МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. М.В. ЛОМОНОСОВА

Механико – математический факультет
Кафедра вычислительной математики

И. О. Арушанян

Избранные задачи для семинарских занятий по численным методам: матричная теория возмущений

Учебное пособие

Москва, 2020

Данное учебное пособие содержит методические материалы для проведения семинарских занятий на четвертом курсе механико-математического факультета МГУ по учебной дисциплине “Численные методы”, на которых рассматривается матричная теория возмущений.

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 число $||\boldsymbol{x} - \tilde{\boldsymbol{x}}||$ — это *абсолютная* ошибка в приближенном решении $\tilde{\boldsymbol{x}}$. 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 $\|A^{-1}\delta A\| \leq 1$ имеет место оценка

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

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

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

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

$$\left\|A^{-1} - (A + \delta A)^{-1}\right\| \leq \frac{\mathrm{cond}(A)}{1 - \|A^{-1}\delta A\|}\,\frac{\|\delta A\|}{\|A\|}.$$

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

$$\left\|A^{-1} - (A + \delta A)^{-1}\right\| \leq \frac{\mathrm{cond}(A)}{1 - \|A^{-1}\|\|\delta A\|}\,\frac{\|\delta A\|}{\|A\|}.$$

Собственные значения. При оценке ошибок, возникающих в процессе решения алгебраической проблемы собственных значений, тоже 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 образуют 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

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

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.

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 $\boldsymbol{b} \ne 0$. “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

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

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

$$||\boldsymbol{x}|| \ge \frac{||\boldsymbol{b}||}{||A||}. \qquad (**)$$

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 небольшую 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 $||\boldsymbol{x}|| \le ||A^{-1}||\,||\boldsymbol{b}|| = ||A^{-1}|| = \operatorname{cond}(A)$. Поэтому если норма $||\boldsymbol{x}||$ велика, то велико и $\operatorname{cond}(A)$. 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 δE — матрица ее возмущений, такая, что $||\delta E|| < 1$. Показать, что матрица $E - \delta E$ невырожденная и выполнена оценка

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

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

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

Следовательно, если $\boldsymbol{x} \neq 0$, то $(E-\delta E)\,\boldsymbol{x} \neq 0$, 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

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

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

Таким образом, задача 2 отвечает на вопрос, насколько малым должно быть возмущение δE единичной матрицы E, чтобы возмущенная матрица $E - \delta E$ оставалась нвырожденной.

Можно также показать, что матрица $E + \delta E$ невырожденная, если для матрицы возмущений выполнено неравенство $||\delta E|| < 1$. Действительно, возьмем произвольный ненулевой вектор $\boldsymbol{x}$ и запишем неравенство

$$||\boldsymbol{x}|| = ||\boldsymbol{x} + \delta E\boldsymbol{x} - \delta E\boldsymbol{x}|| \leq ||\boldsymbol{x} + \delta E\boldsymbol{x}|| + ||\delta E\boldsymbol{x}||.$$

Из этого неравенства и известных неравенств $||\boldsymbol{x} + \boldsymbol{y}|| \geq \big|||\boldsymbol{x}|| - ||\boldsymbol{y}||\big|$ (здесь $\boldsymbol{y}$ — ненулевой вектор) и $||\delta E\boldsymbol{x}|| \leq ||\delta E||\,||\boldsymbol{x}||$ получим следующее неравенство:

$$\begin{aligned} &||(E + \delta E)\boldsymbol{x}|| = ||\boldsymbol{x} + \delta E\boldsymbol{x}|| \geq \big|||\boldsymbol{x}|| - ||\delta E\boldsymbol{x}||\big| \geq \big|||\boldsymbol{x}|| - ||\delta E||||\boldsymbol{x}||\big| = \\ &= \big|(1 - ||\delta E\boldsymbol{x}||)||\boldsymbol{x}||\big| = (1 - ||\delta E\boldsymbol{x}||)||\boldsymbol{x}|| > 0. \end{aligned}$$

Следовательно, матрица $E + \delta E$ не вырождена, поскольку $\boldsymbol{x}$ — произвольный ненулевой вектор.

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

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

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

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

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 о значает, 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

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

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 $||Q^T||_2 = ||Q^T E||_2 = ||E||_2 = 1$. 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}:

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

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

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

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

$$\|A^{-1}\|_\infty = n, \quad |a| = 1$$

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

$$\operatorname{cond}_\infty(A) = (|a|+1)\,n, \quad |a| = 1.$$

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 $\operatorname{cond}_\infty(A) \approx 10^{14}$.

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.

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 $i \leq j$, 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. Рассмотрим пример линейной системы $Ax = \boldsymbol{b}$ с хорошо обусловленной матрицей

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

где ε — произвольное малое число.

Показать, что при решении системы наблюдается значительный рост значений промежуточных элементов, если применяется метод LU-разложения без выбора ведущего элемента.

Решение. Сначала вычислим число обусловленности матрицы A, чтобы показать ее хорошую обусловленность. Для этого выпишем матрицу A^{-1}, решая две системы

$$A\boldsymbol{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad A\boldsymbol{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Полученные векторы $\boldsymbol{x}$ и $\boldsymbol{y}$ образуют столбцы матрицы A^{-1}:

$$A^{-1} = \begin{pmatrix} -\dfrac{1}{1-\varepsilon} & \dfrac{1}{1-\varepsilon} \\ \dfrac{1}{1-\varepsilon} & -\dfrac{\varepsilon}{1-\varepsilon} \end{pmatrix} = \frac{1}{1-\varepsilon}\begin{pmatrix} -1 & 1 \\ 1 & -\varepsilon \end{pmatrix}.$$

Поскольку ε — малое число, то можно приближенно записать

$$\mathrm{cond}_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty \approx 4.$$

Таким образом, мы показали хорошую обусловленность матрицы A.

Напомним, что метод LU-разложения без выбора ведущего элемента состоит из вычисления представления матрицы A в виде $A = LU$ с последующим решением двух систем $L\boldsymbol{y} = \boldsymbol{b}$ и $U\boldsymbol{x} = \boldsymbol{y}$. Здесь L — нижняя треугольная матрица с единичной диагональю, а U — верхняя треугольная матрица.

Опустив промежуточные выкладки, для нашего случая получим следующее разложение:

$$A = \begin{pmatrix} \varepsilon & 1 \\ 1 & 1 \end{pmatrix} = LU = \begin{pmatrix} 1 & 0 \\ \dfrac{1}{\varepsilon} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon & 1 \\ 0 & 1 - \dfrac{1}{\varepsilon} \end{pmatrix}.$$

Вычислив L^{-1} и U^{-1}, получим

$$\mathrm{cond}_\infty(L) \approx \frac{2}{\varepsilon}, \quad \mathrm{cond}_\infty(U) \approx \frac{1}{\varepsilon},$$

т.е. множители L и U в выписанном выше разложении очень плохо обусловлены и в них наблюдается значительный рост элементов. Это означает, что вычислительная схема без выбора ведущих элементов может быть неустойчивой для линейных систем с хорошо обусловленными матрицами.

Если применить вычислительную схему с выбором ведущих элементов по столбцам, то получим следующее LU-разложение после перестановки двух строк матрицы A:

$$A = \begin{pmatrix} \varepsilon & 1 \\ 1 & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 \\ \varepsilon & 1 \end{pmatrix} = LU = \begin{pmatrix} 1 & 0 \\ \varepsilon & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 - \varepsilon \end{pmatrix}.$$

Вычислив L^{-1} и U^{-1}, получим

$$\mathrm{cond}_\infty(L) \approx 1, \quad \mathrm{cond}_\infty(U) \approx 4,$$

т.е. множители L и U во втором разложении хорошо обусловлены и в них нет роста элементов.

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 так возмущенной матрицы Уилкинсона имеет вид

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

В условии задачи значение ε выбрано таким образом, что свободный член в этом уравнении равен нулю. Следовательно, наименьшее собственное значение возмущенной матрицы Уилкинсона тоже равно нулю, т.е. при таком возмущении матрица Уилкинсона становится вырожденной.

11. Пусть дана матрица

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

где $a > 1$. 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.

13. Пусть A — вещественная квадратная матрица порядка n. Говорят, что матрица A имеет *строчное диагональное преобладание*, если

$$|a_{ii}| > \sum_{j=1}^{n} |a_{ij}|, \quad i = 1, \dots, n, \quad j \neq i,$$

и *столбцовое диагональное преобладание*, если

$$|a_{jj}| > \sum_{i=1}^{n} |a_{ij}|, \quad j = 1, \dots, n, \quad i \neq j.$$

Показать, что матрица, имеющая строчное или столбцовое диагональное преобладание, является невырожденной.

Решение. Рассмотрим случай строчного диагонального преобладания и введем следующие обозначения:

$$\operatorname{diag}(A) = \operatorname{diag}(a_{11}, \ldots, a_{nn}), \quad \operatorname{off}(A) = A - \operatorname{diag}(A).$$

Тогда матрицу A можно представить в виде $A = \operatorname{diag}(A) + \operatorname{off}(A)$, где матрица $\operatorname{off}(A)$ может рассматриваться как возмущение матрицы $\operatorname{diag}(A)$. Из неравенства

$$\sum_{j=1}^{n} |a_{ij}| < |a_{ii}|, \quad i = 1, \ldots, n, \quad j \neq i$$

следует неравенство

$$\sum_{j=1}^{n} \frac{|a_{ij}|}{|a_{ii}|} < \sum_{j=1}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| < 1, \quad i = 1, \ldots, n, \quad j \neq i,$$

В самом последнем неравенстве стоит сумма модулей элементов i-й строки матрицы $(\operatorname{diag}(A))^{-1} \operatorname{off}(A)$, которая меньше 1. Поскольку эта сумма меньше 1 для всех i, то она меньше 1 и для того i, для которого она максимальна. Это означает, что выполнено неравенство

$$\|(\operatorname{diag}(A))^{-1} \operatorname{off}(A)\|_\infty < 1.$$

Это неравенство можно доказать по-другому. В силу строчного диагонального преобладания матрицы A неравенство $|a_{ii}| > \sum_{j=1}^{n} |a_{ij}|$ $(j \neq i)$ выполнено для всех i, в том числе и для того значения i, при котором правая часть этого неравенства достигает своего максимума: следовательно, $|a_{ii}| > \|\operatorname{off} A\|_\infty$. Это последнее неравенство выполнено для всех диагональных элементов матрицы A, в том числе и для максимального по модулю: $\|\operatorname{diag}(A)\|_\infty > \|\operatorname{off} A\|_\infty$. Отсюда заключаем, что

$$\|\operatorname{diag}(A)\|_\infty^{-1} \|\operatorname{off} A\|_\infty < 1.$$

Поскольку бесконечная матричная норма мультипликативна, то имеем

$$\|(\operatorname{diag}(A))^{-1} \operatorname{off}(A)\|_\infty \leq \|\operatorname{diag}(A)\|_\infty^{-1} \|\operatorname{off} A\|_\infty < 1.$$

В задаче 4 было показано, что если некоторая матрица A не вырождена, то возмущенная матрица $A + \delta A$ тоже не вырождена, если $\|A^{-1}\delta A\| < 1$. В нашем случае матрица $\operatorname{diag}(A)$ не вырождена в силу диагонального преобладания матрицы A, а ее возмущение $\operatorname{off}(A)$ таково, что в соответствии с задачей 4 возмущенная матрица

$$\operatorname{diag}(A) + \operatorname{off}(A)$$

остается невырожденной. Отсюда заключаем, что рассматриваемая в этой задаче 13 матрица $A = \operatorname{diag}(A) + \operatorname{off}(A)$ не вырождена.

Для случая столбцового диагонального преобладания у матрицы A аналогично доказывается невырожденность транспонированной матрицы A^T, которая имеет строчное диагональное преобладание. Отсюда заключаем, что матрица A тоже будет невырожденной.

14. Пусть E — единичная матрица и δE — матрица ее возмущений, такая, что $||\delta E|| < 1$. Показать, что матрица $E + \delta E$ невырожденная.

Решение. Пусть λ — собственное значение матрицы δE: $\delta E\boldsymbol{x} = \lambda\boldsymbol{x}$. Из неравенства $|\lambda| \leq ||\delta E|| < 1$ заключаем, что все собственные значения матрицы δE по модулю меньше 1.

Собственные значения матрицы $E + \delta E$ равны $1 + \lambda$, поскольку

$$(E + \delta E)\boldsymbol{x} = E\boldsymbol{x} + \delta E\boldsymbol{x} = E\boldsymbol{x} + \lambda\boldsymbol{x} = (1 + \lambda)\boldsymbol{x},$$

где $\boldsymbol{x}$ — собственный вектор матрицы δE, соответствующий ее собственному значению λ. Так как $|\lambda| < 1$, то $1 + \lambda \neq 0$; следовательно, матрица $E + \delta E$ не вырождена.

15. Пусть A — квадратная невырожденная матрица. Основываясь на задаче 14 показать, что возмущенная матрица $A + \delta A$ не вырождена, если $\|A^{-1}\delta A\| < 1$.

Решение. Представим матрицу $A + \delta A$ в следующем виде:

$$A + \delta A = A(E + A^{-1}\delta A).$$

В этом произведении первая матрица A не вырождена по условию, а вторая матрица не вырождена в силу задачи 14. Следовательно, матрица $A + \delta A$ тоже не вырождена.

16. Пусть A — квадратная невырожденная матрица порядка n. Доказать неравенство

$$\frac{1}{n} \leq \frac{\mathrm{cond}_1(A)}{\mathrm{cond}_2(A)} \leq n.$$

Решение. Рассмотрим соотношение эквивалентности для первой нормы матрицы A и ее спектральной нормы:

$$\frac{1}{\sqrt{n}}\|A\|_2 \leq \|A\|_1 \leq \sqrt{n}\|A\|_2.$$

Для обратной матрицы A^{-1} аналогичное соотношение эквивалентности записывается в виде

$$\frac{1}{\sqrt{n}}\|A^{-1}\|_2 \leq \|A^{-1}\|_1 \leq \sqrt{n}\|A^{-1}\|_2.$$

Перемножив эти два соотношения эквивалентности, получим

$$\frac{1}{n}\|A\|_2\|A^{-1}\|_2 \leq \|A\|_1\|A^{-1}\|_1 \leq n\|A\|_2\|A^{-1}\|_2.$$

Поделив последнее соотношение на $\|A\|_2\|A^{-1}\|_2$, получим решение задачи:

$$\frac{1}{n} \leq \frac{\|A\|_1\|A^{-1}\|_1}{\|A\|_2\|A^{-1}\|_2} = \frac{\mathrm{cond}_1(A)}{\mathrm{cond}_2(A)} \leq n.$$

17. Показать, что матрица

$$A_n = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & \dots & -1 & 2 \end{pmatrix}$$

положительно определена.

Решение. В соответствии с критерием Сильвестра симметричная матрица A_n положительно определена, если все ее угловые миноры D_n положительны. Разложив D_n по первой строке, получим рекуррентное соотношение

$$D_n = 2D_{n-1} - D_{n-2}.$$

Для упрощения дальнейших выкладок положим здесь $n = n + 2$. Тогда для угловых миноров имеем разностное уравнение

$$D_{n+2} - 2D_{n-1} + D_n = 0, \quad n = 1, 2, 3, \dots,$$

с начальными условиями $D_1 = 2$, $D_2 = 3$.

Характеристическое уравнение $q^2 - 2q + 1 = 0$ имеет кратные корни $q_{1,2} = 1$. Следовательно, $D_n = C_1 q_1^n + C_2 q_2^n = C_1 + nC_2$. Константы C_2 и C_2 ищем из начальных условий и получаем $C_1 = C_2 = 1$. Таким образом, угловые миноры D_n равны $1 + n$ для всех $n = 1, 2, 3, \dots$. Следовательно, выписанная выше матрица положительно определена в соответствии с критерием Сильвестра.

18. Для матрицы A_n из задачи 17 найти собственные значения и оценить ее число обусловленности $\mathrm{cond}_2(A_n)$.

Решение. Искомые собственные значения являются корнями характеристического уравнения $D_n = \det(A_n - \lambda E) = 0$, где E — единичная матрица порядка n. Как и в задаче 17, получим рекуррентное соотношение

$$D_n = (2 - \lambda)D_{n-1} - D_{n-2}, \quad n = 3, 4, 5, \dots,$$

с начальными условиями

$$D_1 = 2 - \lambda, \quad D_2 = (2 - \lambda)^2 - 1.$$

Запишем это соотношение в виде разностного уравнения

$$D_n - (2 - \lambda)D_{n-1} + D_{n-2}.$$

Общее решение этого уравнения записывается в виде

$$D_n = \alpha q_1^n + \beta q_2^n,$$

где q_1 и q_2 — корни характеристического уравнения

$$q^2 - (2 - \lambda)q + 1 = 0.$$

Решая это уравнение, получим

$$q_{1,2} = \frac{2-\lambda}{2} \pm i\sqrt{1-\left(\frac{2-\lambda}{2}\right)}.$$

Теперь запишем эти корни в нормальной тригонометрической форме:

$$q_{1,2} = \cos\varphi \pm i\sin\varphi, \quad \cos\varphi = \frac{2-\lambda}{2}.$$

Круги Гершгорина для матрицы A_n дают два неравенства $|2-\lambda| \le 1$ и $|2-\lambda| \le 2$. Из первого неравенства получим $1 \le \lambda \le 3$, а из второго неравенства получим $0 \le \lambda \le 4$. Поскольку матрица A_n положительно определена (см задачу 17), то $\lambda \ne 0$ и $\cos\varphi \ne 1$. Таким образом, мы пришли к следующей локализации собственных значений: $0 < \lambda \le 4$.

Общее решение разностного уравнения примет вид

$$\begin{aligned} D_n &= \alpha(\cos n\varphi + i\sin n\varphi) + \beta(\cos n\varphi - i\sin n\varphi) = \\ &= (\alpha+\beta)\cos n\varphi + i(\alpha-\beta)\sin n\varphi = \alpha'\cos n\varphi + \beta'\sin n\varphi. \end{aligned}$$

Теперь выпишем уравнения для определения констант α' и β'. Поскольку $\cos\varphi = \frac{2-\lambda}{2}$, начальные условия можно переписать в виде

$$D_1 = 2-\lambda = 2\cos\varphi, \quad D_2 = (2-\lambda)^2 - 1 = 4\cos^2\varphi - 1.$$

Отсюда приходим к искомой системе уравнений:

$$\begin{aligned} &\alpha'\cos\varphi + \beta'\sin\varphi = 2\cos\varphi, \\ &\alpha'\cos 2\varphi + \beta'\sin 2\varphi = 4\cos^2\varphi - 1. \end{aligned}$$

Решив эту систему, получим

$$\alpha' = 1, \quad \beta' = \frac{\cos\varphi}{\sin\varphi}.$$

Следовательно, характеристический определитель принимает вид

$$D_n = \cos n\varphi + \frac{\cos\varphi}{\sin\varphi}\sin n\varphi = \frac{\sin(n+1)\varphi}{\sin\varphi}.$$

Выше мы отметили, что $\cos\varphi = \frac{2-\lambda}{2} \ne 1$, поскольку матрица A_n положительно определена и все $\lambda > 0$. Следовательно, $\sin\varphi \ne 0$. Это означает, что из характеристического уравнения $D_n = 0$ мы получим уравнение

$$\sin(n+1)\varphi = 0,$$

решение которого запишем в виде

$$\varphi_k = \frac{\pi k}{n+1}, \quad k = 1, \ldots, n.$$

Здесь значения $k = 0$ и $k = n + 1$ мы не учитываем, поскольку при этих значениях k имеем $\sin\varphi = 0$, что невозможно в силу положительной определенности матрицы A_n.

Поскольку
$$\frac{2 - \lambda_k}{2} = \cos\frac{\pi k}{n+1},$$
то
$$\lambda_k = 2\left(1 - \cos\frac{\pi k}{n+1}\right) = 4\sin^2\frac{\pi k}{2(n+1)}.$$
Следовательно, значения λ_k лежат в интервале от 0 до 4 и сгущаются к концам этого интервала.

Поскольку матрица A_n положительно определена, то
$$\mathrm{cond}_2(A_n) = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{4\sin^2\frac{\pi n}{2(n+1)}}{4\sin^2\frac{\pi}{2(n+1)}} \approx \frac{1}{\left(\frac{\pi}{2(n+1)}\right)^2} \approx \frac{4}{\pi^2}\, n^2.$$

19. Пусть A — квадратная матрица порядка n. Доказать, что
$$\mathrm{cond}_2(A^T A) = \mathrm{cond}_2^2(A).$$

Решение. Сначала покажем справедливость равенства $(A^{-1})^T = (A^T)^{-1}$. Предположим, что это равенство справедливо и умножим обе его части справа на A^T. В результате получим тождество:
$$(A^{-1})^T A^T = E, \quad (AA^{-1})^T = E, \quad E^T = E,$$
где E — единичная матрица порядка n. Таким образом, справедливость равенства показана.

Напомним следующие два свойства спектральной нормы: $\|A^T\|_2 = \|A\|_2$ и $\|A\|_2^2 = \|A^T A\|_2$. Кроме того, собственные значения и спектральные нормы матриц $A^T A$ и AA^T совпадают. Сказанное выше позволяет выписать следующую цепочку равенств:
$$\|A^{-1}\|_2^2 = \|(A^{-1})^T A^{-1}\|_2 = \|(A^T)^{-1}A^{-1}\|_2 = \|(AA^T)^{-1}\|_2 = \|(A^T A)^{-1}\|_2.$$
Отсюда получаем решение задачи:
$$\mathrm{cond}_2^2(A) = \|A\|_2^2\|A^{-1}\|_2^2 = \|A^T A\|_2\|(A^T A)^{-1}\|_2 = \mathrm{cond}_2(A^T A).$$

Литература

1. *Арушанян И.О., Чижонков Е.В.* Материалы семинарских занятий по курсу “Методы вычислений” / под ред. Арушаняна О.Б. М.: Изд-во ЦПИ при механико-математическом факультете МГУ, 1999.

2. *Бахвалов Н.С., Корнев А.А., Чижонков Е.В.* Численные методы решения задач и упражнения. М.: Дрофа, 2009.

3. *Бахвалов Н.С.* Численные методы. М.: Наука, 1975.

4. *Бахвалов Н.С., Жидков Н.П., Кобельков Г.М.* Численные методы. М.: Наука, 1987.

5. Библиотека НИВЦ МГУ решения типовых задач численного анализа (http://num-anal.srcc.msu.ru/).

6. *Демидович Б.П., Марон И.А.* Основы вычислительной математики. М.: Физматгиз, 1963.

7. *Калиткин Н.Н.* Численные методы. М.: Наука, 1978.

8. *Каханер Д., Моулер К., Нэш С.* Численные методы и программное обеспечение. М.: Мир, 1998.

9. *Крылов В.И., Бобков В.В., Монастырный П.И.* Вычислительные методы. М.: Наука, 1977.