5.41. Пусть A — матрица размерности $n \times n$, $\rho(A)$ — ее спектральный радиус и задано число $\varepsilon > 0$. Доказать, что существует по крайней мере одна матричная норма, для которой имеют место оценки

$$\rho(A) \leqslant \|A\| \leqslant \rho(A) + \varepsilon.$$

$\lhd$ Из курса линейной алгебры (теорема Шура об унитарной триангуляции) известно, что найдутся такие унитарная матрица $U\,(U^* = U^{-1})$ и верхняя треугольная матрица R, что $A = U\,R\,U^*$. Положим $D_t = \mathrm{diag}\,(t, t^2, t^3, \ldots, t^n)$ и вычислим матрицу

$$D_t R D_t^{-1} = \begin{pmatrix} \lambda_1 & t^{-1} r_{12} & t^{-2} r_{13} & \ldots & t^{-n+1} r_{1n} \\ 0 & \lambda_2 & t^{-1} r_{23} & \ldots & t^{-n+2} r_{2n} \\ 0 & 0 & \lambda_3 & \ldots & t^{-n+3} r_{3n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{pmatrix}.$$

При достаточно большом $t > 0$ сумма модулей наддиагональных элементов матрицы $D_t R D_t^{-1}$ не превосходит ε. В частности, это приводит к неравенству $\|D_t R D_t^{-1}\|_1 \leqslant \rho(A) + \varepsilon$. Теперь определим матричную норму с помощью формулы

$$\|A\| = \|D_t U^* A\, U\, D_t^{-1}\|_1 = \|\left(U\, D_t^{-1}\right)^{-1} A\, \left(U\, D_t^{-1}\right)\|_1.$$

Таким образом, выбор достаточно большого t в приведенной выше формуле приводит к оценке сверху, а оценка снизу следует из 5.6. $\rhd$

5.2. Элементы теории возмущений

Рассмотрим систему линейных алгебраических уравнений

$$A\,\mathbf{x} = \mathbf{b}$$

с квадратной невырожденной матрицей A. При ее решении в результате вычислений с конечной разрядностью вместо $\mathbf{x}$ получается *приближенное* решение $\tilde{\mathbf{x}}$, которое можно рассматривать как *точное* решение *возмущенной* системы

$$(A + \delta A)\,\tilde{\mathbf{x}} = \mathbf{b},$$

где матрица возмущений δA мала в каком-либо смысле.

Другой источник ошибок в $\tilde{\mathbf{x}}$ определяется возмущениями δA и $\delta \mathbf{b}$ в элементах матрицы A и в компонентах вектора правой части $\mathbf{b}$ (например, вследствие ошибок округлений, возникающих в процессе ввода вещественных чисел в память компьютера).

Чтобы оценить насколько приближенное решение $\tilde{\mathbf{x}}$ отличается от точного решения $\mathbf{x}$, используют нормы векторов и подчиненные нормы матриц.

Пусть в системе $A\mathbf{x} = \mathbf{b}$ возмущается только вектор $\mathbf{b}$, т. е. вместо исходной системы решается возмущенная система $A\tilde{\mathbf{x}} = \tilde{\mathbf{b}} \equiv \mathbf{b} + \delta\mathbf{b}$, и пусть $\tilde{\mathbf{x}}$ — точное решение возмущенной системы. Тогда для относительной ошибки в $\tilde{\mathbf{x}}$ верна оценка

$$\frac{||\mathbf{x}-\tilde{\mathbf{x}}||}{||\mathbf{x}||} \leqslant ||A||\,||A^{-1}||\ \frac{||\mathbf{b}-\tilde{\mathbf{b}}||}{||\mathbf{b}||} = ||A||\,||A^{-1}||\ \frac{||\mathbf{b}-A\,\tilde{\mathbf{x}}||}{||\mathbf{b}||}\ .$$

Величину $||A||\,||A^{-1}||$ называют *числом обусловленности* матрицы A и часто обозначают $\mathrm{cond}(A)$. Для вырожденных матриц $\mathrm{cond}(A)=\infty$. Конкретное значение $\mathrm{cond}(A)$ зависит от выбора матричной нормы, однако в силу их эквивалентности при практических оценках этим различием можно пренебречь.

Из приведенного выше неравенства следует, что даже если *вектор невязки* $\mathbf{r}=\mathbf{b}-A\tilde{\mathbf{x}}$ мал, относительные возмущения в решении могут быть большими, если $\mathrm{cond}(A)$ велико (такие матрицы называют *плохо обусловленными*).

5.42. Доказать неравенство

$$\frac{||\mathbf{x}-\tilde{\mathbf{x}}||}{||\mathbf{x}||} \leqslant \mathrm{cond}(A)\ \frac{||\mathbf{b}-\tilde{\mathbf{b}}||}{||\mathbf{b}||} = \mathrm{cond}(A)\ \frac{||\mathbf{b}-A\,\tilde{\mathbf{x}}||}{||\mathbf{b}||}\ .$$

$\triangleleft$ Из равенства $A^{-1}\mathbf{r}=A^{-1}\mathbf{b}-A^{-1}A\tilde{\mathbf{x}}=\mathbf{x}-\tilde{\mathbf{x}}$ следует, что

$$||\mathbf{x}-\tilde{\mathbf{x}}|| \leqslant ||A^{-1}||\,||\mathbf{r}||. \tag{5.1}$$

Из равенства $\mathbf{b}=A\mathbf{x}$ имеем, что $||\mathbf{b}||=||A\mathbf{x}||\leqslant||A||\,||\mathbf{x}||$, т. е.

$$||\mathbf{x}|| \geqslant \frac{||\mathbf{b}||}{||A||}\,. \tag{5.2}$$

Разделив неравенство (5.1) на неравенство (5.2), получим

$$\frac{||\mathbf{x}-\tilde{\mathbf{x}}||}{||\mathbf{x}||} \leqslant ||A||\,||A^{-1}||\ \frac{||\mathbf{r}||}{||\mathbf{b}||} = \mathrm{cond}(A)\ \frac{||\mathbf{r}||}{||\mathbf{b}||} = \mathrm{cond}(A)\ \frac{||\mathbf{b}-A\tilde{\mathbf{x}}||}{||\mathbf{b}||}\,.$$

Отсюда видно, что если матрица A плохо обусловлена, то даже очень маленькая невязка не может гарантировать малость относительной ошибки в $\tilde{\mathbf{x}}$. С другой стороны, может оказаться так, что достаточно точное решение имеет большую невязку. Рассмотрим пример:

$$A=\begin{pmatrix}1,000 & 1,001\\ 1,000 & 1,000\end{pmatrix},\qquad \mathbf{b}=\begin{pmatrix}2,001\\ 2,000\end{pmatrix}$$

Точное решение системы $A\mathbf{x}=\mathbf{b}$ имеет вид $\mathbf{x}=(1,1)^T$. Однако вектор $\tilde{\mathbf{x}}=(2,0)^T$, который никак нельзя назвать близким к $\mathbf{x}$, дает маленькую невязку $\mathbf{r}=(10^{-3},0)^T$.

Возьмем теперь $\mathbf{b}=(1,0)^T$. Тогда вектор $\mathbf{x}=(-1000,1000)^T$ — точное решение системы. Вектор $\tilde{\mathbf{x}}=(-1001,1000)^T$ достаточно близок к $\mathbf{x}$ в смысле относительной погрешности, однако $\tilde{\mathbf{x}}$ дает большую невязку $\mathbf{r}=(1,1)^T$, близкую по норме к вектору $\mathbf{b}$. $\triangleright$

5.43. Найти решения двух систем с близкими коэффициентами:

$$\begin{cases} x+3y & = & 4\,, \\ x+3,00001y & = & 4,00001\,; \end{cases} \qquad \begin{cases} x+3y & = & 4\,, \\ x+2,99999y & = & 4,00001 \end{cases}$$

и объяснить полученный результат.

Ответ: $(1,1)^T$ и $(7,-1)^T$. Обе матрицы получены малыми возмущениями одной вырожденной матрицы. В данном случае это приводит к большой разнице в решениях систем.

5.44. Показать, что $\mathrm{cond}(A) \geqslant 1$ для любой матрицы A и $\mathrm{cond}_2(Q) = 1$ для ортогональной матрицы Q.

$\lhd$ Так как $I = AA^{-1}$, то

$$1 = ||I|| = ||A\,A^{-1}|| \leqslant ||A||\,||A^{-1}|| = \mathrm{cond}(A)\,.$$

Умножение матрицы на ортогональную не меняет ее спектральную норму, поэтому

$$||Q||_2 = ||QI||_2 = ||I||_2 = 1 \quad \text{и} \quad ||Q^T||_2 = ||Q^TI||_2 = ||I||_2 = 1\,.$$

Таким образом, $\mathrm{cond}_2(Q) = ||Q||_2\,||Q^{-1}||_2 = ||Q||_2\,||Q^T||_2 = 1\,.$ $\rhd$

5.45. Можно ли утверждать, что если определитель матрицы мал, то матрица плохо обусловлена?

$\lhd$ Пусть дана диагональная матрица $D = \varepsilon I$, где $\varepsilon > 0$ — малое число и I — единичная матрица. Определитель $\det(D) = \varepsilon^n$ мал, тогда как матрица D хорошо обусловлена, поскольку

$$\mathrm{cond}(D) = ||D||\,||D^{-1}|| = \varepsilon||I||\,\varepsilon^{-1}||I^{-1}|| = 1\,.$$

Рассмотрим теперь матрицу

$$A = \begin{pmatrix} 1 & -1 & -1 & \dots & -1 \\ 0 & 1 & -1 & \dots & -1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

у которой определитель равен 1, и вычислим ее число обусловленности. Для этого возьмем произвольный вектор $\mathbf{b} \neq 0$ и, решая систему $A\mathbf{x} = \mathbf{b}$ с помощью обратной подстановки, построим элементы обратной матрицы A^{-1}:

$$\begin{aligned} x_n &= b_n, \\ x_{n-1} &= b_{n-1} + b_n, \\ x_{n-2} &= b_{n-2} + b_{n-1} + 2b_n, \\ x_{n-3} &= b_{n-3} + b_{n-2} + 2b_{n-1} + 2^2 b_n, \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ x_1 &= b_1 + b_2 + 2b_3 + \dots + 2^{n-3}b_{n-1} + 2^{n-2}b_n. \end{aligned}$$

Запишем полученную обратную матрицу:

$$A^{-1} = \begin{pmatrix} 1 & 1 & 2 & 4 & \dots & 2^{n-3} & 2^{n-2} \\ 0 & 1 & 1 & 2 & \dots & 2^{n-4} & 2^{n-3} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

Следовательно, $\|A^{-1}\|_\infty = 1 + 1 + 2 + 2^2 + \dots + 2^{n-2} = 2^{n-1}$. Так как $\|A\|_\infty = n$, то $\mathrm{cond}_\infty(A) = n\, 2^{n-1}$, т. е. матрица A плохо обусловлена, хотя $\det(A) = 1$.

Рассмотренные примеры показывают, что обусловленность матрицы зависит не только от величины определителя. ▷

5.46. Пусть дана матрица A размерности $n \times n$ с параметром $|a| \neq 1$

$$A = \begin{pmatrix} 1 & a & 0 & \dots & 0 & 0 \\ 0 & 1 & a & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & a \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

Вычислить $\mathrm{cond}_\infty(A)$ и оценить возмущение в компоненте x_1 решения системы $A\mathbf{x} = \mathbf{b}$, если компонента b_n вектора $\mathbf{b}$ возмущена на ε.

◁ Как в 5.45, методом обратной подстановки получим обратную матрицу

$$A^{-1} = \begin{pmatrix} 1 & -a & (-a)^2 & \dots & (-a)^{n-2} & (-a)^{n-1} \\ 0 & 1 & -a & \dots & (-a)^{n-3} & (-a)^{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & -a \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

Тогда

$$\|A\|_\infty = 1 + |a|,$$

$$\|A^{-1}\|_\infty = 1 + |a| + a^2 + \dots + |a|^{n-1} = \frac{|a|^n - 1}{|a| - 1},$$

$$\mathrm{cond}_\infty(A) = \frac{(|a| + 1)\,(|a|^n - 1)}{|a| - 1}.$$

Отсюда видно, что матрица A плохо обусловлена при $|a| > 1$ и хорошо обусловлена при $|a| < 1$. Например, при $n = 20$ и $a = 5$ имеем $\mathrm{cond}_\infty(A) \approx 10^{14}$.

Пусть компонента b_n задана с ошибкой ε. Тогда вычисленное значение $\tilde{x}_1$ компоненты x_1 имеет вид

$$\tilde{x}_1 = b_1 - ab_2 + \dots + (-a)^{n-2} b_{n-1} + (-a)^{n-1}(b_n + \varepsilon) = x_1 + (-a)^{n-1}\varepsilon\,.$$

Следовательно, при $|a| > 1$ возмущение в b_n увеличивается в компоненте x_1 в $|a|^{n-1}$ раз, а при $|a| < 1$ во столько же раз уменьшается. ▷

5.47. Пусть A — матрица размерности $n \times n$ с элементами $a_{ij} = \{p$ для $i = j,\quad q$ для $i = j-1,\quad 0$ для остальных индексов$\}$. Вычислить матрицу A^{-1} и показать, что при $|q| < |p|$ матрица A хорошо обусловлена, а при $|q| > |p|$ и больших значениях n — плохо обусловлена.

Указание. Воспользоваться решением 5.46.

5.48. Пусть матрица A определена как в 5.47. Выразить явно решение системы $A\mathbf{x} = \mathbf{b}$ через правую часть.

Указание. Воспользоваться решением 5.46.

5.49. Решается система $A\mathbf{x} = \mathbf{b}$ с матрицей

$$A = \begin{pmatrix} 1 & -1 & 1 \\ -1 & \varepsilon & \varepsilon \\ 1 & \varepsilon & \varepsilon \end{pmatrix}, \quad |\varepsilon| \ll 1.$$

В результате замены $x_1' = x_1$, $x_2' = \varepsilon x_2$, $x_3' = \varepsilon x_3$ для нахождения новых неизвестных $\mathbf{x}'$ имеем систему $A'\mathbf{x}' = \mathbf{b}'$ с матрицей

$$A' = \begin{pmatrix} \varepsilon & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

В каком случае число обусловленности меньше?

◁ Имеем

$$A^{-1} = \begin{pmatrix} 0 & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1-\varepsilon}{4\varepsilon} & \frac{1+\varepsilon}{4\varepsilon} \\ \frac{1}{2} & \frac{1+\varepsilon}{4\varepsilon} & \frac{1-\varepsilon}{4\varepsilon} \end{pmatrix}, \quad (A')^{-1} = \begin{pmatrix} 0 & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1-\varepsilon}{4} & \frac{1+\varepsilon}{4} \\ \frac{1}{2} & \frac{1+\varepsilon}{4} & \frac{1-\varepsilon}{4} \end{pmatrix},$$

поэтому число обусловленности исходной матрицы стремится к бесконечности при $\varepsilon \to 0$, а число обусловленности матрицы A' остается ограниченным. ▷

5.50. Пусть $A = A^T > 0$, $\lambda(A) \in [m, M]$ и $A \neq \beta I$, где I — единичная матрица. Доказать, что $\mathrm{cond}_2(A + \alpha I)$ монотонно убывает по α при $\alpha > 0$.

Указание. Вычислить $\mathrm{cond}_2(A + \alpha I) = \dfrac{M+\alpha}{m+\alpha} = 1 + \dfrac{M-m}{m+\alpha}$.

5.51. Существуют ли несимметричные матрицы, для которых справедливо $\mathrm{cond}^2(A) = \mathrm{cond}(A^2) > 1$?

Ответ: примером такой матрицы является

$$A = \begin{pmatrix} 10^3 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 10^{-3} \end{pmatrix},$$

$$\lambda(A^T A) \in \left\{10^6, 10^{-6}, 4,5 \pm \sqrt{4,25}\right\},$$

$$\mathrm{cond}(A^2) = \|A^2\|\|A^{-2}\| = 10^{12}; \quad \mathrm{cond}(A) = \|A\|\|A^{-1}\| = 10^6.$$

5.52. Доказать неравенство для квадратных невырожденных матриц размерности $n \times n$
$$\frac{1}{n} \leqslant \frac{\mathrm{cond}_1(A)}{\mathrm{cond}_2(A)} \leqslant n.$$

$\lhd$ Воспользовавшись неравенством для векторных норм
$$||\mathbf{x}||_2 \leqslant ||\mathbf{x}||_1 \leqslant \sqrt{n}\,||\mathbf{x}||_2,$$
получим $\frac{1}{\sqrt{n}}\,||A||_2 \leqslant ||A||_1 \leqslant \sqrt{n}\,||A||_2$, откуда и следует требуемый результат. $\rhd$

5.53. Оценить снизу и сверху $\mathrm{cond}_\infty(A)$ невырожденной матрицы A размерности $n \times n$, используя границы собственных чисел матрицы $A^T A$: $\lambda(A^T A) \in [\alpha, \beta]$.

$\lhd$ Из неравенства для чисел обусловленности в матричных нормах $\|\cdot\|_\infty$ и $\|\cdot\|_2$ и равенства $\|A\|_2^2 = \lambda_{\max}(A^T A)$ следует, что
$$\frac{1}{n}\,\sqrt{\frac{\beta}{\alpha}} \leqslant \mathrm{cond}_\infty(A) \leqslant n\,\sqrt{\frac{\beta}{\alpha}}. \qquad \rhd$$

5.54. Получить неравенство $\mathrm{cond}(A) \geqslant \left|\frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}\right|$ для произвольной невырожденной матрицы A и любой матричной нормы, используемой при определении числа обусловленности.

Указание. Воспользоваться решением 5.6.

5.55. Доказать, что $\mathrm{cond}(AB) \leqslant \mathrm{cond}(A)\,\mathrm{cond}(B)$ для любой заданной нормы в определении числа обусловленности и для любых невырожденных квадратных матриц.

5.56. Оценить $\mathrm{cond}_2(A)$ матрицы A размерности $n \times n$
$$A = \begin{pmatrix} 2 & -1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 \end{pmatrix}.$$

Указание. Воспользоваться для собственных векторов $\mathbf{y}^{(j)}$, $j = 1, \dots, n$, матрицы A явной формулой (см. 2.86): $\mathbf{y}_k^{(j)} = \sin\frac{\pi\,j\,k}{n+1}$. Соответствующие собственные числа: $\lambda^{(j)} = 4\sin^2\frac{\pi\,j}{2(n+1)}$, так что
$$\mathrm{cond}_2(A) = \mathrm{ctg}^2\,\frac{\pi}{2(n+1)} \approx \frac{4(n+1)^2}{\pi^2}\,.$$

5.57. Оценить $\mathrm{cond}_2(A)$ матрицы A размерности $n \times n$

$$A = \frac{1}{6}\begin{pmatrix} 4 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 4 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 4 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 4 & 1 \\ 0 & 0 & 0 & 0 & \dots & 1 & 4 \end{pmatrix}.$$

Указание. Для собственных векторов $\mathbf{y}^{(j)}$, $j = 1, \dots, n$ матрицы A найти явную формулу (см. 2.86). Получим $\mathbf{y}_k^{(j)} = \sin\dfrac{\pi\, j\, k}{n+1}$. Соответствующие собственные числа: $\lambda^{(j)} = \frac{2}{3} + \frac{1}{3}\cos\dfrac{\pi\, j}{n+1}$, так что $\mathrm{cond}_2(A) \leqslant 3$.

5.58. Пусть I — единичная матрица, δI — ее возмущение и $||\delta I|| < 1$. Показать, что матрица $I - \delta I$ невырожденная и выполнена оценка

$$||(I - \delta I)^{-1}|| \leqslant \frac{1}{1 - ||\delta I||}\,.$$

$\lhd$ Возьмем произвольный вектор $\mathbf{x} \neq 0$. Так как $1 - ||\delta I|| > 0$ и $||\mathbf{x}|| = ||(\mathbf{x} - \delta I\mathbf{x}) + \delta I\mathbf{x}|| \leqslant ||\mathbf{x} - \delta I\mathbf{x}|| + ||\delta I\mathbf{x}||$, то

$$\begin{gathered} ||(I - \delta I)\,\mathbf{x}|| = ||\mathbf{x} - \delta I\mathbf{x}|| \geqslant ||\mathbf{x}|| - ||\delta I\mathbf{x}|| \geqslant \\ \geqslant ||\mathbf{x}|| - ||\delta I||\,||\mathbf{x}|| = (1 - ||\delta I||)\,||\mathbf{x}|| > 0\,. \end{gathered}$$

Следовательно, если $\mathbf{x} \neq 0$, то $(I - \delta I)\,\mathbf{x} \neq 0$, т. е. матрица $I - \delta I$ невырожденная. Из тождества $(I - \delta I)\,(I - \delta I)^{-1} = I$ получаем $(I - \delta I)^{-1} = I + \delta I\,(I - \delta I)^{-1}$. Отсюда

$$||(I - \delta I)^{-1}|| \leqslant ||I|| + ||\delta I||\,||(I - \delta I)^{-1}|| = 1 + ||(I - \delta I)^{-1}||\,||\delta I||\,.$$

Из этого неравенства следует решение задачи (ее называют *задачей о возмущении единичной матрицы*). $\rhd$

5.59. Пусть I — единичная матрица, δI — ее возмущение и $||\delta I|| < 1$. Получить оценку отклонения матрицы I от матрицы $(I - \delta I)^{-1}$.

$\lhd$ Из $(I - \delta I)^{-1} = I + \delta I(I - \delta I)^{-1}$ (см. 5.58) получим $I - (I - \delta I)^{-1} = -\delta I(I - \delta I)^{-1}$. Отсюда в силу неравенства из 5.58

$$||I - (I - \delta I)^{-1}|| \leqslant ||\delta I||\,||(I - \delta I)^{-1}|| \leqslant \frac{||\delta I||}{1 - ||\delta I||}\,. \qquad \rhd$$

5.60. Пусть A — невырожденная матрица, δA — ее возмущение и $||A^{-1}\delta A|| < 1$. Показать, что матрица $A + \delta A$ невырожденная и выполнена оценка

$$||(A + \delta A)^{-1}|| \leqslant \frac{||A^{-1}||}{1 - ||A^{-1}\delta A||}\,.$$

$\lhd$ Имеем $A+\delta A = A(I+A^{-1}\delta A)$. Поскольку $||A^{-1}\delta A|| < 1$, из 5.58 следует, что матрица $I + A^{-1}\delta A$ невырожденная. Это означает, что и матрица $A + \delta A$ также не вырождена.

Из равенства $(A + \delta A)^{-1} = (I + A^{-1}\delta A)^{-1}A^{-1}$, в силу неравенства из 5.58, следует, что

$$||(A+\delta A)^{-1}|| \leqslant ||(I + A^{-1}\delta A)^{-1}||\,||A^{-1}|| \leqslant \frac{||A^{-1}||}{1 - ||A^{-1}\delta A||}. \qquad \rhd$$

5.61. Пусть A — невырожденная матрица, δA — ее возмущение и $||A^{-1}\delta A|| < 1$. Получить оценку отклонения матрицы $(A + \delta A)^{-1}$ от A^{-1}.

$\lhd$ Из равенства $(A + \delta A)^{-1} = (I + A^{-1}\delta A)^{-1}A^{-1}$ следует, что $A^{-1} - (A + \delta A)^{-1} = (I - (I + A^{-1}\delta A)^{-1})A^{-1}$. Тогда в силу неравенства из 5.59

$$||A^{-1} - (A+\delta A)^{-1}|| \leqslant ||I - (I + A^{-1}\delta A)^{-1}||\,||A^{-1}|| \leqslant \frac{||A^{-1}\delta A||}{1 - ||A^{-1}\delta A||}\,||A^{-1}||.$$

Относительная ошибка в матрице $(A+\delta A)^{-1}$ оценивается неравенством

$$\frac{||A^{-1} - (A+\delta A)^{-1}||}{||A^{-1}||} \leqslant \frac{||A^{-1}||\,||\delta A||}{1 - ||A^{-1}\delta A||} = \frac{\mathrm{cond}(A)}{1 - ||A^{-1}\delta A||}\,\frac{||\delta A||}{||A||}. \qquad \rhd$$

5.62. Оценить снизу число обусловленности $\mathrm{cond}_2(A)$ матрицы:

$$1)\ A = \begin{pmatrix} 10 & 10 & 30 \\ 0,1 & 0,5 & 0,1 \\ 0,03 & 0,01 & 0,01 \end{pmatrix}; \quad 2)\ A = \begin{pmatrix} 1 & 20 & -400 \\ 0,2 & -2 & -20 \\ -0,04 & -0,2 & 1 \end{pmatrix}.$$

5.63. Система $A\mathbf{x} = \mathbf{b}$, где

$$A = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 10^{-10} & 10^{-10} \\ 1 & 10^{-10} & 10^{-10} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2(1+10^{-10}) \\ -10^{-10} \\ 10^{-10} \end{pmatrix},$$

имеет решение $\mathbf{x} = (10^{-10}, -1, 1)^T$. Доказать, что если $(A + \delta A)\mathbf{y} = \mathbf{b}$, $\|\delta A\| \leqslant 10^{-8}\|A\|$, то $\|\mathbf{x} - \mathbf{y}\| \leqslant 10^{-7}$. Это означает, что относительно малые изменения в элементах матрицы A не приводят к большим изменениям в решении, хотя $\mathrm{cond}_\infty(A)$ имеет порядок 10^{10}.

5.64. Пусть $A = \begin{pmatrix} 100 & 99 \\ 99 & 98 \end{pmatrix}$. Доказать, что данная матрица имеет наибольшее число обусловленности $\mathrm{cond}_2(A)$ из всех невырожденных матриц второго порядка, элементами которых являются положительные целые числа, меньшие или равные 100.

$\lhd$ Введем обозначения для элементов матрицы A

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

и найдем $\mathrm{cond}_2(A)$ в явном виде

$$\|A\|_2 = \sqrt{\max\lambda(A^T A)}, \quad \|A^{-1}\|_2 = \sqrt{\max\lambda((A^{-1})^T A^{-1})} = \\ = \sqrt{\max\lambda((A^T A)^{-1})} = \frac{1}{\sqrt{\min\lambda(A^T A)}}\,.$$

Имеем

$$\mathrm{cond}_2(A) = \sqrt{\frac{\max\lambda(A^T A)}{\min\lambda(A^T A)}}.$$

Введем матрицу

$$B = A^T A = \begin{pmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{pmatrix}$$

и запишем ее характеристический многочлен

$$p_B(\lambda) = \lambda^2 - \lambda\,\mathrm{tr}\,B + \det B, \ \ \mathrm{tr}B = a^2+b^2+c^2+d^2.$$

Его корни равны $\lambda_{1,2} = \dfrac{\mathrm{tr}\,B \pm \sqrt{\mathrm{tr}^2 B - 4\det B}}{2}$. Так как tr $B\ >\ 0$, то

$$\mathrm{cond}_2(A) = \sqrt{\frac{\mathrm{tr}\,B + \sqrt{\mathrm{tr}^2 B - 4\det B}}{\mathrm{tr}\,B - \sqrt{\mathrm{tr}^2 B - 4\det B}}} = \frac{\mathrm{tr}\,B + \sqrt{\mathrm{tr}^2 B - 4\det B}}{\sqrt{4\det B}} = \\ = \frac{\mathrm{tr}\,B}{2\sqrt{\det B}} + \sqrt{\frac{\mathrm{tr}^2 B}{4\det B} - 1}.$$

Таким образом, значение $\mathrm{cond}_2(A)$ максимально, если максимально $\dfrac{\mathrm{tr}^2\,(A^T A)}{\det\left(A^T A\right)}$. Имеем $\mathrm{tr}^2(A^T A) = \left(a^2+b^2+c^2+d^2\right)^2$,

$$\det\left(A^T A\right) = \left(a^2+c^2\right)\left(b^2+d^2\right) - \left(ab+cd\right)^2 = \\ = a^2d^2 + b^2c^2 - 2abcd = \left(ad-bc\right)^2 = \left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right|^2 \equiv \det{}^2 A\,,$$

следовательно, значение $\dfrac{a^2+b^2+c^2+d^2}{\det^2 A}$ должно быть максимальным. Отсюда получаем, что выражение $a^2+b^2+c^2+d^2$ максимально при условии $\det A = \pm 1$. Действительно, если модуль определителя больше 1, то $\mathrm{tr}B$ необходимо увеличить больше, чем в два раза. При ограничении $a_{ij} \leqslant 100$ это невозможно. Таким образом, при $n = 98$ можно воспользоваться любой из следующих матриц:

$$A_1 = \begin{pmatrix} n+2 & n+1 \\ n+1 & n \end{pmatrix}, \quad A_2 = \begin{pmatrix} n+1 & n+2 \\ n & n+1 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} n+1 & n \\ n+2 & n+1 \end{pmatrix}, \quad A_4 = \begin{pmatrix} n & n+1 \\ n+1 & n+2 \end{pmatrix}. \qquad \triangleright$$

5.65. Пусть при некотором $1 > q > 0$ для элементов каждой строки i невырожденной матрицы A выполнено неравенство $q\,|a_{ii}| \geqslant \sum\limits_{j\neq i} |a_{ij}|$. Оценить снизу и сверху $\mathrm{cond}_\infty(A)$, используя только диагональные элементы матрицы и параметр q.

$\triangleleft$ Отметим оценки

$$\max_i |a_{ii}| \leqslant \|A\|_\infty \leqslant (1+q)\max_i |a_{ii}|.$$

Введем обозначение $C = A^{-1}$ и заметим, что для $\forall i, j$ справедливо $|c_{ij}| \leqslant \|C\|_\infty$. При каждом i имеем ($A\,C = I$)

$$\sum_{k=1}^n a_{ik}c_{ki} = 1, \quad 1 \leqslant \sum_{k=1}^n |a_{ik}||c_{ki}| \leqslant |a_{ii}|\,(1+q)\,\|C\|_\infty.$$

Отсюда получаем оценку снизу для нормы матрицы A^{-1}

$$\|A^{-1}\|_\infty = \|C\|_\infty \geqslant \frac{1}{(1+q)\min\limits_i |a_{ii}|}\,,$$

следовательно,

$$\mathrm{cond}_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty \geqslant \frac{1}{1+q}\,\frac{\max\limits_i |a_{ii}|}{\min\limits_i |a_{ii}|}\,.$$

Если правая часть неравенства не превышает единицы, то полученная оценка малосодержательна.

В силу невырожденности матрицы A, все диагональные элементы a_{ii} отличны от нуля, поэтому можно построить матрицы

$$J = \mathrm{diag}\,(a_{11}^{-1}, a_{22}^{-1}, \ldots, a_{nn}^{-1}), \quad B = JA - I\,.$$

Отметим, что $\|B\|_\infty \leqslant q < 1$ в силу цепочки неравенств

$$\max_i |b_{i1}x_1 + \cdots + b_{in}x_n| \leqslant \max_i \sum_k |b_{ik}x_k| \leqslant \|\mathbf{x}\|_\infty \max_i \sum_k |b_{ik}| \leqslant q\,\|\mathbf{x}\|_\infty.$$

Отсюда следует справедливость представления

$$A^{-1} = (I+B)^{-1}J = (I - B + B^2 - B^3 + \cdots)\,J,$$

так как ряд является сходящимся.

Далее для произвольного вектора $\mathbf{x}$ получаем оценку

$$\begin{aligned}\|A^{-1}\mathbf{x}\|_\infty &= \|(I - B + B^2 - B^3 + \cdots)\,J\mathbf{x}\|_\infty \leqslant \\ &\leqslant \|J\mathbf{x}\|_\infty + q\,\|J\mathbf{x}\|_\infty + q^2\|J\mathbf{x}\|_\infty + \cdots = \frac{1}{1-q}\,\|J\mathbf{x}\|_\infty \leqslant \\ &\leqslant \frac{1}{1-q}\,\frac{1}{\min\limits_i |a_{ii}|}\,\|\mathbf{x}\|_\infty.\end{aligned}$$

Следовательно,

$$\mathrm{cond}_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty \leqslant \frac{1+q}{1-q}\,\frac{\max\limits_i |a_{ii}|}{\min\limits_i |a_{ii}|}\,. \qquad \triangleright$$

О т в е т: $\dfrac{1}{1+q}\,\dfrac{\max |a_{ii}|}{\min |a_{ii}|} \leqslant \mathrm{cond}_\infty(A) \leqslant \dfrac{1+q}{1-q}\,\dfrac{\max |a_{ii}|}{\min |a_{ii}|}\,.$

5.66. Пусть R — верхняя треугольная матрица размерности $n \times n$, у которой: 1) $|r_{ij}| \leqslant 1$ для всех i, j; 2) $r_{ii} = 1$ для всех i. Найти максимально возможное значение числа обусловленности $\operatorname{cond}_\infty(R)$.

$\lhd$ Рассмотрим вспомогательные матрицы A_k размерности $(k+1)\times(k+1)$ с элементами $|a_{ij}| \leqslant 1$ следующей структуры:

$$a_{ij} = \begin{cases} a_{ii} & \text{при} \quad i = j, \\ 1 & \text{при} \quad i = j+1, \\ 0 & \text{—} \quad \text{иначе.} \end{cases}$$

Для определителя A_k из разложения по первому столбцу следует оценка

$$|\det(A_k)| \leqslant |a_{11}| \left|\det\left(A^{(1)}_{k-1}\right)\right| + \left|\det\left(A^{(2)}_{k-1}\right)\right| \leqslant 2\,|\det(A_{k-1})| \leqslant$$
$$\leqslant 4\,|\det(A_{k-2})| \leqslant \ldots \leqslant 2^k,$$

поскольку

$$|\det(A_1)| = \left|\det\begin{pmatrix} a_{11} & a_{12} \\ 1 & a_{22} \end{pmatrix}\right| \leqslant 2, \quad |\det(A_0)| = |a_{11}| \leqslant 1.$$

Выше было использовано обозначение $A^{(l)}_{k-1}$ $(l = 1, 2)$ для подматриц k-го порядка, получающихся из исходной матрицы A_k вычеркиванием первого столбца и l-й строки.

Рассмотрим теперь обратную к R матрицу R^{-1} с элементами

$$r^{(-1)}_{ij} = \begin{cases} 1 & \text{при} \quad i = j\,, \\ 0 & \text{при} \quad i > j\,, \\ q_{ij} & \text{при} \quad i < j\,. \end{cases}$$

Так как $\det(R) = 1$, то q_{ij} имеет смысл алгебраического дополнения элемента r_{ji} в определителе матрицы R. При этом его значение равно (с точностью до знака) определителю матрицы, у которой диагональные элементы не превышают единицы, на нижней побочной диагонали ровно $k = j - i - 1$ единиц, а остальные элементы равны нулю. Отсюда имеем

$$|q_{ij}| \leqslant |\det(A_{j-i-1})| \leqslant 2^{j-i-1}.$$

Рассмотрим случай максимально возможных значений q_{ij}:

$$R^{-1} = \begin{pmatrix} 1 & 1 & 2 & \ldots & 2^{n-3} & 2^{n-2} \\ 0 & 1 & 1 & \ldots & 2^{n-4} & 2^{n-3} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}.$$

При этом исходная матрица R однозначно определяется как

$$R = \begin{pmatrix} 1 & -1 & -1 & \ldots & -1 \\ 0 & 1 & -1 & \ldots & -1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

Легко проверить, что

$$\|R^{-1}\|_\infty = 1 + 1 + 2 + \cdots + 2^{n-2} = 2^{n-1}, \quad \|R\|_\infty = n,$$

т. е. построена матрица, на которой одновременно достигаются максимально возможные значения как $\|R\|_\infty$, так и $\|R^{-1}\|_\infty$ среди всех матриц из заданного класса. ▷

Ответ: $\max \operatorname{cond}_\infty(R) = n\, 2^{n-1}$.

5.67. Показать, что определитель D_n *матрицы Коши* K_n с элементами $k_{ij} = \dfrac{1}{a_i + b_j}$, $1 \leqslant i, j \leqslant n$ равен

$$D_n = \prod_{1 \leqslant i < j \leqslant n} (a_i - a_j)(b_i - b_j) \left(\prod_{1 \leqslant i,j \leqslant n} (a_i + b_j) \right)^{-1}.$$

◁ Вычтем первый столбец определителя последовательно из второго, третьего, ..., n-го столбцов, а затем вынесем $b_1 - b_2$ за знак определителя из второго столбца, $b_1 - b_3$ — из третьего и т. д. Затем вынесем $(a_1 + b_1)^{-1}$ из первой строки, $(a_2 + b_1)^{-1}$ — из второй строки и т. д. Далее вычтем первую строку последовательно из второй, третьей, ..., n-й строки, а затем вынесем за знак определителя

$$(a_1 - a_2) \ldots (a_1 - a_n)(a_1 + b_2)^{-1} \ldots (a_1 + b_n)^{-1}.$$

В результате останется определитель матрицы Коши $(n-1)$-го порядка. Поэтому искомая формула получается по индукции. ▷

5.68. Пусть задана матрица Гильберта H_n с элементами $h_{ij} = \dfrac{1}{i + j - 1}$, $1 \leqslant i, j \leqslant n$. Показать, что элементами матрицы H_n^{-1} являются целые числа, которые можно вычислить по формуле

$$a_{ij} = (-1)^{i+j} \frac{(i + n - 1)!\,(j + n - 1)!}{[(i-1)!]^2 [(j-1)!]^2 (n-i)!(n-j)!(i+j-1)}.$$

◁ Рассмотрим матрицу Коши K_n с элементами $k_{ij} = \dfrac{1}{a_i + b_j}$, $1 \leqslant i, j \leqslant n$. Ее определитель вычислен в 5.67. Элементы матрицы K_n^{-1} являются отношениями алгебраических дополнений к определителю исходной матрицы. Миноры матрицы Коши снова являются матрицами Коши. Поэтому можно получить явные выражения для элементов K_n^{-1}:

$$b_{ij} = \prod_{k=1}^{n} (a_j + b_k)(a_k + b_i) \left[(a_j + b_i) \prod_{k \neq j} (a_j - a_k) \prod_{k \neq i} (b_i - b_k) \right]^{-1}.$$

Полагая $a_i = i$, $b_i = i - 1$, получим частный случай матрицы Коши — матрицу Гильберта и искомую формулу для элементов H_n^{-1}. ▷

5.69. Оценить рост числа обусловленности $\mathrm{cond}_\infty(H_n)$ матрицы Гильберта с элементами $h_{ij} = \dfrac{1}{i+j-1}$, $1 \leqslant i, j \leqslant n$ относительно параметра размерности n.

Указание. Величина $\dfrac{(i+n-1)!}{((i-1)!)^2(n-i)!}$ принимает максимальное значение при $i = \left[\dfrac{n}{\sqrt{2}}\right]$, поэтому для элементов a_{ij} матрицы H_n^{-1} (см. 5.68) по формуле Стирлинга

$$n! = \sqrt{2\,\pi\,n}\,n^n \mathrm{e}^{-n} \mathrm{e}^{\theta(n)}\,, \quad |\theta(n)| \leqslant \frac{1}{12\,n}\,,$$

имеем асимптотику

$$\max_{i,j} |a_{ij}| = \frac{1}{4\sqrt{2}\pi^2 n} \left(\sqrt{2}+1\right)^{4n} \left(1 + O\left(\frac{1}{n}\right)\right)\,.$$

Отсюда следует равенство

$$\|H_n^{-1}\|_\infty = \frac{1}{(2\,\pi)^{3/2} 2^{7/4} \sqrt{n}} \left(\sqrt{2}+1\right)^{4n} \left(1 + O\left(\frac{1}{n}\right)\right)\,.$$

Так как

$$\ln n \leqslant \|H_n\|_\infty = \sum_{j=1}^{n} \frac{1}{j} \leqslant 3\,\ln n \quad (\text{для}\, n \geqslant 2)\,,$$

то главный член асимптотики $\mathrm{cond}_\infty(H_n)$ имеет вид $\mathrm{const} \cdot 4^n \ln \dfrac{n}{\sqrt{n}}$.

5.70. Доказать *неравенство Адамара* для квадратных матриц вида $A = A^T > 0$:

$$\det(A) \leqslant \prod_{i=1}^{n} a_{ii}\,.$$

$\lhd$ Положим $d_i = \dfrac{1}{\sqrt{a_{ii}}}$ и пусть $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$. Неравенство $\det(A) \leqslant a_{11}a_{22}\ldots a_{nn}$ равносильно условию $\det(DAD) \leqslant 1$, и в дальнейшем достаточно рассматривать матрицу A, все диагональные элементы которой равны единице. Если $\lambda_1, \lambda_2, \ldots, \lambda_n$ — собственные значения матрицы A (обязательно положительные), то

$$\det(A) = \prod_{i=1}^{n} \lambda_i \leqslant \left(\frac{1}{n}\sum_{i=1}^{n} \lambda_i\right)^n = \left(\frac{1}{n}\,\mathrm{tr}A\right)^n = 1\,.$$

Здесь мы воспользовались неравенством между арифметическим и геометрическим средними неотрицательных чисел. Равенство средних имеет место тогда и только тогда, когда все $\lambda_i = 1$. В силу симметрии, матрица A диагонализуема. При единичных диагональных элементах и собственных значениях это равносильно тому, что A является единичной матрицей. Соответственно равенство в исходном неравенстве достигается тогда и только тогда, когда A — диагональная матрица. $\rhd$

5.71. Показать, что для произвольной квадратной матрицы C справедливы неравенства

$$|\det(C)| \leqslant \prod_{i=1}^{n} \left(\sum_{j=1}^{n} |c_{ij}|^2 \right)^{1/2}, \quad |\det(C)| \leqslant \prod_{j=1}^{n} \left(\sum_{i=1}^{n} |c_{ij}|^2 \right)^{1/2},$$

а равенства в них достигаются тогда и только тогда, когда строки (соответственно столбцы) матрицы C попарно ортогональны.

$\triangleleft$ Если матрица C вырождена, то доказывать нечего. В случае невырожденной матрицы C нужно применить неравенство из 5.70 к положительно определенной матрице $A = C\,C^T$ и извлечь квадратный корень из обеих частей неравенства. Правая часть доказываемого неравенства — квадратный корень из произведения диагональных элементов матрицы A, а левая часть — квадратный корень из определителя этой матрицы. Строки матрицы C попарно ортогональны тогда и только тогда, когда A — диагональная матрица, а это и есть случай равенства в 5.70. Второе искомое неравенство получается применением первого к матрице C^T. $\triangleright$

5.3. Точные методы

К точным методам решения системы $A\mathbf{x} = \mathbf{b}$ линейных алгебраических уравнений относятся алгоритмы, которые при отсутствии ошибок округления, позволяют точно вычислить искомый вектор $\mathbf{x}$ за конечное число логических и арифметических операций. Если число ненулевых элементов матрицы имеет порядок n^2, то большинство алгоритмов такого рода позволяют найти решение за $O(n^3)$ арифметических действий. Данная оценка, а также необходимость хранения всех элементов матрицы в памяти компьютера накладывают существенное ограничение на область применимости точных методов. Однако для решения задач размерности n менее 10^4 разумно применять точные алгоритмы. При численном решении задач математической физики часто требуется обращать матрицы блочно-диагонального вида. В этом случае удается построить точные методы с меньшим по порядку числом арифметических действий. К таким алгоритмам относят методы прогонки, стрельбы, Фурье (базисных функций).

Наиболее известным из точных методов, применяемых для решения задач с матрицами общего вида, является *метод исключения Гаусса.* В предположении, что коэффициент $a_{11} \neq 0$, уравнения исходной системы заменяем следующими:

$$\begin{cases} x_1 + \sum\limits_{j=2}^{n} \dfrac{a_{1j}}{a_{11}}\, x_j = \dfrac{b_1}{a_{11}}, \\ \sum\limits_{j=2}^{n} \left(a_{ij}x_j - \dfrac{a_{1j}}{a_{11}}\, a_{i1}x_j \right) = b_i - \dfrac{b_1}{a_{11}}\, a_{i1}, \quad i = 2, \ldots, n, \end{cases}$$

т. е. первое уравнение делим на a_{11}, затем, умноженное на соответствующий коэффициент a_{i1}, вычитаем из последующих уравнений. В полученной системе $A^{(1)}\mathbf{x} = \mathbf{b}^{(1)}$ неизвестное x_1 исключено из всех уравнений,