тим, что выбранные в качестве характеристик ЭВМ числа γ и ε_i соответствуют характеристикам машины ЕС-1050, если для представления вещественных переменных используются двойные слова. Оценки δ_i получены при условии, что в алгоритме скалярные произведения векторов накапливаются с обычной точностью.

ЛИТЕРАТУРА

1. **Булгаков А. Я.** Эффективно вычисляемый параметр качества устойчивости систем линейных дифференциальных уравнений с постоянными коэффициентами.— Сиб. мат. журн., 1980, т. XXI, № 3, с. 32—41.
2. **Ward R. C.** Numerical computation of the matrix exponential with accuracy estimate.— SIAM J. Numerical Anal., 1977, v. 14, N 4, p. 600—610,
3. **Moler C., Van Loan C.** Nineteen dubious ways to compute the exponential of a matrix.— SIAM Review, 1978, v. 20, N 4, p. 801—836.
4. **Повзнер А. Я., Павлов Б. В.** Об одном методе численного интегрирования систем обыкновенных дифференциальных уравнений.— Журн. вычисл. математики и мат. физики, 1973, т. 13, № 4, с. 256—259.
5. **Godounov S. K., Boulgakov A. J.** Difficultés de calcul dans le probleme de Hurwitz et méthodes pour les surmonter.— In: Analysis and optimization of Systems, Versailles, 1982.— Proceedings (Lecture Notes in Control and Information Sciences, 44). Springer Verlag, 1982, p. 843—851.
6. **Булгаков А. Я., Годунов С. К.** Численное определение одного из критериев качества устойчивости систем линейных дифференциальных уравнений с постоянными коэффициентами.— Новосибирск, 1981.— 58 с. (Препринт/АН СССР, Сиб. отд-ние, ИМ).
7. **Годунов С. К.** Решение систем линейных уравнений.— Новосибирск: Наука. Сиб. отд-ние, 1980.— 177 с.
8. **Уилкинсон Дж. Х.** Алгебраическая проблема собственных значений.— М.: Наука, 1970.— 564 с.
9. **Levis A. H.** Some Computational Aspects of the Matrix Exponential.— IEEE Trans. Automatic Control, 1969, v. AC-14, N 4, p. 410—411.

РАСЧЕТ ПОЛОЖИТЕЛЬНО ОПРЕДЕЛЕННЫХ РЕШЕНИЙ УРАВНЕНИЯ ЛЯПУНОВА

А. Я. БУЛГАКОВ, С. К. ГОДУНОВ

ВВЕДЕНИЕ

В большинстве приложений представляют интерес лишь положительно определенные решения H матричного уравнения Ляпунова

$$A^*H + HA + I = 0, \tag{1}$$

где I — единичная матрица размерности $N \times N$. Положительная определенность имеет место только тогда, когда A гурвицева, т. е. если нулевое решение системы обыкновенных уравнений

$$\dot{x} = Ax \tag{2}$$

асимптотически устойчиво.

В работе предлагается детальный алгоритм расчета H, который сопровождается анализом гурвицевости A. Если оказалось, что A не гурвицева, то процесс завершается без указания H. В противном случае указывается положительно определенное, приближенное с машинной точностью решение уравнения (1).

Численный анализ гурвицевости использует числовую характеристику устойчивости

$$\varkappa(A) = 2\|A\| \max_{\dot{x}=Ax} \left\{ \int_0^\infty \|x(t)\|^2 \, dt / \|x(0)\|^2 \right\}, \tag{3}$$

где $\|x\|$ — евклидова норма вектора x; $\|A\|$ — спектральная норма матрицы A. Параметр качества устойчивости ($\varkappa(A)$) введен в работе [1] как результат решения некоторой алгебраической задачи. А именно, анализируя решение матричного уравнения (1), полагается $\varkappa(A)=\infty$, если решение H не оказалось положительно определено или если это уравнение не разрешимо. В противном случае $\varkappa(A)$ определяется формулой $\varkappa(A)=2\|A\|\cdot\|H\|$. С. К. Годуновым в работе [2] замечено, что введенный параметр качества устойчивости допускает представление вида (3). В [1] показано, что для решения системы (2) выполнена оценка

$$\|x(t)\| \leqslant \sqrt{\varkappa(A)}\, e^{-t\|A\|/\varkappa(A)}\|x(0)\|.$$

Там же на основе приведенной оценки доказано важное неравенство

$$\|e^{tA}\| \leqslant \sqrt{\varkappa(A)}\, e^{-t\|A\|/\varkappa(A)}. \qquad (4)$$

Для решения уравнения Ляпунова (1) используется вариант известного метода [2, 3], основанного на итерационном удвоении интервала интегрирования $(0, 2^k h)$ в интегральном представлении решения H:

$$H = \lim_{k\to\infty} H_k = \lim_{k\to\infty}\left(H_k = \int_0^{2^k h} e^{tA^*}e^{tA}\,dt\right).$$

Положив

$$H_0 = \int_0^h e^{tA^*}e^{tA}dt;\ B_k = e^{2^k hA} = B_{k-1}^2;\ B_0 = e^{hA},$$

будем иметь $H_k = H_{k-1} + B_{k-1}^* H_{k-1} B_{k-1}$, что приводит к простой вычислительной схеме.

В обзоре [4] отмечена высокая эффективность этого метода и предлагается в случаях плохой сходимости или расходимости процесса, варьируя параметр h, добиваться улучшения сходимости. Из проведенного в работе анализа влияния ошибок округления на процесс вытекает, что его сходимость зависит от величины $\varkappa(A)$, а не от параметра h. Поэтому при реализации процесса всегда будем брать $h = 1/(2\|A\|)$. Такой выбор h удобен при вычислении матриц B_0 и H_0, основанном на тейлоровском разложении.

Алгоритм расчета H для уравнения (1) опирается на численное определение $\varkappa(A)$. Если при вычислении установлено, что $\varkappa(A)$ превышает некоторое значение $\varkappa_{гр}$, порядок величины которого диктуется разрядностью используемой ЭВМ, то дальнейший расчет не проводится и констатируется «практическая» неустойчивость системы (2).

Понятие «практической» устойчивости неоднократно вводилось разными авторами (см., например, [5, 6]). Здесь используется один из возможных вариантов этого понятия, допускающий численную оценку.

В § 1 выведены оценки относительной погрешности вычисления матрицы H и параметра $\varkappa(A)$. Кроме того, получены неравенства, играющие немаловажную роль в схеме расчета H для «практически» устойчивых матриц A, изложенной в § 3. Само понятие «практической» устойчивости детально обсуждается в § 2.

Отметим, что идея II этапа предлагаемой схемы является развитием соображений, предложенных в дипломной работе А. М. Исраилова, выполненной в 1980 г. под руководством С. К. Годунова в Новосибирском государственном университете. На данном этапе используются оценки погрешностей вычисления матричных экспонент от асимптотически устойчивых матриц, выведенных в работе [7]. Оценки получены в предположении применения специальных приемов программной реализации алгоритма (дополнительные нормировки матриц и векторов, вычисления с двойной точностью). Совокупность таких приемов названа «арифметикой вынесенных порядков». Здесь мы также будем использовать эту

«арифметику». В § 4 выведены оценки точности вычисления матрицы

$$\widetilde{H}_0 = \int_0^1 e^{t/(2\|A\|)A^*} e^{t/(2\|A\|)A} dt,$$

которая вместе с приближениями к матрицам $e^{2^k/(2\|A\|)A}$ позволяет вычислить матрицу $(H = 1/(2\|A\|)\widetilde{H})$:

$$\widetilde{H} = \sum_{k=0}^{\infty} e^{k/(2\|A\|)A^*} \left[\int_0^1 e^{t/(2\|A\|)A^*} e^{t/(2\|A\|)A} dt \right] e^{k/(2\|A\|)A}.$$

В § 5 выведены погрешности вычисления матрицы $X + Y^*XY$, § 6, носящем вспомогательный характер,— неравенства, играющие основную роль при учете влияния ошибок округления на процесс вычисления приближений к матрице H, приведенный в § 7. В § 8 описан детальный алгоритм, предлагаемый для расчета положительно определенных решений уравнения Ляпунова (1). В § 9 приведена сводка численных экспериментов.

§ 1. ПОГРЕШНОСТИ ВЫЧИСЛЕНИЯ И НЕКОТОРЫЕ НЕРАВЕНСТВА

Выводятся оценки погрешности определения $\varkappa(A)$ по известной погрешности приближенного решения уравнения Ляпунова $A^*H + HA + I = 0$.

Теорема 1. *Если* $X = X^*$ *— положительно определенная матрица,* $A^*X + XA + I - C = 0$, *и выполнено неравенство* $\|C\| < 1$, *то матрица* A *гурвицева и для нее*

$$|\varkappa(A) - 2\|X\| \cdot \|A\|| / (2\|X\| \cdot \|A\|) \leqslant \|C\| / (1 - \|C\|);$$
$$|\varkappa(A) - 2\|X\| \cdot \|A\|| / \varkappa(A) \leqslant \|C\|; \quad \|H - X\| / \|H\| \leqslant \|C\|.$$

Приступая к доказательству, заметим, что асимптотическая устойчивость A по теореме Ляпунова вытекает из положительной определенности X и $I - C$.

Из уравнений $A^*H + HA + I = 0$ и $A^*X + XA + I - C = 0$ получаем уравнение Ляпунова для разности $H - X$: $A^*(H - X) + (H - X)A + C = 0$, решение которого при гурвицевой A, как известно, допускает представление следующей интегральной формулой:

$$H - X = \int_0^{\infty} e^{tA^*} C e^{tA} dt.$$

Из этого представления следует очевидная цепочка неравенств:

$$|\|H\| - \|X\|| \leqslant \|H - X\| = \max_{\|x\|=1} ([H - X]x, x) =$$

$$= \max_{\|x\|=1} \int_0^{\infty} (Ce^{tA}x, e^{tA}x)\, dt \leqslant \max_{\|x\|=1} \left(\left[\int_0^{\infty} e^{tA^*} e^{tA} dt \right] x, x \right) \|C\| = \|H\| \cdot \|C\|,$$

получение которой доказывает справедливость теоремы 1.

В дальнейшем на протяжении всего параграфа будем считать, что матрица A асимптотически устойчива $(\varkappa = \varkappa(A) < \infty)$.

Введем обозначения. Положим

$$H_T(C) = \int_0^T e^{tA^*} C e^{tA} dt,$$

а если $C = I$, то $H_T = H_T(I)$. Хорошо известно (см. [1]), что

$$H(C) = H_\infty(C) = \int_0^\infty e^{tA^*} C e^{tA} dt$$

является решением уравнения Ляпунова $A^*H(C) + H(C)A + C = 0$. В таком случае

$$H(C) - H_T(C) = \int_T^\infty e^{tA^*} C e^{tA} dt.$$

Лемма 1. *Если* $C = C^* > 0$, *то*

$$\lambda_{\min}(H_T(C)) \geqslant \lambda_{\min}(C)/(2\|A\|)(1 - e^{-2T\|A\|}).$$

Воспользовавшись очевидным неравенством $\sigma_1(e^{tA}) \geqslant \|e^{-tA}\|^{-1} \geqslant e^{-t\|A\|}$ и определением матрицы $H_T(C)$, получаем

$$(H_T(C)x, x) = \int_0^T (Ce^{tA}x, e^{tA}x)\,dt \geqslant \sigma_1(C)\|x\|^2 \int_0^T \sigma_1^2(e^{tA})\,dt \geqslant$$

$$\geqslant \|x\|^2 \lambda_{\min}(C) \int_0^T e^{-2t\|A\|} dt \geqslant \frac{\lambda_{\min}(C)}{2\|A\|}(1 - e^{-2T\|A\|})\|x\|^2.$$

В заключение доказательства леммы осталось заметить, что

$$\lambda_{\min}(H_T(C)) = \min_{\|x\|=1}(H_T(C)x, x) \geqslant \lambda_{\min}(C)/(2\|A\|)(1 - e^{-2T\|A\|}).$$

Для ρ из интервала $(0, 1)$ определим число $T_\rho = \varkappa/(2\|A\|)\ln[4\varkappa/\rho]$. Очевидно, что при этом для всех $T > T_\rho$ будет выполнена оценка

$$\|e^{TA}\|^2 \leqslant \varkappa e^{-2T\|A\|/\varkappa} \leqslant \varkappa e^{-2T_\rho\|A\|/\varkappa} = \rho/4.$$

Справедливо представление

$$A^*H_T(C) + H_T(C)A + C = A^* \int_0^T e^{tA^*} C e^{tA} dt + \int_0^T e^{tA^*} C e^{tA} dt\, A + C =$$

$$= \int_0^T [e^{tA^*} C e^{tA}]'\,dt + C = e^{TA^*} C e^{TA}, \qquad (1.1)$$

позволяющее заключить, что при $T > T_\rho$

$$\|A^*H_T(C) + H_T(C)A + C\| = \|e^{TA^*} C e^{TA}\| \leqslant \|e^{TA}\|^2 \|C\| \leqslant \|C\|\rho/4.$$

Теорема 2. *Для всех матриц* $X = X^*$, *удовлетворяющих неравенству* $(\rho < 1)$: $\|H_T(C) - X\| \leqslant \|C\|/(2\|A\|)\cdot\rho/4$, *при* $T > T_\rho = \varkappa/(2\|A\|)\ln(4\varkappa/\rho)$ *справедлива оценка* $\|A^*X + XA + C\| \leqslant \|C\|\cdot\rho/2$.

Теорема 3. *Для всех матриц* $X = X^*$, *удовлетворяющих неравенству* $(\rho < 1)$: $\|H_T - X\| \leqslant 1/(2\|A\|)\cdot\rho/4$, *при* $T > T_\rho = \varkappa/(2\|A\|)\ln(4\varkappa/\rho)$ *справедлива оценка* $\lambda_{\min}(X) \geqslant (1 - \rho/2)/(2\|A\|)$.

Переходя к доказательству теоремы 2, заметим, что из (1.1) вытекает равенство

$$A^*X + XA + C = e^{TA^*} C e^{TA} + A^*[X - H_T(C)] + [X - H_T(C)]A,$$

которое в силу условия теоремы и неравенства

$$\varkappa e^{-2T\|A\|/\varkappa} \leqslant \varkappa e^{-2T\rho\|A\|/\varkappa} = \rho/4 \qquad (1.2)$$

приводит к завершающей доказательство цепочке неравенств

$$\|A^*X + XA + C\| \leqslant \|e^{TA^*} C e^{TA}\| + 2\|A\|\cdot\|X - H_T(C)\| \leqslant$$

$$\leqslant \varkappa e^{-2T\|A\|/\varkappa}\|C\| + \|C\|\rho/4 \leqslant \|C\|\rho/2.$$

Доказательство теоремы 3 следует из цепочки неравенств

$$\lambda_{\min}(X) \geqslant \lambda_{\min}(H_T) - \|H_T - X\| \geqslant$$
$$\geqslant (1 - e^{-2T\|A\|})/(2\|A\|) - \rho/(8\|A\|) \geqslant (1 - \rho/2)/(2\|A\|),$$

обосновываемой оценкой леммы 1 и неравенством (1.2).

§ 2. ПРАКТИЧЕСКАЯ УСТОЙЧИВОСТЬ

Понятие практической устойчивости нулевого решения системы $\dot{x} = Ax$ неоднократно вводилось разными авторами (см., например, [5, 6]). Рассмотрим один из возможных вариантов этого понятия. Задавшись числом $\varkappa_{\text{гр}}$ ($\varkappa_{\text{гр}} > 1$), будем считать систему (2) «практически» устойчивой, если для любого начального вектора $x(0)$ справедлива оценка

$$\varkappa(A) = \max_{\dot{x}=Ax} \left\{ \int_0^\infty \|x(t)\|^2 dt / \|x(0)\|^2 \right\} \cdot 2\|A\| \leqslant \varkappa_{\text{гр}}. \tag{2.1}$$

В противном случае констатируется «практическая» неустойчивость системы (2). Если в процессе численного анализа матрицы удается на некотором этапе установить, что неравенство (2.1) не выполнено, то процесс останавливается и его результатом считается неравенство $\varkappa(A) > \varkappa_{\text{гр}}$.

При выборе $\varkappa_{\text{гр}}$ необходимо учитывать, во-первых, точность задания элементов матрицы A и, во-вторых, структуру разрядной сетки ЭВМ, используемой для анализа устойчивости A.

В самом деле, в работе [1] при условии $\|B\|/\|A\| < \varkappa^{-2}/2$ выведена оценка $\varkappa(A+B) \leqslant 4\varkappa^3(A)\|B\|/\|A\| + \varkappa(A)$, которую можно огрубить до неравенства $\varkappa(A+B) < 3\varkappa(A)$. Следовательно, зная отношение нормы матрицы B — матрицы допустимых отклонений элементов A — к норме матрицы A, можно, задавшись величиной $\varkappa_{\text{гр}}$, удовлетворяющей неравенству $\varkappa_{\text{гр}} < [\|A\|/(2\|B\|)]^{1/2}$, гарантировать, что $\varkappa(A+B) < 3\varkappa(A)$ в случае «практической» устойчивости A.

В [1] предложено вычислять $\varkappa(A)$ из анализа решения уравнения Ляпунова

$$\mathfrak{L}H \equiv A^*H + HA + I = 0. \tag{2.2}$$

Если это уравнение не разрешимо или если его решение не положительно определено, то $\varkappa(A)$ полагалось равным ∞. В противном случае $\varkappa(A) = 2\|A\| \cdot \|H\|$.

Разрешимость линейного уравнения (2.2) зависит от числа обусловленности

$$\mu(\mathfrak{L}) = \left\{ \max_X [\|\mathfrak{L}X\|/\|X\|] \Big/ \min_X [\|\mathfrak{L}X\|/\|X\|] \right\}.$$

В работе [8] выведена оценка $\mu(\mathfrak{L}) \leqslant N\varkappa^2(A)$, позволяющая заключить, что уравнение (2.2) плохо разрешимо только в случае плохого качества устойчивости матрицы A. Параметр $\varkappa(A)$ выступает аналогом числа обусловленности для задачи расчета положительно определенных решений уравнения (2.2), а значение $\varkappa_{\text{гр}}$ является допустимым уровнем обусловленности, который выбирается с учетом влияния ошибок округления в конкретном методе решения (2.2).

Для рассматриваемого метода решения уравнения $A_1^*\widetilde{H} + \widetilde{H}A_1 + I = 0$, $\|A_1\| = 1/2$, в § 7 показано, что он позволяет получать матрицы $\widetilde{H}_k$ так, что

$$\left\| \widetilde{H}_k - \int_0^{2^k} e^{tA_1^*} e^{tA_1} dt \right\| \leqslant 14\sqrt{N}\,\varepsilon_1 (\varkappa^*)^{7/2},$$

где $\varkappa(A) < \varkappa^*$. Следовательно, предположив, что $\varkappa_{\text{гр}}$ выбирается из усло-

вия $14\sqrt{N}\varepsilon_1[\varkappa_{\text{гр}}]^{7/2} = \rho/4$, где ρ — некоторое число из интервала $(0,1)$, в силу теорем 2, 3 (см. § 1) можно гарантировать, что если $\varkappa(A) < \varkappa_{\text{гр}}$, то

$$\lambda_{\min}(\widetilde{H}_k) \geqslant 1 - \rho/2; \quad \|A_1^*\widetilde{H}_k + \widetilde{H}_k A_1 + I\| \leqslant \rho/2.$$

Если хотя бы одно из этих неравенств не выполнено, то гарантировано выполнение неравенства $\varkappa(A) > \varkappa_{\text{гр}}$, т. е. матрица A «практически» неустойчива. Иначе, в силу теоремы 1 (см. § 1) выполнены неравенства

$$\|1/(2\|A\|)\widetilde{H}_k - H\|/\|H\| \leqslant \rho/2; \qquad |\varkappa(A) - \|\widetilde{H}_k\|\|/\varkappa(A) \leqslant \rho/2.$$

§ 3. СХЕМА АЛГОРИТМА РАСЧЕТА ПОЛОЖИТЕЛЬНО ОПРЕДЕЛЕННЫХ РЕШЕНИЙ УРАВНЕНИЯ ЛЯПУНОВА

Описывается грубая схема расчета положительно определенных решений уравнения Ляпунова. Впоследствии будут подробно описаны и исследованы все его детали.

Процесс решения уравнения Ляпунова $A^*H + HA + I = 0$ сопровождается анализом гурвицевости матрицы A, обеспечивающей положительную определенность матрицы H. При этом, если выясняется, что $\varkappa(A) > \varkappa_{\text{гр}}$, то матрицу уже не стоит причислять к числу гурвицевых, а систему $\dot{x} = Ax$ можно считать «практически» неустойчивой (см. § 2). Поэтому если в процессе анализа A удается на некотором этапе установить такое неравенство, то процесс останавливается и его результатом считается неравенство $\varkappa(A) > \varkappa_{\text{гр}}$. Если же установить справедливость неравенства не удается, то процесс доводится до вычисления положительно определенной матрицы $\widetilde{H}$, приближающей решение H с точностью порядка ε_1 относительной точности машинного представления чисел.

Мы предлагаем выбирать $\varkappa_{\text{гр}}$ так: возьмем ρ из интервала $(0, 1)$, затем, используя функцию $\Delta(\varkappa) = 14\sqrt{N}\varkappa^{3/2}\varepsilon_1$, определенную в § 7, выберем $\varkappa_{\text{гр}}$ из условия $\Delta(\varkappa_{\text{гр}})\,\varkappa_{\text{гр}}^2 = \rho/4$, т. е. $\varkappa_{\text{гр}} = (56\sqrt{N}\varepsilon_1/\rho)^{-2/7}$.

Схему вычислений разобьем на пять этапов. Перед началом расчета нужно еще задаться величиной ρ_0 ($\rho_0 \leqslant \rho$), которая должна характеризовать желаемую точность расчета H (в случае, если $\varkappa(A) < \varkappa_{\text{гр}}$).

I этап состоит в вычислении сингулярных чисел $\sigma_N(A) = \|A\|$ и $\sigma_1(A) = \|A^{-1}\|^{-1}$ с гарантированной точностью [9]. Если оказывается, что $\sigma_N(A) \geqslant \sigma_1(A)\,\varkappa_{\text{гр}}^{3/2}$, то в силу неравенства $[\varkappa(A)]^{3/2} \geqslant \mu(A)$ [8] можем заключить, что $\varkappa(A) > \varkappa_{\text{гр}}$, и закончить исследование гурвицевости матрицы A. В противном случае, зная $\|A\|$, нормируем A так: $A_1 = 1/(2\sigma_N(A)) \cdot A$. Отметим, что качество устойчивости при этом не меняется (см. [1]).

II этап состоит из следующих основных шагов:

1°. Выберем целое число $\tilde{k}$, удовлетворяющее неравенствам $2^{\tilde{k}-1} \leqslant T_\rho = \varkappa_{\text{гр}} \ln[4\varkappa_{\text{гр}}/\rho] \leqslant 2^{\tilde{k}}$. Вычисляются матрицы

$$B_0 = e^{A_1}, \ldots, B_k = e^{2^k A_1} = B_{k-1}^2, \ldots, B_{\tilde{k}} = B_{\tilde{k}-1}^2 = e^{2^{\tilde{k}} A_1}$$

и их нормы $\|B_0\|, \|B_1\|, \ldots, \|B_{\tilde{k}}\|$. В реальном процессе будут вычислены близкие к требуемым матрицы $\{(B_k)_{\text{выч}}\}$ и числа $\{\|(B_k)_{\text{выч}}\|\}$. Полученная последовательность матриц является вспомогательной для реализуемого на III этапе итерационного процесса решения уравнения Ляпунова.

2°. При всех $k = 0, 1, 2, \ldots, \tilde{k}$ проверяется неравенство

$$\|(B_k)_{\text{выч}}\| \leqslant (\sqrt{\varkappa_{\text{гр}}} + 1)\exp\{-2^{k-1}/\varkappa_{\text{гр}}\}.$$

Если оно нарушено для некоторого k, то процесс завершается утверждением $\varkappa(A) > \varkappa_{\text{гр}}$. В противном случае в силу выбора $\tilde{k}$

$$\|(B_{\tilde{k}})_{\text{выч}}\| \leqslant (1 + 1/\sqrt{\varkappa_{\text{гр}}})\exp\left\{-2^{\tilde{k}-1}/\varkappa_{\text{гр}}\right\} \leqslant (1 + 1/\sqrt{\varkappa_{\text{гр}}})\sqrt{\rho}/2 < 1,$$

что обеспечивает получение приближенного решения $\widetilde{\widetilde{H}}$ на III этапе с точностью $\|\widehat{H}-\widetilde{\widetilde{H}}\|/\|\widehat{H}\| \leqslant \rho/2$.

Для обоснования заключения пункта 2° заметим, что если $\varkappa(A) < \varkappa_{\text{гр}}$, то для всех $k \leqslant k_0$ выполнены оценки

$$\|B_k\| = \left\|e^{2^k A_1}\right\| \leqslant \sqrt{\varkappa(A)}\, e^{-2^k \|A_1\|/\varkappa(A)} = \sqrt{\varkappa(A)}\, e^{-2^{k-1}/\varkappa(A)};$$
$$\|B_k - (B_k)_{\text{выч}}\| \leqslant \exp\{-2^{k-1}/\varkappa\}. \tag{3.1}$$

В § 6 отмечено, что эти неравенства справедливы, если $k \leqslant k_0$, где k_0 — целое число, удовлетворяющее неравенствам $2^{k_0} < 1/(4r_0) + 1/2 < 2^{k_0+1}$. Покажем, что $k_0 > \tilde{k}$. Рассмотрим цепочку неравенств

$$2^{k_0} \geqslant (16,16\sqrt{N}\varepsilon_1 \varkappa_{\text{гр}}^{3/2})^{-1} + \frac{1}{2} > (17\sqrt{N}\varepsilon_1 \varkappa_{\text{гр}}^{3/2})^{-1} =$$
$$= \frac{14}{17\Delta(\varkappa_{\text{гр}})} = \frac{14}{17}\cdot\frac{4}{\rho}\varkappa_{\text{гр}}^2 > 2\varkappa_{\text{гр}}\varkappa_{\text{гр}}/\rho > 2\varkappa_{\text{гр}} \ln[4\varkappa_{\text{гр}}/\rho] \geqslant 2^{\tilde{k}},$$

из которой следует, что $k_0 > \tilde{k}$.

Из (3.1) нетрудно заключить, что при всех $k \leqslant \tilde{k}$

$$\|(B_k)_{\text{выч}}\| \leqslant (\sqrt{\varkappa(A)} + 1)\exp\{-2^{k-1}/\varkappa(A)\}. \tag{3.2}$$

Данное неравенство вместе с оценкой

$$\sqrt{\varkappa}\exp\{-2^{\tilde{k}-1}/\varkappa_{\text{гр}}\} \leqslant \sqrt{\rho}/2,$$

следующей из определения $\tilde{k}$, обосновывает заключение пункта 2°.

III этап состоит в получении приближенного решения уравнения Ляпунова $\widehat{H}A_1 + A_1^*\widehat{H} + C_0 = 0$, $C_0 = I$. В дальнейшем на этапе IV будут вычисляться аналогичные приближения для решений $H^{(j)}$ при правых частях c_j, отличных от $C_0 = I$. Здесь использованы числа ρ и $\tilde{k}$, определенные ранее. Рассматриваемый этап разбивается на следующие основные шаги:

1°. Вычисляется матрица

$$H_0 = \int_0^1 e^{tA_1^*} C_0 e^{tA_1} dt.$$

Детальный алгоритм определения H_0 описан в § 4. Там же проведен анализ влияния ошибок округления.

2°. Вычисляются матрицы ($k = 1, 2, \ldots, \tilde{k}$)

$$H_k = H_{k-1} + (B_{k-1})_{\text{выч}}^* H_{k-1} (B_{k-1})_{\text{выч}},$$

где $\{(B_{k-1})_{\text{выч}}\}$ получены и исследованы на предыдущем этапе. В реальном процессе будут получены близкие к требуемым H_k матрицы $\{(H_k)_{\text{выч}}\}$.

3°. При всех $k = 0, 1, 2, \ldots, \tilde{k}$ проверяется неравенство $\|(H_k)_{\text{выч}}\| < \rho/4 + \varkappa_{\text{гр}}$. Если оно нарушено для некоторого k, то процесс завершается утверждением $\varkappa(A) > \varkappa_{\text{гр}}$.

4°. Определяются величины $\Lambda = \lambda_{\max}[(H_{\tilde{k}})_{\text{выч}}]$ и $\lambda = \lambda_{\min}[(H_{\tilde{k}})_{\text{выч}}]$ — максимальное и минимальное собственные числа симметричной матрицы $(H_{\tilde{k}})_{\text{выч}}$; $C_1 = A_1^*(H_{\tilde{k}})_{\text{выч}} + (H_{\tilde{k}})_{\text{выч}} A + I$ — невязка полученного приближенного решения уравнения Ляпунова, $|C_1|$ — ее спектральная норма.

5°. Проверяются неравенства $\lambda \geqslant (1-\rho/2)$; $\|C_1\| \leqslant \rho/2$, $\varkappa_{\text{гр}} \leqslant \Lambda - \rho/2$. Если хотя бы одно из них не выполнено, то исследование матрицы A

завершается утверждением $\varkappa(A) > \varkappa_{\text{гр}}$. В противном случае гарантировано выполнение неравенств

$$\|C_1\| = \left\|A_1^*(H_{\tilde{k}})_{\text{выч}} + (H_{\tilde{k}})_{\text{выч}} A_1 + C_0\right\| \leqslant \|C_0\| \cdot \rho/2 = \rho/2;$$

$$\left\|(H_{\tilde{k}})_{\text{выч}} - \widehat{H}\right\| / \left\|\widehat{H}\right\| \leqslant \rho/2; \quad |\varkappa(A) - \Lambda| / \varkappa(A) \leqslant \rho/2.$$

6°. После вычисления $\|C_1\|$ проверяется выполнение неравенства $\|C_1\| \leqslant \rho_0$. Если оно выполнено, то расчет считается законченным. Можно гарантировать

$$\left\|(H_{\tilde{k}})_{\text{выч}} - \widehat{H}\right\| / \left\|\widehat{H}\right\| < \rho_0; \quad |\varkappa(A) - \Lambda| / \varkappa(A) < \rho_0.$$

Это дает основание положить $\bar{\varkappa}(A) = \Lambda$ и затем передать управление этапу V.

Если $\|C_1\| > \rho_0$, т. е. если полученная точность не достаточна, то в работу включается IV этап — итерационного уточнения искомых величин.

Перейдем к обоснованию заключений п. 3°—6°. Прежде всего отметим, что в силу результатов § 7 последовательности

$$\{(H_k)_{\text{выч}}\} \text{ и } \left\{\int_0^{2^k} e^{tA_1^*} C_0 e^{tA_1} dt\right\}$$

связаны неравенствами

$$\left\|(H_k)_{\text{выч}} - \int_0^{2^k} e^{tA_1^*} C_0 e^{tA_1} dt\right\| \leqslant \Delta(\varkappa)\, \varkappa^2 \|C_0\|, \tag{3.3}$$

если $k \leqslant k_1$, где k_1 — максимальное целое число, удовлетворяющее неравенству $(2^k - 1) r_0 < 1/2$.

Покажем, что $\tilde{k} < k_1$. Из определения следует, что $2^{k_1} < 1/(2r_0) + 1 < 2^{k_1+1}$ и, значит, $k_1 = k_0 + 1$. При обосновании заключений II этапа показано, что $k_0 > \tilde{k}$, следовательно, $k_1 = k_0 + 1 > \tilde{k}$. Далее, поскольку $\Delta(\varkappa)$ — растущая функция аргумента $\varkappa$, то из предположения $\varkappa(A) < \varkappa_{\text{гр}}$ и неравенства (3.3) вытекает

$$\left\|(H_k)_{\text{выч}} - \int_0^{2^k} e^{tA_1^*} C_0 e^{tA_1} dt\right\| \leqslant \Delta(\varkappa_{\text{гр}})\, \varkappa_{\text{гр}}^2 \|C_0\| = \rho/4.$$

Полученное неравенство вместе с оценкой

$$\left\|\int_0^{2^k} e^{tA_1^*} C_0 e^{tA_1} dt\right\| \leqslant \|C_0\| \cdot \varkappa(A)$$

позволяет заключить, что если $\varkappa(A) < \varkappa_{\text{гр}}$, то при всех $k \leqslant \tilde{k}$ верно неравенство $\|(H_k)_{\text{выч}}\| \leqslant \rho/4 + \varkappa_{\text{гр}}$, а значит, получено обоснование шага 3°.

Наконец отметим, что справедливость оставшихся заключений этого этапа следует из теорем 1 — 3, доказанных в § 1.

IV этап состоит в уточнении полученного на III этапе приближенного решения уравнения $A_1^* \widehat{H} + \widehat{H} A_1 + I = 0$, совпадающего с известной процедурой итерационного уточнения приближенного решения линейной системы алгебраических уравнений (см., например, [10]).

Для начала процесса задаются матрицы

$$H_{\tilde{k}}^0 = (H_{\tilde{k}})_{\text{выч}}; \quad C_1 = A_1^* (H_{\tilde{k}})_{\text{выч}} + (H_{\tilde{k}})_{\text{выч}} A_1 + I,$$

найденные на III этапе.

Сформулируем i-й шаг ($i \geqslant 1$) процесса уточнения. На этом шаге вычисляется приближенное решение уравнения $A_1^* X^{(i)} + X^{(i)} A_1 + C_i = 0$. После $(i-1)$-го шага определены матрицы $H_{\tilde{k}}^{(i-1)}$ и C_i.

1°. Вычисляется матрица

$$X_0^{(i)} = \int_0^1 e^{tA_1^*} C_i e^{tA_1} dt.$$

2°. Используя известные матрицы $\{(B_0)_{\text{выч}}\}$, $\{(B_1)_{\text{выч}}\}$, ..., $\{(B_{\tilde{k}})_{\text{выч}}\}$, строится последовательность $(k = 1, 2, \ldots, \tilde{k})$

$$X_k^{(i)} = X_{k-1}^{(i)} + (B_{k-1})_{\text{выч}}^* X_{k-1}^{(i)} (B_{k-1})_{\text{выч}},$$

которая сходится в силу утверждений этапов II и III.

3°. Определяется матрица $H_{\tilde{k}}^{(i)} = H_{\tilde{k}}^{(i-1)} + X_{\tilde{k}}^{(i)}$.

4°. Вычисляется невязка C_{i+1} и ее норма $\|C_{i+1}\|$:

$$C_{i+1} = A_1^* X_{\tilde{k}}^{(i)} + X_{\tilde{k}}^{(i)} A_1 + C_i.$$

Заметим, что из результатов предыдущих этапов следует неравенство $\|C_{i+1}\| \leqslant \rho/2 \|C_i\|$, которое обеспечивает сходимость процесса итерационного уточнения.

5°. Проверяется неравенство $\|C_{i+1}\| \leqslant \rho_0$, если оно не выполнено, то начинается $(i+1)$-й шаг процесса уточнения. В противном случае матрица $H_{\tilde{k}}^{(i_0)}$ $(i = i_0)$ и есть искомое приближение к решению уравнения $A_1^* \widehat{H} + \widehat{H} A_1 + I = 0$.

V этап — завершающий, на нем мы имеем матрицу $\widehat{H}$ приближенного решения уравнения (3.4). Это либо матрица $(H_{\tilde{k}})_{\text{выч}}$, найденная на этапе III, либо матрица $H_{\tilde{k}}^{(i_0)}$, полученная на IV этапе. Этап состоит из двух шагов.

1°. Определяется $\lambda_{\max}(\widehat{H})$ — максимальное собственное число матрицы $\widehat{H}$ и полагается $\overline{\varkappa}(A) = \lambda_{\max}(\widehat{H})$. При этом в силу результатов предыдущих этапов верно неравенство $|\varkappa(A) - \overline{\varkappa}(A)|/\varkappa(A) \leqslant \rho_0$.

2°. Вычисляется матрица $\widetilde{H} = 1/(2\sigma_N(A))\widehat{H}$, которая и есть искомое положительно определенное приближение решения матричного уравнения $A^*H + HA + I = 0$, так как верно неравенство $\|H - \widetilde{H}\|/\|H\| < \rho_0$.

§ 4. УЧЕТ ПОГРЕШНОСТЕЙ ОКРУГЛЕНИЯ ПРИ ВЫЧИСЛЕНИИ МАТРИЦЫ $\int_0^1 e^{tA^*} C e^{tA} dt$

Пусть A — матрица малой нормы $(1/2 \leqslant \|A\| \leqslant 1)$. Для вычисления $\int_0^1 e^{tA^*} C e^{tA} dt$ используем один из известных алгоритмов (см., например, [4]): $(k > 1)$; $T_1 = C$; $Q_1 = C$;

$$T_{k+1} = 1/(k+1)\{A^* T_k + T_k A\}; \quad Q_{k+1} = Q_k + T_{k+1}. \tag{4.1}$$

Заметим, что если $\mathfrak{L}$ — оператор Ляпунова, заданный на пространстве симметричных матриц размерности $N \times N$: $\mathfrak{L}X \equiv A^*X + XA$, то из (4.1) следует

$$Q_m = \sum_{k=1}^{m} 1/k! \, \mathfrak{L}^{(k-1)} C.$$

Это, в свою очередь, позволяет, используя представление

$$\int_0^1 e^{tA^*} C e^{tA} dt = \sum_{k=1}^{\infty} 1/k! \, \mathfrak{L}^{(k-1)} C, \tag{4.2}$$

заключить, что

$$\left\| \int_0^1 e^{tA^*} C e^{tA} dt - C_m \right\| = \left\| \sum_{k=m+1}^{\infty} 1/k!\, \mathfrak{L}^{(k-1)} C \right\| \leqslant$$

$$\leqslant \|C\| \sum_{k=m+1}^{\infty} (2\|A\|)^{k-1}/k! = \|C\| (2\|A\|)^m/(m+1)!\, e^{2\|A\|}. \qquad (4.3)$$

Выведенное неравенство (4.3) дает оценку точности приближения матрицей Q_m интеграла $\int_0^1 e^{tA^*} C e^{tA} dt$ при любом m. Для получения оптимального приближения матрицы $\int_0^1 e^{tA^*} C e^{tA} dt$ необходимо остановить процесс (4.1) после k_0 шагов, когда гарантированная погрешность вычисления $\bar{Q}_{k_0}$ станет одного порядка с погрешностью приближения матрицей Q_{k_0} интеграла $\int_0^1 e^{tA^*} C e^{tA} dt$. В конце параграфа указан алгоритм выбора k_0 в каждом конкретном случае. Для учета погрешностей, возникающих при реализации процесса (4.1), достаточно предположить, что матрицы связаны соотношениями $\bar{T}_1 = C$; $\bar{Q}_1 = C$;

$$\bar{T}_m = 1/m\{A^* \bar{T}_{m-1} + \bar{T}_{m-1} A\} + \psi_m; \quad \bar{Q}_m = \bar{Q}_{m-1} + \bar{T}_m + \varphi_m, \qquad (4.4)$$

в которых φ_m и ψ_m — квадратные матрицы размерности $N \times N$, обозначают погрешности. Использование при расчете по формулам (4.1) «арифметики вынесенных порядков» позволяет гарантировать выполнение неравенств

$$\|\psi_m\| \leqslant 2\varepsilon_1 d_1/m \|A\| \cdot \|\bar{T}_{m-1}\|; \quad \|\varphi_m\| \leqslant d_2 \varepsilon_1 (\|\bar{T}_m\| + \|\bar{Q}_{m-1}\|), \qquad (4.5)$$

где ε_1 — машинная константа; d_1, d_2 — некоторые постоянные, зависящие от N. Если предположить, что скалярные произведения векторов накапливаются с двойной точностью при реализации (4.1), то в силу результатов § 2 работы [7] в качестве d_1 и d_2 можем взять

$$d_1 = d_2 = 1{,}01\sqrt{N}. \qquad (4.6)$$

Оценим разность $\bar{T}_m - 1/m!\mathfrak{L}^{(m-1)} C$. Из (4.4) следует равенство

$$\bar{T}_m - \frac{1}{m!}\, \mathfrak{L}^{(m-1)} C = \frac{1}{m}\, \mathfrak{L}\left[\bar{T}_{m-1} - \frac{1}{(m-1)!}\, \mathfrak{L}^{(m-2)} C\right] + \psi_m,$$

позволяющее вывести цепочку неравенств ($\|\mathfrak{L}\|$ — норма оператора $\mathfrak{L}$; $\|\mathfrak{L}\| \leqslant 2\|A\|$):

$$\|\bar{T}_m - 1/m!\mathfrak{L}^{(m-1)} C\| \leqslant \|\mathfrak{L}\|/m \cdot \|\bar{T}_{m-1} - 1/(m-1)!\mathfrak{L}^{(m-2)} C\| + \|\psi_m\| \leqslant$$
$$\leqslant 2\|A\|/m \|\bar{T}_{m-1} - 1/(m-1)!\mathfrak{L}^{(m-2)} C\| + 2d_1\varepsilon_1/m \cdot \|A\| \cdot \|\bar{T}_{m-1}\| \leqslant$$
$$\leqslant 2\|A\|/m(1 + d_1\varepsilon_1)\|\bar{T}_{m-1} - 1/(m-1)!\mathfrak{L}^{(m-2)} C\| +$$
$$+ d_1\varepsilon_1 2\|A\|/m \|1/(m-1)!\mathfrak{L}^{(m-2)} C\| \leqslant d_1\varepsilon_1 2\|A\| \cdot (2\|A\|)^{m-2}/(m-1)!\|C\| =$$
$$= d_1\varepsilon_1 (2\|A\|)^{m-1}/(m-1)!\|C\|.$$

Опираясь на полученную оценку и используя неравенства (4.5), докажем по индукции, что если

$$\sum_{m=2}^{k} \varepsilon_1 (1{,}01 d_2 m + 1{,}01 d_1 2\|A\|) \leqslant 0{,}01,$$

то

$$\|\bar{Q}_k - Q_k\| \leqslant 1{,}01 (d_2 k + 2\|A\| d_1) \|C\|/(2\|A\|) e^{2\|A\|} \varepsilon_1. \qquad (4.7)$$

В самом деле, при $k = 2$ неравенство (4.7) справедливо в силу (4.2)—(4.5). Предположим, что оно верно при всех $k < j$, и докажем, что тогда

(4.7) верно при $k=j$. Сделанные предположения позволяют написать цепочку неравенств

$$\|\tilde{Q}_j - Q_j\| \leqslant \left\| \sum_{m=2}^{j} \tilde{T}_m - 1/m!\, \mathfrak{L}^{(m-1)}C + \varphi_m \right\| \leqslant$$

$$\leqslant \sum_{m=2}^{j} \{\|\tilde{T}_m - 1/m!\, \mathfrak{L}^{(m-1)}C\|(1 + d_2\varepsilon_1) + d_2\varepsilon_1 \|1/m!\, \mathfrak{L}^{(m-1)}C\| +$$

$$+ d_2\varepsilon_1 \|Q_{m-1}\| + d_2\varepsilon_1 \|\tilde{Q}_{m-1} - Q_{m-1}\|\} \leqslant$$

$$\leqslant (1 + d_2\varepsilon_1) \sum_{m=2}^{j} d_1\varepsilon_1 (2\|A\|)^{m-1}/(m-1)!\|C\| +$$

$$+ d_2\varepsilon_1 \sum_{m=2}^{j} \sum_{k=1}^{m} \|1/k!\, \mathfrak{L}^{(k-1)}C\| + \sum_{m=2}^{j} d_2\varepsilon_1 \|C\|/(2\|A\|)\, e^{2\|A\|} \times$$

$$\times 1{,}01\,(d_2 m + 2\|A\|d_1) \leqslant 1{,}01\,(d_2 j + 2\|A\|d_1)\|C\|/(2\|A\|)\, e^{2\|A\|}\varepsilon_1.$$

Полученная оценка вместе с (4.3) дает возможность заключить, что наиболее подходящим выбором значения k_0 для лучшего приближения матричного интеграла $\int_0^1 e^{tA*} C e^{tA} dt$ является минимальное среди всех целых k, удовлетворяющих оценке $1{,}01\,(d_2 k + 2\|A\|d_1) \geqslant (2\|A\|)^{k+1}/(k+1)!$, где d_1 и d_2 заданы (4.6). Возникающая при этом погрешность оценивается неравенством

$$\left\| \tilde{Q}_{k_0} - \int_0^1 e^{tA*} C e^{tA} dt \right\| \leqslant 2{,}02\,(d_2 k_0 + 2\|A\|d_1)\|C\|/(2\|A\|)\, e^{2\|A\|}\varepsilon_1.$$

§ 5. ОЦЕНКА ТОЧНОСТИ ВЫЧИСЛЕНИЯ $X+Y*XY$

Рассмотрим алгоритм вычисления произведения матриц $V = Y*XY$, использующего «арифметику вынесенных порядков». Здесь, как и в работе [7], для формального описания арифметики будут использоваться операторы $\mathfrak{M}$ и $\mathscr{P}$, а матрицы V, X и Y задаваться своими каноническими парами: $[V^0, V^1]$, $[X^0, X^1]$ и $[Y^0, Y^1]$ (см. § 2 работы [7]).

Последовательность операций

$$\tilde{\tilde{V}} = X^0 Y^0; \quad \tilde{\tilde{\tilde{V}}} = Y^{0*}\tilde{\tilde{V}}; \quad V^0 = \mathfrak{M}(\tilde{\tilde{\tilde{V}}}); \quad V^1 = \mathscr{P}(\tilde{\tilde{\tilde{V}}}) + X^1 + 2Y^1,$$

позволяет вычислить каноническую пару $[V^0, V^1]$ матрицы $V = Y*XY$.

Заметим, что при вычислении $\tilde{\tilde{V}}$ применяются машинные числа с удвоенной мантиссой. Удобно это требование реализовать на языке АССЕМБЛЕР ЕСЭВМ или АВТОКОД для БЭСМ-6. Итак, как показано, например, в § 21 работы [9], используя машинные числа с удвоенной длиной мантиссы для вычисления элементов матрицы $\tilde{\tilde{V}}$ и числа с удвоенной мантиссой для хранения элементов $\tilde{\tilde{\tilde{V}}}$, можно вычислить их с точностью

$$|(\tilde{\tilde{V}}_{ij})_{\text{выч}} - \tilde{\tilde{V}}_{ij}| \leqslant \left[\sum_{k=1}^{N} (X_{ik}^0)^2\right]^{1/2} \cdot \left[\sum_{k=1}^{N} (Y_{kj}^0)^2\right]^{1/2} \cdot N/\gamma\varepsilon_1^2 + N/\gamma^2\varepsilon_2,$$

где ε_1, ε_2 — машинные константы (см. § 2 [7]); γ — основание системы счисления используемой ЭВМ. Выведенные неравенства позволяют заключить, что

$$\|(\tilde{\tilde{\tilde{V}}})_{\text{выч}} - \tilde{\tilde{\tilde{V}}}\| \leqslant 2N^2/\gamma^2\varepsilon_2 + 2N/\gamma\varepsilon_1^2 \|X^0\|_E \cdot \|Y^0\|_E.$$

Далее, умножая матрицу $\tilde{\tilde{\tilde{V}}}$ на матрицу Y^{0*}, полученный результат будем хранить в виде массива чисел со стандартной длиной мантиссы.

В этом случае, возможно, необходимо будет привлечение машинных команд для перемножения чисел с мантиссами разной длины. Такая процедура вычисления гарантирует выполнение неравенств

$$\|(Y^{0*}(\widetilde{V})_{\text{выч}})_{\text{выч}} - Y^{0*}(\widetilde{V})_{\text{выч}}\|_E \leqslant \varepsilon_1 \|Y^{0*}(\widetilde{V})_{\text{выч}}\|_E +$$
$$+ 2N/\gamma\varepsilon_1^2 \|Y^0\|_E \|(\widetilde{V})_{\text{выч}}\|_E + 2N^2/\gamma^2\varepsilon_2 \leqslant \varepsilon_1 \|Y^{0*}X^0Y^0\|_E +$$
$$+ \varepsilon_1 \|Y^0\|_E \cdot \|(\widetilde{V})_{\text{выч}} - \widetilde{V}\|_E + 2N^2/\gamma^2\varepsilon_2 + N/\gamma\varepsilon_1^2 \|Y^0\|_E \cdot \|(\widetilde{V})_{\text{выч}} - \widetilde{V}\|_E +$$
$$+ N/\gamma\varepsilon_1^2 \|Y^0\|_E \cdot \|X^0Y^0\|_E \leqslant \varepsilon_1 \|Y^{0*}X^0Y^0\|_E +$$
$$+ N/\gamma\varepsilon_1^2 \|Y^0\|_E \cdot \|X^0Y^0\|_E + 2N^2/\gamma^2\varepsilon_2 + (N/\gamma\varepsilon_1^2 \|Y^0\|_E +$$
$$+ \varepsilon_1 \|Y^0\|_E (2N^2/\gamma^2\varepsilon_2 + N/\gamma\varepsilon_1^2 \|Y^0\|_E \cdot \|X^0\|_E).$$

Так как всегда можно считать, что $1/\gamma \leqslant \|Y^0\|_E \leqslant N$; $1/\gamma \leqslant \|X^0\|_E \leqslant N$, то цепочку неравенств можно продолжить:

$$\|(Y^{0*}(\widetilde{V})_{\text{выч}})_{\text{выч}} - Y^{0*}(\widetilde{V})_{\text{выч}}\|_E \leqslant \varepsilon_1 \|Y^{0*}X^0Y^0\|_E +$$
$$+ (4N/\gamma\varepsilon_1^2 + 4N^2\gamma\varepsilon_2) \|X^0\|_E \cdot \|Y^0\|_E^2.$$

Это неравенство, в свою очередь, позволяет гарантировать выполнение оценки

$$\|(Y^*XY)_{\text{выч}} - Y^*XY\|_E \leqslant \varepsilon_1 \|Y^*XY\|_E + 4N(\varepsilon_1^2/\gamma + 4\gamma\varepsilon_2) \|X^0\|_E \cdot \|Y^0\|_E^2,$$

которая на основании результатов § 2 работы [7] обеспечивает следующую оценку точности вычисления матрицы $X + Y^*XY$:

$$\|(X + Y^*XY)_{\text{выч}} - (X + Y^*XY)\|_E \leqslant$$
$$\leqslant 2N/\gamma\varepsilon_2 (\|X\|_E + \|Y^*XY\|_E) + \varepsilon_1 \|X + Y^*XY\|_E +$$
$$+ \|(Y^*XY)_{\text{выч}} - Y^*XY\|_E \leqslant (2N/\gamma\varepsilon_2 + 2\varepsilon_1) \|X + Y^*XY\|_E +$$
$$+ (\varepsilon_1 + 4N/\gamma\varepsilon_2 + 4N/\gamma\varepsilon_1^2 \|Y\|_E^2 + 4N^2\gamma\varepsilon_2 \|Y\|_E^2) \|X\|_E \leqslant$$
$$\leqslant 2{,}01\varepsilon_1 \|X + Y^*XY\|_E + \varepsilon_1 \|X\|_E \{1 + 4N\varepsilon_2/(\gamma\varepsilon_1) + 4N(\varepsilon_1/\gamma +$$
$$+ 4N\varepsilon_2/\varepsilon_1) \|Y\|_E^2\}.$$

Значит, если предположить, что

$$4N\varepsilon_2/(\gamma\varepsilon_1) + (4N/\gamma\varepsilon_1 + 4N^2\gamma\varepsilon_2/\varepsilon_1) \|Y\|_E^2 < 0{,}01,$$

то

$$\|(X + Y^*XY)_{\text{выч}} - (X + Y^*XY)\|_E \leqslant 2{,}01\varepsilon_1 \|X + Y^*XY\|_E + 1{,}01\varepsilon_1 \|X\|_E.$$

§ 6. УЧЕТ ПОГРЕШНОСТИ ВЫЧИСЛЕНИЯ e^{pA}

Напомним некоторые утверждения, доказанные в работе [7], и на их основании получим обобщающие неравенства. В данном параграфе матрица A асимптотически устойчива с $\varkappa = \varkappa(A) < \varkappa^* < \infty$ и нормирована условием $\|A\| = 1/2$.

Для начала возьмем целое число k_0, удовлетворяющее неравенствам $2^{k_0} < 1/(4r_0) + 1/2 < 2^{k_0+1}$, где $r_0 \ll 1$.

Рассмотрим последовательность матриц размерности $N \times N$: B_0, B_1, ..., B_{k_0}, связанную при помощи квадратных матриц ($j = 0, 1, 2, \ldots, k_0$) рекуррентными соотношениями

$$B_0 = e^A + \Phi_0; \quad B_{m+1} = B_m^2 + \Phi_{m+1}, \tag{6.1}$$

где $\|\Phi_0\| \leqslant r_0/(2\varkappa^*)\, e^{-1/(2\varkappa^*)}$; $\|\Phi_m\| \leqslant r_0/(2\varkappa^{*3/2}) \|B_{m-1}^2\|$.
Для ее членов по теореме 1 § 4 работы [7] верны неравенства ($k \leqslant k_0$)

$$\|B_k - \exp\{-2^k A\}\| \leqslant \delta(k) \exp\{-2^k/(2\varkappa^*)\}, \tag{6.2}$$

где $\delta(k) = (2^k - 1) r_0/\{1 - (2^k - 1) r_0\}$.

Выбор k_0 обеспечивает выполнение неравенства $\delta(k) \leqslant \delta(k_0) < 1/2$, которое позволяет огрубить (6.2):

$$\|B_k - \exp\{2^k A\}\| < \exp\{-2^k/(2\varkappa^*)\}.$$

Прежде чем переходить к оценке близости степеней e^{pA} с различными произведениями матриц B_i, напомним одно обозначение из [7]. Раскроем скобки в следующем из (6.1) выражении $B_{m+1} = (\ldots((e^A + \Phi_0)^2 + \Phi_1)^2 + \ldots + \Phi_m)^2 + \Phi_{m+1}$ и соберем в R^i_{m+1} все слагаемые, содержащие Φ_j в суммарной степени, равной i. В этом случае верно представление

$$B_{m+1} = \exp\{2^{m+1}A\} + R^1_{m+1} + \ldots + R^{2^{m+1}}_{m+1}.$$

В § 4 работы [7] показано, что для всех p, $q > 0$ верны оценки ($k \leqslant k_0$)

$$\|e^{pA} R^j_k e^{qA}\| \leqslant [(2^{k+1} - 1) r_0]^j \exp\{-(2^k + p + q)/(2\varkappa^*)\}. \qquad (6.3)$$

Перейдем к оценке близости степеней матричной экспоненты e^{pA} с различными произведениями матриц B_i.

Лемма 1. *Если* m, $l \leqslant k_0$, *то*

$$\|B_m B_l - e^{(2^m + 2^l)A}\| \leqslant \frac{2r_0(2^m + 2^l - 1)}{1 - 2r_0(2^m + 2^l - 1)} \exp\{-(2^m + 2^l)\|A\|/\varkappa^*\}.$$

Так как

$$B_m B_l = [e^{2^m A} + R^1_m + \ldots + R^{2^m}_m][e^{2^l A} + R^1_l + \ldots + R^{2^l}_l],$$

то, обозначив через G_j сумму всех слагаемых, содержащих $R^{i_1}_m$ и $R^{i_2}_l$ с суммарным верхним индексом, равным j ($i_1 + i_2 = j$), будем иметь равенство

$$B_m B_l = \exp\{(2^m + 2^l)A\} + G_1 + \ldots + G_{2^m + 2^l}. \qquad (6.4)$$

Для более удобного доказательства леммы введем матрицы $R^i_m = 0$ при $i > 2^m$; $R^j_l = 0$ при $j > 2^l$, которые позволяют выписать простые соотношения ($j = 2, 3, \ldots, 2^m + 2^l$)

$$G_1 = \exp\{2^m A\} R^1_l + R^1_m \exp\{2^l A\};$$
$$G_j = \exp\{2^m A\} R^j_l + R^j_m \exp\{2^l A\} + \sum_{\tau=1}^{j-1} R^\tau_m R^{j-\tau}_l. \qquad (6.5)$$

Из (6.3) и (6.5) получаем, что при всех p, $q > 0$ верны неравенства

$$\|e^{pA} G_1 e^{qA}\| \leqslant [2^{l+1} + 2^{m+1} - 2] r_0 \exp\{-(2^m + 2^l + p + q)\|A\|/\varkappa^*\}.$$

Аналогично из (6.3) и (6.5) следует, что ($j = 2, 3, \ldots, 2^m + 2^l$)

$$\|e^{pA} G_j e^{qA}\| \leqslant \exp\{-(2^m + 2^l + p + q)\|A\|/\varkappa^*\} \cdot \{[(2^{l+1} - 1) r_0]^j +$$
$$+ [(2^{m+1} - 1) r_0]^j + \sum_{\tau=1}^{j-1} [(2^{m+1} - 1) r_0]^\tau [(2^{l+1} - 1) r_0]^{j-\tau}\} \leqslant$$
$$\leqslant \{[2^{l+1} + 2^{m+1} - 2] r_0\}^j \exp\{-(2^m + 2^l + p + q)\|A\|/\varkappa^*\}.$$

Полученные неравенства и (6.4) позволяют сделать вывод о справедливости утверждения леммы 1.

Из доказанной леммы следует, что если $m_{1i}, m_2, \ldots, m_i \leqslant k_0$ и если обозначить $m_{1i} = 2^{m_1} + 2^{m_2} + \ldots + 2^{m_i}$, то имеет место представление

$$\prod_{j=1}^{i} B_{m_j} - e^{m_{1i} A} = G_1 + G_2 + \ldots + G_{m_{1i}},$$

где для матриц $\{G_k\}_{k=1}^{m_{1i}}$ при всех p, $q > 0$ выполнены оценки

$$\|e^{pA} G_m e^{qA}\| \leqslant r_0 (m_{1i} - i) \exp\{-(p + q + m_{1i})\|A\|/\varkappa^*\},$$

позволяющие заключить, что

$$\left\| e^{pA}\left[\prod_{j=1}^{i} B_{m_j} - e^{m_{1i}A}\right] e^{qA}\right\| \leqslant$$

$$\leqslant r_0(2m_{1i}-1)/\{1-r_0(2m_{1i}-1)\}\exp\{-(p+q+m_{1i})\|A\|/\varkappa^*\}.$$

Пусть $t = m_{1i} + \tau$, где $0 \leqslant \tau \leqslant 1$, $m_j \leqslant k_0$. В этом случае

$$\left\| e^{\tau A}\prod_{j=1}^{i} B_{m_j} - e^{(\tau+m_{1i})A}\right\| \leqslant \delta_1(t)\exp\{-t\|A\|/\varkappa^*\},$$

где $\delta_1(t) = r_0(t-1)/(1-r_0(t-1))$, $t > 1$.

Полученные результаты потребуются в следующем параграфе, здесь же укажем, каким образом выбирается r_0 при конкретной машинной реализации рассмотренных ранее алгоритмов.

Пусть ε_1 — машинная постоянная. В § 3 [7] предложен алгоритм вычисления e^A, позволяющий гарантировать оценку

$$\|B_0 - e^A\| \leqslant \alpha_0\varepsilon_1\exp\{-1/(2\varkappa^*)\},$$

где

$$\alpha_0 = 2{,}02\sqrt{N}(2k_1 + 1{,}01/2)\exp\{(1+1/\varkappa^*)/2\},$$

а k_1 определяется как минимальное целое k, при котором верна оценка: $1{,}01\sqrt{N}(k+1{,}01/2)\varepsilon_1 \geqslant 1/[2^{k+1}\cdot(k+1)!]$. В § 2 этой же работы показано, что если Φ_{m+1} — погрешность вычисления B_m^2 и если скалярные произведения векторов накапливаются с двойной точностью, то верно неравенство $\|\Phi_{m+1}\| \leqslant 1{,}01\sqrt{N}\varepsilon_1\|B_m^2\|$.

Приведенные оценки позволяют заключить, что если взять

$$r_0 = 2\varkappa^*\varepsilon_1\max\{\alpha_0,\ \sqrt{\varkappa^*}\cdot 1{,}01\sqrt{N}\} =$$

$$= 2{,}02\varkappa^*\sqrt{N}\varepsilon_1\max\{\sqrt{\varkappa^*},\ (4k_1 + 1{,}01\exp(1+1/(2\varkappa^*)))\},$$

то все выведенные оценки будут справедливы при машинной реализации рассмотренных алгоритмов. Более того, если взять $\varkappa^* > [4k_1 + 1{,}01e]^2$, то нас удовлетворит выбор $r_0 = 2{,}02\sqrt{N}\varepsilon_1\varkappa^{*3/2}$.

§ 7. УЧЕТ ПОГРЕШНОСТИ ВЫЧИСЛЕНИЯ $\int_0^{2^k} e^{tA^*}Ce^{tA}dt$

Пусть A — асимптотически устойчивая матрица с $\varkappa(A) < \varkappa^* < \infty$ и $\|A\| = 1/2$. На одном из основных этапов схемы § 3 вычисляются матрицы ($m = 1, 2, \ldots, k_0$)

$$H_m = H_{m-1} + B_{m-1}^* H_{m-1} B_{m-1}, \tag{7.1}$$

где k_0 задана в § 3, алгоритм вычисления матрицы

$$H_0 = \int_0^1 e^{tA^*}Ce^{tA}dt$$

с оценкой накопившейся погрешности приведен в § 4, а матрицы B_m — приближающие матрицы $\exp\{2^mA\}$. Алгоритм вычисления B_m с оценкой близости $\exp\{2^mA\}$ показан в работе [7]. Более детально последовательность $\{B_m\}$ рассмотрена в § 6.

Здесь будут выведены оценки близости матрицы $\int_0^{2^k} e^{tA^*}Ce^{tA}dt$ с матрицами $\tilde{H}_k$, полученными при машинной реализации процесса (7.1) при условии, что ($m = 0, 1, 2, \ldots, k$)

$$\|\tilde{H}_m\| < 4/3\varkappa^*\|C\|. \tag{7.2}$$

Если для некоторого m $(0 \leqslant m \leqslant k_0)$ выполнено неравенство $\|\widetilde{H}_m\| > 4/3\varkappa^*\|C\|$, то процесс (7.1) завершается ввиду слишком большого роста погрешности. В самом деле, верно неравенство

$$\left\| \int_0^{2m} e^{tA^*} C e^{tA} dt \right\| \leqslant \left\| \int_0^{\infty} e^{tA^*} C e^{tA} dt \right\| \leqslant \|C\| \frac{\varkappa(A)}{2\|A\|} < \|C\| \varkappa^*.$$

Для учета погрешностей, возникающих при реализации процесса (7.1), достаточно предположить, что полученные матрицы связаны соотношениями

$$\widetilde{H}_0 = \int_0^1 e^{tA^*} C e^{tA} dt + \varphi_0;$$

$$\widetilde{H}_m = \widetilde{H}_{m-1} + B_{m-1}^* \widetilde{H}_{m-1} B_{m-1} + \varphi_m, \tag{7.3}$$

в которых квадратные $N \times N$ матрицы $\{\varphi_i\}$ обозначают погрешности, допущенные при вычислении.

Исходя из результатов § 4 и 5 можно гарантировать выполнение неравенств $(m \geqslant 1)$

$$\|\varphi_m\| < 3\sqrt{N}\varepsilon_1 \|\widetilde{H}_m\|; \quad \|\varphi_0\| \leqslant 5{,}6\sqrt{N}\varepsilon_1 (m_0 + 1)\|C\|,$$

где m_0 определяется как минимальное среди всех m, удовлетворяющих неравенству $2{,}03(m+1)\sqrt{N}\varepsilon_1 \geqslant 1/(m+1)!$. Если взять

$$\varkappa^* > 5{,}6(m_0 + 1), \tag{7.4}$$

(напомним, что в этом параграфе $\varkappa^* > \varkappa(A)$), то из полученных неравенств следует $(m \geqslant 0)$

$$\|\varphi_m\| \leqslant 4\sqrt{N}\varkappa^*\varepsilon_1\|C\|. \tag{7.5}$$

Для дальнейшего изложения потребуется теорема 4, доказательство которой приведено в конце параграфа.

Теорема 4. *Пусть A — асимптотически устойчивая матрица с $\varkappa(A) < \varkappa^* < \infty$ и $\|A\| = 1/2$. Возьмем $r_0 \ll 1$ и k_1 такое целое число, что $(2^{k_1} - 1) r_0 < 1/2 < (2^{k_1+1} - 1) r_0$. Рассмотрим последовательность матриц $N \times N$: $B_0, B_1, \ldots, B_{k_1}$, связанную при помощи квадратных матриц $\{\Phi_j\}$ рекуррентными соотношениями*

$$B_0 = e^A + \Phi_0; \quad B_{m+1} = B_m^2 + \Phi_{m+1},$$

где

$$\|\Phi_0\| \leqslant r_0/(2\varkappa^*) \exp\{-1/(2\varkappa^*)\}; \quad \|\Phi_j\| \leqslant r_0/(2\varkappa^{*3/2}) \|B_{j-1}^2\|.$$

С ее помощью, отправляясь от матрицы

$$\widehat{H} = \int_0^1 e^{tA^*} C e^{tA} dt + \varphi_0,$$

получим из рекуррентных соотношений $(m = 1, 2, \ldots, k_1;\ \alpha \ll 1)$

$$\widehat{H}_m = \widehat{H}_{m-1} + B_{m-1}^* \widehat{H}_{m-1} B_{m-1} + \varphi_m; \quad \|\varphi_m\| \leqslant \alpha\varkappa^* \|C\|$$

и при условии $\|\widehat{H}_m\| < 4/3\|C\|\varkappa^*$ последовательность матриц $\widehat{H}_0, \widehat{H}_1, \ldots, \widehat{H}_{k_1}$, для которой верны оценки

$$\left\| \widehat{H}_{k_1} - \int_0^{2^{k_1}} e^{tA^*} C e^{tA} dt \right\| \leqslant \Delta_{k_1}(\varkappa^*) \varkappa^{*2} \|C\|,$$

где

$$\Delta_{k_1}(\varkappa^*) = [6r_0 + 8\varkappa^*\alpha(k_1 + 1)] \exp\{-1/\varkappa^*\} + 2\alpha(k_1 + 1)/\varkappa^*.$$

Из теоремы 4 следует, что если $\varkappa^*$ удовлетворяет неравенствам $2(k+1)/\varkappa^{*2}+8(k+1)\exp\{-1/\varkappa^*\}<\sqrt{\varkappa^*}/4$, то $\Delta_k(\varkappa^*)$ и $\Delta(\varkappa^*)=6r_0+1/4\varkappa^{*3/2}$ будут связаны неравенством $\Delta_k(\varkappa^*)<\Delta(\varkappa^*)$ при $k<k_1$.

Нетрудно видеть, что, взяв $r_0=2{,}02\varepsilon_1\sqrt{N}\varkappa^{*3/2}$ и $\alpha=\sqrt{N}\varepsilon_1$, можно, опираясь на результаты § 6 и 7, использовать для оценки близости матриц $\int_0^{2^k} e^{tA^*}Ce^{tA}dt$ и $\widetilde{H}_k$ теорему 4. В этом случае $\Delta(\varkappa^*)=14\varepsilon_1\sqrt{N}\varkappa^{*3/2}$, и значит, если k_1 удовлетворяет неравенствам $(2^{k_1}-1)r_0<1/2<r_0(2^{k_1+1}-1)$, то либо для некоторого k $(k\leqslant k_1)$ выполнено неравенство $\|\widetilde{H}_k\|>4/3\|C\|\cdot\varkappa^*$, либо для всех k $(k\leqslant k_1)$

$$\left\|\widetilde{H}_k-\int_0^{2^k}e^{tA^*}Ce^{tA}dt\right\|\leqslant 14\sqrt{N}\varkappa^{*3/2}\|C\|.$$

Прежде чем переходить к доказательству теоремы 4, введем обозначения

$$\xi_p=B_{k_1}\dots B_{k_s}-e^{pA};\quad p=2^{k_1}+2^{k_2}+\dots+2^{k_s};\quad 0\leqslant\tau\leqslant 1;$$

$$\zeta_p(\tau)=(e^{\tau A}\xi_p)^*Ce^{\tau A}\xi_p+(e^{\tau A}\xi_p)^*Ce^{(\tau+p)A}+e^{(\tau+p)A^*}Ce^{\tau A}\xi_p;$$

$$\eta_q=B_{n_1}\dots B_{n_j}-e^{qA};\quad q=2^{n_1}+\dots+2^{n_j};$$

$$k-1\geqslant k_s>k_{s-1}>\dots>k_1\geqslant 0;\quad k-1\geqslant n_j>n_{j-1}>\dots>n_1\geqslant m.$$

В силу следствия леммы 1 имеют место неравенства

$$\|e^{\tau A}\xi_p\|\leqslant\delta_1(t)\exp\{-t/(2\varkappa^*)\};\quad \|e^{\tau' A}\eta_q\|\leqslant\delta_1(t')\exp\{-t'/(2\varkappa^*)\},\quad (7.6)$$

где $\delta_1(\tilde{t})=(\tilde{t}-1)r_0/(1-(\tilde{t}-1)r_0)$; $t'=q+\tau'$; $t=p+\tau$; $1>\tau$, $\tau'\geqslant 0$. Отметим, что при $0<(\tilde{t}-1)r_0\leqslant 1/2$ выполнена оценка

$$\delta_1^2(t)+2\delta_1(t)\leqslant 6(t-1)r_0.\quad (7.7)$$

Перейдем к доказательству теоремы, используя введенные обозначения. Из ее условия следует равенство

$$\widehat{H}_k=\widehat{H}_0+\sum_{p=1}^{2^k-1}(\xi_p+e^{pA})^*\widehat{H}_0(\xi_p+e^{pA})+$$

$$+\sum_{m=0}^{k-1}\sum_{k-1\geqslant n_j>\dots>n_1\geqslant m}\sum_{j=1}^{k-1-m}B_{n_j}^*\cdot\dots\cdot B_{n_1}^*\varphi_m B_{n_1}\cdot\dots\cdot B_{n_j}=$$

$$=\int_0^{2^k}e^{tA^*}Ce^{tA}dt+\sum_{p=0}^{k-1}\int_0^1\zeta_p(\tau)\,d\tau+\varphi_k+$$

$$+\sum_{m=0}^{k-1}\left\{\varphi_m+\sum_{q=2^m}^{2^k-2^m}(e^{qA}+\eta_q)^*\varphi_m(e^{qA}+\eta_q)\right\}.\quad (7.8)$$

Применяя неравенства (7.6) и (7.7), можем записать, что

$$\dots\left\|\int_0^1\zeta_p(\tau)\,d\tau\right\|\leqslant\|C\|\int_0^1\left[\delta_1^2(\tau+p)+2\delta_1(\tau+p)\right]\exp\left[-(\tau+p)/\varkappa^*\right]d\tau\leqslant$$

$$\leqslant\|C\|\int_0^1 6(\tau+p-1)r_0\exp\left[-(\tau+p)/\varkappa^*\right]d\tau=$$

$$=\|C\|\int_p^{p+1}6(t-1)r_0\exp\{-t/\varkappa^*\}\,dt.$$

Полученное неравенство подтверждает справедливость оценки

$$\left\|\sum_{p=1}^{2^k-1}\int_0^1 \zeta_p(\tau)\,d\tau\right\| \leqslant \|C\| \int_1^{2^k} 6(t-1)\,r_0 \exp(-t/\varkappa^*)\,dt =$$

$$= \|C\|\{6r_0\varkappa^{*2}\exp(-1/\varkappa^*) - [6r_0(2^k-1)\varkappa^* + 6r_0\varkappa^{*2}]\exp(-2^k/\varkappa^*)\}. \tag{7.9}$$

Вывод неравенств, оценивающих оставшиеся слагаемые в (7.8), основывается на следующей цепочке:

$$\left\|\varphi_m + \sum_{q=2^m}^{2^k-2^m} (e^{qA}+\eta_q)^*\varphi_m(e^{qA}+\eta_q)\right\| \leqslant$$

$$\leqslant \|\varphi_m\|\left\{1 + \sum_{q=2^m}^{2^k-2^m} [\delta_1(q) + \sqrt{\varkappa^*}]^2 \exp(-q/\varkappa^*)\right\} \leqslant$$

$$\leqslant \|\varphi_m\|\left[1 + \sum_{q=2^m}^{2^k-2^m} 4\varkappa^* \exp(-q/\varkappa^*)\right].$$

В самом деле, в силу (7.5) верна цепочка неравенств

$$\left\|\sum_{m=0}^{k-1}\left[\varphi_m + \sum_{q=2^m}^{2^k-2^m} (e^{qA}+\eta_q)^*\varphi_m(e^{qA}+\eta_q)\right]\right\| \leqslant$$

$$\leqslant \sum_{m=0}^{k-1}\left[1 + \sum_{q=2^m}^{2^k-2^m} 4\varkappa^* \exp(-q/\varkappa^*)\right]\alpha\varkappa^*\|C\| \leqslant$$

$$\leqslant \|C\|\alpha\varkappa^* k[1 + 4\varkappa^{*2}\exp(-1/\varkappa^*)],$$

завершающая вместе с (7.8) и (7.9) доказательство теоремы 4.

§ 8. АЛГОРИТМ РАСЧЕТА ПОЛОЖИТЕЛЬНО ОПРЕДЕЛЕННЫХ РЕШЕНИЙ УРАВНЕНИЯ ЛЯПУНОВА

Опишем детальный алгоритм расчета положительно определенных решений матричного уравнения Ляпунова

$$A^*H + HA + I = 0, \tag{8.1}$$

все этапы которого подробно рассматривались в предыдущих параграфах. Особое внимание уделим арифметике процесса. Ранее (см. [7, 8]) было предложено для реализации матричных процессов использовать «арифметику вынесенных порядков». Для ее формального описания введем операторы $\mathfrak{M}$, $\mathscr{P}$, $\mathfrak{m}$ и p.

Пусть γ — основание системы счисления используемой ЭВМ. Определим для каждого заданного в машине числа α пару чисел $\mathfrak{m}(\alpha)$ и $p(\alpha)$ таких, чтобы выполнялись условия $1/\gamma \leqslant |\mathfrak{m}(\alpha)| < 1$, $p(\alpha)$ — целое, $\alpha = \mathfrak{m}(\alpha)\cdot\gamma^{p(\alpha)}$. Назовем $[\alpha^0, \alpha^1]$ канонической парой, представляющей число α, если $1/\gamma \leqslant |\alpha^0| < 1$, α^1 — целое и $\alpha = \alpha^0\gamma^{\alpha^1}$ (α, равное 0, задается парой [0, 0]). Операторы $\mathfrak{m}$ и p позволяют по каждому заданному в машине числу α вычислить его каноническую пару $[\mathfrak{m}(\alpha), p(\alpha)]$. Аналогично для матрицы X определяется каноническая пара $[X^0, X^1]$ такая, что $1/\gamma \leqslant \max\limits_{i,j} |X^0_{ij}| < 1$, X^1 — целое число и $X = \gamma^{X^1}\cdot X^0$. Нулевой матрице соответствует пара [0, 0]. Определим на пространстве матриц операторы $\mathfrak{M}$ и $\mathscr{P}$ такие, что $\mathfrak{M}(X) = X^0$ и $\mathscr{P}(X) = X^1$.

Ниже перечислены основные этапы алгоритма, выполняемые последовательно один за другим. Отметим, что при реализации схемы § 3 $\rho = 1$.

1°. Входные данные. В машину вводятся следующие величины: N — целое число; $A=\{A_{ij}\}$ — исходная матрица размерности $N\times N$; ε_1 — машинная постоянная; ρ_0 — вещественное — требуемая точность определения решения.

2°. Вычисляется значение параметра практической устойчивости: $\varkappa_{\text{гр}}=(196N\varepsilon_1^2)^{-1/7}$.

3°. Находятся максимальное ($\sigma_N(A)$) и минимальное ($\sigma_1(A)$) сингулярные числа матрицы A.

4°. Проверяется неравенство $\sigma_N(A)>[\varkappa_{\text{гр}}]^{3/2}\sigma_1(A)$. Если оно нарушено, то делается заключение о практической неустойчивости A и результатом работы процесса является неравенство $\varkappa(A)>\varkappa_{\text{гр}}$. В противном случае матрица нормируется: $A_1=1/(2\|A\|)A$.

5°. Определяется величина $r_0=2{,}02[\varkappa_{\text{гр}}]^{3/2}\sqrt{N}\varepsilon_1$.

6°. Опишем алгоритм вычисления матрицы

$$B_0=I+\frac{1}{1!}A_1+\frac{1}{2!}A_1^2+\ldots+\frac{1}{k_2!}A_1^{k_2}.$$

6°.1. Получаем целое число k_2 как минимальное k, для которого справедливо неравенство $2{,}02\sqrt{N}(k+0{,}5)\varepsilon_1\geqslant 1/(k+1)!$.

6°.2. Вычисление B_0 осуществляется по рекуррентным формулам, стандартный шаг которых описан в 6°.3. Для начала процесса задаются матрицы $[U_1^0, U_1^1]=[I+A_1, 0]$; $[V_1^0, V_1^1]=[A_1, 0]$. Напомним, что означают эти формальные равенства ($i, j=1, 2, \ldots, N$):

$$(U_1^0)_{ij}=\begin{cases}(A_1)_{ij}, \text{ если } i\neq j; & (V_1^0)_{ij}=(A_1)_{ij};\\ 1+(A_1)_{ii}, \text{ если } i=j; & V_1^1=U_1^1=0.\end{cases}$$

6°.3. Опишем k-й шаг ($k=2, 3, \ldots, k_2$) процесса вычисления матрицы B_0. Пусть канонические пары $[U_{k-1}^0, U_{k-1}^1]$, $[V_{k-1}, V_{k-1}^1]$ получены на предыдущем шаге.

6°.3.1. Находится матрица V_k ($i, j=1, 2, \ldots, N$):

$$(\tilde{V}_k)_{ij}=\sum_{l=1}^{N}(A_1)_{il}(V_{k-1}^0)_{lj}/k;$$

$$V_k^0=\mathfrak{M}(\tilde{V}_k);\quad V_k^1=\mathscr{P}(\tilde{V}_k)+V_{k-1}^1.$$

Здесь каждая из N^2 сумм накапливается с двойной точностью.

6°.3.2. Вычисляется матрица U_k:

$$q=\max\{U_{k-1}^1, V_k^1\};\quad i, j=1, 2, \ldots, N;$$

$$(\tilde{U}_k)_{ij}=\gamma^{U_{k-1}^1-q}(U_{k-1}^0)_{ij}+\gamma^{V_k^1-q}(V_k^0)_{ij};$$

$$U_k^0=\mathfrak{M}(\tilde{U}_k);\quad U_k^1=\mathscr{P}(\tilde{U}_k)+q.$$

После завершения процесса определена выходная информация пункта 6°.3: $[B_0^0, B_0^1]=[U_{k_2}^0, U_{k_2}^1]$ — каноническая пара, представляющая матрицу B_0.

7°. Опишем алгоритм вычисления матрицы

$$H_0=I+\frac{1}{2!}\mathfrak{L}I+\ldots+\frac{1}{k_3!}\mathfrak{L}^{k_3}I,$$

где $\mathfrak{L}$ — оператор Ляпунова.

7°.1. Определяется целое число k_3 как минимальное k, для которого справедливо неравенство $2{,}02\sqrt{N}(1+k)\varepsilon_1\geqslant 1/(k+1)!$.

7°.2. Вычисление H_0 осуществляется по рекуррентным формулам, стандартный шаг которых описан в 7°.3. Для начала процесса задаются матрицы $[G_1^0, G_1^1]=[I, 0]$; $[Q_1^0, Q_1^1]=[I, 0]$.

7°.3. Опишем k-й шаг ($k=2, 3, \ldots, k_3$) процесса вычисления H_0. Пусть канонические пары $[G_{k-1}^0, G_{k-1}^1]$, $[Q_{k-1}^0, Q_{k-1}^1]$ получены на предыдущем шаге.

7°.3.1. Находится матрица G_k. С двойной точностью накапливаются суммы $(i, j = 1, 2, \ldots, N)$:

$$(\tilde{G}_k)_{ij} = \sum_{m=1}^{N} (G^0_{k-1})_{im} (A_1)_{mj}.$$

Затем определяются матрицы

$$\tilde{\tilde{G}}_k = \frac{1}{k}[\tilde{G}^*_k + \tilde{G}_k]; \quad G^0_k = \mathfrak{M}\left(\tilde{\tilde{G}}_k\right); \quad G^1_k = \mathscr{P}\left(\tilde{\tilde{G}}_k\right) + G^1_{k-1}.$$

7°.3.2. Вычисляется матрица Q_k:

$$q = \max\{Q^1_{k-1}, G^1_k\}; \quad i, j = 1, 2, \ldots, N;$$

$$(\tilde{Q}_k)_{ij} = \gamma^{Q^1_{k-1}-q}(Q^0_{k-1})_{ij} + \gamma^{G^1_k-q}(G^0_k)_{ij};$$

$$Q^0_k = \mathfrak{M}(\tilde{Q}_k); \quad Q^1_k = q + \mathscr{P}(\tilde{Q}_k).$$

После завершения процесса определена выходная информация пункта 7°.3: $[H^0_0, H^1_0] = [Q^1_{k_3}, Q^1_{k_3}]$ — каноническая пара, представляющая матрицу H_0.

8°. Находится целое число $\tilde{k}$, удовлетворяющее неравенству $2^{\tilde{k}-1} \leqslant \leqslant \varkappa_{\text{гр}} \ln[4\varkappa_{\text{гр}}] \leqslant 2^{\tilde{k}}$.

9°. Вычисляется приближение к матрице $\int_0^{2^{\tilde{k}}} e^{tA^*} e^{tA} dt$. Для этого используются рекуррентные соотношения, приведенные в § 4. Начальные значения для последовательного получения матриц, приближающих искомую, найдены в п. 6° и 7°: $[B^0_0, B^1_0]$, $[H^0_0, H^1_0]$. Здесь опишем k-й шаг $(k = 1, 2, \ldots, \tilde{k})$ процесса вычисления матрицы $H_{\tilde{k}}$.

Пусть матричные пары $[B^0_{k-1}, B^1_{k-1}]$, $[H^0_{k-1}, H^1_{k-1}]$ получены на предыдущем шаге и представляют матрицы B_{k-1} и H_{k-1}.

9°.1. Определяется последовательно по столбцам матрица $B^*_{k-1}H_{k-1}B_{k-1}$. Опишем вычисление m-го столбца $(1 \leqslant m \leqslant N)$.

9°.1.1. Пусть $(y_m)_l = (B^0_{k-1})_{lm}$ $(l = 1, 2, \ldots, N)$. С двойной точностью накапливаются суммы

$$(\tilde{y}_m)_i = \sum_{l=1}^{N} (H^0_{k-1})_{il} (y_m)_l.$$

По технологии, предложенной в § 5, для использования на этом этапе необходимо для запоминания вектора $\tilde{y}_m$ отвести два массива длины N.

9°.1.2. Найдем вектор $(i = 1, 2, \ldots, N)$

$$\left(\tilde{\tilde{y}}_m\right)_i = \sum_{l=1}^{N} (B^0_{k-1})_{li} (\tilde{y}_m)_l,$$

где сумма накапливается с двойной точностью, а вычисление произведения $(B^0_{k-1})_{li}(\tilde{y}_m)_l$ требует учесть, что вектор $\tilde{y}_m$ хранится в двух массивах длины N. Определенный вектор $\tilde{\tilde{y}}_m$ для хранения и использования заносят в m-й столбец матрицы H_k. Пара $[\mathfrak{M}(H_k), \mathscr{P}(H_k) + 2B^1_{k-1}]$ задает искомую матрицу $B^*_{k-1}H_{k-1}B_{k-1}$.

9°.2. Вычисляется матрица $H_k = H_{k-1} + B^*_{k-1}H_{k-1}B_{k-1}$ и число $h_k = \|H_k\|_E$ $(i, j = 1, 2, \ldots, N)$:

$$\left(\tilde{\tilde{H}}_k\right)_{ij} = (H^0_{k-1})_{ij}\gamma^{-2B^1_{k-1}} + (\tilde{H}_k)_{ij};$$

$$\tilde{h}_k = \sum_{i,j=1}^{N} \left(\tilde{\tilde{H}}_k\right)^2_{ij}; \quad h_k = \tilde{h}_k \gamma^{2B^1_{k-1}+H^1_{k-1}};$$

$$H^0_k = \mathfrak{M}\left(\tilde{\tilde{H}}_k\right); \quad H^1_k = \mathscr{P}\left(\tilde{\tilde{H}}_k\right) + 2B^1_{k-1} + H^1_{k-1}.$$

9°.3. Проверяется неравенство $h_k < (1/4 + \varkappa_{\text{гр}})\sqrt{N}$. Если оно нарушено, то процесс заканчивается утверждением $\varkappa(A) > \varkappa_{\text{гр}}$, так как в этом случае гарантировано выполнение неравенства $\|H_k\| > 1/4 + \varkappa_{\text{гр}}$ (см. этап III п. 3° § 3). Здесь вместо спектральной нормы вычисляется фробениусова норма матрицы, равная квадратному корню из суммы квадратов ее элементов, так как это более «дешевая» операция.

9°.4. Определяются матрица $B_k = B_{k-1}^2$ и число $b_k = \|B_k\|_E$. С двойной точностью накапливаются суммы ($i, j = 1, 2, \ldots, N$):

$$(\tilde{B}_k)_{ij} = \sum_{l=1}^{N} (B_{k-1}^0)_{il} (B_{k-1}^0)_{lj};$$

$$b_k = \left(\sum_{i,j=1}^{N} (\tilde{B}_k)_{ij}^2 \right)^{1/2} \gamma^{2B_{k-1}^1}; \quad B_k^1 = \mathscr{P}(\tilde{B}_k) + 2B_{k-1}^1; \quad B_k^0 = \mathfrak{M}(\tilde{B}_k).$$

9°.5. Проверяется неравенство

$$b_k \leqslant (\varkappa_{\text{гр}} + 1)\sqrt{N} \exp\{-2^{k-1}/\varkappa_{\text{гр}}\}.$$

Если оно нарушено, то процесс заканчивается утверждением $\varkappa(A) > \varkappa_{\text{гр}}$, так как при этом гарантировано выполнение неравенства

$$\|B_k\| > (\varkappa_{\text{гр}} + 1) \exp\{-2^{k-1}/\varkappa_{\text{гр}}\}$$

(см. этап II п. 2° § 3).

Итак, если за $\tilde{k}$ шагов не произошло прерывания процесса с сообщением $\varkappa(A) > \varkappa_{\text{гр}}$, то в нашем распоряжении в качестве выходной информации п. 9° будет матричная пара $[H_{\tilde{k}}^0, H_{\tilde{k}}^1]$, задающая матрицу $[H_{\tilde{k}}]$.

10°. Находится матрица C_1 — невязка полученного приближенного решения. С двойной точностью накапливаются суммы ($i, j = 1, 2, \ldots, N$):

$$(\tilde{C}_1)_{ij} = \sum_{m=1}^{N} \left(H_{\tilde{k}}^0\right)_{im} (A_1)_{mj}.$$

Затем определяются матрицы

$$\tilde{\tilde{C}}_1 = \tilde{C}_1^* + \tilde{C}_1 + \gamma^{-H_{\tilde{k}}^1} I; \quad C_1^0 = \mathfrak{M}\left(\tilde{\tilde{C}}_1\right); \quad C_1^1 = \mathscr{P}\left(\tilde{\tilde{C}}_1\right) + H_{\tilde{k}}^1,$$

где I — единичная матрица.

11°. Вычисляются Λ — максимальное, λ — минимальное собственные числа матрицы $H_{\tilde{k}}$, а также $\sigma(C_1)$ — спектральная норма матрицы C_1.

12°. Проверяются неравенства

$$\lambda \geqslant 1/2; \quad \sigma(C_1) \leqslant 1/2; \quad \varkappa_{\text{гр}} \leqslant \Lambda - 1/2.$$

Если хотя бы одно из них не выполнено, то процесс завершается и его результатом является неравенство $\varkappa(A) > \varkappa_{\text{гр}}$. В противном случае гарантировано, что

$$\left\| A_1^* H_{\tilde{k}} + H_{\tilde{k}} A_1 + I \right\| \leqslant \sigma(C_1) \leqslant 1/2;$$

$$\lambda_{\min}(H_{\tilde{k}}) > 0; \quad |\varkappa(A) - \Lambda| / \varkappa(A) \leqslant 1/2.$$

13°. Проверяется неравенство $\sigma(C_1) \leqslant \rho_0$. Если оно выполнено, то получаем положительно определенную матрицу приближенного решения уравнения Ляпунова (8.1)

$$\tilde{H}_{ij} = \gamma^{H_{\tilde{k}}^1} \Big/ (2\sigma_N(A)) \cdot \left(H_{\tilde{k}}^0\right)_{ij}; \quad i, j = 1, 2, \ldots, N,$$

для которого выполнены оценки

$$\|\tilde{H} - H\| / \|H\| \leqslant \rho_0; \quad \|A^* \tilde{H} + \tilde{H} A + I\| \leqslant \rho_0; \quad |\varkappa(A) - \Lambda| / \varkappa(A) \leqslant \rho_0,$$

и процесс завершается. В противном случае управление передается п. 14° для уточнения решения.

14°. Описывается алгоритм итерационного уточнения полученного решения, стандартный шаг (k-й) которого приведен в п. 14°.1.

Для начала процесса имеются матричные пары $[C_1^0, C_1^1]$, $[\breve{H}_0^0, \breve{H}_0^1] = [H_k^0, H_k^1]$, найденные ранее.

14°.1. Предположим, что получены канонические матричные пары $[\breve{H}_{k-1}^0, \breve{H}_{k-1}^1]$ и $[C_k^0, C_k^1]$.

14°.1.1. Повторяются вычисления п. 6° и 7°, где в 7° вместо I берется матрица C_k^0. Затем выполняются п. 9°.1, 9°.2, 9°.4. В результате матричная пара $[H_k^0(k), H_k^1(k)]$ оказывается найденной.

14°.1.2. Вычисляется матрица C_{k+1} — невязка вновь полученного приближения. С двойной точностью накапливаются суммы ($i, j = 1, 2, \ldots, N$):

$$(\tilde{C}_{k+1})_{ij} = \sum_{m=1}^{N} (H_k^0(k))_{im} (A_1)_{mj}.$$

Затем определяются матрицы

$$\tilde{\tilde{C}}_{k+1} = \tilde{C}_{k+1}^* + \tilde{C}_{k+1} + \gamma^{-H_k^1(k)} \tilde{C}_k;$$

$$C_{k+1}^1 = \mathscr{P}(\tilde{\tilde{C}}_{k+1}) + H_k^1(k) + C_k^1; \quad C_{k+1}^0 = \mathfrak{M}(\tilde{\tilde{C}}_{k+1}).$$

14°.1.3. Находится новое приближение $\breve{H}_k$ к решению уравнения Ляпунова

$$[\breve{H}_k^0, \breve{H}_k^1] = [\breve{H}_{k-1}^0, \breve{H}_{k-1}^1] + 1/(2\sigma_N(A)) [H_k^0(k), H_k^1(k)].$$

14°.1.4. Вычисляется $\sigma_N(C_{k+1})$ — спектральная норма матрицы, определенной в п. 14°.1.2 в виде канонической пары $[C_{k+1}^0, C_{k+1}^1]$.

14°.1.5. Проверяется неравенство $\sigma_N(C_{k+1}) > \rho_0$. Если оно нарушено, то начинает работу п. 14°.1.6. В противном случае управление передается п. 14°.1 с изменением k на $k+1$.

14°.1.6. Вычисляется $\overline{\varkappa}(A)$ — максимальное собственное число матрицы $[\breve{H}_k^0, \breve{H}_k^1]$, найденной в п. 14°.1.3; дальше будем обозначать ее $\breve{H}$. На этом процесс заканчивается.

15°. Выходная информация. В результате работы алгоритма выдаются следующие данные: 1) максимальное $\sigma_N(A)$ и минимальное $\sigma_1(A)$ — сингулярные числа матрицы A; 2) $\breve{H} = \{\breve{H}_{ij}\}$ ($i, j = 1, 2, \ldots, N$) — положительно определенное решение уравнения Ляпунова $A^*H + HA + I = 0$. Для него выполнены оценки:

$$\|A^*\breve{H} + \breve{H}A + I\| \leqslant \rho_0; \quad \|\breve{H} - H\|/\|H\| \leqslant \rho_0;$$

3) $\overline{\varkappa}(A)$ — приближение к параметру качества устойчивости:

$$|\varkappa(A) - \overline{\varkappa}(A)|/\varkappa(A) \leqslant \rho_0.$$

Если выяснилось, что $\varkappa(A) > \varkappa_{\text{гр}}$, то информация 1) и 2) не определена.

§ 9. ПРИМЕР ИСПОЛЬЗОВАНИЯ

Значения входных параметров: $N = 4$; $\varepsilon_1 = 0.2_{10} - 14$; $\rho_0 = 0.1_{10} - 5$;

$$A_{ij} = \begin{cases} -1; & i = j, \quad i = 1, 2, 3, 4; \\ 2; & j = i + 1, \quad i = 1, 2, 3; \\ 0 & \text{при всех остальных } i, j. \end{cases}$$

Результаты расчета: $\varkappa(A) = 105.76$; $\varkappa_{\text{гр}} = 0.5_{10}4$;

$$\breve{H} = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1.5 & 2.0 & 2.5 \\ 0.5 & 2.0 & 4.5 & 7.0 \\ 0.5 & 2.5 & 7.0 & 14.5 \end{bmatrix}.$$

ЛИТЕРАТУРА

1. **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.
2. **Godounov S. K., Boulgakov A. J.** Difficultés de calcul dans le probleme de Hurwitz et méthodes pour les surmonter.— In: Analysis and optimization of Systems, Versailles, 1982.— Proceedings (Lecture Notes in Control and Information sciences, 44). Springer Verlag, 1982, p. 843—851.
3. **Davison E. J., Man F. T.** The numerical solution of $A^*Q + QA = -C$.— IEEE Trans. Automatic Control., 1968, v. 13, p. 448.
4. **Per Hagander.** Numerical solution of $A^*S + SA + Q = 0$.— Information Sciences, 1972, № 4, p. 45—50.
5. **Ла-Салль Ж., Лефшетц С.** Исследование устойчивости прямым методом Ляпунова.— М.: Мир, 1964.— 168 с.
6. **Карачаров К. А., Пилютик А. Г.** Введение в техническую теорию устойчивости движения.— М.: Физматгиз, 1962.— 244 с.
7. **Булгаков А. Я.** Вычисление экспоненты от асимптотически устойчивой матрицы.— В кн.: Вычислительные методы линейной алгебры. Новосибирск: Наука, 1985, с. 4—17.
8. **Булгаков А. Я., Годунов С. К.** Численное определение одного из критериев качества устойчивости систем линейных дифференциальных уравнений с постоянными коэффициентами.— Новосибирск, 1981.— 58 с.— (Препринт/АН СССР, Сиб. отд-ние, ИМ).
9. **Годунов С. К.** Решение систем линейных уравнений.— Новосибирск: Наука, 1980.— 177 с.
10. **Форсайт Дж., Молер К.** Численное решение систем линейных алгебраических уравнений.— М.: Мир, 1969.— 167 с.

УЧЕТ ВЫЧИСЛИТЕЛЬНЫХ ПОГРЕШНОСТЕЙ В ОДНОМ ВАРИАНТЕ МЕТОДА СОПРЯЖЕННЫХ ГРАДИЕНТОВ

А. Я. БУЛГАКОВ, С. К. ГОДУНОВ

ВВЕДЕНИЕ

В работе рассматривается вариант метода сопряженных градиентов для решения операторного уравнения

$$\mathfrak{L}H = C, \qquad (1)$$

где $\mathfrak{L}$ — несамосопряженный линейный оператор, действующий в N-мерном евклидовом пространстве. Этот вариант использовался авторами при решении матричного уравнения Ляпунова [1]. Предлагаемый алгоритм — один из возможных вариантов метода сопряженных градиентов. Следует отметить, что разные варианты метода по-разному реагируют на возмущения, возникающие от приближенного выполнения арифметических операций на ЭВМ (см., например, [2—5]). Выбор конкретного варианта сделан с учетом ряда проделанных вычислительных экспериментов.

В первых двух параграфах описывается указанный алгоритм, для которого в предположении точного вычисления всех формул выведены оценки уменьшения нормы невязки за первые k шагов и конкретно за k-й шаг. В процессе работы алгоритма строится базис, в котором оператор $\mathfrak{L}$ записывается в виде произведения ортогональной и двухдиагональ-