2. МАТРИЧНАЯ ТЕОРИЯ ВОЗМУЩЕНИЙ

2.1. Предварительные замечания

Линейные системы. Пусть дана система линейных алгебраических уравнений (линейная система)

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

с квадратной невырожденной матрицей A.

Вследствие ошибок округлений в результате решения линейной системы мы получаем *приближенное* решение $\tilde{\boldsymbol{x}}$, которое можно рассматривать как *точное* решение *возмущенной* системы

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

где матрица возмущений δA мала в каком-либо смысле.

Второй источник ошибок в $\tilde{\boldsymbol{x}}$ определяется возмущениями δA и $\delta\boldsymbol{b}$ в элементах матрицы A и в компонентах вектора правой части $\boldsymbol{b}$ (например, вследствие ошибок измерений или ошибок округлений, возникающих в процессе ввода вещественных чисел в память вычислительной машины).

Для оценки того, насколько приближенное решение $\tilde{\boldsymbol{x}}$ отличается от точного решения $\boldsymbol{x}$, водится понятие *меры* такого отличия. В качестве меры используются нормы векторов и согласованные нормы матриц, для которых норма единичной матрицы равна 1.

Пусть задана какая-то векторная норма. Тогда говорят, что число $||\boldsymbol{x} - \tilde{\boldsymbol{x}}||$ есть *абсолютная* ошибка в $\tilde{\boldsymbol{x}}$. Если $\boldsymbol{x} \neq 0$, то число $||\boldsymbol{x} - \tilde{\boldsymbol{x}}||/||\boldsymbol{x}||$ называют *относительной* ошибкой в $\tilde{\boldsymbol{x}}$. Вектор $\boldsymbol{r} = \boldsymbol{x} - \tilde{\boldsymbol{x}}$ называется *вектором невязки*. Относительная ошибка в ∞-норме может рассматриваться как оценка количества верных значащих цифр в $\tilde{\boldsymbol{x}}$: если

$$\frac{||\boldsymbol{x} - \tilde{\boldsymbol{x}}||_\infty}{||\boldsymbol{x}||_\infty} \approx 10^{-p},$$

то наибольшая по модулю компонента в $\tilde{\boldsymbol{x}}$ имеет примерно p значащих цифр.

Как правило, при оценках отклонения вычисленного решения $\tilde{\boldsymbol{x}}$ заданной системы $A\,\boldsymbol{x} = \boldsymbol{b}$ от ее точного решения $\boldsymbol{x}$ применяется *обратный анализ ошибок*, когда $\tilde{\boldsymbol{x}}$ рассматривается как *точное* решение возмущенной системы

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

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

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

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

Из приведенного выше неравенства следует, что даже если вектор невязки $\boldsymbol{r} = \boldsymbol{b} - A\tilde{\boldsymbol{x}}$ мал, относительные возмущения в решении могут быть большими, если $\mathrm{cond}(A)$ велико (такие матрицы называют *плохо обусловленными*). Следовательно, число обусловленности может рассматриваться как мера *чувствительности* решения к возмущениям системы.

Если в системе $A\boldsymbol{x} = \boldsymbol{b}$ возмущены матрица A и вектор $\boldsymbol{b}$, т.е. в действительности решается возмущенная система $(A + \delta A)\,\tilde{\boldsymbol{x}} = \boldsymbol{b} + \delta\boldsymbol{b}$, то при условии $\|A^{-1}\delta A\| \le 1$ имеет место оценка

$$\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \le \frac{\operatorname{cond}(A)}{1 - \|A^{-1}\delta A\|}\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}\right).$$

Если $\|A^{-1}\|\|\delta A\| \le 1$, то имеет место более грубая оценка

$$\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \le \frac{\operatorname{cond}(A)}{1 - \|A^{-1}\|\|\delta A\|}\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}\right).$$

Обращение матриц. Пусть невырожденная вещественная матрица A возмущена на матрицу δA. Тогда если $\|A^{-1}\delta A\| \le 1$, то возмущенная матрица $A + \delta A$ не вырождена и имеет место оценка

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

Если $\|A^{-1}\|\|\delta A\| \le 1$, то имеет место более грубая оценка

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

Собственные значения. При оценке ошибок, возникающих в процессе решения алгебраической проблемы собственных значений, также применим обратный анализ ошибок, когда вычисленные собственные значения рассматриваются как точные для возмущенной матрицы.

Пусть λ — спектр квадратной матрицы A порядка n, а $\tilde{\lambda}$ — спектр возмущенной матрицы $A + \delta A$. Тогда анализ возмущений собственных значений может быть выполнен следующим образом.

Вначале рассмотрим случай, когда матрица A имеет простую структуру, т.е. имеет n линейно независимых собственных векторов. Такая матрица подобна диагональной матрице:

$$X^{-1} A\, X = \operatorname{diag}(\lambda_1,\, \ldots,\, \lambda_n)\,,$$

где X — матрица, столбцы которой являются собственными векторами матрицы A. Тогда для любого $\tilde{\lambda}_i \in \tilde{\lambda}$ выполнено неравенство

$$\min_{1\le j\le n}\left|\tilde{\lambda}_i - \lambda_j\right| \le ||X^{-1}||_2\, ||X||_2\, ||\delta A||_2 = \operatorname{cond}_2(X)\, ||\delta A||_2\,,$$

где $\lambda_j \in \lambda$. Величину $\operatorname{cond}_2(X)$ называют *спектральным числом обусловленности* по отношению к проблеме собственных значений.

Из этого неравенства следует, что возмущения в собственных значениях матрицы A прямо пропорциональны числу обусловленности матрицы ее собственных векторов. Если матрица симметрична, то матрица X ортогональна и $\operatorname{cond}_2(X) = 1$, т.е. для этих матриц проблема собственных значений всегда хорошо обусловлена.

Если матрица A не имеет n линейно независимых собственных векторов, то она подобна блочно-диагональной матрице J, блоки которой есть канонические клетки Жордана:

$$P^{-1} A\, P = J\,,$$

где P — матрица, столбцы которой являются корневыми векторами матрицы A (т.е. матрицу A можно представить в канонической форме Жордана). Тогда для каждого $\tilde{\lambda}_i \in \tilde{\lambda}$ существует такое $\lambda_j \in \lambda$, что

$$\frac{|\tilde{\lambda}_i - \lambda_j|^m}{1 + |\tilde{\lambda}_i - \lambda_j| + \cdots + |\tilde{\lambda}_i - \lambda_j|^{m-1}} \le ||P^{-1}||_2\, ||P||_2\, ||\delta A||_2\,,$$

где m — максимальный порядок жордановых клеток, отвечающих λ_j. Отсюда следует, что если k — максимальный порядок всех жордановых клеток и $\delta A = \varepsilon B$, где $\varepsilon > 0$ — малая величина и $||B||_2 = 1$, то любое собственное значение возмущенной матрицы отличается от некоторого собственного значения исходной матрицы на величину порядка $\varepsilon^{1/k}$.

Выписанные оценки дают верхнюю границу обусловленности каждого из собственных значений матрицы A, хотя различные собственные значения имеют различную степень обусловленности.

Переопределенные линейные системы. Пусть заданы вещественная прямоугольная $(n \times m)$-матрица A, $n > m$, и вещественный n-вектор $\boldsymbol{x}$. Тогда линейная система $A\boldsymbol{x} = \boldsymbol{b}$ называется *переопределенной*, а под ее решением понимается такой вектор $\boldsymbol{x}$, для которого норма вектора невязки $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ минимальна в том смысле, что

$$\|\boldsymbol{x}\|_2 = \min_{\boldsymbol{x}} \|\boldsymbol{b} - A\boldsymbol{x}\|_2.$$

Если матрица A имеет *полный* ранг, т.е. ее ранг равен m, то решение системы в указанном смысле единственно и совпадает с решением *нормальной* системы $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$.

Запишем решение нормальной системы в виде $\boldsymbol{x} = (A^T A)^{-1} A^T \boldsymbol{b}$ и введем обозначение $A^+ = (A^T A)^{-1} A^T$. Матрица A^+ называется *псевдообратной*. Решение переопределенной системы можно записать в виде $\boldsymbol{x} = A^+\boldsymbol{b}$.

Пусть в переопределенной системе возмущается вектор правой части, т.е. решается возмущенная система $A\boldsymbol{x} = \boldsymbol{b} + \delta\boldsymbol{b} = \tilde{\boldsymbol{b}}$. Обозначим через $\boldsymbol{b}_1$ и $\tilde{\boldsymbol{b}}_1$ проекции векторов $\boldsymbol{b}$ и $\tilde{\boldsymbol{b}}$ на пространство $R(A)$, образованное столбцами матрицы A. Тогда для относительной ошибки в вычисленном решении имеет место неравенство

$$\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}\|_2} \le \frac{\|A^+\boldsymbol{b} - A^+\tilde{\boldsymbol{b}}\|_2}{\|A^+\boldsymbol{b}\|} \le \operatorname{cond}_2(A)\,\frac{\|\boldsymbol{b}_1 - \tilde{\boldsymbol{b}}_1\|}{\|\boldsymbol{b}_1\|},$$

где $\operatorname{cond}_2(A) = \|A\|_2\|A^+\|_2$ называется числом обусловленности прямоугольной матрицы.

2.2. Задачи и решения

1. Рассмотрим задачу вычисления *апостериорных* оценок приближенного решения $\tilde{\boldsymbol{x}}$ системы $A\boldsymbol{x} = \boldsymbol{b}$, где A — невырожденная матрица и $\boldsymbol{b} \neq 0$. “Естественная” проверка того, насколько хорошо $\tilde{\boldsymbol{x}}$ удовлетворяет системе, состоит в анализе вектора невязки $\boldsymbol{r} = \boldsymbol{b} - A\tilde{\boldsymbol{x}}$. Если $\boldsymbol{r} = 0$, то $\tilde{\boldsymbol{x}}$ есть точное решение $\boldsymbol{x}$. Если же вектор $\boldsymbol{r}$ мал, то можно ли ожидать, что вектор $\tilde{\boldsymbol{x}}$ близок к $\boldsymbol{x}$?

Решение. Из равенства $A^{-1}\boldsymbol{r} = A^{-1}\boldsymbol{b} - A^{-1}A\tilde{\boldsymbol{x}} = \boldsymbol{x} - \tilde{\boldsymbol{x}}$ следует, что

$$||\boldsymbol{x} - \tilde{\boldsymbol{x}}|| \le ||A^{-1}||\,||\boldsymbol{r}||. \tag{$*$}$$

Из $\boldsymbol{b} = A\boldsymbol{x}$ следует, что $||\boldsymbol{b}|| = ||A\boldsymbol{x}|| \le ||A||\,||\boldsymbol{x}||$, т.е.

$$||\boldsymbol{x}|| \ge \frac{||\boldsymbol{b}||}{||A||}. \tag{$**$}$$

Поделим неравенство $(*)$ на неравенство $(**)$. Тогда получим

$$\frac{||\boldsymbol{x} - \tilde{\boldsymbol{x}}||}{||\boldsymbol{x}||} \le ||A||\,||A^{-1}||\,\frac{||\boldsymbol{r}||}{||\boldsymbol{b}||} = \operatorname{cond}(A)\,\frac{||\boldsymbol{r}||}{||\boldsymbol{b}||} = \operatorname{cond}(A)\,\frac{||\boldsymbol{b} - A\tilde{\boldsymbol{x}}||}{||\boldsymbol{b}||}.$$

Отсюда видно, что если матрица A плохо обусловлена, то даже очень маленькая невязка *не может* гарантировать малость относительной ошибки в $\tilde{\boldsymbol{x}}$. Хуже того, может так оказаться, что достаточно точное решение будет иметь большую невязку. Действительно, рассмотрим пример

$$A = \begin{pmatrix} 1.000 & 1.001 \\ 1.000 & 1.000 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 2.001 \\ 2.000 \end{pmatrix}.$$

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

Возьмем теперь $\boldsymbol{b} = (1, 0)^T$. Тогда вектор $\boldsymbol{x} = (-1000, 1000)^T$ является точным решением системы. Вектор $\tilde{\boldsymbol{x}} = (-1001, 1000)^T$ достаточно близок к $\boldsymbol{x}$ в смысле относительной погрешности, однако $\tilde{\boldsymbol{x}}$ дает большую невязку $\boldsymbol{r} = (0, -1)^T$, которая имеет порядок правой части.

Сделаем два полезных замечания. Первым признаком плохой обусловленности линейной системы является появление малых ведущих элементов в процессе применения гауссова исключения. Для большинства матриц это достаточно надежный признак. Однако существуют плохо обусловленные матрицы (например, с диагональным преобладанием), для которых малые ведущие элементы *не появляются*. Вторым признаком плохой обусловленности может служить появление большого решения. Пусть, например, $||A|| = ||\boldsymbol{b}|| = 1$. Тогда $||\boldsymbol{x}|| \le ||A^{-1}||\,||\boldsymbol{b}|| = ||A^{-1}|| = \mathrm{cond}(A)$. Поэтому если норма $||\boldsymbol{x}||$ велика, то велико и $\mathrm{cond}(A)$. К сожалению, плохо обусловленные системы могут иметь небольшие решения, которые дают маленькие невязки.

Теперь покажем, что *округленное* точное решение линейной системы может иметь *большую* невязку. Пусть вычисления проводятся в системе счисления с основанием 10 и с t разрядами мантиссы, а $\boldsymbol{x}$ — каким-либо образом полученное точное решение линейной системы порядка n. Тогда округленное решение $\tilde{\boldsymbol{x}}$ запишется в виде $\tilde{\boldsymbol{x}} = \big(x_1(1+\varepsilon_1), \ldots, x_n(1+\varepsilon_n)\big)$, где $|\varepsilon_i| \le 10^{-t}$, $i = 1, \ldots, n$. Следовательно, $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{e}$, где $\boldsymbol{e} = (x_1\varepsilon_1, \ldots, x_n\varepsilon_n)$. Заметим, что $|e_i| = |x_i\varepsilon_i| \le |x_i|\,|\varepsilon_i| \le |x_i|\,10^{-t} \le ||\boldsymbol{x}||_\infty\,10^{-t}$. Поскольку это неравенство выполнено для всех $i = 1, \ldots, n$, то оно выполнено и для того i, при котором его левая часть достигает максимума. Это означает, что $||\boldsymbol{e}||_\infty \le ||\boldsymbol{x}||_\infty\,10^{-t}$. Из

$$\boldsymbol{r} = \boldsymbol{b} - A\,\tilde{\boldsymbol{x}} = \boldsymbol{b} - A\,\boldsymbol{x} - A\,\boldsymbol{e} = -A\,\boldsymbol{e}$$

и $||\boldsymbol{r}||_\infty \le ||A||_\infty ||\boldsymbol{e}||_\infty \le |A||_\infty ||\boldsymbol{x}||_\infty 10^{-t}$ следует, что

$$\frac{||\boldsymbol{r}||_\infty}{||A||_\infty} \le ||\boldsymbol{x}||_\infty 10^{-t}.$$

Если матрица A плохо обусловлена и норма $||\boldsymbol{x}||$ велика, то невязка будет большой. Это заключение иллюстрируется второй частью приведенного выше примера.

2. Пусть E — единичная матрица и $||\delta E|| < 1$. Показать, что матрица $E - \delta E$ невырожденная и выполнена оценка

$$||(E - \delta E)^{-1}|| \le \frac{1}{1 - ||\delta E||}\,.$$

Решение. Возьмем произвольный вектор $\boldsymbol{x} \ne 0$. Поскольку $1 - ||\delta E|| > 0$ и $||\boldsymbol{x}|| = ||(\boldsymbol{x} - \delta E\boldsymbol{x}) + \delta E\boldsymbol{x}|| \le ||\boldsymbol{x} - \delta E\boldsymbol{x}|| + ||\delta E\boldsymbol{x}||$, то

$$||(E - \delta E)\,\boldsymbol{x}|| = ||\boldsymbol{x} - \delta E\boldsymbol{x}|| \ge ||\boldsymbol{x}|| - ||\delta E\boldsymbol{x}|| \ge ||\boldsymbol{x}|| - ||\delta E||\,||\boldsymbol{x}|| \ge (1 - ||\delta E||)\,||\boldsymbol{x}|| > 0\,.$$

Следовательно, если $\boldsymbol{x} \ne 0$, то $(E - \delta E)\,\boldsymbol{x} \ne 0$, т.е. матрица $E - \delta E$ не вырождена.

Из тождества $(E - \delta E)\,(E - \delta E)^{-1} = E$ получим $(E - \delta E)^{-1} = E + \delta E\,(E - \delta E)^{-1}$. Отсюда

$$||(E - \delta E)^{-1}|| \le ||E|| + ||\delta E||\,||(E - \delta E)^{-1}|| = 1 + ||(E - \delta E)^{-1}||\,||\delta E||\,.$$

Из этого неравенства следует решение задачи (часто ее называют задачей о возмущении единичной матрицы).

3. Пусть E — единичная матрица и $||\delta E|| < 1$. Получить оценку отклонения матрицы E от матрицы $(E - \delta E)^{-1}$.

Решение. Из $(E - \delta E)^{-1} = E + \delta E(E - \delta E)^{-1}$ (см. задачу 2) получим $E - (E - \delta E)^{-1} = -\delta E(E - \delta E)^{-1}$. Отсюда

$$||E - (E - \delta E)^{-1}|| \le ||\delta E||\,||(E - \delta E)^{-1}|| \le \frac{||\delta E||}{1 - ||\delta E||}$$

в силу неравенства из задачи 2.

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

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

Решение. Имеем $A+\delta A = A(E+A^{-1}\delta A)$. Поскольку $||A^{-1}\delta A|| < 1$, то из задачи 2 следует, что матрица $E+A^{-1}\delta A$ невырожденная. Это означает, что и матрица $A+\delta A$ также не вырождена.

Из равенства $(A+\delta A)^{-1} = (E+A^{-1}\delta A)^{-1}A^{-1}$ следует, что

$$||(A+\delta A)^{-1}|| \le ||(E+A^{-1}\delta A)^{-1}||\,||A^{-1}|| \le \frac{||A^{-1}||}{1-||A^{-1}\delta A||}$$

в силу неравенства из задачи 2.

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

Решение. Из равенства $(A+\delta A)^{-1} = (E+A^{-1}\delta A)^{-1}A^{-1}$ следует, что $A^{-1}-(A+\delta A)^{-1} = \big(E-(E+A^{-1}\delta A)^{-1}\big)A^{-1}$. Тогда

$$||A^{-1}-(A+\delta A)^{-1}|| \le ||E-(E+A^{-1}\delta A)^{-1}||\,||A^{-1}|| \le \frac{||A^{-1}\delta A||}{1-||A^{-1}\delta A||}\,||A^{-1}||$$

в силу неравенства из задачи 3.

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

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

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

Решение. Поскольку $E = AA^{-1}$, то

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

Далее, так как умножение матрицы на ортогональную не меняет ее спектральную норму, то $||Q||_2 = ||QE||_2 = ||E||_2 = 1$ и $||Q^T||_2 = ||Q^TE||_2 = ||E||_2 = 1$. Тогда

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

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

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

$$\mathrm{cond}(D) = ||D||\,||D^{-1}||_2 = \varepsilon||E||\,\varepsilon^{-1}||E^{-1}|| = 1\,.$$

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 $\boldsymbol{b} \neq 0$ и, решая систему $A\boldsymbol{x} = \boldsymbol{b}$ при помощи обратной подстановки, построим элементы обратной матрицы A^{-1}:

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

Выпишем полученную обратную матрицу:

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

Следовательно,

$$\|A^{-1}\|_\infty = 1 + 1 + 2 + 2^2 + \cdots + 2^{n-2} = 2^{n-1}.$$

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

Эти два примера показывают, что обусловленность матрицы не зависит от величины определителя.

8. Пусть дана жорданова клетка порядка n:

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

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

Решение. Как и в задаче 7, методом обратной подстановки получим обратную матрицу:

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

Тогда

$$\begin{aligned} &\|A\|_\infty = 1 + |a|, \\ &\|A^{-1}\|_\infty = 1 + |a| + a^2 + \cdots + |a|^{n-1} = \frac{|a|^n - 1}{|a| - 1}, \quad |a| \neq 1 \\ &\|A^{-1}\|_\infty = n, \quad |a| = 1 \\ &\mathrm{cond}_\infty(A) = \frac{(|a|+1)\,(|a|^n - 1)}{|a| - 1}, \quad |a| \neq 1 \\ &\mathrm{cond}_\infty(A) = (|a|+1)\,n, \quad |a| = 1. \end{aligned}$$

Отсюда видно, что матрица A плохо обусловлена при $|a| > 1$ и хорошо обусловлена при $|a| \leq 1$.

Например, при $n = 20$ и $a = 5$ будем иметь $\mathrm{cond}_\infty(A) \approx 10^{14}$.

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

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

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.

Легко видеть, что в случае плохой обусловленности жордановой клетки влияние возмущения в b_n в компонентах x_i уменьшается с ростом i. Значит, чувствительность каждой отдельной компоненты решения линейной системы при возмущениях правой части может быть различной. Кроме того, возмущение в компоненте b_j приводит к возмущениям только тех x_i, для которых $i \le j$, и не затрагивает других x_i, причем самые незначительные последствия будет иметь возмущение в b_1. Следовательно, даже в случае плохой обусловленности не всякие возмущения правой части системы приводят к большим возмущениям в решении.

9. Пусть система $Ax = b$ с матрицей

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

решается методом LU-разложения:

$$A = LU,$$

$$Ly = b, \quad Ux = y.$$

Вычислить $\mathrm{cond}_\infty(L)$ и $\mathrm{cond}_\infty(U)$, если LU-разложение строится при помощи компактной схемы Гаусса

а) без выбора ведущего элемента;

б) с выбором ведущего элемента.

Решение. а) Применим компактную схему без выбора ведущего элемента:

$$\begin{pmatrix} \varepsilon & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix}.$$

Отсюда для определения элементов матриц L и U получаем систему линейных алгебраических уравнений:

$$\begin{cases} l_{11}\,u_{11} = \varepsilon, \\ l_{11}\,u_{12} = 1, \\ l_{21}\,u_{11} = 1, \\ l_{21}\,u_{12} + l_{22}\,u_{22} = 1. \end{cases}$$

Для определенности положим $l_{11} = l_{22} = 1$. Тогда

$$L = \begin{pmatrix} 1 & 0 \\ \dfrac{1}{\varepsilon} & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} \varepsilon & 1 \\ 0 & 1 - \dfrac{1}{\varepsilon} \end{pmatrix},$$

$$L^{-1} = \begin{pmatrix} 1 & 0 \\ -\dfrac{1}{\varepsilon} & 1 \end{pmatrix}, \quad U^{-1} = \begin{pmatrix} \dfrac{1}{\varepsilon} & -\dfrac{1}{\varepsilon - 1} \\ 0 & \dfrac{\varepsilon}{\varepsilon - 1} \end{pmatrix}.$$

Отсюда

$$\mathrm{cond}_\infty(L) = \left(1 + \frac{1}{\varepsilon}\right)^2, \quad \mathrm{cond}_\infty(U) = \frac{1+\varepsilon}{\varepsilon^2(1-\varepsilon)}.$$

б) LU-разложение с выбором ведущего элемента:

$$\begin{pmatrix} 1 & 1 \\ \varepsilon & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ l_{21} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix},$$

$$L = \begin{pmatrix} 1 & 0 \\ \varepsilon & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 1 & 1 \\ 0 & 1-\varepsilon \end{pmatrix},$$

$$L^{-1} = \begin{pmatrix} 1 & 0 \\ -\varepsilon & 1 \end{pmatrix}, \quad U^{-1} = \begin{pmatrix} 1 & -\dfrac{1}{1-\varepsilon} \\ 0 & \dfrac{1}{1-\varepsilon} \end{pmatrix}.$$

Отсюда

$$\operatorname{cond}_\infty(L) = (1+\varepsilon)^2, \quad \operatorname{cond}_\infty(U) = 2\left(1 + \frac{1}{1-\varepsilon}\right).$$

10. Матрица Уилкинсона

$$A = \begin{pmatrix} 20 & 20 & 0 & 0 & \dots & 0 & 0 \\ 0 & 19 & 20 & 0 & \dots & 0 & 0 \\ 0 & 0 & 18 & 20 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 2 & 20 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

имеет наименьшее по модулю собственное значение, равное 1. Как оно изменится в результате возмущения первого элемента последней строки на величину $\varepsilon = 20^{-19} \cdot 20! \approx 5 \cdot 10^{-7}$?

Решение. Характеристическое уравнение для возмущенной матрицы Уилкинсона имеет вид:

$$\det(A - \lambda E) = (20-\lambda)(19-\lambda)\cdots(1-\lambda) - 20^{19} \cdot \varepsilon = 0.$$

Свободный член в этом уравнении равен 0 и, следовательно, наименьшее собственное значение также равно 0.

11. Пусть дана матрица

$$A = \begin{pmatrix} n & a & 0 & 0 & \dots & 0 & 0 \\ 0 & n-1 & a & 0 & \dots & 0 & 0 \\ 0 & 0 & n-2 & a & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 2 & a \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

порядка n и $a > 1$. Показать, что для этой матрицы проблема собственных значений плохо обусловлена.

Решение. Как и в предыдущей задаче, возмутим матрицу A, добавив малое ε к первому элементу последней строки. Разложив определитель возмущенной матрицы по последней строке, получим следующее характеристическое уравнение:

$$(n-\lambda)\big((n-1)-\lambda\big)\cdots(2-\lambda)(1-\lambda) + (-1)^{n+1}a^{n-1}\varepsilon = 0\,.$$

Обозначим через λ_i собственные значения матрицы A, а через $\lambda_i(\varepsilon)$ собственные значения возмущенной матрицы. Поскольку матрица A имеет различные собственные значения, то $\lambda_i(\varepsilon)$ представляется в виде степенного ряда по параметру ε:

$$\lambda_i(\varepsilon) = \lambda_i + \sum_{j=1}^{n} \mu_{ij}\varepsilon^j = i + \sum_{j=1}^{n} \mu_{ij}\varepsilon^j\,.$$

Ограничимся членами первого порядка малости относительно ε и подставим $\lambda_i(\varepsilon) = i + \nu_i\varepsilon$ в характеристическое уравнение, где $\nu_i = \mu_{i1}$. Отбросим в получившемся выражении члены второго порядка малости и выше относительно ε. В результате получим следующие соотношения на ν_i, которые определяют степень возмущения каждого собственного значения при возмущении исходной матрицы:

$$\nu_i = \frac{(-1)^{n-i+1}a^{n-1}}{(n-i)!(i-1)!}\,.$$

Для матрицы Уилкинсона ($n = 20$ и $a = 20$) наименьшие возмущения собственных значений будут при $i = 1$ и $i = 20$, а наибольшие при $i = 10$ и $i = 11$: $\nu_1 = 20^{19}/19! \approx 4.31 \cdot 10^7$ и $\nu_{11} = 20^{19}/9!10! \approx 3.98 \cdot 10^{12}$.

Таким образом, для этой матрицы проблема собственных значений плохо обусловлена, хотя различные собственные значения имеют различную степень обусловленности.

Легко видеть, что если $a < 1$, то плохой обусловленности нет.

12. Пусть дана жорданова клетка порядка n

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

и пусть первый элемент последней строки возмущается на малую величину $\varepsilon > 0$. Выполнить анализ возмущений собственных значений.

Решение. Характеристическое уравнение возмущенной матрицы имеет вид

$$(\lambda - 1)^n - (-1)^n a^{n-1}\varepsilon = 0.$$

Пусть ν_i — корень уравнения $\nu^n - (-1)^n = 0$. Тогда для собственных значений $\tilde{\lambda}$ возмущенной матрицы получим

$$\tilde{\lambda}_i = 1 + \nu_i a^{(n-1)/n}\varepsilon^{1/n}, \quad i = 1, \dots n\,.$$

Отсюда видно, что задача собственных значений для клетки Жордана плохо обусловлена, причем обусловленность ухудшается с ростом n. Даже если $|a| < 1$, когда клетка Жордана хорошо обусловлена по отношению к задачам решения линейных систем и обращения матриц, она остается плохо обусловленной по отношению к задаче на собственные значения.

Для примера рассмотрим случай, когда $n = 10$, $a = 1$ и $\varepsilon = 10^{-10}$. Тогда возмущение в $\lambda_1 = 1$ будет равно $|\varepsilon|^{1/n} = 0.1$, т.е. возмущение порядка 10^{-10} в одном элементе матрицы внесло в собственное значение λ_1 ошибку, равную 0.1.