ортогональные матрицы вращения

$$C_i=\begin{bmatrix} 1 & & & & & & \\ & 1 & & & & 0 & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & c_i & -s_i & \\ & & & & s_i & c_i & \\ & & & & & & 1 \\ & 0 & & & & & \ddots \\ & & & & & & & 1 \end{bmatrix}, \quad i=2, \ldots, m,$$

так, чтобы вектор $u^{(i)}=C_iu^{(i-1)}$ имел вид $u^{(i)}=(0, 0, \ldots, 0, u_i^{(i)}, u_{i+1}, \ldots, u_m)^T$. Другими словами, параметры s_i, $c_i\,(s_i^2+c_i^2=1)$ находятся как решение уравнения $c_iu_{i-1}^{(i-1)}-s_iu_i=0$. При этом компонента $u_i^{(i)}=s_iu_{i-1}^{(i-1)}+c_iu_i$. Одно из решений уравнения $c_iu_{i-1}^{(i-1)}-s_iu_i=0$ есть

$$s_i=\frac{u_{i-1}^{(i-1)}}{\sqrt{[u_{i-1}^{(i-1)}]^2+u_i^2}}; \quad c_i=\frac{u_i}{\sqrt{[u_{i-1}^{(i-1)}]^2+u_i^2}}.$$

Тогда $u_i^{(i)} = \sqrt{[u_{i-1}^{(i-1)}]^2 + u_i^2}$ и выражения для s_i, c_i можно переписать так:

$$s_i = \frac{u_{i-1}^{(i-1)}}{u_i^{(i)}}; \quad c_i = \frac{u_i}{u_i^{(i)}}. \tag{2}$$

Отметим, что знаменатель в этих формулах не равен нулю, так как, очевидно, $u_1 \neq 0$ и, следовательно, $u_i^{(i)} = \sqrt{u_1^2 + u_2^2 + \ldots + u_i^2} \neq 0$. Поделим обе части i-го ($i = 1, 2, \ldots, m-1$) из равенств (1) на $u_{i+1}^{(i+1)}$ и последнего — на $u_m^{(m)}$. В результате простых преобразований, использующих формулы (2), получим

$$\begin{gathered} s_2(d_1 - \lambda) + c_2 b_2 = 0; \\ s_3 s_2 b_2 + s_3 c_2 (d_2 - \lambda) + c_3 b_3 = 0; \\ s_{i+1} s_i c_{i-1} b_i + s_{i+1} c_i (d_i - \lambda) + c_{i+1} b_{i+1} = 0, \quad i = 3, \ldots, m-1; \\ s_m c_{m-1} b_m + c_m (d_m - \lambda) = 0. \end{gathered} \tag{3}$$

С помощью этих равенств покажем, что матрица $\overline{A} = C_m C_{m-1} \cdots C_3 C_2 A C_2^* \times \times C_3^* \cdots C_{m-1}^* C_m^*$ имеет вид

$$\overline{A} = \begin{bmatrix} \bar{d}_1 & \bar{b}_2 & & & & & \\ \bar{b}_2 & \bar{d}_2 & \bar{b}_3 & & & 0 & \\ & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & 0 & & \bar{b}_{m-2} & \bar{d}_{m-2} & \bar{b}_{m-1} & \\ & & & & \bar{b}_{m-1} & \bar{d}_{m-1} & 0 \\ & & & & & 0 & \lambda \end{bmatrix}, \tag{4}$$

т. е. матрица A в результате применения к ней $m-1$ подобных преобразований вращения приводится к клеточно-диагональной форме, в которой одномерная клетка совпадает с собственным значением λ.

Обозначив $A_i = C_i C_{i-1} \cdots C_3 C_2 A C_2^* C_3^* \cdots C_{i-1}^* C_i^*$, покажем сначала, что матрица A_i выглядит так:

$$A_i = \begin{bmatrix} \bar{d}_1 & \bar{b}_2 & & & & & & & \\ \bar{b}_2 & \bar{d}_2 & \bar{b}_3 & & & & & & \\ & \ddots & \ddots & \ddots & & & & & \\ & & \ddots & \ddots & \ddots & & 0 & & \\ & & \bar{b}_{i-1} & \bar{d}_{i-1} & u_i & -s_i b_{i+1} & & & \\ & & & u_i & w_i & c_i b_{i+1} & & & \\ & & & -s_i b_{i+1} & c_i b_{i+1} & d_{i+1} & b_{i+2} & & \\ & & & & & \ddots & \ddots & & \\ & & & & & & \ddots & \ddots & \ddots \\ & 0 & & & & & & b_{m-1} & d_{m-1} & b_m \\ & & & & & & & b_m & d_m \end{bmatrix}, \tag{5}$$

где для элементов u_i, w_i имеют место формулы

$$u_i = -s_i[s_i c_{i-1} b_i + c_i(d_i - \lambda)]; \quad w_i = c_i[s_i c_{i-1} b_i + c_i(d_i - \lambda)] + \lambda.$$

Для этого предположим, что матрица A_{i-1} имеет такой же вид, если значение индекса i заменить на $i-1$, и вычислим матрицу $A_i = C_i A_{i-1} C_i^*$. Нетрудно проверить, что вычисленная по этой формуле матрица A_i записывается следующим образом:

$$A_i = \begin{bmatrix} \bar{d}_1 & \bar{b}_2 & & & & & & & & \\ \bar{b}_2 & \bar{d}_2 & \bar{b}_3 & & & & & & & \\ & \ddots & \ddots & \ddots & & & & & & \\ & & \bar{b}_{i-2} & \bar{d}_{i-2} & \bar{b}_{i-1} & z_{i-1} & & & & \\ & & & \bar{b}_{i-1} & \bar{d}_{i-1} & u_i & -s_i b_{i+1} & & & \\ & & & z_{i-1} & u_i & w_i & c_i b_{i+1} & & & \\ & & & & -s_i b_{i+1} & c_i b_{i+1} & d_{i+1} & b_{i+2} & & \\ & & 0 & & & & \ddots & \ddots & \ddots & \\ & & & & & & & b_{m-1} & d_{m-1} & b_m \\ & & & & & & & & b_m & d_m \end{bmatrix}$$

где использованы обозначения

$$\begin{aligned} &\bar{b}_{i-1} = c_i u_{i-1} + s_i s_{i-1} b_i; \quad z_{i-1} = s_i u_{i-1} - c_i s_{i-1} b_i; \\ &\bar{d}_{i-1} = c_i (c_i w_{i-1} - s_i c_{i-1} b_i) - s_i (c_i c_{i-1} b_i - s_i d_i); \\ &u_i = c_i (s_i w_{i-1} + c_i c_{i-1} b_i) - s_i (s_i c_{i-1} b_i + c_i d_i); \\ &w_i = s_i (s_i w_{i-1} + c_i c_{i-1} b_i) + c_i (s_i c_{i-1} b_i + c_i d_i). \end{aligned} \tag{6}$$

Подставляя выражения для u_{i-1}, w_{i-1} в формулы для z_{i-1}, u_i, w_i, находим

$$z_{i-1} = -s_{i-1}[s_i s_{i-1} c_{i-2} b_{i-1} + s_i c_{i-1}(d_{i-1} - \lambda) + c_i b_i];$$

$$u_i = c_i c_{i-1}[s_i s_{i-1} c_{i-2} b_{i-1} + s_i c_{i-1}(d_{i-1} - \lambda) + c_i b_i] - s_i[s_i c_{i-1} b_i + c_i (d_i - \lambda)];$$

$$w_i = s_i c_{i-1}[s_i s_{i-1} c_{i-2} b_{i-1} + s_i c_{i-1}(d_{i-1} - \lambda) + c_i b_i] + c_i[s_i c_{i-1} b_i + c_i (d_i - \lambda)] + \lambda.$$

Используя теперь равенства (3), видим, что $z_{i-1} = 0$, а u_i, w_i имеют требуемые выражения

$$u_i = -s_i[s_i c_{i-1} b_i + c_i (d_i - \lambda)]; \quad w_i = c_i[s_i c_{i-1} b_i + c_i (d_i - \lambda)] + \lambda.$$

Таким образом, предположение, что матрица A_i имеет вид (5), обосновано.

Рассмотрим теперь матрицу A_i при $i = m$, т. е. матрицу $A_m = \bar{A}$. Чтобы убедиться в том, что она имеет вид (4), выпишем формулы для элементов u_m, w_m:

$$u_m = -s_m[s_m c_{m-1} b_m + c_m (d_m - \lambda)];$$

$$w_m = c_m[s_m c_{m-1} b_m + c_m (d_m - \lambda)] + \lambda.$$

Выражение, заключенное в квадратные скобки, представляет собой в точности левую часть последнего из равенств (3). Следовательно, $u_m = 0$, $w_m = \lambda$ и матрица $\bar{A}$ действительно имеет вид (4). Переход от матрицы A к подобной ей матрице $\bar{A}$ (4) и составляет сущность алгоритма исчерпывания.

Особо подчеркнем, что при получении матрицы $\bar{A}$ использованы все равенства (3). В то же время последнее из них является в некотором смысле лишним. Конкретнее, для определения $(m-1)$ пар неизвестных s_i, c_i достаточно первых $(m-1)$ из этих равенств, а последнее должно выполняться автоматически, если λ — точное собственное значение. Если же λ мало отличается от точной величины собственного значения, то, казалось бы, это последнее уравнение должно быть почти выполненным. В действительности же бывают случаи, когда это не так, т. е. последнее уравнение имеет большую (по отношению к норме решения) невязку, хотя все остальные уравнения решены точно. По-видимому, первым обратил на это внимание Уилкинсон [см. [3], с. 286]. Один из наиболее показательных в этом отношении примеров рассмотрен в [4].

На практике параметры s_i, c_i подбираются именно из первых $(m-1)$ равенств (3). При этом из-за того, что последнее уравнение может иметь большую невязку, значения элементов u_m, w_m могут сильно отличаться от их точных значений, в частности значение элемента u_m может оказать-

ся недостаточно малым, чтобы им можно было пренебречь. Таким образом, первая причина неустойчивости алгоритма исчерпывания связана с недооценкой роли последнего уравнения, которое обычно считается выполненным. На первый взгляд, эту причину можно устранить, вычислив собственный вектор так, чтобы хорошо удовлетворялись все уравнения (1),— а значит, и уравнения (3), так как они являются следствием (1),— включая последнее, и определяя параметры s_i, c_i по формулам (2), т. е. непосредственно через компоненты собственного вектора. Однако простое преобразование этих формул к виду

$$s_i = \frac{1}{\sqrt{1 + c_{i-1}^2 \left(\frac{u_i}{u_{i-1}}\right)^2}}; \quad c_i = \frac{\frac{u_i}{u_{i-1}}}{\sqrt{1 + c_{i-1}^2 \left(\frac{u_i}{u_{i-1}}\right)^2}} \tag{7}$$

показывает, что параметры s_i, c_i определяются не столько самими компонентами собственного вектора, сколько их отношениями u_i/u_{i-1}. В таком случае ясно, что даже чрезвычайно точное определение собственного вектора в метрике евклидова пространства не гарантирует правильности выбора параметров s_i, c_i, поскольку оно все равно может приводить к неправильным отношениям компонент данного вектора. Это происходит, например, в том случае, когда среди компонент наряду с большими есть очень малые. Таким образом, вторая причина неустойчивости связана с недостаточным определением отношений компонент собственного вектора.

Задача вычисления отношений компонент собственного вектора с высокой точностью решена в [4], результаты которой использованы в данной работе. В частности, в главе 2 показано, что использование этих отношений позволяет добиться достаточно точного удовлетворения соотношений (3). При этом найдены некоторые «идеальные» параметры s_i, c_i $(s_i^2 + c_i^2 = 1)$, которые удовлетворяют соотношениям (3) приближенно, т. е. с некоторыми возмущениями коэффициентов d_i, b_i. Приведена оценка этих возмущений. Параметры s_i, c_i названы идеальными по той причине, что они в точности удовлетворяют соотношению $s_i^2 + c_i^2 = 1$ и используются только при выводе формул алгоритма, т. е. только в теоретических рассмотрениях. Реально же в памяти машины хранятся лишь некоторые их «машинные» приближения s_i', c_i', мало (в смысле относительной погрешности) отличающиеся от s_i, c_i. Чтобы различие между параметрами s_i, c_i и s_i', c_i' стало понятным, скажем, что s_i, c_i определяются в результате точных вычислений по формулам (7), в которых вместо c_{i-1} используется c_{i-1}', а s_i', c_i' представляют собой результаты машинной реализации этих формул. Обеспечение высокой относительной точности при вычислении s_i', c_i' важно по нескольким причинам. Во-первых, от величины погрешности параметров c_i' зависит величина возмущений коэффициентов d_i, b_i, с которыми выполняются равенства (3) для параметров s_i, c_i. Во-вторых, параметры s_i, c_i используются при построении явных формул, задающих элементы $\bar{d}_i$, $\bar{b}_i$ преобразованной матрицы $\bar{A}$ [см. (4)]. Понятно, что при реальных вычислениях в этих формулах можно использовать лишь приближения s_i', c_i'. Поэтому от их точности зависит точность определения преобразованной матрицы $\bar{A}$. Наконец, важно обеспечить, чтобы приближенная матрица C'_i (см. с. 115) была близка к ортогональной, ведь результатами работы алгоритма, кроме преобразованной матрицы, являются и матрицы, задающие требуемые преобразования вращения. Алгоритм вычисления величин s_i', c_i', обеспечивающий их высокую относительную точность, описан в § 4 главы 2; § 3 той же главы посвящен непосредственно решению уравнений (3) относительно параметров s_i, c_i.

В главе 1 на основе соотношений (3) выводятся явные формулы для элементов преобразованной матрицы $\bar{A}$. В традиционно используемом ал-

$$C_i' = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & 0 & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & c_i' & -s_i' & \\ & & & & s_i' & c_i' & \\ & 0 & & & & & 1 \\ & & & & & & & \ddots \\ & & & & & & & & 1 \end{bmatrix}.$$

горитме исчерпывания формулы для этих элементов имеют, как нетрудно видеть из (6), рекуррентный характер, что, конечно, затрудняет анализ погрешностей, возникающих при их реализации на вычислительной машине. В данной работе приведены окончательные формулы для этих элементов, так что сами преобразования вращения выполняются не в процессе вычислений, а аналитически. Поскольку уравнения (3) могут быть решены только приближенно, преобразованная матрица уже не будет иметь клеточно-диагональной формы (4), а окажется заполненной. В работе показано, что эта заполненная матрица может быть представлена в виде суммы двух матриц, одна из которых имеет требуемый вид (4), а другая, называемая матрицей погрешностей,— малую норму. Выводу формул для элементов преобразованной матрицы и оценке нормы матрицы погрешностей посвящен § 1.

При вычислении элементов преобразованной матрицы по полученным формулам неизбежно возникновение погрешностей. Эти погрешности имеют двоякую природу. Во-первых, как отмечено, вместо участвующих в этих формулах параметров s_i, c_i можно использовать только их машинные приближения s_i', c_i'. Во-вторых, погрешности возникают и при непосредственном выполнении арифметических операций, предписываемых этими формулами. Оба источника погрешностей учитываются в § 5 главы 3.

Проблемы, аналогичные описанным, имеют место в алгоритме исчерпывания двухдиагональной матрицы

$$A = \begin{bmatrix} a_1 & b_2 & & & \\ & a_2 & b_3 & & 0 \\ & & \ddots & \ddots & \\ & 0 & & a_{N-1} & b_N \\ & & & & a_N \end{bmatrix}.$$

Кратко опишем его, предполагая элементы a_i, b_i отличными от нуля. Пусть известно сингулярное число σ этой матрицы. Напомним, что сингулярными числами матрицы A называются N наибольших собственных значений составной матрицы

$$S = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}.$$

Алгоритм состоит в определении ортогональных матриц вращения C_i, $\bar{C}_i$ $(i = 2, \ldots, N)$

$$C_i = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & 0 & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & c_i & -s_i & \\ & & & & s_i & c_i & \\ & 0 & & & & & 1 \\ & & & & & & & \ddots \\ & & & & & & & & 1 \end{bmatrix};$$

$$\bar{C}_i = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & 0 & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & c_i & -\bar{s}_i & \\ & & & & \bar{s}_i & \bar{c} & \\ & 0 & & & & 1 & \\ & & & & & & \ddots \\ & & & & & & & 1 \end{bmatrix}$$

так, чтобы матрица $\bar{A} = \bar{C}_N \bar{C}_{N-1} \cdots \bar{C}_3 \bar{C}_2 A C_2^* C_3^* \cdots C_{N-1}^* C_N^*$ имела вид

$$\bar{A} = \begin{bmatrix} \bar{a}_1 & \bar{b}_2 & & & \\ \bar{a}_2 & \bar{b}_3 & & 0 & \\ & \ddots & \ddots & & \\ & & \bar{a}_{N-2} & \bar{b}_{N-1} & \\ 0 & & & \bar{a}_{N-1} & 0 \\ & & & & \pm\sigma \end{bmatrix}. \tag{8}$$

Аналогично случаю трехдиагональной матрицы нетрудно показать, что параметры s_i, c_i, $\bar{s}_i$, $\bar{c}_i$ $(s_i^2 + c_i^2 = 1,\ \bar{s}_i^2 + \bar{c}_i^2 = 1)$ должны удовлетворять равенствам

$$\left.\begin{aligned} & s_2 a_1 - \bar{s}_2 \frac{\sigma^2}{a_1} + c_2 b_2 = 0; \\ & \bar{s}_2 b_2 - c_2 \frac{\bar{s}_2}{s_2} a_1 + \bar{c}_2 a_2 = 0; \\ & s_{i+1} c_i a_i - \bar{c}_i \frac{s_{i+1} s_i \cdots s_2}{\bar{s}_i \cdots \bar{s}_2} \frac{\sigma^2}{a_1} + c_{i+1} b_{i+1} = 0, \\ & \bar{s}_{i+1} \bar{c}_i b_{i+1} - c_{i+1} \frac{\bar{s}_{i+1} \cdots \bar{s}_2}{s_{i+1} \cdots s_2} a_1 + \bar{c}_{i+1} a_{i+1} = 0 \end{aligned}\right\} i = 2, \ldots, N-1, \tag{9}$$

$$c_N a_N - \bar{c}_N \frac{s_N \cdots s_2}{\bar{s}_N \cdots \bar{s}_2} \frac{\sigma^2}{a_1} = 0.$$

Первое из этих равенств можно рассматривать как уравнение для определения параметров s_2, c_2, второе — как уравнение для $\bar{s}_2$, $\bar{c}_2$ и т. д. При этом последнее равенство остается неиспользованным. Если σ известно с некоторой, пусть даже достаточно малой, погрешностью, то неиспользованное уравнение может иметь большую невязку. Это приводит к тому, что элемент матрицы $\bar{A}$, стоящий в позиции $(N-1, N)$, теоретически равный нулю, в практических расчетах оказывается недостаточно малым, чтобы им можно было пренебречь. Поэтому возникает задача научиться достаточно точно решать уравнения (9). При ее решении опять будем опираться на возможность предварительного определения отношений компонент собственного вектора с высокой точностью (см. [4]). В данном случае речь идет о собственном векторе трехдиагональной симметрической матрицы

$$T = \begin{bmatrix} 0 & a_1 & & & & & \\ a_1 & 0 & b_2 & & & 0 & \\ & b_2 & 0 & a_2 & & & \\ & & a_2 & 0 & b_3 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & a_{N-1} & 0 & b_N \\ & 0 & & & & b_N & 0 & a_N \\ & & & & & & a_N & 0 \end{bmatrix},$$

отвечающем ее почти собственному значению σ. Отметим, что матрица T получается из матрицы S с помощью одноименных перестановок строк и столбцов. Задачам определения параметров s_i, c_i, $\bar{s}_i$, $\bar{c}_i$, удовлетворяющих (приближенным) равенствам (9), и вычисления их достаточно точных машинных приближений s_i', c_i', $\bar{s}_i'$, $\bar{c}_i'$ посвящены § 3, 4 главы 2, где они решаются параллельно с аналогичными вопросами для случая трехдиагональной матрицы. Отметим, что с достаточной точностью удалось решить уравнения (9) только в случае, когда σ является наибольшим сингулярным числом матрицы A.

Поскольку уравнения (9) можно решить только приближенно, т. е. они выполняются с некоторыми возмущенными коэффициентами a_i, b_i, матрица $\bar{A}$ уже не может иметь вида (8), а оказывается заполненной. Однако, как показано в данной работе, эту заполненную матрицу можно представить в виде суммы двух матриц, одна из которых имеет требуемую форму (8), а вторая (матрица погрешностей) — малую норму. Для элементов $\bar{a}_i$, $\bar{b}_i$ первой матрицы получены простые явные формулы. Их вывод и оценка нормы матрицы погрешностей составляют содержание § 2 главы 1; § 6 главы 3 посвящен анализу погрешностей, возникающих при непосредственном вычислении элементов $\bar{a}_i$, $\bar{b}_i$ по полученным формулам. В результате анализа находится оценка, показывающая, насколько вычисленная (т. е. рассматриваемая уже как таблица машинных чисел) преобразованная матрица отличается от некоторой матрицы, ортогонально эквивалентной исходной матрице A.

В § 7 и 8 главы 4 приведены общие схемы алгоритмов исчерпывания соответственно трехдиагональных симметрических и двухдиагональных матриц.

Автор считает приятным долгом отметить неизменную поддержку и внимание к его работе со стороны чл.-кор. АН СССР С. К. Годунова и выражает ему искреннюю признательность.

Глава 1

ФОРМУЛЫ АЛГОРИТМОВ ИСЧЕРПЫВАНИЯ. АНАЛИЗ ПОГРЕШНОСТЕЙ, СВЯЗАННЫХ С НЕТОЧНОСТЬЮ РЕШЕНИЯ УРАВНЕНИЙ ОТНОСИТЕЛЬНО ПАРАМЕТРОВ ВРАЩЕНИЯ

§ 1. Алгоритм исчерпывания симметрической трехдиагональной матрицы

Рассмотрим симметрическую трехдиагональную матрицу

$$A = \begin{bmatrix} d_1 & b_2 & & & \\ b_2 & d_2 & b_3 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \\ 0 & & b_{m-1} & d_{m-1} & b_m \\ & & & b_m & d_m \end{bmatrix} \tag{1.1}$$

такую, что $b_i \neq 0$. Предположим, что η — некоторое приближение к собственному значению λ этой матрицы, удовлетворяющее неравенству

$$|\eta - \lambda| \leqslant \gamma \mathcal{M}(A), \quad \gamma \ll 1, \tag{1.2}$$

и числа s_i, c_i удовлетворяют соотношениям

$$c_i^2 + s_i^2 = 1, \quad i = 2, \ldots, m, \tag{1.3}$$

$$s_i > 0, \quad i = 2, \ldots, m; \tag{1.4}$$

$$
\begin{gathered}
s_2(d_1+\alpha_1-\eta)+c_2b_2=0;\\
s_3s_2b_2(1+\beta_2)+s_3c_2(d_2+\alpha_2-\eta)+c_3b_3=0;\\
s_{i+1}s_ic_{i-1}b_i(1+\beta_i)+s_{i+1}c_i(d_i+\alpha_i-\eta)+c_{i+1}b_{i+1}=0,\quad i=3,\ \ldots,\ m-1;\\
s_mc_{m-1}b_m(1+\beta_m)+c_m(d_m+\alpha_m-\eta)=0,
\end{gathered}
\tag{1.5}
$$

где возмущения α_i и β_i удовлетворяют оценкам

$$
\begin{gathered}
|\alpha_i|\leqslant\bar{\alpha}\mathscr{M}(A),\quad \bar{\alpha}\ll 1;\\
|\beta_i|\leqslant\bar{\beta},\qquad \bar{\beta}\ll 1,
\end{gathered}
\tag{1.6}
$$

а $\mathscr{M}(A)$ обозначает одну из норм матрицы A:

$$
\mathscr{M}(A)=\max\begin{cases}|d_1|+|b_2|,\\ \max\limits_{2\leqslant i\leqslant m-1}(|d_i|+|b_i|+|b_{i+1}|),\\ |d_m|+|b_m|.\end{cases}
$$

Отметим, что получение соотношений (1.5) подробно описано в § 3 главы 2. Там же приведены формулы для нахождения возмущений α_i и β_i.

Определим ортогональные матрицы

$$
C_i=\begin{bmatrix}1&&&&&&\\&\ddots&&&&0&\\&&1&&&&\\&&&c_i&-s_i&&\\&&&s_i&c_i&&\\&0&&&&1&\\&&&&&&\ddots\\&&&&&&1\end{bmatrix},\quad i=2,\ \ldots,\ m,
$$

и их произведение $C=C_mC_{m-1}\cdots C_3C_2$.

Теорема 1. *Матрицу CAC^*, ортогонально подобную матрице A, можно представить в виде суммы двух матриц $CAC^*=\bar{A}+R$, где матрица $\bar{A}$ имеет вид*

$$
\bar{A}_i=\begin{bmatrix}\bar{d}_1&\bar{b}_2&&&&&\\\bar{b}_2&\bar{d}_2&\bar{b}_3&&0&&\\&\ddots&\ddots&\ddots&&&\\&&\bar{b}_{m-2}&\bar{d}_{m-2}&\bar{b}_{m-1}&&\\&0&&\bar{b}_{m-1}&\bar{d}_{m-1}&0&\\&&&&0&\eta\end{bmatrix},
\tag{1.7}
$$

а для спектральной нормы матрицы R имеет место оценка

$$
\|R\|\leqslant 2(\tilde{\alpha}+\sqrt{m}\tilde{\beta})\mathscr{M}(A).
\tag{1.8}
$$

При этом для элементов $\bar{d}_i$, $\bar{b}_i$ матрицы $\bar{A}$ справедливы формулы

$$
\begin{aligned}
&\bar{d}_1=d_2-\frac{c_2b_2}{s_2}+\frac{c_3c_2b_3}{s_3};\\
&\bar{d}_i=d_{i+1}-\frac{c_{i+1}c_ib_{i+1}}{s_{i+1}}+\frac{c_{i+2}c_{i+1}b_{i+2}}{s_{i+2}},\quad i=2,\ \ldots,\ m-2;\\
&\bar{d}_{m-1}=d_m-\frac{c_mc_{m-1}b_m}{s_m};\\
&\bar{b}_i=\frac{s_ib_{i+1}}{s_{i+1}},\quad i=2,\ \ldots,\ m-1.
\end{aligned}
\tag{1.9}
$$

Доказательство. Обозначив

$$
d_i'=d_i+\alpha_i,\ i=1,\ \ldots,\ m,
\tag{1.10}
$$

введем в рассмотрение вспомогательную матрицу

$$A' = \begin{bmatrix} d_1' & b_2 & & & & \\ b_2 & d_2' & b_3 & & 0 & \\ & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & 0 & & b_{m-1} & d_{m-1}' & b_m \\ & & & & b_m & d_m' \end{bmatrix}$$

и вычислим матрицу $CA'C^*$. Сформулируем рекуррентное правило для вычисления этой матрицы:

$$A_1 = A', \quad A_i = C_i A_{i-1} C_i^*, \quad i = 2, \ldots, m, \quad A_m \equiv CA'C^*.$$

Легко проверить, что матрица $A_2 = C_2 A_1 C_2^*$ имеет вид

$$A_2 = \begin{bmatrix} u_1 & v_2 & -s_2 b_3 & & & \\ v_2 & w_2 & c_2 b_3 & & 0 & \\ -s_2 b_3 & c_2 b_3 & d_3' & b_4 & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & 0 & & b_{m-1} & d_{m-1}' & b_m \\ & & & & b_m & d_m' \end{bmatrix},$$

где

$$u_1 = c_2(c_2 d_1' - s_2 b_2) - s_2(c_2 b_2 - s_2 d_2');$$
$$v_2 = c_2(s_2 d_1' + c_2 b_2) - s_2(s_2 b_2 + c_2 d_2');$$
$$w_2 = s_2(s_2 d_1' + c_2 b_2) + c_2(s_2 b_2 + c_2 d_2').$$

Преобразуем выражения для u_1, v_2 и w_2:

$$u_1 = d_2' + \frac{c_2^2}{s_2}[s_2(d_1' - \eta) + c_2 b_2] - \frac{c_2 b_2}{s_2} - \frac{c_2}{s_2}[s_3 s_2 b_2(1 + \beta_2) + s_3 c_2(d_2' - \eta)] + s_2 c_2 b_2 \beta_2;$$

$$v_2 = c_2[s_2(d_1' - \eta) + c_2 b_2] - \frac{s_2}{s_3}[s_3 s_2 b_2(1 + \beta_2) + s_3 c_2(d_2' - \eta)] + s_2^2 b_2 \beta_2;$$

$$w_2 = \eta + s_2[s_2(d_1' - \eta) + c_2 b_2] + \frac{c_2}{s_3}[s_3 s_2 b_2(1 + \beta_2) + s_3 c_2(d_2' - \eta)] - s_2 c_2 b_2 \beta_2.$$

Принимая во внимание первые два из соотношений (1.5) и обозначения (1.9), (1.10), получаем

$$u_1 = \bar{d}_1 + \alpha_2 + s_2 c_2 b_2 \beta_2; \quad v_2 = c_2 \bar{b}_2 + s_2^2 b_2 \beta_2;$$

$$w_2 = \eta - \frac{c_2 c_2 b_2}{s_3} - s_2 c_2 b_2 \beta_2.$$

Введем обозначения

$$b_{11} = s_2 c_2 b_2 \beta_2; \quad b_{21}' \equiv b_{12}' = s_2^2 b_2 \beta_2; \quad b_{22}' = -s_2 c_2 b_2 \beta_2. \qquad (1.11)$$

Тогда матрицу A_2 можно записать так:

$$A_2' = \begin{bmatrix} \bar{d}_1 + \alpha_2 + b_{11} & c_2 \bar{b}_2 + b_{12}' & -s_2 b_3 & & & \\ c_2 \bar{b}_2 + b_{21}' & \eta - \frac{c_2 c_2 b_2}{s_3} + b_{22}' & c_2 b_3 & & 0 & \\ -s_2 b_3 & c_2 b_3 & d_3' & b_4 & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & 0 & & b_{m-1} & d_{m-1}' & b_m \\ & & & & b_m & d_m' \end{bmatrix}.$$

Предположим теперь, что матрица A_i выглядит следующим образом:

$$A_i = \left[\begin{array}{c|cc} \bar{A}_{(i-1)} + X_{(i)} + B_{(i-1)} & \begin{matrix} b'_{1i} \\ \vdots \\ b'_{i-2,i} \end{matrix} & 0 \\ \hline \begin{matrix} b'_{i1} \dots b'_{i,i-2}\, c_{i+1}\bar{b}_i + b'_{i,i-1} \\ 0 \end{matrix} & \begin{matrix} c_{i+1}\bar{b}_i + b'_{i-1,i} & -s_i b_{i+1} \\ -s_i b_{i+1} & c_i b_{i+1} \end{matrix} & \begin{matrix} t_i + b'_{ii} & c_i b_{i+1} \\ c_i b_{i+1} & A'_{(i+1)} \end{matrix} \end{array}\right]. \quad (I)$$

Здесь через $\bar{A}_{(i-1)}$, $X_{(i)}$, $B_{(i-1)}$ обозначены матрицы порядка $i-1$:

$$\bar{A}_{(i-1)} = \begin{bmatrix} \bar{d}_1 & \bar{b}_2 & & & 0 \\ \bar{b}_2 & \bar{d}_2 & \bar{b}_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \\ 0 & & \bar{b}_{i-2} & \bar{d}_{i-2} & \bar{b}_{i-1} \\ & & & \bar{b}_{i-1} & d_{i-1} \end{bmatrix},$$

$X_{(i)} = \operatorname{diag}[a_2, a_3, \dots, a_i]$, $B_{(i-1)} = (b_{kl})$ $(k, l = 1, \dots, i-1)$;

через $A'_{(i+1)}$ — трехдиагональная матрица порядка $m-i$:

$$A'_{(i+1)} = \begin{bmatrix} d'_{i+1} & b_{i+2} & & & 0 \\ b_{i+2} & d'_{i+2} & b_{i+3} & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \\ 0 & & b_{m-1} & d'_{m-1} & b_m \\ & & & b_m & d'_m \end{bmatrix},$$

а элемент t_i имеет выражение $t_i = \eta - c_{i+1}c_i b_{i+1}/s_{i+1}$. Ясно, что матрица A_i должна быть симметрической, поэтому $b_{kj} = b_{jk}$ $(k = 2, \dots, i-1;\ j = 1, \dots, k-1)$; $b'_{ij} = b'_{ji}$ $(j = 1, \dots, i-1)$. Легко проверить, что матрица $A_{i+1} = C_{i+1}A_iC^*_{i+1}$ имеет следующий вид:

$$A_{i+1} = \left[\begin{array}{c|cc} \bar{A}_{(i-1)} + X_{(i)} + B_{(i-1)} & \begin{matrix} c_{i+1}b'_{1i} & s_{i+1}b'_{1i} \\ \vdots & \vdots \\ c_{i+1}b'_{i-2,i} & s_{i+1}b'_{i-2,i} \\ z_i & y_i \end{matrix} & 0 \\ \hline \begin{matrix} c_{i+1}b'_{i1} \dots c_{i+1}b'_{i,i-2}\ z_i \\ s_{i+1}b'_{i1} \dots s_{i+1}b'_{i,i-2}\ y_i \\ 0 \end{matrix} & \begin{matrix} u_i & v_{i+1} & -s_{i+1}b_{i+2} \\ v_{i+1} & w_{i+1} & c_{i+1}b_{i+2} \\ -s_{i+1}b_{i+2} & c_{i+1}b_{i+2} & \end{matrix} & A'_{(i+2)} \end{array}\right],$$

где для элементов z_i, y_i, u_i, v_{i+1} и w_{i+1} приняты обозначения

$$\begin{aligned} z_i &= c^2_{i+1}\bar{b}_i + s_{i+1}s_i b_{i+1} + c_{i+1}b'_{i-1,i}; \\ y_i &= s_{i+1}c_{i+1}\bar{b}_i - c_{i+1}s_i b_{i+1} + s_{i+1}b'_{i-1,i}; \\ u_i &= c_{i+1}\left[c_{i+1}(t_i + b'_{ii}) - s_{i+1}c_i b_{i+1}\right] - s_{i+1}(c_{i+1}c_i b_{i+1} - s_{i+1}d'_{i+1}); \quad (1.12) \\ v_{i+1} &= c_{i+1}\left[s_{i+1}(t_i + b'_{ii}) + c_{i+1}c_i b_{i+1}\right] - s_{i+1}(s_{i+1}c_i b_{i+1} + c_{i+1}d'_{i+1}); \\ w_{i+1} &= s_{i+1}\left[s_{i+1}(t_i + b'_{ii}) + c_{i+1}c_i b_{i+1}\right] + c_{i+1}(s_{i+1}c_i b_{i+1} + c_{i+1}d'_{i+1}). \end{aligned}$$

Из определения $\bar{b}_i$ следует, что

$$z_i = \bar{b}_i + c_{i+1}b'_{i-1,i};\ y_i = s_{i+1}b'_{i-1,i}.$$

С учетом обозначения для t_i преобразуем выражения для u_i, v_{i+1} и w_{i+1}:

$$u_i = d'_{i+1} - \frac{c_{i+1}c_ib_{i+1}}{s_{i+1}} - \frac{c_{i+1}}{s_{i+2}}\left[s_{i+2}s_{i+1}c_ib_{i+1}(1+\beta_{i+1}) + s_{i+2}c_{i+1}(d'_{i+1}-\eta)\right] +$$
$$+ c^2_{i+1}b'_{ii} + c_{i+1}s_{i+1}c_ib_{i+1}\beta_{i+1};$$

$$v_{i+1} = -\frac{s_{i+1}}{s_{i+2}}\left[s_{i+2}s_{i+1}c_ib_{i+1}(1+\beta_{i+1}) + s_{i+2}c_{i+1}(d'_{i+1}-\eta)\right] +$$
$$+ c_{i+1}s_{i+1}b'_{ii} + s^2_{i+1}c_ib_{i+1}\beta_{i+1};$$

$$w_{i+1} = \eta + \frac{c_{i+1}}{s_{i+2}}\left[s_{i+2}s_{i+1}c_ib'_{i+1}(1+\beta_{i+1}) + s_{i+2}c_{i+1}(d'_{i+1}-\eta)\right] +$$
$$+ s^2_{i+1}b'_{ii} - c_{i+1}s_{i+1}c_ib_{i+1}\beta_{i+1}.$$

Учитывая соотношения (1.5) и принимая во внимание обозначения (1.9) и (1.10), получаем

$$u_i = \bar{d}_i + \alpha_{i+1} + c^2_{i+1}b'_{ii} + c_{i+1}s_{i+1}c_ib^i{}_{+1}\beta_{i+1};$$
$$v_{i+1} = c_{i+2}\bar{b}_{i+1} + c_{i+1}s_{i+1}b'_{ii} + s^2_{i+1}c_ib_{i+1}\beta_{i+1};$$
$$w_{i+1} = t_{i+1} + s^2_{i+1}b'_{ii} - c_{i+1}s_{i+1}c_ib_{i+1}\beta_{i+1}.$$

После введения обозначений

$$\left.\begin{aligned}
&b_{ii} = c^2_{i+1}b'_{ii} + c_{i+1}s_{i+1}c_ib_{i+1}\beta_{i+1};\\
&b'_{i,i+1} \equiv b'_{i+1,i} = c_{i+1}s_{i+1}b'_{ii} + s^2_{i+1}c_ib_{i+1}\beta_{i+1};\\
&b'_{i+1,i+1} = s^2_{i+1}b'_{ii} - c_{i+1}s_{i+1}c_ib_{i+1}\beta_{i+1};\\
&b_{ij} \equiv b_{ji} = c_{i+1}b'_{ij},\\
&b'_{i+1,j} \equiv b'_{j,i+1} = s_{i+1}b'_{ij}
\end{aligned}\right\}\quad j = 1, \ldots, i-1, \tag{1.13}$$

становится понятным, что матрица A_{i+1} записывается так же, как и матрица A_i, если только значение индекса i заменить на $i+1$. Таким образом, предположение, что матрица A_i $(i = 2, \ldots, m-1)$ имеет вид (I), обосновано.

Непосредственное вычисление матрицы A_m по формуле $A_m = C_mA_{m-1}C^*_m$ дает

$$A_m = \left[\begin{array}{c|cc}
\bar{A}_{(m-2)} + X_{(m-1)} + B_{(m-2)} & \begin{matrix} c_mb'_{1,m-1} \\ \vdots \\ c_mb'_{m-3,m-1} \\ z_{m-1} \end{matrix} & \begin{matrix} s_mb'_{1,m-1} \\ \vdots \\ s_mb'_{m-3,m-1} \\ y_{m-1} \end{matrix} \\
\hline
\begin{matrix} c_mb'_{m-1,1} \ldots c_mb'_{m-1,m-3}\ z_{m-1} \\ s_mb'_{m-1,1} \ldots s_mb'_{m-1,m-3}\ y_{m-1} \end{matrix} & \begin{matrix} u_{m-1} \\ v_m \end{matrix} & \begin{matrix} v_m \\ w_m \end{matrix}
\end{array}\right],$$

где для элементов z_{m-1}, y_{m-1}, u_{m-1}, v_m и w_m справедливы формулы (1.12) при $i = m-1$. Легко видеть, что $z_{m-1} = \bar{b}_{m-1} + c_mb'_{m-2,m-1}$, $y_{m-1} = s_mb'_{m-2,m-1}$. Выражения для u_{m-1}, v_m и w_m после простых преобразований могут быть записаны в виде

$$u_{m-1} = d'_m - \frac{c_mc_{m-1}b_m}{s_m} - c_m\left[s_mc_{m-1}b_m(1+\beta_m) + c_m(d'_m-\eta)\right] +$$
$$+ c^2_mb'_{m-1,m-1} + c_ms_mc_{m-1}b_m\beta_m;$$

$$v_m = -s_m\left[s_m c_{m-1} b_m(1+\beta_m) + c_m(d'_m - \eta)\right] + c_m s_m b'_{m-1,m-1} + s_m^2 c_{m-1} b_m \beta_m;$$
$$w_m = \eta + c_m\left[s_m c_{m-1} b_m(1+\beta_m) + c_m(d'_m - \eta)\right] + s_m^2 b'_{m-1,m-1} - c_m s_m c_{m-1} b_m \beta_m.$$

Учет последнего из соотношений (1.5), а также обозначений (1.9) и (1.10) позволяет заключить, что

$$u_{m-1} = \bar{d}_{m-1} + \alpha_m + c_m^2 b'_{m-1,m-1} + c_m s_m c_{m-1} b_m \beta_m;$$
$$v_m = c_m s_m b'_{m-1,m-1} + s_m^2 c_{m-1} b_m \beta_m;$$
$$w_m = \eta + s_m^2 b'_{m-1,m-1} - c_m s_m c_{m-1} b_m \beta_m.$$

Введем обозначения

$$\left.\begin{aligned} &\mathrm{b}_{m-1,m-1} = c_m^2 b'_{m-1,m-1} + c_m s_m c_{m-1} b_m \beta_m;\\ &\mathrm{b}'_{m-1,m} \equiv \mathrm{b}'_{m,m-1} = c_m s_m \mathrm{b}'_{m-1,m-1} + s_m^2 c_{m-1} b_m \beta_m;\\ &\mathrm{b}'_{mm} = s_m^2 \mathrm{b}'_{m-1,m-1} - c_m s_m c_{m-1} b_m \beta_m;\\ &\mathrm{b}_{m-1,j} \equiv \mathrm{b}_{j,m-1} = c_m \mathrm{b}'_{m-1,j},\\ &\mathrm{b}'_{mj} \equiv \mathrm{b}'_{jm} = s_m \mathrm{b}'_{m-1,j}, \end{aligned}\right\} \quad j = 1, \ldots, m-2,$$

которые, очевидно, совпадают с обозначениями (1.13) при $i = m-1$, а также положим

$$\left.\begin{aligned} &\mathrm{b}_{m-1,m} \equiv \mathrm{b}_{m,m-1} = \mathrm{b}'_{m-1,m};\\ &\mathrm{b}_{mm} = \mathrm{b}'_{mm};\\ &\mathrm{b}_{mj} \equiv \mathrm{b}_{jm} = \mathrm{b}'_{mj}, \quad j = 1, \ldots, m-2. \end{aligned}\right\} \tag{1.14}$$

Ясно, что матрицу A_m можно теперь записать в виде

$$A_m = \bar{A} + X + B, \tag{1.15}$$

где $X = \mathrm{diag}[\alpha_2, \alpha_3, \ldots, \alpha_m, 0]$, а элементы b_{ij} $(i, j = 1, 2, \ldots, m)$ матрицы B вычисляются в соответствии с рекуррентными формулами (1.11), (1.13) (при $i = 2, \ldots, m-1$) и (1.14). С другой стороны, в силу обозначений (1.10) понятно, что матрицу $A_m \equiv CA'C^*$ можно представить как

$$A_m = CAC^* + CYC^*, \tag{1.16}$$

если обозначить $Y = \mathrm{diag}[\alpha_1, \alpha_2, \ldots, \alpha_m]$. Сравнивая между собой (1.15) и (1.16), получим $CAC^* = \bar{A} + X - CYC^* + B$. Введем обозначение $R = X - CYC^* + B$. Так как

$$\|R\| \leqslant \|X\| + \|Y\| + \|B\|, \tag{1.17}$$

для оценки нормы матрицы R достаточно оценить нормы матриц X, Y и B. Очевидно, что

$$\left.\begin{aligned} &\|X\| = \max_{2 \leqslant i \leqslant m} |\alpha_i| \leqslant \tilde{\alpha}\mathcal{M}(A);\\ &\|Y\| = \max_{1 \leqslant i \leqslant m} |\alpha_i| \leqslant \tilde{\alpha}\mathcal{M}(A). \end{aligned}\right\} \tag{1.18}$$

Для оценки нормы матрицы B введем в рассмотрение вспомогательную матрицу

$$D = \begin{bmatrix} 0 & -b_2\beta_2 & & & \\ & 0 & -b_3\beta_3 & & 0 \\ & & \ddots & \ddots & \\ & & & \ddots & \\ 0 & & & 0 & -b_m\beta_m \\ & & & & 0 \end{bmatrix}$$

и построим матрицу CDC^*. Принимая во внимание определение матрицы C и выписывая для элементов матрицы CDC^* рекуррентные формулы аналогично тому, как это было сделано для элементов матрицы B, нетрудно убедиться, что элементы последней, стоящие на главной диагонали и выше ее, совпадают с соответствующими элементами матрицы CDC^*. Кроме того, в силу симметричности матрицы B ясно, что ее элементы, стоящие ниже главной диагонали, совпадают с соответствующими элементами матрицы CD^*C^*. Представим матрицу B в виде суммы B_1 и B_2, где матрица B_1 — верхняя треугольная, содержащая на главной диагонали и выше ее соответствующие элементы матрицы B, а B_2 — нижняя треугольная с нулевой главной диагональю, содержащая ниже ее соответствующие элементы матрицы B. Очевидно, что для фробениусовых норм матриц B_1 и B_2 имеют место оценки $\|B_1\|_E \leqslant \|CDC^*\|_E = \|D\|_E$, $\|B_2\|_E \leqslant \|CD^*C^*\|_E = \|D\|_E$ и, следовательно, $\|B\| \leqslant \|B\| \leqslant \|B_1\|_E \leqslant \|B_2\|_E \leqslant 2\|D\|_E$. Так как

$$\|D\|_E = \sqrt{\sum_{i=2}^{m} (b_i\beta_i)^2} \leqslant \max_{2\leqslant i\leqslant m} |\beta_i| \sqrt{\sum_{i=2}^{m} b_i^2} \leqslant$$

$$\leqslant \tilde{\beta} \sqrt{\sum_{i=2}^{m} b_i^2} \leqslant \sqrt{m}\,\tilde{\beta} \max_{2\leqslant i\leqslant m} |\bar{b}_i| \leqslant \sqrt{m}\,\tilde{\beta}\mathcal{M}(A),$$

оценка нормы матрицы B принимает вид $\|B\| \leqslant 2\sqrt{m}\tilde{\beta}\mathcal{M}(A)$. С помощью полученной оценки и оценки (1.18) из (1.17) следует окончательная оценка

$$\|R\| \leqslant 2(\tilde{\alpha} + \sqrt{m}\tilde{\beta})\mathcal{M}(A).$$

Теорема доказана.

§ 2. Алгоритм исчерпывания двухдиагональной матрицы

Рассмотрим двухдиагональную матрицу

$$A = \begin{bmatrix} a_1 & b_2 & & & \\ & a_2 & b_3 & & 0 \\ & & \ddots & \ddots & \\ & & & a_{N-1} & b_N \\ 0 & & & & a_N \end{bmatrix}, \tag{2.1}$$

относительно которой предположим, что $a_i \neq 0$, $b_i \neq 0$. Пусть $\sigma > 0$ — некоторое достаточно хорошее приближение к наибольшему сингулярному числу $\sigma_N(A) = \|A\|$ матрицы A, т. е.

$$|\sigma - \|A\|| \leqslant \delta\|A\|, \quad \delta \ll 1, \tag{2.2}$$

и числа c_i, s_i, $\bar{c}_i$, $\bar{s}_i$ удовлетворяют соотношениям

$$\left.\begin{aligned} c_i^2 + s_i^2 &= 1, \\ \bar{c}_i^2 + \bar{s}_i^2 &= 1, \end{aligned}\right\} \quad i = 2, \ldots, N, \tag{2.3}$$

$$s_i > 0, \quad \bar{s}_i > 0, \quad i = 2, \ldots, N, \tag{2.4}$$

$$s_2 a_1 (1 + \hat{\alpha}_1) - s_2 \frac{\sigma^2}{a_1} + c_2 b_2 (1 + \check{\beta}_2) = 0;$$

$$\bar{s}_2 b_2 (1 + \hat{\beta}_2) - c_2 \frac{\bar{s}_2}{s_2} a_1 + \bar{c}_2 a_2 (1 + \check{\alpha}_2) = 0;$$

$$\left.\begin{aligned} s_{i+1}c_i a_i(1+\hat{\alpha}_i) - \bar{c}_i \frac{s_{i+1}\cdot s_i \cdots s_2}{\bar{s}_i \cdots \bar{s}_2}\frac{\sigma^2}{a_1} + c_{i+1}b_{i+1}(1+\check{\beta}_{i+1}) = 0, \\ \bar{s}_{i+1}\bar{c}_i b_{i+1}(1+\hat{\beta}_{i+1}) - c_{i+1}\frac{\bar{s}_{i+1}\cdots \bar{s}_2}{s_{i+1}\cdots s_2}a_1 + \bar{c}_{i+1}a_{i+1}(1+\check{\alpha}_{i+1}) = 0, \end{aligned}\right\} i=2,\ldots,N-1; \tag{2.5}$$

$$c_N a_N(1+\hat{\alpha}_N) - \bar{c}_N \frac{s_N \cdots s_2}{\bar{s}_N \cdots \bar{s}_2}\frac{\sigma^2}{a_1} = 0,$$

где возмущения $\hat{\alpha}_i$, $\check{\alpha}_i$, $\hat{\beta}_i$, $\check{\beta}_i$ удовлетворяют оценкам

$$|\hat{\alpha}_i| \leqslant \varepsilon, \quad |\check{\alpha}_i| \leqslant \varepsilon, \quad |\hat{\beta}_i| \leqslant \varepsilon, \quad |\check{\beta}_i| \leqslant \varepsilon. \tag{2.6}$$

Отметим, что получение соотношений (2.5) подробно описано в § 3 главы 2. Там же приведено значение величины ε, с которым удовлетворяются оценки (2.6) [см. (3.20)].

Определим ортогональные матрицы C_i, $\bar{C}_i$ $(i=2,\ldots,N)$ и C_{N+1}:

$$C_i = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c_i & -s_i & & \\ & & & s_i & c_i & & \\ & & & & & 1 & \\ & & & & & & \ddots \\ & & & & & & & 1 \end{bmatrix}, \quad \bar{C}_i = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \bar{c}_i & -\bar{s}_i & & \\ & & & \bar{s}_i & \bar{c}_i & & \\ & & & & & 1 & \\ & & & & & & \ddots \\ & & & & & & & 1 \end{bmatrix},$$

$$C_{N+1} = \operatorname{diag}[1, 1, \ldots, 1, \operatorname{sign} a_1],$$

а также их произведения $C = C_{N+1}\cdot C_N \cdots C_3 \cdot C_2$, $\bar{C} = \bar{C}_N \cdots \bar{C}_3 \cdot \bar{C}_2$.

Теорема 2. *Матрица $\bar{C}AC^*$, ортогонально эквивалентная матрице A, может быть представлена в виде суммы двух матриц $\bar{A}$ и R:*

$$\bar{C}AC^* = \bar{A} + R, \tag{2.7}$$

где матрица $\bar{A}$ имеет вид

$$\bar{A} = \begin{bmatrix} \bar{a}_1 & \bar{b}_2 & & & & 0 \\ & \bar{a}_2 & \bar{b}_3 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & & \\ & & 0 & & \bar{a}_{N-2} & \bar{b}_{N-1} \\ & & & & & \bar{a}_{N-1} \quad 0 \end{bmatrix}, \tag{2.8}$$

а для спектральной нормы матрицы R имеет место оценка

$$\|R\| \leqslant 2\sqrt{2}(\sqrt{N}+2)\varepsilon\|A\|. \tag{2.9}$$

При этом элементы матрицы $\bar{A}$ можно вычислить по формулам

$$\begin{aligned} \bar{a}_i &= \frac{s_{i+1}a_{i+1}}{\bar{s}_{i+1}}, \quad i=1,\ldots,N-1; \\ \bar{b}_i &= \frac{\bar{s}_i b_{i+1}}{s_{i+1}}, \quad i=2,\ldots,N-1. \end{aligned} \tag{2.10}$$

Доказательство. Для доказательства потребуются равенства

$$s_2a_1 + c_2b_2 = s_2\frac{\sigma^2}{a_1} + \bar{\varphi}_1;$$

$$\bar{s}_2s_2\frac{\sigma^2}{a_1} + \bar{c}_2c_2a_2 = \frac{\bar{s}_2}{s_2}a_1 + \varphi_2;$$

$$\left.\begin{aligned} s_{i+1}\frac{s_i \dots s_2}{\bar{s}_i \dots \bar{s}_2}a_1 + c_{i+1}\bar{c}_ib_{i+1} &= \frac{s_{i+1}s_i \dots s_2}{\bar{s}_i \dots \bar{s}_2}\frac{\sigma^2}{a_1} + \bar{\varphi}_i, \\ \bar{s}_{i+1}\frac{s_{i+1}s_i \dots s_2}{\bar{s}_i \dots \bar{s}_2}\frac{\sigma^2}{a_1} + \bar{c}_{i+1}c_{i+1}a_{i+1} &= \frac{\bar{s}_{i+1} \dots \bar{s}_2}{s_{i+1} \dots s_2}a_1 + \varphi_{i+1}, \end{aligned}\right\} \quad i = 2, \dots, N-1, \tag{2.11}$$

которые достаточно просто следуют из соотношений (2.3) и (2.5), если для величин φ_i, $\bar{\varphi}_i$ справедливы следующие рекуррентные формулы:

$$\bar{\varphi}_1 = -s_2a_1\hat{\alpha}_1 - c_2b_2\check{\beta}_2; \tag{2.12}$$

$$\varphi_2 = -\bar{s}_2\bar{\varphi}_1 - c_2\left(\bar{s}_2b_2\hat{\beta}_2 + \bar{c}_2a_2\check{\alpha}_2\right);$$

$$\left.\begin{aligned} \bar{\varphi}_i &= -s_{i+1}\varphi_i - \bar{c}_i\left(s_{i+1}c_ia_i\hat{\alpha}_i + c_{i+1}b_{i+1}\check{\beta}_{i+1}\right), \\ \varphi_{i+1} &= -\bar{s}_{i+1}\bar{\varphi}_i - c_{i+1}\left(\bar{s}_{i+1}\bar{c}_ib_{i+1}\hat{\beta}_{i+1} + \bar{c}_{i+1}a_{i+1}\check{\alpha}_{i+1}\right), \end{aligned}\right\} \quad i = 2, \dots, N-1.$$

Действительно, первое из этих равенств просто совпадает с первым из соотношений (2.5), если положить $\bar{\varphi}_1 = -s_2a_1\hat{\alpha}_1 - c_2b_2\check{\beta}_2$. Используя теперь первое равенство (2.11) и второе равенство (2.5), а также (2.3), находим

$$\bar{s}_2s_2\frac{\sigma^2}{a_1} + \bar{c}_2c_2a_2 = \bar{s}_2\left(s_2a_1 + c_2b_2 - \bar{\varphi}_1\right) + \bar{c}_2c_2a_2 =$$

$$= \bar{s}_2s_2a_1 + c_2\left(\bar{s}_2b_2 + \bar{c}_2a_2\right) - \bar{s}_2\bar{\varphi}_1 = s_2^2\frac{\bar{s}_2}{s_2}a_1 + c_2^2\frac{\bar{s}_2}{s_2}a_1 -$$

$$- \bar{s}_2\bar{\varphi}_1 - c_2\left(\bar{s}_2b_2\hat{\beta}_2 + \bar{c}_2a_2\check{\alpha}_2\right) = \frac{\bar{s}_2}{s_2}a_1 - \bar{s}_2\bar{\varphi}_1 - c_2\left(\bar{s}_2b_2\hat{\beta}_2 + \bar{c}_2a_2\check{\alpha}_2\right).$$

Вводя обозначение $\varphi_2 = -\bar{s}_2\bar{\varphi}_1 - c_2\left(\bar{s}_2b_2\hat{\beta}_2 + \bar{c}_2a_2\check{\alpha}_2\right)$, будем иметь второе из равенств (2.11)

$$\bar{s}_2s_2\frac{\sigma^2}{a_1} + \bar{c}_2c_2a_2 = \frac{\bar{s}_2}{s_2}a_1 + \varphi_2.$$

Аналогичным образом получаются все остальные равенства (2.11), в которых использованы обозначения (2.12).

Переходя непосредственно к доказательству теоремы, введем обозначения

$$A_1 = A, \quad A_i' = A_{i-1}C_i^*, \quad A_i = \bar{C}_iA_i', \quad i = 2, \dots, N,$$

из которых следует, что $\bar{C}AC^* = A_NC_{N+1}^*$. Матрица $A_2' = A_1C_2^*$, очевидно, имеет вид

$$A_2' = \begin{bmatrix} c_2a_1 - s_2b_2 & s_2a_1 + c_2b_2 & & & & \\ -s_2a_2 & c_2a_2 & b_3 & & & \\ & & a_3 & b_4 & & 0 \\ & & & \ddots & \ddots & \\ & 0 & & & \ddots & \\ & & & & a_{N-1} & b_N \\ & & & & & a_N \end{bmatrix}.$$

Используя соотношения (2.5) и (2.11), а также обозначения (2.10), преобразуем выражения для элементов матрицы A_2':

$$c_2a_1 - s_2b_2 = \frac{s_2}{\bar{s}_2}\left[c_2\frac{\bar{s}_2}{s_2}a_1 - \bar{s}_2\bar{b}_2(1+\hat{\beta}_2)\right] + s_2b_2\hat{\beta}_2 = \bar{c}_2\bar{a}_1(1+\check{\alpha}_1) + s_2b_2\hat{\beta}_2,$$

$$s_2a_1 + c_2b_2 = s_2\frac{\sigma^2}{a_1} + \bar{\varphi}_1.$$

Введем обозначения

$$\mathrm{b}_{11}' = s_2b_2\hat{\beta}_2, \quad \mathrm{b}_{12}'' = \bar{\varphi}_1. \tag{2.13}$$

Матрица A_2' принимает теперь вид

$$A_2' = \begin{bmatrix} \bar{c}_2\bar{a}_1(1+\check{\alpha}_2) + \mathrm{b}_{11}' & s_2\dfrac{\sigma^2}{a_1} + \mathrm{b}_{12}'' & & \\ -s_2a_2 & c_2a_2 & b_3 & 0 \\ & & a_3 & b_4 \\ & 0 & & \ddots \\ & & & a_{N-1} \quad b_N \\ & & & a_N \end{bmatrix}.$$

По индукции легко показать, что матрицы A_i' $(i=3, \ldots, N)$ и A_i $(i=2, \ldots, N-1)$ выглядят следующим образом:

$$A_i' = \left[\begin{array}{c|cc} A_{(i-1)}' + X_{(i)}' + B_{(i-1)}' & \begin{matrix}\mathrm{b}_{1i}' \\ \mathrm{b}_{2i}' \\ \vdots \\ \mathrm{b}_{i-2,i}'\end{matrix} & 0 \\ \hline \mathrm{b}_{i-1,1}' \ldots \mathrm{b}_{i-1,i-2}'\ t_i + \mathrm{b}_{i-1,i-1}'\ r_i + \mathrm{b}_{i-1,i}'' & & \\ 0 \qquad -s_ia_i \qquad c_ia_i & & \widetilde{A}_{(i+1)}' \end{array}\right]; \tag{II}$$

$$A_i = \left[\begin{array}{c|cc} \bar{A}_{(i-1)} + X_{(i)} + B_{(i-1)} & \begin{matrix}\mathrm{b}_{1i}' \\ \vdots \\ \mathrm{b}_{i-2,i}'\end{matrix} & 0 \\ \hline \mathrm{b}_{i1}' \ldots \mathrm{b}_{i,i-2}'\ \mathrm{b}_{i,i-1}' & x_i + \mathrm{b}_{i-1,i}' \quad -\bar{s}_ib_{i+1} & \\ 0 & y_i + \mathrm{b}_{ii}'' \quad \bar{c}_ib_{i+1} & \widetilde{A}_{(i+1)} \end{array}\right], \tag{III}$$

где приняты обозначения

$$\bar{A}_{(i-1)}' = \begin{bmatrix} \bar{a}_1 & \bar{b}_2 & & \\ & a_2 & \bar{b}_3 & 0 \\ 0 & & \ddots & \\ & & \bar{a}_{i-2} & \bar{b}_{i-1} \end{bmatrix}, \quad \widetilde{A}_{(i+1)} = \begin{bmatrix} a_{i+1} & b_{i+2} & & \\ & \ddots & \ddots & 0 \\ 0 & & a_{N-1} & b_N \\ & & & a_N \end{bmatrix};$$

$$X_{(i)}' = \begin{bmatrix} \bar{a}_1\check{\alpha}_2 & \bar{b}_2\check{\beta}_3 & & \\ & \bar{a}_2\check{\alpha}_3 & \bar{b}_3\check{\beta}_4 & 0 \\ 0 & & \ddots & \\ & & \bar{a}_{i-2}\check{\alpha}_{i-1} & \bar{b}_{i-1}\check{\beta}_i \end{bmatrix};$$

$$B'_{(i-1)} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1,i-2} & b_{1,i-1} \\ b_{21} & b_{22} & \dots & b_{2,i-2} & b_{2,i-1} \\ \dots & \dots & \dots & \dots & \dots \\ b_{i-2,1} & b_{i-2,2} & \dots & b_{i-2,i-2} & b_{i-2,i-1} \end{bmatrix};$$

$$\bar{A}_{(i-1)} = \left[\begin{array}{c} \bar{A}'_{(i-1)} \\ \hline 0 \dots 0\, \bar{a}_{i-1} \end{array}\right], \quad \tilde{A}'_{(i+1)} = \left[\begin{array}{c} b_{i+1}\, 0 \dots 0 \\ \hline \tilde{A}_{(i+1)} \end{array}\right];$$

$$X_{(i)} = \left[\begin{array}{c} X'_{(i)} \\ \hline 0 \dots 0\, \bar{a}_{i-1}\breve{\alpha}_i \end{array}\right], \quad B_{(i-1)} = \left[\begin{array}{c} B'_{(i-1)} \\ \hline b_{i-1,1} b_{i-1,2} \dots b_{i-1,i-1} \end{array}\right],$$

$$t_i = \bar{c}_i \bar{a}_{i-1}(1 + \breve{\alpha}_i); \qquad r_i = \frac{s_i \cdot s_{i-1} \dots s_2}{\bar{s}_{i-1} \dots \bar{s}_2} \frac{\sigma^2}{a_1};$$

$$x_i = c_{i+1} \bar{b}_i (1 + \breve{\beta}_{i+1}); \quad y_i = \frac{\bar{s}_i \dots \bar{s}_2}{s_i \dots s_2} a_1.$$

Продемонстрируем только переход от матрицы A'_i к матрице A_i. Вычисления по формуле $A_i = \bar{C}_i A'_i$ дают

$$A_i = \left[\begin{array}{ccc|c} & & & b'_{1i} \\ & \bar{A}'_{(i-1)} + X'_{(i)} + B'_{(i-1)} & & b'_{2i} \\ & & & \vdots \\ & & & b'_{i-2,i} \\ \hline \bar{c}_i b'_{i-1,1} & \dots & \bar{c}_i b'_{i-1,i-2}\ u_{i-1} & v_i \\ \bar{s}_i b'_{i-1,1} & \dots & \bar{s}_i b'_{i-1,i-2}\ z_i & w_i \\ & 0 & & \end{array} \right. \left. \begin{array}{c} 0 \\ \\ \\ \\ -\bar{s}_i b_{i+1} \\ \bar{c}_i b_{i+1} \\ \hline \tilde{A}_{(i+1)} \end{array} \right],$$

где элементы u_{i-1}, z_i, v_i, w_i имеют следующие выражения:

$$\begin{aligned} u_{i-1} &= \bar{c}_i (t_i + b'_{i-1,i-1}) + \bar{s}_i s_i a_i; \\ z_i &= \bar{s} (t_i + b'_{i-1,i-1}) - \bar{c}_i s_i a_i; \\ v_i &= \bar{c}_i (r_i + b''_{i-1,i}) - \bar{s}_i c_i a_i; \\ w_i &= \bar{s}_i (r_i + b''_{i-1,i}) + \bar{c}_i c_i a_i. \end{aligned} \tag{2.14}$$

Преобразуем эти выражения. В силу определения $\bar{a}_{i-1}$ и обозначения для t_i легко видеть, что $u_{i-1} = \bar{a}_{i-1}(1 + \breve{\alpha}_i) + \bar{c}_i b'_{i-1,i-1} - \bar{s}_i s_i a_i \breve{\alpha}_i$, $z_i = \bar{s}_i b'_{i-1,i-1} + \bar{c}_i s_i a_i \breve{\alpha}_i$. Выражение для v_i с учетом обозначения для r_i преобразуем к виду

$$v_i = \frac{\bar{s}_i}{s_{i+1}} \left[\bar{c}_i \frac{s_{i+1} s_i \dots s_2}{\bar{s}_i \dots \bar{s}_2} \frac{\sigma^2}{a_1} - s_{i+1} c_i a_i (1 + \hat{\alpha}_i) \right] + \bar{c}_i b''_{i-1,i} + \bar{s}_i c_i a_i \hat{\alpha}_i,$$

откуда, принимая во внимание (2.5), определение $\bar{b}_{i-1}$ и обозначение для x_i, получим $v_i = x_i + \bar{c}_i b''_{i-1,i} + s_i c_i a_i \hat{\alpha}_i$. Наконец, из (2.11) и обозначения для y_i следует, что $w_i = y_i + \bar{s}_i b''_{i-1,i} + \varphi_i$. После введения обозначений

$$\begin{aligned} &b_{i-1,i-1} = \bar{c}_i b'_{i-1,i-1} - \bar{s}_i s_i a_i \breve{\alpha}_i, \ b'_{i,i-1} = \bar{s}_i b'_{i-1,i-1} + \bar{c}_i s_i \breve{\alpha}_i; \\ &b'_{i-1,i} = \bar{c}_i b''_{i-1,i} + \bar{s}_i c_i a_i \hat{\alpha}_i, \ b''_{ii} = \bar{s}_i b''_{i-1,i} + \varphi_i; \\ &\left.\begin{aligned} b_{i-1,j} &= \bar{c}_i b'_{i-1,j}, \\ b'_{ij} &= \bar{s}_i b'_{i-1,j}, \end{aligned}\right\} \quad j = 1, \dots, i-2, \end{aligned} \tag{2.15}$$

становится понятным, что матрица A_i действительно принимает вид (III).

Аналогичным образом выполняется переход от матрицы A_i к матрице A'_{i+1} по формуле $A'_{i+1} = A_i C^*_{i+1}$. Нетрудно проверить, что A'_{i+1} записывается в виде (II), если только значение индекса i заменено на $i+1$. При этом должны быть сделаны следующие обозначения:

$$\left.\begin{aligned} &b_{i-1,i} = c_{i+1} b'_{i-1,i} - s_{i+1}\bar{s}_i b_{i+1}\breve{\beta}_{i+1};\\ &b'_{i-1,i+1} = s_{i+1} b'_{i-1,i} + c_{i+1}\bar{s}_i b_{i+1}\breve{\beta}_{i+1};\\ &b'_{ii} = c_{i+1} b''_{ii} + s_{i+1}\bar{c}_i b_{i+1}\hat{\beta}_{i+1},\ b''_{i,i+1} = s_{i+1} b''_{ii} + \bar{\varphi}_i;\\ &b_{ji} = c_{i+1} b'_{ji},\\ &b_{j,i+1} = s_{i+1} b'_{ji},\end{aligned}\right\}\ j = 1, \ldots, i-2. \qquad (2.16)$$

Рассмотрим еще переход от матрицы A'_N к матрице A_N по формуле $A_N = \bar{C}_N A'_N$. Непосредственно проверяется, что A_N имеет вид

$$A_N = \left[\begin{array}{c|c} \bar{A}'_{(N-1)} + X'_{(N)} + B'_{(N-1)} & \begin{matrix} b'_{1N} \\ \vdots \\ b'_{N-2,N} \end{matrix} \\ \hline \begin{matrix} \bar{c}_N b'_{N-1,1} \ldots \bar{c}_N b'_{N-1,N-2} & u_{N-1} \\ \bar{s}_N b'_{N-1,1} \ldots \bar{s}_N b'_{N-1,N-2} & z_N \end{matrix} & \begin{matrix} v_N \\ w_N \end{matrix} \end{array}\right],$$

где для элементов u_{N-1}, z_N, v_N и w_N справедливы формулы (2.14) при $i = N$. Из определения $\bar{a}_{N-1}$ легко заключить, что $u_{N-1} = \bar{a}_{N-1}(1+\breve{\alpha}_N) + \bar{c}_N b'_{N-1,N-1} - \bar{s}_N s_N a_N \breve{\alpha}_N$, $z_N = \bar{s}_N b'_{N-1,N-1} + \bar{c}_N s_N a_N \breve{\alpha}_N$. Выражение для v_N с учетом последнего из соотношений (2.5) преобразуется в $v_N = \bar{c}_N b''_{N-1,N} + \bar{s}_N c_N a_N \hat{\alpha}_N$. Использование того же соотношения позволяет получить выражение для w_N:

$$w_N = \frac{\bar{s}_N \ldots \bar{s}_2}{s_N \ldots s_2}\frac{\sigma^2}{a_1} + \bar{s}_N b''_{N-1,N} - \bar{c}_N c_N a_N \hat{\alpha}_N.$$

С другой стороны, из последнего из соотношений (2.11) непосредственно вытекает, что

$$w_N = \frac{\bar{s}_N \ldots \bar{s}_2}{s_N \ldots s_2} a_1 + \bar{s}_N b''_{N-1,N} + \varphi_N.$$

Вводя обозначение

$$p = \frac{\bar{s}_N \ldots \bar{s}_2}{s_N \ldots s_2} a_1 \qquad (2.17)$$

и сравнивая два различных выражения для w_N, получим уравнение для определения величины p:

$$(\sigma^2 - p^2)/p = \varphi_N + \bar{c}_N c_N a_N \hat{\alpha}_N, \qquad (2.18)$$

которое исследуем далее.

Воспользуемся обозначениями (2.15) при значении индекса $i = N$ и введем новые обозначения

$$\left.\begin{aligned} &b_{Nj} = b'_{Nj},\\ &b_{jN} = b'_{jN},\end{aligned}\right\}\ j = 1, \ldots, N-1,$$
$$b_{NN} = b''_{NN}. \qquad (2.19)$$

Выражения для элементов u_{N-1}, z_N, v_N и w_N теперь можно переписать следующим образом: $u_{N-1} = \bar{a}_{N-1}(1+\breve{\alpha}_N) + b_{N-1,N-1}$, $z_N = b_{N,N-1}$, $v_N =$

$= b_{N-1,N}$, $w_N = p + b_{NN}$. Легко понять, что матрица A_N принимает вид

$$A_N = \left[\begin{array}{c|c} \overline{A}_{(N-1)} + X_{(N)} + B_{(N-1)} & \begin{matrix} b_{1N} \\ \vdots \\ b_{N-1,N} \end{matrix} \\ \hline b_{N1} \ldots b_{N,N-1} & p + b_{NN} \end{array}\right].$$

Отсюда следует, что окончательную матрицу $\overline{C}AC^* \equiv A_N C^*_{N+1}$ можно представить в виде суммы $\overline{C}AC^* = \overline{A} + X + BC^*_{N+1} + Y$, где $\overline{A}$ имеет требуемый вид (2.8)

$$X = \left[\begin{array}{c|c} X_{(N)} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline 0 \ldots 0 \; 0 & \end{array}\right]; \quad B = \left[\begin{array}{c|c} B_{(N-1)} & \begin{matrix} b_{1N} \\ \vdots \\ b_{N-1,N} \end{matrix} \\ \hline b_{N1} \ldots b_{N,N-1} \; b_{NN} & \end{array}\right],$$

а единственный ненулевой элемент матрицы Y расположен в ее нижнем правом углу и равен $p \operatorname{sign} a_1 - \sigma$. Обозначив $R = X + BC^*_{N+1} + Y$, оценим норму матрицы R. В силу неравенства

$$\|R\| \leqslant \|X\| + \|B\| + \|Y\| \tag{2.20}$$

достаточно найти оценки норм матриц X, B и Y.

Сначала оценим норму матрицы B, предварительно напомнив, что ее элементы b_{ij} $(i, j = 1, \ldots, N)$ вычисляются в соответствии с рекуррентными формулами (2.13), (2.15) при $i = 2, \ldots, N$, (2.16) при $i = 2, \ldots, N-1$ и (2.19). Введем в рассмотрение две вспомогательные двухдиагональные матрицы

$$D_1 = \begin{bmatrix} -a_1\widehat{\alpha}_1 & -d_2\breve{\beta}_2 & & & \\ & -a_2\widehat{\alpha}_2 & -b_3\breve{\beta}_3 & & 0 \\ & & \cdot & \cdot & \\ & & & \cdot & \cdot \\ & 0 & & -a_{N-1}\widehat{\alpha}_{N-1} & -b_N\breve{\beta}_N \\ & & & & -a_N\widehat{\alpha}_N \end{bmatrix},$$

$$D_2 = \begin{bmatrix} 0 & -b_2\widehat{\beta}_2 & & & \\ & -a_2\breve{\alpha}_2 & -b_3\widehat{\beta}_3 & & 0 \\ & & \cdot & \cdot & \\ & & & \cdot & \cdot \\ & 0 & & -a_{N-1}\breve{\alpha}_{N-1} & -b_N\widehat{\beta}_N \\ & & & & -a_N\breve{\alpha}_N \end{bmatrix}$$

и положим

$$E = \overline{C}_N \cdots \overline{C}_3\overline{C}_2 D_1 C_2 C_3 \cdots C_N, \quad F = \overline{C}_N \cdots \overline{C}_3\overline{C}_2 D_2 C_2 C_3 \cdots C_N. \tag{2.21}$$

Если для элементов матриц E и F выписать рекуррентные формулы аналогично тому, как это сделано для элементов матрицы B, и сравнить получающиеся формулы с (2.13), (2.15), (2.16) и (2.19), то нетрудно убедиться, что имеют место равенства

$$b_{ij} = \begin{cases} e_{ij}, & \text{если } j > i, \\ f_{ij}, & \text{если } j \leqslant i. \end{cases} \tag{2.22}$$

Здесь e_{ij} и f_{ij} — элементы матриц E и F соответственно. Разобьем теперь матрицу B на сумму двух треугольных матриц B_1 и B_2, отнеся в верхнюю треугольную B_1, имеющую нулевую главную диагональ, элементы

матрицы B, расположенные выше ее главной диагонали, а в нижнюю треугольную B_2 — остальные элементы матрицы B. Ясно, что $\|B\| \leqslant \|B_1\| + \|B_2\|$. В силу определения матриц B_1 и B_2 и равенств (2.22) для фробениусовых норм этих матриц имеют место оценки $\|B_1\|_E \leqslant \|E\|_E$, $\|B_2\|_E \leqslant \|F\|_E$. Отсюда, учитывая определение (2.21) матриц E и F и ортогональность матриц C_i, $\overline{C}_i$, а также связь между фробениусовой и спектральной нормами, находим $\|B_1\| \leqslant \|B_1\|_E \leqslant \|E\|_E \leqslant \sqrt{N}\|E\| = \sqrt{N}\|D_1\|$ и аналогично $\|B_2\| \leqslant \sqrt{N}\|D_2\|$. Получим еще оценки норм матриц D_1 и D_2 через норму исходной матрицы A. Для нормы двухдиагональной матрицы D_1 имеет место оценка (см., например, (2.30) из [4]):

$$\|D_1\| \leqslant \max \begin{cases} \max\limits_{1 \leqslant i \leqslant N-1} (|a_i \widehat{\alpha}_i| + |b_{i+1} \breve{\beta}_{i+1}|), \\ \max\limits_{2 \leqslant i \leqslant N} (a_i \widehat{\alpha}_i| + |b_i \breve{\beta}_i|), \end{cases}$$

из которой в силу неравенств (2.6) следует

$$\|D_1\| \leqslant \varepsilon \max \begin{cases} \max\limits_{1 \leqslant i \leqslant N-1} (|a_i| + |b_{i+1}|), \\ \max\limits_{2 \leqslant i \leqslant N} (|a_i| + |b_i|). \end{cases} \tag{2.23}$$

Далее, известно (см., например, [5], § 3), что для нормы двухдиагональной матрицы A справедливы оценки снизу $\sqrt{a_i^2 + b_i^2} \leqslant \|A\|$, $\sqrt{a_i^2 + b_{i+1}^2} \leqslant \|A\|$, из которых очевидным образом следуют неравенства

$$|a_i| + |b_i| \leqslant \sqrt{2}\|A\|, \quad |a_i| + |b_{i+1}| \leqslant \sqrt{2}\|A\|. \tag{2.24}$$

Огрубляя с помощью этих неравенств оценку (2.23), получим $\|D_1\| \leqslant \sqrt{2}\varepsilon\|A\|$. Аналогичным образом запишем оценку нормы матрицы D_2: $\|D_2\| \leqslant \sqrt{2}\varepsilon\|A\|$. В результате окончательная оценка нормы матрицы B принимает вид $\|B\| \leqslant 2\sqrt{2}\sqrt{N}\varepsilon\|A\|$.

Для оценки нормы матрицы X воспользуемся упомянутой оценкой (2.30) из [4]:

$$\|X\| \leqslant \max \begin{cases} \max\limits_{2 \leqslant i \leqslant N-1} (|\overline{a}_{i-1} \breve{\alpha}_i| + |\overline{b}_i \breve{\beta}_{i+1}|), \\ \max\limits_{2 \leqslant i \leqslant N-1} (|\overline{a}_i \breve{\alpha}_{i+1}| + |\overline{b}_i \breve{\beta}_{i+1}|) \end{cases} \leqslant \sqrt{2}\varepsilon\|\overline{A}\|.$$

Для оценки нормы матрицы Y потребуется некоторое представление элемента p, который, как уже отмечалось, удовлетворяет квадратному уравнению (2.18). Прежде чем приступить к исследованию решений этого уравнения, получим удобное выражение и оценку для его правой части $\varphi_N + \overline{c}_N c_N a_N \widehat{\alpha}_N$. Оказывается, и это нетрудно проверить непосредственными вычислениями, что величина $\varphi_N + \overline{c}_N c_N a_N \widehat{\alpha}_N$ совпадает с элементом матрицы $G = \overline{C}_N \cdots \overline{C}_3 \overline{C}_2 (-D_1 + D_2) C_2 C_3 \cdots C_N$, стоящим в ее нижнем правом углу. Отсюда сразу следует оценка $|\varphi_N + \overline{c}_N c_N a_N \widehat{\alpha}_N| \leqslant \|G\| = \|-D_1 + D_2\| \leqslant 2\sqrt{2}\varepsilon\|A\|$, очевидно, эквивалентная равенству $\varphi_N + \overline{c}_N c_N a_N \widehat{\alpha}_N = \gamma_1\|A\|$, которое выполнено при некотором γ_1 таком, что $|\gamma_1| \leqslant 2\sqrt{2}\varepsilon$. Заметим, что оценку (2.2) также можно записать в виде равенства $\sigma = (1 + \gamma_2)\|A\|$, справедливого при некотором γ_2, удовлетворяющем оценке $|\gamma_2| \leqslant \delta$. Уравнение (2.18) теперь перепишем в виде $(\sigma^2 - p^2)/p = \sigma\gamma_1/(1 + \gamma_2)$ или $(\sigma^2 - p^2)/p = \gamma_3\sigma$, где $|\gamma_3| \leqslant 2\sqrt{2}\varepsilon/(1 - \delta)$. Решая это уравнение, получим $p_{1,2} = \pm\sigma\left[\sqrt{1 + \gamma_3^2/4} \mp \gamma_3/2\right]$, откуда следует, что $p_1 = \sigma(1 + \gamma_4)$, $p_2 = -\sigma(1 + \gamma_5)$, где $|\gamma_4| \leqslant |\gamma_3| \leqslant 2\sqrt{2}\varepsilon/(1 - \sigma)$, $|\gamma_5| \leqslant 2\sqrt{2}\varepsilon/(1 - \delta)$. Из определения (2.17) p и положительности чисел s_i, $\overline{s}_i$ вытекает, что знак корня следует выбирать равным знаку элемента

a_1, вследствие чего

$$p = \begin{cases} \sigma(1+\gamma_4), & \text{если } a_1 > 0, \\ -\sigma(1+\gamma_5), & \text{если } a_1 < 0. \end{cases}$$

Последнее равенство, очевидно, можно записать в виде $p = \sigma \operatorname{sign} a_1 (1+\gamma_6)$, где $|\gamma_6| \leqslant 2\sqrt{2}\varepsilon/(1-\delta)$. Из этого равенства следует, что $p \operatorname{sign} a_1 - \sigma = \sigma\gamma_6$. Помня, что $p \operatorname{sign} a_1 - \sigma$ есть единственный ненулевой элемент матрицы Y, легко получить оценку ее нормы

$$\|Y\| = |p \operatorname{sign} a_1 - \sigma| = \sigma|\gamma_6| \leqslant \frac{2\sqrt{2}\varepsilon(1+\delta)}{1-\delta}\|A\|.$$

Подставляя полученные оценки норм матриц B, X и Y в оценку (2.20), находим

$$\|R\| \leqslant 2\sqrt{2}\sqrt{N}\varepsilon\|A\| + \sqrt{2}\varepsilon\|\bar{A}\| + \frac{2\sqrt{2}\varepsilon(1+\delta)}{1-\delta}\|A\|. \qquad (2.25)$$

С помощью этого неравенства и равенства $R = \bar{C}AC^* - \bar{A}$, следующего из (2.7), можно найти оценку нормы матрицы $\bar{A}$ в терминах нормы исходной матрицы A. Действительно, $\|R\| \geqslant \|\bar{A}\| - \|\bar{C}AC^*\| = \|\bar{A}\| - \|A\|$, следовательно,

$$\|\bar{A}\| \leqslant \frac{1 + 2\sqrt{2}\sqrt{N}\varepsilon + [2\sqrt{2}\varepsilon(1+\delta)]/(1-\delta)}{1-\sqrt{2}\varepsilon}\|A\|.$$

Подстановка полученной оценки в (2.25) дает

$$\|R\| \leqslant \frac{\sqrt{2}[(2\sqrt{N}+1)(1-\delta)+2(1+\delta)]}{(1-\delta)(1-\sqrt{2}\varepsilon)}\varepsilon\|A\|.$$

Предполагая величины ε и δ настолько малыми, что выполнено неравенство $5\delta + 2\sqrt{2}(\sqrt{N}+2)(1-\delta)\varepsilon \leqslant 1$, полученную оценку огрубим следующим образом:

$$\|R\| \leqslant 2\sqrt{2}(\sqrt{N}+2)\varepsilon\|A\|.$$

Теорема доказана.

Глава 2

РЕШЕНИЕ УРАВНЕНИЙ ДЛЯ ПАРАМЕТРОВ, ЗАДАЮЩИХ ОРТОГОНАЛЬНЫЕ ПРЕОБРАЗОВАНИЯ ВРАЩЕНИЯ, И ИХ МАШИННОЕ ВЫЧИСЛЕНИЕ

§ 3. Решение уравнений для параметров, определяющих преобразования вращения

В этом параграфе показано, как найти параметры c_i, s_i, удовлетворяющие соотношениям (1.3)—(1.5), а также параметры c_i, s_i, $\bar{c}_i$, $\bar{s}_i$, удовлетворяющие соотношениям (2.3)—(2.5), т. е. рассмотрены случаи трехдиагональной симметрической и двухдиагональной матриц. Получены оценки возмущений α_i, β_i, с которыми удовлетворяются уравнения (1.5), и возмущений $\hat{\alpha}_i$, α_i, $\hat{\beta}_i$, $\check{\beta}_i$, с которыми выполняются уравнения (2.5).

В случае трехдиагональной матрицы A (1.1) будем предполагать, что известны достаточно хорошее приближение η к собственному значению λ матрицы A и отношения $\bar{\mathcal{P}}_2, \bar{\mathcal{P}}_3, \ldots, \bar{\mathcal{P}}_m$ последовательных компонент почти собственного вектора $\bar{u} = (\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m)^T$ этой матрицы, отвечающего ее почти собственному значению η. Конкретнее, предполо-

жим, что имеют место неравенство

$$|\eta-\lambda|\leqslant\gamma\mathscr{M}(A),\quad \gamma\ll 1, \tag{3.1}$$

и соотношения

$$\begin{gathered}d_1+\bar{\alpha}_1-\eta+b_2\bar{\mathscr{P}}_2=0;\\ \frac{b_i(1+\bar{\beta}_i)}{\bar{\mathscr{P}}_i}+d_i+\bar{\alpha}_i-\eta+b_{i+1}\bar{\mathscr{P}}_{i+1}=0,\ i=2,\ldots,m-1;\\ \frac{b_m(1+\bar{\beta}_m)}{\bar{\mathscr{P}}_m}+d_m+\bar{\alpha}_m-\eta=0,\end{gathered} \tag{3.2}$$

где возмущения $\bar{\alpha}_i$ и $\bar{\beta}_i$ удовлетворяют оценкам

$$\begin{gathered}|\bar{\alpha}_i|\leqslant\bar{\alpha}\mathscr{M}(A),\quad \bar{\alpha}\ll 1;\\ |\bar{\beta}_i|\leqslant\bar{\beta},\quad \bar{\beta}\ll 1;\end{gathered} \tag{3.3}$$

а $\mathscr{M}(A)$ обозначает одну из норм матрицы A:

$$\mathscr{M}(A)=\max\begin{cases}|d_1|+|b_2|,\\ \max\limits_{2<i<m-1}(|d_i|+|b_i|+|b_{i+1}|),\\ |d_m|+|b_m|.\end{cases}$$

Вычисление η — стандартная задача, оно подробно описано, например, в [5]. Возможность же получения чисел $\bar{\mathscr{P}}_2$, $\bar{\mathscr{P}}_3$, ..., $\bar{\mathscr{P}}_m$, удовлетворяющих соотношениям (3.2), является в точности утверждением теоремы 2, доказанной в [4]. Отметим для дальнейшего, что из [4] следует, что числа $\bar{\mathscr{P}}_i$ удовлетворяют неравенствам

$$\frac{\varepsilon_2}{\alpha^2}<\frac{\varepsilon_1^2}{2}\leqslant|\bar{\mathscr{P}}_i|\leqslant\frac{2}{\varepsilon_1^2}<\frac{1}{\varepsilon_2}, \tag{3.4}$$

гарантирующим отличие их от нуля, а также возможность размещения в ячейке ЭВМ ЕС, не вызывая переполнения разрядной сетки. Здесь α — основание системы счисления, принятой в ЭВМ, а ε_1 и ε_2 — характеристики ее разрядной сетки (об их определении см., например, [4], § 4). Определение этих характеристик напомним в следующем параграфе, где они существенно используются при анализе выполнения арифметических операций на вычислительной машине. А пока приведем их значения для ЭВМ ЕС в случае использования чисел с «двойной точностью»: $\varepsilon_1\approx 0{,}22\cdot 10^{-15}$, $1/\varepsilon_2\approx 0{,}72\cdot 10^{76}$. Основание α системы счисления, принятой в этих машинах, равно 16. Напомним также одно обозначение, введенное в § 3 работы [4]. Если a — арифметическое выражение, то под $(a)_{\text{м}}$ понимается результат вычисления значения этого выражения в вычислительной машине.

Введем теперь в рассмотрение последовательности чисел s_i, c_i, c_i', определив их следующим образом:

$$\begin{gathered}s_2=\frac{1}{\sqrt{1+\bar{\mathscr{P}}_2^2}},\ c_2=\frac{\bar{\mathscr{P}}_2}{\sqrt{1+\bar{\mathscr{P}}_2^2}},\qquad c_2'=(c_2)_{\text{м}};\\ s_i=\frac{1}{\sqrt{1+(c_{i-1}'\bar{\mathscr{P}}_i)^2}},\quad c_i=\frac{c_{i-1}'\bar{\mathscr{P}}_i}{\sqrt{1+(c_{i-1}'\bar{\mathscr{P}}_i)^2}},\quad c_i'=(c_i)_{\text{м}},\ i=3,\ldots,m.\end{gathered} \tag{3.5}$$

Предположим, что c_i' можно вычислить так, что

$$c_i'=c_i(1+\varkappa_i),\ |\varkappa_i|\leqslant\varkappa\ll 1. \tag{3.6}$$

Это предположение обосновано в § 4, где приведен алгоритм вычисле-

ния c_i' и анализируются погрешности, возникающие при его реализации на вычислительной машине.

Из (3.5) непосредственно следуют равенства

$$c_2 = s_2\overline{\mathscr{P}}_2,\ c_i = s_i c_{i-1}'\overline{\mathscr{P}}_i, \quad i = 3, \ldots, m, \tag{3.7}$$

которые с учетом (3.6) легко приводят к равенствам

$$\frac{c_2'}{\overline{\mathscr{P}}_2} = s_2(1+\varkappa_2);\ \frac{c_i'}{\overline{\mathscr{P}}_i} = s_i c_{i-1}(1+\varkappa_{i-1})(1+\varkappa_i), \quad i = 3, \ldots, m-1; \tag{3.8}$$

$$\frac{c_m}{\overline{\mathscr{P}}_m} = s_m c_{m-1}(1+\varkappa_{m-1}).$$

Умножая первое из соотношений (3.2) на s_2, i-е — на $s_{i+1}c_i'$, последнее — на c_m и учитывая равенства (3.6) — (3.8), получим

$$s_2(d_1 + \overline{\alpha}_1 - \eta) + c_2 b_2 = 0;$$

$$s_3 s_2 b_2 (1+\overline{\beta}_2)(1+\varkappa_2) + s_3 c_2 (d_2 + \overline{\alpha}_2 - \eta)(1+\varkappa_2) + c_3 b_3 = 0;$$

$$s_{i+1}s_i c_{i-1} b_i (1+\overline{\beta}_i)(1+\varkappa_{i-1})(1+\varkappa_i) + s_{i+1}c_i(d_i + \overline{\alpha}_i - \eta)(1+\varkappa_i) + c_{i+1}b_{i+1} = 0, \quad i = 3, \ldots, m-1;$$

$$s_m c_{m-1} b_m (1+\overline{\beta}_m)(1+\varkappa_{m-1}) + c_m(d_m + \overline{\alpha}_m - \eta) = 0.$$

Вводя обозначения

$$\begin{aligned}
&\alpha_1 = \overline{\alpha}_1; \quad \beta_2 = (1+\overline{\beta}_2)(1+\varkappa_2) - 1;\\
&\alpha_i = (d_i - \eta)\varkappa_i + \overline{\alpha}_i(1+\varkappa_i), \quad i = 2, \ldots, m-1;\\
&\beta_i = (1+\overline{\beta}_i)(1+\varkappa_{i-1})(1+\varkappa_i) - 1, \quad i = 3, \ldots, m-1;\\
&\alpha_m = \overline{\alpha}_m; \quad \beta_m = (1+\overline{\beta}_m)(1+\varkappa_{m-1}) - 1,
\end{aligned} \tag{3.9}$$

полученные соотношения можно записать в виде

$$s_2(d_1 + \alpha_1 - \eta) + c_2 b_2 = 0;$$

$$s_3 s_2 b_2 (1+\beta_2) + s_3 c_2 (d_2 + \alpha_2 - \eta) + c_3 b_3 = 0;$$

$$s_{i+1}s_i c_{i-1} b_i (1+\beta_i) + s_{i+1}c_i(d_i + \alpha_i - \eta) + c_{i+1}b_{i+1} = 0, \quad i = 3, \ldots, m-1;$$

$$s_m c_{m-1} b_m (1+\beta_m) + c_m(d_m + \alpha_m - \eta) = 0.$$

Они и являются основным результатом данного параграфа и, очевидно, полностью совпадают с соотношениями (1.5). Что касается соотношений (1.3) и неравенств (1.4), то их выполнение непосредственно вытекает из определения (3.5).

Обратимся теперь к случаю двухдиагональной матрицы A (2.1). Предположим, что нам известны некоторое достаточно хорошее приближение σ к наибольшему сингулярному числу $\sigma_N(A) = \|A\|$ матрицы A, а также отношения $\mathscr{P}_1, \mathscr{R}_2, \mathscr{P}_2, \ldots, \mathscr{R}_N, \mathscr{P}_N$ последовательных компонент почти* собственного вектора $w = (u_1, v_1, u_2, v_2, \ldots, u_N, v_N)^T$ трехдиагональной симметрической матрицы

$$S = \begin{bmatrix}
0 & a_1 & & & & & \\
a_1 & 0 & b_2 & & & 0 & \\
 & b_2 & 0 & a_2 & & & \\
 & & a_2 & 0 & b_3 & & \\
 & & & \ddots & \ddots & \ddots & \\
 & & & & a_{N-1} & 0 & b_N \\
 & 0 & & & & b_N & 0 & a_N \\
 & & & & & & a_N & 0
\end{bmatrix},$$

отвечающего ее почти собственному значению σ. Более точно, предполагается, что имеют место неравенство

$$|\sigma - \|A\|| \leqslant \delta \|A\|, \quad \delta \ll 1, \tag{3.10}$$

и соотношения

$$\left.\begin{aligned} &-\sigma + a_1(1+\check{\zeta}_1)\mathscr{P}_1 = 0;\\ &\frac{a_{i-1}(1+\hat{\zeta}_{i-1})}{\mathscr{P}_{i-1}} - \sigma + b_i(1+\check{\xi}_i)\mathscr{R}_i = 0,\\ &\frac{b_i(1+\hat{\xi}_i)}{\mathscr{R}_i} - \sigma + a_i(1+\check{\zeta}_i)\mathscr{P}_i = 0,\\ &\frac{a_N(1+\hat{\zeta}_N)}{\mathscr{P}_N} - \sigma = 0, \end{aligned}\right\} \quad i = 2, \ldots, N; \tag{3.11}$$

где возмущения $\hat{\zeta}_i$, $\check{\zeta}_i$, $\hat{\xi}_i$, $\check{\xi}_i$ удовлетворяют оценкам

$$|\hat{\zeta}_i| \leqslant \rho, \quad |\check{\zeta}_i| \leqslant \rho, \quad |\hat{\xi}_i| \leqslant \rho, \quad |\check{\xi}_i| \leqslant \rho, \quad \rho \ll 1. \tag{3.12}$$

Заметим, что основанием для соотношений (3.11) служит теорема 1 из [4], которую достаточно просто переформулировать на данный случай. Подробно получение этих соотношений описано в § 8 главы 4, где приводится общая схема предлагаемого алгоритма исчерпывания. Особо подчеркнем, что под σ понимается приближение именно к наибольшему сингулярному числу, в противном случае нельзя гарантировать малости возмущений $\hat{\zeta}_i$, $\check{\zeta}_i$, $\hat{\xi}_i$, $\check{\xi}_i$.

Умножая второе из соотношений (3.11) на $\mathscr{P}_1$, третье — на $\mathscr{R}_2$ и вообще $(2i-1)$-е — на $\mathscr{R}_i$, $2i$-е — на $\mathscr{P}_i$, преобразуем их к виду

$$\left.\begin{aligned} &-\sigma + a_i(1+\check{\zeta}_i)\mathscr{P}_i = 0;\\ &a_{i-1}(1+\hat{\zeta}_{i-1}) - \sigma\mathscr{P}_{i-1} + b_i(1+\check{\xi}_i)\mathscr{P}_{i-1}\mathscr{R}_i = 0,\\ &b_i(1+\hat{\xi}_i) - \sigma\mathscr{R}_i + a_i(1+\check{\zeta}_i)\mathscr{R}_i\mathscr{P}_i = 0,\\ &a_N(1+\hat{\zeta}_N) - \sigma\mathscr{P}_N = 0. \end{aligned}\right\} \quad i = 2, \ldots, N; \tag{3.13}$$

Введем в рассмотрение последовательности чисел s_i, c_i, c_i', $\bar{s}_i$, $\bar{c}_i$, $\bar{c}_i'$, определив их следующим образом:

$$\left.\begin{aligned} &s_2 = \frac{1}{\sqrt{1+(\mathscr{P}_1\mathscr{R}_2)^2}}, \quad c_2 = \frac{\mathscr{P}_1\mathscr{R}_2}{\sqrt{1+(\mathscr{P}_1\mathscr{R}_2)^2}}, \quad c_2' = (c_2)_{\text{м}};\\ &\bar{s}_2 = \frac{1}{\sqrt{1+(\mathscr{R}_2\mathscr{P}_2)^2}}, \quad \bar{c}_2 = \frac{\mathscr{R}_2\mathscr{P}_2}{\sqrt{1+(\mathscr{R}_2\mathscr{P}_2)^2}}, \quad \bar{c}_2' = (\bar{c}_2)_{\text{м}};\\ &s_i = \frac{1}{\sqrt{1+(c_{i-1}'\mathscr{P}_{i-1}\mathscr{R}_i)^2}}, \quad c_i = \frac{c_{i-1}'\mathscr{P}_{i-1}\mathscr{R}_i}{\sqrt{1+(c_{i-1}'\mathscr{P}_{i-1}\mathscr{R}_i)^2}}, \quad c_i' = (c_i)_{\text{м}},\\ &\bar{s}_i = \frac{1}{\sqrt{1+(\bar{c}_{i-1}'\mathscr{R}_i\mathscr{P}_i)^2}}, \quad \bar{c}_i = \frac{\bar{c}_{i-1}'\mathscr{R}_i\mathscr{P}_i}{\sqrt{1+(\bar{c}_{i-1}'\mathscr{R}_i\mathscr{P}_i)^2}}, \quad \bar{c}_i' = (\bar{c}_i)_{\text{м}}, \end{aligned}\right\} \quad i=3,\ldots,N. \tag{3.14}$$

Предположим, что величины c_i' и $\bar{c}_i'$ можно вычислить так, что

$$c_i' = c_i(1+\tau_i), \quad \bar{c}_i' = \bar{c}_i(1+\bar{\tau}_i), \tag{3.15}$$

где $|\tau_i| \leqslant \tau$, $|\bar{\tau}_i| \leqslant \tau$, $\tau \ll 1$. Из определения (3.14) ясно, что $s_i > 0$, $\bar{s}_i > 0$. А из того, что числа $\mathscr{P}_i$, $\mathscr{R}_i$ отличны от нуля, и из предположений (3.15) легко следует, что и все числа c_i, c_i', $\bar{c}_i$, $\bar{c}_i'$ отличны от нуля.

Оставляя первое и последнее из соотношений (3.13) без изменений, а второе, третье и вообще $2i$-е и $(2i+1)$-е умножая соответственно на s_2, $\bar{s}_2$, $s_{i+1}c_i'$ и $\bar{s}_{i+1}\bar{c}_i'$, преобразуем их к виду

$$
\begin{aligned}
&-\sigma + a_1(1+\check{\zeta}_1)\mathscr{P}_1 = 0;\\
&s_2 a_1(1+\hat{\zeta}_1) - s_2\sigma\mathscr{P}_1 + c_2 b_2(1+\check{\xi}_2) = 0;\\
&\bar{s}_2 b_2(1+\hat{\xi}_2) - \bar{s}_2\sigma\mathscr{R}_2 + \bar{c}_2 a_2(1+\check{\zeta}_2) = 0;
\end{aligned}
\tag{3.16}
$$

$$
\left.
\begin{aligned}
&s_{i+1}c_i a_i(1+\hat{\zeta}_i) - s_{i+1}c_i\sigma\mathscr{P}_i + c_{i+1}b_{i+1}\frac{1+\check{\xi}_{i+1}}{1+\tau_i} = 0,\\
&\bar{s}_{i+1}\bar{c}_i b_{i+1}(1+\hat{\xi}_{i+1}) - \bar{s}_{i+1}\bar{c}_i\sigma\mathscr{R}_{i+1} + \bar{c}_{i+1}a_{i+1}\frac{1+\check{\zeta}_{i+1}}{1+\bar{\tau}_i} = 0,
\end{aligned}
\right\} i = 2, \ldots, N-1;
$$

$$
a_N(1+\hat{\zeta}_N) - \sigma\mathscr{P}_N = 0.
$$

При получении этих соотношений учтены равенства

$$
c_2 = s_2\mathscr{P}_1\mathscr{R}_2; \quad \bar{c}_2 = \bar{s}_2\mathscr{R}_2\mathscr{P}_2;
$$

$$
c_{i+1} = s_{i+1}c_i'\mathscr{P}_i\mathscr{R}_{i+1}; \quad \bar{c}_{i+1} = \bar{s}_{i+1}\bar{c}_i'\mathscr{R}_{i+1}\mathscr{P}_{i+1}, \quad i = 2, \ldots, N-1, \tag{3.17}
$$

непосредственно следующие из (3.14), а также предположения (3.15).

Покажем теперь, как, пользуясь первым из полученных соотношений, исключить $\mathscr{P}_1$, $\mathscr{R}_2$, $\mathscr{P}_2$, ..., $\mathscr{R}_N$, $\mathscr{P}_N$ из остальных соотношений (3.16). Из формул (3.17) легко следуют рекуррентные соотношения для $\mathscr{P}_i$ и $\mathscr{R}_i$:

$$
\mathscr{P}_2 = \frac{\bar{c}_2}{c_2}\frac{s_2}{\bar{s}_2}\mathscr{P}_1, \quad \mathscr{P}_i = \frac{\bar{c}_i}{c_i}\frac{c_{i-1}'}{\bar{c}_{i-1}'}\frac{s_i}{\bar{s}_i}\mathscr{P}_{i-1}, \quad i = 3, \ldots, N;
$$

$$
\mathscr{R}_2 = \frac{c_2}{s_2}\frac{1}{\mathscr{P}_1}, \quad \mathscr{R}_2 = \frac{c_3}{c_2}\frac{1}{c_2'}\frac{\bar{s}_2}{s_3}\mathscr{R}_2, \quad \mathscr{R}_i = \frac{c_i}{c_{i-1}}\frac{\bar{c}_{i-2}'}{c_{i-1}'}\frac{\bar{s}_{i-1}}{s_i}\mathscr{R}_{i-1}, \quad i = 4, \ldots, N,
$$

последовательное применение которых дает

$$
\mathscr{P}_i = \frac{\bar{c}_i}{c_i}\frac{c_{i-1}'}{c_{i-1}}\frac{\bar{c}_{i-1}}{\bar{c}_{i-1}'}\frac{c_{i-2}'}{c_{i-2}}\frac{\bar{c}_{i-2}}{\bar{c}_{i-2}'}\cdots\frac{c_2'}{c_2}\frac{\bar{c}_2}{\bar{c}_2'}\frac{s_i\cdots s_2}{\bar{s}_i\cdots\bar{s}_2}\mathscr{P}_1, \quad i = 3, \ldots, N;
$$

$$
\mathscr{R}_i = \frac{c_i}{\bar{c}_{i-1}}\frac{c_{i-1}}{c_{i-1}'}\frac{\bar{c}_{i-2}'}{\bar{c}_{i-2}}\frac{c_{i-2}}{c_{i-2}'}\cdots\frac{c_3}{c_3'}\frac{\bar{c}_2'}{\bar{c}_2}\frac{c_2}{c_2'}\frac{\bar{s}_{i-1}\cdots\bar{s}_2}{s_i s_{i-1}\cdots s_2}\frac{1}{\mathscr{P}_1}, \quad i = 3, \ldots, N.
$$

Из первого соотношения (3.16) следует формула $\mathscr{P}_1 = \sigma/[a_1(1+\check{\zeta}_1)]$. Подставляя ее в выражения для $\mathscr{P}_i$ и $\mathscr{R}_i$ и учитывая (3.15), получим

$$
\begin{aligned}
&\mathscr{R}_2 = c_2\frac{1}{s_2}\frac{a_1(1+\check{\zeta}_1)}{\sigma};\\
&\mathscr{P}_2 = \frac{\bar{c}_2}{c_2}\frac{s_2}{\bar{s}_2}\frac{\sigma}{a_1(1+\check{\zeta}_1)};\\
&\mathscr{R}_3 = \frac{c_3}{\bar{c}_2}\frac{1}{1+\tau_2}\frac{\bar{s}_2}{s_3 s_2}\frac{a_1(1+\check{\zeta}_1)}{\sigma};
\end{aligned}
\tag{3.18}
$$

$$
\mathscr{P}_i = \frac{c_i}{\bar{c}_i}\frac{(1+\tau_{i-1})\cdots(1+\tau_2)}{(1+\bar{\tau}_{i-1})\cdots(1+\bar{\tau}_2)}\frac{s_i\cdots s_2}{\bar{s}_i\cdots\bar{s}_2}\frac{\sigma}{a_1(1+\check{\zeta}_1)}, \quad i = 3, \ldots, N;
$$

$$
\mathscr{R}_i = \frac{c_i}{\bar{c}_{i-1}}\frac{(1+\bar{\tau}_{i-2})\cdots(1+\bar{\tau}_2)}{(1+\tau_{i-1})(1+\tau_{i-2})\cdots(1+\tau_2)}\frac{\bar{s}_{i-1}\cdots\bar{s}_2}{s_i s_{i-1}\cdots s_2}\frac{a_1(1+\check{\zeta}_1)}{\sigma}, \quad i = 4, \ldots, N.
$$

Введем обозначения

$$\hat{\alpha}_1 = (1+\hat{\zeta}_1)(1+\check{\zeta}_1) - 1;$$

$$\check{\beta}_2 = (1+\check{\xi}_2)(1+\check{\zeta}_1) - 1, \quad \hat{\beta}_2 = \frac{1+\hat{\xi}_2}{1+\check{\zeta}_1} - 1;$$

$$\check{\alpha}_2 = \frac{1+\check{\zeta}_2}{1+\check{\zeta}_1} - 1, \quad \hat{\alpha}_2 = (1+\hat{\zeta}_2)(1+\check{\zeta}_1) - 1;$$

$$\check{\beta}_3 = \frac{(1+\check{\xi}_3)(1+\check{\zeta}_1)}{1+\tau_2} - 1, \quad \hat{\beta}_3 = \frac{(1+\hat{\xi}_3)(1+\tau_2)}{1+\check{\zeta}_1} - 1; \tag{3.19}$$

$$\left.\begin{aligned} \check{\alpha}_i &= \frac{(1+\check{\zeta}_i)(1+\tau_{i-1})\dots(1+\tau_2)}{(1+\check{\zeta}_1)(1+\bar{\tau}_{i-1})\dots(1+\bar{\tau}_2)} - 1, \\ \hat{\alpha}_i &= \frac{(1+\hat{\zeta}_i)(1+\check{\zeta}_1)(1+\bar{\tau}_{i-1})\dots(1+\bar{\tau}_2)}{(1+\tau_{i-1})\dots(1+\tau_2)} - 1, \end{aligned}\right\} i = 3, \dots, N;$$

$$\left.\begin{aligned} \check{\beta}_i &= \frac{(1+\check{\xi}_i)(1+\check{\zeta}_1)(1+\bar{\tau}_{i-2})\dots(1+\bar{\tau}_2)}{(1+\tau_{i-1})(1+\tau_{i-2})\dots(1+\tau_2)} - 1, \\ \hat{\beta}_i &= \frac{(1+\hat{\xi}_i)(1+\tau_{i-1})(1+\tau_{i-2})\dots(1+\tau_2)}{(1+\check{\zeta}_1)(1+\bar{\tau}_{i-2})\dots(1+\bar{\tau}_2)} - 1, \end{aligned}\right\} i = 4, \dots, N.$$

С учетом этих обозначений и формул (3.18) второе, третье и следующие соотношения (3.16) перепишем в виде

$$s_2 a_1 (1+\hat{\alpha}_1) - s_2 \frac{\sigma^2}{a_1} + c_2 b_2 (1+\check{\beta}_2) = 0;$$

$$\bar{s}_2 b_2 (1+\hat{\beta}_2) - c_2 \frac{\bar{s}_2}{s_2} a_1 + \bar{c}_2 a_2 (1+\check{\alpha}_2) = 0;$$

$$\left.\begin{aligned} s_{i+1} c_i a_i (1+\hat{\alpha}_i) - \bar{c}_i \frac{s_{i+1} s_i \dots s_2}{\bar{s}_i \dots \bar{s}_2} \frac{\sigma^2}{a_1} + c_{i+1} b_{i+1} (1+\check{\beta}_{i+1}) = 0, \\ \bar{s}_{i+1} \bar{c}_i b_{i+1} (1+\hat{\beta}_{i+1}) - c_{i+1} \frac{\bar{s}_{i+1} \dots \bar{s}_2}{s_{i+1} \dots s_2} a_1 + \bar{c}_{i+1} a_{i+1} (1+\check{\alpha}_{i+1}) = 0, \end{aligned}\right\} i = 2, \dots, N-1;$$

$$c_N a_N (1+\hat{\alpha}_N) - \bar{c}_N \frac{s_N \dots s_2}{\bar{s}_N \dots \bar{s}_2} \frac{\sigma^2}{a_1} = 0.$$

Они, очевидно, полностью совпадают с соотношениями (2.5). Выполнение же соотношений (2.3) непосредственно следует из определения (3.14) чисел c_i, s_i, $\bar{c}_i$, $\bar{s}_i$.

Для завершения рассмотрений, связанных с соотношениями (2.5), дадим оценки возмущений $\hat{\alpha}_i$, $\check{\alpha}_i$, $\hat{\beta}_i$ и $\check{\beta}_i$. Для примера оценим величину $\check{\alpha}_i$. Очевидно, что

$$\check{\alpha}_i \leqslant \frac{(1+\rho)(1+\tau)^{i-2}}{(1-\rho)(1-\tau)^{i-2}} - 1 \leqslant \frac{(1+\rho)(1+\tau)^{N-2}}{(1-\rho)(1-\tau)^{N-2}} - 1.$$

В предположениях $N^2\tau \leqslant 1$, $2(2N\tau+\rho)\rho \leqslant \tau$, которые при всех разумных порядках N совершенно необременительны, легко получить оценку $\check{\alpha}_i \leqslant 2(N\tau+\rho)$. Подобным же образом оценивается $\check{\alpha}_i$ снизу: $\check{\alpha}_i \geqslant -2(N\tau+\rho)$. Следовательно, $|\check{\alpha}_i| \leqslant 2(N\tau+\rho)$. Аналогичные оценки имеют место и для остальных возмущений. Таким образом, обозначив

$$\varepsilon = 2(N\tau+\rho), \tag{3.20}$$

можно записать $|\hat{\alpha}_i| \leqslant \varepsilon$, $|\check{\alpha}_i| \leqslant \varepsilon$, $|\hat{\beta}_i| \leqslant \varepsilon$, $|\check{\beta}_i| \leqslant \varepsilon$.

Как уже отмечалось, значение ρ оценивает величину возмущений $\hat{\zeta}_i$, $\check{\zeta}_i$, $\hat{\xi}_i$, $\check{\xi}_i$, с которыми удовлетворены уравнения (3.11). Получение этих уравнений и оценка величины ρ — самостоятельная задача, решению которой посвящена работа [4]. Величина же τ, оценивающая погрешность машинного вычисления чисел c_i и $\bar{c}_i$, оценивается аналогично тому, как в следующем параграфе определена погрешность вычисления чисел c_i в случае трехдиагональной матрицы.

§ 4. Машинное вычисление параметров, задающих преобразования вращения

Основной целью данного параграфа является обоснование предположений (3.6), обеспечивающих высокую относительную точность вычисления величин c_i [см. формулы (3.5)], а также получение аналогичного результата относительно вычисления величин s_i. Подобные результаты сформулированы и для параметров c_i, s_i, $\bar{c}_i$, $\bar{s}_i$, используемых в алгоритме исчерпывания двухдиагональной матрицы [см. формулы (3.14)].

Предварительно напомним определение параметров ε_1 и ε_2, которыми будем характеризовать разрядную сетку вычислительной машины [см. [4], § 4]. При этом будем рассматривать машины, использующие представление чисел с «плавающей точкой». Каждое число $x \neq 0$ в памяти таких машин хранится в виде $x = \alpha^t s$, где α — основание системы счисления, принятой в машине, t — целочисленный порядок, s $(1/\alpha \leqslant \leqslant |s| < 1)$ — мантисса числа x. Число разрядов, отводимых для размещения мантиссы, ограничивает относительную точность представления чисел в машине. В качестве характеристики этой точности используют число ε_1, определяемое как наименьшее положительное машинное число, для которого $(1+\varepsilon_1)_м > 1$. Число разрядов, отводимых для размещения порядка, задает величину наибольшего и наименьшего по модулю чисел, которые могут быть представлены в машине. Обозначим наибольшее по модулю машинное число через $1/\varepsilon_2$. Тогда наименьшим по модулю отличным от нуля машинным числом будет

$$\frac{\varepsilon_2}{\alpha^2}\left(1-\frac{\varepsilon_1}{\alpha}\right) < \frac{\varepsilon_2}{\alpha^2}.$$

Погрешность выполнения арифметических операций определяется следующим неравенством:

$$|(a * b)_м - (a * b)| \leqslant \varepsilon_1 |a * b| + \varepsilon_2/\alpha^2, \tag{4.1}$$

где под знаком «$*$» подразумевается один из знаков арифметических операций сложения, вычитания, умножения или деления. Первое слагаемое $\varepsilon_1|a * b|$ в правой части (4.1) определяет погрешность, возникающую из-за ограниченности числа разрядов, отводимых для размещения мантиссы результата операции. Второе слагаемое описывает погрешность, возникающую в том случае, когда модуль результата операции меньше, чем наименьшее по модулю отличное от нуля машинное число. В этом случае $(a * b)_м$ принудительно полагается равным нулю, а возникающая из-за этого погрешность, очевидно, не превосходит ε_2/α^2. В дальнейшем будем называть ее погрешностью исчезновения порядка.

Перейдем непосредственно к машинному вычислению параметров, определяющих преобразования вращения. Представим числа c'_{i-1} и $\overline{\mathscr{P}}_i$ $(c'_{i-1} \neq 0,\ \overline{\mathscr{P}}_i \neq 0)$ в виде

$$\begin{aligned} c'_{i-1} &= \alpha^{k_i - 1} q_{i-1}, \quad \frac{1}{\alpha} \leqslant |q_{i-1}| < 1; \\ \overline{\mathscr{P}}_i &= \alpha^{m_i} p_i, \qquad \frac{1}{\alpha} \leqslant |p_i| < 1 \end{aligned} \tag{4.2}$$

и сформулируем алгоритм вычисления c_i по формуле

$$c_i = \frac{c'_{i-1}\overline{\mathscr{P}}_i}{\sqrt{1+(c'_{i-1}\overline{\mathscr{P}}_i)^2}}$$

следующим образом:

1) $x = (q_{i-1}p_i)_м$;
2) $j = k_{i-1} + m_i$;
3) $y = (xx)_м$;
4) представим y в виде
$$y = \alpha^t z, \quad 1/\alpha \leqslant |z| < 1; \tag{4.3}$$
5) $l = t + 2j$;
6) $g = \max(1, l)$;
7) если g нечетно, положим $g = g + 1$;
8) $u = (\alpha^{1-g}/\alpha)_м$, $v = (\alpha^{l-g}z)_м$;
9) $w = (u+v)_м$;
10) $\mu = (\sqrt{w})_м$;
11) $\nu = (x/\mu)_м$;
12) представим ν в виде $\nu = \alpha^h q_i$, $1/\alpha \leqslant |q_i| < 1$;
13) $k_i = h + j - g/2$.

Оценим погрешности, возникающие при реализации описанного алгоритма, предварительно заметив, что п. 2, 4—7, 12 и 13 выполняются на машине абсолютно точно. Кроме того, использование арифметики вынесенных порядков позволяет при выполнении п. 1, 3, 9—11 избежать погрешности исчезновения порядка, в результате чего оценки погрешностей, возникающих при выполнении этих пунктов, упрощаются. Учитывая высказанные замечания, получим

$$|x - q_{i-1}p_i| \leqslant \varepsilon_1 |q_{i-1}p_i|, \tag{4.4}$$
$$|y - (q_{i-1}p_i)^2| \leqslant |y - x^2| + |x^2 - (q_{i-1}p_i)^2| \leqslant$$
$$\leqslant \varepsilon_1 x^2 + |x - q_{i-1}p_i||x + q_{i-1}p_i| \leqslant [(1+\varepsilon_1)^3 - 1](q_{i-1}p_i)^2 <$$
$$< \varepsilon_1 (3 + 4\varepsilon_1)(q_{i-1}p_i)^2.$$

Остановимся на выполнении п. 8 и 9. При вычислении одной из величин u или v возможно появление погрешности исчезновения порядка. Эта погрешность возникает всякий раз, когда порядки 1 и l отличаются на достаточно большую величину (например, для машин серии ЕС ЭВМ эта величина равна 64, если число $\max(1, l)$ четно, и 63 — в противном случае). В силу определения g ясно, что хотя бы одна из величин u или v вычисляется точно. Предположим для определенности, что точно вычисляется u, а при вычислении v происходит исчезновение порядка, т. е. $\alpha^{l-g}z < \varepsilon_2/\alpha^2$. В этом случае $v = 0$, $u = \alpha^{1-g}/\alpha$, операция сложения u и v осуществляется, очевидно, точно и, следовательно,

$$\left|w - \left(\alpha^{1-g}\frac{1}{\alpha} + \alpha^{l-g}z\right)\right| = \alpha^{l-g}z < \frac{\varepsilon_2}{\alpha^2} = \frac{\varepsilon_2}{\alpha^{2-g}}\alpha^{1-g}\frac{1}{\alpha} <$$
$$< \frac{\varepsilon_2}{\alpha^{2-g}}\left(\alpha^{1-g}\frac{1}{\alpha} + \alpha^{l-g}z\right) < \varepsilon_1\left(\alpha^{1-g}\frac{1}{\alpha} + \alpha^{l-g}z\right).$$

При выводе последнего неравенства использовано то обстоятельство, что в данном случае $\max(1, l) = 1$, вследствие чего $2 - g = 0$, и неравенство $\varepsilon_2 \ll \varepsilon_1$. Случай, когда v вычисляется точно, а исчезновение порядка происходит при вычислении u, анализируется аналогично. Наконец, если вычисление и u и v осуществляется без исчезновения порядка, т. е. точно, погрешность выполнения п. 9 оценивается стандартным образом:

$$\left|w - \left(\alpha^{1-g}\frac{1}{\alpha} + \alpha^{l-g}z\right)\right| = |w - (u+v)| \leqslant \varepsilon_1 (u+v) = \varepsilon_1\left(\alpha^{1-g}\frac{1}{\alpha} + \alpha^{l-g}z\right).$$

Итак, во всех рассматриваемых случаях имеет место оценка

$$\left|w-\left(\alpha^{1-g}\frac{1}{\alpha}+\alpha^{l-g}z\right)\right|\leqslant\varepsilon_1\left(\alpha^{1-g}\frac{1}{\alpha}+\alpha^{l-g}z\right),$$

которая очевидным образом преобразуется к виду

$$|\alpha^g w-(1+\alpha^l z)|\leqslant\varepsilon_1(1+\alpha^l z).$$

Учитывая определение l и z и оценки погрешностей выполнения предыдущих пунктов, теперь нетрудно получить оценку

$$|\alpha^g w-[1+\alpha^{2j}(q_{i-1}p_i)^2]|\leqslant\varepsilon_1(4+8\varepsilon_1)[1+\alpha^{2j}(q_{i-1}p_i)^2]. \qquad (4.5)$$

Перейдем к оценкам погрешностей, допускаемых на следующих пунктах алгоритма:

$$|\mu-\sqrt{w}|\leqslant\varepsilon_1\sqrt{w};$$

$$\left|\nu-\frac{x}{\sqrt{w}}\right|\leqslant\left|\nu-\frac{x}{\mu}\right|+\left|\frac{x}{\mu}-\frac{x}{\sqrt{w}}\right|\leqslant\varepsilon_1\left|\frac{x}{\mu}\right|+\frac{|x||\mu-\sqrt{w}|}{\mu\sqrt{w}}\leqslant\frac{2\varepsilon_1}{1-\varepsilon_1}\frac{|x|}{\sqrt{w}}.$$

Последнее неравенство после умножения обеих его частей на $\alpha^{-g/2}$ принимает вид

$$\left|\alpha^{-g/2}\nu-\frac{x}{\sqrt{\alpha^g w}}\right|\leqslant\frac{2\varepsilon_1}{1-\varepsilon_1}\frac{|x|}{\sqrt{\alpha^g w}}. \qquad (4.6)$$

Наконец, на основе неравенств (4.4)—(4.6) с помощью несложных, но довольно громоздких выкладок можно получить оценку

$$\left|\alpha^{-g/2}\nu-\frac{q_{i-1}p_i}{\sqrt{1+\alpha^{2j}(q_{i-1}p_i)^2}}\right|\leqslant\varepsilon_1(5+35\varepsilon_1)\frac{q_{i-1}p_i}{\sqrt{1+\alpha^{2j}(q_{i-1}p_i)^2}}.$$

Умножая обе части этой оценки на α^j и учитывая формулы (4.2), а также формулы п. 2, 12 и 13 алгоритма (4.3), ее легко преобразовать к виду

$$\left|\alpha^{h_i}q_i-\frac{c'_{i-1}\overline{\mathcal{P}}_i}{\sqrt{1+(c'_{i-1}\overline{\mathcal{P}}_i)^2}}\right|\leqslant\varepsilon_1(5+35\varepsilon_1)\frac{c'_{i-1}\overline{\mathcal{P}}_i}{\sqrt{1+(c'_{i-1}\overline{\mathcal{P}}_i)^2}}.$$

Обозначив теперь величину $\alpha^{h_i}q_i$ через c'_i, получим

$$|c'_i-c_i|\leqslant\varepsilon_1(5+35\varepsilon_1)|c_i|. \qquad (4.7)$$

Особо подчеркнем, что при определении числа $c'_i\equiv(c_i)_{\text{м}}$ по формуле $c'_i=\alpha^{h_i}q_i$ операция умножения на машине не выполняется, т. е. подразумевается, что c'_i хранится в памяти машины в виде двух чисел — порядка k_i и мантиссы q_i.

Пользуясь малостью ε_1, оценку (4.7) можно огрубить, например, следующим образом: $|c'_i-c_i|\leqslant 5{,}001\varepsilon_1|c_i|$. Очевидно, что эта оценка эквивалентна равенству $c'_i=c_i(1+\varkappa_i)$, которое имеет место при некотором $\varkappa_i$ таком, что $|\varkappa_i|\leqslant 5{,}001\varepsilon_1$. Следовательно, в качестве значения $\varkappa$, оценивающего сверху величину погрешности $\varkappa_i$ [см. формулы (3.6)], можно взять $\varkappa=5{,}001\varepsilon_1$.

Для вычисления s_i по формуле [см. (3.5)]

$$s_i=\frac{1}{\sqrt{1+(c'_{i-1}\overline{\mathcal{P}}_i)^2}}$$

можно сформулировать алгоритм, аналогичный (4.3). Этот алгоритм совпадает с (4.3) по всем пунктам, за исключением п. 11—13. Эти пункты формулируются теперь следующим образом:

11′) $\nu=(1/\mu)_{\text{м}}$;

12′) представим ν в виде $\nu=\alpha^h r_i$, $1/\alpha\leqslant|r_i|<1$;

13′) $n_i=h-g/2$.

В результате применения этого алгоритма находятся числа n_i и r_i, которые и определяют величину $s_i' \equiv (s_i)_м : s_i' = \alpha^{n_i} r_i$. Анализ погрешностей, проводимый аналогично предыдущему, позволяет утверждать, что $s_i' = s_i(1 + \omega_i)$, где $|\omega_i| \leqslant \omega$, а в качестве значения ω можно взять $\omega = 4{,}001\varepsilon_1$.

Обратимся теперь к вычислению чисел c_i, s_i, $\bar{c}_i$, $\bar{s}_i$, используемых в алгоритме исчерпывания двухдиагональной матрицы [см. формулы (3.14)]. Представив числа c_{i-1}', $\mathscr{P}_{i-1}$ и $\mathscr{R}_i$ в виде

$$c_{i-1}' = \alpha^{k_i - 1} q_{i-1}, \quad \frac{1}{\alpha} \leqslant |q_{i-1}| < 1;$$

$$\mathscr{P}_{i-1} = \alpha^{m_{i-1}} p_{i-1}, \quad \frac{1}{\alpha} \leqslant |p_{i-1}| < 1;$$

$$\mathscr{R}_i = \alpha^{n_i} r_i, \quad \frac{1}{\alpha} \leqslant |r_i| < 1;$$

алгоритм вычисления c_i по формуле

$$c_i = \frac{c_{i-1}' \mathscr{P}_{i-1} \mathscr{R}_i}{\sqrt{1 + (c_{i-1}' \mathscr{P}_{i-1} \mathscr{R}_i)^2}}$$

сформулируем следующим образом:

1) $x = (q_{i-1} p_{i-1} r_i)_м$;
2) $j = k_{i-1} + m_{i-1} + n_i$;
3) $y = (xx)_м$;
4) представим y в виде $y = \alpha^t z$, $1/\alpha \leqslant |z| < 1$;
5) $l = t + 2j$;
6) $g = \max(1, l)$;
7) если g нечетно, положим $g = g + 1$;
8) $u = \left(\alpha^{1-g} \frac{1}{\alpha}\right)_м$, $v = (\alpha^{l-g} z)_м$;
9) $w = (u + v)_м$;
10) $\mu = (\sqrt{w})_м$;
11) $\nu = (x/\mu)_м$;
12) представим ν в виде $\nu = \alpha^h q_i$, $1/\alpha \leqslant |q_i| < 1$;
13) $k_i = h + j - g/2$.

Легко показать, что величина $c_i' = \alpha^{k_i} q_i$, получаемая в результате работы этого алгоритма, связана с c_i соотношением $c_i' = c_i(1 + \tau_i)$, где $|\tau_i| \leqslant \tau$, а в качестве τ можно взять $\tau = 7{,}001\varepsilon_1$. Совершенно аналогичный алгоритм можно предложить для вычисления величины

$$\bar{c}_i = \frac{\bar{c}_{i-1}' \mathscr{R}_i \mathscr{P}_i}{\sqrt{1 + (\bar{c}_{i-1}' \mathscr{R}_i \mathscr{P}_i)^2}}.$$

При этом вычисленная величина $\bar{c}_i' \equiv (\bar{c}_i)_м$ оказывается связанной с величиной $\bar{c}_i$ соотношением $\bar{c}_i' = \bar{c}_i(1 + \bar{\tau}_i)$, где $|\bar{\tau}_i| \leqslant \tau$.

Отметим, наконец, что и для вычисления величин s_i и $\bar{s}_i$ можно сформулировать соответствующие алгоритмы, результатами выполнения которых являются числа s_i' и $\bar{s}_i'$, удовлетворяющие равенствам $s_i' = s_i(1 + \pi_i)$, $\bar{s}_i' = \bar{s}_i(1 + \bar{\pi}_i)$, где $|\pi_i| \leqslant \pi$, $|\bar{\pi}_i| \leqslant \pi$, а в качестве π можно взять $\pi = 5{,}001\varepsilon_1$.

Глава 3

МАШИННОЕ ВЫЧИСЛЕНИЕ ПРЕОБРАЗОВАННЫХ МАТРИЦ

§ 5. Анализ погрешностей, возникающих при вычислении элементов преобразованной трехдиагональной матрицы

В этом параграфе описано вычисление элементов $\bar{d}_i$, $\bar{b}_i$ матрицы (1.7) $\bar{A}$, получаемой в результате применения алгоритма исчерпывания к трехдиагональной симметрической матрице A, и проанализированы возникающие при этом погрешности. Напомним формулы, по которым определяются эти элементы [см. (1.9)],

$$\bar{d}_1 = d_2 - \frac{c_2 b_2}{s_2} + \frac{c_3 c_2 b_3}{s_3};$$

$$\bar{d}_i = d_{i+1} - \frac{c_{i+1} c_i b_{i+1}}{s_{i+1}} + \frac{c_{i+2} c_{i+1} b_{i+2}}{s_{i+2}}, \quad i = 2, \ldots, m-2;$$

$$\bar{d}_{m-1} = d_m - \frac{c_m c_{m-1} b_m}{s_m};$$

$$\bar{b}_i = \frac{s_i b_{i+1}}{s_{i+1}}, \quad i = 2, \ldots, m-1.$$

Как уже отмечалось, погрешности, возникающие при вычислении по этим формулам, имеют двоякую природу. Прежде всего, они определяются тем, что вместо точных значений параметров s_i и c_i, фигурирующих в формулах, в памяти машины имеются лишь некоторые их приближения s_i' и c_i', так что реальные вычисления могут использовать только эти приближенные значения. В таком случае ясно, что вместо точных элементов $\bar{d}_i$ и $\bar{b}_i$ можно вычислить только элементы $\bar{d}_i'$ и $\bar{b}_i'$, определяемые по формулам

$$\bar{d}_i' = d_{i+1} - \frac{c_{i+1}' c_i' b_{i+1}}{s_{i+1}'} + \frac{c_{i+2}' c_{i+1}' b_{i+2}}{s_{i+2}'};$$

$$\bar{b}_i' = \frac{s_i' b_{i+1}}{s_{i+1}'}. \tag{5.1}$$

Второй источник погрешностей связан с непосредственной реализацией на вычислительной машине операций, предписываемых формулами (5.1).

Приступая к анализу погрешностей, напомним, что в § 4 разработан алгоритм вычисления величин s_i' и c_i', гарантирующий выполнение равенств

$$s_i' = s_i(1 + \omega_i), \quad c_i' = c_i(1 + \varkappa_i). \tag{5.2}$$

Эти равенства имеют место при некоторых ω_i и $\varkappa_i$ таких, что $|\omega_i| \leqslant \omega = 4{,}001\varepsilon_1$, $|\varkappa_i| \leqslant \varkappa = 5{,}001\varepsilon_1$. Здесь ε_1, а также используемая ниже величина ε_2 — характеристики разрядной сетки используемой ЭВМ (см. § 4). Важно отметить, что упомянутый алгоритм позволяет определить каждую из величин s_i' и c_i' с помощью двух чисел: порядка и мантиссы, так что

$$s_i' = m(s_i')\,\alpha^{p(s_i')}, \quad c_i' = m(c_i')\,\alpha^{p(c_i')}, \tag{5.3}$$

где $p(s_i')$ и $p(c_i')$ — порядки чисел s_i' и c_i' соответственно, а $m(s_i')$ и $m(c_i')$ — их мантиссы, удовлетворяющие неравенствам

$$\frac{1}{\alpha} \leqslant |m(s_i')| < 1, \quad \frac{1}{\alpha} \leqslant |m(c_i')| < 1. \tag{5.4}$$

Под α здесь понимается основание системы счисления, принятой в вычислительной машине. Отметим также, что в силу (5.3) и (5.4) $s_i' \neq 0$, $c_i' \neq 0$.

Получим оценки величин $\bar{d}_i' - \bar{d}_i$ и $\bar{b}_i' - b_i$. В силу равенств (5.2) ясно, что для этих величин имеют место формулы

$$\bar{d}_i' - \bar{d}_i = -\frac{c_{i+1}c_i b_{i+1}}{s_{i+1}}\left[\frac{(1+\varkappa_{i+1})(1+\varkappa_i)}{1+\omega_{i+1}} - 1\right] + \frac{c_{i+2}c_{i+1}b_{i+2}}{s_{i+2}}\left[\frac{(1+\varkappa_{i+2})(1+\varkappa_{i+1})}{1+\omega_{i+2}} - 1\right];$$

$$\bar{b}_i' - \bar{b}_i = \frac{s_i b_{i+1}}{s_{i+1}}\left(\frac{1+\omega_i}{1+\omega_{i+1}} - 1\right),$$

из которых легко следуют оценки

$$|\bar{d}_i' - \bar{d}_i| \leqslant \left[\frac{(1+\varkappa)^2}{1-\omega} - 1\right]\left(\left|\frac{c_{i+1}c_i b_{i+1}}{s_{i+1}}\right| + \left|\frac{c_{i+2}c_{i+1}b_{i+2}}{s_{i+2}}\right|\right),$$

$$|\bar{b}_i' - \bar{b}_i| \leqslant \frac{2\omega}{1-\omega}\left|\frac{s_i b_{i+1}}{s_{i+1}}\right|.$$

Вводя обозначения

$$\hat{\varkappa} = \frac{(1+\varkappa)^2}{1-\omega} - 1, \quad \hat{\omega} = \frac{2\omega}{1-\omega} \tag{5.5}$$

и учитывая формулу для $\bar{b}_i$, перепишем полученные оценки в виде

$$\begin{aligned} &|\bar{d}_i' - \bar{d}_i| \leqslant \hat{\varkappa}\left(\left|\frac{c_{i+1}c_i b_{i+1}}{s_{i+1}}\right| + \left|\frac{c_{i+2}c_{i+1}b_{i+2}}{s_{i+2}}\right|\right); \\ &|\bar{b}_i' - \bar{b}_i| \leqslant \hat{\omega}|\bar{b}_i|. \end{aligned} \tag{5.6}$$

Перейдем теперь к анализу погрешностей, связанных с вычислением величин $\bar{d}_i'$ и $\bar{b}_i'$ по формулам (5.1). Представив числа b_{i+1} и b_{i+2} как

$$b_{i+1} = m(b_{i+1})\alpha^{p(b_{i+1})}, \quad \frac{1}{\alpha} \leqslant |m(b_{i+1})| < 1;$$

$$b_{i+2} = m(b_{i+2})\alpha^{p(b_{i+2})}, \quad \frac{1}{\alpha} \leqslant |m(b_{i+2})| < 1$$

и обозначив через $(\bar{d}_i')_{\text{м}}$ результат машинного вычисления величины $\bar{d}_i'$, сформулируем алгоритм вычисления:

1) $x = (m(c_{i+1}')\, m(c_i')\, m(b_{i+1})/m(s_{i+1}'))_{\text{м}}$;
2) $j = p(c_{i+1}') + p(c_i') + p(b_{i+1}) - p(s_{i+1}')$;
3) $u = (x\alpha^j)_{\text{м}}$;
4) $y = (m(c_{i+2}')\, m(c_{i+1}')\, m(b_{i+2})/m(s_{i+2}'))_{\text{м}}$;
5) $k = p(c_{i+2}') + p(c_{i+1}') + p(b_{i+2}) - p(s_{i+2}')$;
6) $v = (y\alpha^k)_{\text{м}}$;
7) $w = (v-u)_{\text{м}}$;
8) $(\bar{d}_i')_{\text{м}} = (d_{i+1} + w)_{\text{м}}$.

Оценим погрешности, возникающие при реализации этого алгоритма. Заметим, что вынесение порядков, используемое в алгоритме, позволяет при выполнении п. 1 и 4 избежать исчезновения порядка (см. § 4). В таком случае нетрудно проверить, что оценки погрешностей, допускаемых при выполнении этих пунктов, имеют вид

$$\left|x - \frac{m(c_{i+1}')\, m(c_i')\, m(b_{i+1})}{m(s_{i+1}')}\right| \leqslant [(1+\varepsilon_1)^3 - 1]\left|\frac{m(c_{i+1}')\, m(c_i')\, m(b_{i+1})}{m(s_{i+1}')}\right|;$$

$$\left|y - \frac{m(c_{i+2}')\, m(c_{i+1}')\, m(b_{i+2})}{m(s_{i+2}')}\right| \leqslant [(1+\varepsilon_1)^3 - 1]\left|\frac{m(c_{i+2}')\, m(c_{i+1}')\, m(b_{i+2})}{m(s_{i+2}')}\right|.$$

Умножая обе части первого из полученных неравенств на α^j, а обе части второго — на α^k, перепишем их в виде

$$\left|\alpha^j x - \frac{c'_{i+1}c'_i b_{i+1}}{s'_{i+1}}\right| \leqslant [(1+\varepsilon_1)^3 - 1]\left|\frac{c'_{i+1}c'_i b_{i+1}}{s'_{i+1}}\right|;$$
$$\left|\alpha^k y - \frac{c'_{i+2}c'_{i+1}b_{i+2}}{s'_{i+2}}\right| \leqslant [(1+\varepsilon_1)^3 - 1]\left|\frac{c'_{i+2}c'_{i+1}b_{i+2}}{s'_{i+2}}\right|. \tag{5.7}$$

Остановимся на п. 3 и 6. При выполнении каждого из них возможно появление погрешности исчезновения порядка. Например, при выполнении п. 3 она возникает в том случае, если $|x\alpha^j| < \varepsilon_2/\alpha^2$. При этом величина u полагается равной нулю и, следовательно, имеет место оценка

$$|u - x\alpha^j| < \varepsilon_2/\alpha^2. \tag{5.8}$$

Если при выполнении п. 3 порядок не исчезает, то величина u вычисляется точно и оценка (5.8) остается справедливой. Аналогично оценивается погрешность выполнения п. 6:

$$|v - y\alpha^k| < \varepsilon_2/\alpha^2. \tag{5.9}$$

Погрешность выполнения п. 7 определяется стандартным образом. $|w-(v-u)| \leqslant \varepsilon_1|v-u| + \varepsilon_2/\alpha^2$. Из этой оценки с помощью (5.8) и (5.9) легко получить $|w-(y\alpha^k - x\alpha^j)| \leqslant \varepsilon_1|x\alpha^j| + \varepsilon_1|y\alpha^k| + (3+2\varepsilon_1)\varepsilon_2/\alpha^2$. Отсюда, в свою очередь, в силу (5.7) следует

$$\left|w - \left(\frac{c'_{i+2}c_{i+1}b_{i+2}}{s'_{i+2}} - \frac{c'_{i+1}c'_i b_{i+1}}{s'_{i+1}}\right)\right| \leqslant [(1+\varepsilon_1)^4 - 1]\left(\left|\frac{c'_{i+1}c'_i b_{i+1}}{s'_{i+1}}\right| + \left|\frac{c'_{i+2}c'_{i+1}b_{i+2}}{s'_{i+2}}\right|\right) + \frac{(3+2\varepsilon_1)\varepsilon_2}{\alpha^2}. \tag{5.10}$$

Оценим, наконец, погрешность выполнения п. 8: $|(\overline{d}'_i)_м - (d_{i+1} + w)| \leqslant \varepsilon_1|d_{i+1} + w| + \varepsilon_2/\alpha^2$. Из этой оценки с помощью (5.10) нетрудно получить

$$|(\overline{d}'_i)_м - \overline{d}'_i| \leqslant \varepsilon_1|d_{i+1}| + [(1+\varepsilon_1)^5 - 1]\left(\left|\frac{c'_{i+1}c'_i b_{i+1}}{s'_{i+1}}\right| + \left|\frac{c'_{i+2}c'_{i+1}b_{i+2}}{s'_{i+2}}\right|\right) + \frac{[4+\varepsilon_1(5+2\varepsilon_1)]\varepsilon_2}{\alpha^2}.$$

Учитывая равенства (5.2) и обозначение (5.5) для $\widehat{\varkappa}$, полученную оценку преобразуем к виду

$$|(\overline{d}'_i)_м - \overline{d}'_i| \leqslant \varepsilon_1|d_{i+1}| + [(1+\varepsilon_1)^5 - 1](1+\widehat{\varkappa})\left(\left|\frac{c_{i+1}c_i b_{i+1}}{s_{i+1}}\right| + \left|\frac{c_{i+2}c_{i+1}b_{i+2}}{s_{i+2}}\right|\right) + \frac{[4+\varepsilon_1(5+2\varepsilon_1)]\varepsilon_2}{\alpha^2}. \tag{5.11}$$

Теперь можно оценить отличие вычисленной величины $(\overline{d}'_i)_м$ от $\overline{d}_i$, а именно из (5.6) и (5.11) следует оценка

$$|(\overline{d}'_i)_м - \overline{d}_i| \leqslant \varepsilon_1|d_{i+1}| + \Delta\left(\left|\frac{c_{i+1}c_i b_{i+1}}{s_{i+1}}\right| + \left|\frac{c_{i+2}c_{i+1}b_{i+2}}{s_{i+2}}\right|\right) + \frac{[4+\varepsilon_1(5+2\varepsilon_1)]\varepsilon_2}{\alpha^2}, \tag{5.12}$$

где через Δ обозначена величина

$$\Delta = [(1+\varepsilon_1)^5 - 1](1+\widehat{\varkappa}) + \widehat{\varkappa}. \tag{5.13}$$

Покажем, что выражения $c_{i+1}c_ib_{i+1}/s_{i+1}$, $c_{i+2}c_{i+1}b_{i+2}/s_{i+2}$ ограничены. Точнее говоря, попытаемся оценить эти выражения через величину элементов исходной матрицы A. Для этого рассмотрим соотношение $s_{i+1}s_ic_{i-1}b_i(1+\beta_i)+s_{i+1}c_i(d_i+\alpha_i-\eta)+c_{i+1}b_{i+1}=0$, использованное в § 1. Умножая обе части соотношения на c_i/s_{i+1} и перенося первое и второе слагаемые в правую часть, получим

$$\frac{c_{i+1}c_ib_{i+1}}{s_{i+1}} = -c_is_ic_{i-1}b_i(1+\beta_i) - c_i^2(d_i+\alpha_i-\eta).$$

Отсюда с учетом равенства $c_i^2+s_i^2=1$ следует, что

$$\left|\frac{c_{i+1}c_ib_{i+1}}{s_{i+1}}\right| \leqslant \frac{1}{2}|b_i|(1+\beta_i)+|d_i|+|\eta|+|\alpha_i|.$$

В силу неравенств (1.6) эту оценку перепишем в виде

$$\left|\frac{c_{i+1}c_ib_{i+1}}{s_{i+1}}\right| \leqslant \frac{1}{2}|b_i|(1+\tilde{\beta})+|d_i|+|\eta|+\tilde{\alpha}\mathcal{M}(A).$$

Аналогичным образом будем иметь

$$\left|\frac{c_{i+2}c_{i+1}b_{i+2}}{s_{i+2}}\right| \leqslant \frac{1}{2}|b_{i+1}|(1+\tilde{\beta})+|d_{i+1}|+|\eta|+\tilde{\alpha}\mathcal{M}(A).$$

С помощью полученных оценок из (5.12) нетрудно вывести

$$|(\bar{d}_i')_{\text{м}}-\bar{d}_i| \leqslant \varepsilon_1|d_{i+1}|+\frac{1}{2}\Delta(|b_i|+|d_i|+|b_{i+1}|)(1+\tilde{\beta})+$$

$$+\Delta\left(\frac{1}{2}|d_i|+|d_{i+1}|+2|\eta|\right)+2\Delta\tilde{\alpha}\mathcal{M}(A)+\frac{[4+\varepsilon_1(5+2\varepsilon_1)]\varepsilon_2}{\alpha^2}. \quad (5.14)$$

Из определения величины $\mathcal{M}(A)$ легко следует, что $|d_i|\leqslant\mathcal{M}(A)$, $|d_{i+1}|\leqslant\mathcal{M}(A)$, $|b_i|+|d_i|+|b_{i+1}|\leqslant\mathcal{M}(A)$. Кроме того, поскольку $\mathcal{M}(A)$ — одна из норм матрицы A, а модуль любого собственного значения матрицы не превосходит любой ее нормы, из неравенства (1.2) следует $|\eta|\leqslant(1+\gamma)\mathcal{M}(A)$. Прежде чем огрубить оценку (5.14) с помощью полученных неравенств, оценим последнее слагаемое в ее правой части. Предполагая, что матрица A нормирована, т. е. $\mathcal{M}(A)=O(1)$, и для характеристик вычислительной машины ε_1 и ε_2 выполнено соотношение $\varepsilon_2\ll\varepsilon_1$, оценим это слагаемое следующим образом: $[4+\varepsilon_1(5+2\varepsilon_1)]\varepsilon_2/\alpha^2\leqslant\varepsilon_1\mathcal{M}(A)$. Из (5.14) теперь следует оценка $|(\bar{d}_i')_{\text{м}}-\bar{d}_i|\leqslant 2(2\Delta+\varepsilon_1)(1+\tilde{\beta})(1+\tilde{\alpha}+\gamma)\mathcal{M}(A)$. Вводя обозначение

$$\tilde{\Delta}=2(2\Delta+\varepsilon_1)(1+\tilde{\beta})(1+\tilde{\alpha}+\gamma), \quad (5.15)$$

эту оценку окончательно запишем в виде

$$|(\bar{d}_i')_{\text{м}}-\bar{d}_i|\leqslant\tilde{\Delta}\mathcal{M}(A). \quad (5.16)$$

Обозначим теперь через $(\bar{b}_i')_{\text{м}}$ результат машинного вычисления величины $\bar{b}_i'$ по формуле (5.1) и сформулируем алгоритм этого вычисления следующим образом:

1) $x=(m(s_i')\,m(b_{i+1})/m(s_{i+1}'))_{\text{м}}$;
2) $j=p(s_i')+p(b_{i+1})-p(s_{i+1}')$;
3) $(\bar{b}_i')_{\text{м}}=(x\alpha^j)_{\text{м}}$.

Учтем погрешности, возникающие при реализации этого алгоритма на вычислительной машине. Поскольку вынесение порядков, используемое в алгоритме, позволяет при выполнении п. 1 избежать исчезновения порядка, погрешность выполнения этого пункта оценивается совсем просто:

$$\left|x-\frac{m(s_i')\,m(b_{i+1})}{m(s_{i+1}')}\right| \leqslant \varepsilon_1(2+\varepsilon_1)\left|\frac{m(s_i')\,m(b_{i+1})}{m(s_{i+1}')}\right|.$$

Принимая во внимание формулу (5.1) для $\bar{b}'_i$, легко проверить, что полученная оценка после умножения обеих ее частей на α^j принимает вид

$$|x\alpha^j - \bar{b}'_i| \leqslant \varepsilon_1 (2 + \varepsilon_1) |\bar{b}'_i|. \tag{5.17}$$

Погрешность выполнения п. 3 оценивается точно так же, как в алгоритме вычисления величины $\bar{d}'_i$. В результате получим оценку

$$|(\bar{b}'_i)_м - x\alpha^j| < \frac{\varepsilon_2}{\alpha^2}. \tag{5.18}$$

Из (5.17) и (5.18) легко следует $|(\bar{b}'_i)_м - \bar{b}'_i| \leqslant \varepsilon_1 (2 + \varepsilon_1) |\bar{b}'_i| + \varepsilon_2/\alpha^2$. Из этой оценки, в свою очередь, с помощью (5.6) нетрудно получить

$$|(\bar{b}'_i)_м - \bar{b}'_i| \leqslant \varepsilon_1 (2 + \varepsilon_1)(1 + \hat{\omega}) |\bar{b}_i| + \varepsilon_2/\alpha^2. \tag{5.19}$$

Теперь можно оценить отличие вычисленной величины $(\bar{b}'_i)_м$ от $\bar{b}_i$. Используя еще раз оценки (5.6) и (5.19), получаем $|(\bar{b}'_i)_м - \bar{b}_i| \leqslant [\varepsilon_1 (2 + \varepsilon_1)(1 + \hat{\omega}) + \hat{\omega}] |\bar{b}_i| + \varepsilon_2/\alpha^2$. Вводя обозначение

$$\tilde{\omega} = \varepsilon_1 (2 + \varepsilon_1)(1 + \hat{\omega}) + \hat{\omega}, \tag{5.20}$$

полученную оценку окончательно запишем в виде

$$|(\bar{b}'_i)_м - \bar{b}_i| \leqslant \tilde{\omega} |\bar{b}_i| + \varepsilon_2/\alpha^2. \tag{5.21}$$

Имея оценки погрешностей, допускаемых при вычислении каждого из элементов $\bar{d}_i$, $\bar{b}_i$ матрицы $\bar{A}$, можно оценить погрешность всей вычисленной матрицы, т. е. матрицы с элементами $(\bar{d}'_i)_м$ и $(\bar{b}'_i)_м$. Обозначим эту матрицу через $(\bar{A})_м$ и оценим норму разности $(\bar{A})_м - \bar{A}$. Для этого матрицу $(\bar{A})_м - \bar{A}$ представим в виде суммы двух матриц R_1 и R_2:

$$(\bar{A})_м - \bar{A} = R_1 + R_2, \tag{5.22}$$

где $R_1 = \operatorname{diag}[(\bar{d}'_1)_м - \bar{d}_1,\ (\bar{d}'_2)_м - \bar{d}_2,\ \ldots,\ (\bar{d}'_{m-1})_м - \bar{d}_{m-1},\ 0]$, а матрица R_2 имеет вид

$$R_2 = \begin{bmatrix} 0 & (\bar{b}'_2)_м - \bar{b}_2 & & & & \\ (\bar{b}'_2)_м - \bar{b}_2 & 0 & (\bar{b}'_3)_м - \bar{b}_3 & & 0 & \\ & \ddots & \ddots & \ddots & & \\ & & (\bar{b}'_{m-2})_м - \bar{b}_{m-2} & 0 & (\bar{b}'_{m-1})_м - \bar{b}_{m-1} & \\ & 0 & & (\bar{b}'_{m-1})_м - \bar{b}_{m-1} & 0 & 0 \\ & & & & 0 & 0 \end{bmatrix}.$$

В силу (5.22)

$$\|(\bar{A})_м - \bar{A}\| \leqslant \|R_1\| + \|R_2\|, \tag{5.23}$$

следовательно, достаточно оценить нормы матриц R_1 и R_2. Учитывая (5.16), оценим норму матрицы R_1:

$$\|R_1\| = \max_{1 \leqslant i \leqslant m-1} |(\bar{d}'_i)_м - \bar{d}_i| \leqslant \tilde{\Delta} \mathscr{M}(A).$$

Для оценки нормы матрицы R_2 воспользуемся оценками (2.30) из [4] и (5.21) из настоящего параграфа:

$$\|R_2\| \leqslant \max_{2 \leqslant i \leqslant m-2} \left(|(\bar{b}'_i)_м - \bar{b}_i| + |(\bar{b}'_{i+1})_м - \bar{b}_{i+1}| \right) \leqslant$$

$$\leqslant \tilde{\omega} \max_{2 \leqslant i \leqslant m-2} \left(|\bar{b}_i| + |\bar{b}_{i+1}| \right) + \frac{2\varepsilon_2}{\alpha^2}.$$

Введем в рассмотрение норму

$$\mathscr{M}(\bar{A}) = \max \begin{cases} |\bar{d}_1| + |\bar{b}_2|, \\ \max\limits_{2 \leqslant i \leqslant m-2} (|\bar{b}_i| + |\bar{d}_i| + |\bar{b}_{i+1}|), \\ |\bar{d}_{m-1}| + |\bar{b}_{m-1}|, \\ |\eta| \end{cases}$$

матрицы $\bar{A}$, аналогичную норме $\mathscr{M}(A)$ матрицы A. Используя эту норму, оценку нормы матрицы R_2 можно огрубить следующим образом: $\|R_2\| \leqslant \tilde{\omega}\mathscr{M}(\bar{A}) + 2\varepsilon_2/\alpha^2$. Предполагая выполненным неравенство $2\varepsilon_2/\alpha^2 \leqslant \varepsilon_1 \mathscr{M}(\bar{A})$, еще раз огрубим полученную оценку

$$\|R_2\| \leqslant (\tilde{\omega} + \varepsilon_1)\mathscr{M}(A). \tag{5.24}$$

Здесь использовано уже упоминавшееся предположение о нормированности матрицы A, $\mathscr{M}(A) = O(1)$, вследствие которого норма $\mathscr{M}(\bar{A})$ также не может быть слишком малой.

Легко показать, что спектральная норма матрицы $\bar{A}$ и норма $\mathscr{M}(\bar{A})$ связаны соотношением

$$\mathscr{M}(\bar{A}) \leqslant 2\|\bar{A}\|. \tag{5.25}$$

Действительно, очевидно, что $\|\bar{A}\| \geqslant \sqrt{\bar{b}_i^2 + \bar{d}_i^2 + \bar{b}_{i+1}^2}$. Но $\sqrt{\bar{b}_i^2 + \bar{d}_i^2 + \bar{b}_{i+1}^2} \geqslant (1/2)(|\bar{b}_i| + |\bar{d}_i| + |\bar{b}_{i+1}|)$. Следовательно, $|\bar{b}_i| + |\bar{d}_i| + |\bar{b}_{i+1}| \leqslant 2\|\bar{A}\|$. Аналогично можно показать, что $|\bar{d}_1| + |\bar{b}_2| \leqslant 2\|\bar{A}\|$, $|\bar{d}_{m-1}| + |\bar{b}_{m-1}| \leqslant 2\|\bar{A}\|$. Кроме того, очевидно, что $|\eta| \leqslant 2\|\bar{A}\|$. Следовательно, неравенство (5.25) выполнено. Огрубляя с его помощью оценку (5.24), получим

$$\|R_2\| \leqslant 2(\tilde{\omega} + \varepsilon_1)\|\bar{A}\|. \tag{5.26}$$

Оценим теперь величину $\|\bar{A}\|$ через величину $\mathscr{M}(A)$. Для этого воспользуемся оценкой (1.8), которая в силу равенства $CAC^* = \bar{A} + H$ записывается в виде

$$\|\bar{A} - CAC^*\| \leqslant 2(\tilde{\alpha} + \sqrt{m}\tilde{\beta})\mathscr{M}(A). \tag{5.27}$$

Отсюда следует, что $\|\bar{A}\| \leqslant \|A\| + 2(\tilde{\alpha} + \sqrt{m}\tilde{\beta})\mathscr{M}(A)$. Но спектральная норма симметрической матрицы не превосходит любой другой ее нормы. Следовательно, $\|\bar{A}\| \leqslant [1 + 2(\tilde{\alpha} + \sqrt{m}\tilde{\beta})]\mathscr{M}(A)$. Оценку (5.26) теперь можно огрубить следующим образом: $\|R_2\| \leqslant 2(\tilde{\omega} + \varepsilon_1)[1 + 2(\tilde{\alpha} + \sqrt{m}\tilde{\beta})] \times \mathscr{M}(A)$. Вводя обозначение

$$\tilde{\tilde{\Delta}} = 2(\tilde{\omega} + \varepsilon_1)[1 + 2(\tilde{\alpha} + \sqrt{m}\tilde{\beta})], \tag{5.28}$$

полученную оценку окончательно запишем в виде $\|R_2\| \leqslant \tilde{\tilde{\Delta}}\mathscr{M}(A)$. Подставляя полученные оценки норм матриц R_1 и R_2 в (5.23), находим

$$\|(\bar{A})_{\text{м}} - \bar{A}\| \leqslant (\tilde{\Delta} + \tilde{\tilde{\Delta}})\mathscr{M}(A). \tag{5.29}$$

Это и есть требуемая оценка.

В заключение приведем элементарно следующую из (5.27) и (5.29) оценку $\|(\bar{A})_{\text{м}} - CAC^*\| \leqslant [2(\tilde{\alpha} + \sqrt{m}\tilde{\beta}) + \tilde{\Delta} + \tilde{\tilde{\Delta}}]\mathscr{M}(A)$, которая показывает, насколько отличается вычисленная матрица $(\bar{A})_{\text{м}}$ от матрицы CAC^*, ортогонально подобной исходной матрице A. Таким образом, доказана

Теорема 3. *Пусть параметры c_i и s_i, определяющие ортогональные преобразования вращения, и задаваемая с их помощью матрица C опре-*

делены так же, как в § 1. Пусть, далее, $(\bar{A})_{\text{м}}$ *обозначает матрицу*

$$(\bar{A})_{\text{м}} = \begin{bmatrix} (\bar{d}'_1)_{\text{м}} & (\bar{b}'_2)_{\text{м}} & & & & \\ (\bar{b}'_2)_{\text{м}} & (\bar{d}'_2)_{\text{м}} & (\bar{b}'_3)_{\text{м}} & & 0 & \\ & \ddots & \ddots & \ddots & & \\ & & & (\bar{b}'_{m-2})_{\text{м}} & (\bar{d}'_{m-2})_{\text{м}} & (\bar{b}'_{m-1})_{\text{м}} \\ & 0 & & & (\bar{b}'_{m-1})_{\text{м}} & (\bar{d}'_{m-1})_{\text{м}} \; 0 \\ & & & & 0 & \eta \end{bmatrix},$$

элементы которой получены в результате машинного вычисления по формулам

$$\bar{d}'_i = d_{i+1} - \frac{c'_{i+1}c'_i b_{i+1}}{s'_{i+1}} + \frac{c'_{i+2}c'_{i+1}b_{i+2}}{s'_{i+2}};$$

$$\bar{b}'_i = \frac{s'_i b_{i+1}}{s'_{i+1}}.$$

Здесь c'_i *и* s'_i *— имеющиеся в памяти машины приближения к точным значениям параметров* c_i *и* s_i, *связанные с ними равенствами* $c'_i = c_i(1+\varkappa_i)$, $s'_i = s_i(1+\omega_i)$, *где* $|\varkappa_i| \leqslant \varkappa \ll 1$, $|\omega_i| \leqslant \omega \ll 1$. *Тогда имеет место оценка* $\|(\bar{A})_{\text{м}} - CAC^*\| \leqslant [2(\tilde{\alpha} + \sqrt{m}\tilde{\beta}) + \tilde{\Delta} + \tilde{\tilde{\Delta}}]\, \mathcal{M}(A)$.

В этой оценке константы $\tilde{\alpha}$ и $\tilde{\beta}$ определяются точностью решения уравнений (1.5) для параметров c_i и s_i, а константы $\tilde{\Delta}$ и $\tilde{\tilde{\Delta}}$ — в соответствии с (5.15) и (5.28) и зависят в основном от величины значений $\varkappa$ и ω, а также от точности вычисления элементов $\bar{d}'_i$ и $\bar{b}'_i$ по выписанным формулам.

§ 6. Машинное вычисление преобразованной двухдиагональной матрицы

В этом параграфе описано вычисление элементов $\bar{a}_i$, $\bar{b}_i$ матрицы (2.8) $\bar{A}$ по формулам

$$\bar{a}_i = \frac{s_{i+1}a_{i+1}}{\bar{s}_{i+1}}, \quad i = 1, \ldots, N-1;$$
$$\bar{b}_i = \frac{\bar{s}_i b_{i+1}}{s_{i+1}}, \quad i = 2, \ldots, N-1, \tag{6.1}$$

а также анализируются возникающие при этом погрешности. В первую очередь, они определяются тем, что вместо точных значений параметров s_i и $\bar{s}_i$, используемых в формулах (6.1), в памяти машины имеются лишь некоторые их приближения s'_i и $\bar{s}'_i$. Поэтому реальные вычисления можно вести только по формулам

$$\bar{a}'_i = \frac{s'_{i+1}a_{i+1}}{\bar{s}'_{i+1}}, \quad i = 1, \ldots, N-1;$$
$$\bar{b}'_i = \frac{\bar{s}'_i b_{i+1}}{s'_{i+1}}, \quad i = 2, \ldots, N-1, \tag{6.2}$$

а не по формулам (6.1).

Чтобы оценить возникающие при таком изменении формул погрешности, вспомним, что в § 4 разработан простой алгоритм вычисления величин s'_i и $\bar{s}'_i$, гарантирующий выполнение равенств, связывающих эти

вычисленные значения с точными значениями s_i и $\bar{s}_i$,

$$s_i' = s_i(1+\pi_i), \quad \bar{s}_i' = \bar{s}_i(1+\bar{\pi}_i). \tag{6.3}$$

Там же показано, что данные равенства имеют место при некоторых значениях величин π_i и $\bar{\pi}_i$ таких, что

$$|\pi_i| \leqslant \pi, \quad |\bar{\pi}_i| \leqslant \pi, \tag{6.4}$$

где в качестве значения π можно взять $\pi = 5{,}001\varepsilon_1$. Для дальнейшего отметим также, что в упомянутом алгоритме каждая из величин s_i' и $\bar{s}_i'$ определяется с помощью двух чисел: целочисленного порядка и мантиссы, так что

$$s_i' = m(s_i')\,\alpha^{p(s_i')}, \quad \bar{s}_i' = m(\bar{s}_i')\,\alpha^{p(\bar{s}_i')}, \tag{6.5}$$

где $p(s_i')$ и $p(\bar{s}_i')$ — порядки чисел s_i' и $\bar{s}_i'$ соответственно, а $m(s_i')$ и $m(\bar{s}_i')$ — их мантиссы, удовлетворяющие неравенствам

$$1/\alpha \leqslant |m(s_i')| < 1, \quad 1/\alpha \leqslant |m(\bar{s}_i')| < 1. \tag{6.6}$$

Наконец, отметим, что из (6.5) и (6.6) следует отличие от нуля чисел s_i' и $\bar{s}_i'$.

Теперь уже можно оценить погрешности, обусловленные заменой точных формул (6.1) формулами (6.2). В силу равенств (6.3) ясно, что для величин $\bar{a}_i' - \bar{a}_i$ и $\bar{b}_i' - \bar{b}_i$ имеют место выражения

$$\bar{a}_i' - \bar{a}_i = \left(\frac{1+\pi_{i+1}}{1+\bar{\pi}_{i+1}} - 1\right)\bar{a}_i, \quad \bar{b}_i' - \bar{b}_i = \left(\frac{1+\bar{\pi}_i}{1+\pi_{i+1}} - 1\right)\bar{b}_i.$$

С учетом неравенств (6.4) из этих формул следуют оценки $|\bar{a}_i' - \bar{a}_i| \leqslant [2\pi/(1-\pi)]\,|\bar{a}_i|$, $|\bar{b}_i' - \bar{b}_i| \leqslant [2\pi/(1-\pi)]\,|\bar{b}_i|$. Вводя обозначение

$$\hat{\pi} = \frac{2\pi}{1-\pi}, \tag{6.7}$$

перепишем полученные оценки в виде

$$|\bar{a}_i' - \bar{a}_i| \leqslant \hat{\pi}|\bar{a}_i|, \quad |\bar{b}_i' - \bar{b}_i| \leqslant \hat{\pi}|\bar{b}_i|. \tag{6.8}$$

Вторым источником погрешностей, возникающих при вычислении элементов матрицы $\bar{A}$, является непосредственное выполнение арифметических операций, предписываемых формулами (6.2). Во избежание переполнения, а также исчезновения порядка, возможных при выполнении промежуточных операций, предлагается вести вычисления по специальным алгоритмам. Представив числа a_{i+1} и b_{i+1} в виде

$$a_{i+1} = m(a_{i+1})\,\alpha^{p(a_{i+1})}, \quad \frac{1}{\alpha} \leqslant |m(a_{i+1})| < 1;$$

$$b_{i+1} = m(b_{i+1})\,\alpha^{p(b_{i+1})}, \quad \frac{1}{\alpha} \leqslant |m(b_{i+1})| < 1,$$

алгоритмы вычисления $\bar{a}_i'$ и $\bar{b}_i'$ сформулируем следующим образом:

$$\begin{aligned} &1)\ x = \left(m(\bar{s}_{i+1}')\,m(a_{i+1})/m(\bar{s}_{i+1}')\right)_{\text{м}};\\ &2)\ j = p(\bar{s}_{i+1}') + p(a_{i+1}) - p(\bar{s}_{i+1}');\\ &3)\ (\bar{a}_i')_{\text{м}} = (x\alpha^j)_{\text{м}}; \end{aligned} \tag{6.9}$$

$$\begin{aligned} &1)\ y = \left(m(\bar{s}_i')\,m(b_{i+1})/m(\bar{s}_{i+1}')\right)_{\text{м}};\\ &2)\ k = p(\bar{s}_i') + p(b_{i+1}) - p(\bar{s}_{i+1}');\\ &3)\ (\bar{b}_i')_{\text{м}} = (y\alpha^k)_{\text{м}}, \end{aligned} \tag{6.10}$$

Здесь через $(\bar{a}'_i)_м$ и $(\bar{b}'_i)_м$ обозначены результаты машинного вычисления величин $\bar{a}'_i$ и $\bar{b}'_i$ по сформулированным алгоритмам.

Учтем погрешности, возникающие при реализации алгоритма (6.9) на вычислительной машине. Нетрудно проверить, что погрешность выполнения п. 1 оценивается так:

$$\left| x - \frac{m(s'_{i+1})\, m(a_{i+1})}{m(\bar{s}'_{i+1})} \right| \leqslant \varepsilon_1 (2 + \varepsilon_1) \left| \frac{m(s'_{i+1})\, m(a_{i+1})}{m(\bar{s}'_{i+1})} \right|.$$

Принимая во внимание формулы (6.2) и умножая обе части полученной оценки на α^j, перепишем ее в виде $|x\alpha^j - \bar{a}'_i| \leqslant \varepsilon_1 (2 + \varepsilon_1) |\bar{a}'_i|$. Пункт 2, очевидно, выполняется точно. При выполнении же п. 3 может возникнуть только погрешность исчезновения порядка. Эта погрешность появляется в том случае, если $|x\alpha^j| < \varepsilon_2/\alpha^2$. При этом величина $(\bar{a}'_i)_м$ полагается равной нулю, а возникающая из-за этого погрешность, очевидно, не превосходит ε_2/α^2: $|(\bar{a}'_i)_м - x\alpha^j| \leqslant \varepsilon_2/\alpha^2$. Окончательно, оценка погрешности, возникающей при выполнении алгоритма (6.9), примет вид

$$|(\bar{a}'_i)_м - \bar{a}'_i| \leqslant \varepsilon_1 (2 + \varepsilon_1) |\bar{a}'_i| + \varepsilon_2/\alpha^2. \tag{6.11}$$

Для погрешности выполнения алгоритма (6.10), очевидно, имеет место аналогичная оценка

$$|(\bar{b}'_i)_м - \bar{b}'_i| \leqslant \varepsilon_1 (2 + \varepsilon_1) |\bar{b}'_i| + \varepsilon_2/\alpha^2. \tag{6.12}$$

В силу (6.8) из (6.11) и (6.12) легко следуют оценки

$$\begin{aligned} |(\bar{a}'_i)_м - \bar{a}'_i| &\leqslant \varepsilon_1 (2 + \varepsilon_1)(1 + \hat{\pi}) |\bar{a}_i| + \varepsilon_2/\alpha^2; \\ |(\bar{b}'_i)_м - \bar{b}'_i| &\leqslant \varepsilon_1 (2 + \varepsilon_1)(1 + \hat{\pi}) |\bar{b}_i| + \varepsilon_2/\alpha^2. \end{aligned} \tag{6.13}$$

Теперь уже можно достаточно просто оценить отличие вычисленных элементов $(\bar{a}'_i)_м$ и $(\bar{b}'_i)_м$ от $\bar{a}_i$ и $\bar{b}_i$ соответственно. Действительно, из (6.8) и (6.13) легко находим

$$\begin{aligned} |(\bar{a}'_i)_м - \bar{a}_i| &\leqslant [\varepsilon_1 (2 + \varepsilon_1)(1 + \hat{\pi}) + \hat{\pi}] |\bar{a}_i| + \varepsilon_2/\alpha^2; \\ |(\bar{b}'_i)_м - \bar{b}_i| &\leqslant [\varepsilon_1 (2 + \varepsilon_1)(1 + \hat{\pi}) + \hat{\pi}] |\bar{b}_i| + \varepsilon_2/\alpha^2. \end{aligned}$$

После введения обозначения

$$\tilde{\pi} = \varepsilon_1 (2 + \varepsilon_1)(1 + \hat{\pi}) + \hat{\pi} \tag{6.14}$$

перепишем полученные оценки

$$|(\bar{a}'_i)_м - \bar{a}_i| \leqslant \tilde{\pi} |\bar{a}_i| + \frac{\varepsilon_2}{\alpha^2}, \quad |(\bar{b}'_i)_м - \bar{b}_i| \leqslant \tilde{\pi} |\bar{b}_i| + \frac{\varepsilon_2}{\alpha^2}. \tag{6.15}$$

Обозначим теперь через $(\bar{A})_м$ вычисленную матрицу, т. е. матрицу с элементами $(\bar{a}'_i)_м$, $(\bar{b}'_i)_м$, и оценим ее отличие от матрицы $\bar{A}$. Для этого воспользуемся оценкой (2.30) из [4] и только что полученными оценками (6.15):

$$\begin{aligned} \|(\bar{A})_м - \bar{A}\| &\leqslant \max \left\{ \begin{matrix} \max\limits_{1 \leqslant i \leqslant N-2} (|(\bar{a}'_i)_м - \bar{a}_i| + |(\bar{b}'_{i+1})_м - \bar{b}_{i+1}|), \\ \max\limits_{2 \leqslant i \leqslant N-1} (|(\bar{a}'_i)_м - \bar{a}_i| + |(\bar{b}'_i)_м - \bar{b}_i|) \end{matrix} \right\} \leqslant \\ &\leqslant \tilde{\pi} \max \left\{ \begin{matrix} \max\limits_{1 \leqslant i \leqslant N-2} (|\bar{a}_i| + |\bar{b}_{i+1}|), \\ \max\limits_{2 \leqslant i \leqslant N-1} (|\bar{a}_i| + |\bar{b}_i|) \end{matrix} \right\} + \frac{2\varepsilon_2}{\alpha^2}. \end{aligned} \tag{6.16}$$

На основе таких же рассуждений, с помощью которых получены неравенства (2.24), нетрудно показать, что $|\bar{a}_i| + |\bar{b}_{i+1}| \leqslant \sqrt{2}\|\bar{A}\|$, $|\bar{a}_i| + |\bar{b}_i| \leqslant$

$\leqslant \sqrt{2}\|\bar{A}\|$. Тогда из (6.16) легко следует

$$\|(\bar{A})_{\text{м}} - \bar{A}\| \leqslant \sqrt{2}\,\tilde{\pi}\|\bar{A}\| + \frac{2\varepsilon_2}{\alpha^2}. \tag{6.17}$$

Из оценки (2.9), которую можно переписать в виде

$$\|\bar{A} - \bar{C}AC^*\| \leqslant 2\sqrt{2}(\sqrt{N}+2)\varepsilon\|A\|, \tag{6.18}$$

следует, что $\|\bar{A}\| \leqslant [1 + 2\sqrt{2}(\sqrt{N}+2)\varepsilon]\|A\|$. Подставляя последнюю оценку в (6.17), получим $\|(\bar{A})_{\text{м}} - \bar{A}\| \leqslant \sqrt{2}\,\tilde{\pi}[1 + 2\sqrt{2}(\sqrt{N}+2)\varepsilon]\|A\| + 2\varepsilon_2/\alpha^2$. Наконец, предположим, что выполнено неравенство $2\varepsilon_2/\alpha^2 \leqslant \varepsilon_1\|A\|$. Это предположение опирается на условие нормированности матрицы A и неравенство $\varepsilon_2 \ll \varepsilon_1$. Тогда оценку нормы матрицы $(\bar{A})_{\text{м}} - \bar{A}$ можно записать так:

$$\|(\bar{A})_{\text{м}} - \bar{A}\| \leqslant (\sqrt{2}\,\tilde{\pi} + \varepsilon_1)[1 + 2\sqrt{2}(\sqrt{N}+2)\varepsilon]\|A\|. \tag{6.19}$$

В заключение приведем оценку

$$\|(\bar{A})_{\text{м}} - \bar{C}AC^*\| \leqslant \{2\sqrt{2}(\sqrt{N}+2)\varepsilon + (\sqrt{2}\,\tilde{\pi} + \varepsilon_1)[1 + 2\sqrt{2}(\sqrt{N}+2)\varepsilon]\}\|A\|,$$

которая элементарно следует из (6.18) и (6.19) и показывает, насколько вычисленная матрица $(\bar{A})_{\text{м}}$ отличается от матрицы $\bar{C}AC^*$, ортогонально эквивалентной исходной матрице A. Таким образом, нами доказана

Теорема 4. *Пусть параметры c_i, s_i, $\bar{c}_i$ и $\bar{s}_i$, определяющие преобразования вращения, и задаваемые с их помощью ортогональные матрицы C и $\bar{C}$ определены так же, как в* § 2. *Пусть, далее, $(\bar{A})_{\text{м}}$ обозначает матрицу*

$$(\bar{A})_{\text{м}} = \begin{bmatrix} (\bar{a}'_1)_{\text{м}} & (\bar{b}'_2)_{\text{м}} & & & \\ & (\bar{a}'_2)_{\text{м}} & (\bar{b}'_3)_{\text{м}} & & 0 \\ & & \ddots & \ddots & \\ & & & (\bar{a}'_{N-2})_{\text{м}} & (\bar{b}'_{N-1})_{\text{м}} \\ & 0 & & & (\bar{a}'_{N-1})_{\text{м}} \quad 0 \\ & & & & \sigma \end{bmatrix},$$

элементы $(\bar{a}'_i)_{\text{м}}$, $(\bar{b}'_i)_{\text{м}}$ которой получены в результате машинного вычисления по формулам

$$\bar{a}'_i = \frac{\bar{s}'_{i+1}a_{i+1}}{\bar{s}'_{i+1}}, \quad \bar{b}'_i = \frac{\bar{s}'_i b_{i+1}}{s'_{i+1}}.$$

Здесь s'_i и $\bar{s}'_i$ — имеющиеся в памяти машины приближения к точным значениям параметров s_i и $\bar{s}_i$, связанные с ними равенствами $s'_i = s_i(1+\pi_i)$, $\bar{s}'_i = \bar{s}_i(1+\bar{\pi}_i)$, которые имеют место при некоторых значениях π_i и $\bar{\pi}_i$ таких, что

$$|\pi_i| \leqslant \pi \ll 1, \quad |\bar{\pi}_i| \leqslant \pi. \tag{6.20}$$

Тогда справедлива оценка

$$\|(\bar{A})_{\text{м}} - \bar{C}AC^*\| \leqslant \{2\sqrt{2}(\sqrt{N}+2)\varepsilon + (\sqrt{2}\,\tilde{\pi} + \varepsilon_1)[1 + 2\sqrt{2}(\sqrt{N}+2)\varepsilon]\}\|A\|.$$

В этой оценке константа ε определяется точностью решения уравнений (2.5) относительно параметров c_i, s_i, $\bar{c}_i$ и $\bar{s}_i$, а константа $\tilde{\pi}$ определяется по формуле (6.14) и зависит от значения π, с которым удовлетворяются неравенства (6.20), а также от точности вычисления элементов $\bar{a}'_i$, $\bar{b}'_i$ по приведенным формулам.

Глава 4

ОБЩИЕ СХЕМЫ АЛГОРИТМОВ ИСЧЕРПЫВАНИЯ ТРЕХДИАГОНАЛЬНЫХ СИММЕТРИЧЕСКИХ И ДВУХДИАГОНАЛЬНЫХ МАТРИЦ

§ 7. Общая схема исчерпывания симметрической трехдиагональной матрицы

Рассмотрим симметрическую трехдиагональную матрицу

$$\tilde{A}=\begin{bmatrix} \tilde{d}_1 & \tilde{b}_2 & & & \\ \tilde{b}_2 & \tilde{d}_2 & \tilde{b}_3 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \tilde{b}_{m-1} & \tilde{d}_{m-1} & \tilde{b}_m \\ 0 & & & \tilde{b}_m & \tilde{d}_m \end{bmatrix}$$

такую, что

$$\mathscr{M}(\tilde{A})=\max\left\{\begin{array}{c} |\tilde{d}_1|+|\tilde{b}_2|, \\ \max\limits_{2<i<m-1}(|\tilde{b}_i|+|\tilde{d}_i|+|\tilde{b}_{i+1}|), \\ |\tilde{d}_m|+|\tilde{b}_m| \end{array}\right\}=O(1),$$

т. е. матрица $\tilde{A}$ предполагается нормированной. Отметим, что относительно величины элементов $\tilde{b}_i$ ее побочных диагоналей никаких предположений не делается.

Пусть $\lambda(\tilde{A})$ — одно из собственных значений матрицы $\tilde{A}$. В этом параграфе опишем общую схему исчерпывания собственного значения $\lambda(\tilde{A})$ и приведем оценку погрешности, возникающей при этом исчерпывании. Предположим, что η — некоторое приближение к $\lambda(\tilde{A})$ такое, что

$$|\eta-\lambda(\tilde{A})|\leqslant\tilde{\gamma}\mathscr{M}(\tilde{A}). \tag{7.1}$$

Отметим, что приближенное собственное значение η можно получить, например, с помощью алгоритма бисекций, основанного на применении теоремы Штурма. В [5], где этот алгоритм подробно описан и исследован, показано (см. [5], § 10), что при $\mathscr{M}(\tilde{A})=O(1)$ значение η, получаемое в результате работы алгоритма, удовлетворяет оценке (7.1) с $\tilde{\gamma}=6\varepsilon_1$.

Наряду с матрицей $\tilde{A}$ введем в рассмотрение

$$A=\begin{bmatrix} d_1 & b_2 & & & \\ b_2 & d_2 & b_3 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & b_{m-1} & d_{m-1} & b_m \\ 0 & & & b_m & d_m \end{bmatrix},$$

положив $d_i=\tilde{d}_i$, а элементы b_i ее побочных диагоналей определив следующим образом:

$$b_i=\begin{cases} \tilde{b}_i, \text{ если } |\tilde{b}_i|\geqslant\varepsilon_1\mathscr{M}(\tilde{A}); \\ \text{иначе} \begin{cases} \varepsilon_1\mathscr{M}(\tilde{A}), \text{ если } \tilde{b}_i>0; \\ -\varepsilon_1\mathscr{M}(\tilde{A}), \text{ если } \tilde{b}_i\leqslant 0. \end{cases} \end{cases}$$

Легко видеть, что

$$(1-2\varepsilon_1)\mathscr{M}(\tilde{A}) \leqslant \mathscr{M}(A) \leqslant (1+2\varepsilon_1)\mathscr{M}(\tilde{A}). \tag{7.2}$$

Кроме того, нетрудно показать, что соответствующие собственные значения матриц A и $\tilde{A}$ различаются не более чем на $2\varepsilon_1\mathscr{M}(\tilde{A})$. Отсюда и из оценки (7.1) следует, что η является приближением к некоторому собственному значению $\lambda(A)$ матрицы A, удовлетворяющим оценке

$$|\eta - \lambda(A)| \leqslant \gamma\mathscr{M}(\tilde{A}), \tag{7.3}$$

где

$$\gamma = \tilde{\gamma} + 2\varepsilon_1. \tag{7.4}$$

С помощью алгоритма, детально описанного в § 3 работы [4], определим виртуальный почти собственный вектор $u = (u_1, u_2, \ldots, u_m)^T$ матрицы A, отвечающий ее почти собственному значению η. Напомним, что вектор называется виртуальным, если он задается с помощью не своих компонент, а их последовательных отношений. В используемом алгоритме виртуальный почти собственный вектор u определяется двумя последовательностями $\{\mathscr{R}_i\}$, $\{Q_i\}$ и целым числом i_0 так, что отношения $u_2/u_1, u_3/u_2, \ldots, u_{i_0-1}/u_{i_0-2}$ равны соответственно числам $\mathscr{R}_2, \mathscr{R}_3, \ldots, \mathscr{R}_{i_0-1}$, а отношения u_{i_0}/u_{i_0-1}, $u_{i_0+1}/u_{i_0}, \ldots, u_m/u_{m-1}$ — числам $1/Q_{i_0}$, $1/Q_{i_0+1}, \ldots, 1/Q_m$.

Определим $\overline{\mathscr{P}}_2, \overline{\mathscr{P}}_3, \ldots, \overline{\mathscr{P}}_m$ в соответствии со следующим правилом: $\overline{\mathscr{P}}_i = \mathscr{R}_i$, $i = 2, \ldots, i_0 - 1$; $\overline{\mathscr{P}}_i = (1+\psi_i)/Q_i$, $i = i_0, \ldots, m$. Здесь ψ_i моделируют погрешности делений $1/Q_i$, поэтому, очевидно, $|\psi_i| \leqslant \varepsilon_1$. Согласно теореме 2 из § 2 работы [4], числа $\overline{\mathscr{P}}_i$ удовлетворяют соотношениям

$$d_1 - \eta + \bar{\alpha}_1 + b_2\overline{\mathscr{P}}_2 = 0;$$

$$\frac{b_i(1+\bar{\beta}_i)}{\overline{\mathscr{P}}_i} + d_i - \eta + \bar{\alpha}_i + b_{i+1}\overline{\mathscr{P}}_{i+1} = 0, \quad i = 2, \ldots, m-1,$$

$$\frac{b_m(1+\bar{\beta}_m)}{\overline{\mathscr{P}}_m} + d_m - \eta + \bar{\alpha}_m = 0,$$

где для возмущений $\bar{\alpha}_i$ и $\bar{\beta}_i$ имеют место оценки

$$|\bar{\alpha}_1| \leqslant \alpha_1'(1+\varepsilon_1) + \varepsilon_1(|d_1| + |\eta|), \quad |\bar{\alpha}_m| \leqslant \alpha_m'', \quad |\bar{\beta}_m| \leqslant \varepsilon_1;$$

$$\left.\begin{aligned} |\bar{\alpha}_i| &\leqslant \max\left\{\begin{array}{c} \alpha_i'(1+\varepsilon_1) + \varepsilon_1(|d_i| + |\eta|), \\ \dfrac{\alpha_i''(1+\varepsilon_1) + (|\beta_{i+1}''| + \varepsilon_1)(|d_i| + |\eta|)}{1 - |\beta_{i+1}''|} \end{array}\right\}, \\ |\bar{\beta}_i| &\leqslant \max\left\{|\beta_i'|(1+\varepsilon_1) + \varepsilon_1, \frac{|\beta_{i+1}''| + \varepsilon_1(2+\varepsilon_1)}{1 - |\beta_{i+1}''|}\right\}, \end{aligned}\right\} \quad i = 2, \ldots, m-1. \tag{7.5}$$

Используемые здесь величины α_i', α_i'', β_i', β_i'' подчиняются, в свою очередь, следующим оценкам (см. [4], § 4):

$$\gamma\mathscr{M}(\tilde{A}) < \alpha_i' \leqslant \gamma\mathscr{M}(\tilde{A}) + 16\varepsilon_1\mathscr{M}(A), \quad |\beta_i'| \leqslant 10\varepsilon_1;$$

$$\gamma\mathscr{M}(\tilde{A}) < \alpha_i'' \leqslant \gamma\mathscr{M}(\tilde{A}) + 16\varepsilon_1\mathscr{M}(A), \quad |\beta_i''| \leqslant 10\varepsilon_1.$$

Упростим оценки (7.5). Согласно сделанному ранее замечанию относительно возможного значения величины $\tilde{\gamma}$ в оценке (7.1) примем $\tilde{\gamma} = 6\varepsilon_1$. Тогда из (7.1) следует, что в качестве значения γ можно взять $\gamma = 8\varepsilon_1$. Принимая теперь во внимание неравенства (7.2) и пренебрегая членами второго порядка малости, оценки величин $\bar{\alpha}_i$ и $\bar{\beta}_i$ запишем в виде $|\bar{\alpha}_i| \leqslant \bar{\alpha}\mathscr{M}(A)$, $|\bar{\beta}_i| \leqslant \bar{\beta}$, где в качестве значений $\bar{\alpha}$ и $\bar{\beta}$ можно взять $\bar{\alpha} = 46\varepsilon_1$, $\bar{\beta} = 12\varepsilon_1$.

С помощью чисел $\bar{\mathscr{P}}_i$ определим числа c_i, s_i, c_i' и s_i' следующим образом:

$$s_2 = \frac{1}{\sqrt{1+\bar{\mathscr{P}}_2^2}},\quad c_2 = \frac{\bar{\mathscr{P}}_2}{\sqrt{1+\bar{\mathscr{P}}_2^2}},\quad c_2' = (c_2)_{\text{м}},\quad s_2' = (s_2)_{\text{м}},$$
$$s_i = \frac{1}{\sqrt{1+(c_{i-1}'\bar{\mathscr{P}}_i)^2}},\quad c_i = \frac{c_{i-1}'\bar{\mathscr{P}}_i}{\sqrt{1+(c_{i-1}'\bar{\mathscr{P}}_i)^2}},\quad c_i' = (c_i)_{\text{м}}, \tag{7.6}$$
$$s_i' = (s_i)_{\text{м}},\quad i = 3, \ldots, m.$$

В § 4 главы 2 разработан специальный алгоритм вычисления величин c_i и s_i по приведенным формулам. Получаемые в результате применения этого алгоритма машинные числа c_i' и s_i' связаны с c_i и s_i равенствами $c_i' = c_i(1+\varkappa_i)$, $s_i' = s_i(1+\omega_i)$, которые имеют место при некоторых $\varkappa_i$ и ω_i таких, что $|\varkappa_i| \leqslant \varkappa = 5{,}001\varepsilon_1$, $|\omega_i| \leqslant \omega = 4{,}001\varepsilon_1$. Отметим, что в используемом алгоритме каждая из величин c_i' и s_i' определяется с помощью двух чисел — мантиссы и целочисленного порядка, а именно c_i' определяется числами $m(c_i')$ и $p(c_i')$, а s_i' определяется числами $m(s_i')$ и $p(s_i')$ так, что

$$c_i' = m(c_i')\,\alpha^{p(c_i')},\quad 1/\alpha \leqslant |m(c_i')| < 1;$$
$$s_i' = m(s_i')\,\alpha^{p(s_i')},\quad 1/\alpha \leqslant |m(s_i')| < 1.$$

Здесь α — основание системы счисления, принятой в вычислительной машине. Подчеркнем также, что алгоритм гарантирует отличие от нуля чисел c_i' и s_i'.

В § 3 главы 2 показано, что числа c_i и s_i удовлетворяют в таком случае соотношениям

$$s_2(d_1+\alpha_1-\eta)+c_2b_2 = 0;$$
$$s_3s_2b_2(1+\beta_2)+s_3c_2(d_2+\alpha_2-\eta)+c_3b_3 = 0;$$
$$s_{i+1}s_ic_{i-1}b_i(1+\beta_i)+s_{i+1}c_i(d_i+\alpha_i-\eta)+c_{i+1}b_{i+1} = 0,\quad i = 3, \ldots, m-1;$$
$$s_mc_{m-1}b_m(1+\beta_m)+c_m(d_m+\alpha_m-\eta) = 0,$$

где возмущения α_i и β_i определяются формулами

$$\alpha_1 = \bar{\alpha}_1,\quad \beta_2 = (1+\bar{\beta}_2)(1+\varkappa_2)-1;$$
$$\alpha_i = (d_i-\eta)\varkappa_i+\bar{\alpha}_i(1+\varkappa_i),\quad i = 2, \ldots, m-1;$$
$$\beta_i = (1+\bar{\beta}_i)(1+\varkappa_{i-1})(1+\varkappa_i)-1,\quad i = 3, \ldots, m-1;$$
$$\alpha_m = \bar{\alpha}_m,\quad \beta_m = (1+\bar{\beta}_m)(1+\varkappa_{m-1})-1.$$

Из этих равенств следуют оценки $|\alpha_i| \leqslant \tilde{\alpha}\mathscr{M}(A)$, $|\beta_i| \leqslant \tilde{\beta}$, в которых в качестве значений $\tilde{\alpha}$ и $\tilde{\beta}$ с точностью до членов первого порядка малости по ε_1 можно взять

$$\tilde{\alpha} = \bar{\alpha}+2\varkappa = 56{,}002\varepsilon_1,\quad \tilde{\beta} = \bar{\beta}+2\varkappa = 22{,}002\varepsilon_1. \tag{7.7}$$

Определим теперь матрицу

$$(\bar{A})_{\text{м}} = \begin{bmatrix} (\bar{d}_1')_{\text{м}} & (\bar{b}_2')_{\text{м}} & & & & 0 \\ (\bar{b}_2')_{\text{м}} & (\bar{d}_2')_{\text{м}} & (\bar{b}_3')_{\text{м}} & & & \\ & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & (\bar{b}_{m-2}')_{\text{м}} & (\bar{d}_{m-2}')_{\text{м}} & (\bar{b}_{m-1}')_{\text{м}} \\ 0 & & & & (\bar{b}_{m-1}')_{\text{м}} & (\bar{d}_{m-1}')_{\text{м}} \; 0 \\ & & & & & 0 \quad \eta \end{bmatrix}.$$

элементы $(\bar{d}_i')_{\text{м}}$, $(\bar{b}_i')_{\text{м}}$ которой получаются в результате машинного вычисления по формулам

$$\bar{d}_1' = d_2 - \frac{c_2' b_2}{s_2'} + \frac{c_3' c_2' b_3}{s_3'};$$

$$\bar{d}_i' = d_{i+1} - \frac{c_{i+1}' c_i' b_{i+1}}{s_{i+1}'} + \frac{c_{i+2}' c_{i+1}' b_{i+2}}{s_{i+2}'}, \quad i = 2, \ldots, m-2;$$

$$\bar{d}_{m-1}' = d_m - \frac{c_m' c_{m-1}' b_m}{s_m'},$$

$$\frac{\bar{b}_i'}{s_{i+1}'} = \frac{s_i' b_{i+1}}{s_{i+1}'}, \quad i = 2, \ldots, m-1.$$

В § 5 главы 3 детально описаны алгоритмы, с помощью которых проводятся вычисления по этим формулам.

Из рассмотрений § 5 следует, что определенная таким образом матрица $(\bar{A})_{\text{м}}$ по норме мало отличается от некоторой матрицы, подобной матрице A. Точнее говоря, из теоремы 3 следует оценка

$$\|(\bar{A})_{\text{м}} - CAC^*\| \leqslant [2(\tilde{\alpha} + \sqrt{m}\,\tilde{\beta}) + \tilde{\Delta} + \tilde{\tilde{\Delta}}]\mathcal{M}(A). \tag{7.8}$$

Напомним, что матрица C предетавляет собой произведение $C = C_m \cdots C_3 \cdot C_2$ ортогональных матриц вращения

$$C_i = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c_i & -s_i & & \\ & & & s_i & c_i & & \\ & & & & & 1 & \\ & & & & & & \ddots \\ & & & & & & & 1 \end{bmatrix},$$

которые конструируются с помощью параметров c_i и s_i $(c_i^2 + s_i^2 = 1)$, определенных формулами (7.6). Возможные значения констант $\tilde{\alpha}$ и $\tilde{\beta}$ определяются формулами (7.7), а величины $\tilde{\Delta}$ и $\tilde{\tilde{\Delta}}$ — соответственно формулами (5.15) и (5.28). Напомним эти формулы:

$$\tilde{\Delta} = 2(2\Delta + \varepsilon_1)(1 + \tilde{\beta})(1 + \tilde{\alpha} + \gamma), \quad \tilde{\tilde{\Delta}} = 2(\tilde{\omega} + \varepsilon_1)[1 + 2(\tilde{\alpha} + \sqrt{m}\,\tilde{\beta})].$$

Значения величин Δ и $\tilde{\omega}$, в свою очередь, определяются формулами (5.13) и (5.20):

$$\Delta = [(1+\varepsilon_1)^5 - 1](1 + \hat{\varkappa}) + \hat{\varkappa}, \quad \tilde{\omega} = \varepsilon_1(2 + \varepsilon_1)(1 + \hat{\omega}) + \hat{\omega}.$$

Наконец, выпишем еще формулы (5.5), согласно которым определяются величины $\hat{\varkappa}$ и $\hat{\omega}$,

$$\hat{\varkappa} = (1+\varkappa)^2/(1-\omega) - 1, \quad \hat{\omega} = 2\omega/(1-\omega).$$

Пренебрегая в последних формулах членами второго порядка малости по ε_1, в качестве значений $\hat{\varkappa}$ и $\hat{\omega}$ примем

$$\hat{\varkappa} = 2\varkappa + \omega = 14{,}003\varepsilon_1, \quad \hat{\omega} = 2\omega = 8{,}002\varepsilon_1.$$

Аналогичным образом упростим выражения для величин Δ и $\tilde{\omega}$:

$$\Delta = 5\varepsilon_1 + \hat{\varkappa} = 19{,}003\varepsilon_1, \quad \tilde{\omega} = 2\varepsilon_1 + \hat{\omega} = 10{,}002\varepsilon_1.$$

Наконец, пренебрегая членами второго порядка в формулах для $\tilde{\Delta}$ и $\tilde{\tilde{\Delta}}$, получим

$$\tilde{\Delta} = 2(2\Delta + \varepsilon_1) = 78{,}012\varepsilon_1, \quad \tilde{\tilde{\Delta}} = 2(\tilde{\omega} + \varepsilon_1) = 22{,}004\varepsilon_1. \tag{7.9}$$

Итак, оценка (7.8) имеет место при значениях констант $\tilde{\alpha}$ и $\tilde{\beta}$, определяемых формулами (7.7), и значениях констант $\tilde{\Delta}$ и $\tilde{\tilde{\Delta}}$, определяемых формулами (7.9).

В заключение оценим отличие вычисленной матрицы $(\tilde{A})_м$ от $C\tilde{A}C^*$, подобной исходной матрице $\tilde{A}$. Соответствующая оценка легко следует из оценки (7.8) и очевидной оценки $\|A - \tilde{A}\| \leqslant 2\varepsilon_1 \mathscr{M}(\tilde{A})$. Действительно,

$$\|(\tilde{A})_м - C\tilde{A}C^*\| \leqslant \|(\tilde{A})_м - CAC^*\| + \|A - \tilde{A}\| \leqslant$$

$$\leqslant [2(\tilde{\alpha} + \sqrt{m}\tilde{\beta}) + \tilde{\Delta} + \tilde{\tilde{\Delta}}]\mathscr{M}(A) + 2\varepsilon_1 \mathscr{M}(\tilde{A}).$$

Воспользовавшись еще раз неравенствами (7.2) и пренебрегая членами второго порядка, запишем полученную оценку в терминах нормы $\mathscr{M}(\tilde{A})$ исходной матрицы $\tilde{A}$

$$\|(\tilde{A})_м - C\tilde{A}C^*\| \leqslant [2(\tilde{\alpha} + \sqrt{m}\tilde{\beta}) + \tilde{\Delta} + \tilde{\tilde{\Delta}} + 2\varepsilon^1] \ \mathscr{M}(\tilde{A}).$$

Подчеркнем, что к результатам работы описанного алгоритма кроме матрицы $(\tilde{A})_м$ относится и значение выражения, стоящего в правой части полученной оценки. Кроме того, результатами работы алгоритма следует считать и числа c_i и s_i, а также величины $\varkappa$ и ω, оценивающие отличие этих чисел от точных значений параметров c_i и s_i, с помощью которых конструируется ортогональная матрица C.

§ 8. Общая схема исчерпывания двухдиагональной матрицы

Рассмотрим двухдиагональную матрицу

$$\tilde{A} = \begin{bmatrix} \tilde{a}_1 & \tilde{b}_2 & & & 0 \\ & \tilde{a}_2 & \tilde{b}_3 & & \\ & & \ddots & \ddots & \\ & & & \tilde{a}_{N-1} & \tilde{b}_N \\ 0 & & & & \tilde{a}_N \end{bmatrix},$$

относительно которой предположим, что

$$\mathscr{K}(\tilde{A}) = \max \begin{Bmatrix} \max\limits_{1 \leqslant i \leqslant N-1} (|\tilde{a}_i| + |\tilde{b}_{i+1}|), \\ \max\limits_{2 \leqslant i \leqslant N} (|\tilde{a}_i| + |\tilde{b}_i|) \end{Bmatrix} = O(1). \tag{8.1}$$

Пусть $\sigma_N(\tilde{A})$ — наибольшее сингулярное число матрицы $\tilde{A}$, т. е. $\sigma_N(\tilde{A}) = \|\tilde{A}\|$. В этом параграфе описана общая схема исчерпывания сингулярного числа $\sigma_N(\tilde{A})$ и приведена оценка погрешности, возникающей при этом исчерпывании. Заметим, что предположение (8.1) означает, что первым этапом работы алгоритма должно быть масштабирование исходной матрицы. Предположим, что σ — приближение к $\sigma_N(\tilde{A}) = \|\tilde{A}\|$ такое, что

$$|\sigma - \|\tilde{A}\|| \leqslant \tilde{\delta}\mathscr{K}(\tilde{A}). \tag{8.2}$$

Отметим, что если для получения σ воспользоваться алгоритмом бисек-

ций, подробно описанным в § 10 работы [5], то в качестве значения величины $\tilde{\delta}$ в оценке (8.2) можно взять $\tilde{\delta} = 4\varepsilon_1$.

Наряду с матрицей $\tilde{A}$ будем рассматривать матрицу

$$A = \begin{bmatrix} a_1 & b_2 & & & \\ & a_2 & b_3 & & 0 \\ & & \ddots & \ddots & \\ & 0 & & a_{N-1} & b_N \\ & & & & a_N \end{bmatrix},$$

элементы которой определим следующим образом:

$$a_i = \begin{cases} \tilde{a}_i, \text{ если } |\tilde{a}_i| \geqslant \varepsilon_1 \mathcal{K}(\tilde{A}); \\ \text{иначе} \begin{cases} \varepsilon_1 \mathcal{K}(\tilde{A}), \text{ если } \tilde{a}_i > 0; \\ -\varepsilon_1 \mathcal{K}(\tilde{A}), \text{ если } \tilde{a}_i \leqslant 0; \end{cases} \end{cases} \tag{8.3}$$

$$b_i = \begin{cases} \tilde{b}_i, \text{ если } |\tilde{b}_i| \geqslant \varepsilon_1 \mathcal{K}(\tilde{A}); \\ \text{иначе} \begin{cases} \varepsilon_1 \mathcal{K}(\tilde{A}), \text{ если } \tilde{b}_i > 0; \\ -\varepsilon_1 \mathcal{K}(\tilde{A}), \text{ если } \tilde{b}_i \leqslant 0. \end{cases} \end{cases}$$

Введем в рассмотрение величину

$$\mathcal{K}(A) = \max \begin{cases} \max\limits_{1 \leqslant i \leqslant N-1} (|a_i| + |b_{i+1}|), \\ \max\limits_{2 \leqslant i \leqslant N} (|a_i| + |b_i|). \end{cases}$$

Легко видеть, что

$$(1 - 2\varepsilon_1)\mathcal{K}(\tilde{A}) \leqslant \mathcal{K}(A) \leqslant (1 + 2\varepsilon_1)\mathcal{K}(\tilde{A}). \tag{8.4}$$

Так как сингулярные числа матрицы A являются в то же время собственными значениями симметрической трехдиагональной матрицы

$$S = \begin{bmatrix} 0 & a_1 & & & & & \\ a_1 & 0 & b_2 & & & 0 & \\ & b_2 & 0 & a_2 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & 0 & & b_N & 0 & a \\ & & & & & a_N & 0 \end{bmatrix},$$

имеющей порядок $2N$, то нетрудно показать, что сингулярные числа матриц A и $\tilde{A}$ различаются не более чем на $2\varepsilon_1 \mathcal{K}(\tilde{A})$. Отсюда следует, что σ является приближением к наибольшему сингулярному числу $\sigma_N(A) = \|A\|$ матрицы A, удовлетворяющим оценке

$$|\sigma - \|A\|| \leqslant \delta K(\tilde{A}), \tag{8.5}$$

где

$$\delta = \tilde{\delta} + 2\varepsilon_1. \tag{8.6}$$

С помощью алгоритма, описанного в § 3 работы [4], определим виртуальный почти собственный вектор $w = (w_1, w_2, \ldots, w_{2N-1}, w_{2N})^T$ матрицы S, отвечающий ее почти собственному значению σ. Для применения этого алгоритма необходимо заметить только, что используемая в его формулировке величина $\mathcal{M}(S)$, представляющая собой одну из норм матрицы S, совпадает с величиной $\mathcal{K}(A)$. В результате применения этого алгоритма находятся две последовательности чисел $Q_2, Q_3, \ldots$ $\ldots, Q_{2N-1}, Q_{2N}$; $T_2, T_3, \ldots, T_{2N-1}, T_{2N}$ и целое число j_0 такие, что отно-

шения $w_2/w_1, w_3/w_2, \ldots, w_{j_0-1}/w_{j_0-2}$ компонент вектора w равны соответственно числам $\mathcal{Q}_2, \mathcal{Q}_3, \ldots, \mathcal{Q}_{j_0-1}$, а отношения $w_{j_0}/w_{j_0-1}, w_{j_0+1}/w_{j_0}, \ldots$ $\ldots, w_{2N}/w_{2N-1}$ — числам $1/T_{j_0}, 1/T_{j_0+1}, \ldots, 1/T_{2N}$.

Определим числа $\mathcal{P}_1, \mathcal{R}_2, \mathcal{P}_2, \ldots, \mathcal{R}_N, \mathcal{P}_N$ по следующему правилу. Если j_0 четно, то представим j_0 в виде $j_0 = 2i_0$ и положим

$$\mathcal{P}_i = \mathcal{Q}_{2i}, \quad \mathcal{R}_{i+1} = \mathcal{Q}_{2i+1}, \quad i = 1, \ldots, i_0 - 1;$$
$$\mathcal{P}_i = (1+\psi_i)/T_{2i}, \quad \mathcal{R}_{i+1} = (1+\varphi_{i+1})/T_{2i+1}, \quad i = i_0, \ldots, N-1;$$
$$\mathcal{P}_N = (1+\psi_N)/T_{2N}.$$

Если же j_0 нечетно, то представим его в виде $j_0 = 2i_0 + 1$ и положим

$$\mathcal{P}_1 = \mathcal{Q}_2, \quad \mathcal{R}_i = \mathcal{Q}_{2i-1}, \quad \mathcal{P}_i = \mathcal{Q}_{2i}, \quad i = 2, \ldots, i_0;$$
$$\mathcal{R}_i = (1+\varphi_i)/T_{2i-1}, \quad \mathcal{P}_i = (1+\psi_i)/T_{2i}, \quad i = i_0+1, \ldots, N.$$

Здесь возмущения φ_i и ψ_i моделируют погрешности, возникающие при выполнении операций деления $1/T_{2i-1}$ и $1/T_{2i}$ соответственно так, что $|\varphi_i| \leqslant \varepsilon_1$, $|\psi_i| \leqslant \varepsilon_1$. Из доказательства теоремы 1 работы [4] легко получить следующее утверждение. Числа $\mathcal{P}_1, \mathcal{R}_2, \mathcal{P}_2, \ldots, \mathcal{R}_N, \mathcal{P}_N$ удовлетворяют соотношениям

$$\left.\begin{aligned}
&-\sigma + \delta_1'' + a_1(1+\alpha_1'')(1+\varphi_1)\mathcal{P}_1 = 0;\\
&\frac{a_{i-1}(1+\alpha_{i-1}')}{(1+\varphi_{i-1})\mathcal{P}_{i-1}} - \sigma + \delta_{i-1}' + b_i(1+\beta_i'')(1+\psi_i)\mathcal{R}_i = 0,\\
&\frac{b_i(1+\beta_i')}{(1+\psi_i)\mathcal{R}_i} - \sigma + \delta_i'' + a_i(1+\alpha_i'')(1+\varphi_i)\mathcal{P}_i = 0,
\end{aligned}\right\} i = 2, \ldots, N; \quad (8.7)$$
$$\frac{a_N(1+\alpha_N')}{(1+\varphi_N)\mathcal{P}_N} - \sigma + \delta_N' = 0.$$

Согласно теореме 3 из работы [4] возмущения $\delta_i', \delta_i'', \alpha_i', \alpha_i'', \beta_i', \beta_i''$ подчиняются оценкам

$$|\delta_i'| \leqslant \delta\mathcal{K}(\tilde{A}) + 16\varepsilon_1\mathcal{K}(A), \quad |\delta_i''| \leqslant \delta\mathcal{K}(\tilde{A}) + 16\varepsilon_1\mathcal{K}(A);$$
$$|\alpha_i'| \leqslant 10\varepsilon_1, \quad |\alpha_i''| \leqslant 10\varepsilon_1, \quad |\beta_i'| \leqslant 10\varepsilon_1, \quad |\beta_i''| \leqslant 10\varepsilon_1.$$

Заметим, что если учесть специфику матрицы S, а именно наличие у нее нулевой главной диагонали, то можно получить несколько лучшие оценки. Здесь же в целях простоты изложения использована непосредственно упомянутая теорема 3.

Для дальнейшего получим оценки величин δ_i' и δ_i'' в терминах величины σ. Из неравенств (2.24) легко следует оценка

$$\mathcal{K}(\tilde{A}) \leqslant \sqrt{2}\,\|\tilde{A}\|. \quad (8.8)$$

Принимая теперь во внимание неравенства (8.4), имеем

$$|\delta_i'| \leqslant \sqrt{2}\,[\delta + 16\varepsilon_1(1+2\varepsilon_1)]\|\tilde{A}\|. \quad (8.9)$$

Из (8.2) и (8.8) нетрудно получить оценку $\|\tilde{A}\| \leqslant \sigma/(1-\sqrt{2}\,\tilde{\delta})$. Огрубляя с ее помощью оценку (8.9), запишем

$$|\delta_i'| \leqslant \frac{\sqrt{2}\,[\delta + 16\varepsilon_1(1+2\varepsilon_1)]}{1-\sqrt{2}\tilde{\delta}}\,\sigma.$$

Наконец, учитывая замечание о возможном значении константы $\tilde{\delta}$ в оценке (8.2) ($\tilde{\delta} = 4\varepsilon_1$) и определение (8.6) величины δ, полученную оценку можно записать в виде

$$|\delta_i'| \leqslant \mu\sigma, \quad (8.10)$$

где в качестве значения μ принято $\mu = 31{,}5\varepsilon_1$. Оценка (8.10), очевидно, эквивалентна равенству $\delta_i' = \mu_i'\sigma$, которое имеет место при некотором μ_i' таком, что $|\mu_i'| \leqslant \mu$. Аналогичным образом можно получить равенство $\delta_i'' = \mu_i''\sigma$, справедливое при некотором значении μ_i'', удовлетворяющем оценке $|\mu_i''| \leqslant \mu$.

Преобразуем теперь соотношения (8.7). Необходимые преобразования продемонстрируем на примере соотношения

$$\frac{b_i(1+\beta_i')}{(1+\psi_i)\mathscr{R}_i} - \sigma + \delta_i'' + a_i(1+\alpha_i'')(1+\varphi_i)\mathscr{P}_i = 0.$$

Поскольку выражение $-\sigma + \delta_i''$ можно записать в виде $-\sigma + \delta_i'' = -\sigma(1-\mu_i'')$, это соотношение после деления обеих его частей на $1-\mu_i''$ принимает вид

$$\frac{b_i(1+\beta_i')}{(1+\psi_i)(1-\mu_i'')\mathscr{R}_i} - \sigma + \frac{a_i(1+\alpha_i'')(1+\varphi_i)}{1-\mu_i''}\mathscr{P}_i = 0.$$

Вводя обозначения

$$\hat{\xi}_i = \frac{1+\beta_i'}{(1+\psi_i)(1+\mu_i'')} - 1, \quad \check{\zeta}_i = \frac{(1+\alpha_i'')(1+\varphi_i)}{1-\mu_i''} - 1, \tag{8.11}$$

перепишем его в виде

$$\frac{b_i(1+\hat{\xi}_i)}{\mathscr{R}_i} - \sigma + a_i(1+\check{\zeta}_i)\mathscr{P}_i = 0. \tag{8.12}$$

Из формул (8.11) и оценок величин β_i', α_i'', ψ_i, φ_i, μ_i'' легко получить оценки величин $\hat{\xi}_i$ и $\check{\zeta}_i$:

$$|\hat{\xi}_i| \leqslant \rho, \quad |\check{\zeta}_i| \leqslant \rho, \tag{8.13}$$

справедливые при значении $\rho = 43\varepsilon_1$. Итак, соотношение (8.12) имеет место при некоторых значениях величин $\hat{\xi}_i$ и $\check{\zeta}_i$, удовлетворяющих оценкам (8.13). Аналогично преобразуются остальные соотношения (8.7). Таким образом, мы показали, что числа $\mathscr{P}_1$, $\mathscr{R}_2$, $\mathscr{P}_2$, ..., $\mathscr{R}_N$, $\mathscr{P}_N$ удовлетворяют соотношениям

$$\left.\begin{aligned}
&-\sigma + a_1(1+\check{\zeta}_1)\mathscr{P}_1 = 0;\\
&\frac{a_{i-1}(1+\hat{\zeta}_{i-1})}{\mathscr{P}_{i-1}} - \sigma + b_i(1+\check{\xi}_i)\mathscr{R}_i = 0,\\
&\frac{b_i(1+\hat{\xi}_i)}{\mathscr{R}_i} - \sigma + a_i(1+\check{\zeta}_i)\mathscr{P}_i = 0,\\
&\frac{a_N(1+\hat{\zeta}_N)}{\mathscr{P}_N} - \sigma = 0
\end{aligned}\right\} \quad i = 2, \ldots, N; \tag{8.14}$$

при некоторых $\hat{\xi}_i$, $\check{\xi}_i$, $\hat{\zeta}_i$, $\check{\zeta}_i$ таких, что $|\hat{\xi}_i| \leqslant \rho$, $|\check{\xi}_i| \leqslant \rho$, $|\hat{\zeta}_i| \leqslant \rho$, $|\check{\zeta}_i| \leqslant \rho$, где $\rho = 43\varepsilon_1$.

Определим теперь числа c_i, s_i, $\bar{c}_i$, $\bar{s}_i$ и c_i', s_i', $\bar{c}_i'$, $\bar{s}_i'$ следующим образом:

$$s_2 = \frac{1}{\sqrt{1+(\mathscr{P}_1\mathscr{R}_2)^2}}, \quad c_2 = \frac{\mathscr{P}_1\mathscr{R}_2}{\sqrt{1+(\mathscr{P}_1\mathscr{R}_2)^2}}, \quad c_2' = (c_2)_{\text{м}}, \quad s_2' = (s_2)_{\text{м}};$$

$$\bar{s}_2 = \frac{1}{\sqrt{1+(\mathscr{R}_2\mathscr{P}_2)^2}}, \quad \bar{c}_2 = \frac{\mathscr{R}_2\mathscr{P}_2}{\sqrt{1+(\mathscr{R}_2\mathscr{P}_2)^2}}, \quad \bar{c}_2' = (\bar{c}_2)_{\text{м}}, \quad \bar{s}_2' = (\bar{s}_2)_{\text{м}};$$

$$\left.\begin{aligned} s_i &= \frac{1}{\sqrt{1+(c'_{i-1}\mathscr{P}_{i-1}\mathscr{R}_i)^2}}, \quad c_i = \frac{c'_{i-1}\mathscr{P}_{i-1}\mathscr{R}_i}{\sqrt{1+(c'_{i-1}\mathscr{P}_{i-1}\mathscr{R}_i)^2}}, \\ c'_i &= (c_i)_{\text{м}}, \quad s'_i = (s_i)_{\text{м}}, \\ \bar{s}_i &= \frac{1}{\sqrt{1+(\bar{c}'_{i-1}\mathscr{R}_i\mathscr{P}_i)^2}}, \quad \bar{c}_i = \frac{\bar{c}'_{i-1}\mathscr{R}_i\mathscr{P}_i}{\sqrt{1+(\bar{c}'_{i-1}\mathscr{R}_i\mathscr{P}_i)^2}}, \\ \bar{c}'_i &= (\bar{c}_i)_{\text{м}}, \quad \bar{s}'_i = (\bar{s}_i)_{\text{м}}, \end{aligned}\right\} \quad i = 3, \ldots, N. \tag{8.15}$$

Формулы $c'_i = (c_i)_{\text{м}}$, $s'_i = (s_i)_{\text{м}}$, $\bar{c}'_i = (\bar{c}_i)_{\text{м}}$, $\bar{s}'_i = (\bar{s}_i)_{\text{м}}$ означают, что числа c'_i, s'_i, $\bar{c}'_i$, $\bar{s}'_i$ являются результатами машинного вычисления величин c_i, s_i, $\bar{c}_i$, $\bar{s}_i$. При этом предполагается, что вычисления ведутся по специальным алгоритмам, подробно описанным и исследованным в § 4 главы. 2. Анализ этих алгоритмов позволяет гарантировать выполнение равенств

$$c'_i = c_i(1+\tau_i), \quad s'_i = s_i(1+\pi_i), \quad \bar{c}'_i = \bar{c}_i(1+\bar{\tau}_i), \quad \bar{s}'_i = \bar{s}_i(1+\bar{\pi}_i), \tag{8.16}$$

где величины τ_i, π_i, $\bar{\tau}_i$, $\bar{\pi}_i$ удовлетворяют оценкам

$$|\tau_i| \leqslant \tau, \quad |\pi_i| \leqslant \pi, \quad |\bar{\tau}_i| \leqslant \tau, \quad |\bar{\pi}_i| \leqslant \pi$$

при $\tau = 7{,}001\varepsilon_1$, $\pi = 5{,}001\varepsilon_1$. Напомним также, что в указанных алгоритмах каждая из величин c'_i, s'_i, $\bar{c}'_i$, $\bar{s}'_i$ определяется с помощью двух чисел: порядка и мантиссы, хранящихся в разных ячейках памяти машины. Значения мантисс этих величин, как обычно, заключены в промежутке от $1/\alpha$ до 1, где α — основание системы счисления, принятой в машине. В силу этого все числа c'_i, s'_i, $\bar{c}'_i$, $\bar{s}'_i$ отличны от нуля.

В § 3 главы 2 показано, что из соотношений (8.14) и равенств (8.16) следуют соотношения

$$s_2 a_1(1+\breve{\alpha}_1) - s_2\frac{\sigma^2}{a_1} + c_2 b_2(1+\breve{\beta}_2) = 0;$$

$$\bar{s}_2 b_2(1+\hat{\beta}_2) - c_2\frac{\bar{s}_2}{s_2}a_1 + \bar{c}_2 a_2(1+\breve{\alpha}_2) = 0;$$

$$\left.\begin{aligned} &s_{i+1}c_i a_i(1+\hat{\alpha}_i) - \bar{c}_i\frac{s_{i+1}s_i\cdots s_2}{\bar{s}_i\cdots\bar{s}_2}\frac{\sigma^2}{a_1} + c_{i+1}b_{i+1}(1+\breve{\beta}_{i+1}) = 0, \\ &s_{i+1}\bar{c}_i b_{i+1}(1+\hat{\beta}_{i+1}) - c_{i+1}\frac{\bar{s}_{i+1}\cdots\bar{s}_2}{s_{i+1}\cdots s_2}a_1 + \bar{c}_{i+1}a_{i+1}(1+\breve{\alpha}_{i+1}) = 0, \end{aligned}\right\} \quad i = 2, \ldots, N-1;$$

$$c_N a_N(1+\hat{\alpha}_N) - \bar{c}_N\frac{s_N\cdots s_2}{\bar{s}_N\cdots\bar{s}_2}\frac{\sigma^2}{a_1} = 0,$$

которые имеют место при некоторых $\hat{\alpha}_i$, $\breve{\alpha}_i$, $\hat{\beta}_i$, $\breve{\beta}_i$ таких, что

$$|\hat{\alpha}_i| \leqslant \varepsilon, \quad |\breve{\alpha}_i| \leqslant \varepsilon, \quad |\hat{\beta}_i| \leqslant \varepsilon, \quad |\breve{\beta}_i| \leqslant \varepsilon,$$

где в качестве значения константы ε можно взять $\varepsilon = 2(N\tau + \rho)$.

Дальнейшие рассмотрения непосредственно опираются на результаты § 6 главы 3. Определим матрицу

$$(\bar{A})_{\text{м}} = \begin{bmatrix} (\bar{a}'_1)_{\text{м}} & (\bar{b}'_2)_{\text{м}} & & & \\ & (\bar{a}'_2)_{\text{м}} & (\bar{b}'_3)_{\text{м}} & 0 & \\ & & \ddots & \ddots & \\ & & & (\bar{a}'_{N-2})_{\text{м}} & (\bar{b}'_{N-1})_{\text{м}} \\ & 0 & & & (\bar{a}'_{N-1})_{\text{м}} \quad 0 \\ & & & & \sigma \end{bmatrix}.$$

элементы $(\bar{a}_i')_м$, $(\bar{b}_i')_м$ которой получаются в результате машинного вычисления по формулам

$$\bar{a}_i' = \frac{s_{i+1}' a_{i+1}}{\bar{s}_{i+1}'}, \quad i = 1, \ldots, N-1;$$

$$\bar{b}_i' = \frac{\bar{s}_i' b_{i+1}}{s_{i+1}'}, \qquad i = 2, \ldots, N-1.$$

В § 6 детально описаны алгоритмы, с помощью которых проводятся вычисления по этим формулам.

Из рассмотрений § 6 следует, что вычисленная таким образом матрица $(\bar{A})_м$ по норме мало отличается от некоторой матрицы, ортогонально эквивалентной матрице A, а именно из теоремы 4 следует оценка

$$\|(\bar{A})_м - \bar{C}AC^*\| \leqslant \{2\sqrt{2}(\sqrt{N}+2)\varepsilon + (\sqrt{2}\,\tilde{\pi} + \varepsilon_1)[1 + 2\sqrt{2}(\sqrt{N}+2)\varepsilon]\}\|A\|. \tag{8.17}$$

Напомним, что матрицы C и $\bar{C}$ представляют собой произведения ортогональных матриц

$$C = C_{N+1}C_N \cdots C_3C_2, \quad \bar{C} = \bar{C}_N \cdots \bar{C}_3\bar{C}_2,$$

где C_i, $\bar{C}_i$ $(i = 2, \ldots, N)$ — элементарные матрицы вращения, конструируемые с помощью параметров c_i, s_i $(c_i^2 + s_i^2 = 1)$ и $\bar{c}_i$, $\bar{s}_i$ $(\bar{c}_i^2 + \bar{s}_i^2 = 1)$ соответственно, определенных формулами (8.15), а матрица C_{N+1} имеет вид

$$C_{N+1} = \operatorname{diag}[1, 1, \ldots, 1, \operatorname{sign} a_1].$$

Характер константы ε в этой оценке можно представить себе с помощью формулы $\varepsilon = 2(N\tau + \rho)$, в которой значения величин τ и ρ равны соответственно $7{,}001\varepsilon_1$ и $43\varepsilon_1$. Значение константы $\tilde{\pi}$ определяется формулой (6.14) $\tilde{\pi} = \varepsilon_1(2 + \varepsilon_1)(1 + \hat{\pi}) + \hat{\pi}$. В свою очередь, величина $\hat{\pi}$ вычисляется по формуле (6.7): $\hat{\pi} = 2\pi/(1 - \pi)$, где $\pi = 5{,}001\varepsilon_1$. Пренебрегая в этих формулах членами второго порядка малости по ε_1, находим, что в качестве значения $\tilde{\pi}$ можно взять $\tilde{\pi} = 12{,}002\varepsilon_1$.

Итак, оценка (8.17) имеет место при следующих значениях входящих в нее констант ε и $\tilde{\pi}$: $\varepsilon = 2(7{,}001N + 43)\varepsilon_1$, $\tilde{\pi} = 12{,}002\varepsilon_1$.

В заключение оценим отличие вычисленной матрицы $(\bar{A})_м$ от $\bar{C}\tilde{A}C^*$, ортогонально эквивалентной исходной матрице $\tilde{A}$. Из определения (8.3) элементов a_i, b_i матрицы A и неравенства (8.8) с помощью неоднократно упоминавшейся оценки (2.30) из [4] легко получить оценку $\|A - \tilde{A}\| \leqslant 2\sqrt{2}\,\varepsilon_1\|\tilde{A}\|$. Из этой оценки следует, что $\|A\| \leqslant (1 + 2\sqrt{2}\,\varepsilon_1)\|\tilde{A}\|$. Используя (8.17) и полученные оценки, находим

$$\|(\bar{A})_м - \bar{C}\tilde{A}C^*\| \leqslant \|(\bar{A})_м - \bar{C}AC^*\| + \|A - \tilde{A}\| \leqslant$$
$$\leqslant \{2\sqrt{2}(\sqrt{N}+2)\varepsilon + (\sqrt{2}\,\tilde{\pi} + \varepsilon_1)[1 + 2\sqrt{2}(\sqrt{N}+2)\varepsilon]\}(1 + 2\sqrt{2}\,\varepsilon_1)\|\tilde{A}\| +$$
$$+ 2\sqrt{2}\,\varepsilon_1\|\tilde{A}\|.$$

Пренебрегая в этой оценке членами второго порядка, окончательно получим

$$\|(\bar{A})_м - \bar{C}\tilde{A}C^*\| \leqslant \{2\sqrt{2}[(\sqrt{N}+2)\varepsilon + \varepsilon_1] + \sqrt{2}\,\tilde{\pi} + \varepsilon_1\}\|\tilde{A}\|.$$

Подчеркнем, что к результатам работы описанного алгоритма кроме матрицы $(\bar{A})_м$ относятся значение выражения, стоящего в правой части полученной оценки, а также числа c_i', s_i', $\bar{c}_i'$, $\bar{s}_i'$ и значения констант τ и π, оценивающих отличие этих чисел от точных значений параметров c_i, s_i, $\bar{c}_i$, $\bar{s}_i$, с помощью которых конструируются ортогональные матрицы C и $\bar{C}$.

ЛИТЕРАТУРА

1. **Rutishauser H.** On Jacobi rotation patterns.— Proceedings A. M. S. Symposia in Applied Mathematics, 1963, v. 15, p. 219—239.
2. **Golub G., Kahan W.** Calculating the singular values and pseudoinverse of a matrix.— J. SIAM Numer. Anal. Ser. B, 1965, v. 2, N 2, p. 205—224.
3. **Уилкинсон Дж. Х.** Алгебраическая проблема собственных значений.— М.: Наука, 1970.— 564 с.
4. **Годунов С. К., Костин В. И., Митченко А. Д.** Вычисление собственного вектора симметрической трехдиагональной матрицы.— Сиб. мат. журн., 1985, т. XXVI. № 5.
5. **Годунов С. К.** Решение систем линейных уравнений.— Новосибирск: Наука. Сиб. отд-ние, 1980.— 177 с.

ОБОБЩЕННЫЕ РЕШЕНИЯ И РЕГУЛЯРИЗАЦИЯ СИСТЕМ НЕРАВЕНСТВ

В. А. БУЛАВСКИЙ

Рассматривается подход к решению несовместных систем неравенств, включающий построение их обобщенных решений и регуляризацию. Этот подход представляет естественное обобщение развитого А. Н. Тихоновым [1, 2] способа регуляризации систем линейных алгебраических уравнений и обладает следующими свойствами. Во-первых, он не выводит за рамки линейных задач; во-вторых, позволяет построить класс релаксационных алгоритмов, сходящихся независимо от совместности решаемой задачи, который включает в себя, например, методы последовательного проектирования [3—6] и градиентный метод. Наконец, предлагаемый подход применим для коррекции несовместных (включая двойственную несовместность) задач линейного программирования, которым в последнее время уделяется внимание [7]. Материал статьи является развитием ранее опубликованных результатов автора [5, 8, 9] и основан на использовании аппарата задач с условиями дополнительности. Укажем также, что вводимые понятия и значительная часть результатов могут быть перенесены на нелинейный случай.

В заключение введения оговорим некоторые обозначения и терминологию. Все изложение ведется для конечномерных пространств, и векторы трактуются как векторы-столбцы, так что при умножении на матрицу они стоят справа. Транспонирование обозначается штрихом в качестве верхнего индекса, например, если x — вектор-столбец, то x' — вектор-строка. Для положительной и отрицательной частей числа α приняты обозначения α_+ и α_-. Эти же обозначения используются для векторов: символ x_+ обозначает вектор, составленный из положительных частей компонент вектора x. Покомпонентно применяется и знак неравенства между векторами. Наконец, термины «положительно определенная матрица Q» или «положительно полуопределенная матрица Q» не предполагают симметрию матрицы Q, а лишь означают, что выполняются неравенства $z'Qz > 0$ при $z \neq 0$ или $z'Qz \geqslant 0$ при всех z.

§ 1. ЗАДАЧИ С УСЛОВИЯМИ ДОПОЛНИТЕЛЬНОСТИ

Линейной задачей с условиями дополнительности называют задачу

$$Mz + q \geqslant 0, \quad z \geqslant 0, \quad z'(Mz + q) = 0, \tag{1.1}$$

где $z \in R^s$ — искомый вектор, а матрица $M \in R^{s \times s}$ и свободный член $q \in R^s$ заданы. В последние годы задачам такого типа посвящено много исследований. Обзор результатов и библиографию можно найти, напри-