

ГАРАНТИРОВАННАЯ ТОЧНОСТЬ В СПЕКТРАЛЬНЫХ ЗАДАЧАХ ЛИНЕЙНОЙ АЛГЕБРЫ

В 1988 г. в издательстве «Наука» вышла из печати монография С. К. Годунова, А. Г. Антонова, О. П. Кирилюка и В. И. Костина «Гарантированная точность решения систем линейных уравнений в евклидовых пространствах», посвященная изложению теоретических основ и практических аспектов алгоритмов линейной алгебры, объединенных общей идеей получения ответа с гарантированной точностью. Список основных задач, для решения которых пригодны излагаемые алгоритмы, включает:

- расчет собственных значений симметричных матриц;
- решение хорошо обусловленной системы линейных уравнений с исследованием обусловленности;
- построение ортонормированного базиса из собственных векторов симметричной матрицы;
- построение сингулярного разложения матрицы;
- нахождение обобщенного r -решения произвольной системы линейных уравнений.

Во всех алгоритмах предусмотрено исследование разрешимости задач в процессе их решения и получение эффективной гарантированной оценки погрешности ответа.

Методика исследования алгоритмов, используемая в монографии, основывается, с одной стороны, на применении ортогональных преобразований для упрощения вида матриц, с другой — на обратном анализе ошибок округления и подходящей теории возмущений. Такой подход ранее успешно применялся в классическом труде Дж. Х. Уилкинсона «Алгебраическая проблема собственных значений». В книге [1] эта методика доведена до полного совершенства и математической строгости. Теория возмущений симметричных спектральных задач базируется на использовании последовательности сингулярных значений матриц.

Настоящая работа посвящена исследованию на ЭВМ с гарантированной точностью несимметричных регулярных спектральных задач линейной алгебры. Сложность этих задач заключается прежде всего в том, что малые возмущения входных данных могут привести к большим ошибкам при определении собственных значений. А входные данные, как правило, являются результатом некоторых измерений или предыдущих вычислений и содержат в себе ошибки этих измерений или вычислений.

Стандартный пример такой неустойчивости к малым возмущениям — анализ устойчивости по Ляпунову двухдиагональной матрицы A размеров 20×20 с -1 по главной диагонали и 10 под диагональю. Внеся возмущение величины ε в правый верхний угол этой матрицы. Тогда определитель матрицы $A - \lambda I$, где I — единичная матрица размеров 20×20 , равен $(-1 - \lambda)^{20} - \varepsilon \cdot 10^{19}$. При $\varepsilon = 10^{-18}$, что меньше относительной погрешности арифметических операций многих современных компьютеров, получим корень полинома $\det(A - \lambda I)$ вида $\lambda = \sqrt[20]{10} - 1 > 0$. Таким образом, устойчивая матрица A с собственными значениями, равными -1 , малым возмущением элементов становится неустойчивой. Эти обстоятельства позволяют причислить матрицу A к семейству «плохо устойчивых» или «практически неустойчивых» матриц.

Числовой критерий качества устойчивости матриц, эффективно вычисляемый на ЭВМ, был предложен в работах [2, 3]. Развитие исследований этих работ в итоге привело к критериям качества дихотомии мат-

ричного спектра мнимой осью и единичной окружностью [4—6]. Под задачей о дихотомии спектра матрицы замкнутым контуром комплексной плоскости подразумевается выяснение вопроса о том, есть ли собственные числа матрицы на данном контуре и вблизи от него, и в случае отсутствия таких точек спектра вычисление проекторов на инвариантные подпространства, отвечающие собственным числам внутри контура и вне. Критерии качества дихотомии являются количественными характеристиками устойчивости этих инвариантных подпространств по отношению к возмущениям входных данных.

При таком подходе методика исследования спектра матрицы сводится к разбиению комплексной плоскости кругами и прямыми на части, определению параметра качества дихотомии для каждого круга или прямой и вычислению инвариантных подпространств для точек спектра внутри и вне круга или полуплоскости, если параметр дихотомии не слишком велик. После этого можно определить инвариантные подпространства для частей спектра в пересечении некоторых кругов или полуплоскостей.

Предложенная выше методика исследования несимметричных спектральных задач линейной алгебры свободна от трудностей, связанных с наличием кратных собственных значений: вместо отыскания индивидуальных собственных значений определяются окрестности изолированных групп точек спектра и соответствующие им инвариантные подпространства. Наш подход цементируется эффективно вычислимым параметром дихотомии, который позволяет построить эффективную теорию возмущений спектральных характеристик и проанализировать скорость сходимости и устойчивость некоторых эффективных численных алгоритмов. Параметр дихотомии можно считать мерой обусловленности спектральных задач.

Параметр дихотомии — это положительное вещественное число, которое характеризует, насколько устойчиво определены инвариантные подпространства матрицы для двух частей спектра, разделенных данным замкнутым контуром. Чем больше величина этого параметра, тем менее устойчивы соответствующие инвариантные подпространства. Устойчивость уменьшается, например, если собственные числа подходят ближе к контуру, или уменьшается угол между инвариантными подпространствами. Если вычисления ведутся на ЭВМ, то можно считать, что контур «практически» не делит спектр, когда параметр дихотомии больше некоторого достаточно большого числа. Величина этой границы зависит главным образом от относительной погрешности округлений и выбирается волевым образом, исходя из запросов практики или требования гарантированной точности решения.

Классическими спектральными характеристиками квадратной матрицы являются набор собственных чисел и базис, составленный из жордановых цепочек. В методике исследования спектра с помощью дихотомии спектральными характеристиками, имеющими первостепенное значение, являются параметр дихотомии данным замкнутым контуром и, если этот параметр конечен, инвариантные подпространства и обобщенная функция Ляпунова, отвечающие группам спектра внутри и вне контура.

Настоящая работа является дальнейшим развитием исследований [4—7]. Основное внимание в ней уделяется вычислению спектральных характеристик регулярного линейного пучка матриц $\lambda B - A$. В качестве контура, разделяющего спектр пучка, рассматривается единичная окружность комплексной плоскости.

По сравнению со спектральными проблемами для одной матрицы в случае пучков возникают новые особенности. Например, регулярный пучок может иметь бесконечные собственные значения. Для пучков также нужно различать правые и левые спектральные характеристики.

В работе предлагается новый ортогональный метод исследования спектра регулярного линейного пучка матриц, который имеет квадратичную скорость сходимости и вычисляет как проекторы на инвариантные подпространства пучка, так и параметры качества устойчивости этих подпространств. Для этого метода показана вычислительная устойчивость к ошибкам округления при не слишком больших значениях параметра дихотомии.

Все изложение работы ведется для вещественных матриц. С минимальными поправками оно годится для комплекснозначных матриц.

Анализ погрешностей округления проводится в предположении, что вычисления выполняются в арифметике вещественных чисел ЭВМ с плавающей точкой.

Некоторые обозначения

I_k — единичная $k \times k$ -матрица (если индекс k опущен, то размеры матрицы определяются по смыслу);

A^* — комплексно сопряженная матрица к матрице A ;

$\|A\|$ — спектральная норма матрицы A , т. е. максимальное сингулярное значение A ;

$\|A\|_F$ — фробениусова норма матрицы A , т. е. $\sqrt{\sum_{i,j} A_{ij}^2}$ с суммированием по всем i, j ;

$\sigma_{\max}(A)$ — максимальное сингулярное значение матрицы A ;

$\sigma_{\min}(A)$ — минимальное сингулярное значение матрицы A ;

ε_0 — абсолютная погрешность нуля вещественной арифметики ЭВМ с плавающей точкой;

ε_1 — относительная погрешность вещественной арифметики ЭВМ с плавающей точкой;

$\varepsilon_i, \varepsilon_{ii}, o_i$ — коэффициенты погрешностей округления при вычислении скалярного произведения на ЭВМ с плавающей точкой;

ε_r, o_r — коэффициенты погрешностей округления при выполнении преобразований отражения Хаусхолдера на ЭВМ с плавающей точкой.

§ 1. СПЕКТРАЛЬНЫЕ ХАРАКТЕРИСТИКИ РЕГУЛЯРНОГО ОТНОСИТЕЛЬНО ЕДИНИЧНОЙ ОКРУЖНОСТИ ЛИНЕЙНОГО ПУЧКА МАТРИЦ

Пусть A и B — матрицы размеров $N \times N$ над полем вещественных чисел. Назовем пучок $\lambda B - A$ *регулярным относительно единичной окружности*, если $\det(\lambda B - A) \neq 0$ для любого комплексного числа λ , лежащего на единичной окружности $|\lambda| = 1$. Все пучки матриц, рассматриваемые в этом параграфе, предполагаются регулярными относительно единичной окружности. Здесь и в дальнейших параграфах нас будут интересовать только правые спектральные характеристики.

1.1. Инвариантные подпространства пучка матриц, отвечающие группам спектра внутри и вне единичного круга

Вначале напомним понятия собственных чисел и соответствующих им жордановых цепочек регулярного линейного пучка матриц $\lambda B - A$.

Комплексные числа λ и $1/\mu$ (если $\mu = 0$, то $1/\mu = \infty$) называются *собственными числами пучка* $\lambda B - A$, если $\det(\lambda B - A) = 0$ или $\det(B - \mu A) = 0$.

Векторы $x_1, x_2, \dots, x_t$ образуют *жорданову цепочку пучка* $\lambda B - A$, отвечающую собственному числу λ , если $(\lambda B - A)x_1 = 0$, $(\lambda B - A)x_l + Bx_{l-1} = 0$, $l = 2, \dots, t$, причем, линейная система $(\lambda B - A)z + Bx_t = 0$ не имеет решения z . Соответственно векторы $x_1, x_2, \dots, x_t$ образуют

жорданову цепочку пучка $\lambda B - A$, отвечающую собственному числу $1/\mu$, если $(B - \mu A)x_1 = 0$, $(B - \mu A)x_l - Ax_{l-1} = 0$, $l = 2, \dots, t$, причем, линейная система $(B - \mu A)z - Ax_t = 0$ не имеет решения z .

Линейная оболочка $\mathcal{L}_0$ всех жордановых цепочек, отвечающих собственным числам λ , удовлетворяющим условию $|\lambda| < 1$, называется *инвариантным подпространством пучка $\lambda B - A$* , отвечающим собственным числам внутри единичного круга. Линейная оболочка $\mathcal{L}_\infty$ всех жордановых цепочек, отвечающих собственным числам $1/\mu$, удовлетворяющим условию $|\mu| < 1$, называется *инвариантным подпространством пучка $\lambda B - A$* , отвечающим собственным числам вне единичного круга.

Основные свойства и связи введенных выше спектральных характеристик выявляются в следующей теореме о кронекеровой форме регулярного относительно единичной окружности линейного пучка матриц.

Теорема 1.1. Если $\det(\lambda B - A) \neq 0$ при любом комплексном λ , удовлетворяющем условию $|\lambda| = 1$, то существуют такие невырожденные $N \times N$ -матрицы Q_l и Q_r , что

$$\begin{aligned} Q_l(\lambda B - A)Q_r &= \text{block diag} \{ \lambda I - J_0, \lambda J_\infty - I \}, \\ J_0 &= \text{block diag} \{ J_{i_1}(\lambda_1), \dots, J_{i_m}(\lambda_m) \}, \quad |\lambda_i| < 1, \\ J_\infty &= \text{block diag} \{ J_{j_1}(\mu_1), \dots, J_{j_n}(\mu_n) \}, \quad |\mu_j| < 1, \end{aligned} \quad (1.1)$$

где I — единичная матрица подходящих размеров, а

$$J_t(v) = \begin{bmatrix} v & 1 & & 0 \\ & \cdot & \cdot & \\ & & \cdot & \cdot \\ 0 & & & 1 & \\ & & & \cdot & v \end{bmatrix}$$

— блок Жордана размеров $t \times t$.

Доказательство. По существу доказательство теоремы 1.1 состоит в применении подходящим способом канонической формы Жордана. Перепишем пучок матриц $\lambda B - A$ следующим образом:

$$\lambda B - A = (B - A) [(\lambda - 1)(B - A)^{-1}B + I].$$

Существует такая невырожденная матрица F , что

$$\lambda B - A = (B - A)F[(\lambda - 1)J + I]F^{-1},$$

где J — каноническая форма Жордана матрицы $(B - A)^{-1}B$.

Разобьем матрицу J на два диагональных блока: $J = \text{block diag} \{J_1, J_2\}$, где J_1 — блок, содержащий все клетки, отвечающие собственным числам пучка $\lambda J + I - J$ внутри единичного круга, а J_2 — вне единичного круга. Легко убедиться в том, что матрицы J_1 и $J_2 - I$ обратимы.

Пучок $\lambda J + I - J$ представим в следующем виде:

$$\lambda J - (J - I) = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 - I \end{pmatrix} \left[\lambda \begin{pmatrix} I & 0 \\ 0 & (J_2 - I)^{-1}J_2 \end{pmatrix} - \begin{pmatrix} J_1^{-1}(J_1 - I) & 0 \\ 0 & I \end{pmatrix} \right].$$

Приведем матрицы $J_1^{-1}(J_1 - I)$ и $(J_2 - I)^{-1}J_2$ к формам Жордана $F_1 J_0 F_1^{-1}$ и $F_2 J_\infty F_2^{-1}$ соответственно, тогда

$$\lambda B - A = (B - A)F \begin{pmatrix} J_1 F_1 & 0 \\ 0 & (J_2 - I) F_2 \end{pmatrix} \begin{pmatrix} \lambda I - J_0 & 0 \\ 0 & \lambda J_\infty - I \end{pmatrix} \begin{pmatrix} F_1^{-1} & 0 \\ 0 & F_2^{-1} \end{pmatrix} F^{-1}.$$

Теорема доказана

Запишем матрицу Q_r в блочном виде

$$Q_r = [Q_1 \dots Q_m Q_{m+1} \dots Q_{m+n}],$$

где разбиение на блочные столбцы соответствует размерам диагональных блоков в разложении (1.1). Тогда, очевидно, столбцы каждого блока Q_i будут образовывать жорданову цепочку.

Можно показать, что инвариантное подпространство $\mathcal{L}_0$ пучка $\lambda B - A$, отвечающее собственным числам внутри единичного круга, порождается столбцами блоков $Q_1, Q_2, \dots, Q_m$, а инвариантное подпространство $\mathcal{L}_\infty$, отвечающее собственным числам вне единичного круга, порождается столбцами блоков $Q_{m+1}, \dots, Q_{m+n}$. Следовательно, пересечение подпространств $\mathcal{L}_0$ и $\mathcal{L}_\infty$ равно нулю, а сумма — всему пространству $\mathbb{R}^N$.

Векторные подпространства удобно определять с помощью проекторов на эти подпространства. Пусть N_0 — размерность подпространства $\mathcal{L}_0$, а $N_\infty = N - N_0$ — размерность подпространства $\mathcal{L}_\infty$. Рассмотрим матрицы P_0 и P_∞ , построенные при помощи разложения (1.1):

$$P_0 = Q_r \begin{pmatrix} I_{N_0} & 0 \\ 0 & 0 \end{pmatrix} Q_r^{-1}, \quad P_\infty = Q_r \begin{pmatrix} 0 & 0 \\ 0 & I_{N_\infty} \end{pmatrix} Q_r^{-1}. \quad (1.2)$$

Несложно убедиться в том, что P_0 — проектор на инвариантное подпространство $\mathcal{L}_0$ пучка $\lambda B - A$, т. е. $P_0^2 = P_0$, и равенство $P_0 x = x$ равносильно принадлежности вектора x этому подпространству. Соответственно матрица P_∞ является проектором на инвариантное подпространство $\mathcal{L}_\infty$.

Заметим, что $P_0 + P_\infty = I$. Это тождество означает, что P_0 и P_∞ — параллельные проекторы на $\mathcal{L}_0$ и $\mathcal{L}_\infty$.

Кроме параллельных проекторов P_0 и P_∞ полезно также рассмотреть ортогональные проекторы Π_0 и Π_∞ на соответствующие инвариантные подпространства пучка $\lambda B - A$. Приведем формулы, выражающие матрицы Π_0 и Π_∞ через матрицы P_0 и P_∞ . Умножим слева тождество $P_0^2 = P_0$ на P_0^* , а тождество $P_\infty P_0 = 0$ на P_∞^* . Складывая результаты и группируя левую часть, получим тождество

$$(P_0^* P_0 + P_\infty^* P_\infty) P_0 = P_0^* P_0,$$

с помощью которого выводим равенства

$$P_0 = (P_0^* P_0 + P_\infty^* P_\infty)^{-1} P_0^* P_0, \quad P_0 = P_0 (P_0^* P_0 + P_\infty^* P_\infty)^{-1} P_0^* P_0.$$

Последнее из этих равенств позволяет проверить, что матрица

$$\Pi_0 = P_0 (P_0^* P_0 + P_\infty^* P_\infty)^{-1} P_0^* \quad (1.3)$$

удовлетворяет системе матричных уравнений

$$\Pi_0^2 = \Pi_0 = \Pi_0^*,$$

$$P_0 \Pi_0 = \Pi_0,$$

$$\Pi_0 P_0 = P_0,$$

которая однозначно определяет ортогональный проектор Π_0 на инвариантное подпространство $\mathcal{L}_0$ пучка $\lambda B - A$. Аналогично выводится представление

$$\Pi_\infty = P_\infty (P_0^* P_0 + P_\infty^* P_\infty)^{-1} P_\infty^*. \quad (1.4)$$

В дальнейшем нам пригодятся следующие тождества:

$$I - \Pi_0 = P_\infty^* (P_0 P_0^* + P_\infty P_\infty^*)^{-1} P_\infty, \quad (1.5)$$

$$I - \Pi_\infty = P_0^* (P_0 P_0^* + P_\infty P_\infty^*)^{-1} P_0.$$

Покажем, к примеру, справедливость второй из этих формул. Умножим справа тождество $P_0^2 = P_0$ на P_0^* , а тождество $P_0 P_\infty = 0$ на P_∞^* и сложим результаты. Получим цепочку равенств

$$P_0 (P_0 P_0^* + P_\infty P_\infty^*) = P_0 P_0^*,$$

$$P_0 = P_0 P_0^* (P_0 P_0^* + P_\infty P_\infty^*)^{-1},$$

$$P_0 = P_0 P_0^* (P_0 P_0^* + P_\infty P_\infty^*)^{-1} P_0.$$

Последнее из этих равенств позволяет доказать, что матрица $I - \Pi_\infty = P_0^*(P_0 P_0^* + P_\infty P_\infty^*)^{-1} P_0$ удовлетворяет системе матричных уравнений

$$\begin{aligned}(I - \Pi_\infty)^2 &= I - \Pi_\infty = (I - \Pi_\infty)^*, \\ P_0 &= P_0(I - \Pi_\infty), \\ I - \Pi_\infty &= (I - \Pi_\infty)P_0,\end{aligned}$$

которая единственным образом определяет матрицу $I - \Pi_\infty$.

В заключение отметим связь нормы проекторов с углом между нетривиальными инвариантными подпространствами $\mathcal{L}_0$ и $\mathcal{L}_\infty$. Нетривиальность этих подпространств означает, что $\mathcal{L}_0 \neq 0$ и $\mathcal{L}_\infty \neq 0$.

Определим угол φ , $0 < \varphi \leq \pi/2$, между инвариантными подпространствами с параллельными проекторами P_0 и P_∞ по формуле

$$\cos \varphi = \max_{\substack{P_0 x \neq 0 \\ P_\infty y \neq 0}} \frac{|(P_0 x, P_\infty y)|}{\|P_0 x\| \|P_\infty y\|}. \quad (1.6)$$

Используя тождества типа $(I - \Pi_\infty)P_\infty = 0$, получим цепочку соотношений

$$\begin{aligned}\cos \varphi &= \max_{x, y} \frac{|(\Pi_\infty P_0 x, P_\infty y)|}{\|P_0 x\| \|P_\infty y\|} = \max_x \frac{\|\Pi_\infty P_0 x\|}{\|P_0 x\|}, \\ \sin \varphi &= \min_x \frac{\|(I - \Pi_\infty)P_0 x\|}{\|P_0 x\|} = \min_x \frac{\|(I - \Pi_\infty)x\|}{\|P_0 x\|}, \\ \sin \varphi &\leq \min_x \frac{\|x\|}{\|P_0 x\|} = \frac{1}{\|P_0\|}, \\ \sin \varphi &= \min_x \frac{\|(I - \Pi_\infty)x\|}{\|P_0(I - \Pi_\infty)x\|} \geq \frac{1}{\|P_0\|}.\end{aligned}$$

В итоге

$$\sin \varphi = \frac{1}{\|P_0\|} = \frac{1}{\|P_\infty\|}. \quad (1.7)$$

Можно также показать, что

$$\sin \varphi = \sigma_{\min}(\Pi_0 - \Pi_\infty), \quad (1.8)$$

где $\sigma_{\min}(\Pi_0 - \Pi_\infty)$ — минимальное сингулярное число матрицы $\Pi_0 - \Pi_\infty$.

Для различных манипуляций с проекторами очень полезен выбор специальных ортогональных базисов. Продемонстрируем эту технику для доказательства тождества (1.8). Приведем матрицу проектора P_0 к форме Шура

$$P_0 = U^* \begin{pmatrix} X_1 & X \\ 0 & X_0 \end{pmatrix} U, \quad UU^* = I,$$

причем, верхнетреугольная матрица X_1 содержит на диагонали все единицы, а верхнетреугольная X_0 имеет над диагонали все нули. Тогда тождества $X_1^2 = X_1$ и $X_0^2 = X_0$ обеспечивают равенства $X_1 = I$, $X_0 = 0$. Следовательно, в таком базисе Шура

$$P_0 = U^* \begin{pmatrix} I & L \\ 0 & 0 \end{pmatrix} U, \quad P_\infty = U^* \begin{pmatrix} 0 & -L \\ 0 & I \end{pmatrix} U \quad (1.9)$$

с некоторой матрицей $L (=X)$.

Если $\mathcal{L}_0 \neq 0$ и $\mathcal{L}_\infty \neq 0$, то несложно показать, что $\|P_0\| = \|P_\infty\| = \sqrt{1 + \|L\|^2}$.

Выпишем, используя (1.9), выражения для ортогональных проекторов Π_0 и Π_∞ . Так как

$$P_0^* P_0 + P_\infty^* P_\infty = U^* \begin{pmatrix} I & L \\ L^* & I + 2L^* L \end{pmatrix} U,$$

$$(P_0^* P_0 + P_\infty^* P_\infty)^{-1} = U^* \begin{pmatrix} I + L(I + L^* L)^{-1} L^* & -L(I + L^* L)^{-1} \\ -(I + L^* L)^{-1} L^* & (I + L^* L)^{-1} \end{pmatrix} U,$$

то

$$\Pi_0 = U^* \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} U,$$

$$\Pi_\infty = U^* \begin{pmatrix} L(I + L^* L)^{-1} L^* & -L(I + L^* L)^{-1} \\ -(I + L^* L)^{-1} L^* & (I + L^* L)^{-1} \end{pmatrix} U,$$

$$\Pi_0 - \Pi_\infty = U^* \begin{pmatrix} I - L(I + L^* L)^{-1} L^* & L(I + L^* L)^{-1} \\ (I + L^* L)^{-1} L^* & -(I + L^* L)^{-1} \end{pmatrix} U.$$

Далее воспользуемся сингулярным разложением матрицы L :

$$L = V_1^* (\Sigma \ 0) V_2 \text{ или } L = V_1^* \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V_2,$$

где Σ — диагональная матрица с положительными элементами по диагонали, а V_1 и V_2 — ортогональные матрицы. Если L — квадратная матрица, то нулевых блоков в сингулярном разложении нет, т. е. $L = V_1^* \Sigma V_2$. Рассмотрим, к примеру, случай, когда $L = V_1^* (\Sigma 0) V_2$. Тогда

$$\Pi_0 - \Pi_\infty = U^* \begin{pmatrix} V_1^* & 0 \\ 0 & V_2^* \end{pmatrix} \begin{bmatrix} (I + \Sigma^2)^{-1} & (I + \Sigma^2)^{-1} \Sigma & 0 \\ \Sigma (I + \Sigma^2)^{-1} & -(I + \Sigma^2)^{-1} & 0 \\ 0 & 0 & -I \end{bmatrix} \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} U,$$

и $\sigma_{\min}(\Pi_0 - \Pi_\infty) = 1/\sqrt{1 + \|L\|^2}$. Оставшиеся случаи для сингулярного разложения матрицы L разбираются аналогично.

1.2. Канонический вид регулярного относительно единичной окружности линейного пучка матриц

Воспользуемся обозначениями теоремы 1.1 для того, чтобы определить последовательность матриц

$$\begin{aligned} P_{+0} = P_0 &= Q_r \begin{pmatrix} I_{N_0} & 0 \\ 0 & 0 \end{pmatrix} Q_r^{-1}, \quad P_{-0} = P_\infty = Q_r \begin{pmatrix} 0 & 0 \\ 0 & I_{N_\infty} \end{pmatrix} Q_r^{-1}, \\ P_k &= Q_r \begin{pmatrix} J_0^k & 0 \\ 0 & 0 \end{pmatrix} Q_r^{-1} \text{ при целых } k \geq 1, \\ P_{-k} &= Q_r \begin{pmatrix} 0 & 0 \\ 0 & J_\infty^k \end{pmatrix} Q_r^{-1} \text{ при целых } -k \leq -1. \end{aligned} \quad (1.10)$$

Для этих матриц P_k выполнены следующие тождества:

$$\begin{aligned} P_k P_l &= P_{k+l}, \text{ если } k \text{ и } l \text{ одного знака,} \\ P_k P_l &= 0, \text{ если } k \text{ и } l \text{ разных знаков.} \end{aligned} \quad (1.11)$$

Индексу $+0$ приписывается знак плюс, а индексу -0 — знак минус.

Полагая в теореме 1.1 $T = Q_r Q_l$, получим важное

Следствие. Если $\det(\lambda B - A) \neq 0$ при любом λ , $|\lambda| = 1$, то пучок $\lambda B - A$ можно представить в каноническом виде

$$\lambda B - A = T^{-1} [\lambda (P_{+0} + P_{-1}) - (P_{-0} + P_{+1})] \quad (1.12)$$

с невырожденной $N \times N$ -матрицей T и $N \times N$ -матрицами P_{+0} , P_{-1} , P_{-0} , P_{+1} , определенными в (1.10).

Заметим, что при $B = I$ матрица T равна $P_{+0} + P_{-1}$.

Канонический вид (1.12) можно определить независимо от разложения (1.1) теоремы 1.1. Такая характеристизация содержится в следующей теореме.

Теорема 1.2. Пусть

$$\begin{aligned} \lambda B - A &= T^{-1} [\lambda (P_{+0} + P_{-1}) - (P_{-0} + P_{+1})], \quad \det T \neq 0, \\ P_{+0}^2 - P_{+0} &= 0, \quad P_{-0} = I - P_{+0}, \\ P_{-0}P_{+1} &= P_{+1}P_{-0} = P_{+0}P_{-1} = P_{-1}P_{+0} = 0 \end{aligned} \quad (1.13)$$

и все собственные числа матрицы $P_{-1} + P_{+1}$ лежат внутри единичного круга. Тогда $\det(\lambda B - A) \neq 0$ при любом комплексном λ с единичной окружности $|\lambda| = 1$ и выполнены представления

$$\begin{aligned} T &= \frac{1}{2\pi} \int_0^{2\pi} (B - e^{i\varphi} A)^{-1} (1 + e^{i\varphi}) d\varphi, \\ P_{+0} &= \left[\frac{1}{2\pi} \int_0^{2\pi} (B - e^{i\varphi} A)^{-1} d\varphi \right] T^{-1}, \\ P_{-0} &= \left[\frac{1}{2\pi} \int_0^{2\pi} (B - e^{i\varphi} A)^{-1} e^{i\varphi} d\varphi \right] T^{-1}. \end{aligned} \quad (1.14)$$

Доказательство. Все рассуждения будем проводить в базисе, в котором матрицы P_{+0} и P_{-0} диагональны. Тогда из системы (1.13) следует, что

$$\begin{aligned} P_{+0} &= \begin{pmatrix} I_{N_0} & 0 \\ 0 & 0 \end{pmatrix}, \quad P_{-0} = \begin{pmatrix} 0 & 0 \\ 0 & I_{N_\infty} \end{pmatrix}, \\ P_{+1} &= \begin{pmatrix} K_0 & 0 \\ 0 & 0 \end{pmatrix}, \quad P_{-1} = \begin{pmatrix} 0 & 0 \\ 0 & K_\infty \end{pmatrix}, \end{aligned}$$

где все собственные числа K_0 и K_∞ лежат внутри единичного круга. Поэтому $\det(B - e^{i\varphi} A) \neq 0$ при любом вещественном φ .

Чтобы вывести формулы (1.14), выясним связь последовательности матриц P_k и последовательности матриц Грина G_k для разностного уравнения $Bx_n - Ax_{n-1} = f_n$. Разложим периодическую матричную функцию

$$\mathcal{D}(\varphi) = (B - e^{i\varphi} A)^{-1} \text{ в ряд Фурье по базису } e^{ik\varphi}: \quad \mathcal{D}(\varphi) = \sum_{k=-\infty}^{\infty} Z_k e^{ik\varphi}.$$

Коэффициенты ряда Фурье вычисляются с помощью интегралов

$$Z_k = \frac{1}{2\pi} \int_0^{2\pi} (B - e^{i\varphi} A)^{-1} e^{-ik\varphi} d\varphi. \quad (1.15)$$

Отсюда легко следует равномерная ограниченность последовательности $\|Z_k\|$ при всех целых k .

Из тождества $(B - e^{i\varphi} A)\mathcal{D}(\varphi) = I$ выводится система уравнений, связывающих матрицы Z_k :

$$BZ_k - AZ_{k-1} = \begin{cases} I, & k = 0, \\ 0, & k \neq 0. \end{cases} \quad (1.16)$$

Поэтому последовательность матриц Z_k является последовательностью матриц Грина G_k , т. е. $Z_k = G_k$.

Свойства матриц P_k позволяют явно сконструировать решение бесконечной системы (1.16) с ограниченными при всех целых k величинами $\|Z_k\|$:

$$\begin{aligned} G_0 &= Z_0 = P_{+0}T, & G_k &= Z_k = P_kT \text{ при } k \geq 1, \\ G_{-1} &= Z_{-1} = -P_{-0}T, & G_k &= Z_k = -P_{k+1}T \text{ при } k \leq -2. \end{aligned} \quad (1.17)$$

Следовательно, $T = G_0 - G_{-1}$.

Представления (1.14) получаются из (1.15) и (1.17). Теорема доказана.

1.3. Параметр ω — критерий отсутствия на единичной окружности и вблизи от нее точек спектра регулярного линейного пучка матриц

Пусть $\det(B - e^{i\varphi}A) \neq 0$ при любом вещественном φ , т. е. пучок $\lambda B - A$ регулярен относительно единичной окружности. Рассмотрим эрмитову матрицу

$$H = \frac{1}{2\pi} \int_0^{2\pi} (B - e^{i\varphi}A)^{-1} (AA^* + BB^*) (B^* - e^{-i\varphi}A^*)^{-1} d\varphi. \quad (1.18)$$

Спектральную норму (наибольшее сингулярное число) матрицы H обозначим через $\omega = \|H\|$ и назовем *критерием отсутствия собственных чисел пучка $\lambda B - A$ на единичной окружности и вблизи от нее*.

Отметим, что при умножении пучка $\lambda B - A$ слева на любую невырожденную матрицу эрмитова матрица H не изменяется. Пусть невырожденная матрица L удовлетворяет уравнению

$$LL^* = AA^* + BB^*. \quad (1.19)$$

Обозначив $A_0 = L^{-1}A$, $B_0 = L^{-1}B$, получим тождество

$$H = \frac{1}{2\pi} \int_0^{2\pi} (B_0 - e^{i\varphi}A_0)^{-1} (B_0^* - e^{-i\varphi}A_0^*)^{-1} d\varphi. \quad (1.20)$$

Пучок $\lambda B_0 - A_0$ удовлетворяет уравнению

$$A_0A_0^* + B_0B_0^* = I. \quad (1.21)$$

Мы будем называть такие пучки $\lambda B_0 - A_0$ *нормированными пучками*, полученными нормировкой пучка $\lambda B - A$ условием (1.21). Очевидно, что все нормированные пучки, полученные нормировкой одного и того же пучка $\lambda B - A$, отличаются на унитарный множитель слева.

Предыдущие замечания наталкивают на мысль о том, что параметр ω отражает спектральные свойства нормированного пучка $\lambda B_0 - A_0$. Действительно, с помощью ω можно оценить $\sigma_{\min}(B_0 - e^{i\varphi}A_0)$.

Теорема 1.3. Пусть $A_0A_0^* + B_0B_0^* = I$ и $\det(B_0 - e^{i\varphi}A_0) \neq 0$ при любом вещественном φ . Тогда

$$\max_{\varphi} \|(B_0 - e^{i\varphi}A_0)^{-1}\| < 5\pi\omega. \quad (1.22)$$

Доказательство. Так как

$$\omega = \frac{1}{2\pi} \max_{\|y\|=1} \int_0^{2\pi} \|(B_0^* - e^{-i\varphi}A_0^*)^{-1}y\|^2 d\varphi,$$

то

$$\omega \geq \frac{1}{2\pi} \int_0^{2\pi} \|(B_0^* - e^{-i\varphi}A_0^*)^{-1}y_0\|^2 d\varphi,$$

где y_0 — любой вектор единичной длины. Зафиксируем $\varphi = \varphi_0$ и выберем вектор y_0 , удовлетворяющий равенству

$$\|(B_0^* - e^{-i\varphi_0} A_0^*)^{-1}\| = \|(B_0^* - e^{-i\varphi_0} A_0^*)^{-1} y_0\|, \quad \|y_0\| = 1.$$

Далее воспользуемся матричным тождеством

$$(K + M)^{-1} = K^{-1} - (K + M)^{-1} M K^{-1}$$

и проведем следующие выкладки:

$$\begin{aligned} \|(B_0^* - e^{i\varphi} A_0^*)^{-1} y_0\| &= \|(B_0^* - e^{-i\varphi_0} A_0^*)^{-1} y_0 - \\ &- (B_0^* - e^{-i\varphi} A_0^*)^{-1} (e^{-i\varphi_0} - e^{-i\varphi}) A_0^* (B_0^* - e^{-i\varphi_0} A_0^*)^{-1} y_0\| \geq \\ &\geq \|(B_0^* - e^{-i\varphi_0} A_0^*)^{-1}\| (1 - \|(B_0^* - e^{-i\varphi} A_0^*)^{-1} (e^{-i\varphi_0} - e^{-i\varphi}) A_0^*\|) \geq \\ &\geq \|(B_0^* - e^{-i\varphi_0} A_0^*)^{-1}\| \left(1 - \frac{\|\varphi - \varphi_0\| \|(B_0^* - e^{-i\varphi_0} A_0^*)^{-1}\|}{1 - \|\varphi - \varphi_0\| \|(B_0^* - e^{-i\varphi_0} A_0^*)^{-1}\|} \right). \end{aligned}$$

В этих выкладках были использованы неравенства

$$\|A_0\| \leq 1, \quad |e^{-i\varphi_0} - e^{-i\varphi}| < |\varphi - \varphi_0|, \quad \|(K + M)^{-1}\| \leq \frac{\|K^{-1}\|}{1 - \|K^{-1}\| \|M\|}.$$

Для уменьшения громоздкости формул обозначим $r = \|(B_0^* - e^{-i\varphi_0} A_0^*)^{-1}\|$, тогда

$$\begin{aligned} \omega &\geq \frac{1}{2\pi} \int_{|\varphi - \varphi_0| < \frac{1}{2r}} r^2 \left(\frac{1 - 2r|\varphi - \varphi_0|}{1 - r|\varphi - \varphi_0|} \right)^2 d\varphi = \frac{r}{\pi} \int_0^{1/2} \left(\frac{1 - 2t}{1 - t} \right)^2 dt = \\ &= \frac{r}{\pi} \int_0^{1/2} \left[4 + \frac{4}{t-1} + \frac{1}{(t-1)^2} \right] dt = \frac{r}{\pi} \left(3 - 4 \int_{1/2}^1 \frac{dt}{t} \right) = \frac{r}{\pi} (3 - 4 \ln 2) > \frac{r}{5\pi}. \end{aligned}$$

Теорема доказана.

Таким образом, собственные числа нормированного пучка $\lambda B_0 - A_0$, а значит и исходного пучка $\lambda B - A$, регулярных относительно единичной окружности, отстоят от единичной окружности на расстоянии не менее $1/5\pi\omega$.

В качестве иллюстрации к теореме 1.3 приведем один пример. Пусть $B = I$, а A — симметричная матрица. Воспользуемся диагональным видом матрицы A : $A = U^* \mathcal{D} U$, где $U^* U = I$. Так как в этом случае

$$\begin{aligned} &\frac{1}{2\pi} \int_0^{2\pi} (B - e^{i\varphi} A)^{-1} (A A^* + B B^*) (B^* - e^{-i\varphi} A^*)^{-1} d\varphi = \\ &= U^* \left[\frac{1}{2\pi} \int_0^{2\pi} (I - e^{i\varphi} \mathcal{D})^{-1} (I + \mathcal{D}^2) (I - e^{-i\varphi} \mathcal{D})^{-1} d\varphi \right] U, \end{aligned}$$

то

$$\omega = \max_i \left(\left| \frac{1}{2\pi} \int_0^{2\pi} \frac{1 + d_i^2}{1 + d_i^2 - 2d_i \cos \varphi} d\varphi \right| \right) = \max_i \left| \frac{1 + d_i^2}{1 - d_i^2} \right|,$$

где d_i — диагональные элементы матрицы $\mathcal{D}$. Один из нормированных пучков $\lambda B_0 - A_0$ равен

$$U^* \frac{\lambda I - \mathcal{D}}{\sqrt{I + \mathcal{D}^2}} U,$$

и, следовательно,

$$\max_{\varphi} \|(B_0 - e^{i\varphi} A_0)^{-1}\| = \max_i \left| \frac{\sqrt{1 + d_i^2}}{1 - |d_i|} \right|.$$

Для этого примера выполняется оценка

$$\max_{\varphi} \|(B_0 - e^{i\varphi} A_0)^{-1}\| \leq \omega \max_{x \geq 0} \frac{1+x}{\sqrt{1+x^2}} = \sqrt{2}\omega.$$

Рассмотренный пример показывает, что степень ω в оценке (1.22) точна.

В вычислительной математике исключительно важен вопрос, насколько устойчив вычисляемый объект к возмущениям входных данных. Как ранее отмечалось, параметр ω отражает спектральные свойства нормированного пучка $\lambda B_0 - A_0$. Поэтому задача об оценке возмущений ω ставится следующим образом. Пусть

$$A_0 A_0^* + B_0 B_0^* = I, \quad \|(\tilde{A}_0 \tilde{B}_0) - (A_0 B_0)\| \leq \delta,$$

$$\omega = \left\| \frac{1}{2\pi} \int_0^{2\pi} (B_0 - e^{i\varphi} A_0)^{-1} (B_0^* - e^{-i\varphi} A_0^*)^{-1} d\varphi \right\|,$$

$$\tilde{\omega} = \left\| \frac{1}{2\pi} \int_0^{2\pi} (\tilde{B}_0 - e^{i\varphi} \tilde{A}_0)^{-1} (\tilde{A}_0 \tilde{A}_0^* + \tilde{B}_0 \tilde{B}_0^*) (\tilde{B}_0^* - e^{-i\varphi} \tilde{A}_0^*)^{-1} d\varphi \right\|.$$

Требуется оценить $|\tilde{\omega} - \omega|$.

Воспользуемся тождеством $(K + M)^{-1} = K^{-1} - (K + M)^{-1} M K^{-1}$.

Так как

$$\begin{aligned} (\tilde{B}_0 - e^{i\varphi} \tilde{A}_0)^{-1} &= (B_0 - e^{i\varphi} A_0)^{-1} - \\ &- (B_0 - e^{i\varphi} A_0)^{-1} [(\tilde{B}_0 - B_0) - e^{i\varphi} (\tilde{A}_0 - A_0)] (\tilde{B}_0 - e^{i\varphi} \tilde{A}_0)^{-1}, \end{aligned}$$

то формула для $\tilde{\omega}$ принимает вид

$$\begin{aligned} \tilde{\omega} &= \left\| \frac{1}{2\pi} \int_0^{2\pi} (B_0 - e^{i\varphi} A_0)^{-1} \{I - [(\tilde{B}_0 - B_0) - e^{i\varphi} (\tilde{A}_0 - A_0)] (\tilde{B}_0 - e^{i\varphi} \tilde{A}_0)^{-1}\} \times \right. \\ &\times (\tilde{A}_0 \tilde{A}_0^* + \tilde{B}_0 \tilde{B}_0^*) \{I - [(\tilde{B}_0 - B_0) - e^{i\varphi} (\tilde{A}_0 - A_0)] (\tilde{B}_0 - \\ &\left. - e^{i\varphi} \tilde{A}_0)^{-1}\}^* (B_0^* - e^{-i\varphi} A_0^*)^{-1} d\varphi \right\|. \end{aligned}$$

Следовательно,

$$|\tilde{\omega} - \omega| \leq \left\| \frac{1}{2\pi} \int_0^{2\pi} (B_0 - e^{i\varphi} A_0)^{-1} \Delta (B_0^* - e^{-i\varphi} A_0^*)^{-1} d\varphi \right\|,$$

где

$$\begin{aligned} \Delta &= \{I - [(\tilde{B}_0 - B_0) - e^{i\varphi} (\tilde{A}_0 - A_0)] (\tilde{B}_0 - e^{i\varphi} \tilde{A}_0)^{-1}\} \\ &(\tilde{A}_0 \tilde{A}_0^* + \tilde{B}_0 \tilde{B}_0^*) \{I - [(\tilde{B}_0 - B_0) - e^{i\varphi} (\tilde{A}_0 - A_0)] (\tilde{B}_0 - e^{i\varphi} \tilde{A}_0)^{-1}\}^* - I. \end{aligned}$$

Далее,

$$\begin{aligned} &\left\| \frac{1}{2\pi} \int_0^{2\pi} (B_0 - e^{i\varphi} A_0)^{-1} \Delta (B_0^* - e^{-i\varphi} A_0^*)^{-1} d\varphi \right\| = \\ &= \sup_{\|x\|=1} \left| \left(\left[\frac{1}{2\pi} \int_0^{2\pi} (B_0 - e^{i\varphi} A_0)^{-1} \Delta (B_0^* - e^{-i\varphi} A_0^*)^{-1} d\varphi \right] x, x \right) \right| = \end{aligned}$$

$$\begin{aligned}
&= \sup_{\|x\|=1} \left| \frac{1}{2\pi} \int_0^{2\pi} (\Delta (B_0^* - e^{-i\varphi} A_0^*)^{-1} x, (B_0^* - e^{-i\varphi} A_0^*)^{-1} x) d\varphi \right| \leq \\
&\leq \|\Delta\| \sup_{\|x\|=1} \left| \frac{1}{2\pi} \int_0^{2\pi} \|(B_0^* - e^{-i\varphi} A_0^*)^{-1} x\|^2 d\varphi \right| = \|\Delta\| \omega.
\end{aligned}$$

Оценим $\|\Delta\|$:

$$\begin{aligned}
&\|(\tilde{B}_0 - B_0) - e^{i\varphi}(\tilde{A}_0 - A_0)\| \leq \delta \left\| -e^{i\varphi} I \right\| = \delta \sqrt{2}, \\
&\|(\tilde{B}_0 - e^{i\varphi} \tilde{A}_0)^{-1}\| \leq \frac{\|(B_0 - e^{i\varphi} A_0)^{-1}\|}{1 - \sqrt{2}\delta \|(B_0 - e^{i\varphi} A_0)^{-1}\|}, \\
&\|\Delta\| \leq \left(1 + \frac{\sqrt{2}\delta \|(B_0 - e^{i\varphi} A_0)^{-1}\|}{1 - \sqrt{2}\delta \|(B_0 - e^{i\varphi} A_0)^{-1}\|} \right)^2 (1 + \delta)^2 - 1 \leq \\
&\leq \frac{1}{1 - 2\delta - 2\sqrt{2}\delta \|(B_0 - e^{i\varphi} A_0)^{-1}\|} - 1 \leq \frac{2\delta + 10\sqrt{2}\pi\delta\omega}{1 - 2\delta - 10\sqrt{2}\pi\delta\omega}.
\end{aligned}$$

В итоге

$$|\tilde{\omega} - \omega| \leq \omega \frac{2\delta + 10\sqrt{2}\pi\delta\omega}{1 - 2\delta - 10\sqrt{2}\pi\delta\omega}. \quad (1.23)$$

Естественно, что эта оценка верна только при выполнении неравенства $2\delta + 10\sqrt{2}\pi\delta\omega < 1$. Учитывая неравенство $\omega \geq 1$, которое будет доказано позднее, оценку (1.23) можно несколько упростить

$$\left| \frac{\tilde{\omega} - \omega}{\omega} \right| < \frac{47\omega\delta}{1 - 47\omega\delta}. \quad (1.24)$$

Целью последующего изложения будет установление структуры матрицы H с помощью последовательности матриц P_k . Напомним, что

$$H = \frac{1}{2\pi} \int_0^{2\pi} (B_0 - e^{i\varphi} A_0)^{-1} (B_0^* - e^{-i\varphi} A_0^*)^{-1} d\varphi,$$

где $A_0 = L^{-1}A$, $B_0 = L^{-1}B$, $LL^* = AA^* + BB^*$. Вследствие равенства $A_0 A_0^* + B_0 B_0^* = I$ матрица T_0 из канонического вида пучка

$$\lambda B_0 - A_0 = \lambda T_0^{-1} (P_{+0} + P_{-1}) - T_0^{-1} (P_{-0} + P_{+1})$$

удовлетворяет тождеству

$$T_0 T_0^* = (P_{-0} + P_{+1})(P_{-0}^* + P_{+1}^*) + (P_{+0} + P_{-1})(P_{+0}^* + P_{-1}^*). \quad (1.25)$$

Из (1.17) следует, что

$$(B_0 - e^{i\varphi} A_0)^{-1} = \sum_{j=-\infty}^{\infty} P_j T_0 e^{i\varphi}. \quad (1.26)$$

По равенству Парсеваля для функции $(B_0 - e^{i\varphi} A_0)^{-1}$ получаем представление

$$H = \sum_{j=-\infty}^{\infty} P_j T_0 T_0^* P_j^*. \quad (1.27)$$

Подставим (1.25) в (1.27) и раскроем скобки, используя (1.11). В результате получим важное тождество

$$H = 2 \sum_{j=-\infty}^{\infty} P_j P_j^* - P_{+0} P_{+0}^* - P_{-0} P_{-0}^*. \quad (1.28)$$

Это тождество позволяет вывести неравенство

$$\omega \geq 1. \quad (1.29)$$

Действительно,

$$\begin{aligned} H &\geq P_{+0}P_{+0}^* + P_{-0}P_{-0}^*, \\ \omega = \max_{\|x\|=1} |(Hx, x)| &\geq \max_{\|x\|=1} |([P_{+0}P_{+0}^* + P_{-0}P_{-0}^*]x, x)| = \\ &= \max_{\|x\|=1} (\|P_{+0}^*x\|^2 + \|P_{-0}^*x\|^2) \geq 1. \end{aligned} \quad (1.30)$$

В последнем неравенстве достаточно взять x , удовлетворяющий равенству $P_{+0}^*x = x$, если $P_{+0} \neq 0$, или $P_{-0}^*x = 0$, если $P_{-0} \neq 0$.

Огромное значение имеет эффективная оценка матриц P_k с помощью параметра ω . Она позволяет обосновать многие вычислительные процедуры для исследования спектра несимметричных задач. Приступим к выводу этой важной оценки.

При $k \geq 1$ справедлива цепочка неравенств

$$\begin{aligned} (P_0HP_0^*P_k^*f, P_k^*f) &= (P_1HP_1^*P_{k-1}^*f, P_{k-1}^*f) = (P_0HP_0^*P_{k-1}^*f, P_{k-1}^*f) - \\ &- ([P_0P_0^* + P_1P_1^*]P_{k-1}^*f, P_{k-1}^*f) = (P_0HP_0^*P_{k-1}^*f, P_{k-1}^*f) - \\ &- (P_{k-1}^*f, P_{k-1}^*f) - (P_k^*f, P_k^*f), \\ (P_k^*f, P_k^*f) + (P_0HP_0^*P_k^*f, P_k^*f) &= (P_{k-1}^*f, P_{k-1}^*f) + (P_0HP_0^*P_{k-1}^*f, P_{k-1}^*f) - \\ &- 2(P_{k-1}^*f, P_{k-1}^*f) \leq \left(1 - \frac{2}{1 + \|H\|}\right) \{(P_{k-1}^*f, P_{k-1}^*f) + \\ &+ (P_0HP_0^*P_{k-1}^*f, P_{k-1}^*f)\}, \\ (P_k^*f, P_k^*f) + (P_0HP_0^*P_k^*f, P_k^*f) &\leq \left(1 - \frac{2}{1 + \|H\|}\right)^k \{(P_0^*f, P_0^*f) + \\ &+ (P_0HP_0^*f, f)\} \leq \left(1 - \frac{2}{1 + \|H\|}\right)^k \cdot 2\|H\| \|f\|^2. \end{aligned}$$

Поэтому при $k \geq 0$ выполнено неравенство

$$(P_k^*f, P_k^*f) = (P_0^*P_k^*f, P_0^*P_k^*f) = (P_0P_0^*P_k^*f, P_k^*f) \leq (P_0HP_0^*P_k^*f, P_k^*f),$$

которое дает следующие оценки:

$$\begin{aligned} \|P_k\|^2 &\leq \|H\| \left(1 - \frac{2}{1 + \|H\|}\right)^k, \\ \|P_k\| &\leq V_{\omega}^{-} \left(1 - \frac{1}{1 + \omega}\right)^k \leq V_{\omega}^{-} e^{-\frac{k}{1 + \omega}}, \quad k \geq 0. \end{aligned}$$

Точно так же рассматривается случай $k \leq 0$. В итоге нами доказана следующая теорема, которая чрезвычайно важна для исследования сходимости степенных методов вычисления спектральных характеристик.

Теорема 1.4. Если $\lambda B - A$ — регулярный относительно единичной окружности пучок матриц, то

$$\|P_k\| \leq V_{\omega}^{-} \left(1 - \frac{1}{1 + \omega}\right)^{|k|} \leq V_{\omega}^{-} e^{-\frac{|k|}{1 + \omega}}. \quad (1.31)$$

Применим полученные выше результаты для оценки числа обусловленности матрицы T_0 из канонической формы нормированного пучка матриц $\lambda B_0 - A_0$. Из тождества (1.25) следует, что

$$\begin{aligned} \sigma_{\max}(T_0) &= \sigma_{\max}([(P_{+0} + P_{-1}), (P_{-0} + P_{+1})]), \\ \|T_0\| &\leq \|P_{+0}P_{+1}\| + \|P_{-1}P_{-0}\| \leq 2V_{\omega}^{-}. \end{aligned}$$

Аналогично

$$\sigma_{\min}(T_0) = \sigma_{\min}([(P_{+0} + P_{-1}), (P_{-0} + P_{+1})]).$$

Обозначим $f = x(P_{+0} + P_{-1})$, $g = x(P_{-0} + P_{+1})$. Тогда $fP_{+0} + gP_{-0} = xP_{+0} + xP_{-0} = x$,

$$\|x\| \leq \|fg\| \left\| \begin{pmatrix} P_{+0} \\ P_{-0} \end{pmatrix} \right\|, \quad \sigma_{\min}(T_0) \geq 1 \left\| \begin{pmatrix} P_{+0} \\ P_{-0} \end{pmatrix} \right\|.$$

Покажем, что $\left\| \begin{pmatrix} P_{+0} \\ P_{-0} \end{pmatrix} \right\| = \|P_{+0}P_{-0}\|$. Для этого воспользуемся формой Шура (1.9) матриц P_{+0} и P_{-0} . Имеем

$$P_{+0}P_{+0}^* + P_{-0}P_{-0}^* = U^* \begin{pmatrix} I + 2LL^* & -L \\ -L^* & I \end{pmatrix} U,$$

$$P_{+0}^*P_{+0} + P_{-0}^*P_{-0} = U^* \begin{pmatrix} I & L \\ L^* & I + 2L^*L \end{pmatrix} U.$$

С помощью сингулярного разложения матрицы L можно показать, что

$$\left\| \begin{pmatrix} I + 2LL^* & -L \\ -L^* & I \end{pmatrix} \right\| = \left\| \begin{pmatrix} I & L \\ L^* & I + 2L^*L \end{pmatrix} \right\|.$$

Пусть, к примеру, $L = U^*(\Sigma 0)V$, $U^*U = I$, $V^*V = I$, Σ — диагональная матрица с положительными элементами на диагонали. Тогда

$$\begin{pmatrix} I + 2LL^* & -L \\ -L^* & I \end{pmatrix} = \begin{pmatrix} U^* & 0 \\ 0 & V^* \end{pmatrix} \begin{pmatrix} I + 2\Sigma^2 & -\Sigma & 0 \\ -\Sigma & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix},$$

$$\begin{pmatrix} I & L \\ L^* & I + 2L^*L \end{pmatrix} = \begin{pmatrix} U^* & 0 \\ 0 & V^* \end{pmatrix} \begin{pmatrix} I & \Sigma & 0 \\ \Sigma & I + 2\Sigma^2 & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix},$$

откуда следует требуемое равенство норм.

Подытожим все эти рассуждения:

$$\sigma_{\max}(T_0) \leq 2\sqrt{\omega}, \quad \sigma_{\min}(T_0) \geq 1/\sqrt{\omega}. \quad (1.32)$$

Обусловленность матрицы T_0 удовлетворяет оценке $\mu(T_0) \leq 2\omega$.

§ 2. АЛГОРИТМ ВЫЧИСЛЕНИЯ ПРОЕКТОРОВ НА ИНВАРИАНТНЫЕ ПОДПРОСТРАНСТВА РЕГУЛЯРНОГО ОТНОСИТЕЛЬНО ЕДИНИЧНОЙ ОКРУЖНОСТИ ЛИНЕЙНОГО ПУЧКА МАТРИЦ

2.1. Описание алгоритма

Начнем с разбора наводящих соображений, которые приводят к конструкции предлагаемого алгоритма вычисления инвариантных подпространств регулярного относительно единичной окружности пучка матриц $\lambda B - A$.

Систему разностных уравнений (1.16) для бесконечной последовательности матриц G_k заменим, пользуясь соотношениями (1.17), системой разностных уравнений, связывающих матрицы

$$\dots, P_{-2}, P_{-1}, P_{-0}, P_{+0}, P_{+1}, P_{+2}, \dots$$

Эта система имеет вид

$$\begin{aligned} BP_k - AP_{k-1} &= 0, \quad k \leq -1, \\ P_{+0} + P_{-0} &= I, \\ BP_k - AP_{k-1} &= 0, \quad k \geq 1. \end{aligned} \quad (2.1)$$

Для численного решения система (2.1) непригодна из-за бесконечного числа уравнений. Чтобы преодолеть этот недостаток, предлагается модификация системы (2.1), основанная на факте быстрого убывания нормы матриц P_k при $|k| \rightarrow \infty$.

Определим при целых k матрицы $P_k^{(n)} = \sum_{p=-\infty}^{\infty} P_{k+np}$, где n — достаточно большое натуральное четное число. Для $k = np$ определяются две матрицы $P_{k-0}^{(n)}$ и $P_{k+0}^{(n)}$. Пользуясь теоремой 1.4, можно показать, что при $-n/2 \leq k \leq n/2$ справедливо неравенство

$$\|P_k - P_k^{(n)}\| \leq 2 \sqrt{\omega} \frac{e^{-\frac{n}{2(1+\omega)}}}{1 - e^{-\frac{n}{1+\omega}}},$$

означающее при большом n близость матриц P_k и $P_k^{(n)}$ для $|k| \leq n/2$.

Последовательность матриц $P_k^{(n)}$ периодическая с периодом n и удовлетворяет системе матричных уравнений

$$\begin{aligned} BP_k^{(n)} - AP_{k-1}^{(n)} &= 0, \quad k \neq np + 0, \\ P_{k+0}^{(n)} + P_{k-0}^{(n)} &= I, \quad k = np + 0. \end{aligned}$$

Рассмотрим эту систему на периоде $[+0, n - 0]$:

$$\begin{aligned} Z_0 + Z_n &= I, \\ BZ_k - AZ_{k-1} &= 0, \quad 1 \leq k \leq n, \end{aligned} \quad (2.2)$$

где $Z_k = P_k^{(n)}$. При достаточно большом n матрицы $Z_0, Z_1, \dots, Z_{n/2}$ являются приближениями для $P_{+0}, P_1, \dots, P_{n/2}$, а матрицы $Z_{n/2}, \dots, Z_n$ — приближениями матриц $P_{-n/2}, \dots, P_{-0}$.

Нас интересует главным образом проекторы $P_{+0} = P_0$ и $P_{-0} = P_\infty$. В этом случае, оказывается, можно организовать быстрый ортогональный процесс вычисления приближений к этим проекторам — матриц Z_0 и Z_n . Пусть $n = 2^{m_0}$ и при некотором m , $0 \leq m < m_0$, выполнены разностные соотношения вида

$$B_m Z_k - A_m Z_{k-2^m} = 0. \quad (2.3)$$

Выпишем уравнение (2.3) для матриц Z_{k+2^m} и Z_k :

$$B_m Z_{k+2^m} - A_m Z_k = 0. \quad (2.4)$$

Умножим слева уравнение (2.3) на некоторую матрицу X_m , а уравнение (2.4) на матрицу Y_m и результаты сложим:

$$Y_m B_m Z_{k+2^m} + (X_m B_m - Y_m A_m) Z_k - X_m A_m Z_{k-2^m} = 0. \quad (2.5)$$

В качестве X_m и Y_m возьмем $N \times N$ -матрицы, удовлетворяющие системе матричных уравнений:

$$\begin{aligned} X_m X_m^* + Y_m Y_m^* &= I, \\ X_m B_m - Y_m A_m &= 0. \end{aligned} \quad (2.6)$$

Способ нахождения таких матриц X_m и Y_m будет описан ниже.

Используя обозначения $A_{m+1} = X_m A_m$, $B_{m+1} = Y_m B_m$, уравнение (2.5) перепишем следующим образом:

$$B_{m+1} Z_k - A_{m+1} Z_{k-2^{m+1}} = 0.$$

В конце итерационного процесса ортогонального исключения получим линейную систему, связывающую матрицы Z_0 и $Z_{2^{m_0}}$

$$\begin{aligned} Z_0 + Z_{2^{m_0}} &= I, \\ B_{m_0} Z_{2^{m_0}} - A_{m_0} Z_0 &= 0. \end{aligned} \quad (2.7)$$

Решением этой системы, как нетрудно убедиться, будут матрицы $Z_0 = (A_{m_0} + B_{m_0})^{-1} B_{m_0}$, $Z_{m_0} = (A_{m_0} + B_{m_0})^{-1} A_{m_0}$. При большом m_0 матрица $(A_{m_0} + B_{m_0})^{-1} B_{m_0}$ будет близка к проектору P_0 , а матрица $(A_{m_0} + B_{m_0})^{-1} A_{m_0}$ — к P_∞ .

Подводя итоги этих рассуждений, сформулируем основные этапы вычислительного процесса для нахождения проекторов P_0 и P_∞ на инвариантные подпространства регулярного относительно единичной окружности пучка матриц $\lambda B - A$, отвечающие собственным числам внутри единичного круга и вне.

Этап I. Пучок $\lambda B - A$ нормируется подходящим образом, т. е. находятся такие матрицы B_0 и A_0 , что $\lambda B - A = L(\lambda B_0 - A_0)$ с невырожденной матрицей L . Как правило, это будет нормировка условием $A_0 A_0^* + B_0 B_0^* = I$.

Этап II. Выполняются итерации перехода от пары матриц A_m, B_m к паре A_{m+1}, B_{m+1} для $0 \leq m < m_0$: $A_{m+1} = X_m A_m$, $B_{m+1} = Y_m B_m$, где X_m и Y_m удовлетворяют системе (2.6). Алгоритмически вычисление матриц A_{m+1} и B_{m+1} сводится к подбору ортогонального преобразования, применение которого слева к матрице

$$\begin{pmatrix} -A_m & B_m & 0 \\ 0 & \boxed{-A_m} & B_m \end{pmatrix}$$

аннулирует блок, обведенный прямоугольной рамкой. В результате этого применения получаем матрицу

$$\begin{pmatrix} C_{m+1} & R_{m+1} & D_{m+1} \\ -A_{m+1} & 0 & B_{m+1} \end{pmatrix}$$

с некоторыми матрицами C_{m+1}, R_{m+1} и D_{m+1} , которые, впрочем, нас не интересуют. Нетрудно видеть, что матрицы A_{m+1}, B_{m+1} являются результатом вычислений по формулам $A_{m+1} = X_m A_m$, $B_{m+1} = Y_m B_m$ с некоторыми X_m и Y_m , связанными соотношениями (2.6). Сами матрицы X_m и Y_m в процессе работы алгоритма явно не вычисляются. Они будут использоваться только в обосновании предлагаемого алгоритма.

Этап III. Выбрав некоторое достаточно большое m_0 , решим линейную систему

$$\begin{aligned} Z_0 + Z_\infty &= I, \\ A_{m_0} Z_0 - B_{m_0} Z_\infty &= 0 \end{aligned} \quad (2.8)$$

относительно $N \times N$ -матриц Z_0 и Z_∞ . При $m_0 \rightarrow \infty$ матрицы Z_0 и Z_∞ будут все более точными приближениями к проекторам P_0 и P_∞ пучка матриц $\lambda B - A$.

Если есть необходимость в вычислении ортогональных проекторов Π_0 и Π_∞ на инвариантные подпространства $\mathcal{L}_0$ и $\mathcal{L}_\infty$ пучка $\lambda B - A$, то можно вместо вычисления проекторов P_0 и P_∞ и последующего применения формул

$$\Pi_0 = P_0 (P_0^* P_0 + P_\infty^* P_\infty)^{-1} P_0^*, \quad \Pi_\infty = P_\infty (P_0^* P_0 + P_\infty^* P_\infty)^{-1} P_\infty^*$$

поступить следующим образом. Сначала вычислим нормированный пучок $\lambda \bar{B}_{m_0} - \bar{A}_{m_0}$ из пучка $\lambda B_{m_0} - A_{m_0}$, где $\bar{A}_{m_0} \bar{A}_{m_0}^* + \bar{B}_{m_0} \bar{B}_{m_0}^* = I$, а затем определим матрицы $S_0 = I - \bar{A}_{m_0}^* \bar{A}_{m_0}$, $S_\infty = I - \bar{B}_{m_0}^* \bar{B}_{m_0}$. Утверждается, что при $m_0 \rightarrow \infty$ матрицы S_0 и S_∞ стремятся к Π_0 и Π_∞ соответственно.

2.2. Обоснование сходимости алгоритма

Начнем с исследования главного второго этапа алгоритма. Напомним, что перед этим этапом пара A_0, B_0 имеет канонический вид $A_0 = T_0^{-1}(P_{-0} + P_{+1})$, $B_0 = T_0^{-1}(P_{+0} + P_{-1})$ с невырожденной $N \times N$ -матрицей T_0 . Покажем, что в дальнейшем для матриц A_m, B_m справедливо следующее представление:

$$A_m = T_m^{-1}(P_{-0} + P_{2m}), B_m = T_m^{-1}(P_{+0} + P_{-2m}). \quad (2.9)$$

Чтобы продемонстрировать это, перепишем уравнение $X_m B_m - Y_m A_m = 0$, используя разложение (1.1):

$$\begin{aligned} (X_m T_m^{-1})(P_{+0} + P_{-2m}) - (Y_m T_m^{-1})(P_{-0} + P_{2m}) &= 0, \\ \left[(X_m T_m^{-1} Q_r) \begin{pmatrix} I_{N_0} & 0 \\ 0 & J_\infty^{2m} \end{pmatrix} - (Y_m T_m^{-1} Q_r) \begin{pmatrix} J_0^{2m} & 0 \\ 0 & I_{N_\infty} \end{pmatrix} \right] Q_r^{-1} &= 0. \end{aligned}$$

Отсюда следует, что все решения уравнения $X_m B_m - Y_m A_m = 0$ представимы в виде

$$X_m = F_{m+1} \begin{pmatrix} J_0^{2m} & 0 \\ 0 & I_{N_\infty} \end{pmatrix} Q_r^{-1} T_m, \quad Y_m = F_{m+1} \begin{pmatrix} I_{N_0} & 0 \\ 0 & J_\infty^{2m} \end{pmatrix} Q_r^{-1} T_m$$

с произвольной матрицей F_{m+1} .

Так как матрицы X_m и Y_m должны удовлетворять уравнению $X_m X_m^* + Y_m Y_m^* = I$, то матрица F_{m+1} должна быть невырожденной. Следовательно, решение системы (2.6) представимо в виде

$$X_m = T_{m+1}^{-1}(P_{-0} + P_{2m}) T_m, \quad Y_m = T_{m+1}^{-1}(P_{+0} + P_{-2m}) T_m, \quad (2.10)$$

где $T_{m+1} = Q_r F_{m+1}^{-1}$.

Подставим (2.10) в тождество $X_m X_m^* + Y_m Y_m^* = I$, после простых преобразований получим рекуррентное соотношение для матриц T_m :

$$\begin{aligned} T_{m+1} T_{m+1}^* &= (P_{-0} + P_{2m}) T_m T_m^* (P_{-0}^* + P_{2m}^*) + \\ &+ (P_{+0} + P_{-2m}) T_m T_m^* (P_{+0}^* + P_{-2m}^*). \end{aligned} \quad (2.11)$$

Из формулы (2.11) следует, что решение системы (2.6) единственно с точностью до ортогонального множителя слева, т. е. все решения системы (2.6) представимы в виде $X_m = U_m \bar{X}_m$, $Y_m = U_m \bar{Y}_m$, где U_m — произвольная унитарная матрица, а $\bar{X}_m, \bar{Y}_m$ — некоторое частное решение системы (2.6).

Приступим к исследованию третьего этапа алгоритма. Используя представление (2.9), матричную систему (2.8) перепишем следующим образом:

$$\begin{aligned} Z_0 + Z_\infty &= I, \\ T_{m_0}^{-1} \{ (P_{-0} + P_{2m_0}) Z_0 - (P_{+0} + P_{-2m_0}) Z_\infty \} &= 0. \end{aligned} \quad (2.12)$$

Пусть матрица ξ такова, что $Z_0 = P_{+0} + \xi$, а $Z_\infty = P_{-0} - \xi$. Тогда из второго уравнения системы (2.12) получим

$$\begin{aligned} (P_{-0} + P_{2m_0})(P_{+0} + \xi) - (P_{+0} + P_{-2m_0})(P_{-0} - \xi) &= 0, \\ (I + P_{2m_0} + P_{-2m_0}) \xi &= P_{-2m_0} - P_{2m_0}. \end{aligned}$$

Теорема 1.4 гарантирует оценку

$$\|P_{\pm 2m_0}\| \leq \delta = \sqrt{\omega e^{-\frac{2m_0}{1+\omega}}}.$$

Предположим, что m_0 достаточно велико и выполнена оценка $2\delta < 1$. Тогда

$$\|P_{-2m_0} - P_{2m_0}\| \leq 2\delta, \quad \sigma_{\min}(I + P_{2m_0} + P_{-2m_0}) \geq 1 - 2\delta, \quad \|\xi\| \leq \frac{2\delta}{1-\delta}.$$

В итоге получаем выражение для скорости сходимости алгоритма:

$$\|Z_{\infty} - P_{\infty}\| \leq \frac{2\sqrt{\omega}e^{-\frac{2m_0}{1+\omega}}}{1 - 2\sqrt{\omega}e^{-\frac{2m_0}{1+\omega}}}. \quad (2.13)$$

Эта оценка справедлива для тех значений m_0 , для которых положительны знаменатель.

Теперь остановимся на вычислении ортогональных проекторов Π_0 и Π_{∞} . Напомним, что вначале проводится нормировка пучка $\lambda B_{m_0} - A_{m_0}$ условием $\bar{A}_{m_0} \bar{A}_{m_0}^* + \bar{B}_{m_0} \bar{B}_{m_0}^* = I$. Обозначим каноническую форму нормированного пучка $\lambda \bar{B}_{m_0} - \bar{A}_{m_0}$ так:

$$\bar{A}_{m_0} = \bar{T}_{m_0}^{-1}(P_{-0} + P_{2m_0}), \quad \bar{B}_{m_0} = \bar{T}_{m_0}^{-1}(P_{+0} + P_{-2m_0}).$$

Тогда матрица $\bar{T}_{m_0}$ будет удовлетворять тождеству

$$\bar{T}_{m_0} \bar{T}_{m_0}^* = (P_{-0} + P_{2m_0})(P_{-0}^* + P_{2m_0}^*) + (P_{+0} + P_{-2m_0})(P_{+0}^* + P_{-2m_0}^*).$$

Отсюда следует, что

$$\begin{aligned} I - \bar{A}_{m_0}^* \bar{A}_{m_0} &= I - (P_{-0}^* + P_{2m_0}^*) [(P_{-0} + P_{2m_0})(P_{-0}^* + P_{2m_0}^*) + \\ &\quad + (P_{+0} + P_{-2m_0})(P_{+0}^* + P_{-2m_0}^*)]^{-1} (P_{-0} + P_{2m_0}), \\ I - \bar{B}_{m_0}^* \bar{B}_{m_0} &= I - (P_{+0}^* + P_{-2m_0}^*) [(P_{-0} + P_{2m_0})(P_{-0}^* + P_{2m_0}^*) + \\ &\quad + (P_{+0} + P_{-2m_0})(P_{+0}^* + P_{-2m_0}^*)]^{-1} (P_{+0} + P_{-2m_0}). \end{aligned}$$

При $m_0 \rightarrow \infty$, очевидно, $I - \bar{A}_{m_0}^* \bar{A}_{m_0} \rightarrow \Pi_0$, $I - \bar{B}_{m_0}^* \bar{B}_{m_0} \rightarrow \Pi_{\infty}$. Оценим скорость сходимости к ортогональным проекторам. Обозначим $R = P_{-0} \times \times P_{-0}^* + P_{+0} P_{+0}^*$, тогда $R \geq \frac{1}{2} I$. Действительно,

$$\|xP_{+0}\|^2 + \|xP_{-0}\|^2 = xRx^* \geq \frac{1}{2} (\|xP_{+0}\| + \|xP_{-0}\|)^2 \geq \frac{1}{2} \|x\|^2.$$

Далее,

$$\begin{aligned} &\|(P_{-0} + P_{2m_0})(P_{-0}^* + P_{2m_0}^*) + (P_{+0} + P_{-2m_0})(P_{+0}^* + P_{-2m_0}^*) - R\| \leq \\ &\leq 2\|P_{2m_0}\|\|P_{-0}\| + \|P_{2m_0}\|^2 + 2\|P_{-2m_0}\|\|P_{+0}\| + \|P_{-2m_0}\|^2 \leq \\ &\leq 4\sqrt{\omega}\sqrt{\omega}e^{-\frac{2m_0}{1+\omega}} + 2\omega e^{-\frac{2m_0}{1+\omega}} \leq \frac{4\omega e^{-\frac{2m_0}{1+\omega}}}{1 - \frac{1}{2}e^{-\frac{2m_0}{1+\omega}}} = \delta_1, \end{aligned}$$

$$\begin{aligned} &\|[(P_{-0} + P_{2m_0})(P_{-0}^* + P_{2m_0}^*) + (P_{+0} + P_{-2m_0})(P_{+0}^* + P_{-2m_0}^*)]^{-1} - R^{-1}\| \leq \\ &\leq \frac{4\delta_1}{1 - 2\delta_1} < \frac{16\omega e^{-\frac{2m_0}{1+\omega}}}{1 - 9\omega e^{-\frac{2m_0}{1+\omega}}} = \delta_2, \end{aligned}$$

$$\begin{aligned}
& \|I - \bar{A}_{m_0}^* \bar{A}_{m_0} - \Pi_0\| \leq \delta_2 \|P_{-0}\|^2 + 4 \|P_{2m_0}\| \|P_{-0}\| + 2\delta_2 \|P_{2m_0}\| \|P_{-0}\| + \\
& + 2 \|P_{2m_0}\|^2 + \delta_2 \|P_{2m_0}\|^2 \leq \delta_2 \omega + 4\omega e^{-\frac{2m_0}{1+\omega}} + 2\delta_2 \omega e^{-\frac{2m_0}{1+\omega}} + 2\omega e^{-\frac{2 \cdot 2m_0}{1+\omega}} + \\
& + \delta_2 \omega e^{-\frac{2 \cdot 2m_0}{1+\omega}} = \delta_2 \omega \left(1 + e^{-\frac{2m_0}{1+\omega}}\right)^2 + 2\omega \left[\left(1 + e^{-\frac{2m_0}{1+\omega}}\right)^2 - 1\right] < \\
& < \frac{16\omega^2 e^{-\frac{2m_0}{1+\omega}}}{1 - 9\omega e^{-\frac{2m_0}{1+\omega}}} \frac{1}{1 - 2e^{-\frac{2m_0}{1+\omega}}} + 2\omega \frac{2e^{-\frac{2m_0}{1+\omega}}}{1 - 2e^{-\frac{2m_0}{1+\omega}}} < \\
& < \frac{16\omega^2 e^{-\frac{2m_0}{1+\omega}} + 4\omega e^{-\frac{2m_0}{1+\omega}}}{1 - 9\omega e^{-\frac{2m_0}{1+\omega}} - 2e^{-\frac{2m_0}{1+\omega}}} < \frac{20\omega^2 e^{-\frac{2m_0}{1+\omega}}}{1 - 11\omega e^{-\frac{2m_0}{1+\omega}}}. \tag{2.14}
\end{aligned}$$

Аналогично

$$\|I - \bar{B}_{m_0}^* \bar{B}_{m_0} - \Pi_\infty\| < \frac{20\omega^2 e^{-\frac{2m_0}{1+\omega}}}{1 - 11\omega e^{-\frac{2m_0}{1+\omega}}}. \tag{2.15}$$

Естественно, что оценки (2.14) и (2.15) справедливы при достаточно большом m_0 , когда $1 - 11\omega e^{-\frac{2m_0}{1+\omega}} > 0$.

Кроме вычисления проекторов на инвариантные подпространства пучка матриц предложенный алгоритм имеет еще одно неожиданное свойство: на третьем этапе легко вычисляется матрица H .

Проследим за поведением матриц T_m . По индукции из (2.11) легко вывести формулу

$$T_k T_k^* = \sum_{j=0}^{2^k-1} (P_j + P_{j+1-2^k}) T_0 T_0^* (P_j^* + P_{j+1-2^k}^*), \tag{2.16}$$

где $P_{j+1-2^k} = P_{-0}$ при $j = 2^k - 1$. Для $k = 0$ формула (2.16), очевидно, справедлива. Переход от k к $k+1$ обосновывается следующей выкладкой:

$$\begin{aligned}
& (P_{-0} + P_{2^k}) \left[\sum_{j=0}^{2^k-1} (P_j + P_{j+1-2^k}) T_0 T_0^* (P_j^* + P_{j+1-2^k}^*) \right] (P_{-0}^* + P_{2^k}^*) + \\
& + (P_{+0} + P_{-2^k}) \left[\sum_{j=0}^{2^k-1} (P_j + P_{j+1-2^k}) T_0 T_0^* (P_j^* + P_{j+1-2^k}^*) \right] (P_{+0}^* + P_{-2^k}^*) = \\
& = \sum_{j=0}^{2^k-1} (P_{j+2^k} + P_{j+1-2^k}) T_0 T_0^* (P_{j+2^k}^* + P_{j+1-2^k}^*) + \\
& + \sum_{j=0}^{2^k-1} (P_j + P_{j+1-2^{k+1}}) T_0 T_0^* (P_j^* + P_{j+1-2^{k+1}}^*) = \\
& = \sum_{j=0}^{2^{k+1}-1} (P_j + P_{j+1-2^{k+1}}) T_0 T_0^* (P_j^* + P_{j+1-2^{k+1}}^*).
\end{aligned}$$

Перейдем к пределу при $k \rightarrow \infty$ в тождестве (2.16). Для этого оценим

разность

$$\begin{aligned} & \left\| T_k T_k^* - \sum_{j=-\infty}^{\infty} P_j T_0 T_0^* P_j^* \right\| \leq \left\| \sum_{|j| \geq 2^k} P_j T_0 T_0^* P_j^* \right\| + 2 \left\| \sum_{j=0}^{2^k-1} P_j T_0 T_0^* P_{j+1-2^k}^* \right\| \leq \\ & \leq \|T_0\|^2 \cdot 2 \cdot \left[\frac{e^{\frac{2^k+1}{1+\omega}}}{1 - e^{-\frac{2}{1+\omega}}} + \omega \sum_{j=0}^{2^k-1} e^{-\frac{2^k-1}{1+\omega}} \right] = 2\omega \|T_0\|^2 \left[\frac{e^{\frac{2^k+1}{1+\omega}}}{1 - e^{-\frac{2}{1+\omega}}} + 2^k e^{-\frac{2^k-1}{1+\omega}} \right]. \end{aligned}$$

Последняя оценка позволяет утверждать, что

$$T_k T_k^* \rightarrow \sum_{j=-\infty}^{\infty} P_j T_0 T_0^* P_j^* \quad (2.17)$$

при $k \rightarrow \infty$. Из представления (2.9) следует формула

$$(A_{m_0} + B_{m_0})^{-1} = (I + P_{2^k m_0} + P_{-2^k m_0})^{-1} T_{m_0},$$

откуда при $m_0 \rightarrow \infty$ выводим, что

$$(A_{m_0} + B_{m_0})^{-1} (A_{m_0}^* + B_{m_0}^*)^{-1} \rightarrow \sum_{j=-\infty}^{\infty} P_j T_0 T_0^* P_j^*. \quad (2.18)$$

Так как $A_0 A_0^* + B_0 B_0^* = I$, то из (1.27) следует, что

$$\begin{aligned} & (A_{m_0} + B_{m_0})^{-1} (A_{m_0}^* + B_{m_0}^*)^{-1} \rightarrow \\ & \rightarrow H = \frac{1}{2\pi} \int_0^{2\pi} (B - e^{i\varphi} A)^{-1} (A A^* + B B^*) (B^* - e^{-i\varphi} A^*)^{-1} d\varphi. \quad (2.19) \end{aligned}$$

В заключение займемся уточнением скорости сходимости в (2.19). Справедлива цепочка равенств

$$\begin{aligned} T_k T_k^* &= (P_{-0} P_{-0}^* + 2P_{-1} P_{-1}^* + \dots + 2P_{-2^k+1} P_{-2^k+1}^* + P_{-2^k} P_{-2^k}^*) + \\ &+ (P_{+0} P_{+0}^* + 2P_1 P_1^* + \dots + 2P_{2^k-1} P_{2^k-1}^* + P_{2^k} P_{2^k}^*) + \\ &+ (P_{-2^k} P_{+0}^* + 2P_{-2^k+1} P_1^* + \dots + 2P_{-1} P_{2^k-1}^* + P_{-0} P_{2^k}^*) + \\ &+ (P_{2^k} P_{-0}^* + 2P_{2^k-1} P_{-1}^* + \dots + 2P_{+1} P_{-2^k+1}^* + P_{+0} P_{-2^k}^*), \\ H - T_k T_k^* &= P_{-2^k} H P_{-2^k}^* + P_{2^k} H P_{2^k}^* - \\ &- (P_{-2^k} P_{+0}^* + 2P_{-2^k+1} P_1^* + \dots + 2P_{-1} P_{2^k-1}^* + P_{-0} P_{2^k}^*) - \\ &- (P_{2^k} P_{-0}^* + 2P_{2^k-1} P_{-1}^* + \dots + 2P_1 P_{-2^k+1}^* + P_{+0} P_{-2^k}^*). \end{aligned}$$

Отсюда следует неравенство

$$\|H - T_k T_k^*\| \leq 2\omega^2 e^{-\frac{2^k+1}{1+\omega}} + 2^{k+2} \omega e^{-\frac{2^k}{1+\omega}}.$$

Обозначим $H_k = (A_k + B_k)^{-1} (A_k^* + B_k^*)^{-1}$, тогда

$$H_k - T_k T_k^* = \Delta_k T_k T_k^* + T_k T_k^* \Delta_k^* + \Delta_k T_k T_k^* \Delta_k^*$$

с $\Delta_k = (I + P_{2^k} + P_{-2^k})^{-1} - I$. Воспользуемся теоремой 1.4 для получения оценки

$$\|\Delta_k\| \leq \frac{\|P_{2^k} + P_{-2^k}\|}{1 - \|P_{2^k} + P_{-2^k}\|} \leq \frac{2\sqrt{\omega} e^{-\frac{2^k}{1+\omega}}}{1 - 2\sqrt{\omega} e^{-\frac{2^k}{1+\omega}}}.$$

Далее,

$$\begin{aligned} \|H_k - T_k T_k^*\| &\leq \|\Delta\| (2 + \|\Delta\|) \|T_k T_k^*\| \leq \\ &\leq \|\Delta\| (2 + \|\Delta\|) [\|H\| + \|H - T_k T_k^*\|], \\ \|H - H_k\| &\leq \|H - T_k T_k^*\| + \|H_k - T_k T_k^*\| \leq \\ &\leq \|\Delta\| (2 + \|\Delta\|) \|H\| + (1 + \|\Delta\|)^2 \|H - T_k T_k^*\| = \\ &= [(1 + \|\Delta\|)^2 - 1] \|H\| + (1 + \|\Delta\|)^2 \|H - T_k T_k^*\|. \end{aligned}$$

В итоге

$$\begin{aligned} \|H - (A_k + B_k)^{-1} (A_k^* + B_k^*)^{-1}\| &\leq \\ &\leq \frac{4\omega^{3/2} e^{-\frac{2k}{1+\omega}} + 2\omega^2 e^{-\frac{2k+1}{1+\omega}} + 2^{k+2}\omega e^{-\frac{2k}{1+\omega}}}{1 - 4\sqrt{\omega} e^{-\frac{2k}{1+\omega}}}. \end{aligned} \quad (2.20)$$

Эта оценка, конечно, справедлива, когда знаменатель $1 - 4\sqrt{\omega} e^{-\frac{2k}{1+\omega}} > 0$.

Рассмотрим вопрос о безаварийности вычислений третьего этапа алгоритма. Напомним, что решение системы (2.8) выражается формулами $Z_0 = (A_{m_0} + B_{m_0})^{-1} B_{m_0}$, $Z_\infty = (A_{m_0} + B_{m_0})^{-1} A_{m_0}$. Поэтому встает вопрос об обусловленности матрицы $A_{m_0} + B_{m_0}$. Так как $\|A_{m+1}\| \leq \|A_m\|$, $\|B_{m+1}\| \leq \|B_m\|$, то

$$\|A_{m_0} + B_{m_0}\| \leq \|A_0\| + \|B_0\| \leq 2. \quad (2.24)$$

Если m_0 достаточно велико, то

$$\|(A_{m_0} + B_{m_0})^{-1}\| \leq \sqrt{\omega} + 1. \quad (2.22)$$

В случае вычисления ортогональных проекторов важна обусловленность матрицы $(A_{m_0} B_{m_0})$. Легко показать, что

$$\|(A_{m_0} + B_{m_0})\| \leq \sqrt{2}. \quad (2.23)$$

Можно получить следующие неравенства:

$$\begin{aligned} 1/\sigma_{\min}(A_{m_0} B_{m_0}) &\leq \|T_{m_0}\| \left\| \left[(P_{+0} + P_{-2}^{m_0}) (P_{+0}^* + P_{-2}^{*m_0}) + \right. \right. \\ &\quad \left. \left. + (P_{-0} + P_{2}^{m_0}) (P_{-0}^* + P_{2}^{*m_0}) \right]^{-\frac{1}{2}} \right\| \leq \sqrt{2\omega} + 1, \end{aligned} \quad (2.24)$$

если m_0 достаточно велико.

§ 3. ПОГРЕШНОСТИ ВЕКТОРНЫХ ВЫЧИСЛЕНИЙ НА ЭВМ С ПЛАВАЮЩЕЙ ТОЧКОЙ

Этот параграф в основном носит справочный характер. Некоторые подробности об излагаемых фактах можно найти в книге [1]. Цель изложения в этом параграфе — собрать все необходимые результаты, касающиеся особенностей вещественной арифметики ЭВМ, которые нужны для обоснования и реализации нашей методики исследования спектра. Часть результатов сопровождается подробными доказательствами.

3.1. Погрешности выполнения арифметических операций

Ненулевые вещественные числа, представимые в ЭВМ (в дальнейшем они будут называться машинными числами), имеют вид $z = \pm \gamma^p \cdot m$, где $\pm$ — знак числа z , γ — целое положительное число, называемое *основанием системы счисления*, p — целое число в пределах $p_- \leq p \leq p_+$, на-

зывается *порядком числа* z , m — вещественное число вида $0, d_1 d_2 \dots d_t$ с целыми d_i , причем $0 \leq d_i < \gamma$ и $d_1 \neq 0$. Число m называется (*нормализованной*) *мантиссой машинного числа* z . Вещественный нуль в ЭВМ представляется специальным образом.

По существу, разрядная сетка машинных вещественных чисел характеризуется четырьмя целыми числами: γ , p_- , p_+ и t . Для численного анализа удобнее использовать другие параметры, характеризующие множество машинных вещественных чисел: γ , ε_0 , ε_1 и ε_∞ . Целое число γ уже было определено. Вещественные положительные числа ε_0 , ε_1 и ε_∞ определяют соответственно наименьшее ненулевое положительное машинное число, относительную погрешность округлений и максимальное машинное число:

$$\varepsilon_0 = \gamma^{-1+p_-}, \quad \varepsilon_1 = \gamma^{1-t}, \quad \varepsilon_\infty = \gamma^{p_+}(1 - \gamma^{-t}). \quad (3.1)$$

Константа ε_1 допускает еще следующее описание: между числами 1 и $1 + \varepsilon_1$ нет других машинных чисел.

Машинный результат арифметических операций $+$, $-$, $\times$, $/$ над вещественными машинными числами удовлетворяет следующим свойствам: если точный результат операции по абсолютной величине превышает ε_∞ , то в качестве машинного результата регистрируется ошибочное состояние «переполнения»; если точный результат по абсолютной величине меньше ε_0 , то машинный результат полагается равным нулю; если точный результат по абсолютной величине лежит в замкнутом интервале $[\varepsilon_0, \varepsilon_\infty]$ и является машинным числом, то машинный результат и будет этим точным результатом; и наконец, если точный результат по абсолютной величине лежит в интервале $(\varepsilon_0, \varepsilon_\infty)$, но не является машинным числом, то в качестве машинного результата берется одно из ближайших двух к точному результату машинных чисел.

Можно показать, что для операции $\square \in \{+, -, \times, /\}$, реализованной на ЭВМ описанным выше способом, справедливо тождество

$$(a \square b)_{\text{маш}} = (a \square b)(1 + \varphi) + \psi, \quad (3.2)$$

где $(a \square b)_{\text{маш}}$ — машинный результат операции над машинными вещественными числами a и b , причем

$$|\varphi| \leq \varepsilon_1, \quad |\psi| \leq \varepsilon_0, \quad \varphi\psi = 0. \quad (3.3)$$

Отсюда следуют неравенства

$$|a \square b - (a \square b)_{\text{маш}}| \leq \max\{\varepsilon_0, \varepsilon_1 |a \square b|\} \leq \varepsilon_0 + \varepsilon_1 |a \square b|. \quad (3.4)$$

Будем предполагать, что выполняется также оценка

$$|\sqrt{z} - (\sqrt{z})_{\text{маш}}| \leq \varepsilon_1 \sqrt{z}, \quad z \geq 0. \quad (3.5)$$

Естественно, что оценки погрешностей округления машинных операций справедливы только тогда, когда не возникает ошибочной ситуации «переполнения». В дальнейшем всегда предполагается, что при организации вычислений приняты меры, чтобы не допустить возникновения ситуации «переполнения».

3.2. Погрешности выполнения элементарных матричных операций

В качестве нормы вектора $x \in \mathbb{R}^L$ мы будем использовать евклидову норму $\|x\| = \sqrt{\sum_{i=1}^L x_i^2}$. Из матричных норм нам понадобятся главным образом спектральная

$$\|A\| = \sup_{\|x\|=1} \|Ax\|$$

и фробениусова

$$\|A\|_F = \sqrt{\sum_{i=1}^M \sum_{j=1}^N A_{ij}^2},$$

где A — вещественная $M \times N$ -матрица. Легко показать, что

$$\|A\| \leq \|A\|_F \leq \|A\| \cdot \min(\sqrt{M}, \sqrt{N}).$$

При размещении элементов матрицы вещественных чисел в памяти ЭВМ, среди которых есть не машинные числа, поступают следующим образом: машинные числа представляются точно; числа, меньшие ε_0 по абсолютной величине, представляются машинным нулем; числа, по абсолютной величине лежащие в интервале $(\varepsilon_0, \varepsilon_\infty)$, представляются одним из ближайших машинных чисел. Элементы матрицы, абсолютная величина которых превышает ε_∞ , нельзя представить в ЭВМ. Таким образом, погрешность при размещении $M \times N$ -матрицы A в памяти ЭВМ удовлетворяет оценке

$$\|A - A_{\text{маш}}\|_F \leq \varepsilon_1 \|A\|_F + \varepsilon_0 \sqrt{MN}, \quad (3.6)$$

где $A_{\text{маш}}$ — представление матрицы A в ЭВМ. В дальнейшем все операции проводятся над матрицами и векторами, состоящими из машинных чисел.

Пусть A и B — вещественные $M \times N$ -матрицы, а λ — вещественный скаляр. Тогда

$$\|A \pm B - (A \pm B)_{\text{маш}}\|_F \leq \varepsilon_1 \|A \pm B\|_F + \varepsilon_0 \sqrt{MN}, \quad (3.7)$$

$$\|\lambda A - (\lambda A)_{\text{маш}}\|_F \leq \varepsilon_1 \|\lambda A\|_F + \varepsilon_0 \sqrt{MN}. \quad (3.8)$$

Скалярное произведение векторов $x, y \in \mathbb{R}^L$ будем вычислять при помощи следующего алгоритма:

$$S_0 = 0; \quad S_k = S_{k-1} + x_k y_k, \quad k = 1, 2, \dots, L; \quad (x, y) = S_L.$$

По индукции можно показать справедливость оценки

$$|(x, y) - (x, y)_{\text{маш}}| \leq \frac{L\varepsilon_1}{1 - \frac{L-1}{2}\varepsilon_1} \|x\| \|y\| + \frac{L\varepsilon_0}{1 - \frac{L-1}{2}\varepsilon_1}. \quad (3.9)$$

С целью повышения точности можно вычислять выражение $S_{k-1} + x_k y_k$, используя мантиссы длиной $2t$, т. е. с двойной точностью. Окончательный результат S_L округляется до одинарной точности. Двойная точность приводит к относительной погрешности округлений $\varepsilon_1^{(2)} = \varepsilon_1^2 / \gamma$. Легко показать, что при вычислении скалярного произведения с двойной точностью выполнена оценка

$$|(x, y) - (x, y)_{\text{маш}}| \leq \varepsilon_1 |(x, y)| + \frac{(1 + \varepsilon_1) L \varepsilon_1^{(2)}}{1 - \frac{L-1}{2} \varepsilon_1^{(2)}} \|x\| \|y\| + \frac{(1 + \varepsilon_1) L \varepsilon_0}{1 - \frac{L-1}{2} \varepsilon_1^{(2)}}. \quad (3.10)$$

В итоге, при вычислении скалярного произведения векторов $x, y \in \mathbb{R}^L$ на ЭВМ выполнена оценка погрешностей округлений

$$|(x, y) - (x, y)_{\text{маш}}| \leq \varepsilon_i |(x, y)| + \varepsilon_{ii} \|x\| \|y\| + o_i, \quad (3.11)$$

где в случае одинарной точности

$$\varepsilon_i = 0, \quad \varepsilon_{ii} = \frac{L\varepsilon_1}{1 - \frac{L-1}{2}\varepsilon_1}, \quad o_i = \frac{L\varepsilon_0}{1 - \frac{L-1}{2}\varepsilon_1},$$

а в случае двойной точности

$$\varepsilon_i = \varepsilon_1, \quad \varepsilon_{ii} = \frac{(1 + \varepsilon_1) L \varepsilon_1^{(2)}}{1 - \frac{L-1}{2} \varepsilon_1^{(2)}}, \quad o_i = \frac{(1 + \varepsilon_1) L \varepsilon_0}{1 - \frac{L-1}{2} \varepsilon_1^{(2)}},$$

Пусть A — вещественная $M \times N$ -матрица, а B — $N \times K$ -матрица. Погрешность округлений при вычислении произведения этих матриц удовлетворяет оценке

$$\|AB - (AB)_{\text{маш}}\|_F \leq \varepsilon_i \|AB\|_F + \varepsilon_{ii} \|A\|_F \|B\|_F + o_i \sqrt{MK}, \quad (3.12)$$

где ε_i , ε_{ii} , o_i вычисляются для $L = N$.

Для вычисления на ЭВМ нормы вектора $x \in \mathbb{R}^L$ предлагается следующий алгоритм:

определим $\|x\|_\infty = \max_i |x_i|$;

вычислим скалярное произведение векторов $x/\|x\|_\infty$ и $x/\|x\|_\infty$;

извлекаем квадратный корень из $(x, x)/\|x\|_\infty^2$;

результат предыдущего шага умножим на $\|x\|_\infty$.

Можно показать, что для такого алгоритма выполнена оценка

$$|\|x\| - \|x\|_{\text{маш}}| \leq \varepsilon_n \|x\| \quad (3.13)$$

$$\varepsilon_n = \frac{3\varepsilon_1 + \varepsilon_0 \sqrt{L-1} + (\varepsilon_i + \varepsilon_{ii} + o_i)/2}{1 - 2\varepsilon_1 - (\varepsilon_i + \varepsilon_{ii} + o_i)/2}.$$

3.3. Погрешности выполнения ортогональных преобразований

Для выполнения ортогональных преобразований мы будем использовать преобразования отражения Хаусхолдера

$$Px = x - \frac{2(x, p)}{(p, p)} p,$$

где p — вектор нормали к гиперплоскости отражения.

В матричных вычислениях применяется такое преобразование отражения, которое переводит заданный вектор $u \in \mathbb{R}^L$ в вектор $v = Pu$, имеющий нулевые компоненты с n по L , где $2 \leq n \leq L$. Если n -ю компоненту вектора v определить по формуле

$$v_n = -\text{sign}(u_n) \sqrt{\sum_{j=n}^L u_j^2},$$

где $\text{sign}(z)$ — знак вещественного числа z , то нормаль p длины $\sqrt{2}$ будет удовлетворять соотношениям

$$p_i = 0, \quad i < n; \quad p_n = \frac{u_n - v_n}{\rho}; \quad p_i = \frac{u_i}{\rho}, \quad i > n,$$

а $\rho = \sqrt{v_n(v_n - u_n)}$. На ЭВМ эти формулы можно реализовать следующим образом:

определим $r = \max_{j \geq n} |u_j|$;

вычислим вектор $\tilde{u} = (\tilde{u}_n, \dots, \tilde{u}_L) = \frac{1}{r}(u_n, u_{n+1}, \dots, u_L)$;

вычислим величины $\tilde{v}_n = -\text{sign}(\tilde{u}_n) \|\tilde{u}\|$, $v_n = r\tilde{v}_n$, $\rho = \sqrt{v_n(v_n - \tilde{u}_n)}$;

вычислим компоненты вектора нормали

$$p_n = \frac{\tilde{u}_n - \tilde{v}_n}{\rho}, \quad p_i = \frac{\tilde{u}_i}{\rho} \quad \text{при} \quad i \geq n.$$

Если $r = 0$, то возьмем $v_n = 0$, $p = 0$.

Преобразование отражения P для вектора $x \neq u$ вычисляется по формуле $Px = x - (p, x)p$.

При реализации на ЭВМ предложенного алгоритма получается вектор p , норма которого приближенно равна $\sqrt{2}$ (кроме вырожденного случая $p = 0$). Следовательно, точное преобразование отражения с полученной нормалью p будет приближением к машинному результату вычисления по формуле $Px = x - (p, x)p$.

Обозначим через Px точное преобразование отражения с полученным машинным вектором p , а через $(Px)_{\text{маш}}$ — результат вычислений по формуле $x - (p, x)p$ для $x \neq u$ и $v_{\text{маш}}$ для $x = u$. Тогда выполнена оценка

$$\|Px - (Px)_{\text{маш}}\| \leq \varepsilon_r \|x\| + o_r.$$

Приступим к вычислению коэффициентов ε_r и o_r . Пусть $v_n \neq 0$, тогда ($n \leq k \leq L$)

$$\begin{aligned} |\tilde{u}_{k_{\text{маш}}} - \tilde{u}_k| &\leq \varepsilon_1 |\tilde{u}_k| + \varepsilon_0, \quad |\tilde{u}_k| \leq 1, \quad |\tilde{u}_{k_{\text{маш}}}| \leq 1, \\ |v_{n_{\text{маш}}} - v_n| &\leq \varepsilon_n |v_n|, \quad |\tilde{v}_{n_{\text{маш}}} - \tilde{v}_n| \leq (\varepsilon_n - \varepsilon_1) |\tilde{v}_n|. \end{aligned}$$

Далее

$$\begin{aligned} &|(\tilde{v}_{n_{\text{маш}}} - \tilde{u}_{n_{\text{маш}}})_{\text{маш}} - (\tilde{v}_n - \tilde{u}_n)| \leq \varepsilon_1 |\tilde{v}_{n_{\text{маш}}} - \tilde{u}_{n_{\text{маш}}}| + \\ &+ |\tilde{v}_{n_{\text{маш}}} - \tilde{v}_n| + |\tilde{u}_{n_{\text{маш}}} - \tilde{u}_n| \leq \varepsilon_1 |\tilde{v}_n - \tilde{u}_n| + (1 + \varepsilon_1) |\tilde{v}_{n_{\text{маш}}} - \tilde{v}_n| + \\ &+ (1 + \varepsilon_1) |\tilde{u}_{n_{\text{маш}}} - \tilde{u}_n| \leq \varepsilon_1 |\tilde{v}_n - \tilde{u}_n| + (1 + \varepsilon_1)(\varepsilon_n - \varepsilon_1) |\tilde{v}_n| + \\ &+ (1 + \varepsilon_1)(\varepsilon_1 |\tilde{u}_n| + \varepsilon_0) \leq [\varepsilon_1 + (1 + \varepsilon_1)(\varepsilon_n - \varepsilon_1 + \varepsilon_0)] |\tilde{v}_n - \tilde{u}_n|. \end{aligned}$$

Обозначим $\tilde{v}_{n_{\text{маш}}} = \tilde{v}_n(1 + \xi)$, $(\tilde{v}_{n_{\text{маш}}} - \tilde{u}_{n_{\text{маш}}})_{\text{маш}} = (\tilde{v}_n - \tilde{u}_n)(1 + \eta)$, где $|\xi| \leq \varepsilon_n - \varepsilon_1$, $|\eta| \leq \varepsilon_1 + (1 + \varepsilon_1)(\varepsilon_n - \varepsilon_1 + \varepsilon_0)$. Тогда

$$|p_n - p_{n_{\text{маш}}}| = |p_n| \left| \frac{(1 + \eta)(1 + \beta)}{(1 + \alpha) \sqrt{(1 + \delta)(1 + \xi)(1 + \eta)}} - 1 \right| + \varphi$$

с $|\beta| \leq \varepsilon_1$, $|\alpha| \leq \varepsilon_1$, $|\delta| \leq \varepsilon_1$, $|\varphi| \leq \varepsilon_0$.

Следовательно,

$$|p_n - p_{n_{\text{маш}}}| \leq |p_n| \frac{2\varepsilon_n + 2\varepsilon_1}{1 - 2\varepsilon_n - 2\varepsilon_1} + \varepsilon_0.$$

Подобным образом

$$\begin{aligned} |p_i - p_{i_{\text{маш}}}| &= |p_i| \left| \frac{(1 + \beta)(1 + \theta)}{(1 + \alpha) \sqrt{(1 + \delta)(1 + \xi)(1 + \eta)}} - 1 \right| + \\ &+ \left| \frac{\psi(1 + \beta)}{(1 + \alpha) \sqrt{(1 + \delta)(1 + \xi)(1 + \eta)}} \right| + \varphi, \quad |\theta| \leq \varepsilon_1, \quad |\psi| \leq \varepsilon_0, \\ |p_i - p_{i_{\text{маш}}}| &\leq \frac{2\varepsilon_n + 3\varepsilon_1 + \varepsilon_0}{1 - 2\varepsilon_n - 2\varepsilon_1 - \varepsilon_0} |p_i| + \frac{2\varepsilon_0}{1 - 2\varepsilon_n - 2\varepsilon_1 - \varepsilon_0}. \end{aligned}$$

Подытожим эти оценки:

$$\|p - p_{\text{маш}}\| \leq \|p\| \varepsilon_p = \|p\| \frac{2\varepsilon_n + 3\varepsilon_1 + \varepsilon_0(1 + \sqrt{2L})}{1 - 2\varepsilon_n - 2\varepsilon_1 - \varepsilon_0}.$$

Справедлива цепочка выкладок:

$$\begin{aligned} \|Px - (Px)_{\text{маш}}\| &\leq \|x - (p, x)p\|_{\text{маш}} - \|x - ((p, x)p)_{\text{маш}}\| + \\ &+ \left\| ((p, x)p)_{\text{маш}} - \frac{2(p, x)}{(p, p)} p \right\| \leq \varepsilon_1 \|x - ((p, x)p)_{\text{маш}}\| + \\ &+ \left\| ((p, x)p)_{\text{маш}} - \frac{2(p, x)}{(p, p)} p \right\| + \sqrt{L} \varepsilon_0 \leq \varepsilon_1 \left\| x - \frac{2(p, x)}{(p, p)} p \right\| + \\ &+ \sqrt{L} \varepsilon_0 + (1 + \varepsilon_1) \left\| ((p, x)p)_{\text{маш}} - \frac{2(p, x)}{(p, p)} p \right\| \leq \varepsilon_1 \|x\| + \\ &+ \sqrt{L} \varepsilon_0 + (1 + \varepsilon_1) \left\| ((p, x)p)_{\text{маш}} - \frac{2(p, x)}{(p, p)} p \right\|, \end{aligned}$$

$$\begin{aligned}
& \left\| ((p, x) p)_{\text{маш}} - \frac{2(p, x)}{(p, p)} p \right\| \leq \left\| ((p, x) p)_{\text{маш}} - (p, x)_{\text{маш}} p \right\| + \\
& + \left\| (p, x)_{\text{маш}} p - \frac{2(p, x)}{(p, p)} p \right\| \leq \varepsilon_1 \|p\| |(p, x)_{\text{маш}}| + \sqrt{L} \varepsilon_0 + \\
& + \|p\| |(p, x)_{\text{маш}} - (p, x)| + \left\| (p, x) p - \frac{2(p, x)}{(p, p)} p \right\| \leq \sqrt{L} \varepsilon_0 + \\
& + \varepsilon_1 \|p\| |(p, x)| + (1 + \varepsilon_1) \|p\| |(p, x)_{\text{маш}} - (p, x)| + \left\| (p, x) p - \frac{2(p, x)}{(p, p)} p \right\|, \\
& \left\| (p, x) p - \frac{2(p, x)}{(p, p)} p \right\| = \left\| \frac{2(p, x)}{(p, x)} p \right\| \left| \frac{(p, p)}{2} - 1 \right| = \\
& = \|x - y\| \left| \frac{(p, p)}{2} - 1 \right| \leq \|x\| |(p, p) - 2|, \\
& \|Px - (Px)_{\text{маш}}\| \leq \varepsilon_1 \|x\| + \sqrt{L} \varepsilon_0 + (1 + \varepsilon_1) \{ \sqrt{L} \varepsilon_0 + \|x\| \cdot \|p\|^2 - 2 | + \\
& + (1 + \varepsilon_1) \|p\| o_i + \|p\|^2 \|x\| [\varepsilon_1 + (1 + \varepsilon_1) (\varepsilon_i + \varepsilon_{ii})] \} \leq \sqrt{L} \varepsilon_0 (2 + \\
& + \varepsilon_1) + (1 + \varepsilon_1)^2 \|p\| o_i + \|x\| \{ \varepsilon_1 + (1 + \varepsilon_1) \|p\|^2 - 2 | + (1 + \\
& + \varepsilon_1) \|p\|^2 [\varepsilon_1 + (1 + \varepsilon_1) (\varepsilon_i + \varepsilon_{ii})] \} \leq \sqrt{L} \varepsilon_0 (2 + \varepsilon_1) + \\
& + (1 + \varepsilon_1)^2 \sqrt{2} (1 + \varepsilon_p) o_i + \|x\|^2 \{ \varepsilon_1 + (1 + \varepsilon_1) 2\varepsilon_p (2 + \varepsilon_p) + \\
& + 2(1 + \varepsilon_1) (1 + \varepsilon_p)^2 [\varepsilon_1 + (1 + \varepsilon_1) (\varepsilon_i + \varepsilon_{ii})] \}.
\end{aligned}$$

В итоге

$$\|Px - (Px)_{\text{маш}}\| \leq \varepsilon_r \|x\| + o_r \quad (3.14)$$

с $\varepsilon_r = \frac{3\varepsilon_1 + 4\varepsilon_p + 2\varepsilon_i + 2\varepsilon_{ii}}{1 - 2\varepsilon_p - 2\varepsilon_1}$, $o_r = \frac{2\sqrt{L}\varepsilon_0 + \sqrt{2}o_i}{1 - \varepsilon_p - 2\varepsilon_1}$. Нетрудно проверить, что неравенство (3.14) выполнено в случае, когда $x = u$. Действительно,

$$\begin{aligned}
\|Pu - v_{\text{маш}}\| & \leq \|Pu - P_{\text{точн}}u\| + \|v_{\text{маш}} - v\| \leq \\
& \leq 2\varepsilon_p \|u\| + \varepsilon_n \|v\| = (\varepsilon_n + 2\varepsilon_p) \|u\|.
\end{aligned}$$

Формулы для ε_r и o_r можно несколько упростить:

$$\begin{aligned}
\varepsilon_r & < \frac{39\varepsilon_1 + 20\varepsilon_0 \sqrt{L} + 6(\varepsilon_i + \varepsilon_{ii} + o_i)}{1 - 30\varepsilon_1 - 13\varepsilon_0 \sqrt{L} - 4(\varepsilon_i + \varepsilon_{ii} + o_i)}, \\
o_r & < \frac{\sqrt{2}o_i + 2\varepsilon_0 \sqrt{L}}{1 - 21\varepsilon_1 - 8\varepsilon_0 \sqrt{L} - 3(\varepsilon_i + \varepsilon_{ii} + o_i)}.
\end{aligned}$$

Оценим точность применения преобразования отражения к $M \times N$ -матрице A :

$$\|AP - (AP)_{\text{маш}}\|_F \leq \varepsilon_r \|A\|_F + o_r \sqrt{M}, \quad (3.15)$$

где ε_r и o_r определяются при $L = N$;

$$\|PA - (PA)_{\text{маш}}\|_F \leq \varepsilon_r \|A\|_F + o_r \sqrt{N}, \quad (3.16)$$

где ε_r и o_r определяется при $L = M$.

В заключение получим оценку погрешностей при применении нескольких преобразований отражения к $M \times N$ -матрице A . Тактика применения преобразований отражения для приведения матриц к трехдиагональному и двухдиагональному виду подробно описана в [1].

Пусть $B = PAQ$, $P = P_k P_{k-1} \dots P_1$, $Q = Q_1 Q_2 \dots Q_l$, где P_j и Q_i — преобразования отражения. Предположим сначала, что эти преобразования применяются к матрице A в следующем порядке: вначале все преобразования P_j , а затем все преобразования Q_i .

Обозначим $A_0 = A$, $A_j = (P_j A_{j-1})_{\text{маш}}$. Тогда $A_j = P_j A_{j-1} + H_j$, где $\|H_j\|_F \leq \varepsilon_r \|A_{j-1}\|_F + \sqrt{N} o_r$, причем, ε_r и o_r определяются при $L = M$. Так

как $A_j - P_j \dots P_1 A = P_j (A_{j-1} - P_{j-1} \dots P_1 A) + H_j$, то

$$\begin{aligned} \|A_j - P_j \dots P_1 A\|_F &\leq \|A_{j-1} - P_{j-1} \dots P_1 A\|_F + \varepsilon_r \|A_{j-1}\|_F + \\ &+ \sqrt{N} o_r \leq (1 + \varepsilon_r) \|A_{j-1} - P_{j-1} \dots P_1 A\|_F + \varepsilon_r \|A\|_F + \sqrt{N} o_r, \\ \|A_j - P_j \dots P_1 A\|_F &\leq \sum_{i=0}^{j-1} (1 + \varepsilon_r)^i \{ \varepsilon_r \|A\|_F + \sqrt{N} o_r \} \leq \\ &\leq \frac{j}{1 - \frac{j-1}{2} \varepsilon_r} (\varepsilon_r \|A\|_F + \sqrt{N} o_r). \end{aligned}$$

Аналогично рассматривается применение преобразований отражения Q_i . В итоге получим оценку погрешностей вида

$$\|B_{\text{маш}} - PAQ\|_F \leq \frac{k+l}{1 - \frac{k+l-1}{2} \varepsilon_r} (\varepsilon_r \|A\|_F + o_r \sqrt{\max(M, N)}) \quad (3.17)$$

с ε_r и o_r , вычисленными при $L = \max(M, N)$. Можно показать, что это неравенство выполняется также и во всех других случаях независимо от порядка выполнения преобразований P_j и Q_i .

3.4. Погрешности вычисления обратной матрицы

Вычисление обратной матрицы на ЭВМ будем реализовывать по следующей схеме: сначала приводим матрицу преобразованиями отражения к двухдиагональному виду; затем обращаем полученную двухдиагональную матрицу; в завершение применяем обратные преобразования отражения.

Начнем с вывода оценок погрешностей решений линейных систем с двухдиагональными матрицами. Пусть $\mathcal{D}$ — невырожденная двухдиагональная $N \times N$ -матрица

$$\mathcal{D} = \begin{bmatrix} d_1 & b_2 & & 0 \\ & d_2 & b_3 & \\ & & \ddots & \ddots \\ 0 & & & d_{N-1} & b_N \\ & & & & d_N \end{bmatrix},$$

а $x = (x_1, x_2, \dots, x_N)^T$. Решение системы $\mathcal{D}x = f$ вычисляется по формулам

$$x_N = f_N/d_N, \quad x_{k-1} = (f_{k-1} - b_k x_k)/d_{k-1} \quad (k = N, \dots, 2).$$

Смоделируем ошибки округлений при вычислениях на ЭВМ по этим формулам:

$$\begin{aligned} \tilde{x}_N &= f_N(1 + \varphi_N)/d_N + \xi_N, \\ (b_k \tilde{x}_k)_{\text{маш}} &= b_k(1 + \beta_k) \tilde{x}_k + \psi_{k-1}, \\ (f_{k-1} - (b_k \tilde{x}_k)_{\text{маш}})_{\text{маш}} &= [f_{k-1} - (b_k \tilde{x}_k)_{\text{маш}}] (1 - \varphi_{k-1}) + \chi_{k-1}, \\ \tilde{x}_{k-1} &= (f_{k-1} - (b_k \tilde{x}_k)_{\text{маш}})_{\text{маш}} (1 + \delta_k)/d_{k-1} + \xi_{k-1}, \end{aligned}$$

где $\tilde{x}_k$ — результат вычислений на ЭВМ и

$$|\varphi_k|, |\beta_k|, |\delta_k| \leq \varepsilon_1; |\xi_k|, |\chi_k|, |\psi_k| \leq \varepsilon_0.$$

Обозначим

$$\begin{aligned} \tilde{d}_N &= d_N, \quad \tilde{d}_k = d_k/(1 + \delta_k) \quad (1 \leq k \leq N-1), \\ \tilde{b}_k &= b_k(1 + \beta_k)(1 + \varphi_{k-1}) \quad (2 \leq k \leq N), \\ \tilde{f}_N &= (1 + \varphi_N)f_N + d_N \xi_N, \quad \tilde{f}_k = f_k(1 + \varphi_k) + \chi_k - \psi_k(1 + \varphi_k) + \\ &+ d_k \xi_k/(1 + \delta_k) \quad (1 \leq k \leq N-1). \end{aligned}$$

Легко убедиться в том, что вектор $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N)^T$ удовлетворяет возмущенной линейной системе $\tilde{\mathcal{D}}x = \tilde{f}$ с матрицей

$$\tilde{\mathcal{D}} = \begin{bmatrix} \tilde{d}_1 & \tilde{b}_2 & & 0 \\ & \tilde{d}_2 & \cdot & \\ & & \ddots & \\ 0 & & & \tilde{b}_N \\ & & & & \tilde{d}_N \end{bmatrix}.$$

Выполнены неравенства

$$\max_k \left(\left| \frac{\tilde{d}_k - d_k}{d_k} \right|, \left| \frac{\tilde{b}_k - b_k}{b_k} \right| \right) \leq \frac{2\varepsilon_1}{1 - \varepsilon_1}, \quad (3.18)$$

$$|\tilde{f}_k - f_k| \leq \varepsilon_1 |f_k| + \varepsilon_0 [2 + \varepsilon_1 + \|\mathcal{D}\|/(1 - \varepsilon_1)]. \quad (3.19)$$

Теперь приступим к полной схеме алгоритма обращения матрицы A . На первом шаге получаем такую двухдиагональную матрицу $\mathcal{D}$, что выполняется равенство $\mathcal{D} = P(A + \Phi_1)Q$ с оценкой

$$\|\Phi_1\| = \|\mathcal{D} - PAQ\| \leq \frac{(2N-3)\sqrt{N}}{1 - (N-2)\varepsilon_r} (\varepsilon_r \|A\| + o_r),$$

где ε_r и o_r определяются при $L = N$. Напомним, что P и Q — ортогональные матрицы. На втором шаге решаем N систем линейных уравнений $\mathcal{D}x_j = e_j$, $j = 1, 2, \dots, N$, с правыми частями вида $e_j = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix}^T$. Приближенные решения обозначим через $\tilde{x}_j$. Если $\tilde{X}$ — матрица, составленная из столбцов $\tilde{x}_j$, $j = 1, 2, \dots, N$, то $\tilde{X} = (\mathcal{D} + \Phi_2)^{-1}(I + \Psi_2)$, а матрицы возмущений Φ_2 и Ψ_2 удовлетворяют оценкам

$$\|\Phi_2\| \leq \frac{2\sqrt{2}\varepsilon_1}{1 - \varepsilon_1} \|\mathcal{D}\|, \quad \|\Psi_2\| \leq \varepsilon_1 + \varepsilon_0 (2 + \varepsilon_1 + \|\mathcal{D}\|/(1 - \varepsilon_1)).$$

На третьем шаге вычисляется матрица $Q\tilde{X}P$. Окончательный результат алгоритма обозначим через B . Тогда справедлива оценка

$$\|B - Q\tilde{X}P\| \leq \frac{(2N-3)\sqrt{N}}{1 - (N-2)\varepsilon_r} (\varepsilon_r \|\tilde{X}\| + o_r).$$

Просуммируем погрешности всего алгоритма:

$$Q\tilde{X}P = Q(\mathcal{D} + \Phi_2)^{-1}(I + \Psi_2)P = (A + \Phi_1 + P^{-1}\Phi_2Q^{-1})^{-1}(I + P^{-1}\Psi_2P) = (A + \Phi_3)^{-1}(I + \Psi_3),$$

где $\Phi_3 = \Phi_1 + P^{-1}\Phi_2Q^{-1}$, $\Psi_3 = P^{-1}\Psi_2P$. Выполнены оценки

$$\|\Phi_3\| \leq \|\Phi_1\| + \frac{2\sqrt{2}\varepsilon_1}{1 - \varepsilon_1} (\|A\| + \|\Phi_1\|),$$

$$\|\Psi_3\| \leq \varepsilon_1 + \frac{\varepsilon_0}{1 - \varepsilon_1} (2 + \|A\| + \|\Phi_1\|),$$

$$\|\tilde{X}\| \leq \frac{1 + \|\Psi_3\|}{\sigma_{\min}(A) - \|\Phi_3\|}.$$

Далее,

$$B - A^{-1} = (B - Q\tilde{X}P) + (A + \Phi_3)^{-1}(I + \Psi_3) - A^{-1} = (B - Q\tilde{X}P) + A^{-1}\Psi_3 - (A + \Phi_3)^{-1}\Phi_3A^{-1}(I + \Psi_3),$$

$$\|B - A^{-1}\| \leq \frac{(2N-3)\sqrt{N}}{1 - (N-2)\varepsilon_r} (\varepsilon_r \|\tilde{X}\| + o_r) +$$

$$+ \frac{\|\Psi_3\|}{\sigma_{\min}(A)} + \frac{\|\Phi_3\|(1 + \|\Psi_3\|)}{\sigma_{\min}(A)(\sigma_{\min}(A) - \|\Phi_3\|)}.$$

Следовательно,

$$\|A^{-1} - (A^{-1})_{\text{маш}}\| \leq \frac{(2N-3) \sqrt{N} \varepsilon_r}{1 - (N-2) \varepsilon_r} + \\ + \frac{1}{\sigma_{\min}(A) - \|\Phi_3\|} \left[\|\Psi_3\| + \frac{\|\Phi_3\| (1 + \|\Psi_3\|)}{\sigma_{\min}(A)} + \frac{(2N-3) \sqrt{N} \varepsilon_r (1 + \|\Psi_3\|)}{1 - (N-2) \varepsilon_r} \right], \quad (3.20)$$

где

$$\|\Phi_3\| \leq \frac{2\sqrt{2} \varepsilon_1 + (2N-3) \sqrt{N} \varepsilon_r}{1 - 2\sqrt{2} \varepsilon_1 - (N-2) \varepsilon_r} \|A\| + \frac{(2N-3) \sqrt{N} \varepsilon_r}{1 - 2\sqrt{2} \varepsilon_1 - (N-2) \varepsilon_r}, \\ \|\Psi_3\| \leq \varepsilon_1 + \frac{\varepsilon_0}{1 - \varepsilon_1 - (N-2) \varepsilon_r} [(1 + (2N-3) \sqrt{N} \varepsilon_r) \|A\| + \\ + (2 + (2N-3) \sqrt{N} \varepsilon_r)].$$

Полученная оценка носит априорный характер. Часто более точной бывает апостериорная оценка по невязке $A_{\text{маш}}^{-1} A - I$. В этом случае выполнено неравенство

$$\|A_{\text{маш}}^{-1} - A^{-1}\| \leq \|A^{-1}\| \|A_{\text{маш}}^{-1} A - I\|. \quad (3.21)$$

3.5. Погрешности вычисления последовательностей Штурма симметричной трехдиагональной матрицы

Обозначим через

$$T = \begin{bmatrix} d_1 & b_2 & & & \\ b_2 & d_2 & & & \\ & & \ddots & \ddots & \\ & & & \ddots & b_N \\ & & & b_N & d_N \end{bmatrix}$$

симметричную трехдиагональную матрицу. Последовательностью Штурма назовем рекуррентную последовательность функций

$$q_1(\lambda) = d_1 - \lambda, \quad q_i(\lambda) = d_i - \lambda - [b_i/q_{i-1}(\lambda)] b_i \quad (i = 2, \dots, N). \quad (3.22)$$

Теорема 3.1 (Штурм). Пусть $q_{i-1}(\lambda) \neq 0$ ($i = 2, \dots, N$) при некотором вещественном λ . Тогда количество собственных чисел матрицы T , меньших λ , равно количеству отрицательных чисел среди $q_i(\lambda)$ ($i = 1, 2, \dots, N$).

Непосредственный счет по формулам (3.22) может привести, с одной стороны, к «переполнению», с другой — к излишнему накоплению погрешностей. Для преодоления этих недостатков необходимо провести предварительную нормировку T .

Предлагается следующий вычислительный процесс:

находим $\rho = \max_{i,j} (|d_i|, |b_j|)$ и вычисляем нормированную матрицу T/ρ ;

вычисляем последовательность величин $q_i(\lambda)$, $i = 1, 2, \dots, N$, для матрицы T/ρ (если какое либо $q_i(\lambda)$ по абсолютной величине меньше

$$\varepsilon_s = \max(\varepsilon_0, 2/\varepsilon_\infty), \quad (3.23)$$

то заменим такое $q_i(\lambda)$ на ε_s).

Проведем анализ вычислительных погрешностей предложенной реализации построения последовательности Штурма на ЭВМ. При нормировке матрицы, очевидно, будет допущена ошибка

$$\|(T/\rho)_{\text{маш}} - T/\rho\|_\infty \leq \varepsilon_1 \|(T/\rho)_{\text{маш}}\|_\infty + 3\varepsilon_0,$$

где норма $\|A\|_\infty$ $M \times N$ -матрицы A определяется по формуле

$$\|A\|_\infty = \max_{1 \leq i \leq M} \left(\sum_{j=1}^N |a_{ij}| \right).$$

Для симметричной трехдиагональной матрицы T справедливы неравенства

$$\|T\| \leq \|T\|_\infty \leq \sqrt{3}\|T\|.$$

Обозначим через $\tilde{d}_i, \tilde{b}_i$ элементы матрицы $(T/\rho)_{\text{маш}}$. Смоделируем погрешности округлений при вычислении последовательности Штурма $q_i(\lambda)$:

$$\begin{aligned} (\tilde{d}_i - \lambda)_{\text{маш}} &= (\tilde{d}_i - \lambda)(1 + \alpha_1) + \bar{\alpha}_1, \quad |\alpha_1| \leq \varepsilon_1, \quad |\bar{\alpha}_1| \leq \varepsilon_0, \\ [\tilde{b}_i/q_{i-1}(\lambda)]_{\text{маш}} &= [\tilde{b}_i/q_{i-1}(\lambda)](1 + \alpha_2) + \bar{\alpha}_2, \quad |\alpha_2| \leq \varepsilon_1, \quad |\bar{\alpha}_2| \leq \varepsilon_0, \\ ([\tilde{b}_i/q_{i-1}(\lambda)]_{\text{маш}} \tilde{b}_i)_{\text{маш}} &= [\tilde{b}_i/q_{i-1}(\lambda)]_{\text{маш}} \tilde{b}_i(1 + \alpha_3) + \\ &\quad + \bar{\alpha}_3, \quad |\alpha_3| \leq \varepsilon_1, \quad |\bar{\alpha}_3| \leq \varepsilon_0, \\ ((\tilde{d}_i - \lambda)_{\text{маш}} - ([\tilde{b}_i/q_{i-1}(\lambda)]_{\text{маш}} \tilde{b}_i)_{\text{маш}})_{\text{маш}} &= \\ &= \bar{\alpha}_4 + \{\bar{\alpha}_1 + (\tilde{d}_i - \lambda)(1 + \alpha_1) - \\ &\quad - ([\tilde{b}_i/q_{i-1}(\lambda)](1 + \alpha_2) + \bar{\alpha}_2) \tilde{b}_i(1 + \alpha_3) - \bar{\alpha}_3\}(1 + \alpha_4), \\ &\quad |\alpha_4| \leq \varepsilon_1, \quad |\bar{\alpha}_4| \leq \varepsilon_0 + \varepsilon_S. \end{aligned}$$

Добавка ε_S в $\bar{\alpha}_4$ обусловлена возможной заменой на втором этапе вычислительного процесса величины $q_i(\lambda)$ на ε_S . Таким образом,

$$\begin{aligned} q_i(\lambda)_{\text{маш}} &= \tilde{d}_i(1 + \alpha_1)(1 + \alpha_4) - \lambda - \frac{\tilde{b}_i^2}{q_{i-1}(\lambda)}(1 + \alpha_2)(1 + \alpha_3)(1 + \alpha_4) + \\ &\quad + \{\bar{\alpha}_4 + (\bar{\alpha}_1 - \bar{\alpha}_3)(1 + \alpha_4) - \bar{\alpha}_2 \tilde{b}_i(1 + \alpha_3) - \lambda[(1 + \alpha_1)(1 + \alpha_4) - 1]\}. \end{aligned}$$

Это дает основание утверждать, что вычисленная последовательность Штурма $q_i(\lambda)_{\text{маш}}$ является последовательностью Штурма для матрицы, близкой к $(T/\rho)_{\text{маш}}$. Обозначим эту матрицу через $\hat{T}/\rho$ и оценим норму разности $\Delta = \hat{T}/\rho - (T/\rho)_{\text{маш}}$:

$$\begin{aligned} \|\Delta\|_\infty &\leq \max(|(1 + \alpha_1)(1 + \alpha_4) - 1|, |\sqrt{(1 + \alpha_2)(1 + \alpha_3)(1 + \alpha_4)} - \\ &\quad - 1|) \|(T/\rho)_{\text{маш}}\|_\infty + \varepsilon_0 + \varepsilon_S + 2\varepsilon_0(1 + \varepsilon_1) + \varepsilon_0(1 + \varepsilon_1) + \\ &\quad + |\lambda|[(1 + \varepsilon_1)^2 - 1] \leq \frac{2\varepsilon_1}{1 - \varepsilon_1} \|(T/\rho)_{\text{маш}}\|_\infty + \frac{\varepsilon_S + 4\varepsilon_0}{1 - \varepsilon_1} + \frac{2\varepsilon_1}{1 - \varepsilon_1} |\lambda|. \end{aligned}$$

Следовательно,

$$\begin{aligned} \|\hat{T}/\rho - T/\rho\|_\infty &\leq \|\Delta\|_\infty + \varepsilon_1 \|(T/\rho)_{\text{маш}}\|_\infty + 3\varepsilon_0 \leq \frac{3\varepsilon_1}{1 - \varepsilon_1} \|(T/\rho)_{\text{маш}}\|_\infty + \\ &\quad + \frac{2\varepsilon_1}{1 - \varepsilon_1} |\lambda| + \frac{\varepsilon_S + 7\varepsilon_0}{1 - \varepsilon_1}. \end{aligned} \quad (3.24)$$

Рассмотрим вопрос о безаварийности вычислений последовательности Штурма. При точных вычислениях $|q_i(\lambda)| \leq 1 + |\lambda| + \varepsilon_\infty/2$, поэтому для $|\lambda| \leq 3$ величина $|q_i(\lambda)|$ намного меньше ε_∞ , так как $\varepsilon_\infty \gg 8$. При вычислении на ЭВМ величина $|q_i(\lambda)_{\text{маш}}|$ мало отличается от $|q_i(\lambda)|$.

Используя теорему Штурма, можно вычислять собственные числа методом бисекций. Оценка (3.24) гарантирует величины погрешностей.

Метод бисекций годится и для вычисления сингулярных чисел двухдиагональной матрицы

$$\mathcal{D} = \begin{bmatrix} -d_1 & b_2 & & 0 \\ & d_2 & \cdot & \\ & & \ddots & \\ 0 & & & b_N \\ & & & & d_N \end{bmatrix},$$

так как они совпадают с N наибольшими собственными числами матрицы

$$T = \begin{bmatrix} -0 & d_1 & & 0 \\ d_1 & 0 & b_2 & \\ & b_2 & 0 & \cdot \\ & & \cdot & \ddots & \\ & & & \cdot & d_N \\ 0 & & & & d_N & 0 \end{bmatrix}.$$

В этом случае оценка (3.24) может быть несколько улучшена:

$$\|\widehat{T}/\rho - T/\rho\|_\infty \leq \frac{3\varepsilon_1}{1-\varepsilon_1} \|(T/\rho)_{\text{маш}}\|_\infty + \frac{\varepsilon_1}{1-\varepsilon_1} |\lambda| + \frac{\varepsilon_S + 5\varepsilon_0}{1-\varepsilon_1}. \quad (3.25)$$

3.6. Разложение Холесского положительно определенной симметричной матрицы

Пусть S — симметричная положительно определенная $N \times N$ -матрица. Разложением Холесского матрицы S называется представление $S = LL^*$ с нижней треугольной матрицей L . Удобно не ограничивать себя условием треугольности матрицы L . Поэтому поставим задачу вычисления матрицы R общего вида, удовлетворяющей тождеству

$$S = RR^*. \quad (3.26)$$

Предлагается следующий алгоритм. Сначала приведем симметричную матрицу S к трехдиагональному виду $T = QSQ^*$, $QQ^* = I$. У полученной трехдиагональной симметричной матрицы T найдем разложение Холесского $T = LL^*$. В конце вычислим $R = Q^{-1}L$.

Рассмотрим подробно процесс вычисления разложения Холесского $T = LL^*$. Пусть

$$T = \begin{bmatrix} -d_1 & b_2 & & 0 \\ b_2 & d_2 & \cdot & \\ & \cdot & \ddots & \\ 0 & & & b_N \\ & & & & d_N \end{bmatrix}, \quad L = \begin{bmatrix} -l_1 & & & 0 \\ c_2 & l_2 & & \\ & \cdot & \ddots & \\ 0 & & & c_N & l_N \end{bmatrix}, \quad l_j > 0.$$

Нетрудно убедиться в том, что $l_i = \sqrt{q_i(0)}$, где $q_i(\lambda)$ — последовательность Штурма для матрицы T , а $c_h = b_h/l_{h-1}$.

Проведем анализ ошибок округления при вычислении матрицы L . Как было показано в 3.5, вычисленная последовательность Штурма $q_i(0)_{\text{маш}}$ соответствует возмущенной матрице $\widetilde{T}$ с оценкой

$$\|\widetilde{T} - T\|_\infty \leq \frac{2\varepsilon_1}{1-\varepsilon_1} \|T\|_\infty + \frac{\varepsilon_S + 3\varepsilon_0}{1-\varepsilon_1}. \quad (3.27)$$

Теперь нужно учесть ошибки при извлечении квадратного корня $\sqrt{q_i(0)_{\text{маш}}}$ и при делении b_h/l_{h-1} . Смоделируем погрешности извлечения

квадратного корня: $\tilde{l}_j = \tilde{l}_j(1 + \alpha)$, $|\alpha| \leq \varepsilon_1$, где $\tilde{L}\tilde{L}^* = \tilde{T}$,

$$\tilde{L} = \begin{bmatrix} \tilde{l}_1 & & & 0 \\ \tilde{c}_2 & \tilde{l}_2 & & \\ & \cdot & \cdot & \\ 0 & \cdot & \tilde{c}_N & \tilde{l}_N \end{bmatrix}.$$

Погрешность деления смоделируем следующим образом:

$$\hat{c}_k = \frac{\tilde{b}_k(1 + \beta)}{\tilde{l}_{k-1}(1 + \alpha)} + \bar{\beta}, \quad |\alpha| \leq \varepsilon_1, \quad |\beta| \leq \varepsilon_1, \quad |\bar{\beta}| \leq \varepsilon_0.$$

Оценим разность $\hat{L} - \tilde{L}$, где $\hat{L}$ — вычисленная матрица разложения Хо-лесского:

$$\|\hat{L} - \tilde{L}\|_\infty \leq \frac{2\varepsilon_1}{1 - \varepsilon_1} \|\tilde{L}\|_\infty + \varepsilon_0. \quad (3.28)$$

Обозначим $\hat{T} = \hat{L}\hat{L}^*$, тогда при $\varepsilon_1 < 0.05$

$$\begin{aligned} \|\hat{T} - \tilde{T}\| &\leq \|(\hat{L} - \tilde{L})(\hat{L}^* - \tilde{L}^*)\| + \|\tilde{L}(\hat{L}^* - \tilde{L}^*)\| + \\ &+ \|(\hat{L} - \tilde{L})\tilde{L}^*\| \leq \|\hat{L} - \tilde{L}\|^2 + 2\|\tilde{L}\| \|\hat{L} - \tilde{L}\| \leq \\ &\leq (3\varepsilon_1 \|\tilde{L}\| + \varepsilon_0)^2 + 2\|\tilde{L}\| (3\varepsilon_1 \|\tilde{L}\| + \varepsilon_0) = \\ &= (6\varepsilon_1 + 9\varepsilon_1^2) \|\tilde{L}\|^2 + (2\varepsilon_0 + 6\varepsilon_1\varepsilon_0) \|\tilde{L}\| + \varepsilon_0^2. \end{aligned}$$

Так как $2\varepsilon_0 \|\tilde{L}\| \leq \varepsilon_0 \|\tilde{L}\|^2 + \varepsilon_0$, то

$$\begin{aligned} \|\hat{T} - \tilde{T}\| &\leq \frac{6\varepsilon_1 + \varepsilon_0}{1 - 3\varepsilon_1} \|\tilde{T}\| + \frac{\varepsilon_0 + \varepsilon_0^2}{1 - 3\varepsilon_1}, \\ \|\hat{T} - T\| &\leq \|\hat{T} - \tilde{T}\| + \|\tilde{T} - T\| \leq \frac{9\varepsilon_1 + \varepsilon_0}{1 - 6\varepsilon_1} \|T\| + \frac{\varepsilon_S + 4\varepsilon_0}{1 - 7\varepsilon_1 - \varepsilon_0}. \end{aligned} \quad (3.29)$$

Просуммируем погрешности полного алгоритма. Поскольку $T = Q(S + \Phi)Q^*$ и

$$\|\Phi\| \leq \frac{(2N - 4) \sqrt{N}}{1 - (N - 2) \varepsilon_r} (\varepsilon_r \|S\| + o_r), \quad (3.30)$$

то матрица $\hat{R} = Q^{-1}\hat{L}$ удовлетворяет тождеству $\hat{R}\hat{R}^* = S + \Phi + \Psi$ и

$$\begin{aligned} \|\Psi\| &\leq \frac{9\varepsilon_1 + \varepsilon_0}{1 - 6\varepsilon_1} (\|S\| + \|\Phi\|) + \frac{\varepsilon_S + 4\varepsilon_0}{1 - 7\varepsilon_1 - \varepsilon_0}, \\ \|\Phi + \Psi\| &\leq \left[\frac{9\varepsilon_1 + \varepsilon_0}{1 - 6\varepsilon_1} + \frac{(2N - 4) \sqrt{N} \varepsilon_r}{1 - (N - 2) \varepsilon_r - 3\varepsilon_1 - \varepsilon_0} \right] \|S\| + \\ &+ \left[\frac{\varepsilon_S + 4\varepsilon_0}{1 - 7\varepsilon_1 - \varepsilon_0} + \frac{(2N - 4) \sqrt{N} o_r}{1 - (N - 2) \varepsilon_r - 3\varepsilon_1 - \varepsilon_0} \right]. \end{aligned} \quad (3.31)$$

Таким образом, вычисленная матрица $\hat{L}$ такова, что $Q^{-1}\hat{L}\hat{L}^*Q$ является возмущением матрицы S . Определим матрицу R , близкую к $\hat{R}$, удовлетворяющую тождеству $RR^* = S$:

$$\begin{aligned} \hat{R}^{-1}S\hat{R}^{-*} &= I - \hat{R}^{-1}(\Phi + \Psi)\hat{R}^{-*}, \\ I - \hat{R}^{-1}(\Phi + \Psi)\hat{R}^{-*} &= (I + \Theta)(I + \Theta)^*, \\ R &= \hat{R}(I + \Theta). \end{aligned}$$

Выполнены оценки вида

$$\|\Theta\| \leq \frac{\|\Phi + \Psi\| \|(S + \Phi + \Psi)^{-1}\|}{2 - \|\Phi + \Psi\| \|(S + \Phi + \Psi)^{-1}\|}, \quad (3.32)$$

$$\|\widehat{R} - R\| \leq \|R\| \|(I + \Theta)^{-1} - I\| \leq \|R\| \frac{\|\Phi + \Psi\| \|(S + \Phi + \Psi)^{-1}\|}{2(1 - \|\Phi + \Psi\| \|(S + \Phi + \Psi)^{-1}\|)}.$$

Погрешность вычисления матрицы $R_{\text{маш}} = (Q^{-1}\widehat{L})_{\text{маш}}$ равна

$$\|R_{\text{маш}} - \widehat{R}\| \leq \frac{(N-2)\sqrt{N}}{1 - \frac{N-3}{2}\varepsilon_r} (\varepsilon_r \|\widehat{R}\|_{0_r}). \quad (3.33)$$

Следовательно,

$$\begin{aligned} \|R_{\text{маш}} - R\| &\leq \|R_{\text{маш}} - \widehat{R}\| + \|\widehat{R} - R\| \leq \frac{(N-2)\sqrt{N}o_r}{1 - \frac{N-3}{2}\varepsilon_r} + \\ &+ \|R\| \left[\frac{\|\Phi + \Psi\| \|(S + \Phi + \Psi)^{-1}\|}{2(1 - \|\Phi + \Psi\| \|(S + \Phi + \Psi)^{-1}\|)} + \right. \\ &\left. + \frac{(N-2)\sqrt{N}\varepsilon_r}{1 - \frac{N-3}{2}\varepsilon_r - \|\Phi + \Psi\| \|(S + \Phi + \Psi)^{-1}\|} \right]. \quad (3.34) \end{aligned}$$

В заключение отметим, что анализ алгоритма разложения Холецкого был проведен без проверки безаварийности. Чтобы алгоритм был безаварийным и не было излишнего накопления погрешностей, нужно вначале нормировать матрицу S условием $\|S\| \sim 1$. Также необходимо, чтобы обусловленность матрицы S была не слишком большой. В тех случаях, когда мы будем использовать разложение Холецкого, эти требования выполнены. У нас будет $\|S\| < 2$, $\|S^{-1}\| < 2$.

§ 4. УСТОЙЧИВОСТЬ АЛГОРИТМА К ПОГРЕШНОСТЯМ ВЫЧИСЛЕНИЙ

В этом параграфе проводится полный анализ влияния ошибок округления при реализации алгоритма вычисления проекторов регулярного пучка матриц на ЭВМ с плавающей точкой. При этом для оценок погрешностей приводятся точные выражения, а не грубые прикидки порядков величин.

Конечно, наш анализ учитывает на всех этапах самые неблагоприятные ситуации и получаемые оценки будут существенно завышены по сравнению с реальными. Тем не менее, эти оценки достаточно эффективны, если задача о вычислении проекторов P_0 и P_∞ пучка $\lambda B - A$ хорошо поставлена. Под хорошей постановкой задачи понимается прежде всего устойчивость к возмущениям входных данных. У нас мерой качества устойчивости спектральных характеристик пучка $\lambda B - A$ являются величины параметра ω и обусловленности матрицы (AB) .

Содержание § 4 можно рассматривать как доказательство корректности алгоритма при реализации его на ЭВМ. Корректным алгоритмом следует называть тот алгоритм, который хорошо справляется с хорошо поставленными задачами.

Очень важно подчеркнуть, что система оценок, полученных в этом параграфе, все-таки не дает гарантии точности вычисленных проекторов P_0 и P_∞ и матрицы H . Вопрос о гарантиях достоверности результатов счета будет обсуждаться в § 5.

Излагаемые ниже результаты показывают, что параметр ω является мерой качества устойчивости предложенного алгоритма. Полученные оценки могут быть использованы для определения множества хорошо поставленных задач для конкретной ЭВМ.

4.1. Некоторые леммы

Для проведения обратного анализа ошибок округления и погрешностей входных данных полезно выделить три технических леммы.

Лемма 4.1. Пусть R — невырожденная $N \times N$ -матрица, а ξ — матрица размеров $n \times N$. Тогда существует такая $(Nn) \times (Nn)$ -матрица S , что

$$(I + S) \begin{pmatrix} R \\ \xi \end{pmatrix} = \begin{pmatrix} \tilde{R} \\ 0 \end{pmatrix}, \quad (4.1)$$

$$(I + S^*) (I + S) = I, \quad (4.2)$$

причем выполнены оценки

$$\|S\| \leq \|\xi R^{-1}\| \leq \|\xi\| \|R^{-1}\|. \quad (4.3)$$

Доказательство. Подходящая матрица S конструируется явно. Рассмотрим матрицу

$$\tilde{S} = \begin{pmatrix} 0 & (\xi R^{-1})^* \\ -\xi R^{-1} & 0 \end{pmatrix}.$$

Легко показать, что имеют место матричные тождества

$$(I + \tilde{S}) \begin{pmatrix} R \\ \xi \end{pmatrix} = \begin{pmatrix} R + (R^{-1})^* \xi^* \xi \\ 0 \end{pmatrix},$$

$$(I + \tilde{S})(I + \tilde{S}^*) = \begin{pmatrix} I + (\xi R^{-1})^* \xi R^{-1} & 0 \\ 0 & I + \xi R^{-1} (\xi R^{-1})^* \end{pmatrix}.$$

Матрица

$$S = \begin{pmatrix} I + (\xi R^{-1})^* \xi R^{-1} & 0 \\ 0 & I + \xi R^{-1} (\xi R^{-1})^* \end{pmatrix} (I + \tilde{S}) - I$$

удовлетворяет условиям леммы.

Осталось доказать оценку $\|S\|$. Пусть $U^* \mathcal{D} V$ — сингулярное разложение матрицы ξR^{-1} . Рассмотрим, например, случай, когда $\mathcal{D} = (\Sigma \ 0)$, где Σ — диагональная матрица с положительными элементами по диагонали. Несложные выкладки дают следующее представление матрицы S :

$$S = \begin{pmatrix} V^* & 0 \\ 0 & U^* \end{pmatrix} \begin{bmatrix} (I + \Sigma^2)^{-\frac{1}{2}} - I & 0 & (I + \Sigma^2)^{-\frac{1}{2}} \Sigma \\ 0 & 0 & 0 \\ -(I + \Sigma^2)^{-\frac{1}{2}} \Sigma & 0 & (I + \Sigma^2)^{-\frac{1}{2}} - I \end{bmatrix} \begin{pmatrix} V & 0 \\ 0 & U \end{pmatrix}.$$

Отсюда следует нужная нам оценка:

$$\|S\| \leq \sqrt{2 \left(1 - \frac{1}{\sqrt{1 + \|\xi R^{-1}\|^2}} \right)} \leq \|\xi R^{-1}\|.$$

Аналогично рассматривается случай, когда $\mathcal{D} = (\Sigma \ 0)^T$. Лемма доказана.

Следующая лемма — о единственности блочного QR -разложения.

Лемма 4.2. Пусть

$$A = Q_1 \begin{pmatrix} K_1 & L_1 \\ 0 & M_1 \end{pmatrix} = Q_2 \begin{pmatrix} K_2 & L_2 \\ 0 & M_2 \end{pmatrix},$$

где Q_1 и Q_2 — ортогональные матрицы, K_1 и K_2 — невырожденные квадратные матрицы одинаковых размеров, M_1 и M_2 — матрицы одинаковых размеров и полного ранга. Тогда существуют такие ортогональные матрицы $\mathcal{P}$ и Q , что $K_2 = \mathcal{P} K_1$, $L_2 = \mathcal{P} L_1$, $M_2 = Q M_1$.

Доказательство. Воспользуемся тождеством

$$A^*A = \begin{pmatrix} K_1^*K_1 & K_1^*L_1 \\ L_1^*K_1 & L_1^*L_1 + M_1^*M_1 \end{pmatrix} = \begin{pmatrix} K_2^*K_2 & K_2^*L_2 \\ L_2^*K_2 & L_2^*L_2 + M_2^*M_2 \end{pmatrix}.$$

В силу невырожденности матриц K_1 и K_2 из равенства $K_1^*K_1 = K_2^*K_2$ выводится представление $K_2 = \mathcal{P}K_1$ с ортогональной матрицей $\mathcal{P}$. Действительно, $(K_1K_2^{-1})^*(K_1K_2^{-1}) = I$, $K_2 = (K_1K_2^{-1})^*K_1$, т. е. $\mathcal{P} = (K_1K_2^{-1})^*$. Из равенства $K_1^*L_1 = K_2^*L_2$ следует представление $L_2 = \mathcal{P}L_1$. Отсюда вытекает тождество $M_1^*M_1 = M_2^*M_2$, связывающее матрицы M_1 и M_2 полного ранга.

Вначале предположим, что число строк в матрицах M_1 и M_2 не меньше числа столбцов. В этом случае, воспользовавшись QR разложением матриц M_1 и M_2 , получим тождество $\widetilde{M}_1^*\widetilde{M}_1 = \widetilde{M}_2^*\widetilde{M}_2$ для треугольных невырожденных матриц $\widetilde{M}_1$ и $\widetilde{M}_2$, из которого, аналогично случаю матриц K_1 и K_2 , следует утверждение леммы 4.2.

Теперь рассмотрим ситуацию, когда число строк в матрицах M_1 и M_2 меньше числа столбцов. Сингулярное разложение этих матриц имеет вид $M_1 = V_1^*\mathcal{D}_1U_1$, $M_2 = V_2^*\mathcal{D}_2U_2$, где U_1 , U_2 , V_1 и V_2 — ортогональные матрицы, а $\mathcal{D}_1 = \mathcal{D}_2 = (\Sigma \ 0)$ с квадратной невырожденной диагональной матрицей Σ . В этих обозначениях выполнено равенство

$$U_2U_1^{-1} \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} U_2U_1^{-1},$$

откуда имеем следующую структуру матрицы $U_2U_1^{-1}$:

$$U_2U_1^{-1} = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix},$$

причем $W_1\Sigma^2 = \Sigma^2W_1$. Следовательно,

$$\begin{aligned} M_2 &= V_2^*(\Sigma W_1, 0)U_1 = V_2^*(\Sigma^{-1}W_1\Sigma)(\Sigma \ 0)U_1 = \\ &= V_2^*(\Sigma^{-1}W_1\Sigma)V_1 \cdot V_1^*(\Sigma \ 0)U_1. \end{aligned}$$

Осталось показать ортогональность матриц $\Sigma^{-1}W_1\Sigma$. Действительно,

$$(\Sigma^{-1}W\Sigma)(\Sigma W_1^*\Sigma^{-1}) = \Sigma^{-1}W_1\Sigma^2W_1^*\Sigma^{-1} = \Sigma^{-1}\Sigma^2W_1W_1^*\Sigma^{-1} = I.$$

Лемма доказана.

Лемма 4.3. Предположим, что пучок $N \times N$ -матриц $\lambda B - A$ регулярен относительно единичной окружности, т. е. $\det(B - e^{i\varphi}A) \neq 0$ при любом вещественном φ . Пусть

$$r = \begin{bmatrix} B & & & 0 \\ -A & B & & \\ & \cdot & \cdot & \\ & & \cdot & B \\ 0 & -A & & -A \end{bmatrix}$$

— блочная двухдиагональная матрица размеров $[N(n+1)] \times [Nn]$, где $n \geq 1$ — любое целое число. Тогда для любого вектора $x \in \mathbb{R}^{nN}$ справедливо неравенство

$$\|rx\| \geq \frac{\|x\|}{\max_{\varphi} \|(B - e^{i\varphi}A)^{-1}\|}. \quad (4.4)$$

Доказательство. Применим способ сведения к бесконечномерному случаю. Для этого рассмотрим множество последовательностей век-

торов из $\mathbb{R}^N$ с элементами вида $Y = (\dots, y_{-1}, y_0, y_1, \dots)$, где $y_i \in \mathbb{R}^N$. Определим норму элемента Y :

$$\|Y\|^2 = \sum_{i=-\infty}^{\infty} \|y_i\|^2.$$

Элементы Y с конечной нормой образуют линейное нормированное пространство.

Вектору $x = (x_1, \dots, x_n)$ поставим в соответствие последовательность $X = (\dots, x_1, \dots, x_n, \dots)$ с $x_i = 0$ при $i \geq 0$ и при $i > n$. Соответственно вектору $f = rx = (f_0, f_1, \dots, f_n)$ сопоставим последовательность $F = (\dots, f_0, \dots, f_n, \dots)$ с $f_i = 0$ при $i < 0$ и при $i > n$.

Определим оператор R , действующий на последовательность Y по правилу

$$RY = G = (\dots, g_{-1}, g_0, g_1, \dots), \quad g_i = By_{i+1} - Ay_i.$$

Очевидно, что $RX = F$, $\|X\| = \|x\|$, $\|F\| = \|f\|$. Следовательно, оценку леммы можно получить с помощью оценки нормы оператора R^{-1} .

Рассмотрим ряд Фурье для последовательности $Y = (\dots, y_{-1}, y_0, y_1, \dots)$:

$$\widehat{Y}(\varphi) = \sum_{j=-\infty}^{\infty} y_j e^{ij\varphi}.$$

По равенству Парсеваля

$$\|Y\|^2 = \frac{1}{2\pi} \int_0^{2\pi} \|\widehat{Y}(\varphi)\|^2 d\varphi.$$

Так как уравнение $RX = F$ эквивалентно уравнению $(Be^{-i\varphi} - A)\widehat{X}(\varphi) = \widehat{F}(\varphi)$, то $\widehat{X}(\varphi) = (Be^{-i\varphi} - A)^{-1}\widehat{F}(\varphi)$.

Следовательно,

$$\begin{aligned} \|X\|^2 &= \frac{1}{2\pi} \int_0^{2\pi} \|\widehat{X}(\varphi)\|^2 d\varphi \leq \max_{\varphi} \| (Be^{-i\varphi} - A)^{-1} \|^2 \cdot \frac{1}{2\pi} \int_0^{2\pi} \|\widehat{F}(\varphi)\|^2 d\varphi = \\ &= \max_{\varphi} \| (B - e^{i\varphi}A)^{-1} \|^2 = \|F\|^2, \\ \|x\| &\leq \max_{\varphi} \| (B - e^{i\varphi}A)^{-1} \| \|rx\|. \end{aligned}$$

Лемма 4.3 доказана.

4.2. Оценки погрешностей на первом этапе алгоритма

Анализ вычислительных погрешностей алгоритма разбивается на две, не связанные между собой, части: анализ погрешностей нормировки исходного пучка матриц и анализ погрешностей в итерационном процессе. Обусловленность первой части анализа оценивается обусловленностью нормировки, т. е. числом обусловленности матрицы $(A \ B)$, а обусловленность второй части дается параметром дихотомии ω .

В этом пункте подробно исследуется процесс нормировки исходного пучка матриц $\lambda B - A$, чтобы получить нормированный пучок $\lambda B_0 - A_0$, удовлетворяющий равенству $A_0 A_0^* + B_0 B_0^* = I$. Для этого предлагается следующая вычислительная процедура:

сформируем $N \times 2N$ -матрицу $(A \ B)$;

применим преобразования отражения Хаусхолдера слева к матрице $(A \ B)$, чтобы получить матричное равенство

$$(A \ B)Q = (L \ 0) \quad (4.5)$$

с ортогональной матрицей Q и нижней треугольной L ;

вычислим матрицу

$$(A_0 \ B_0) = (I \ 0)Q^{-1}. \quad (4.6)$$

При реализации на ЭВМ этого процесса в формуле (4.5) вместо нулевой матрицы будет стоять некоторая матрица малой нормы. Чтобы от нее избавиться, воспользуемся леммой 4.1. Опишем эту процедуру подробнее. В силу результатов из 3.3 выполнена оценка

$$\|(A \ B)Q - (L_{\text{маш}}0)\|_F \leq \frac{N}{1 - \frac{N-1}{2} \varepsilon_r} (\varepsilon_r \|(A \ B)\|_F + o_r \sqrt{2N}), \quad (4.7)$$

где ε_r и o_r вычисляются при $L = 2N$. Матрица Q определена точно по вычисленным на ЭВМ векторам нормалей к гиперплоскостям отражений. Обозначим $(\eta \ \xi) = (A \ B)Q - (L_{\text{маш}}0)$. Тогда матрица $(\eta \ \xi)$ будет иметь малую норму (4.7). Применим лемму 4.1 к матрице $(L_{\text{маш}} + \eta \ \xi)$. Существует такая ортогональная матрица $I + S$, что

$$(L_{\text{маш}} + \eta \ \xi)(I + S) = (\tilde{L} \ 0),$$

$$\begin{aligned} \|S\| &\leq \|\xi\| \|(L_{\text{маш}} + \eta)^{-1}\| \leq \frac{\|\eta\xi\|}{\sigma_{\min}(A \ B) - \|\eta\xi\|} \leq \\ &\leq \frac{N^{3/2}(\varepsilon_r \|(A \ B)\| + \sqrt{2}o_r)}{\left(1 - \frac{N-1}{2} \varepsilon_r\right) \sigma_{\min}(A \ B) - N^{3/2}(\varepsilon_r \|A \ B\| + \sqrt{2}o_r)}. \end{aligned}$$

Обозначим $(A_0 \ B_0) = (I \ 0)(I + S)^{-1}Q^{-1}$. Известно, что выполнена оценка

$$\|(I \ 0)Q^{-1} - (A_0 \ B_0)_{\text{маш}}\|_F \leq \frac{N^{3/2}(\varepsilon_r + \sqrt{2}o_r)}{1 - \frac{N-1}{2} \varepsilon_r}.$$

Следовательно,

$$\begin{aligned} \|(A_0 \ B_0)_{\text{маш}} - (A_0 \ B_0)\| &\leq \|(A_0 \ B_0)_{\text{маш}} - (I \ 0)Q^{-1}\| + \\ &+ \|(I \ 0)Q^{-1} - (I \ 0)(I + S)^{-1}Q^{-1}\| \leq \|S\|(I + S)^{-1} + \frac{N^{3/2}(\varepsilon_r + o_r \sqrt{2})}{1 - \frac{N-1}{2} \varepsilon_r} \leq \\ &\leq [N^{3/2}(\varepsilon_r \|(A \ B)\| + \sqrt{2}o_r)] / \left[\left(1 - \frac{N-1}{2} \varepsilon_r\right) \sigma_{\min}(A \ B) - \right. \\ &\quad \left. - N^{3/2}(\varepsilon_r \|(A \ B)\| + \sqrt{2}o_r)\right] + \frac{N^{3/2}(\varepsilon_r + \sqrt{2}o_r)}{1 - \frac{N-1}{2} \varepsilon_r}. \quad (4.8) \end{aligned}$$

Оценка (4.8) имеет априорный характер. Попробуем получить апостериорную оценку, используя невязки. Пусть

$$\Phi = (A \ B) - L_{\text{маш}}(A_0 \ B_0)_{\text{маш}},$$

$$\Psi = (A_0 \ B_0)_{\text{маш}} \cdot (A_0 \ B_0)_{\text{маш}}^* - I.$$

Обозначим $(\hat{A}_0 \ \hat{B}_0) = (I + \Psi)^{-1/2}(A_0 \ B_0)_{\text{маш}}$, $\hat{L} = L_{\text{маш}}(I + \Psi)^{1/2}$.

В качестве значения квадратного корня берется главная ветвь функции $\sqrt{1+z}$. В 5.1 будет показано, что при $\|\Psi\| < 1$

$$(I + \Psi)^{1/2} = I + \Psi_1, \quad \|\Psi_1\| \leq \frac{\|\Psi\|}{2 - \|\Psi\|}.$$

Дополним матрицу $(\hat{A}_0 \ \hat{B}_0)$ до ортогональной

$$\hat{Q}^* = \begin{pmatrix} \hat{A}_0 & \hat{B}_0 \\ \hat{C}_0 & \hat{D}_0 \end{pmatrix}.$$

Из равенства $(A \ B) = (\hat{L} \ 0)\hat{Q} + \Phi$ следует $(A \ B)\hat{Q}^* = (\hat{L} \ 0) + \Phi\hat{Q}^*$.

Применим лемму 4.1 к матрице $(\widehat{L} \ 0) + \Phi \widehat{Q}^*$. Обозначим $(A_0 \ B_0) = (I \ 0)(I + S)^{-1}\widehat{Q}^*$. Тогда выполнены оценки

$$\begin{aligned} & \| (A_0 \ B_0)_{\text{маш}} - (A_0 \ B_0) \| \leq \\ & \leq \| (A_0 \ B_0) - (\widehat{A}_0 \ \widehat{B}_0) \| + \| (\widehat{A}_0 \ \widehat{B}_0) - (A_0 \ B_0)_{\text{маш}} \| \leq \| S \| + \| (I + \Psi)^{1/2} - I \| \leq \\ & \leq \frac{\| \Phi \|}{\sigma_{\min}(\widehat{L}) - \| \Phi \|} + \frac{\| \Psi \|}{2 - \| \Psi \|}, \end{aligned} \quad (4.9)$$

$$\| (A_0 \ B_0)_{\text{маш}} - (A_0 \ B_0) \| \leq \frac{\| \Psi \|}{2 - \| \Psi \|} + \frac{\| \Phi \|}{\sigma_{\min}(L_{\text{маш}})(1 - \| \Psi \|) - \| \Phi \|}.$$

Оценка (4.9), конечно, справедлива, если только $\| \Psi \| < 1$ и $\sigma_{\min}(L_{\text{маш}}) \times (1 - \| \Psi \|) - \| \Phi \| > 0$.

4.3. Оценки погрешностей на втором этапе алгоритма

После первого этапа алгоритма мы получаем почти нормированный пучок матриц $\lambda B_{0 \text{ маш}} - A_{0 \text{ маш}} = \lambda \widetilde{B}_0 - \widetilde{A}_0$, который удовлетворяет неравенству

$$\| (\widetilde{A}_0 \ \widetilde{B}_0) - (A_0 \ B_0) \| \leq \beta_0, \quad (4.10)$$

где β_0 — малая константа, а $\lambda B_0 - A_0$ получается при точном нормировании исходного пучка матриц $\lambda B - A$.

Для анализа устойчивости итераций второго этапа алгоритма воспользуемся наводящими соображениями из 2.4. Рассмотрим блочную двухдиагональную матрицу размеров $[N \cdot 2^{m_0}] \times [N(2^{m_0} + 1)]$:

$$\mathcal{A}_0 = \begin{bmatrix} -A_0 & B_0 & & & \\ & -A_0 & B_0 & & \\ & & \ddots & \ddots & \\ & & & \ddots & \\ & & & & -A_0 & B_0 \end{bmatrix}. \quad (4.11)$$

Оказывается, что вычислительный процесс второго этапа алгоритма можно проинтерпретировать как применение серии ортогональных преобразований слева к матрице $\mathcal{A}_0$.

Разобьем матрицу $\mathcal{A}_0$ на блоки

$$\begin{pmatrix} -A_0 & B_0 & 0 \\ 0 & -A_0 & B_0 \end{pmatrix}, \quad (4.12)$$

расположенные по диагонали матрицы $\mathcal{A}_0$. Таких блоков будет, очевидно, 2^{m_0-1} . Во время первой итерации второго этапа алгоритма все эти блоки одновременно преобразуем умножением слева на ортогональную блочнодиагональную матрицу $\mathcal{P}_1$ с диагональными блоками вида

$$\begin{pmatrix} X_0 & Y_0 \\ X_0 & Y_0 \end{pmatrix},$$

где X_0 и Y_0 удовлетворяют уравнению $X_0 B_0 - Y_0 A_0 = 0$. Диагональные блоки (4.12) матрицы $\mathcal{A}_0$ после этого преобразуются в блоки вида

$$\begin{pmatrix} C_1 & R_1 & D_1 \\ -A_1 & 0 & B_1 \end{pmatrix}.$$

Во время второй итерации у матрицы $\mathcal{A}_1 = \mathcal{P}_1 \mathcal{A}_0$ выделим блоки по

диагонали вида

$$\begin{pmatrix} -A_1 & 0 & B_1 & 0 & 0 \\ 0 & 0 & -A_1 & 0 & B_1 \end{pmatrix}. \quad (4.13)$$

Строки, содержащие блоки $(C_1 \ R_1 \ D_1)$, не участвуют в дальнейших преобразованиях. С выделенными блоками (4.13) для пары матриц $(A_1 \ B_1)$ сделаем аналогичные операции, что и для пары $(A_0 \ B_0)$.

Продолжая такой процесс ортогонального исключения, получим последовательность ортогональных матриц $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_{m_0}$, которые применяются слева к матрицам $\mathcal{A}_k = \mathcal{P}_k \dots \mathcal{P}_1 \mathcal{A}_0$. После m_0 итераций получим матрицу

$$\mathcal{A}_{m_0} = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -A_{m_0} & 0 & \cdot & \cdot & 0B_{m_0} \end{bmatrix}$$

с нижней блочной строкой, содержащей между матрицами A_{m_0} и B_{m_0} только нулевые блоки. Перепишем матрицу $\mathcal{A}_{m_0}$ в следующем виде:

$$\mathcal{A}_{m_0} = \left(\begin{array}{c|c|c} F_1 & F & F_2 \\ \hline -A_{m_0} & 0 & B_{m_0} \end{array} \right),$$

где матрицы F_1 и F_2 имеют размеры $[N(2^{m_0} - 1)] \times N$, а F — квадратная матрица. Фактически, в результате описанной выше интерпретации второго этапа алгоритма мы приводим матрицу $\mathcal{A}_0$ к блочному верхнетреугольному виду ортогональными преобразованиями слева. Это становится совсем очевидным, если у матрицы $\mathcal{A}_{m_0}$ поменять местами блочные столбцы, содержащие блоки F_1 и F . В результате такой перестановки получится матрица

$$\begin{pmatrix} F & F_1 & F_2 \\ 0 & -A_{m_0} & B_{m_0} \end{pmatrix} \quad (4.14)$$

с квадратным блоком F .

Заметим еще, что матрица $\begin{pmatrix} F \\ 0 \end{pmatrix}$ получается из блочной двухдиагональной матрицы

$$\mathcal{D} = \begin{bmatrix} B_0 & & \\ -A_0 & B_0 & \\ \cdot & \cdot & \\ \cdot & \cdot & \cdot \\ -A_0 & B_0 & \\ & & -A_0 \end{bmatrix}$$

применением ортогонального преобразования слева. По лемме 4.3

$$\|F^{-1}\| = 1/\sigma_{\min}(\mathcal{D}) \leq \max_{\varphi} \|(B_0 - e^{i\varphi} A_0)^{-1}\| < 5\pi\omega. \quad (4.15)$$

И наконец, применяя лемму 4.2 к матрице (4.14), можно утверждать, что, несмотря на неоднозначность выбора ортогональных преобразований $\mathcal{P}_j$, $j = 1, 2, \dots, m_0$, матрица $(-A_{m_0} \ B_{m_0})$ единственна с точностью до ортогонального преобразования слева.

Попытаемся смоделировать погрешности округлений в предложенной интерпретации вычислений второго этапа алгоритма. Имеются ошибки двух сортов: с одной стороны — начальные возмущения матриц A_0 и B_0 , с другой стороны — погрешности применения преобразований $\mathcal{P}_k$. На-

чальные возмущения, которые возникают на выходе первого этапа алгоритма, мы уже учли в виде неравенства

$$\|(\tilde{A}_0 \tilde{B}_0) - (A_0 B_0)\| \leq \beta_0$$

с малым β_0 . Знаком $\sim$ в дальнейшем будем обозначать результаты машинных вычислений. Легко показать, что матрица

$$\tilde{\mathcal{A}}_0 = \begin{bmatrix} -\tilde{A}_0 & \tilde{B}_0 & & \\ & -\tilde{A}_0 & \tilde{B}_0 & \\ & & \ddots & \ddots \\ & & & -\tilde{A}_0 & \tilde{B}_0 \end{bmatrix}$$

отличается от матрицы $\mathcal{A}_0$ на $\sqrt{2}\beta_0$, точнее, выполнено неравенство $\|\tilde{\mathcal{A}}_0 - \mathcal{A}_0\| \leq \sqrt{2}\beta_0$.

Далее, в результате применения машинного ортогонального преобразования, которое «почти ортогонально» в определенном смысле, блочная матрица

$$\begin{pmatrix} -\tilde{A}_0 & \tilde{B}_0 & 0 \\ 0 & -\tilde{A}_0 & \tilde{B}_0 \end{pmatrix} \quad (4.16)$$

переходит в матрицу

$$\begin{pmatrix} \tilde{C}_1 & \tilde{R}_1 & \tilde{D}_1 \\ -\tilde{A}_1 & 0 & \tilde{B}_1 \end{pmatrix}.$$

В 3.3 было показано, что существует точное ортогональное преобразование, которое переводит (4.16) в блок

$$\begin{pmatrix} \hat{C}_1 & \hat{R}_1 & \hat{D}_1 \\ -\hat{A}_1 & \Delta_1 & \hat{B}_1 \end{pmatrix},$$

причем выполняет неравенство

$$\left\| \begin{pmatrix} \hat{C}_1 & \hat{R}_1 & \hat{D}_1 \\ -\hat{A}_1 & \Delta_1 & \hat{B}_1 \end{pmatrix} - \begin{pmatrix} \tilde{C}_1 & \tilde{R}_1 & \tilde{D}_1 \\ -\tilde{A}_1 & 0 & \tilde{B}_1 \end{pmatrix} \right\| \leq \alpha \left\| \begin{pmatrix} -\tilde{A}_0 & \tilde{B}_0 & 0 \\ 0 & -\tilde{A}_0 & \tilde{B}_0 \end{pmatrix} \right\| + \beta$$

$$\text{с} \quad \alpha = \frac{N^{3/2}\varepsilon_r}{1 - \frac{N-1}{2}\varepsilon_r}, \quad \beta = \frac{N^{3/2}o_r}{1 - \frac{N-1}{2}\varepsilon_r},$$

где ε_r и o_r определены при $L = 2N$.

Аналогично можно показать, что существует точное ортогональное преобразование, переводящее блок

$$\begin{pmatrix} -\tilde{A}_m & \tilde{B}_m & 0 \\ -0 & -\tilde{A}_m & \tilde{B}_m \end{pmatrix}$$

в блок

$$\begin{pmatrix} \hat{C}_{m+1} & \hat{R}_{m+1} & \hat{D}_{m+1} \\ -\hat{A}_{m+1} & \Delta_{m+1} & \hat{B}_{m+1} \end{pmatrix},$$

Причем выполнено неравенство

$$\left\| \begin{pmatrix} -\hat{A}_{m+1} & \Delta_{m+1} & \hat{B}_{m+1} \end{pmatrix} - \begin{pmatrix} -\tilde{A}_{m+1} & 0 & \tilde{B}_{m+1} \end{pmatrix} \right\| \leq \alpha \left\| \begin{pmatrix} -\tilde{A}_m & \tilde{B}_m & 0 \\ 0 & -\tilde{A}_m & \tilde{B}_m \end{pmatrix} \right\| + \beta.$$

Матрицы точных ортогональных преобразований размеров $(N \cdot 2^{m_0}) \times (N \cdot 2^{m_0})$ на m -й итерации обозначим через $\widehat{\mathcal{P}}_m$. Определим матрицы $\widetilde{\mathcal{A}}_m$ и $\widetilde{\mathcal{A}}_m$ следующим образом. Пусть известна матрица $\widetilde{\mathcal{A}}_{m-1}$, тогда обозначим $\widetilde{\mathcal{A}}_m = \widehat{\mathcal{P}}_m \widetilde{\mathcal{A}}_{m-1}$. Строки матрицы $\widetilde{\mathcal{A}}_m$, содержащие блоки $-\widehat{A}_m$, Δ_m , $\widehat{B}_m$, заменим на строки с матрицами $-\widetilde{A}_m$, 0 , $\widetilde{B}_m$ соответственно. Тогда строки матриц $\Psi_m = \widetilde{\mathcal{A}}_m - \widetilde{\mathcal{A}}_m$ содержат либо все нулевые блоки, либо блоки

$$-(\widehat{A}_m - \widetilde{A}_m), \Delta_m, (\widehat{B}_m - \widetilde{B}_m). \quad (4.17)$$

Нетрудно проверить, что структура матриц Ψ_m позволяет доказать оценку

$$\|\Psi_m\| \leq \sqrt{2}\alpha \|\widetilde{\mathcal{A}}_{m-1}\| + \sqrt{2}\beta.$$

Это следует из того, что строки, содержащие блоки (4.17), имеют следующее «зацепление»:

$$\begin{bmatrix} -(\widehat{A}_m - \widetilde{A}_m) & \Delta_m & (\widehat{B}_m - \widetilde{B}_m) & 0 & 0 \\ 0 & 0 & -(\widehat{A}_m - \widetilde{A}_m) & \Delta_m & (\widehat{B}_m - \widetilde{B}_m) \end{bmatrix}.$$

Нас будет интересовать оценка величины $\|\widetilde{\mathcal{A}}_{m_0} - \widehat{\mathcal{P}}_{m_0} \dots \widehat{\mathcal{P}}_1 \mathcal{A}_0\|$. Она может быть получена следующим способом:

$$\begin{aligned} \|\widetilde{\mathcal{A}}_{m+1} - \widehat{\mathcal{P}}_{m+1} \dots \widehat{\mathcal{P}}_1 \mathcal{A}_0\| &= \|\widetilde{\mathcal{A}}_{m+1} - \Psi_{m+1} - \widehat{\mathcal{P}}_{m+1} \dots \widehat{\mathcal{P}}_1 \mathcal{A}_0\| = \\ &= \|\widehat{\mathcal{P}}_{m+1}(\widetilde{\mathcal{A}}_m - \widehat{\mathcal{P}}_m \dots \widehat{\mathcal{P}}_1 \mathcal{A}_0) - \Psi_{m+1}\| \leq \|\Psi_{m+1}\| + \|\widetilde{\mathcal{A}}_m - \widehat{\mathcal{P}}_m \dots \widehat{\mathcal{P}}_1 \mathcal{A}_0\| \leq \\ &\leq \|\widetilde{\mathcal{A}}_m - \widehat{\mathcal{P}}_m \dots \widehat{\mathcal{P}}_1 \mathcal{A}_0\| + \sqrt{2}\alpha \|\widetilde{\mathcal{A}}_m\| + \sqrt{2}\beta \leq \\ &\leq (1 + \alpha\sqrt{2}) \|\widetilde{\mathcal{A}}_m - \widehat{\mathcal{P}}_m \dots \widehat{\mathcal{P}}_1 \mathcal{A}_0\| + \sqrt{2}\alpha \|\mathcal{A}_0\| + \sqrt{2}\beta, \\ \|\widetilde{\mathcal{A}}_{m_0} - \widehat{\mathcal{P}}_{m_0} \dots \widehat{\mathcal{P}}_1 \mathcal{A}_0\| &\leq (1 + \alpha\sqrt{2})^{m_0} \|\widetilde{\mathcal{A}}_0 - \mathcal{A}_0\| + \\ &+ \sum_{j=0}^{m_0-1} (1 + \alpha\sqrt{2})^j \{\alpha\sqrt{2} \|\mathcal{A}_0\| + \beta\sqrt{2}\} \leq \\ &\leq \frac{\|\widetilde{\mathcal{A}}_0 - \mathcal{A}_0\| + m_0(\alpha\sqrt{2} \|\mathcal{A}_0\| + \beta\sqrt{2})}{1 - \alpha\sqrt{2}m_0} \leq \frac{\sqrt{2}}{1 - \alpha\sqrt{2}m_0} (\beta_0 + m_0\alpha \|\mathcal{A}_0\| + m_0\beta). \end{aligned}$$

Обозначим $\widehat{\mathcal{P}} = \widehat{\mathcal{P}}_{m_0} \dots \widehat{\mathcal{P}}_1$, тогда матрица $\widehat{\mathcal{P}}\mathcal{A}_0$ мало отличается от матрицы $\widetilde{\mathcal{A}}_{m_0}$:

$$\|\widetilde{\mathcal{A}}_{m_0} - \widehat{\mathcal{P}}\mathcal{A}_0\| \leq \frac{\sqrt{2}}{1 - \alpha\sqrt{2}m_0} (\beta_0 + m_0\alpha \|\mathcal{A}_0\| + m_0\beta). \quad (4.18)$$

Заметим, что при точных вычислениях мы бы получили равенство $\mathcal{A}_{m_0} = \mathcal{P}\mathcal{A}_0$ с ортогональной матрицей $\mathcal{P}$ вместо неравенства (4.18).

Легко убедиться в том, что матрица $\widetilde{\mathcal{A}}_{m_0}$ имеет следующую блочную структуру:

$$\widetilde{\mathcal{A}}_{m_0} = \begin{bmatrix} \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ -\widetilde{A}_{m_0} & 0 & \cdots & 0 & \widetilde{B}_{m_0} \end{bmatrix}.$$

Матрицу $\widehat{\mathcal{P}}\mathcal{A}_0$ представим в следующем блочном виде:

$$\widehat{\mathcal{P}}\mathcal{A}_0 = \begin{pmatrix} \widehat{F}_1 & \widehat{F} & \widehat{F}_2 \\ -\widehat{A}_{m_0} & \Phi & \widehat{B}_{m_0} \end{pmatrix}.$$

Тогда матрица $\begin{pmatrix} -\dot{A}_{m_0} & \dot{B}_{m_0} \end{pmatrix}$ мало отличается от $\begin{pmatrix} -\tilde{A}_{m_0} & \tilde{B}_{m_0} \end{pmatrix}$, а матрица Φ близка к нулевой. Теперь вспомним, что матрица $\begin{pmatrix} \widehat{F} \\ \Phi \end{pmatrix}$ получается из матрицы

$$\mathcal{D} = \begin{bmatrix} B_0 & & & \\ -A_0 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & B_0 \\ & & & -A_0 \end{bmatrix}$$

ортогональным преобразованием слева. По лемме 4.1 существует ортогональная матрица $I + S$, умножением на которую слева матрица $\begin{pmatrix} \widehat{F} \\ \Phi \end{pmatrix}$ приводится к $\begin{pmatrix} \widehat{F} \\ 0 \end{pmatrix}$. Более того, выполнена оценка

$$\|S\| \leq \|\Phi\| \|\widehat{F}^{-1}\| \leq \frac{\|\Phi\|}{\sigma_{\min}(\mathcal{D}) - \|\Phi\|}.$$

По лемме 4.3

$$\sigma_{\min}(\mathcal{D}) \geq \frac{1}{\max_{\varphi} \|(B_0 - e^{i\varphi} A_0)^{-1}\|},$$

а $\|\Phi\| \leq \|\tilde{\mathcal{A}}_{m_0} - \widehat{\mathcal{P}}\mathcal{A}_0\|$. Обозначим $(I + S)\widehat{\mathcal{P}}\mathcal{A}_0 = \mathcal{P}\mathcal{A}_0$, тогда можно вывести неравенства

$$\begin{aligned} \|\mathcal{P}\mathcal{A}_0 - \tilde{\mathcal{A}}_{m_0}\| &\leq \|S\| \|\mathcal{A}_0\| + \|\widehat{\mathcal{P}}\mathcal{A}_0 - \tilde{\mathcal{A}}_{m_0}\| \leq \\ &\leq \|\Phi\| + \frac{\|\Phi\| \|\mathcal{A}_0\|}{\frac{1}{\max_{\varphi} \|(B_0 - e^{i\varphi} A_0)^{-1}\|} - \|\Phi\|} \leq \\ &\leq \|\Phi\| \frac{1 + \|\mathcal{A}_0\| \max_{\varphi} \|(B_0 - e^{i\varphi} A_0)^{-1}\|}{1 - \|\Phi\| \max_{\varphi} \|(B_0 - e^{i\varphi} A_0)^{-1}\|}. \end{aligned} \quad (4.19)$$

Учитывая неравенства $\|\mathcal{A}_0\| \leq \sqrt{2}$, $\max_{\varphi} \|(B_0 - e^{i\varphi} A_0)^{-1}\| < 5\pi\omega$ и (4.18), оценку (4.19) можно превратить в следующую оценку:

$$\|\tilde{\mathcal{A}}_{m_0} - \mathcal{P}\mathcal{A}_0\| \leq \frac{\sqrt{2}(1 + 5\sqrt{2}\pi\omega)(\beta_0 + \alpha\sqrt{2}m_0 + \beta m_0)}{1 - 2\sqrt{2}\alpha m_0 - 5\sqrt{2}\pi\omega(\beta_0 + \alpha\sqrt{2}m_0 + \beta m_0)}. \quad (4.20)$$

Матрица $\mathcal{A}_0$ имеет полный ранг, так как пучок $\lambda B_0 - A_0$ регулярен относительно единичной окружности. Следовательно, по лемме 4.2 существует некоторый точный вычислительный процесс, в котором получаются матрицы A_{m_0} и B_{m_0} , удовлетворяющие оценке

$$\|(-\tilde{A}_{m_0} \tilde{B}_{m_0}) - (-A_{m_0} B_{m_0})\| \leq \|\tilde{\mathcal{A}}_{m_0} - \mathcal{P}\mathcal{A}_0\|.$$

4.4. Оценки погрешностей на третьем этапе алгоритма

На этом этапе погрешности складываются из ошибок предыдущих двух этапов, ошибок округления при реализации вычислений самого третьего этапа и ошибок, связанных с тем, что итерации не ведутся до бесконечности, а обрываются при $m = m_0$.

Ошибки, возникающие из-за конечности итерационного процесса, были тщательно исследованы в 4.3. Поэтому можно считать, что эти ошибки уже учтены в погрешностях входных данных третьего этапа алгоритма. Либо можно их добавить к погрешностям выходных матриц H_{m_0} и Z_0, Z_∞ .

Отправной точкой в последующем анализе послужит нам неравенство

$$\|(-\tilde{A}_{m_0}\tilde{B}_{m_0}) - (-A_{m_0}B_{m_0})\| \leq \delta \quad (4.21)$$

$$\delta = \frac{\sqrt{2}(1 + 5\sqrt{2}\pi\omega)(\beta_0 + \alpha\sqrt{2}m_0 + \beta m_0)}{1 - 2\sqrt{2}\alpha m_0 - 5\sqrt{2}\pi\omega(\beta_0 + \alpha\sqrt{2}m_0 + \beta m_0)},$$

где α, β, β_0 определены в п. 4.3. На третьем этапе конструируются матрицы

$$\begin{aligned} H_{m_0} &= (A_{m_0} + B_{m_0})^{-1} (A_{m_0}^* + B_{m_0}^*)^{-1}, \\ Z_0 &= (A_{m_0} + B_{m_0})^{-1} B_{m_0}, \\ Z_\infty &= I - Z_0 = (A_{m_0} + B_{m_0})^{-1} A_{m_0}. \end{aligned} \quad (4.22)$$

Требуется оценить нормы разностей $\tilde{H}_{m_0} - H_{m_0}$ и $\tilde{Z}_0 - Z_0$, где $\tilde{H}_{m_0}$ и $\tilde{Z}_0$ — вычисленные по формулам (4.22) матрицы. Естественно, что вычисления эти проводятся на ЭВМ с плавающей точкой и с входными матрицами $\tilde{A}_{m_0}$ и $\tilde{B}_{m_0}$.

Введем обозначения $F = (A_{m_0} + B_{m_0})^{-1}$, $\tilde{F} = (\tilde{A}_{m_0} + \tilde{B}_{m_0})^{-1}$. Из формулы

$$\tilde{F} = F - \tilde{F}[(\tilde{A}_{m_0} + \tilde{B}_{m_0}) - (A_{m_0} + B_{m_0})]F$$

следует оценка

$$\|\tilde{F} - F\| \leq \delta\sqrt{2}\|\tilde{F}\|\|F\|.$$

Здесь мы воспользовались простым неравенством

$$\|(\tilde{A}_{m_0} + \tilde{B}_{m_0}) - (A_{m_0} + B_{m_0})\| \leq \delta\sqrt{2}.$$

В силу оценок

$$\begin{aligned} \sigma_{\min}(\tilde{A}_{m_0} + \tilde{B}_{m_0}) &\geq \sigma_{\min}(A_{m_0} + B_{m_0}) - \delta\sqrt{2}, \\ \|\tilde{F}\| &\leq \frac{1}{\sigma_{\min}(F^{-1}) - \delta\sqrt{2}} \leq \frac{\|F\|}{1 - \delta\sqrt{2}\|F\|} \end{aligned}$$

справедлива цепочка неравенств

$$\|\tilde{F} - F\| \leq \frac{\delta\sqrt{2}\|F\|^2}{1 - \delta\sqrt{2}\|F\|}, \quad \|F_{\text{маш}} - F\| \leq \frac{\delta\sqrt{2}\|F\|^2}{1 - \delta\sqrt{2}\|F\|} + \Delta(\tilde{F}),$$

где через $\Delta(\tilde{F})$ мы обозначили погрешность обращения на ЭВМ матрицы $\tilde{F}^{-1}$, исследованную в 3.4.

Далее, используя результаты п. 3.2, получим неравенства

$$\begin{aligned} \|(F_{\text{маш}}F_{\text{маш}}^*)_{\text{маш}} - FF^*\| &\leq (\varepsilon_i\sqrt{N} + \varepsilon_{ii}N)\|F_{\text{маш}}\|^2 + o_i + \\ &+ \|F_{\text{маш}}F_{\text{маш}}^* - FF^*\| \leq \|F_{\text{маш}} - F\|^2 + 2\|F\|\|F_{\text{маш}} - F\| + \\ &+ (\varepsilon_i\sqrt{N} + \varepsilon_{ii}N)\|F_{\text{маш}}\|^2 + o_i \leq o_i + (\varepsilon_i\sqrt{N} + \varepsilon_{ii}N)\|F\|^2 + \\ &+ [(\varepsilon_i\sqrt{N} + \varepsilon_{ii}N)(2\|F\| + \|F_{\text{маш}} - F\|) + 2\|F\| + \|F_{\text{маш}} - F\|]\|F_{\text{маш}} - \\ &- F\| \leq o_i + (\varepsilon_i\sqrt{N} + \varepsilon_{ii}N)\|F\|^2 + \\ &+ \|F_{\text{маш}} - F\|(2\|F\| + \|F_{\text{маш}} - F\|)(1 + \varepsilon_i\sqrt{N} + \varepsilon_{ii}N). \end{aligned}$$

В итоге выводим следующую оценку:

$$\begin{aligned} & \|\tilde{H}_{m_0} - H_{m_0}\| \leq \alpha_i + (\varepsilon_i \sqrt{N} + \varepsilon_{ii}N) \|F\|^2 + \\ & + \left[\frac{\delta \sqrt{2} \|F\|^2}{1 - \delta \sqrt{2} \|F\|} + \Delta(\tilde{F}) \right] (1 + \varepsilon_i \sqrt{N} + \varepsilon_{ii}N) \left(\frac{2 \|F\|}{1 - \delta \sqrt{2} \|F\|} + \Delta(\tilde{F}) \right). \end{aligned}$$

Аналогично можно вывести оценку погрешности вычисления матрицы Z_0 , которая является приближением к проектору P_0 :

$$\begin{aligned} \|\tilde{Z}_0 - Z_0\| &= \|(F_{\text{маш}} \tilde{B}_{m_0})_{\text{маш}} - FB_{m_0}\| \leq \varepsilon_i \sqrt{N} \|F_{\text{маш}} \tilde{B}_{m_0}\| + \\ &+ \varepsilon_{ii}N \|F_{\text{маш}}\| \|\tilde{B}_{m_0}\| + \|F_{\text{маш}} - F\| \|B_{m_0}\| + \|F\| \|\tilde{B}_{m_0} - B_{m_0}\| + \\ &+ \|F_{\text{маш}} - F\| \|\tilde{B}_{m_0} - B_{m_0}\| \leq \varepsilon_i \sqrt{N} \|Z_0\| + \varepsilon_{ii}N \|F\| \|B_{m_0}\| + \\ &+ (1 + \varepsilon_i \sqrt{N} + \varepsilon_{ii}N) \left[\frac{\delta \sqrt{2} \|F\|^2 \|B_{m_0}\| + \delta \|F\|}{1 - \delta \sqrt{2} \|F\|} + \Delta(\tilde{F}) (\|B_{m_0}\| + \delta) \right]. \end{aligned}$$

Перепишем предыдущие оценки, учитывая равенство $\|F\| = \|H_{m_0}\|^{1/2}$ и неравенство $\|B_{m_0}\| \leq 1$:

$$\begin{aligned} \|\tilde{H}_{m_0} - H_{m_0}\| &\leq \alpha_i + (\varepsilon_i \sqrt{N} + \varepsilon_{ii}N) \|H_{m_0}\| + (1 + \varepsilon_i \sqrt{N} + \varepsilon_{ii}N) \times \\ &\times \frac{[\delta \sqrt{2} \|H_{m_0}\| + \Delta(\tilde{F})] [2 \|H_{m_0}\|^{1/2} + \Delta(\tilde{F})]}{1 - 2 \sqrt{2} \delta \|H_{m_0}\|^{1/2}}, \end{aligned} \quad (4.23)$$

$$\begin{aligned} \|\tilde{Z}_0 - Z\| &\leq \varepsilon_i \sqrt{N} \|Z_0\| + \varepsilon_{ii}N \|H_{m_0}\|^{1/2} + \\ &+ \frac{1 + \varepsilon_i \sqrt{N} + \varepsilon_{ii}N}{1 - \delta \sqrt{2} \|H_{m_0}\|^{1/2}} [\delta \sqrt{2} \|H_{m_0}\| + \delta \|H_{m_0}\|^{1/2} + \Delta(\tilde{F}) (1 + \delta)]. \end{aligned} \quad (4.24)$$

Оставшуюся часть параграфа посвятим оценкам возмущений спектральных характеристик пучка $\lambda B - A$, которые можно получить, используя полученные в этом параграфе результаты анализа алгоритма. Предположим теперь, что вычисления ведутся точно, без округления арифметических операций. Пусть вначале имеются два пучка: $\lambda B - A$ и $\lambda \tilde{B} - \tilde{A}$, который является возмущением первого:

$$\|(\tilde{A} \ \tilde{B}) - (A \ B)\| \leq \delta \|(A \ B)\|. \quad (4.25)$$

После первого этапа алгоритма получим оценку, которая выводится аналогично (4.8):

$$\|(\tilde{A}_0 \ \tilde{B}_0) - (A_0 \ B_0)\| \leq \frac{\delta \|(A \ B)\|}{\sigma_{\min}(A \ B) - \delta \|(A \ B)\|} \leq \frac{\delta}{\frac{\delta_{\min}(A \ B)}{\sigma_{\max}(A \ B)} - \delta}. \quad (4.26)$$

Предположим, что итерационный процесс на втором этапе алгоритма бесконечный. Обозначим через A_∞ и B_∞ матрицы A_{m_0} и B_{m_0} при бесконечно большом m_0 . Оценка (4.21) при $\alpha = \beta = 0$ превращается в оценку

$$\|(-\tilde{A}_\infty \ \tilde{B}_\infty) - (-A_\infty \ B_\infty)\| < \frac{33\omega\beta_0}{1 - 23\omega\beta_0}, \quad (4.27)$$

где $\beta_0 = \delta / [\sigma_{\min}(A \ B) / \sigma_{\max}(A \ B) - \delta]$.

Так как $\sigma_{\min}(A_\infty + B_\infty) = 1/\sqrt{\omega}$, то

$$\left| \frac{1}{\sqrt{\tilde{\omega}}} - \frac{1}{\sqrt{\omega}} \right| < \frac{33\omega\beta_0}{1 - 23\omega\beta_0}.$$

Несложные выкладки приводят к оценке

$$\left| \frac{\tilde{\omega} - \omega}{\omega} \right| \leq V\bar{\omega} \left| \frac{1}{V\bar{\omega}} - \frac{1}{V\bar{\omega}} \right| \left| \left(1 - 2 V\bar{\omega} \left| \frac{1}{V\bar{\omega}} - \frac{1}{V\bar{\omega}} \right| \right) \right| < \\ < \frac{66\omega^{3/2}\beta_0}{1 - 89\omega^{3/2}\beta_0}. \quad (4.28)$$

В сравнении с оценкой (4.24) эта оценка, конечно, хуже.

Что касается оценки возмущений матрицы P_0 , то из (4.24) следует, что

$$\|\tilde{P}_0 - P_0\| \leq (1 + V\bar{2}) \omega \frac{33\omega\beta_0}{1 - 23\omega\beta_0} \left| \left(1 - V\bar{2}\omega \frac{33\omega\beta_0}{1 - 23\omega\beta_0} \right) \right| < \frac{80\omega^2\beta_0}{1 - 70\omega^{3/2}\beta_0}. \quad (4.29)$$

Таким образом, в дополнение к оценке (4.24) получена оценка (4.29) возмущения проектора пучка матриц с помощью параметра ω . Тем самым показано, что параметры ω и обусловленность нормировки $\sigma_{\max}(A B)/\sigma_{\min}(A B)$ позволяют оценить качество устойчивости задачи о вычислении проекторов P_0 и P_∞ регулярного относительно единичной окружности пучка матриц.

Не представляет особых трудностей, поступая аналогично рассуждениям п. 2.2, получить оценки возмущений от начальных данных для ортогональных проекторов Π_0 и Π_∞ .

§ 5. ОЦЕНКА ТОЧНОСТИ ВЫЧИСЛЕННЫХ СПЕКТРАЛЬНЫХ ХАРАКТЕРИСТИК РЕГУЛЯРНОГО ЛИНЕЙНОГО ПУЧКА МАТРИЦ

Как уже говорилось ранее, анализ вычислительных погрешностей, проведенный в § 4, не дает гарантии точности вычисленного значения параметра ω . Для замыкания системы оценок § 4 нужна оценка сверху параметра ω .

В настоящем параграфе предлагается методика оценки точности вычисленных спектральных характеристик пучка матриц $\lambda B - A$, матриц H , P_0 и P_∞ , основанная на обобщении теории уравнения Ляпунова. Такой способ был впервые применен А. Я. Булгаковым для оценки точности решения задачи о дихотомии спектра матрицы мнимой осью [7].

Предлагаемая методика оценки точности состоит в вычислении нескольких невязок матричной алгебраической системы уравнений и применении теории возмущений.

Простой пример для демонстрации этого подхода доставляет классическое уравнение Ляпунова

$$A^*H + HA = -I, \quad H = H^*.$$

Предположим, что матрица $\tilde{H}$ является приближенным решением этого уравнения, т. е. $C = A^*\tilde{H} + \tilde{H}A + I$ — матрица малой нормы. Если $\tilde{H} = \tilde{H}^* > 0$ и $\|C\| < 1$, то по теореме Ляпунова матрица A асимптотически устойчива. Более того,

$$\tilde{H} = \int_0^\infty e^{tA^*} (I - C) e^{tA} dt, \\ \|\tilde{H} - H\| = \left\| \int_0^\infty e^{tA^*} C e^{tA} dt \right\| = \max_{\|x\|=1} \left| \left(\int_0^\infty e^{tA^*} C e^{tA} dt x, x \right) \right| \leq \\ \leq \|C\| \max_{\|x\|=1} \left| \left(\int_0^\infty e^{tA^*} e^{tA} dt x, x \right) \right| = \|C\| \|H\|.$$

Следовательно, норма невязки уравнения Ляпунова $\|A^*\tilde{H} + \tilde{H}A + I\|$ дает относительную погрешность вычисленной матрицы $\tilde{H}$.

5.1. Обобщение уравнения Ляпунова

Сначала поясним классические результаты для матрицы A , имеющей все собственные числа внутри единичного круга. Рассматривается уравнение Ляпунова

$$AXA^* - X = -C. \quad (5.1)$$

Это матричное уравнение относительно X имеет единственное решение, если все попарные произведения собственных чисел матрицы A не равны 1, что, конечно, справедливо для матриц со спектром внутри единичного круга.

Принципиально важную роль в теории устойчивости играет теорема Ляпунова.

Теорема 5.1. Пусть матрицы $X = X^* > 0$ и $C = C^* > 0$ удовлетворяют уравнению (5.1). Тогда все собственные числа матрицы A лежат внутри единичного круга.

Легко убедиться в том, что, если все собственные числа матрицы A лежат в единичном круге, то решение уравнения (5.1) задается формулой

$$X = \sum_{j=0}^{\infty} A^j C (A^*)^j. \quad (5.2)$$

Приступим к изложению одного из обобщений результатов Ляпунова на пучки матриц, не имеющих точек спектра на единичной окружности. Выбор предлагаемого варианта обобщения определяется прежде всего возможностью построения для него достаточно простой и наглядной теории возмущений.

Рассмотрим следующую систему матричных уравнений относительно тройки квадратных $N \times N$ -матриц (P, T, X) :

$$\begin{aligned} P^2 - P &= 0, \quad TA(I - P) - (I - P) = 0, \quad (I - P)TA - (I - P) = 0, \\ TVP - P &= 0, \quad PTB - P = 0, \quad PXP^* + (I - P)X(I - P^*) - X = 0, \\ TA X(TA)^* - TB X(TB)^* &= (I - P)C(I - P^*) - PCP^*. \end{aligned} \quad (5.3)$$

Теорема 5.2. Предположим, что существуют матрицы P, T, X, C , удовлетворяющие системе уравнений (5.3), причем $X = X^* > 0, C = C^* > 0$. Тогда $\det(B - e^{i\varphi}A) \neq 0$ при любом вещественном φ , P — проектор на инвариантное подпространство пучка $\lambda B - A$, отвечающее собственным числам внутри единичного круга, $I - P$ — проектор на инвариантное подпространство пучка $\lambda B - A$, отвечающее собственным числам вне единичного круга.

Доказательство. Обозначим $P_{+0} = P, P_{-0} = I - P, P_{+1} = TAP, P_{-1} = TB(I - P)$. Тогда легко проверить выполнение тождеств $P_{-0}P_{+1} = P_{+1}P_{-0} = P_{+0}P_{-1} = P_{-1}P_{+0} = 0$. Справедливо также представление $A = T^{-1}(P_{-0} + P_{+1}), B = T^{-1}(P_{+0} + P_{-1})$. Из двух последних уравнений системы (5.3) выведем уравнение Ляпунова

$$X - (P_{-1} + P_{+1})X(P_{-1} + P_{+1})^* = P_{+0}CP_{+0}^* + P_{-0}CP_{-0}^*.$$

Действительно,

$$\begin{aligned} & (P_{-0} + P_{+1})X(P_{-0} + P_{+1})^* - (P_{+0} + P_{-1})X(P_{+0} + P_{-1})^* = \\ &= P_{-0}XP_{-0}^* + P_{+1}XP_{+1}^* - P_{+0}XP_{+0}^* - P_{-1}XP_{-1}^* = P_{-0}CP_{-0}^* - P_{+0}CP_{+0}^*, \\ & P_{-0}XP_{-0}^* - P_{-1}XP_{-1}^* = P_{-0}CP_{-0}^*, \quad P_{+0}XP_{+0}^* - P_{+1}XP_{+1}^* = P_{+0}CP_{+0}^*, \\ & P_{-0}XP_{-0}^* + P_{+0}XP_{+0}^* = X, \quad X - P_{-1}XP_{-1}^* - P_{+1}XP_{+1}^* = \\ &= X - (P_{-1} + P_{+1})X(P_{-1} + P_{+1})^*. \end{aligned}$$

Так как $X = X^* > 0$ и

$$P_{+0}CP_{+0}^* + P_{-0}CP_{-0}^* \geq \frac{\sigma_{\min}(C)}{2} I > 0,$$

то по теореме 5.1 матрица $P_{+1} + P_{-1}$ имеет все собственные числа внутри единичного круга. Осталось воспользоваться теоремой 1.2. Теорема 5.2 доказана.

Непосредственной проверкой можно убедиться, что матрица X при выполнении условий теоремы 5.2 представима в виде суммы ряда

$$X = \sum_{j=-\infty}^{\infty} P_j C P_j^*. \quad (5.4)$$

Сходимость этого ряда при любой матрице C гарантируется оценкой (1.31) из теоремы 1.4. Единственность решения системы (5.3), матриц P , T , X следует из теоремы 5.2 и единственности решения уравнения Ляпунова

$$X - (P_{-1} + P_{+1}) X (P_{-1} + P_{+1})^* = P_{+0} C P_{+0}^* + P_{-0} C P_{-0}^*.$$

В этом параграфе нас будет интересовать главным образом решение системы (5.3) при $C = I$. В этом случае

$$X = \sum_{j=-\infty}^{\infty} P_j P_j^* = \frac{1}{2} (H + P_{+0} P_{+0}^* + P_{-0} P_{-0}^*), \quad (5.5)$$

где

$$H = \frac{1}{2\pi} \int_0^{2\pi} (B - e^{i\varphi} A)^{-1} (A A^* + B B^*) (B^* - e^{-i\varphi} A^*)^{-1} d\varphi.$$

В дальнейшем мы воспользуемся следующей модификацией теоремы 5.2.

Теорема 5.3. Пусть матрицы P , T и X удовлетворяют системе уравнений

$$\begin{aligned} P^2 - P &= 0, \\ TA(I - P) - (I - P) &= 0, \\ (I - P)TA - (I - P) &= 0, \\ TBP - P &= 0, \\ PTB - P &= 0, \\ PXP^* + (I - P)X(I - P^*) - X &= 0, \\ TAX(TA)^* - TBX(TB)^* + P + P^* - I &= \Phi \end{aligned} \quad (5.6)$$

с $\|\Phi\| < 1$. Тогда $\det(B - e^{i\varphi} A) \neq 0$ при любом вещественном φ , а P и $I - P$ — проекторы пучка $\lambda B - A$ на инвариантные подпространства, отвечающие собственным числам соответственно внутри и вне единичного круга. Более того,

$$\left\| X - \sum_{j=-\infty}^{\infty} P_j P_j^* \right\| \leq \left\| \sum_{j=-\infty}^{\infty} P_j P_j^* \right\| \|\Phi\|. \quad (5.7)$$

Доказательство. Следуя доказательству теоремы 5.2, необходимо показать, что

$$W = P(I - \Phi)P^* + (I - P)(I + \Phi)(I - P^*) > 0.$$

Действительно, $W \geq \frac{1 - \|\Phi\|}{2} I > 0$. Следовательно, $\det(B - e^{i\varphi} A) \neq 0$ для любого $\varphi \in \mathbf{R}$, а P и $I - P$ — параллельные проекторы пучка $\lambda B - A$.

Покажем справедливость оценки (5.7). Воспользуемся опять доказательством теоремы 5.2. Матрица X удовлетворяет уравнению Ляпунова

$$(P_{-1} + P_{+1})X(P_{-1} + P_{+1})^* - X = -W$$

с устойчивой матрицей $P_{-1} + P_{+1}$. Непосредственной подстановкой в это уравнение легко убедиться, что $X = \sum_{j=-\infty}^{\infty} P_j W P_j^*$. Следовательно,

$$X - \sum_{j=-\infty}^{\infty} P_j P_j^* = \sum_{j \leq -1} P_j \Phi P_j^* - \sum_{j \geq +1} P_j \Phi P_j^*,$$

$$\left\| X - \sum_{j=-\infty}^{\infty} P_j P_j^* \right\| \leq \|\Phi\| \left\| \sum_{j=-\infty}^{\infty} P_j P_j^* \right\|.$$

Таким образом, теорема 5.3 доказана.

5.2. Теория возмущений для обобщенного уравнения Ляпунова

Предположим, что все уравнения системы (5.6) выполнены приближенно, т. е. некоторые матрицы $\tilde{P}$, $\tilde{T}$ и $\tilde{X}$ удовлетворяют равенствам

$$\begin{aligned} \tilde{P}^2 - \tilde{P} &= \Phi_1, \\ \tilde{T}A(I - \tilde{P}) - (I - \tilde{P}) &= \Phi_2, \\ (I - \tilde{P})\tilde{T}A - (I - \tilde{P}) &= \Phi_3, \\ \tilde{T}B\tilde{P} - \tilde{P} &= \Phi_4, \\ \tilde{P}\tilde{T}B - \tilde{P} &= \Phi_5, \\ \tilde{P}\tilde{X}\tilde{P}^* + (I - \tilde{P})\tilde{X}(I - \tilde{P}^*) - \tilde{X} &= \Phi_6, \\ \tilde{T}A\tilde{X}(\tilde{T}A)^* - \tilde{T}B\tilde{X}(\tilde{T}B)^* + \tilde{P} + \tilde{P}^* - I &= \Phi_7 \end{aligned} \quad (5.8)$$

с матрицами $\Phi_1, \Phi_2, \Phi_3, \Phi_4, \Phi_5, \Phi_6, \Phi_7$ малой нормы. Будем также предполагать, что $\tilde{X} = \tilde{X}^* \geq \rho I$ с положительной константой ρ . Требуется оценить величины $\|\tilde{X} - \sum P_j P_j^*\|$ и $\|\tilde{P} - P_{+0}\|$, а также получить условия на матрицы $\Phi_i, i = 1, 2, \dots, 7$, и константу ρ , при которых возможны такие оценки. Матрицы $P_j, j = -\infty, \dots, +\infty$, определяются для пучка $\lambda B - A$.

Идея получения таких оценок состоит в построении матриц $\widehat{T}, \widehat{P}, \widehat{A}, \widehat{B}, \widehat{X}$, мало отличающихся от матриц $\tilde{T}, \tilde{P}, A, B, \tilde{X}$, но уже удовлетворяющих условиям теоремы 5.3. Начнем с первого уравнения

$$\tilde{P}^2 - \tilde{P} = \Phi_1. \quad (5.9)$$

Воспользуемся формой Шура матрицы $\tilde{P}$:

$$\tilde{P} = U^* \begin{bmatrix} \alpha_1 & & \\ & \alpha_2 & * \\ & & \ddots \\ 0 & & & \alpha_N \end{bmatrix} U, \quad U^* U = I.$$

Из (5.9) следуют неравенства $|\alpha_i^2 - \alpha_i| \leq \|\Phi_1\|$. Поэтому при $\|\Phi_1\| \leq 1/4$ либо $|\alpha_i| \leq \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|}$, либо $|\alpha_i - 1| \leq \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|}$. Теперь всюду ниже предполагаем, что $\|\Phi_1\| < 1/4$.

Таким образом, собственные числа матрицы $\tilde{P}$ разбиваются на две непересекающиеся группы: в одной находятся собственные значения на расстоянии меньше $1/2$ от нуля, в другой — меньше $1/2$ от единицы. Форму Шура матрицы $\tilde{P}$ выберем с учетом этого разбиения:

$$\tilde{P} = U^* \begin{pmatrix} S_0 & S \\ 0 & S_1 \end{pmatrix} U,$$

где S_0 — верхнетреугольная матрица с собственными числами в окрестности нуля, а S_1 — верхнетреугольная матрица с собственными числами в окрестности единицы. В этих обозначениях уравнение (5.9) примет следующий вид:

$$\begin{pmatrix} S_0^2 - S_0 & S_0 S + S S_1 - S \\ 0 & S_1^2 - S_1 \end{pmatrix} = U \Phi_1 U^* = \begin{pmatrix} F_0 & F \\ 0 & F_1 \end{pmatrix},$$

из которого получим два уравнения на матрицы S_0 и S_1 :

$$S_0^2 - S_0 = F_0, \quad S_1^2 - S_1 = F_1.$$

Выпишем решение первого из этих уравнений: $S_0 = \frac{1}{2}(I - \sqrt{I + 4F_0})$, где квадратный корень из матрицы $I + 4F_0$ берется в смысле главного значения, т. е.

$$\sqrt{I + 4F_0} = I + \frac{1}{2}(4F_0) + \frac{\frac{1}{2}\left(\frac{1}{2} - 1\right)}{2!}(4F_0)^2 + \dots$$

Следовательно,

$$\|S_0\| \leq \frac{1}{2}(1 - \sqrt{1 - 4\|F_0\|}) \leq \frac{\|F_0\|}{1 - 2\|F_0\|} \leq \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|}. \quad (5.10)$$

Аналогичным способом можно получить неравенство

$$\|S_1 - I\| \leq \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|}. \quad (5.11)$$

Матрицу $\hat{P}$ построим по формуле

$$\hat{P} = U^* \begin{pmatrix} 0 & S \\ 0 & I \end{pmatrix} U,$$

тогда, очевидно, выполняется тождество $\hat{P}^2 = \hat{P}$ и выполнено неравенство

$$\|\tilde{P} - \hat{P}\| = \left\| \begin{pmatrix} S_0 & 0 \\ 0 & S_1 - I \end{pmatrix} \right\| \leq \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|}. \quad (5.12)$$

Рассмотрим уравнения

$$\tilde{T}A(I - \tilde{P}) - (I - \tilde{P}) = \Phi_2, \quad (5.13)$$

$$(I - \tilde{P})\tilde{T}A - (I - \tilde{P}) = \Phi_3. \quad (5.14)$$

Матрицу $\tilde{T}A$ перепишем в базисе U :

$$\tilde{T}A = U^* \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} U.$$

Уравнение (5.13) дает следующие соотношения на матрицы A_{11} , A_{12} , A_{21} , A_{22} :

$$\begin{bmatrix} (A_{11} - I)(I - S_0) & -(A_{11} - I)S + A_{12}(I - S_1) \\ A_{21}(I - S_0) & -A_{21}S + (A_{22} - I)(I - S_1) \end{bmatrix} = U\Phi_2U^*.$$

Подобным образом из (5.14) следует уравнение

$$\begin{bmatrix} (I - S_0)(A_{11} - I) - SA_{21} & (I - S_0)A_{12} - S(A_{22} - I) \\ (I - S_1)A_{21} & (I - S_1)(A_{22} - I) \end{bmatrix} = U\Phi_3U^*.$$

Суммируя второе и четвертое уравнения системы (5.8) и предполагая, что выполнено неравенство

$$\|\Phi_2\| + \|\Phi_4\| < 1, \quad (5.15)$$

из уравнения

$$\tilde{T}[A(I - \tilde{P}) + B\tilde{P}] = I - \Phi_2 - \Phi_4$$

следует обратимость матрицы $\tilde{T}$. В дальнейшем предполагается, что (5.15) имеет место.

Выберем матрицу $\hat{A}$ так, чтобы выполнялось равенство

$$\tilde{T}\hat{A} = U^* \begin{pmatrix} I & S(A_{22} - I) \\ 0 & A_{22} \end{pmatrix} U.$$

Тогда получим оценку

$$\begin{aligned} \|\tilde{T}A - \tilde{T}\hat{A}\| &= \left\| \begin{pmatrix} A_{11} - I & A_{12} - S(A_{22} - I) \\ A_{21} & 0 \end{pmatrix} \right\| \leq \\ &\leq \|\Phi_2\| \|(I - S_0)^{-1}\| + \|\Phi_3\| + \|S_0 A_{12}\| \leq \|\Phi_2\|/(1 - 2\|\Phi_1\|) + \|\Phi_3\| + \\ &+ \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|} \|\tilde{T}A\| \leq \delta_A = \frac{\|\Phi_2\| + \|\Phi_3\| + \|\Phi_1\| \|\tilde{T}A\|}{1 - 2\|\Phi_1\|}. \end{aligned} \quad (5.16)$$

При этом удовлетворяются уравнения

$$\tilde{T}\hat{A}(I - \hat{P}) - (I - \hat{P}) = 0, \quad (I - \hat{P})\tilde{T}\hat{A} - (I - \hat{P}) = 0.$$

Аналогично существует такая матрица $\hat{B}$, что

$$\tilde{T}\hat{B}\hat{P} - \hat{P} = 0, \quad \hat{P}\tilde{T}\hat{B} - \hat{P} = 0,$$

и выполнена оценка вида

$$\|\tilde{T}B - \tilde{T}\hat{B}\| \leq \delta_B = \frac{\|\Phi_4\| + \|\Phi_5\| + \|\Phi_1\| \|\tilde{T}B\|}{1 - 2\|\Phi_1\|}. \quad (5.17)$$

Перейдем к анализу уравнения

$$\tilde{P}\tilde{X}\tilde{P}^* + (I - \tilde{P})\tilde{X}(I - \tilde{P}^*) - \tilde{X} = \Phi_6.$$

Обозначим

$$\hat{X} = \hat{P}\tilde{X}\hat{P}^* + (I - \hat{P})\tilde{X}(I - \hat{P}^*), \quad (5.18)$$

Тогда, очевидно, справедливо тождество

$$\hat{P}\hat{X}\hat{P}^* + (I - \hat{P})\hat{X}(I - \hat{P}^*) - \hat{X} = 0.$$

Оценим $\|\tilde{X} - \hat{X}\|$. Так как

$$\begin{aligned} \tilde{X} - \hat{X} &= -(\tilde{P} - \hat{P})\tilde{X} - \tilde{X}(\tilde{P} - \hat{P})^* + 2(\tilde{P} - \hat{P})\tilde{X}(\tilde{P} - \hat{P})^* + \\ &+ 2(\tilde{P} - \hat{P})\hat{P}\tilde{X}\hat{P}^* + 2(\tilde{P} - \hat{P})(\tilde{X} - \hat{X})\hat{P}^* + 2\hat{P}(\tilde{X} - \hat{X})(\tilde{P} - \hat{P})^* + \\ &+ 2\hat{P}\hat{X}\hat{P}^*(\tilde{P} - \hat{P})^* - \Phi_6, \end{aligned}$$

то при $\tilde{X} > 0$

$$\begin{aligned} \|\tilde{X} - \hat{X}\| &\leq \|\Phi_6\| + \frac{2\|\Phi_1\|\|\tilde{X}\|}{1 - 2\|\Phi_1\|} + \frac{2\|\Phi_1\|^2\|\tilde{X}\|}{(1 - 2\|\Phi_1\|)^2} + \\ &+ \frac{4\|\Phi_1\|}{1 - 2\|\Phi_1\|} (\|\tilde{X}\| + \|\tilde{X} - \hat{X}\|) + \frac{4\|\Phi_1\|\|\tilde{X} - \hat{X}\|}{1 - 2\|\Phi_1\|} \left(\|\tilde{P}\| + \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|} \right). \end{aligned}$$

Следовательно, выполнена оценка

$$\begin{aligned} \|\tilde{X} - \hat{X}\| &\leq \left(\|\Phi_6\| + \frac{6\|\Phi_1\|}{1 - 2\|\Phi_1\|} \|\tilde{X}\| + \frac{2\|\Phi_1\|^2}{(1 - 2\|\Phi_1\|)^2} \|\tilde{X}\| \right) / \left[1 - \right. \\ &- \frac{4\|\Phi_1\|}{1 - 2\|\Phi_1\|} \left(1 + \|\tilde{P}\| + \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|} \right) \left. \right] \leq (\|\Phi_6\| + 6\|\Phi_1\|\|\tilde{X}\|) / \left[1 - \right. \\ &- 2\|\Phi_1\| - \frac{\|\Phi_1\|^3}{1 - 2\|\Phi_1\|} - \frac{4\|\Phi_1\|}{1 - 2\|\Phi_1\|} \left(1 + \|\tilde{P}\| + \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|} \right) \left. \right] \leq \delta_X = \\ &= (\|\Phi_6\| + 6\|\Phi_1\|\|\tilde{X}\|) / (1 - 8\|\Phi_1\| - 4\|\Phi_1\|\|\tilde{P}\|). \end{aligned} \quad (5.19)$$

Оценка (5.19) справедлива только при $8\|\Phi_1\| + 4\|\Phi_1\| \|\tilde{P}\| < 1$ и $\tilde{X} > 0$.

Преобразуем последнее уравнение системы (5.8) к виду

$$\begin{aligned} & \tilde{T}\hat{A}\tilde{X}(\tilde{T}\hat{A})^* - \tilde{T}\hat{B}\tilde{X}(\tilde{T}\hat{B})^* + \tilde{P} + \tilde{P}^* - I = \Phi_7 - (\tilde{P} - \hat{P}) - (\tilde{P} - \hat{P})^* - \\ & - (\tilde{T}A - \tilde{T}\hat{A})\tilde{X}(\tilde{T}A)^* - \tilde{T}A\tilde{X}(\tilde{T}A - \tilde{T}\hat{A})^* + (\tilde{T}A - \tilde{T}\hat{A})\tilde{X}(\tilde{T}A - \tilde{T}\hat{A})^* + \\ & + (\tilde{T}B - \tilde{T}\hat{B})\tilde{X}(\tilde{T}B)^* + \tilde{T}B\tilde{X}(\tilde{T}B - \tilde{T}\hat{B})^* - (\tilde{T}B - \tilde{T}\hat{B})\tilde{X}(\tilde{T}B - \tilde{T}\hat{B})^* = \Phi. \end{aligned}$$

Заметим, что

$$\tilde{T}\hat{A}\tilde{X}(\tilde{T}\hat{A})^* - \tilde{T}\hat{B}\tilde{X}(\tilde{T}\hat{B})^* = \tilde{T}\hat{A}\tilde{X}(\tilde{T}\hat{A})^* - \tilde{T}\hat{B}\tilde{X}(\tilde{T}\hat{B})^*.$$

Следовательно, выполнено уравнение

$$\tilde{T}\hat{A}\tilde{X}(\tilde{T}\hat{A})^* - \tilde{T}\hat{B}\tilde{X}(\tilde{T}\hat{B})^* = \Phi,$$

причем

$$\begin{aligned} \|\Phi\| & \leq \|\Phi_7\| + 2\|\tilde{P} - \hat{P}\| + 2\|\tilde{T}A\tilde{X}\|\|\tilde{T}A - \tilde{T}\hat{A}\| + 2\|\tilde{T}B\tilde{X}\|\|\tilde{T}B - \tilde{T}\hat{B}\| + \\ & + \|\tilde{X}\|(\|\tilde{T}A - \tilde{T}\hat{A}\|^2 + \|\tilde{T}B - \tilde{T}\hat{B}\|^2) \leq \delta = \|\Phi_7\| + 2\|\Phi_1\|/(1 - 2\|\Phi_1\|) + \\ & + \|\tilde{X}\|(\delta_A^2 + \delta_B^2) + 2\|\tilde{T}A\tilde{X}\|\delta_A + 2\|\tilde{T}B\tilde{X}\|\delta_B. \end{aligned} \quad (5.20)$$

Предположим, что выполнены ограничения

$$\begin{aligned} & \|\Phi_2\| + \|\Phi_4\| < 1, \quad 8\|\Phi_1\| + 4\|\Phi_1\| \|\tilde{P}\| < 1, \\ & \|\Phi_7\| + 2\|\Phi_1\|/(1 - 2\|\Phi_1\|) + \|\tilde{X}\| \left\{ \left(\frac{\|\Phi_2\| + \|\Phi_3\| + \|\Phi_1\|\|\tilde{T}A\|}{1 - 2\|\Phi_1\|} \right)^2 + \right. \\ & + \left(\frac{\|\Phi_4\| + \|\Phi_5\| + \|\Phi_1\|\|\tilde{T}B\|}{1 - 2\|\Phi_1\|} \right)^2 \Big\} + \|\tilde{T}A\tilde{X}\| \frac{\|\Phi_2\| + \|\Phi_3\| + \|\Phi_1\|\|\tilde{T}A\|}{1 - 2\|\Phi_1\|} + \\ & + \|\tilde{T}B\tilde{X}\| \frac{\|\Phi_4\| + \|\Phi_5\| + \|\Phi_1\|\|\tilde{T}B\|}{1 - 2\|\Phi_1\|} < 1, \\ & \tilde{X} > \left(\frac{\|\Phi_6\| + 6\|\Phi_1\|\|\tilde{X}\|}{1 - 8\|\Phi_1\| - 4\|\Phi_1\|\|\tilde{P}\|} \right) I. \end{aligned} \quad (5.21)$$

Тогда по теореме 5.3 пучок матриц $\lambda\hat{B} - \hat{A}$ регулярен относительно единичной окружности и имеет параллельные проекторы $\hat{P}$ и $I - \hat{P}$ на инвариантные подпространства, отвечающие собственным числам соответственно внутри и вне единичного круга. Более того, справедливо неравенство

$$\left\| \hat{X} - \sum_{j=-\infty}^{\infty} \hat{P}_j \hat{P}_j^* \right\| \leq \delta \left\| \sum_{j=-\infty}^{\infty} \hat{P}_j \hat{P}_j^* \right\|,$$

где δ определено в (5.20). Из этого неравенства следуют оценки

$$\begin{aligned} & \left\| \tilde{X} - \sum_{j=-\infty}^{\infty} \hat{P}_j \hat{P}_j^* \right\| \leq \delta_X + \delta \left\| \sum_{j=-\infty}^{\infty} \hat{P}_j \hat{P}_j^* \right\|, \\ & \frac{\|\tilde{X}\| - \delta_X}{1 + \delta} \leq \left\| \sum_{j=-\infty}^{\infty} \hat{P}_j \hat{P}_j^* \right\| \leq \frac{\|\tilde{X}\| + \delta_X}{1 - \delta}, \\ & \left\| \tilde{X} - \sum_{j=-\infty}^{\infty} \hat{P}_j \hat{P}_j^* \right\| \leq \frac{\delta_X + \delta \|\tilde{X}\|}{1 - \delta}. \end{aligned} \quad (5.22)$$

Подведем итоги. Результатом теории возмущений, разработанной в этом пункте, является построение пучка матриц $\lambda\hat{B} - \hat{A}$, который имеет параллельные проекторы $\hat{P}$ и $I - \hat{P}$ и обобщенную функцию Ляпунова $\hat{X}$, удовлетворяющие оценкам

$$\|\hat{P} - \tilde{P}\| \leq \|\Phi_1\|/(1 - 2\|\Phi_1\|),$$

$$\begin{aligned} \|(\widehat{P}_{-0} + \widehat{P}_{+1}) - \widetilde{T}A\| &\leq \delta_A = \frac{\|\Phi_2\| + \|\Phi_3\| + \|\Phi_1\| \|\widetilde{T}A\|}{1 - 2\|\Phi_1\|}, \\ \|(\widehat{P}_{+0} + \widehat{P}_{-1}) - \widetilde{T}B\| &\leq \delta_B = \frac{\|\Phi_4\| + \|\Phi_5\| + \|\Phi_1\| \|\widetilde{T}B\|}{1 - 2\|\Phi_1\|}, \end{aligned} \quad (5.23)$$

$$\lambda \widehat{B} - \widehat{A} = \lambda(\widehat{P}_{+0} + \widehat{P}_{-1}) - (\widehat{P}_{-0} + \widehat{P}_{+1}),$$

$$\left\| \sum_{j=-\infty}^{\infty} \widehat{P}_j \widehat{P}_j^* - \widetilde{X} \right\| \leq (\delta_X + \delta \|\widetilde{X}\|)/(1 - \delta),$$

где

$$\delta_X = \frac{\|\Phi_6\| + 6\|\Phi_1\| \|\widetilde{X}\|}{1 - 8\|\Phi_1\| - 4\|\Phi_1\| \|\widetilde{P}\|},$$

$$\delta = \|\Phi_7\| + 2\|\Phi_1\|/(1 - 2\|\Phi_1\|) + \|\widetilde{X}\|(\delta_A^2 + \delta_B^2) + 2\|\widetilde{T}A\widetilde{X}\|\delta_A + 2\|\widetilde{T}B\widetilde{X}\|\delta_B.$$

Эта теория возмущений уравнения Ляпунова предполагает справедливость ограничений (5.21).

5.3. Вариант теории возмущений спектральных характеристик пучка матриц

Пусть пучок матриц имеет следующее строение:

$$\lambda \widehat{B} - \widehat{A} = \lambda(\widehat{P}_{+0} + \widehat{P}_{-1}) - (\widehat{P}_{-0} + \widehat{P}_{+1}),$$

где $\widehat{P}_{+0}^2 = \widehat{P}_{+0}$, $\widehat{P}_{-0} = I - \widehat{P}_{+0}$, $\widehat{P}_{-1}\widehat{P}_{+0} = \widehat{P}_{+0}\widehat{P}_{-1} = \widehat{P}_{-0}\widehat{P}_{+1} = \widehat{P}_{+1}\widehat{P}_{-0} = 0$ и все собственные числа матрицы $\widehat{P}_{-1} + \widehat{P}_{+1}$ лежат внутри единичного круга. Определим матрицы $\widehat{P}_j = \widehat{P}_{+1}^j$ при $j > 1$ и $\widehat{P}_j = \widehat{P}_{-1}^{-j}$ при $j < -1$.

Обозначим $\widehat{X} = \sum_{j=-\infty}^{\infty} \widehat{P}_j \widehat{P}_j^*$, $\widehat{\omega}_0 = \|\widehat{X}\|$. Предположим, что пучок $\lambda B - A$ является возмущением пучка $\lambda \widehat{B} - \widehat{A}$, так что выполнены оценки

$$\|A - (\widehat{P}_{-0} + \widehat{P}_{+1})\| \leq \delta_A, \quad \|B - (\widehat{P}_{+0} + \widehat{P}_{-1})\| \leq \delta_B$$

с некоторыми малыми δ_A и δ_B . Пусть

$$\lambda B - A = T^{-1}[\lambda(P_{+0} + P_{-1}) - (P_{-0} + P_{+1})]$$

— канонический вид пучка $\lambda B - A$. Требуется оценить $\|\widehat{P}_{+0} - P_{+0}\|$, $\|\widehat{X} - X\|$, где $X = \sum_{j=-\infty}^{\infty} P_j P_j^*$.

Прежде всего, с помощью параметра $\widehat{\omega}_0$ можно получить оценку

$$\|\widehat{P}_j\| \leq \sqrt{\widehat{\omega}_0} \left(1 - \frac{1}{2\widehat{\omega}_0}\right)^{|j|} \leq \sqrt{\widehat{\omega}_0} e^{-\frac{|j|}{2\widehat{\omega}_0}}. \quad (5.24)$$

Доказательство этой оценки почти полностью повторяет доказательство теоремы 1.4. Используя (5.24), легко вывести неравенство

$$\begin{aligned} \|(\widehat{B} - \widehat{A}e^{i\varphi})^{-1}\| &= \left\| \sum_{j=-\infty}^{\infty} \widehat{P}_j e^{ij\varphi} \right\| \leq \\ &\leq 2 \sum_{j=0}^{\infty} \|\widehat{P}_j\|^2 \leq 2 \sqrt{\widehat{\omega}_0} \sum_{j=0}^{\infty} \left(1 - \frac{1}{2\widehat{\omega}_0}\right)^j \leq 4\widehat{\omega}_0^{3/2}. \end{aligned} \quad (5.25)$$

Разложим функцию $(B - e^{i\varphi}A)^{-1}$ в ряд Фурье: $(B - Ae^{i\varphi})^{-1} = \sum_{n=-\infty}^{\infty} G_n e^{in\varphi}$.

Тогда, очевидно, выполнены тождества

$$BG_n - AG_{n-1} = \delta_n I, \quad \widehat{B}\widehat{P}_n - \widehat{A}\widehat{P}_{n-1} = \delta_n I$$

с $\delta_n = 0$ при $n \neq 0$ и $\delta_0 = 1$. Эти тождества дают свертку

$$G_n = \widehat{P}_n - \sum_{k=-\infty}^{\infty} \widehat{P}_{n-k} [(B - \widehat{B}) G_k - (A - \widehat{A}) G_{k-1}], \quad (5.26)$$

которая позволяет оценить разность $G_n - \widehat{P}_n$ и $\left\| \sum_{n=-\infty}^{\infty} G_n G_n^* \right\|$:

$$\begin{aligned} \|G_n - \widehat{P}_n\| &\leq \sqrt{\widehat{\omega}_0}(\delta_A + \delta_B) \sqrt{\left\| \sum_{j=-\infty}^{\infty} G_j^* G_j \right\|} \leq \\ &\leq \sqrt{\widehat{\omega}_0}(\delta_A + \delta_B) \sqrt{N \left\| \sum_{j=-\infty}^{\infty} G_j G_j^* \right\|}, \\ \widehat{\omega}_0 [1 - (\delta_A + \delta_B) \|(B - e^{i\varphi} A)^{-1}\|]^2 &\leq \left\| \sum_{n=-\infty}^{\infty} G_n G_n^* \right\| \leq \\ &\leq \widehat{\omega}_0 [1 + (\delta_A + \delta_B) \|(B - e^{i\varphi} A)^{-1}\|]^2. \end{aligned}$$

Так как

$$\|(B - Ae^{i\varphi})^{-1}\| \leq \frac{4\widehat{\omega}_0^{3/2}}{1 - 4\widehat{\omega}_0^{3/2}(\delta_A + \delta_B)},$$

то

$$\begin{aligned} \|G_n - \widehat{P}_n\| &\leq \widehat{\omega}_0 \sqrt{N}(\delta_A + \delta_B) \left[1 + \frac{4\widehat{\omega}_0^{3/2}(\delta_A + \delta_B)}{1 - 4\widehat{\omega}_0^{3/2}(\delta_A + \delta_B)} \right] = \\ &= \frac{\widehat{\omega}_0(\delta_A + \delta_B) \sqrt{N}}{1 - 4\widehat{\omega}_0^{3/2}(\delta_A + \delta_B)}. \end{aligned} \quad (5.27)$$

Из оценок для $\|G_0 - \widehat{P}_{+0}\|$ и $\|G_{-1} - \widehat{P}_{-0}\|$ следует, что матрица $T = G_0 - G_{-1}$ удовлетворяет неравенству

$$\|T - I\| \leq \frac{2\widehat{\omega}_0(\delta_A + \delta_B) \sqrt{N}}{1 - 4\widehat{\omega}_0^{3/2}(\delta_A + \delta_B)}. \quad (5.28)$$

Таким образом,

$$\begin{aligned} \|P_{+0} - \widehat{P}_{+0}\| &\leq \|G_0 - \widehat{P}_{+0}\| + \|P_{+0} - \widehat{P}_{+0}\| \|T - I\| + \|\widehat{P}_{+0}\| \|T - I\| \leq \\ &\leq (\|G_0 - \widehat{P}_{+0}\| + \|\widehat{P}_{+0}\| \|T - I\|) / (1 - \|T - I\|) \leq \\ &\leq \frac{\widehat{\omega}_0(\delta_A + \delta_B) \sqrt{N}(1 + 2\sqrt{\widehat{\omega}_0})}{1 - 2\widehat{\omega}_0(\delta_A + \delta_B) \sqrt{N} - 4\widehat{\omega}_0^{3/2}(\delta_A + \delta_B)}. \end{aligned} \quad (5.29)$$

Обозначим $\omega_0 = \left\| \sum_{k=-\infty}^{\infty} P_k P_k^* \right\|$, тогда из (5.28) выведем неравенства

$$\begin{aligned} \omega_0 \left(1 - \frac{2\widehat{\omega}_0(\delta_A + \delta_B) \sqrt{N}}{1 - 4\widehat{\omega}_0^{3/2}(\delta_A + \delta_B)} \right)^2 &\leq \left\| \sum_{n=-\infty}^{\infty} G_n G_n^* \right\| \leq \\ &\leq \omega_0 \left(1 + \frac{2\widehat{\omega}_0(\delta_A + \delta_B) \sqrt{N}}{1 - 4\widehat{\omega}_0^{3/2}(\delta_A + \delta_B)} \right)^2. \end{aligned}$$

В итоге

$$\begin{aligned} & \widehat{\omega}_0 \left[\frac{1 - 8\widehat{\omega}_0^{3/2} (\delta_A + \delta_B)}{1 - 4\widehat{\omega}_0^{3/2} (\delta_A + \delta_B) + 2\widehat{\omega}_0 (\delta_A + \delta_B) \sqrt{N}} \right]^2 \leq \\ & \leq \omega_0 \leq \widehat{\omega}_0 \left[\frac{1}{1 - 4\widehat{\omega}_0^{3/2} (\delta_A + \delta_B) - 2\widehat{\omega}_0 (\delta_A + \delta_B) \sqrt{N}} \right]^2, \\ & \left| \frac{\omega_0 - \widehat{\omega}_0}{\widehat{\omega}_0} \right| \leq \frac{8\widehat{\omega}_0^{3/2} (\delta_A + \delta_B) + 4\widehat{\omega}_0 (\delta_A + \delta_B) \sqrt{N}}{1 - 8\widehat{\omega}_0^{3/2} (\delta_A + \delta_B) - 4\widehat{\omega}_0 (\delta_A + \delta_B) \sqrt{N}}. \end{aligned} \quad (5.30)$$

5.4. Еще один способ контроля точности вычисления спектральных характеристик пучка матриц

В 5.1—5.3 построена теория возмущений обобщенного уравнения Ляпунова (5.3). Эта теория позволяет проконтролировать точность вычисленных приближенных матриц P и X по семи невязкам алгебраических уравнений. Матрица X является приближенным значением к

$$\sum_{j=-\infty}^{\infty} P_j P_j^* = \frac{1}{2} (H + PP^* + (I - P)(I - P^*)).$$

В этом пункте мы построим другую систему контроля точности, применимую непосредственно к «почти» нормированному пучку $(\lambda B_0 - A_0)_{\text{маш}}$ и матрицам $P_{\text{маш}}, H_{\text{маш}}$.

Рассмотрим уравнения

$$\begin{aligned} P^2 - P &= 0, \quad HP^* - PH = 0, \\ AHA^* - BHB^* - A(I - P - P^*)A^* - B(I - P - P^*)B^* &= \Phi \end{aligned} \quad (5.31)$$

и соотношения

$$\text{rank}(AP, BP) \leq \text{rank } P, \quad \text{rank}[A(I - P), B(I - P)] \leq \text{rank}(I - P), \quad (5.32)$$

где (AP, BP) и $[A(I - P), B(I - P)]$ — матрицы размеров $N \times 2N$.

Теорема 5.4. Предположим, что $H = H^* > 0$,

$$\sqrt{\|\Phi\|} < \sigma_{\min}(A, B), \quad (5.33)$$

выполнены уравнения (5.31) и соотношения (5.32). Тогда $\det(B - e^{i\varphi}A) \neq 0$ при всех вещественных φ , а матрицы P и $I - P$ являются проекторами на инвариантные пространства пучка $\lambda B - A$, отвечающие собственным числам соответственно внутри и вне единичного круга. Более того,

$$\begin{aligned} & \left\| H - \frac{1}{2\pi} \int_0^{2\pi} (B - e^{i\varphi}A)^{-1} (AA^* + BB^*) (B^* - e^{-i\varphi}A^*)^{-1} d\varphi \right\| \leq \\ & \leq \frac{\|\Phi\|}{\sigma_{\min}^2(A, B)} \left\| \frac{1}{2\pi} \int_0^{2\pi} (B - e^{i\varphi}A)^{-1} (AA^* + BB^*) (B^* - e^{-i\varphi}A^*)^{-1} d\varphi \right\|. \end{aligned} \quad (5.34)$$

Доказательство. Сначала воспользуемся неравенствами (5.32). Пусть $\text{rank } P = r$, $\text{rank } I = N$. Выберем базис пространства, в котором матрица P диагональная:

$$P = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Так как $\text{rank}(AP, BP) \leq \text{rank } P$, то существуют такие матрицы S_1 размеров $N \times r$ и $r \times r$ -матрицы K_1, K_2 , что $AP = (S_1 K_1, 0)$, $BP = (S_1 K_2, 0)$, где 0 — нулевые матрицы подходящих размеров. Аналогично существу-

ют такие $N \times (N-r)$ -матрица S_2 и $(N-r) \times (N-r)$ -матрицы M_1, M_2 , что $A(I-P) = (0, S_2 M_1)$, $B(I-P) = (0, S_2 M_2)$. Таким образом,

$$A = S \begin{pmatrix} K_1 & 0 \\ 0 & M_1 \end{pmatrix}, \quad B = S \begin{pmatrix} K_2 & 0 \\ 0 & M_2 \end{pmatrix}$$

с $S = (S_1, S_2)$.

Теперь используем условие $\text{rank}(AB) = N$, которое следует из (5.33), чтобы убедиться в том, что матрица S невырожденная. Обратимся к системе (5.31). Второе уравнение этой системы означает, что в выбранном базисе

$$H = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix}.$$

Преобразуем третье уравнение с помощью матриц $S, K_1, K_2, M_1, M_2, H_1$ и H_2 :

$$\begin{aligned} & S \begin{pmatrix} K_1 H_1 K_1^* - K_2 H_1 K_2^* & 0 \\ 0 & M_1 H_2 M_1^* - M_2 H_2 M_2^* \end{pmatrix} S^* - \\ & - S \begin{pmatrix} -K_1 K_1^* - K_2 K_2^* & 0 \\ 0 & M_1 M_1^* + M_2 M_2^* \end{pmatrix} S^* = \Phi. \end{aligned}$$

Следовательно,

$$\begin{aligned} & \begin{pmatrix} K_1 H_1 K_1^* - K_2 H_1 K_2^* & 0 \\ 0 & M_2 H_2 M_2^* - M_1 H_2 M_1^* \end{pmatrix} = \\ & = -PS^{-1}(AA^* + BB^*)S^{-*}P^* - (I-P)S^{-1}(AA^* + BB^*)S^{-*}(I-P^*) + \\ & + PS^{-1}\Phi S^{-*}P^* - (I-P)S^{-1}\Phi S^{-*}(I-P^*). \end{aligned} \quad (5.35)$$

Потребуем, чтобы правая часть уравнения (5.35) была отрицательно определенной. Для этого достаточно выполнения неравенства

$$AA^* + BB^* > \|\Phi\|I,$$

которое удовлетворяется в силу условия (5.33) теоремы.

Теперь приступим к изучению двух уравнений

$$K_1 H_1 K_1^* - K_2 H_1 K_2^* = -C_1, \quad M_2 H_2 M_2^* - M_1 H_2 M_1^* = -C_2$$

с положительно определенными матрицами C_1 и C_2 . Прежде всего, отметим, что матрицы K_2 и M_1 невырожденные. Пусть от противного, к примеру, матрица K_2 вырождена. Тогда существует ненулевой вектор x , удовлетворяющий равенству $xK_2 = 0$. Следовательно, $xK_1 H_1 K_1^* x^* = -xC_1 x^* < 0$, что противоречит положительности матрицы H_1 . В таком случае матрицы $K_2^{-1}K_1$ и $M_1^{-1}M_2$ имеют все собственные числа в единичном круге, т. е. матрицы P и $I-P$ являются проекторами на инвариантные подпространства пучка $\lambda B - A$, отвечающие собственным числам соответственно внутри и вне единичного круга.

Чтобы вывести оценку (5.34), опять рассмотрим последнее уравнение системы (5.31). Пусть пучок $\lambda B - A$ имеет следующий канонический вид:

$$\lambda B - A = T^{-1}[\lambda(P_{+0} + P_{-1}) - (P_{-0} + P_{+1})].$$

Тогда нетрудно проверить, что матрица решения H может быть представлена в виде ряда

$$H = \sum_{j=-\infty}^{\infty} P_j T (AA^* + BB^*) T^* P_j^* + \sum_{j=-\infty}^{-1} P_j T \Phi T^* P_j^* - \sum_{j=+0}^{\infty} P_j T \Phi T^* P_j^*. \quad (5.36)$$

Действительно, подставив это выражение в последнее уравнение системы (5.31), получим тождество

$$\begin{aligned} AHA^* - BHB^* - A(I - P - P^*)A^* - B(I - P - P^*)B^* = \\ = T^{-1}P_{+0}T\Phi T^*P_{+0}^*T^{-*} + T^{-1}P_{-0}T\Phi T^*P_{-0}^*T^{-*}. \end{aligned}$$

Нужно показать, что

$$T^{-1}P_{+0}T\Phi T^*P_{+0}^*T^{-*} + T^{-1}P_{-0}T\Phi T^*P_{-0}^*T^{-*} = \Phi.$$

Воспользуемся для этого каноническим видом пучка $\lambda B - A$ и всеми тремя уравнениями системы (5.31):

$$\begin{aligned} T^{-1}P_{+0}T[AHA^* - BHB^* - A(I - P - P^*)A^* - \\ - B(I - P - P^*)B^*]T^*P_{+0}^*T^{-*} + T^{-1}P_{-0}T[AHA^* - BHB^* - \\ - A(I - P - P^*)A^* - B(I - P - P^*)B^*]T^*P_{-0}^*T^{-*} = T^{-1}[P_{+1}HP_{+1}^* - \\ - P_{+0}HP_{+0}^* - P_{+1}(I - P - P^*)P_{+1}^* - P_{+0}(I - P - P^*)P_{+0}^*]T^{-*} + \\ + T^{-1}[P_{-0}HP_{-0}^* - P_{-1}HP_{-1}^* - P_{-0}(I - P - P^*)P_{-0}^* - \\ - P_{-1}(I - P - P^*)P_{-1}^*]T^{-*} = T^{-1}[(P_{-0} + P_{+1})H(P_{-0} + P_{+1})^* - \\ - (P_{+0} + P_{-1})H(P_{+0} + P_{-1})^* - (P_{-0} + P_{+1})(I - P - P^*)(P_{-0} + P_{+1})^* - \\ - (P_{+0} + P_{-1})(I - P - P^*)(P_{+0} + P_{-1})^*]T^{-*} = AHA^* - BHB^* - \\ - A(I - P - P^*)A^* - B(I - P - P^*)B^* = \Phi. \end{aligned}$$

Формула (5.36) доказана, следовательно,

$$\begin{aligned} \left\| H - \sum_{j=-\infty}^{\infty} P_j T(AA^* + BB^*)T^*P_j^* \right\| \leq \\ \leq \left\| \sum_{j=-\infty}^{-0} P_j T\Phi T^*P_j^* - \sum_{j=+0}^{\infty} P_j T\Phi T^*P_j^* \right\| \leq \\ \leq \frac{\|\Phi\|}{\sigma_{\min}^2(A, B)} \left\| \sum_{j=-\infty}^{\infty} P_j T(AA^* + BB^*)T^*P_j^* \right\|. \end{aligned}$$

Теорема доказана.

Предположим, что система (5.31) и соотношения (5.32) выполнены приближенно. Точнее, имеется система матричных уравнений

$$\tilde{P}^2 - \tilde{P} = \Phi_1, \quad (5.37)$$

$$\tilde{H}\tilde{P}^* - \tilde{P}\tilde{H} = \Phi_2, \quad (5.38)$$

$$A\tilde{H}A^* - B\tilde{H}B^* - A(I - \tilde{P} - \tilde{P}^*)A^* - B(I - \tilde{P} - \tilde{P}^*)B^* = \Phi_3, \quad (5.39)$$

а также условия

$$|\text{trace}(\tilde{P}) - r| \leq 1 - \frac{N\|\Phi_1\|}{1 - 2\|\Phi_1\|}, \quad (5.40)$$

$$\sigma_{N-r}(AP, BP) = \varphi, \quad (5.41)$$

$$\sigma_r[A(I - P), B(I - P)] = \psi, \quad (5.42)$$

где r — целое число, $0 \leq r \leq N$, а $\sigma_1 \leq \sigma_2 \leq \dots \leq \sigma_N$ — сингулярные числа, упорядоченные по возрастанию. Величины $\|\Phi_1\|$, $\|\Phi_2\|$, $\|\Phi_3\|$, φ , ψ достаточно малы.

Построим теорию возмущений для системы (5.31) и неравенств (5.32), использующую теорему 5.4. Сначала исследуем уравнения (5.37) и (5.38). Пусть

$$\tilde{P} = U^* \begin{pmatrix} S_0 & S \\ 0 & S_1 \end{pmatrix} U, \quad U^*U = I,$$

— форма Шура матрицы $\hat{P}$, где

$$\max(\|S_0\|, \|S_1 - I\|) \leq \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|}$$

при $\|\Phi_1\| < 1/4$. Обозначим

$$\hat{P} = U^* \begin{pmatrix} 0 & S \\ 0 & I \end{pmatrix} U,$$

тогда $\hat{P}^2 - \hat{P} = 0$ и

$$\|\tilde{P} - \hat{P}\| \leq \|\Phi_1\|/(1 - 2\|\Phi_1\|). \quad (5.43)$$

Если

$$\tilde{H} = U^* \begin{pmatrix} H_1 & H_2 \\ H_2^* & H_3 \end{pmatrix} U,$$

то уравнение (5.38) эквивалентно следующему:

$$\begin{pmatrix} H_1 S_0^* + H_2 S^* - S_0 H_1 - S H_2^* & H_2 S_1^* - S_0 H_2 - S H_3 \\ H_2^* S_0^* + H_3 S^* - S_1 H_2^* & H_3 S_1^* - S_1 H_3 \end{pmatrix} = U \Phi_2 U^*.$$

Обозначим

$$\hat{H} = U^* \begin{pmatrix} H_1 & S H_3 \\ H_3 S^* & H_3 \end{pmatrix} U,$$

тогда $\hat{H} \hat{P}^* - \hat{P} \hat{H} = 0$ и

$$\|\tilde{H} - \hat{H}\| \leq \|\Phi_2\| + \frac{2\|\Phi_1\|}{1 - 2\|\Phi_1\|} \|\tilde{H}\|. \quad (5.44)$$

Теперь рассмотрим условия (5.44), (5.42). Обозначим $\Delta = (A\tilde{P}, B\tilde{P}) - (A\hat{P}, B\hat{P})$, тогда $\|\Delta\| \leq \|(A, B)\| \|\tilde{P} - \hat{P}\|$. С учетом этого выводим оценки

$$\begin{aligned} \sigma_{N-r}(A\hat{P}, B\hat{P}) &\leq \varphi + \|(AB)\| \|\hat{P} - \tilde{P}\|, \\ \sigma_r[A(I - \hat{P}), B(I - \hat{P})] &\leq \psi + \|(AB)\| \|\tilde{P} - \hat{P}\|. \end{aligned}$$

Пусть матрица $\hat{P}$ имеет форму Шура

$$\hat{P} = \begin{pmatrix} 0 & S \\ 0 & I \end{pmatrix},$$

что не умаляет общности рассуждений. Предположим, что в этом ортогональном базисе

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

Тогда, очевидно,

$$(A\hat{P}, B\hat{P}) = \begin{pmatrix} 0 & A_{11}S + A_{12} & 0 & B_{11}S + B_{12} \\ 0 & A_{21}S + A_{22} & 0 & B_{21}S + B_{22} \end{pmatrix}.$$

Пусть

$$\begin{pmatrix} A_{11}S + A_{12} & B_{11}S + B_{12} \\ A_{21}S + A_{22} & B_{21}S + B_{22} \end{pmatrix} = V_1 \begin{pmatrix} \sigma_N & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \end{pmatrix} V_2$$

— сингулярное разложение. Из матрицы

$$\begin{pmatrix} A_{12} & B_{12} \\ A_{22} & A_{22} \end{pmatrix}$$

вычтем матрицу

$$\Delta_1 = V_1 \left(\begin{array}{c|c} \begin{matrix} -\sigma_{N-r} & \\ & \ddots \\ & & -\sigma_{N-r} \end{matrix} & \\ \hline & \begin{matrix} \sigma_{N-r} & \\ & \ddots \\ & & \sigma_{N-r} \end{matrix} \end{array} \right) V_2.$$

$\begin{matrix} N & N-r \end{matrix}$

В результате этого матрица $(\widehat{A}\widehat{P}, \widehat{B}\widehat{P}) - \Delta_1$ будет иметь все сингулярные числа, меньшие σ_{N-r+1} , равными нулю, а $\sigma_N, \dots, \sigma_{N-r+1}$ увеличатся ровно на σ_{N-r} . Важно здесь заметить, что структура матрицы $(\widehat{A}\widehat{P}, \widehat{B}\widehat{P}) - \Delta_1$ будет иметь вид $(A_1\widehat{P}, B_1\widehat{P})$, где (A_1, B_1) — малое возмущение матрицы (A, B) .

Аналогичным способом можно возмутить матрицу $[A(I - \widehat{P}), B(I - \widehat{P})]$ так, чтобы сингулярные числа, меньшие σ_{r+1} , были равны нулю, а $\sigma_N, \dots, \sigma_{r+1}$ увеличились на σ_r . В результате получим матрицу $[A_2(I - \widehat{P}), B_2(I - \widehat{P})]$. Обозначим $(\widehat{A} \widehat{B}) = (A_1\widehat{P}, B_1\widehat{P}) + [A_2(I - \widehat{P}), B_2(I - \widehat{P})]$. Так как

$$\begin{aligned} \|(\widehat{A}\widehat{P}, \widehat{B}\widehat{P}) - (A\widehat{P}, B\widehat{P})\| &\leq \|\Delta_1\| \leq \varphi + \|(A \ B)\| \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|}, \\ \|[A(I - \widehat{P}), B(I - \widehat{P})] - [A(I - \widehat{P}), B(I - \widehat{P})]\| &\leq \psi + \|(A \ B)\| \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|}, \end{aligned}$$

то

$$\|(\widehat{A} \widehat{B}) - (A \ B)\| \leq \varphi + \psi + \|(A \ B)\| \frac{2\|\Phi_1\|}{1 - 2\|\Phi_1\|}. \quad (5.45)$$

Обозначим $\delta_{AB} = \varphi + \psi + \|(A \ B)\| \frac{2\|\Phi_1\|}{1 - 2\|\Phi_1\|}$. Легко показать, что

$$\text{rank}(\widehat{A}\widehat{P}, \widehat{B}\widehat{P}) \leq \text{rank} \widehat{P}, \quad \text{rank}[\widehat{A}(I - \widehat{P}), \widehat{B}(I - \widehat{P})] \leq \text{rank}(I - \widehat{P}).$$

Теперь рассмотрим уравнение (5.39). Перепишем его следующим образом:

$$\begin{aligned} &\widehat{A}\widehat{H}\widehat{A}^* - \widehat{B}\widehat{H}\widehat{B}^* - \widehat{A}(I - \widehat{P} - \widehat{P}^*)\widehat{A}^* - \widehat{B}(I - \widehat{P} - \widehat{P}^*)\widehat{B}^* = \\ &= \Phi_3 - \widehat{A}(\widetilde{H} - \widehat{H})\widehat{A}^* - (A - \widehat{A})\widetilde{H}A^* - A\widetilde{H}(A - \widehat{A})^* + \\ &+ (A - \widehat{A})\widetilde{H}(A - \widehat{A})^* + \widehat{B}(\widetilde{H} - \widehat{H})\widehat{B}^* + (B - \widehat{B})\widetilde{H}B^* + B\widetilde{H}(B - \widehat{B})^* - \\ &- (B - \widehat{B})\widetilde{H}(B - \widehat{B})^* - \widehat{A}(\widetilde{P} - \widehat{P})\widehat{A}^* - \widehat{A}(\widetilde{P} - \widehat{P})^*\widehat{A}^* + \\ &+ (A - \widehat{A})(I - \widetilde{P} - \widetilde{P}^*)B^* + A(I - \widetilde{P} - \widetilde{P}^*)(A - \widehat{A})^* - \\ &- (A - \widehat{A})(I - \widetilde{P} - \widetilde{P}^*)(A - \widehat{A})^* - \widehat{B}(\widetilde{P} - \widehat{P})\widehat{B}^* - \widehat{B}(\widetilde{P} - \widehat{P})^*\widehat{B}^* + \\ &+ (B - \widehat{B})(I - \widetilde{P} - \widetilde{P}^*)B^* + B(I - \widetilde{P} - \widetilde{P}^*)(B - \widehat{B})^* - \\ &- (B - \widehat{B})(I - \widetilde{P} - \widetilde{P}^*)(B - \widehat{B})^* = \Phi. \end{aligned}$$

Следовательно,

$$\begin{aligned}
& \|\Phi\| \leq \|\Phi_3\| + \|\widehat{A}\|^2 \|\widetilde{H} - \widehat{H}\| + \|\widehat{B}\|^2 \|\widetilde{H} - \widehat{H}\| + 2\|\widehat{A}\|^2 \|\widetilde{P} - \widehat{P}\| + \\
& + 2\|\widehat{B}\|^2 \|\widetilde{P} - \widehat{P}\| + \|\widetilde{H}\| (2\delta_{AB}\|A\| + \delta_{AB}^2 + 2\delta_{AB}\|B\| + \delta_{AB}^2) + \\
& + \|I - \widetilde{P} - \widetilde{P}^*\| (2\delta_{AB}\|A\| + \delta_{AB}^2 + 2\delta_{AB}\|B\| + \delta_{AB}^2) \leq \|\Phi_3\| + \\
& + (\|\widehat{A}\|^2 + \|\widehat{B}\|^2) (\|\widetilde{H} - \widehat{H}\| + 2\|\widetilde{P} - \widehat{P}\|) + \\
& + (2\delta_{AB}\|A\| + 2\delta_{AB}\|B\| + 2\delta_{AB}^2) (\|\widetilde{H}\| + \|I - \widetilde{P} - \widetilde{P}^*\|) \leq \|\Phi_3\| + \\
& + 2\|(\widehat{A} \ \widehat{B})\|^2 (\|\widetilde{H} - \widehat{H}\| + 2\|\widetilde{P} - \widehat{P}\|) + (2\delta_{AB}\|A\| + 2\delta_{AB}\|B\| + \\
& + 2\delta_{AB}^2) (\|\widetilde{H}\| + \|\widetilde{P}\| + 3\|\widetilde{P} - \widehat{P}\|) \leq \delta = \|\Phi_3\| + \\
& + 2(\|(A \ B)\| + \delta_{AB})^2 \left(\|\Phi_2\| + \frac{2\|\Phi_1\|}{1-2\|\Phi_1\|} (1 + \|\widetilde{H}\|) \right) + \\
& + 2\delta_{AB}(2\|(A \ B)\| + \delta_{AB}) \left(\|\widetilde{H}\| + \|\widetilde{P}\| + \frac{3\|\Phi_1\|}{1-2\|\Phi_1\|} \right). \quad (5.46)
\end{aligned}$$

По теореме 5.4, если выполнены ограничения

$$\begin{aligned}
& \widetilde{H} > \left(\|\Phi_2\| + \frac{2\|\Phi_1\|}{1-2\|\Phi_1\|} \|\widetilde{H}\| \right) I, \\
& \forall \delta \leq \sigma_{\min}(A \ B) - \delta_{AB}, \quad (5.47)
\end{aligned}$$

то $\det(\widehat{B} - e^{i\varphi}\widehat{A}) \neq 0$ при всех вещественных φ , и выполнена оценка

$$\begin{aligned}
& \left\| \widehat{H} - \frac{1}{2\pi} \int_0^{2\pi} (\widehat{B} - \widehat{A}e^{i\varphi})^{-1} (\widehat{A}\widehat{A}^* + \widehat{B}\widehat{B}^*) (\widehat{B}^* - e^{-i\varphi}\widehat{A}^*)^{-1} d\varphi \right\| \leq \\
& \leq \delta \left\| \frac{1}{2\pi} \int_0^{2\pi} (\widehat{B} - \widehat{A}e^{i\varphi})^{-1} (\widehat{A}\widehat{A}^* + \widehat{B}\widehat{B}^*) (\widehat{B}^* - e^{-i\varphi}\widehat{A}^*)^{-1} d\varphi \right\|. \quad (5.48)
\end{aligned}$$

Обозначим

$$\widehat{\widehat{H}} = \frac{1}{2\pi} \int_0^{2\pi} (\widehat{B} - \widehat{A}e^{i\varphi})^{-1} (\widehat{A}\widehat{A}^* + \widehat{B}\widehat{B}^*) (\widehat{B}^* - e^{-i\varphi}\widehat{A}^*)^{-1} d\varphi, \quad \widehat{\omega} = \|\widehat{\widehat{H}}\|.$$

Тогда

$$\begin{aligned}
& \|\widehat{\widehat{H}}\| \leq \frac{1}{1-\delta} \left(\|\Phi_2\| + \frac{\|\widetilde{H}\|}{1-2\|\Phi_1\|} \right), \\
& \|\widetilde{H} - \widehat{\widehat{H}}\| \leq \frac{1}{1-\delta} \left(\delta\|\widetilde{H}\| + \|\Phi_2\| + \frac{2\|\Phi_1\|}{1-2\|\Phi_1\|} \|\widetilde{H}\| \right).
\end{aligned}$$

Напомним, что

$$\|H - \widehat{\widehat{H}}\| \leq \frac{47\widehat{\omega}^2\delta_{AB}^\mu}{1-47\widehat{\omega}\delta_{AB}^\mu}, \quad \|P - \widehat{P}\| \leq \frac{80\widehat{\omega}^2\delta_{AB}^\mu}{1-70\widehat{\omega}^{3/2}\delta_{AB}^\mu},$$

где $\mu = \sigma_{\max}(\widehat{A}, \widehat{B})/\sigma_{\min}(\widehat{A}, \widehat{B})$. Следовательно,

$$\begin{aligned}
& \widehat{\omega} \leq \frac{1}{1-\delta} \left(\|\Phi_2\| + \frac{\|\widetilde{H}\|}{1-2\|\Phi_1\|} \right), \\
& \|H - \widetilde{H}\| \leq \frac{1}{1-\delta} \left(\delta\|\widetilde{H}\| + \|\Phi_2\| + \frac{2\|\Phi_1\|}{1-2\|\Phi_1\|} \|\widetilde{H}\| \right) + \\
& + \frac{47\widehat{\omega}^2\delta_{AB}(\|(A \ B)\| + \delta_{AB})}{\sigma_{\min}(A \ B) - \delta_{AB} - 47\widehat{\omega}\delta_{AB}(\|(A \ B)\| + \delta_{AB})}, \quad (5.49) \\
& \|P - \widetilde{P}\| \leq \frac{\|\Phi_1\|}{1-2\|\Phi_1\|} + \frac{80\widehat{\omega}^2\delta_{AB}(\|(A \ B)\| + \delta_{AB})}{\sigma_{\min}(A \ B) - \delta_{AB} - 70\widehat{\omega}^{3/2}\delta_{AB}(\|(A \ B)\| + \delta_{AB})}.
\end{aligned}$$

В заключение выведем оценки погрешностей округлений при вычислении невязок

$$\begin{aligned} & \tilde{P}^2 - \tilde{P}, \quad (A\tilde{P}, B\tilde{P}), \quad [A(I - \tilde{P}), B(I - \tilde{P})], \quad \tilde{H}\tilde{P}^* - \tilde{P}\tilde{H}, \\ & A\tilde{H}A^* - B\tilde{H}B^* - A(I - \tilde{P} - \tilde{P}^*)A^* - B(I - \tilde{P} - \tilde{P}^*)B^*. \end{aligned}$$

Рассмотрим вычисление $(\tilde{P}^2 - \tilde{P})_{\text{маш}}$:

$$\begin{aligned} & \|\tilde{P}_{\text{маш}}^2 - \tilde{P}^2\|_F \leq \varepsilon_i \|\tilde{P}^2\|_F + \varepsilon_{ii} \|\tilde{P}\|_F^2 + o_i N, \\ & \|(\tilde{P}^2 - \tilde{P})_{\text{маш}} - (\tilde{P}^2 - \tilde{P})\|_F \leq \varepsilon_1 \|\tilde{P}_{\text{маш}}^2 - \tilde{P}\|_F + \varepsilon_0 N + \|\tilde{P}_{\text{маш}}^2 - \tilde{P}^2\|_F \leq \\ & \leq \varepsilon_0 N + \varepsilon_1 \|\tilde{P}^2 - \tilde{P}\|_F + (1 + \varepsilon_1) \|\tilde{P}_{\text{маш}}^2 - \tilde{P}^2\|_F \leq \varepsilon_1 \|\tilde{P}^2 - \tilde{P}\|_F + \varepsilon_0 N + \\ & + (1 + \varepsilon_1) o_i N + (1 + \varepsilon_1) \varepsilon_i \|\tilde{P}^2\|_F + (1 + \varepsilon_1) \varepsilon_{ii} \|\tilde{P}\|_F^2 \leq \\ & \leq [\varepsilon_0 N + (1 + \varepsilon_1) o_i N + (1 + \varepsilon_1) \varepsilon_1 \|\tilde{P}\|_F + (1 + \varepsilon_1) \varepsilon_{ii} \|\tilde{P}\|_F^2 + \\ & + (\varepsilon_1 + (1 + \varepsilon_1) \varepsilon_i) \|(\tilde{P}^2 - \tilde{P})_{\text{маш}}\|_F] / (1 - \varepsilon_1 - (1 + \varepsilon_1) \varepsilon_i). \end{aligned}$$

Теперь исследуем вычисление матрицы $(A\tilde{P}, B\tilde{P})$:

$$\begin{aligned} & \|(A\tilde{P})_{\text{маш}} - A\tilde{P}\|_F \leq \varepsilon_i \|A\tilde{P}\|_F + \varepsilon_{ii} \|A\|_F \|\tilde{P}\|_F + o_i N, \\ & \|(B\tilde{P})_{\text{маш}} - B\tilde{P}\|_F \leq \varepsilon_i \|B\tilde{P}\|_F + \varepsilon_{ii} \|B\|_F \|\tilde{P}\|_F + o_i N, \\ & \|(A\tilde{P}, B\tilde{P})_{\text{маш}} - (A\tilde{P}, B\tilde{P})\|_F \leq \sqrt{3} \varepsilon_i \|(A\tilde{P}, B\tilde{P})\|_F + \\ & + \sqrt{3} \varepsilon_{ii} \|(AB)\|_F \|\tilde{P}\|_F + \sqrt{6} o_i N \leq \\ & \leq [\sqrt{3} \varepsilon_i \|(A\tilde{P}, B\tilde{P})_{\text{маш}}\|_F + \sqrt{3} \varepsilon_{ii} \|(AB)\|_F \|\tilde{P}\|_F + \sqrt{6} o_i N] / (1 - \sqrt{3} \varepsilon_i). \end{aligned}$$

Матрицу $[A(I - \tilde{P}), B(I - \tilde{P})]_{\text{маш}}$ будем вычислять по формуле $(A B) - (A\tilde{P}, B\tilde{P})$:

$$\begin{aligned} & \|[A(I - \tilde{P}), B(I - \tilde{P})]_{\text{маш}} - [A(I - \tilde{P}), B(I - \tilde{P})]\|_F \leq \\ & \leq \varepsilon_0 \sqrt{2} N + \varepsilon_1 \|(A B) - (A\tilde{P}, B\tilde{P})_{\text{маш}}\|_F + \|(A\tilde{P}, B\tilde{P})_{\text{маш}} - (A\tilde{P}, B\tilde{P})\|_F \leq \\ & \leq \varepsilon_0 \sqrt{2} N + \varepsilon_1 \|[A(I - \tilde{P}), B(I - \tilde{P})]\|_F + \\ & + (1 + \varepsilon_1) \|(A\tilde{P}, B\tilde{P})_{\text{маш}} - (A\tilde{P}, B\tilde{P})\|_F \leq \\ & \leq \varepsilon_0 \sqrt{2} N + \varepsilon_1 \|[A(I - \tilde{P}), B(I - \tilde{P})]\|_F + \\ & + \{\sqrt{3} \varepsilon_i \|(A\tilde{P}, B\tilde{P})_{\text{маш}}\|_F + \sqrt{6} o_i N + \sqrt{3} \varepsilon_{ii} \|(A B)\|_F \|\tilde{P}\|_F\} / (1 - \varepsilon_1 - \sqrt{3} \varepsilon_i) \leq \\ & \leq \{\varepsilon_0 \sqrt{2} N + \varepsilon_1 \|[A(I - \tilde{P}), B(I - \tilde{P})]_{\text{маш}}\|_F + \sqrt{3} \varepsilon_i \|(A\tilde{P}, B\tilde{P})_{\text{маш}}\|_F + \\ & + \sqrt{3} \varepsilon_{ii} \|(A B)\|_F \|\tilde{P}\|_F + \sqrt{6} o_i N\} / (1 - 2\varepsilon_1 - \sqrt{3} \varepsilon_i). \end{aligned}$$

Оценим погрешности округления при вычислении $(\tilde{H}\tilde{P}^* - \tilde{P}\tilde{H})_{\text{маш}}$:

$$\begin{aligned} & \|(\tilde{P}\tilde{H})_{\text{маш}} - \tilde{P}\tilde{H}\|_F \leq \varepsilon_i \|\tilde{P}\tilde{H}\|_F + \varepsilon_{ii} \|\tilde{P}\|_F \|\tilde{H}\|_F + o_i N, \\ & \|(\tilde{H}\tilde{P}^* - \tilde{P}\tilde{H})_{\text{маш}} - (\tilde{H}\tilde{P}^* - \tilde{P}\tilde{H})\|_F \leq \\ & \leq \varepsilon_1 \|(\tilde{H}\tilde{P}^*)_{\text{маш}} - (\tilde{P}\tilde{H})_{\text{маш}}\|_F + \varepsilon_0 N + 2 \|(\tilde{P}\tilde{H})_{\text{маш}} - \tilde{P}\tilde{H}\|_F \leq \\ & \leq \varepsilon_1 \|\tilde{H}\tilde{P}^* - \tilde{P}\tilde{H}\|_F + \varepsilon_0 N + 2(1 + \varepsilon_1) \|(\tilde{P}\tilde{H})_{\text{маш}} - \tilde{P}\tilde{H}\|_F \leq \\ & \leq \varepsilon_0 N + \varepsilon_1 \|\tilde{H}\tilde{P}^* - \tilde{P}\tilde{H}\|_F + 2(1 + \varepsilon_1) \varepsilon_i \|\tilde{P}\tilde{H}\|_F + \\ & + 2(1 + \varepsilon_1) \varepsilon_{ii} \|\tilde{P}\|_F \|\tilde{H}\|_F + 2(1 + \varepsilon_1) o_i N \leq \\ & \leq [\varepsilon_0 N + \varepsilon_1 \|(\tilde{H}\tilde{P}^* - \tilde{P}\tilde{H})_{\text{маш}}\|_F + 2\varepsilon_1 \|(\tilde{P}\tilde{H})_{\text{маш}}\|_F + \\ & + 2\varepsilon_{ii} \|\tilde{P}\|_F \|\tilde{H}\|_F + 2o_i N] / (1 - 2\varepsilon_1 - \varepsilon_i). \end{aligned}$$

Осталось исследовать ошибки округления в процессе вычисления невязки $A\tilde{H}A^* - B\tilde{H}B^* - A(I - \tilde{P} - \tilde{P}^*)A^* - B(I - \tilde{P} - \tilde{P}^*)B^*$:

$$\begin{aligned} & \|(A\tilde{H})_{\text{маш}} - A\tilde{H}\|_F \leq (\varepsilon_i + \varepsilon_{ii}) \|A\|_F \|\tilde{H}\|_F + o_i N, \\ & \|(A\tilde{H}A^*)_{\text{маш}} - A\tilde{H}A^*\|_F \leq (\varepsilon_i + \varepsilon_{ii}) \|(A\tilde{H})_{\text{маш}}\|_F \|A\|_F + o_i N + \\ & + \|(A\tilde{H})_{\text{маш}} - A\tilde{H}\|_F \|A\|_F \leq (\varepsilon_i + \varepsilon_{ii}) \|\tilde{H}\|_F \|A\|_F^2 + o_i N + \end{aligned}$$

$$\begin{aligned}
& + (1 + \varepsilon_i + \varepsilon_{ii}) \|A\|_F [(\varepsilon_i + \varepsilon_{ii}) \|A\|_F \|\tilde{H}\|_F + o_i N] \leq \\
& \leq (2 + \varepsilon_i + \varepsilon_{ii}) (\varepsilon_i + \varepsilon_{ii}) \|\tilde{H}\|_F \|A\|_F^2 + o_i N [1 + \|A\|_F (1 + \varepsilon_i + \varepsilon_{ii})], \\
& \|(B\tilde{H}B^*)_{\text{маш}} - B\tilde{H}B^*\|_F \leq (2 + \varepsilon_i + \varepsilon_{ii}) (\varepsilon_i + \varepsilon_{ii}) \|\tilde{H}\|_F \|B\|_F^2 + \\
& + o_i N [1 + \|B\|_F (1 + \varepsilon_i + \varepsilon_{ii})], \\
& \|(A\tilde{H}A^* - B\tilde{H}B^*)_{\text{маш}} - (A\tilde{H}A^* - B\tilde{H}B^*)\|_F \leq \\
& \leq \varepsilon_1 \|(A\tilde{H}A^*)_{\text{маш}} - (B\tilde{H}B^*)_{\text{маш}}\|_F + \varepsilon_0 N + \|(A\tilde{H}A^*)_{\text{маш}} - A\tilde{H}A^*\|_F + \\
& + \|(B\tilde{H}B^*)_{\text{маш}} - B\tilde{H}B^*\|_F \leq \varepsilon_1 \|A\tilde{H}A^* - B\tilde{H}B^*\|_F + \varepsilon_0 N + \\
& + (1 + \varepsilon_1) (2 + \varepsilon_i + \varepsilon_{ii}) (\varepsilon_i + \varepsilon_{ii}) \|\tilde{H}\|_F \|A\|_F^2 + \\
& + (1 + \varepsilon_1) (2 + \varepsilon_i + \varepsilon_{ii}) (\varepsilon_i + \varepsilon_{ii}) \|\tilde{H}\|_F \|B\|_F^2 + \\
& + (1 + \varepsilon_1) o_i N [2 + (\|A\|_F + \|B\|_F) (1 + \varepsilon_i + \varepsilon_{ii})] \leq \varepsilon_1 \|A\tilde{H}A^* - B\tilde{H}B^*\|_F + \\
& + \varepsilon_0 N + (1 + \varepsilon_1) (2 + \varepsilon_i + \varepsilon_{ii}) (\varepsilon_i + \varepsilon_{ii}) \|\tilde{H}\|_F \|(AB)\|_F^2 + \\
& + (1 + \varepsilon_1) o_i N [2 + 2\|(AB)\|_F (1 + \varepsilon_i + \varepsilon_{ii})], \\
& \|(\tilde{P} + \tilde{P}^*)_{\text{маш}} - (\tilde{P} + \tilde{P}^*)\|_F \leq \varepsilon_1 \|\tilde{P} + \tilde{P}^*\|_F + \varepsilon_0 N, \\
& \|(I - \tilde{P} - \tilde{P}^*)_{\text{маш}} - (I - \tilde{P} - \tilde{P}^*)\|_F \leq \varepsilon_1 \|I - (\tilde{P} + \tilde{P}^*)_{\text{маш}}\|_F + \\
& + \varepsilon_0 N + \|(\tilde{P} + \tilde{P}^*)_{\text{маш}} - (\tilde{P} + \tilde{P}^*)\|_F \leq \varepsilon_0 N + \varepsilon_1 \|I - \tilde{P} - \tilde{P}^*\|_F + \\
& + (1 + \varepsilon_1) \|(\tilde{P} + \tilde{P}^*)_{\text{маш}} - (\tilde{P} + \tilde{P}^*)\|_F \leq \varepsilon_1 \|I - \tilde{P} - \tilde{P}^*\|_F + \varepsilon_0 N + \\
& + (1 + \varepsilon_1) \varepsilon_1 \cdot 2\|\tilde{P}\|_F + (1 + \varepsilon_1) \varepsilon_0 N, \\
& \|(A(I - \tilde{P} - \tilde{P}^*)A^* + B(I - \tilde{P} - \tilde{P}^*)B^*)_{\text{маш}} - \\
& - (A(I - \tilde{P} - \tilde{P}^*)A^* + B(I - \tilde{P} - \tilde{P}^*)B^*)\|_F \leq \\
& \leq \varepsilon_1 \|(AB)\|_F^2 \|(I - \tilde{P} - \tilde{P}^*)_{\text{маш}}\|_F + \varepsilon_0 N + \\
& + (1 + \varepsilon_1) (2 + \varepsilon_i + \varepsilon_{ii}) (\varepsilon_i + \varepsilon_{ii}) \|(I - \tilde{P} - \tilde{P}^*)_{\text{маш}}\|_F \|(AB)\|_F^2 + \\
& + (1 + \varepsilon_1) o_i N [2 + 2\|(AB)\|_F (1 + \varepsilon_i + \varepsilon_{ii})] + \\
& + \|(AB)\|_F^2 \|(I - \tilde{P} - \tilde{P}^*)_{\text{маш}} - (I - \tilde{P} - \tilde{P}^*)\|_F \leq \\
& \leq \|(AB)\|_F^2 [(\varepsilon_1 + (1 + \varepsilon_1) (2 + \varepsilon_i + \varepsilon_{ii}) (\varepsilon_i + \varepsilon_{ii})) \|(I - \tilde{P} - \tilde{P}^*)_{\text{маш}}\|_F + \\
& + \varepsilon_1 \|(I - \tilde{P} - \tilde{P}^*)_{\text{маш}}\|_F / (1 - \varepsilon_1) + \varepsilon_0 N / (1 - \varepsilon_1) + \\
& + (1 + \varepsilon_1) \varepsilon_1 \cdot 2\|\tilde{P}\|_F / (1 - \varepsilon_1) + (1 + \varepsilon_1) \varepsilon_0 N / (1 - \varepsilon_1)] + \varepsilon_0 N + \\
& + (1 + \varepsilon_1) o_i N [2 + 2\|(AB)\|_F (1 + \varepsilon_i + \varepsilon_{ii})], \\
& \|\Phi_3 - \Phi_{3\text{маш}}\|_F \leq \varepsilon_1 \|\Phi_3\|_F + \varepsilon_0 N + (1 + \varepsilon_1) \|(\tilde{A}H\tilde{A}^* - B\tilde{H}B^*)_{\text{маш}} - \\
& - (A\tilde{H}A^* - B\tilde{H}B^*)\|_F + (1 + \varepsilon_1) \|(A(I - \tilde{P} - \tilde{P}^*)A^* + B(I - \tilde{P} - \tilde{P}^*)B^*)_{\text{маш}} - \\
& - (A(I - \tilde{P} - \tilde{P}^*)A^* + B(I - \tilde{P} - \tilde{P}^*)B^*)\|_F \leq \varepsilon_1 \|\Phi_3\|_F + \varepsilon_0 N + \\
& + (1 + \varepsilon_1) \varepsilon_1 \|\tilde{H}\|_F \|(A \ B)\|_F^2 + (1 + \varepsilon_1) \varepsilon_0 N + \\
& + (1 + \varepsilon_1)^2 (2 + \varepsilon_i + \varepsilon_{ii}) (\varepsilon_i + \varepsilon_{ii}) \|\tilde{H}\|_F \|(A \ B)\|_F^2 + \\
& + (1 + \varepsilon_1)^2 o_i N [2 + 2\|(A \ B)\|_F (1 + \varepsilon_i + \varepsilon_{ii})] + (1 + \varepsilon_1) \varepsilon_0 N + \\
& + (1 + \varepsilon_1)^2 o_i N [2 + 2\|(A \ B)\|_F (1 + \varepsilon_i + \varepsilon_{ii})] + \\
& + (1 + \varepsilon_1) \|(A \ B)\|_F^2 \{2\varepsilon_1 + (2 + \varepsilon_i + \varepsilon_{ii}) (\varepsilon_{ii} + \varepsilon_i)\} \|(I - \tilde{P} - \tilde{P}^*)_{\text{маш}}\|_F + \\
& + (2 + \varepsilon_1) \varepsilon_0 N + (1 + \varepsilon_1) 2\varepsilon_1 \|\tilde{P}\|_F / (1 - \varepsilon_1) \leq \varepsilon_1 \|\Phi_3\|_F + \\
& + \frac{3\varepsilon_0 N + 4o_i N + 4o_i N \|(A \ B)\|_F}{1 - 2\varepsilon_1 - \varepsilon_i - \varepsilon_{ii}} + \|(A \ B)\|_F^2 [\|\tilde{H}\|_F (\varepsilon_1 + 2\varepsilon_i + 2\varepsilon_{ii}) + \\
& + 2\varepsilon_1 \|\tilde{P}\|_F + 2\varepsilon_0 N + \|(I - \tilde{P} - \tilde{P}^*)_{\text{маш}}\|_F (2\varepsilon_1 + 2\varepsilon_i + 2\varepsilon_{ii})] / (1 - 3\varepsilon_1 - \varepsilon_i - \\
& - \varepsilon_{ii}) \leq \{\varepsilon_1 \|\Phi_{3\text{маш}}\|_F + 3\varepsilon_0 N + 4o_i N + 4o_i N \|(A \ B)\|_F + \\
& + \|(A \ B)\|_F^2 [\|\tilde{H}\|_F (\varepsilon_1 + 2\varepsilon_i + 2\varepsilon_{ii}) + 2\varepsilon_0 N + 2\varepsilon_1 \|\tilde{P}\|_F + \\
& + \|(I - \tilde{P} - \tilde{P}^*)_{\text{маш}}\|_F (2\varepsilon_1 + 2\varepsilon_i + 2\varepsilon_{ii})]\} / (1 - 4\varepsilon_1 - \varepsilon_i - \varepsilon_{ii}).
\end{aligned}$$

§ 6. ЗАДАЧА ОБ ОТДЕЛИМОСТИ СПЕКТРА МАТРИЦЫ ОТ МНИМОЙ ОСИ

Разработанная методика вычисления спектральных характеристик регулярного относительно единичной окружности пучка матриц с успехом может быть распространена на задачу о дихотомии спектра матрицы мнимой осью. Обусловленность этой задачи оценивается параметром κ , предложенным С. К. Годуновым. Основные результаты п. 6.1 и 6.4 принадлежат С. К. Годунову и А. Я. Булгакову.

6.1. Спектральные характеристики матрицы, не имеющей собственных чисел на мнимой оси

Пусть A — матрица размеров $N \times N$, не имеющая точек спектра на мнимой оси. Гладкая всюду, кроме нуля, ограниченная матричная функция $G(t)$, удовлетворяющая дифференциальному уравнению

$$\frac{d}{dt} G(t) - AG(t) = \delta(t)I \quad (6.1)$$

и условию на разрыве

$$G(+0) - G(-0) = I, \quad (6.2)$$

называется *функцией Грина дифференциального оператора* $I \frac{d}{dt} - A$. Через $\delta(t)$ мы обозначили дельта-функцию с центром в нуле.

По теореме Жордана существует невырожденная матрица Q , которая приводит матрицу A к каноническому виду

$$Q^{-1}AQ = \text{block diag } \{J_+, J_-\}, \quad (6.3)$$

где матрица J_+ содержит все блоки Жордана с собственными числами в левой полуплоскости, а J_- — в правой полуплоскости.

Пользуясь (6.3), несложно показать, что функция Грина $G(t)$ единственна и задается формулами

$$G(t) = \begin{cases} Q \begin{pmatrix} e^{tJ_+} & 0 \\ 0 & 0 \end{pmatrix} Q^{-1}, & t \geq 0, \\ Q \begin{pmatrix} 0 & 0 \\ 0 & e^{-tJ_-} \end{pmatrix} Q^{-1}, & t \leq 0. \end{cases} \quad (6.4)$$

Отсюда следует, во-первых, что $G(+0)$ — проектор на инвариантное подпространство матрицы A , отвечающее собственным числам в левой полуплоскости, а $-G(-0)$ — проектор на инвариантное подпространство матрицы A , отвечающее собственным числам в правой полуплоскости. Во-вторых, выполнены следующие равенства:

$$\begin{aligned} G(t)G(s) &= G(t+s), & ts \geq 0, \\ G(t)G(s) &= 0, & ts \leq 0. \end{aligned} \quad (6.5)$$

Применим преобразование Фурье к уравнению (6.1):

$$(i\xi I - A)\widehat{G}(\xi) = I. \quad (6.6)$$

Равенство Парсеваля дает матрицу

$$H_A = \int_{-\infty}^{\infty} G^*(t)G(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} (i\xi I - A)^{-*} (i\xi I - A)^{-1} d\xi. \quad (6.7)$$

Назовем число

$$\kappa(A) = 2\|A\| \|H_A\| \quad (6.8)$$

параметром дихотомии матрицы A . Этот параметр является критерием отсутствия собственных чисел матрицы A на мнимой оси и вблизи от нее.

Параметр $\kappa(A)$ позволяет оценить величину $\sigma_{\min}(i\xi I - A)$. Для этого воспользуемся представлением

$$\frac{\kappa(A)}{2\|A\|} = \left\| \frac{1}{2\pi} \int_{-\infty}^{\infty} (i\xi I - A)^{-*} (i\xi I - A)^{-1} d\xi \right\|.$$

Выберем вектор y_0 единичной длины, удовлетворяющий для фиксированного ξ_0 равенству

$$\|(i\xi_0 I - A)^{-1} y_0\| = \|(i\xi_0 I - A)^{-1}\|.$$

Тогда получим неравенство

$$\frac{\kappa(A)}{2\|A\|} \geq \frac{1}{2\pi} \int_{-\infty}^{\infty} \|(i\xi I - A)^{-1} y_0\|^2 d\xi.$$

Далее,

$$\begin{aligned} \|(i\xi I - A)^{-1} y_0\| &= \|(i\xi_0 I - A)^{-1} y_0 - (i\xi I - A)^{-1} (i\xi - i\xi_0) (i\xi_0 I - A)^{-1} y_0\| \geq \\ &\geq \|(i\xi_0 I - A)^{-1}\| (1 - \|(i\xi I - A)^{-1}\| |\xi - \xi_0|) \geq \\ &\geq \|(i\xi_0 I - A)^{-1}\| \left(1 - |\xi - \xi_0| \frac{\|(i\xi_0 I - A)^{-1}\|}{1 - \|(i\xi_0 I - A)^{-1}\|} \right). \end{aligned}$$

Временно обозначим для уменьшения громоздкости выкладок $r = \|(i\xi_0 I - A)^{-1}\|$. Выведем следующие неравенства:

$$\begin{aligned} \frac{\pi\kappa}{\|A\|} &\geq \int_{|\xi - \xi_0| < \frac{1}{2r}} r^2 \left(1 - \frac{|\xi - \xi_0|^r}{1 - |\xi - \xi_0|^r} \right)^2 d\xi = 2 \int_0^{1/2r} r^2 \left(1 - \frac{tr}{1 - tr} \right)^2 dt = \\ &= 2r \int_0^{1/2} \left(\frac{1 - 2t}{1 - t} \right)^2 dt = 2r (3 - 4 \ln 2) > \frac{2}{5} r. \end{aligned}$$

Таким образом,

$$\|(i\xi_0 I - A)^{-1}\| < 5\pi \frac{\kappa(A)}{2\|A\|}. \quad (6.9)$$

В дальнейшем нам потребуются оценки непрерывности параметра $\kappa(A)$ и проекторов $G(+0)$, $-G(-0)$ от возмущений элементов матрицы A . Предположим, что $\tilde{A}$ — возмущение матрицы A , причем выполнена оценка $\|\tilde{A} - A\| \leq \delta \|A\|$ с некоторым достаточно малым δ . Введем следующие обозначения

$$\kappa(\tilde{A}) = 2\|\tilde{A}\| \left\| \frac{1}{2\pi} \int_{-\infty}^{\infty} (i\xi I - \tilde{A})^{-*} (i\xi I - \tilde{A})^{-1} d\xi \right\|,$$

$\tilde{G}(t)$ — функция Грина оператора $\Pi \frac{d}{dt} - \tilde{A}$. Требуется оценить $|\tilde{\kappa} - \kappa|$ и $\|\tilde{G}(+0) - G(+0)\|$.

Сначала разберем случай возмущения параметра κ . Так как

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} (i\xi I - \tilde{A})^{-*} (i\xi I - \tilde{A})^{-1} d\xi &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (i\xi I - A)^{-*} [I - \\ &- (i\xi I - \tilde{A})^{-1} (\tilde{A} - A)]^* [I - (i\xi I - \tilde{A})^{-1} (\tilde{A} - A)] (i\xi I - A)^{-1} d\xi, \end{aligned}$$

то

$$|\tilde{\kappa} - \kappa| \leq 2\|A\| \left\| \frac{1}{2\pi} \int_{-\infty}^{\infty} (i\xi I - A)^{-*} \Delta (i\xi I - A)^{-1} d\xi \right\|,$$

где

$$\Delta = [I - (i\xi I - \tilde{A})^{-1} (\tilde{A} - A)]^* \frac{\|\tilde{A}\|}{\|A\|} [I - (i\xi I - \tilde{A})^{-1} (\tilde{A} - A)] - I.$$

Следовательно,

$$\begin{aligned} \max_{\xi} (\|\tilde{A} - A\| \|(i\xi I - \tilde{A})^{-1}\|) &< \frac{5\pi\kappa\delta}{2 - 5\pi\kappa\delta}, \\ \|\Delta\| &\leq (1 + \delta) \left(1 + \max_{\xi} (\|\tilde{A} - A\| \|(i\xi I - \tilde{A})^{-1}\|)^2 - 1\right) \leq \\ &\leq \frac{1 + \delta}{1 - 5\pi\kappa\delta} - 1 = \frac{\delta(1 + 5\pi\kappa)}{1 - 5\pi\kappa\delta} < \frac{17\kappa\delta}{1 - 16\kappa\delta}, \\ \left| \frac{\tilde{\kappa} - \kappa}{\kappa} \right| &\leq \delta \frac{1 + 5\pi\kappa}{1 - 5\pi\kappa\delta} < \frac{17\kappa\delta}{1 - 16\kappa\delta}. \end{aligned} \quad (6.10)$$

Для оценки возмущений проектора $\tilde{G}(+0)$ воспользуемся тем, что эта матрица удовлетворяет уравнению

$$\frac{d}{dt} \tilde{G}(t) = A \tilde{G}(t) + (\tilde{A} - A) \tilde{G}(t) + \delta(t) I.$$

Так как в этом случае

$$\tilde{G}(t) = \int_{-\infty}^{\infty} G(t-s) [(\tilde{A} - A) \tilde{G}(s) + \delta(s)] ds,$$

то

$$\tilde{G}(t) - G(t) = \int_{-\infty}^{\infty} G(t-s) (\tilde{A} - A) \tilde{G}(s) ds.$$

Следовательно,

$$\begin{aligned} \|\tilde{G}(t) - G(t)\| &\leq \max_{\substack{\|x\|=1 \\ \|y\|=1}} \left| x \int_{-\infty}^{\infty} G(t-s) (\tilde{A} - A) \tilde{G}(s) ds y \right| \leq \\ &\leq \|\tilde{A} - A\| \max_{x,y} \int_{-\infty}^{\infty} \|x G(t-s)\| \|\tilde{G}(s) y\| ds \leq \|\tilde{A} - A\| \left(\left\| \int_{-\infty}^{\infty} G(t) G^*(t) dt \right\| + \right. \\ &+ \left. \left\| \int_{-\infty}^{\infty} \tilde{G}^*(s) \tilde{G}(s) ds \right\| \right)^{1/2} \leq \|\tilde{A} - A\| \cdot \frac{1}{2} \left(\frac{\sqrt{N}\kappa}{2\|A\|} + \frac{\tilde{\kappa}}{2\|\tilde{A}\|} \right) \leq \\ &\leq \frac{\delta}{4} \left(\sqrt{N}\kappa + \frac{\|A\|}{\|\tilde{A}\|} \tilde{\kappa} \right). \end{aligned}$$

Подобно оценке (6.10) можно вывести оценку

$$\begin{aligned} \|H_{\tilde{A}} - H_A\| &\leq \frac{\kappa}{2\|A\|} \left[\left(1 + \max_{\xi} (\|\tilde{A} - A\| \|(i\xi I - \tilde{A})^{-1}\|)^2 - 1\right) \right] < \\ &< \frac{\kappa}{2\|A\|} \left(\frac{1}{1 - 5\pi\kappa\delta} - 1 \right). \end{aligned}$$

В итоге

$$\|\tilde{G}(+0) - G(+0)\| \leq \frac{\delta}{4} \left(\sqrt{N}\kappa + \frac{\kappa}{1 - 5\pi\kappa\delta} \right) \leq \frac{\delta\kappa(1 + \sqrt{N})}{4(1 - 5\pi\kappa\delta)}. \quad (6.11)$$

Докажем фундаментальную теорему (аналог теоремы 1.4) об оценках матрицы Грина $G(t)$ с помощью $\kappa(A)$.

Теорема 6.1. Если $\det(i\xi I - A) \neq 0$ при любом вещественном ξ , то справедлива оценка

$$\|G(t)\| \leq \sqrt{\kappa} e^{-\frac{|t| \cdot \|A\|}{\kappa}}. \quad (6.12)$$

Доказательство. Сначала покажем справедливость двух важных тождеств

$$\begin{aligned} A^*G^*(+0)H_A G(+0) + G^*(+0)H_A G(+0)A &= -G^*(+0)G(+0), \\ A^*G^*(-0)H_A G(-0) + G^*(-0)H_A G(-0)A &= G^*(-0)G(-0). \end{aligned} \quad (6.13)$$

Рассмотрим, к примеру, последнее из них:

$$\begin{aligned} A^*G^*(-0)H_A G(-0) + G^*(-0)H_A G(-0)A &= \\ &= A^* \int_{-\infty}^{-0} G^*(t)G(t)dt + \int_{-\infty}^{-0} G^*(t)G(t)dtA = \\ &= \int_{-\infty}^{-0} \frac{d}{dt} [G^*(t)G(t)]dt = G^*(-0)G(-0). \end{aligned}$$

Далее, будем дифференцировать по t квадратичную функцию $(G^*(+0)H_A G(+0)G(t)x, G(t)x)$ при $t \geq +0$. В результате получим неравенство

$$\begin{aligned} \frac{d}{dt} (G^*(+0)H_A G(+0)G(t)x, G(t)x) &= \\ &= ([A^*G^*(+0)H_A G(+0) + G^*(+0)H_A G(+0)]AGx, Gx) = \\ &= -(G(t)x, G(t)x) \leq -\frac{1}{\|H_A\|} (G^*(+0)H_A G(+0)Gx, Gx). \end{aligned}$$

Следовательно, при $t \geq +0$

$$\begin{aligned} &(G^*(+0)H_A G(+0)G(t)x, G(t)x) \leq \\ &\leq e^{-\frac{t}{\|H_A\|}} (G^*(+0)H_A G(+0)G(+0)x, G(+0)x) \leq \\ &\leq e^{-\frac{t}{\|H_A\|}} \|G^*(+0)H_A G(+0)\| \|x\|^2 \leq \|H_A\| e^{-\frac{t}{\|H_A\|}} \|x\|^2. \end{aligned}$$

Так как при $t \geq +0$

$$G(+0) = e^{-tA}G(t), \quad \|G(t)x\| \geq e^{-t\|A\|}\|G(+0)x\|,$$

то

$$\begin{aligned} (G^*(+0)H_A G(+0)x, x) &= \int_0^\infty \|G(t)x\|^2 dt \geq \\ &\geq \int_0^\infty e^{-2t\|A\|} \|G(+0)x\|^2 dt = \frac{\|G(+0)x\|^2}{2\|A\|}. \end{aligned}$$

Таким образом, при $t \geq 0$

$$\begin{aligned} (G(t)x, G(t)x) &= (G(+0)G(t)x, G(+0)G(t)x) \leq \\ &\leq 2\|A\| (G^*(+0)H_A G(+0)G(t)x, G(t)x) \leq 2\|A\| \|H_A\| e^{-\frac{t}{\|H_A\|}} \|x\|^2, \\ \|G(t)\| &\leq \sqrt{\kappa} e^{-\frac{t\|A\|}{\kappa}}. \end{aligned}$$

Аналогично при $t \leq -0$

$$\|G(t)\| \leq \sqrt{\kappa} e^{-\frac{|t|\|A\|}{\kappa}}.$$

Теорема 6.1 доказана.

6.2. Модификация алгоритма

Алгоритм вычисления проекторов P_0 и P_∞ вместе с обобщенной функцией Ляпунова H для пучка матриц $\lambda B - A$ можно немного изменить так, что будут вычисляться проекторы $G(+0)$, $-G(-0)$ и обобщенная функция Ляпунова H_A для матрицы A , не имеющей точек спектра на мнимой оси. Модифицированный алгоритм также состоит из трех этапов.

Этап I. Вычисляем матричную экспоненту от матрицы $B = A^*/2\|A\|$. Очевидно, что инвариантные подпространства матрицы A^* , отвечающие точкам спектра слева (справа) от мнимой оси, совпадают с инвариантными подпространствами пучка $\lambda I - e^B$, отвечающими собственным числам внутри (вне) единичного круга.

Выберем матрицу L , удовлетворяющую уравнению

$$LL^* = \int_0^1 e^{tB} e^{tB^*} dt. \quad (6.14)$$

Определим пучок матриц

$$\lambda B_0 - A_0 = L^{-1}(\lambda I - e^B). \quad (6.15)$$

Этап II. Этот этап остается без изменения. Проводятся итерации перехода от пары матриц A_{m-1} , B_{m-1} к паре A_m , B_m , $m = 1, 2, \dots, m_0$. Для этого к матрице

$$\begin{pmatrix} -A_{m-1} & B_{m-1} & 0 \\ 0 & \boxed{-A_{m-1}} & B_{m-1} \end{pmatrix}$$

слева применим ортогональные преобразования, аннулирующие блок, обведенный рамкой. В результате получим матрицу

$$\begin{pmatrix} C_m & R_m & D_m \\ -A_m & 0 & B_m \end{pmatrix}.$$

Этап III. Вычислим матрицы

$$H_{m_0} = (A_{m_0} + B_{m_0})^{-1} (A_{m_0}^* + B_{m_0}^*)^{-1}, \quad G_{m_0} = [(A_{m_0} + B_{m_0})^{-1} B_{m_0}]^*. \quad (6.16)$$

При $m_0 \rightarrow \infty$ получим $H_{m_0} \rightarrow 2\|A\| H_A$, $G_{m_0} \rightarrow G(+0)$.

Рассмотрим подробно вопрос о сходимости предложенной модификации алгоритма. Несложно показать, пользуясь оценкой (6.12) и равенством

$$P_k = G^* \left(\frac{k}{2\|A\|} \right), \quad (6.17)$$

что неравенство (2.13) превращается в следующее:

$$\|(A_{m_0} + B_{m_0})^{-1} B_{m_0} - G^*(+0)\| \leq \frac{2\sqrt{\kappa} e^{-2^{m_0-1}/\kappa}}{1 - 2\sqrt{\kappa} e^{-2^{m_0-1}/\kappa}}. \quad (6.18)$$

Что касается предела матрицы H_{m_0} , то тут проведем более подробный анализ. Воспользуемся тем, что

$$\lambda I - e^B = T^{-1} [\lambda (P_{+0} + P_{-1}) - (P_{-0} + P_{+1})]$$

с $T = P_{+0} + P_{-1}$. Следовательно,

$$\lambda B_0 - A_0 = T_0^{-1} [(\lambda (P_{+0} + P_{-1}) - (P_{-0} + P_{+1}))]$$

с $T_0 = TL$. Так как в этом случае

$$\begin{aligned} T_0 T_0^* &= (P_{+0} + P_{-1}) LL^* (P_{+0} + P_{-1})^* = \\ &= [G^*(+0) + G^*(-0)e^{-B}] LL^* [G(+0) + G(-0)e^{-B^*}] = \\ &= G^*(+0) \int_0^1 e^{tB} e^{tB^*} dt G(+0) + G^*(-0) \int_{-1}^0 e^{tB} e^{tB^*} dt G(-0) + \\ &+ G^*(-0) \int_0^1 e^{(t-1)B} e^{tB^*} dt G(+0) + G^*(+0) \int_0^1 e^{tB} e^{(t-1)B^*} dt G(-0), \end{aligned}$$

то

$$\begin{aligned} \sum_{j=-\infty}^{\infty} P_j T_0 T_0^* P_j^* &= G^*(+0) \int_0^1 e^{tB} e^{tB^*} dt G(+0) + \\ &+ G^*(-0) \int_{-\infty}^0 e^{tB} e^{tB^*} dt G(-0) = 2\|A\| \int_{-\infty}^{\infty} G^*(t) G(t) dt. \end{aligned}$$

Из тождества (2.16)

$$T_k T_k^* = \sum_{j=0}^{2^k-1} (P_j + P_{j+1-2^k}) T_0 T_0^* (P_j^* + P_{j+1-2^k}^*)$$

следует, что

$$\begin{aligned} T_k T_k^* &= G^*(+0) \left(\int_0^{2^k} e^{tB} e^{tB^*} dt \right) G(+0) + G^*(-0) \left(\int_{-2^k}^0 e^{tB} e^{tB^*} dt \right) G(-0) + \\ &+ G^*(-0) \left(\int_0^{2^k} e^{(t-2^k)B} e^{tB^*} dt \right) G(+0) + G^*(+0) \left(\int_0^{2^k} e^{tB} e^{(t-2^k)B^*} dt \right) G(-0) = \\ &= 2\|A\| \left\{ \int_{-2^{k/2}\|A\|}^{2^{k/2}\|A\|} G^*(t) G(t) dt + \int_0^{2^{k/2}\|A\|} G^*(t - 2^k/2\|A\|) G(t) dt + \right. \\ &\quad \left. + \int_0^{2^{k/2}\|A\|} G^*(t) G(t - 2^k/2\|A\|) dt \right\}, \\ 2\|A\| H_A - T_k T_k^* &= 2\|A\| \{ G^*(2^k/2\|A\|) H_A G(2^k/2\|A\|) + \\ &+ G^*(-2^k/2\|A\|) H_A G(-2^k/2\|A\|) + \int_0^{2^{k/2}\|A\|} G^*(t - 2^k/2\|A\|) G(t) dt + \\ &+ \int_0^{2^{k/2}\|A\|} G^*(t) G(t - 2^k/2\|A\|) dt \}, \end{aligned}$$

$$\begin{aligned} \|2\|A\| H_A - T_k T_k^*\| &\leq 2\|A\| \left\{ 2\kappa^2 e^{-\frac{2^k}{\kappa}} / 2\|A\| + 2^{k+1} \kappa e^{-2^{k-1}/\kappa} / 2\|A\| \right\} = \\ &= 2\kappa^2 e^{-2^k/\kappa} + 2^{k+1} \kappa e^{-2^{k-1}/\kappa}. \end{aligned}$$

В итоге получим оценку скорости сходимости

$$\begin{aligned} \|2\|A\| H_A - (A_{m_0} + B_{m_0})^{-1} (A_{m_0}^* + B_{m_0}^*)^{-1}\| &\leq \\ &\leq \frac{2\kappa^2 e^{-2^{m_0}/\kappa} + 4\kappa^{3/2} e^{-2^{m_0-1}/\kappa} + 2^{m_0+1} \kappa e^{-2^{m_0-1}/\kappa}}{1 - 4\sqrt{\kappa} e^{-2^{m_0-1}/\kappa}}. \end{aligned} \quad (6.19)$$

В заключение остановимся на обосновании формулы

$$\int_0^1 e^{tB} C e^{tB^*} dt = \sum_{j=1}^{\infty} \frac{\mathcal{L}^j C}{j!}, \quad (6.20)$$

где $\mathcal{L}C \equiv BC + CB^*$ — оператор Ляпунова для матрицы B^* . Формула (6.20) используется при вычислении приближения к матрице $\int_0^1 e^{tB} e^{tB^*} dt$. Действительно, функция $Z(t) = e^{tB} C e^{tB^*}$ является решением задачи Коши

$$\dot{Z} = BZ + ZB^* = \mathcal{L}Z, \quad Z(0) = C.$$

Следовательно,

$$e^{tB} C e^{tB^*} = e^{t\mathcal{L}} C = \sum_{j=0}^{\infty} \frac{(t\mathcal{L})^j}{j!} C,$$

$$\int_0^1 e^{tB} C e^{tB^*} dt = \int_0^1 \left(\sum_{j=0}^{\infty} \frac{(t\mathcal{L})^j}{j!} C \right) dt = \sum_{j=0}^{\infty} \frac{\mathcal{L}^j C}{(j+1)!}.$$

Таким образом, формула (6.20) доказана.

6.3. Оценки вычислительных погрешностей модифицированного алгоритма

Вначале тщательно разберемся с первым этапом модифицированного алгоритма. Пусть $\|B_{\text{маш}} - B\| \leq \delta$ с малым δ , которое можно вычислить, пользуясь результатами из 3.3 и 3.5, относящимся к задаче о вычислении нормы матрицы A . Для вычисления экспоненты от матрицы B с нормой $1/2$ воспользуемся приближением частными суммами ряда Тейлора:

$$e^B = \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{B^k}{k!}.$$

Реализация этого метода на ЭВМ осуществляется по формулам

$$\begin{aligned} K_0 &= M_0 = I, \\ K_m &= K_{m-1} B / m, \\ M_m &= M_{m-1} + K_m, \quad m = 1, \dots, m_1. \end{aligned} \quad (6.21)$$

тогда $M_{m_1} \rightarrow e^B$ при $m_1 \rightarrow \infty$. Исследуем накопление ошибок в этом процессе. Начальная погрешность в матрице $B_{\text{маш}}$ влечет за собой погрешность

$$\|e^{B_{\text{маш}}} - e^B\| \leq e^{\|B\| + \delta} - e^{\|B\|} = e^{\|B\|} (e^\delta - 1) \leq e^{\|B\|} \frac{\delta}{1 - \delta}. \quad (6.22)$$

Реальный процесс вычислений по формулам (6.21) смоделируем следующим образом:

$$\begin{aligned} \tilde{K}_0 &= \tilde{M}_0 = I, \\ \tilde{K}_m &= \left(\frac{\tilde{K}_{m-1}}{m} \right) B_{\text{маш}} + \psi_m, \\ \tilde{M}_m &= \tilde{M}_{m-1} + \tilde{K}_m + \varphi_m, \quad 1 \leq m \leq m_1. \end{aligned}$$

Матрицы погрешностей ψ_m и φ_m оцениваются неравенствами

$$\begin{aligned} \|\psi_m\| &\leq d_1 \|B_{\text{маш}}\| \left\| \frac{\tilde{K}_{m-1}}{m} \right\| + d_2 \|B_{\text{маш}}\| + d_3, \\ d_1 &= \frac{\varepsilon_1 \sqrt{N} + \varepsilon_{ii} N + \varepsilon_1 \sqrt{N}}{1 - \varepsilon_1 \sqrt{N}}, \quad d_2 = \varepsilon_0 N (1 + \varepsilon_i \sqrt{N} + \varepsilon_{ii} N), \quad d_3 = \varepsilon_i N, \\ \|\varphi_m\| &\leq \varepsilon_1 \sqrt{N} (\|\tilde{M}_{m-1}\| + \|\tilde{K}_m\|) + \varepsilon_0 N. \end{aligned}$$

Выведем следующие оценки:

$$\begin{aligned} &\left\| \tilde{K}_m - \frac{B_{\text{маш}}^m}{m!} \right\| \leq \frac{\|B_{\text{маш}}\|}{m} \left\| \tilde{K}_{m-1} - \frac{B_{\text{маш}}^{m-1}}{(m-1)!} \right\| + \|\psi_m\| \leq \\ &\leq \frac{\|B_{\text{маш}}\|}{m} \left\| \tilde{K}_{m-1} - \frac{B_{\text{маш}}^{m-1}}{(m-1)!} \right\| + d_1 \frac{\|B_{\text{маш}}\|}{m} \left(\left\| \tilde{K}_{m-1} - \frac{B_{\text{маш}}^{m-1}}{(m-1)!} \right\| + \frac{\|B_{\text{маш}}\|^{m-1}}{(m-1)!} \right) + \\ &+ d_2 \|B_{\text{маш}}\| + d_3 \leq \frac{\|B_{\text{маш}}\|}{m} (1 + d_1) \left\| \tilde{K}_{m-1} - \frac{B_{\text{маш}}^{m-1}}{(m-1)!} \right\| + d_1 \frac{\|B_{\text{маш}}\|^m}{m!} + \\ &+ d_2 \|B_{\text{маш}}\| + d_3 \leq \frac{\|B_{\text{маш}}\|}{m} (1 + d_1) \left\| \tilde{K}_{m-1} - \frac{B_{\text{маш}}^{m-1}}{(m-1)!} \right\| + d_1 \frac{\|B_{\text{маш}}\|^m}{m!} + \\ &+ d_2 \|B_{\text{маш}}\| + d_3 \leq d_1 \frac{\|B_{\text{маш}}\|^m}{m!} \sum_{i=0}^{m-1} (1 + d_1)^i + \\ &+ \frac{d_2 \|B_{\text{маш}}\| + d_3}{m!} \sum_{i=0}^{m-1} (m-i)! \|B_{\text{маш}}\|^i (1 + d_1)^i \leq \\ &\leq d_1 \frac{\|B_{\text{маш}}\|^m}{m!} \frac{m}{1 - \frac{m-1}{2} d_1} + (d_2 \|B_{\text{маш}}\| + d_3) e^{(1+d_1) \|B_{\text{маш}}\|}. \end{aligned}$$

Далее,

$$\begin{aligned} \|\tilde{M}_m - M_m\| &\leq \|\tilde{M}_{m-1} - M_{m-1}\| + \|\tilde{K}_m - K_m\| + \|\varphi_m\| \leq \\ &\leq \|\tilde{M}_{m-1} - M_{m-1}\| + \|\tilde{K}_m - K_m\| + \varepsilon_0 N + \varepsilon_1 \sqrt{N} \times \\ &\times (\|\tilde{M}_{m-1} - M_{m-1}\| + \|\tilde{K}_m - K_m\| + \|M_{m-1}\| + \|K_m\|) = \\ &= (1 + \varepsilon_1 \sqrt{N}) \|\tilde{M}_{m-1} - M_{m-1}\| - (1 + \varepsilon_1 \sqrt{N}) \|\tilde{K}_m - K_m\| + \\ &+ \varepsilon_1 \sqrt{N} (\|M_{m-1}\| + \|K_m\|) + \varepsilon_0 N. \end{aligned}$$

Так как $\tilde{M}_0 - M_0 = 0$, то

$$\begin{aligned} \|\tilde{M}_m - M_m\| &\leq \sum_{k=1}^m (1 + \varepsilon_1 \sqrt{N})^{m+1-k} \|\tilde{K}_k - K_k\| + \\ &+ \varepsilon_1 \sqrt{N} e^{\|B_{\text{маш}}\|} \sum_{k=0}^{m-1} (1 + \varepsilon_1 \sqrt{N})^k + N \varepsilon_0 \sum_{k=0}^{m-1} (1 + \varepsilon_1 \sqrt{N})^k \leq \\ &\leq (1 + \varepsilon_1 \sqrt{N})^m \sum_{k=1}^m \|\tilde{K}_k - K_k\| + (\varepsilon_1 \sqrt{N} e^{\|B_{\text{маш}}\|} + N \varepsilon_0) \sum_{k=0}^{m-1} (1 + \varepsilon_1 \sqrt{N})^k \leq \\ &\leq \frac{1}{1 - m \varepsilon_1 \sqrt{N}} \frac{d_1 \|B_{\text{маш}}\|}{1 - \frac{m-1}{2} d_1} e^{\|B_{\text{маш}}\|} + \frac{m}{1 - m \varepsilon_1 \sqrt{N}} (d_2 \|B_{\text{маш}}\| + d_3) e^{(1+d_1) \|B_{\text{маш}}\|} + \\ &+ (\varepsilon_1 \sqrt{N} e^{\|B_{\text{маш}}\|} + N \varepsilon_0) \frac{m}{1 - \frac{m-1}{2} \varepsilon_1 \sqrt{N}} \leq \frac{d_1 \|B_{\text{маш}}\| e^{\|B_{\text{маш}}\|}}{1 - \varepsilon_1 m \sqrt{N} - \frac{m-1}{2} d_1} + \\ &+ \frac{m e^{(1+d_1) \|B_{\text{маш}}\|} (d_2 \|B_{\text{маш}}\| + d_3)}{1 - m \varepsilon_1 \sqrt{N}} + \frac{m (\varepsilon_1 \sqrt{N} e^{\|B_{\text{маш}}\|} + N \varepsilon_0)}{1 - \frac{m-1}{2} \varepsilon_1 \sqrt{N}}. \end{aligned}$$

Легко получить неравенство

$$\|e^{B_{\text{маш}}} - M_m\| < \frac{\|B_{\text{маш}}\|^{m+1}}{(m+1)!} e^{\|B_{\text{маш}}\|},$$

где M_m вычислено по формулам (6.21) с $B_{\text{маш}}$. В итоге, выполняется следующая оценка погрешностей вычислений по формулам (6.21):

$$\begin{aligned} \|\widetilde{M}_{m_1} - e^B\| &\leq e^{\|B\|} (e^\delta - 1) + e^{\|B_{\text{маш}}\|} \frac{\|B_{\text{маш}}\|^{m_1+1}}{(m_1+1)!} + \\ &+ \frac{d_1 \|B_{\text{маш}}\| e^{\|B_{\text{маш}}\|}}{1 - \varepsilon_1 \sqrt{N} m_1 - \frac{m_1-1}{2} d_1} + \frac{m_1 e^{(1+d_1)\|B_{\text{маш}}\|} (d_2 \|B_{\text{маш}}\| + d_3)}{1 - m_1 \varepsilon_1 \sqrt{N}} + \\ &+ \frac{m_1 (\varepsilon_1 \sqrt{N} e^{\|B_{\text{маш}}\|} + \varepsilon_0 N)}{1 - \frac{m_1-1}{2} \varepsilon_1 \sqrt{N}}. \end{aligned} \quad (6.23)$$

Для вычисления интеграла $\int_0^1 e^{tB} e^{tB^*} dt$ воспользуемся формулой

$$\int_0^1 e^{tB} e^{tB^*} dt = \lim_{m_2 \rightarrow \infty} \sum_{j=1}^{m_2} \frac{\mathcal{L}^j I}{j!},$$

где $\mathcal{L}C = BC + CB^*$. Предлагается следующий вычислительный процесс:

$$\begin{aligned} K_1 &= M_1 = I, \quad C = I, \\ K_m &= B \frac{K_{m-1}}{m} + \left(B \frac{K_{m-1}}{m} \right)^*, \\ M_m &= M_{m-1} + K_m, \quad 2 \leq m \leq m_2. \end{aligned} \quad (6.24)$$

Тогда $M_{m_2} \rightarrow \int_0^1 e^{tB} e^{tB^*} dt$ при $m_2 \rightarrow \infty$.

Исследуем накопление погрешностей в этом процессе. Начальная погрешность $\|B_{\text{маш}} - B\| \leq \delta$ приводит к погрешности в интеграле

$$\left\| \int_0^1 e^{tB_{\text{маш}}} C e^{tB_{\text{маш}}^*} dt - \int_0^1 e^{tB} C e^{tB^*} dt \right\| \leq \|C\| \frac{(e^{2\|B\|} - 1)(e^{2\delta} - 1)}{2\|B\|}. \quad (6.25)$$

Действительно

$$\begin{aligned} &\left\| \int_0^1 e^{tB_{\text{маш}}} C e^{tB_{\text{маш}}^*} dt - \int_0^1 e^{tB} C e^{tB^*} dt \right\| = \\ &= \left\| \int_0^1 e^{tB} [e^{t(B_{\text{маш}}-B)} C e^{t(B_{\text{маш}}-B)^*} - C] e^{tB^*} dt \right\| \leq \\ &\leq \|e^{t(B_{\text{маш}}-B)} C e^{t(B_{\text{маш}}-B)^*} - C\| \left\| \int_0^1 e^{tB} e^{tB^*} dt \right\| \leq \\ &\leq \|C\| (e^{2\delta} - 1) \left\| \int_0^1 e^{tB} e^{tB^*} dt \right\|. \end{aligned}$$

Так как $e^{-t\|B\|} \leq \|e^{tB}\| \leq e^{t\|B\|}$, то

$$1 - e^{-1} \leq \left\| \int_0^1 e^{tB} e^{tB^*} dt \right\| \leq \frac{e^{2\|B\|} - 1}{2\|B\|} = e - 1, \quad (6.26)$$

откуда следует требуемое неравенство (6.25).

Смоделируем вычисления по формулам (6.24) на ЭВМ следующим образом:

$$\begin{aligned}\tilde{K}_1 &= \tilde{M}_1 = C = I, \\ \tilde{K}_m &= B_{\text{маш}} \frac{\tilde{K}_{m-1}}{m} + \left(B_{\text{маш}} \frac{\tilde{K}_{m-1}}{m} \right)^* + \psi_m, \\ \tilde{M}_m &= \tilde{M}_{m-1} + \tilde{K}_m + \varphi_m, \quad 2 \leq m \leq m_2.\end{aligned}$$

Можно показать, что матрицы погрешностей ψ_m и φ_m удовлетворяют неравенствам

$$\begin{aligned}\|\psi_m\| &\leq 2\bar{d}_1 \frac{\|B_{\text{маш}}\|}{m} \|\tilde{K}_{m-1}\| + 2\bar{d}_2 \|B_{\text{маш}}\| + \bar{d}_3, \\ \bar{d}_1 &= \frac{2\varepsilon_1 \sqrt{N} + \varepsilon_i \sqrt{N} + \varepsilon_{ii} N}{1 - 2\varepsilon_1 \sqrt{N}}, \quad \bar{d}_2 = \varepsilon_0 N \frac{1 + \varepsilon_i \sqrt{N} + \varepsilon_{ii} N}{1 - \varepsilon_1 \sqrt{N}}, \quad \bar{d}_3 = \frac{2\varepsilon_i N + \varepsilon_0 N}{1 - \varepsilon_1 \sqrt{N}}, \\ \|\varphi_m\| &\leq \varepsilon_1 \sqrt{N} (\|\tilde{M}_{m-1}\| + \|\tilde{K}_m\|) + \varepsilon_0 N.\end{aligned}$$

Сначала выведем неравенство

$$\begin{aligned}\left\| \tilde{K}_m - \frac{\mathcal{L}^{m-1}C}{m!} \right\| &\leq \frac{2\|B_{\text{маш}}\|}{m} \left\| \tilde{K}_{m-1} - \frac{\mathcal{L}^{m-2}C}{(m-1)!} \right\| + \|\psi_m\| \leq \\ &\leq \frac{2\|B_{\text{маш}}\|}{m} \left\| \tilde{K}_{m-1} - \frac{\mathcal{L}^{m-2}C}{(m-1)!} \right\| + \\ &+ 2\bar{d}_1 \frac{2\|B_{\text{маш}}\|}{m} \left(\left\| \tilde{K}_{m-1} - \frac{\mathcal{L}^{m-2}C}{(m-1)!} \right\| + \left\| \frac{\mathcal{L}^{m-2}C}{(m-1)!} \right\| \right) + 2\bar{d}_2 \|B_{\text{маш}}\| + \bar{d}_3 \leq \\ &\leq \frac{2\|B_{\text{маш}}\|}{m} (1 + \bar{d}_1) \left\| \tilde{K}_{m-1} - \frac{\mathcal{L}^{m-2}C}{(m-1)!} \right\| + \bar{d}_1 \frac{\|C\|}{2\|B_{\text{маш}}\|} \frac{(2\|B_{\text{маш}}\|)^m}{m!} + \\ &+ 2\bar{d}_2 \|B_{\text{маш}}\| + \bar{d}_3 < \frac{(m-1)\bar{d}_1}{1 - \frac{m-2}{2}\bar{d}_1} \frac{\|C\|}{2\|B_{\text{маш}}\|} \frac{(2\|B_{\text{маш}}\|)^m}{m!} + \\ &+ \frac{2\bar{d}_2 \|B_{\text{маш}}\| + \bar{d}_3}{m!} \sum_{i=0}^{m-2} (m-i)! [2\|B_{\text{маш}}\|(1 + \bar{d}_1)]^i \leq \\ &\leq \frac{(m-1)\bar{d}_1}{1 - \frac{m-2}{2}\bar{d}_1} \frac{\|C\|}{2\|B_{\text{маш}}\|} \frac{(2\|B_{\text{маш}}\|)^m}{m!} + (2\bar{d}_2 \|B_{\text{маш}}\| + \bar{d}_3) e^{2\|B_{\text{маш}}\|(1 + \bar{d}_1)}.\end{aligned}$$

Далее,

$$\begin{aligned}\|\tilde{M}_m - M_m\| &\leq \|\tilde{M}_{m-1} - M_{m-1}\| + \|\tilde{K}_m - K_m\| + \|\varphi_m\| \leq \|\tilde{M}_{m-1} - M_{m-1}\| + \\ &+ \|\tilde{K}_m - K_m\| + \varepsilon_0 N + \varepsilon_1 \sqrt{N} (\|\tilde{M}_{m-1} - M_{m-1}\| + \|\tilde{K}_m - K_m\| + \\ &+ \|M_{m-1}\| + \|K_m\|) \leq (1 + \varepsilon_1 \sqrt{N}) \|\tilde{M}_{m-1} - M_{m-1}\| + \\ &+ (1 + \varepsilon_1 \sqrt{N}) \|\tilde{K}_m - K_m\| + \varepsilon_0 N + \varepsilon_1 \sqrt{N} (\|M_{m-1}\| + \|K_m\|) \leq \\ &\leq (1 + \varepsilon_1 \sqrt{N}) \|\tilde{M}_{m-1} - M_{m-1}\| + (1 + \varepsilon_1 \sqrt{N}) \|\tilde{K}_m - K_m\| + \\ &+ \varepsilon_1 \sqrt{N} \frac{\|C\|}{2\|B_{\text{маш}}\|} \sum_{k=1}^m \frac{(2\|B_{\text{маш}}\|)^k}{k!} + \varepsilon_0 N.\end{aligned}$$

Так как $\tilde{M}_1 - M_1 = 0$, то

$$\begin{aligned}\|\tilde{M}_m - M_m\| &\leq \sum_{k=2}^m (1 + \varepsilon_1 \sqrt{N})^{m+1-k} \|\tilde{K}_k - K_k\| + \\ &+ \left[\varepsilon_1 \sqrt{N} \frac{\|C\|}{2\|B_{\text{маш}}\|} (e^{2\|B_{\text{маш}}\|} - 1) + \varepsilon_0 N \right] \sum_{k=0}^{m-2} (1 + \varepsilon_1 \sqrt{N})^k \leq\end{aligned}$$

$$\leq \frac{1}{1 - (m-1) \varepsilon_1 \sqrt{N}} \left\{ \frac{(m-1) \bar{d}_1}{1 - \frac{m-2}{2} \bar{d}_1} \frac{\|C\|}{2 \|B_{\text{маш}}\|} (e^{2 \|B_{\text{маш}}\|} - 1) + \right. \\ \left. + (m-1) (\bar{d}_2 2 \|B_{\text{маш}}\| + \bar{d}_3) e^{2 \|B_{\text{маш}}\| (1 + \bar{d}_1)} \right\} + \\ + \frac{m-1}{1 - \frac{m-2}{2} \varepsilon_1 \sqrt{N}} \left[\varepsilon_1 \sqrt{N} \frac{\|C\|}{2 \|B_{\text{маш}}\|} (e^{2 \|B_{\text{маш}}\|} - 1) + \varepsilon_0 N \right].$$

Легко проверить, что

$$\left\| M_m - \sum_{h=1}^{\infty} \frac{\mathcal{L}^{h-1} C}{h!} \right\| \leq \frac{\|C\|}{2 \|B_{\text{маш}}\|} \frac{(2 \|B_{\text{маш}}\|)^{m+1}}{(m+1)!} e^{2 \|B_{\text{маш}}\|},$$

если в формулах (6.24) вместо матрицы B используется $B_{\text{маш}}$. В итоге

$$\left\| \int_0^1 e^{tB} C e^{tB^*} dt - \tilde{M}_{m_2} \right\| \leq \frac{\|C\|}{2 \|B_{\text{маш}}\|} \left\{ \frac{(2 \|B_{\text{маш}}\|)^{m_2+1}}{(m_2+1)!} e^{2 \|B_{\text{маш}}\|} + \right. \\ \left. + \frac{(m_2-1) \bar{d}_1 (e^{2 \|B_{\text{маш}}\|} - 1)}{1 - (m_2-1) \varepsilon_1 \sqrt{N} - \frac{m_2-2}{2} \bar{d}_1} + \frac{(m_2-1) \varepsilon_1 \sqrt{N} (e^{2 \|B_{\text{маш}}\|} - 1)}{1 - \frac{m_2-2}{2} \varepsilon_1 \sqrt{N}} \right\} + \\ + \|C\| (e - 1) (e^{2\delta} - 1) + \left\{ \frac{(m_2-1) \varepsilon_0 N}{1 - \frac{m_2-2}{2} \varepsilon_1 \sqrt{N}} + \right. \\ \left. + \frac{(m_2-1) (\bar{d}_2 2 \|B_{\text{маш}}\| + \bar{d}_3) e^{2 \|B_{\text{маш}}\| (1 + \bar{d}_1)}}{1 - (m_2-1) \varepsilon_1 \sqrt{N}} \right\}. \quad (6.27)$$

Для матрицы $\tilde{M}_{m_2}$ вычисляется разложение Холецкого и определяется нормированный пучок $(\lambda B_0 - A_0)_{\text{маш}}$. Можно выписать все оценки погрешностей вычисления этого пучка, используя оценки (6.23), (6.27) и методику оценки ошибок округления из § 3, касающуюся вычислений разложения Холецкого и обратной матрицы. Мы не будем здесь этого делать из-за громоздкости выкладок, которые в основном повторяют вы-

кладки из § 3. Отметим только, что обусловленность матрицы $\int_0^1 e^{tB} e^{tB^*} dt$ не превышает числа e . Это замечание относится к оценке погрешностей при вычислении разложения Холецкого этой матрицы.

Предположим, что

$$\|(A_0 \ B_0)_{\text{маш}} - (A_0 \ B_0)\| \leq \beta_0, \quad (6.28)$$

где $\lambda B_0 - A_0$ — некоторый нормированный пучок, который получается из исходного пучка $\lambda I - e^{A^*/2 \|A\|}$ при точных вычислениях, а β_0 — достаточно малая константа. Существование такого пучка $\lambda B_0 - A_0$ гарантируется предыдущим прямым анализом ошибок округления.

Повторяя рассуждения из 4.3 об оценках погрешностей второго этапа алгоритма, можно выписать следующую оценку для второго этапа модифицированного алгоритма:

$$\|(-\tilde{A}_{m_0} \ \tilde{B}_{m_0}) - (-A_{m_0} \ B_{m_0})\| \leq \\ \leq \|\Phi\| \frac{1 + \sqrt{2} \|(A_0 \ B_0)\|_{\max} \|(B_0 - e^{i\varphi} A_0)^{-1}\|}{1 - \|\Phi\|_{\max} \|(B_0 - e^{i\varphi} A_0)^{-1}\|}, \\ \|\Phi\| \leq \frac{\sqrt{2}}{1 - \alpha \sqrt{2} m_0} (\beta_0 + m_0 \alpha \sqrt{2} \|(A_0 \ B_0)\| + m_0 \beta),$$

где α и β — погрешности выполнения ортогональных преобразований:

$$\alpha = \frac{N^{3/2}\varepsilon_r}{1 - \frac{N-1}{2}\varepsilon_r}, \quad \beta = \frac{N^{3/2}\varepsilon_r}{1 - \frac{N-1}{2}\varepsilon_r},$$

ε_r и ε_r определены при $L = 2N$.

Оценим сверху $\|(A_0 B_0)\|$ и $\max_{\varphi} \|(B_0 - e^{i\varphi} A_0)^{-1}\|$. Так как $\lambda B_0 - A_0 = L^{-1}(\lambda I - e^B)$, то

$$\|B_0\| = \|L^{-1}\| \leq \sqrt{\frac{1}{1 - e^{-1}}} = \sqrt{\frac{e}{e-1}},$$

$$\|(A_0 B_0)\| \leq \sqrt{\|L^{-1}\|^2 (1 + e^{2\|B\|})} \leq \sqrt{\frac{e(e+1)}{e-1}} < 2,5.$$

Справедливо тождество

$$\frac{1}{2\pi} \int_0^{2\pi} (B_0 - A_0 e^{i\varphi})^{-1} (B_0^* - e^{-i\varphi} A_0^*)^{-1} d\varphi = 2\|A\|H_A. \quad (6.29)$$

Из этого тождества можно, рассуждая аналогично доказательству теоремы 1.3, получить неравенство

$$\max_{\varphi} \|(B_0 - e^{i\varphi} A_0)^{-1}\| < 5\pi\kappa \|B_0\|^2 \leq \frac{5\pi\kappa e}{e-1}.$$

Следовательно, после второго этапа модифицированного алгоритма

$$\|(A_{m_0} B_{m_0})_{\text{маш}} - (A_{m_0} B_{m_0})\| \leq \frac{128\kappa(A)(\beta_0 + 4\alpha m_0 + \beta m_0)}{1 - 38\kappa(A)(\beta_0 + 4\alpha m_0 + \beta m_0)}. \quad (6.30)$$

Третий этап модифицированного алгоритма исследуется аналогично 4.4.

6.4. Гарантированная оценка точности вычисленных спектральных характеристик

Анализ вычислительных погрешностей, проведенный в 6.3, обосновывает корректность модифицированного алгоритма, но не дает гарантий точности вычисленных матриц $2\|A\|H_A$ и $G(+0)$, обобщенной функции Ляпунова и проектора на инвариантное подпространство матрицы A , отвечающее собственным числам в левой полуплоскости. Это объясняется тем, что априори не известна эффективная оценка сверху величины параметра $\kappa(A)$. Предлагаемый ниже подход к получению гарантий точности разработан А. Я. Булгаковым и основан на обобщении матричного уравнения Ляпунова.

Рассмотрим следующую матричную алгебраическую систему:

$$\begin{aligned} G^2 - G &= 0, \\ AG - GA &= 0, \\ XG - G^*X &= 0, \\ A^*X + XA + G^*CG - (I - G)^*C(I - G) &= 0. \end{aligned} \quad (6.31)$$

Теорема 6.2. Если матрицы G и $X = X^* > 0$ удовлетворяют системе (6.31) с матрицей $C = C^* > 0$, то матрица A не имеет точек спектра на мнимой оси, а матрица $G(I - G)$ является проектором на инвариантное подпространство матрицы A , отвечающее собственным числам в левой (правой) полуплоскости.

Доказательство. Четвертое уравнение системы (6.31) умножим слева на $2G^*$, а справа — на G , и из полученного результата вычтем чет-

вертое уравнение:

$$0 = 2G^*A^*XG + 2G^*XAG + 2G^2*CG^2 - 2G^*(I - G^*)C(I - G)G - \\ - A^*X - XA - G^*CG + (I - G^*)C(I - G).$$

Преобразуем это уравнение с учетом первых трех уравнений системы (6.31):

$$0 = (2AG - A)^*X + X(2AG - A) + G^*CG + (I - G^*)C(I - G).$$

Так как

$$G^*CG + (I - G^*)C(I - G) \geq \frac{\lambda_{\min}(C)}{2} I > 0,$$

то по теореме Ляпунова матрица $2AG - A$ асимптотически устойчива, т. е. имеет все собственные значения в левой полуплоскости.

Выберем базис пространства $\mathbb{R}^N$, в котором матрица G диагональна:

$$G = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}.$$

Из второго уравнения системы (6.31) следует, что в этом базисе

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix},$$

т. е. G является проектором на инвариантное подпространство матрицы A . Этому подпространству отвечает подматрица A_1 . Легко также убедиться в том, что $I - G$ — также проектор на инвариантное подпространство матрицы A , отвечающее подматрице A_2 . В выбранном базисе

$$2AG - A = \begin{pmatrix} A_1 & 0 \\ 0 & -A_2 \end{pmatrix}.$$

Следовательно, G — проектор на инвариантное подпространство, отвечающее части спектра матрицы A в левой полуплоскости, а $I - G$ — в правой полуплоскости. Теорема 6.2 доказана.

Первое уравнение системы (6.31) означает, что G — некоторый проектор. Второе уравнение означает, что матрицы G и $I - G$ являются проекторами на инвариантные подпространства матрицы A . Третье уравнение разбивает квадратичную форму с матрицей X на клетки в соответствии с инвариантными подпространствами матрицы A , определенными проекторами G и $I - G$. И наконец, четвертое уравнение содержит в себе два уравнения Ляпунова: одно в подпространстве, определяемом проектором G , другое — в инвариантном подпространстве с проектором $I - G$.

В дальнейшем нам понадобится вариант предыдущей теоремы, в котором последнее уравнение системы (6.31) имеет вид

$$A^*X + XA + G + G^* - I = \Phi$$

с матрицей Φ малой нормы.

Теорема 6.3. Пусть матрицы G и $X = X^* > 0$ удовлетворяют системе уравнений

$$\begin{aligned} G^2 - G &= 0, \\ AG - GA &= 0, \\ XG - G^*X &= 0, \\ A^*X + XA + G + G^* - I &= \Phi \end{aligned} \tag{6.32}$$

с $\|\Phi\| < 1$. Тогда $\det(i\xi I - A) \neq 0$ при любом вещественном ξ , G — проектор на инвариантное подпространство матрицы A , отвечающее собственным числам в левой полуплоскости, а $I - G$ — проектор на инвариантное

подпространство для части спектра в правой полуплоскости. Более того,

$$\left\| \frac{1}{2\pi} \int_{-\infty}^{\infty} (i\xi I - A)^{-*} (i\xi I - A)^{-1} d\xi - X \right\| \leq \frac{\kappa(A)}{2\|A\|} \|\Phi\|. \quad (6.33)$$

Доказательство. Следуя доказательству предыдущей теоремы, необходимо показать, что

$$W = 2G^*(G + G^* - I - \Phi)G - (G^* + G - I - \Phi) > 0$$

при $\|\Phi\| < 1$. Так как

$$\begin{aligned} W &= G^*(I - \Phi)G + (I - G)^*(I + \Phi)(I - G) + G^*\Phi + \Phi G - 2G^*\Phi G, \\ G^*\Phi + \Phi G - 2G^*\Phi G &= G^*(A^*X + XA + G^* + G - I) + \\ &+ (A^*X + XA + G^* + G - I)G - 2G^*(A^*X + XA + G^* + G - I)G = 0, \end{aligned}$$

то

$$W = G^*(I - \Phi)G + (I - G)^*(I + \Phi)(I - G) \geq \frac{1 - \|\Phi\|}{2} I > 0.$$

Следовательно, $\det(i\xi I - A) \neq 0$ для любого вещественного ξ , а G и $I - G$ — соответствующие проекторы матрицы A .

Покажем справедливость оценки (6.33). Сначала отметим единственность решения системы (6.32), матриц G и X , когда $X = X^* > 0$, $\|\Phi\| < 1$. Это следует из того, что G — проектор на инвариантное подпространство матрицы A , отвечающее собственным числам в левой полуплоскости, а X — решение уравнения Ляпунова $\bar{A}^*X + X\bar{A} = -W$ с $\bar{A} = 2AG - A$. Матрица H_A удовлетворяет уравнению

$$\bar{A}^*H_A + H_A\bar{A} = -G^*G - (I - G)^*(I - G) = -W_0.$$

Следовательно,

$$X = \int_0^{\infty} e^{t\bar{A}^*} W e^{t\bar{A}} dt, \quad H_A = \int_0^{\infty} e^{t\bar{A}^*} W_0 e^{t\bar{A}} dt,$$

$$\|X - H_A\| \leq \max_{x \neq 0} \left| \frac{([W - W_0]x, x)}{(W_0x, x)} \right| \|H_A\|.$$

Так как $W - W_0 = -G^*\Phi G + (I - G^*)\Phi(I - G)$, то

$$|([W - W_0]x, x)| \leq \|\Phi\| (W_0x, x).$$

Таким образом, оценка (6.33) доказана. Доказательство теоремы завершено.

Предположим теперь, что система (6.32) выполняется приближенно, т. е. матрицы $\tilde{G}$ и $\tilde{X} = \tilde{X}^*$ удовлетворяют системе

$$\tilde{G}^2 - \tilde{G} = \Phi_1. \quad (6.34)$$

$$A\tilde{G} - \tilde{G}A = \Phi_2, \quad (6.35)$$

$$\tilde{X}\tilde{G} - \tilde{G}^*\tilde{X} = \Phi_3, \quad (6.36)$$

$$A^*\tilde{X} + \tilde{X}A + \tilde{G}^* + \tilde{G} - I = \Phi_4 \quad (6.37)$$

с матрицами $\Phi_1, \Phi_2, \Phi_3, \Phi_4$ малой нормы. Целью последующего изложения будет построение матриц $\hat{G}, \hat{A}, \hat{X}$, мало отличающихся от $\tilde{G}, A, \tilde{X}$ и удовлетворяющих теореме 6.3.

Как уже было показано ранее, при $\|\Phi_1\| < 1/4$ существует форма Шура матрицы $\tilde{G}$:

$$\tilde{G} = U^* \begin{pmatrix} S_1 & S \\ 0 & S_0 \end{pmatrix} U, \quad U^*U = I,$$

причем выполнены оценки

$$\max(\|S_0\|, \|S_1 - I\|) \leq \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|} \quad (6.38)$$

Обозначим

$$\widehat{G} = U^* \begin{pmatrix} I & S \\ 0 & 0 \end{pmatrix} U,$$

Тогда $\widehat{G}^2 - \widehat{G} = 0$ и

$$\|\widehat{G} - \widehat{G}\| \leq \|\Phi_1\| / (1 - 2\|\Phi_1\|). \quad (6.39)$$

Пусть

$$A = U^* \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} U.$$

Обозначим

$$\widehat{A} = U^* \begin{pmatrix} A_{11} & A_{11}S - SA_{22} \\ 0 & A_{22} \end{pmatrix} U,$$

Тогда $\widehat{A}\widehat{G} - \widehat{G}\widehat{A} = 0$,

$$\begin{aligned} \|\widetilde{A} - \widehat{A}\| &= \left\| \begin{pmatrix} 0 & A_{12} - A_{11}S + SA_{22} \\ A_{21} & 0 \end{pmatrix} \right\| = \\ &= \max(\|A_{21}\|, \|A_{12} - A_{11}S + SA_{22}\|). \end{aligned}$$

Уравнение (6.35) эквивалентно следующему:

$$\begin{pmatrix} A_{11}S_1 - S_1A_{11} - SA_{21} & A_{11}S + A_{12}S_0 - S_1A_{12} - SA_{22} \\ A_{21}S_1 - S_0A_{21} & A_{21}S + A_{22}S_0 - S_0A_{22} \end{pmatrix} = U\Phi_2U^*.$$

Поэтому

$$\begin{aligned} \|A_{21}\| &\leq \frac{\|\Phi_2\|}{1 - \|S_0\| - \|I - S_1\|} \leq \frac{\|\Phi_2\|}{1 - 4\|\Phi_1\|}, \\ \|A_{12} - A_{11}S + SA_{22}\| &\leq \|\Phi_2\| + \|A_{12}\|(\|S_0\| + \|S_1 - I\|) \leq \\ &\leq \|\Phi_2\| + \|A\| \frac{2\|\Phi_1\|}{1 - 2\|\Phi_1\|}, \\ \|\widetilde{A} - \widehat{A}\| &\leq \frac{\|\Phi_2\|}{1 - 4\|\Phi_1\|} + \|A\| \frac{2\|\Phi_1\|}{1 - 2\|\Phi_1\|} = \delta_A. \end{aligned} \quad (6.40)$$

Пусть

$$\widetilde{X} = U^* \begin{pmatrix} X_{11} & X_{12} \\ X_{12}^* & X_{22} \end{pmatrix} U.$$

Обозначим

$$\widehat{X} = U^* \begin{pmatrix} X_{11} & X_{11}S \\ S^*X_{11} & X_{22} \end{pmatrix} U,$$

Тогда

$$\widehat{X}\widehat{G} - \widehat{G}^*\widehat{X} = 0, \quad \|\widetilde{X} - \widehat{X}\| = \|X_{12} - X_{11}S\|.$$

Уравнение (6.36) перепишем в виде

$$\begin{pmatrix} X_{11}S_1 - S_1^*X_{11} & X_{11}S + X_{12}S_0 - S_1^*X_{12} \\ X_{12}^*S_1 - S^*X_{11} - S_0^*X_{12}^* & X_{12}^*S + X_{22}S_0 - S^*X_{12} - S_0^*X_{22} \end{pmatrix} = U\Phi_3U^*,$$

откуда

$$\begin{aligned} \|X_{12} - X_{11}S\| &\leq \| \Phi_3 \| + \|X_{12}\| (\|S_0\| + \|S_1 - I\|) \leq \| \Phi_3 \| + \\ &+ \frac{2\| \Phi_1 \|}{1 - 2\| \Phi_1 \|} \|X_{12}\| \leq \delta_X = \| \Phi_3 \| + \frac{2\| \Phi_1 \|}{1 - 2\| \Phi_1 \|} \| \tilde{X} \|. \end{aligned} \quad (6.41)$$

Будем предполагать, что $\tilde{X} > \delta_X I$.

Осталось рассмотреть уравнение (6.37). Перепишем его следующим образом:

$$\begin{aligned} \hat{A}^* \tilde{X} + \tilde{X} \hat{A} + \hat{G} + \hat{G}^* - I &= \Phi_4 - (\hat{G} - \tilde{G}) - (\hat{G} - \tilde{G})^* - A^* (\tilde{X} - \hat{X}) - \\ &- (A - \hat{A})^* \tilde{X} - (\tilde{X} - \hat{X}) A - \tilde{X} (A - \hat{A}) + (A - \hat{A})^* (\tilde{X} - \hat{X}) + \\ &+ (\tilde{X} - \hat{X}) (A - \hat{A}) = \Phi. \end{aligned}$$

Следовательно,

$$\begin{aligned} \| \Phi \| &\leq \| \Phi_4 \| + 2\| \tilde{G} - \hat{G} \| + 2\| A \| \| \tilde{X} - \hat{X} \| + 2\| \tilde{X} \| \| A - \hat{A} \| + \\ &+ 2\| A - \hat{A} \| \| \tilde{X} - \hat{X} \| \leq \| \Phi_4 \| + \frac{2\| \Phi_1 \|}{1 - 2\| \Phi_1 \|} + 2\| A \| \delta_X + 2\| \tilde{X} \| \delta_A + 2\delta_A \delta_X. \end{aligned} \quad (6.42)$$

По теореме 6.3, если

$$\delta_\Phi = \| \Phi_4 \| + \frac{2\| \Phi_1 \|}{1 - 2\| \Phi_1 \|} + 2\| A \| \delta_X + 2\| \tilde{X} \| \delta_A + 2\delta_A \delta_X < 1, \quad (6.43)$$

то $\det(i\xi I - \hat{A}) \neq 0$ при всех вещественных ξ , и $\| \tilde{X} - H_{\hat{A}} \| \leq \| \Phi \| \| H_{\hat{A}} \|$.
Итак, при ограничениях

$$\begin{aligned} \| \Phi_1 \| &< 1/4, \\ \delta_X &= \| \Phi_3 \| + \| \tilde{X} \| \frac{2\| \Phi_1 \|}{1 - 2\| \Phi_1 \|} < \lambda_{\min}(\tilde{X}), \\ \delta_\Phi &= \| \Phi_4 \| + \frac{2\| \Phi_1 \|}{1 - 2\| \Phi_1 \|} + 2\| A \| \delta_X + 2(\| \tilde{X} \| + \delta_X) \delta_A < 1, \end{aligned} \quad (6.44)$$

где

$$\delta_A = \frac{\| \Phi_2 \|}{1 - 4\| \Phi_1 \|} + \frac{2\| \Phi_1 \|}{1 - 2\| \Phi_1 \|} \| A \|,$$

выполнено неравенство $\| \tilde{X} - H_{\hat{A}} \| \leq \delta_\Phi \| H_{\hat{A}} \| + \delta_X$, а $\hat{A}$ — матрица близкая к A : $\| A - \hat{A} \| \leq \delta_A$. При этом

$$\| \tilde{G} - \hat{G} \| \leq \frac{\| \Phi_1 \|}{1 - 2\| \Phi_1 \|},$$

где $\hat{G}$ — проектор на инвариантное подпространство матрицы $\hat{A}$, отвечающее части спектра в левой полуплоскости. Легко видеть, что

$$\| H_{\hat{A}} \| \leq \frac{\| \tilde{X} \| + \delta_X}{1 - \delta_\Phi},$$

$$\| 2\| A \| \tilde{X} - 2\| \hat{A} \| H_{\hat{A}} \| \leq \frac{2\delta_A \| \tilde{X} \|}{1 - \delta_\Phi} + 2\| A \| \frac{\delta_\Phi}{1 - \delta_\Phi} \| \tilde{X} \| + 2\delta_X (\| A \| + \delta_A),$$

$$| \kappa(\hat{A}) - 2\| A \| \| \tilde{X} \| | \leq 2(\| A \| + \delta_A) \left(\frac{\| \tilde{X} \|}{1 - \delta_\Phi} + \delta_X \right) - 2\| A \| \| \tilde{X} \|,$$

$$\kappa(\hat{A}) \leq 2(\| A \| + \delta_A) \left(\frac{\| \tilde{X} \|}{1 - \delta_\Phi} + \delta_X \right).$$

В 6.1 были доказаны оценки

$$\|2\|A\|H_A - 2\|\widehat{A}\|H_{\widehat{A}}\| \leq \frac{17\kappa^2(\widehat{A})\delta_A}{1 - 16\kappa(\widehat{A})\delta_A},$$

$$\|G - \widehat{G}\| \leq \frac{\kappa(\widehat{A})(1 + \sqrt{N})\delta_A}{4(1 - 16\kappa(\widehat{A})\delta_A)}$$

при условии, что $16\kappa(\widehat{A})\delta_A < 1$. Следовательно,

$$\|G - \widetilde{G}\| \leq \frac{\|\Phi_1\|}{1 - 2\|\Phi_1\|} + \frac{\kappa(\widehat{A})(1 + \sqrt{N})\delta_A}{4(1 - 16\kappa(\widehat{A})\delta_A)}, \quad (6.45)$$

$$\begin{aligned} \|2\|A\|H_A - 2\|A\|\widetilde{X}\| &\leq 2(\|A\| + \delta_A) \left(\frac{\|\widetilde{X}\|}{1 - \delta_\Phi} + \delta_X \right) - 2\|A\|\|\widetilde{X}\| + \\ &+ \frac{17\kappa^2(\widehat{A})\delta_A}{1 - 16\kappa(\widehat{A})\delta_A} \leq \frac{\kappa(\widehat{A})}{1 - 17\delta_A\kappa(\widehat{A})} - 2\|A\|\|\widetilde{X}\|, \end{aligned} \quad (6.46)$$

где

$$\kappa(\widehat{A}) \leq 2(\|A\| + \delta_A) \left(\frac{\|\widetilde{X}\|}{1 - \delta_\Phi} + \delta_X \right). \quad (6.47)$$

При вычислении невязок (6.34)–(6.37) на ЭВМ требуется учет ошибок округления. Получим оценки этих ошибок, используя результаты § 3.

Рассмотрим вначале уравнение (6.34):

$$\begin{aligned} \|\widetilde{G}_{\text{маш}}^2 - \widetilde{G}^2\|_F &\leq \varepsilon_i \|\widetilde{G}^2\|_F + \varepsilon_{ii} \|\widetilde{G}\|_F^2 + o_i N, \\ \|(\widetilde{G}_{\text{маш}}^2 - \widetilde{G})_{\text{маш}} - (\widetilde{G}^2 - \widetilde{G})\|_F &\leq \varepsilon_0 N + \varepsilon_1 \|\widetilde{G}_{\text{маш}}^2 - \widetilde{G}\|_F + \|\widetilde{G}_{\text{маш}}^2 - \widetilde{G}^2\|_F \leq \\ &\leq \varepsilon_0 N + \varepsilon_1 \|\widetilde{G}^2 - \widetilde{G}\|_F + (1 + \varepsilon_1) \|\widetilde{G}_{\text{маш}}^2 - \widetilde{G}^2\|_F \leq \varepsilon_0 N + \varepsilon_1 \|\widetilde{G}^2 - \widetilde{G}\|_F + \\ &+ (1 + \varepsilon_1) [\varepsilon_i \|\widetilde{G}^2\|_F + \varepsilon_{ii} \|\widetilde{G}\|_F^2 + o_i N] \leq [\varepsilon_0 N + (1 + \varepsilon_1) o_i N] + \\ &+ [\varepsilon_1 + (1 + \varepsilon_1) \varepsilon_i] \|\widetilde{G}^2 - \widetilde{G}\|_F + (1 + \varepsilon_1) \varepsilon_i \|\widetilde{G}\|_F + (1 + \varepsilon_1) \varepsilon_{ii} \|\widetilde{G}\|_F^2. \end{aligned}$$

Следовательно,

$$\begin{aligned} \|\Phi_1\|_F &\leq [\|\Phi_{1\text{маш}}\|_F + \varepsilon_0 N + (1 + \varepsilon_1) o_i N + (1 + \varepsilon_1) \varepsilon_i \|\widetilde{G}\|_F + \\ &+ (1 + \varepsilon_1) \varepsilon_{ii} \|\widetilde{G}\|_F^2] / (1 - \varepsilon_1 - (1 + \varepsilon_1) \varepsilon_i). \end{aligned} \quad (6.48)$$

Теперь исследуем уравнение (6.35):

$$\begin{aligned} \|(A\widetilde{G})_{\text{маш}} - A\widetilde{G}\|_F &\leq \varepsilon_i \|A\widetilde{G}\|_F + \varepsilon_{ii} \|A\|_F \|\widetilde{G}\|_F + o_i N, \\ \|(\widetilde{G}A)_{\text{маш}} - \widetilde{G}A\|_F &\leq \varepsilon_i \|\widetilde{G}A\|_F + \varepsilon_{ii} \|A\|_F \|\widetilde{G}\|_F + o_i N, \\ \|[(A\widetilde{G})_{\text{маш}} - (\widetilde{G}A)_{\text{маш}}]_{\text{маш}} - (A\widetilde{G} - \widetilde{G}A)\|_F &\leq \varepsilon_1 \|(A\widetilde{G})_{\text{маш}} - (\widetilde{G}A)_{\text{маш}}\|_F + \varepsilon_0 N + \\ &+ \|(A\widetilde{G})_{\text{маш}} - A\widetilde{G}\|_F + \|(\widetilde{G}A)_{\text{маш}} - \widetilde{G}A\|_F \leq \varepsilon_0 N + \varepsilon_1 \|A\widetilde{G} - \widetilde{G}A\|_F + \\ &+ (1 + \varepsilon_1) \|(A\widetilde{G})_{\text{маш}} - A\widetilde{G}\|_F + (1 + \varepsilon_1) \|(\widetilde{G}A)_{\text{маш}} - \widetilde{G}A\|_F \leq \varepsilon_0 N + \\ &+ \varepsilon_1 \|A\widetilde{G} - \widetilde{G}A\|_F + (1 + \varepsilon_1) \varepsilon_i \|A\widetilde{G}\|_F + (1 + \varepsilon_1) \varepsilon_i \|\widetilde{G}A\|_F + 2(1 + \varepsilon_1) o_i N + \\ &+ 2(1 + \varepsilon_1) \varepsilon_{ii} \|A\|_F \|\widetilde{G}\|_F \leq [\varepsilon_0 N + 2(1 + \varepsilon_1) o_i N] + \varepsilon_1 \|A\widetilde{G} - \widetilde{G}A\|_F + \\ &+ 2(1 + \varepsilon_1) (\varepsilon_i + \varepsilon_{ii}) \|A\|_F \|\widetilde{G}\|_F. \end{aligned}$$

Следовательно,

$$\begin{aligned} \|\Phi_2\|_F &\leq [\|\Phi_{2\text{маш}}\|_F + [\varepsilon_0 N + 2(1 + \varepsilon_1) o_i N] + \\ &+ 2(1 + \varepsilon_1) (\varepsilon_i + \varepsilon_{ii}) \|A\|_F \|\widetilde{G}\|_F] / (1 - \varepsilon_1). \end{aligned} \quad (6.49)$$

Оценим ошибки округления при вычислении невязки Φ_3 в (6.36):

$$\begin{aligned}
& \|(\tilde{X}\tilde{G})_{\text{маш}} - \tilde{X}\tilde{G}\|_F \leq \varepsilon_i \|\tilde{X}\tilde{G}\|_F + \varepsilon_{ii} \|\tilde{X}\|_F \|\tilde{G}\|_F + o_i N, \\
& \|[(\tilde{X}\tilde{G})_{\text{маш}} - (\tilde{G}^* \tilde{X})_{\text{маш}}]_{\text{маш}} - (\tilde{X}\tilde{G} - \tilde{G}^* \tilde{X})\|_F \leq \varepsilon_1 \|(\tilde{X}\tilde{G})_{\text{маш}} - (\tilde{G}^* \tilde{X})_{\text{маш}}\|_F + \\
& + \varepsilon_0 N + \|(\tilde{X}\tilde{G})_{\text{маш}} - \tilde{X}\tilde{G}\|_F + \|(\tilde{G}^* \tilde{X})_{\text{маш}} - \tilde{G}^* \tilde{X}\|_F \leq \varepsilon_1 \|\tilde{X}\tilde{G} - \tilde{G}^* \tilde{X}\|_F + \\
& + \varepsilon_0 N + 2(1 + \varepsilon_1) \|(\tilde{X}\tilde{G})_{\text{маш}} - \tilde{X}\tilde{G}\|_F \leq \varepsilon_0 N + \varepsilon_1 \|\tilde{X}\tilde{G} - \tilde{G}^* \tilde{X}\|_F + \\
& + 2(1 + \varepsilon_1) \varepsilon_i \|\tilde{X}\tilde{G}\|_F + 2(1 + \varepsilon_1) \varepsilon_{ii} \|\tilde{X}\|_F \|\tilde{G}\|_F + 2(1 + \varepsilon_1) o_i N \leq \\
& \leq [\varepsilon_0 N + 2(1 + \varepsilon_1) o_i N] + \varepsilon_1 \|\tilde{X}\tilde{G} - \tilde{G}^* \tilde{X}\|_F + 2(1 + \varepsilon_1) \varepsilon_{ii} \|\tilde{X}\|_F \|\tilde{G}\|_F + \\
& + 2(1 + \varepsilon_1) [\varepsilon_i \|(\tilde{X}\tilde{G})_{\text{маш}}\|_F + \varepsilon_{ii} \|\tilde{X}\|_F \|\tilde{G}\|_F + o_i N] / (1 - \varepsilon_i) \leq \\
& \leq \left[\varepsilon_0 N + \frac{2(1 + \varepsilon_1) o_i N}{1 - \varepsilon_i} \right] + \varepsilon_1 \|\tilde{X}\tilde{G} - \tilde{G}^* \tilde{X}\|_F + \\
& + 2(1 + \varepsilon_1) \left[\varepsilon_{ii} \frac{\|\tilde{X}\|_F \|\tilde{G}\|_F}{1 - \varepsilon_i} + \varepsilon_i \frac{\|(\tilde{X}\tilde{G})_{\text{маш}}\|_F}{1 - \varepsilon_i} \right].
\end{aligned}$$

Поэтому

$$\begin{aligned}
\|\Phi_3\|_F & \leq \left\{ \|\Phi_{3\text{маш}}\|_F + \left[\varepsilon_0 N + \frac{2(1 + \varepsilon_1) o_i N}{1 - \varepsilon_i} \right] + 2(1 + \varepsilon_1) \varepsilon_{ii} \frac{\|\tilde{X}\|_F \|\tilde{G}\|_F}{1 - \varepsilon_i} + \right. \\
& \left. + 2(1 + \varepsilon_1) \varepsilon_i \frac{\|(\tilde{X}\tilde{G})_{\text{маш}}\|_F}{1 - \varepsilon_i} \right\} / (1 - \varepsilon_1). \quad (6.50)
\end{aligned}$$

Осталось рассмотреть уравнение (6.37):

$$\begin{aligned}
& \|(\tilde{X}A)_{\text{маш}} - \tilde{X}A\|_F \leq \varepsilon_i \|\tilde{X}A\|_F + \varepsilon_{ii} \|\tilde{X}\|_F \|A\|_F + o_i N, \\
& \|[(A^* \tilde{X})_{\text{маш}} + (\tilde{X}A)_{\text{маш}}]_{\text{маш}} - (A^* \tilde{X} + \tilde{X}A)\|_F \leq \varepsilon_0 N + 2\|(\tilde{X}A)_{\text{маш}} - \tilde{X}A\|_F + \\
& + \varepsilon_1 \|(A^* \tilde{X})_{\text{маш}} + (\tilde{X}A)_{\text{маш}}\|_F \leq \varepsilon_0 N + \varepsilon_1 \|A^* \tilde{X} + \tilde{X}A\|_F + \\
& + 2(1 + \varepsilon_1) \|(\tilde{X}A)_{\text{маш}} - \tilde{X}A\|_F \leq [\varepsilon_0 N + 2(1 + \varepsilon_1) o_i N] + \\
& + \varepsilon_1 \|A^* \tilde{X} + \tilde{X}A\|_F + 2(1 + \varepsilon_1) (\varepsilon_i + \varepsilon_{ii}) \|\tilde{X}\|_F \|A\|_F, \\
& \|(\tilde{G} + \tilde{G}^*)_{\text{маш}} - (\tilde{G} + \tilde{G}^*)\|_F \leq \varepsilon_1 \|\tilde{G} + \tilde{G}^*\|_F + \varepsilon_0 N, \\
& \|[(\tilde{G} + \tilde{G}^*)_{\text{маш}} - I]_{\text{маш}} - (\tilde{G} + \tilde{G}^* - I)\|_F \leq \varepsilon_1 \|(\tilde{G} + \tilde{G}^*)_{\text{маш}} - I\|_F + \\
& + \varepsilon_0 N + \|(\tilde{G} + \tilde{G}^*)_{\text{маш}} - (\tilde{G} + \tilde{G}^*)\|_F \leq \varepsilon_1 \|\tilde{G} + \tilde{G}^* - I\|_F + \varepsilon_0 N + \\
& + (1 + \varepsilon_1) \varepsilon_1 \|\tilde{G} + \tilde{G}^*\|_F + (1 + \varepsilon_1) \varepsilon_0 N, \\
& \|\Phi_{4\text{маш}} - \Phi_4\|_F \leq \varepsilon_1 \|[(A^* \tilde{X})_{\text{маш}} + (\tilde{X}A)_{\text{маш}}]_{\text{маш}} + [(\tilde{G} + \tilde{G}^*)_{\text{маш}} - I]_{\text{маш}}\|_F + \\
& + \varepsilon_0 N + \|[(A^* \tilde{X})_{\text{маш}} + (\tilde{X}A)_{\text{маш}}]_{\text{маш}} - (A^* \tilde{X} + \tilde{X}A)\|_F + \\
& + \|[(\tilde{G} + \tilde{G}^*)_{\text{маш}} - I]_{\text{маш}} - (\tilde{G} + \tilde{G}^* - I)\|_F \leq \varepsilon_1 \|A^* \tilde{X} + \tilde{X}A + \tilde{G} + \tilde{G}^* - I\|_F + \\
& + \varepsilon_0 N + (1 + \varepsilon_1) \|[(A^* \tilde{X})_{\text{маш}} + (\tilde{X}A)_{\text{маш}}]_{\text{маш}} - (A^* \tilde{X} + \tilde{X}A)\|_F + \\
& + (1 + \varepsilon_1) \|[(\tilde{G} + \tilde{G}^*)_{\text{маш}} - I]_{\text{маш}} - (\tilde{G} + \tilde{G}^* - I)\|_F \leq \varepsilon_1 \|\Phi_4\|_F + \varepsilon_0 N + \\
& + (1 + \varepsilon_1) [\varepsilon_0 N + 2(1 + \varepsilon_1) o_i N] + (1 + \varepsilon_1) \varepsilon_1 \|A^* \tilde{X} + \tilde{X}A\|_F + \\
& + (1 + \varepsilon_1) (2 + \varepsilon_1) \varepsilon_0 N + 2(1 + \varepsilon_1)^2 (\varepsilon_i + \varepsilon_{ii}) \|\tilde{X}\|_F \|A\|_F + \\
& + (1 + \varepsilon_1)^2 \varepsilon_1 \|\tilde{G} + \tilde{G}^*\|_F + (1 + \varepsilon_1) \|\tilde{G} + \tilde{G}^* - I\|_F \leq [\varepsilon_1 \|\Phi_4\|_F + 4\varepsilon_0 N + \\
& + 2o_i N + 2(\varepsilon_1 + \varepsilon_i + \varepsilon_{ii}) \|\tilde{X}\|_F \|A\|_F + \varepsilon_1 (1 + 4) \|\tilde{G}\|_F] / (1 - 2\varepsilon_1).
\end{aligned}$$

Следовательно,

$$\begin{aligned}
\|\Phi_4\|_F & \leq [\|\Phi_{4\text{маш}}\|_F + 4\varepsilon_0 N + 2o_i N + 2(\varepsilon_1 + \varepsilon_i + \varepsilon_{ii}) \|\tilde{X}\|_F \|A\|_F + \\
& + \varepsilon_1 (1 + 4\|\tilde{G}\|_F)] / (1 - 3\varepsilon_1). \quad (6.51)
\end{aligned}$$

Таким образом, получены оценки ошибок округлений при вычислении невязок в обобщенном уравнении Ляпунова (6.32).

6.5. Замечания к методу матричной сигнум-функции

Для вычисления инвариантных подпространств матрицы, отвечающих собственным числам в левой или правой полуплоскости, широко применяется метод матричной сигнум-функции. Расчетные формулы этого метода чрезвычайно просты:

$$A_0 = A, \quad A_k = \alpha_k A_{k-1} + (1 - \alpha_k) A_{k-1}^{-1}, \quad k = 1, 2, \dots, \quad (6.52)$$

где α_k — специальным образом подбираемый параметр. Самый простой и надежный выбор — это $\alpha_k = 1/2$. Другие стратегии выбора параметра α_k применяются для ускорения сходимости в начале вычислительного процесса. Если матрица A не имеет точек спектра на мнимой оси, то при правильном подборе α_k процесс (6.52) осуществим при всех $k \geq 1$ и сходится при $k \rightarrow \infty$:

$$A_k \rightarrow -G(+0) - G(-0), \quad (6.53)$$

где $G(t)$ — функция Грина дифференциального оператора $I \frac{d}{dt} - A$.

В этом пункте мы будем исследовать формулы (6.52) метода сигнум-функций при $\alpha_k \equiv 1/2$. Оказывается, что в этом случае вычислительный процесс можно проинтерпретировать как метод неортогонального исключения для пучка $\lambda(A - I) - (A + I)$. Легко показать, что инвариантное подпространство этого пучка, отвечающее собственным числам внутри единичного круга, совпадает с инвариантным подпространством матрицы A , отвечающим собственным числам в левой полуплоскости.

Рассмотрим процесс неортогонального исключения. Составим блочную матрицу

$$\begin{bmatrix} -(A_k + I) & A_k - I & 0 \\ 0 & -(A_k + I) & A_k - I \end{bmatrix}.$$

Первую строку этой матрицы умножим слева на матрицу X_k , вторую строку умножим слева на Y_k , а результаты сложим. Если при этом матрицы X_k и Y_k удовлетворяют системе

$$\begin{aligned} X_k(A_k - I) - Y_k(A_k + I) &= 0, \\ X_k + Y_k &= I, \end{aligned} \quad (6.54)$$

то в результате такого исключения получим блочную строку вида

$$-(A_{k+1} + I) \quad 0 \quad (A_{k+1} - I).$$

Рассмотрим подробнее систему матричных уравнений (6.54). В силу регулярности пучка $\lambda(A_k - I) - (A_k + I)$ решение первого уравнения из (6.54) представимо в виде $X_k = T_k^{-1}(A_k + I)$, $Y = T_k^{-1}(A_k - I)$ с произвольной невырожденной матрицей T_k . Второе уравнение системы (6.54) влечет за собой равенство $T_k = 2A_k$. Следовательно,

$$\begin{aligned} X_k(A_k + I) &= (2A_k)^{-1}(A_k^2 + 2A_k + I) = A_{k+1} + I, \\ Y_k(A_k - I) &= (2A_k)^{-1}(A_k^2 - 2A_k + I) = A_{k+1} - I. \end{aligned}$$

Таким образом, метод матричной сигнум-функции является неортогональной версией исключения, подробно описанного в 2.1. Как следствие этих рассуждений, скорость сходимости сигнум-метода оценивается параметром ω для пучка $\lambda(A - I) - (A + I)$.

Пусть

$$A_k - I = S_k^{-1}(P_{+0} + P_{-2k}), \quad A_k + I = S_k^{-1}(P_{-0} + P_{2k})$$

— каноническое разложение пучка $\lambda(A - I) - (A + I)$. Тогда $2I = S_k^{-1}(P_{-0} + P_{2k} - P_{+0} - P_{-2k})$. Оценим величину $\|A_k - (P_{-0} - P_{+0})\|$

при $k \rightarrow \infty$:

$$\begin{aligned}
 A_k &= (P_{-0} + P_{2^k} - P_{+0} - P_{-2^k})^{-1} (P_{+0} + P_{-2^k} + P_{-0} + P_{2^k}) = \\
 &= [(P_{-0} - P_{+0})^{-1} - (P_{-0} + P_{2^k} - P_{+0} - P_{-2^k})^{-1} (P_{2^k} - P_{-2^k}) (P_{-0} - \\
 &\quad - P_{+0})^{-1}] (I + P_{-2^k} + P_{2^k}) = (P_{-0} - P_{+0}) - \\
 &\quad - (P_{-0} - P_{+0} + P_{2^k} - P_{-2^k})^{-1} (P_{2^k} - P_{-2^k}) (P_{-0} - P_{+0}) + \\
 &\quad + (P_{-0} - P_{+0} + P_{2^k} - P_{-2^k})^{-1} (P_{-2^k} + P_{2^k}) = (P_{-0} - P_{+0}) + \\
 &\quad + 2 (P_{-0} - P_{+0} + P_{2^k} - P_{-2^k})^{-1} (P_{-2^k} + P_{2^k}), \quad (6.55) \\
 \|A_k - (P_{-0} - P_{+0})\| &= 2 \| (P_{-0} - P_{+0} + P_{2^k} - P_{-2^k})^{-1} (P_{-2^k} + P_{2^k}) \| \leq \\
 \leq 2 \|P_{-2^k} + P_{2^k}\| \| (P_{-0} - P_{+0} + P_{2^k} - P_{-2^k})^{-1} \| &\leq 2 \|P_{-2^k} + P_{2^k}\| / [\sigma_{\min}(P_{-0} - \\
 - P_{+0}) - \|P_{2^k} - P_{-2^k}\|] &\leq 2 (\|P_{-2^k}\| + \|P_{2^k}\|) (\|P_{+0}\| + \|P_{-0}\|) / [1 - \\
 - (\|P_{+0}\| + \|P_{-0}\|) (\|P_{2^k}\| + \|P_{-2^k}\|)] &\leq 2 \frac{4\omega e^{-2^k/(1+\omega)}}{1 - 4\omega e^{-2^k/(1+\omega)}}.
 \end{aligned}$$

Наконец, остановимся на выяснении вопроса о накоплении погрешностей вычислений в методе матричной сигнум-функции. Погрешности округлений на каждом шаге исключения зависят от величины обусловленности матриц A_k . Рассмотрим, к примеру, начало вычислительного процесса, т. е. оценим обусловленность матрицы $A_0 = A$. При нормировке пучка $\lambda(A - I) - (A + I)$ получим пучок матриц вида $2(I + AA^*)^{-\frac{1}{2}} \times [\lambda(A - I) - (A + I)]$. По теореме 1.3 имеем неравенство

$$\|[(A - I) - e^{i\varphi}(A + I)]^{-1} \sqrt{2(I + AA^*)}\| < 5\pi\omega.$$

Выбирая $e^{i\varphi} = -1$, получим следующую оценку:

$$\|A^{-1}\| < 5\sqrt{2}\pi\omega. \quad (6.56)$$

Можно построить примеры, которые показывают, что оценка (6.56) близка к оптимальной. Следовательно, обусловленность матрицы A оценивается неравенством

$$\|A\| \|A^{-1}\| < 5\sqrt{2}\pi\omega \|A\|.$$

Нетрудно показать, что $\|A_k\| \geq 1$ при $k \geq 1$. Поэтому при больших ω возможно большое накопление вычислительных погрешностей.

Может возникнуть вопрос о том, как зависит ω для пучка $\lambda(A - I) - (A + I)$ от умножения матрицы A на положительный скаляр. Эта зависимость сложная и не имеет практического интереса.

§ 7. ФАКТОРИЗАЦИЯ МАТРИЧНЫХ ПОЛИНОМОВ

Задача, обсуждаемая в этом параграфе, — вычисление разложения матричных полиномов на множители. Эта задача довольно нетривиальна даже в смысле выявления необходимых и достаточных условий существования таких разложений [8].

Будем рассматривать только регулярные матричные полиномы вида $A(\lambda) = A_0 + A_1\lambda + \dots + A_n\lambda^n$, $n \geq 2$, где A_j — квадратные матрицы одинаковых размеров $N \times N$, а $\det A(\lambda)$ не равен тождественно нулю при всех комплексных λ . Ограничимся только такими полиномами, для которых выполняется условие

$$\det(A_0 + A_1 e^{i\varphi} + \dots + A_n e^{in\varphi}) \neq 0 \quad (7.1)$$

при любом вещественном φ , т. е. не имеющими собственных чисел на единичной окружности. Собственные числа регулярного матричного по-

линома $A(\lambda)$ определяются стандартно: корни скалярного полинома $\det A(\lambda)$ называются *конечными собственными значениями* $A(\lambda)$, а нулевые собственные значения матрицы A_n — *бесконечными собственными значениями* полинома $A(\lambda)$.

Для матричного полинома, удовлетворяющего условию (7.1), поставим следующую задачу о факторизации: найти матричный полином $B(\lambda) = B_0 + \dots + B_d \lambda^d$, $B_d = I$ (если это возможно!), который делит $A(\lambda)$ справа, и выполнены условия

$$\det B(\lambda) \neq 0 \quad \text{при} \quad |\lambda| > 1,$$

$$\det(A(\lambda)B^{-1}(\lambda)) \neq 0 \quad \text{при} \quad |\lambda| < 1.$$

Другими словами, требуется разложить $A(\lambda)$ в произведение многочленов $C(\lambda)B(\lambda)$ так, чтобы спектр полинома $B(\lambda)$ лежал внутри единичного круга, а спектр $C(\lambda)$ — вне единичного круга.

7.1. Условия разрешимости задачи о факторизации в случае некратных собственных значений

Предположим, что все собственные числа матричного полинома $A(\lambda)$ различны. Поставим матричному полиному $A(\lambda)$ в соответствие линейный пучок матриц $K + \mu M$, где

$$K = \begin{bmatrix} A_0 & A_1 & \dots & A_{n-1} \\ & A_0 & \dots & A_{n-2} \\ & & \ddots & \vdots \\ 0 & & & A_0 \end{bmatrix}, \quad M = \begin{bmatrix} A_n & & & 0 \\ A_{n-1} & A_n & & \\ \vdots & & \ddots & \\ A_1 & A_2 & \dots & A_n \end{bmatrix}.$$

Нетрудно убедиться в том, что если $A(\lambda)x = 0$ при некоторых λ и $x \in \mathbb{R}^N$, то

$$(K + \lambda^n M) \begin{pmatrix} x \\ \lambda x \\ \vdots \\ \lambda^{n-1}x \end{pmatrix} = 0. \quad (7.3)$$

Таким образом, множество собственных чисел пучка $K + \mu M$ состоит из n -х степеней собственных чисел полинома $A(\lambda)$.

Если задача о факторизации разрешима, то, во-первых, в единичном круге лежит ровно Nd собственных чисел полинома $A(\lambda)$: $\lambda_1, \lambda_2, \dots, \lambda_{Nd}$, а, во-вторых, если $A(\lambda_i)x_i = 0$, $x_i \neq 0$, $i = 1, 2, \dots, Nd$, то матрица

$$W = \begin{bmatrix} x_1 & x_2 & \dots & x_{Nd} \\ \lambda_1 x_1 & \lambda_2 x_2 & \dots & \lambda_{Nd} x_{Nd} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{d-1} x_1 & \lambda_2^{d-1} x_2 & \dots & \lambda_{Nd}^{d-1} x_{Nd} \end{bmatrix} \quad (7.4)$$

невыврожденная. Справедливость первого утверждения очевидна. Для обоснования второго утверждения заметим, что $B(\lambda_i)x_i = 0$, $i = 1, 2, \dots, Nd$, и, следовательно, матрица W диагонализует пучок

$$\begin{bmatrix} B_0 & \dots & B_{d-1} \\ & \ddots & \vdots \\ 0 & & B_0 \end{bmatrix} + \mu \begin{bmatrix} B_d & & 0 \\ \vdots & \ddots & \\ B_1 & \dots & B_d \end{bmatrix}. \quad (7.5)$$

Точнее, матрица W состоит из столбцов, являющихся собственными векторами пучка (7.5). Таким образом, сформулированные выше два ут-

верждения дают необходимые условия разрешимости задачи о факторизации.

Покажем теперь, что эти утверждения являются и достаточными условиями разрешимости задачи о факторизации матричного полинома $A(\lambda)$. Сконструируем линейный пучок вида

$$\lambda I_{Nd} - W \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & & \ddots \\ 0 & & \lambda_{Nd} \end{bmatrix} W^{-1}. \quad (7.6)$$

В силу строения матрицы W получим равенство

$$W \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & & \ddots \\ 0 & & \lambda_{Nd} \end{bmatrix} W^{-1} = \begin{bmatrix} 0 & I & & 0 \\ & 0 & I & \\ & & \ddots & \\ 0 & & 0 & I \\ X_0 & X_1 & \dots & X_{d-1} \end{bmatrix} \quad (7.7)$$

с некоторыми $N \times N$ -матрицами $X_0, X_1, \dots, X_{d-1}$. Обозначим через $X(\lambda)$ матричный полином

$$X(\lambda) = X_0 + X_1 \lambda + \dots + X_{d-1} \lambda^{d-1} + I \lambda^d.$$

Очевидно, что $X(\lambda_i)x_i = 0$ для $i = 1, 2, \dots, Nd$. Докажем, что полином $A(\lambda)$ делится справа на $X(\lambda)$. Для этого разделим $A(\lambda)$ на $X(\lambda)$ с остатком: $A(\lambda) = C(\lambda)X(\lambda) + R(\lambda)$. Полином $R(\lambda)$ имеет степень меньше d . Так как

$$0 = A(\lambda_i)x_i - C(\lambda_i)X(\lambda_i)x_i = R(\lambda_i)x_i,$$

то $R(\lambda)$ имеет Nd собственных чисел. Следовательно, $R(\lambda) \equiv 0$. Подытожим результаты этого пункта.

Предложение 7.1. Задача о факторизации матричного полинома $A(\lambda)$ с некрратным спектром разрешима тогда и только тогда, когда выполнены следующие два условия:

- 1) внутри единичного круга лежит ровно Nd собственных чисел $\lambda_1, \dots, \lambda_{Nd}$ полинома $A(\lambda)$;
- 2) пусть $A(\lambda_i)x_i = 0, x_i \neq 0, i = 1, 2, \dots, Nd$, тогда матрица

$$W = \begin{bmatrix} x_1 & x_2 & \dots & x_{Nd} \\ \lambda_1 x_1 & \lambda_2 x_2 & \dots & \lambda_{Nd} x_{Nd} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{d-1} x_1 & \lambda_2^{d-1} x_2 & \dots & \lambda_{Nd}^{d-1} x_{Nd} \end{bmatrix}$$

невыврожденная.

При этом коэффициенты правого множителя

$$B(\lambda) = B_0 + B_1 \lambda + \dots + B_{d-1} \lambda^{d-1} + I \lambda^d$$

определяются из равенства

$$W \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & & \ddots \\ 0 & & \lambda_{Nd} \end{bmatrix} W^{-1} = \begin{bmatrix} 0 & I & & 0 \\ & 0 & I & \\ & & \ddots & \\ 0 & & 0 & I \\ B_0 & B_1 & \dots & B_{d-1} \end{bmatrix}.$$

7.2. Алгоритм решения задачи о факторизации

Пусть

$$Q_l(K + \mu M)Q_r = \begin{pmatrix} \mu I_{Nd} - J_0 & 0 \\ 0 & \mu J_\infty - I_{Nn-Nd} \end{pmatrix}$$

— кронекерова форма пучка $K + \mu M$, где Q_l и Q_r — невырожденные матрицы, а J_0 и J_∞ — матрицы в жордановой форме с собственными числами из единичного круга. Напомним, что

$$P_0 = Q_r \begin{pmatrix} I_{Nd} & 0 \\ 0 & 0 \end{pmatrix} Q_r^{-1}, \quad P_\infty = I - P_0$$

— проекторы на инвариантные подпространства пучка $K + \mu M$, отвечающие собственным числам соответственно внутри и вне единичного круга.

Вначале предположим, что все собственные числа полинома $A(\lambda)$ различны. Тогда матрица Q_r имеет следующий вид:

$$Q_r = \left(\begin{array}{c|c} W & * \\ \hline V & * \end{array} \right).$$

Можно показать, что

$$\Pi_0 = \begin{pmatrix} W \\ V \end{pmatrix} \left[(W^* \ V^*) \begin{pmatrix} W \\ V \end{pmatrix} \right]^{-1} (W^* \ V^*). \quad (7.8)$$

Условие невырожденности матрицы W эквивалентно невырожденности подматрицы Π_0^d матрицы

$$\Pi_0 = \left(\begin{array}{c|c} \Pi_0^d & * \\ \hline * & * \end{array} \right)$$

ортогонального проектора Π_0 для пучка матриц $K + \mu M$.

Матрица P_1 для пучка $K + \mu M$ имеет следующую структуру:

$$P_1 = Q_r \begin{pmatrix} J_0 & 0 \\ 0 & 0 \end{pmatrix} Q_r^{-1}.$$

Поэтому

$$P_1 \Pi_0 = \begin{pmatrix} W \\ V \end{pmatrix} J_0 \left[(W^* \ V^*) \begin{pmatrix} W \\ V \end{pmatrix} \right]^{-1} (W^* \ V^*).$$

Обозначим через P_1^d подматрицу размеров $Nd \times Nd$ матрицы

$$P_1 \Pi_0 = \left(\begin{array}{c|c} P_1^d & * \\ \hline * & * \end{array} \right).$$

Так как $\Pi_0^d = W S^{-1} W^*$, $P_1^d = W J_0 S^{-1} W^*$, где $S = W^* W + V^* V$, то

$$P_1^d (\Pi_0^d)^{-1} = W J_0 W^{-1}.$$

По предложению 7.1

$$P_1^d (\Pi_0^d)^{-1} = \begin{bmatrix} -0 & I & & \\ & 0 & I & 0 \\ & & \ddots & \\ 0 & & & I \\ B_0 & B_1 & \dots & B_{d-1} \end{bmatrix}.$$

В итоге мы получили следующую теорему.

Теорема 7.2. Задача о факторизации полинома $A(\lambda)$ разрешима тогда и только тогда, когда

1) ранг ортогонального проектора Π_0 пучка

$$\begin{bmatrix} A_0 & \dots & A_{n-1} \\ 0 & \ddots & \vdots \\ & & A_0 \end{bmatrix} + \mu \begin{bmatrix} A_n & 0 \\ \vdots & \ddots \\ A_1 & \dots & A_n \end{bmatrix}$$

на инвариантное подпространство, отвечающее собственным числам в единичном круге, равен Nd ;

2) подматрица Π_0^d проектора Π_0 невырожденная.

В случае разрешимости задачи о факторизации множитель $B(\lambda) = B_0 + \dots + I\lambda^d$ определяется из равенства

$$P_1^d (\Pi_0^d)^{-1} = \begin{bmatrix} -A_0 & I & & \\ & 0 & I & 0 \\ 0 & \ddots & \ddots & \vdots \\ & & 0 & I \\ -B_0 & B_1 & \dots & B_{d-1} \end{bmatrix},$$

где P_1^d — подматрица матрицы $P_1 \Pi_0$, а $P_1 = (KP_\infty + MP_0)^{-1} KP_0$.

Теорема 7.2 доказана в предположении, что матричный полином $A(\lambda)$ имеет некрatные собственные значения. Учитывая плотность множества полиномов с некрatным спектром и соображения непрерывности, нетрудно убедиться в справедливости теоремы 7.2, когда полином $A(\lambda)$ может иметь кратные собственные значения.

Опишем схему алгоритма решения задачи о факторизации матричного полинома $A(\lambda) = A_0 + A_1\lambda + \dots + A_n\lambda^n$:

сконструируем линейный пучок матриц

$$K + \mu M \begin{bmatrix} -A_0 & A_1 & \dots & A_{n-1} \\ & A_0 & \dots & A_{n-2} \\ 0 & \ddots & \ddots & \vdots \\ & & & A_0 \end{bmatrix} + \mu \begin{bmatrix} -A_n & & & \\ A_{n-1} & A_n & & 0 \\ \vdots & & \ddots & \vdots \\ A_1 & A_2 & \dots & A_n \end{bmatrix};$$

вычислим параллельные проекторы P_0 и $P_\infty = I - P_0$ пучка $K + \mu M$, если $\det(K + e^{i\varphi}M) \neq 0$ для любого вещественного φ ;

если $\text{rang } P_0 = Nd$, то проверяем невырожденность подматрицы Π_0^d ортогонального проектора Π_0 ;

если величина $\|\Pi_0^d\| \|(\Pi_0^d)^{-1}\|$ не слишком большая, то вычислим подматрицу P_1^d матрицы $(KP_\infty + MP_0)^{-1} K \Pi_0$ и решим линейную систему $Y \Pi_0^d = P_1^d$.

В результате всех этих вычислений получим матрицу

$$Y = \begin{bmatrix} 0 & I & & \\ & 0 & I & 0 \\ 0 & \ddots & \ddots & \vdots \\ & & 0 & I \\ -B_0 & B_1 & \dots & B_{d-1} \end{bmatrix},$$

где $B(\lambda) = B_0 + B_1\lambda + \dots + B_{d-1}\lambda^{d-1} + I\lambda^d$ — решение задачи о факторизации.

1. Гарантированная точность решения систем линейных уравнений/Годунов С. К., Антонов А. Г., Кириллук О. П., Костин В. П.— Новосибирск: Наука. Сиб. отд-ние, 1988.
2. Булгаков А. Я. Эффективно вычисляемый параметр качества устойчивости систем линейных дифференциальных уравнений с постоянными коэффициентами // Сиб. мат. журн.— 1980.— Т. 21, № 3.— С. 32—41.
3. Godunov S. K., Boulgakov A. J. Difficultes calculatives dans le problemes de Hurwitz et metodes a les surmontes // Analysis and Optimization of Systems: Proc./Fifth Intern. Conf. Versailles; 1982.— Berlin a. o.: Springer, 1982.— P. 845—851.— (Lecture Notes in Control and Information Sciences; 44).
4. Годунов С. К. Задача о дихотомии спектра матрицы // Сиб. мат. журн.— 1986.— Т. 27, № 5.— С. 24—37.
5. Булгаков А. Я., Годунов С. К. Круговая дихотомия матричного спектра.— Новосибирск, 1987.— 34 с.— (Препринт/АН СССР. Сиб. отд-ние. Ин-т математики; № 5).
6. Булгаков А. Я. Оценка матрицы Грина, теоремы непрерывности матрицы Грина и параметра дихотомии.— Новосибирск, 1987.— 36 с.— (Препринт/АН СССР. Сиб. отд-ние. Ин-т математики; № 6).
7. Булгаков А. Я. Точность расчета проекторов $P_+(A)$, $P_-(A)$ — Новосибирск, 1987.— 49 с.— (Препринт/АН СССР. Сиб. отд-ние. Ин-т математики; № 14).
8. Малышев А. Н. Факторизация матричных полиномов // Сиб. мат. журн.— 1982.— Т. 23, № 3.— С. 136—146.

С. И. ФАДЕЕВ

ПРОГРАММА ЧИСЛЕННОГО РЕШЕНИЯ НЕЛИНЕЙНЫХ КРАЕВЫХ ЗАДАЧ ДЛЯ СИСТЕМ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ПАРАМЕТРОМ

В предлагаемой работе рассматривается вычислительная сторона проблемы продолжения решения нелинейной краевой задачи по скалярному параметру Q . Краевая задача формулируется для системы обыкновенных дифференциальных уравнений на конечном отрезке с двухточечными краевыми условиями. Во многих приложениях этой проблемы параметр Q входит естественным образом в формулировку краевой задачи и при этом имеет место непрерывная зависимость решения от параметра. Метод продолжения может оказаться полезным в случае, когда требуется изучить поведение решения в зависимости от значений Q . В другом случае метод продолжения применяется для нахождения достаточно хорошего начального приближения, позволяющего численно построить решение краевой задачи при заданном значении Q итерационным способом. В силу нелинейности краевой задачи и в том и другом случае при продолжении решения по параметру могут возникнуть затруднения вычислительного характера, связанные с проблемой множественности решений.

По существу, в работе излагаются метод и программная реализация численного решения системы трансцендентных уравнений с параметром Q , являющейся дискретной моделью краевой задачи. Продолжение решения по параметру в сочетании с методом Ньютона — Канторовича сопровождается процедурой параметризации. Это позволяет, решая систему трансцендентных уравнений, построить часть гладкой пространственной кривой, задаваемой системой при помощи параметра Q , между двумя особыми точками, если таковые имеют место. Таким образом, в работе рассматривается только та множественность решений, которая возникает в случае неоднозначной зависимости функций от параметра Q , описывающих гладкую пространственную кривую.