

**SOME NUMERICAL
METHODS FOR SOLVING
LINEAR ALGEBRAIC
EQUATIONS**

V. V. DIKOUSSAR

The brief review of some direct and iterative methods is given for well-posed linear problems. A special attention has been focused on ill-posed systems. The examples are presented.

Дается краткий обзор некоторых прямых и итеративных методов для решения хорошо обусловленных линейных задач. Особое внимание уделено некоторым некорректным системам. Приводятся примеры.

**НЕКОТОРЫЕ ЧИСЛЕННЫЕ
МЕТОДЫ РЕШЕНИЯ ЛИНЕЙНЫХ
АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ**

В. В. ДИКУСАР

Московский физико-технический институт,
Долгопрудный Московской обл.

ВВЕДЕНИЕ

Проблема численного решения линейных уравнений интересует математиков уже несколько столетий. Первые математические результаты появились в XVIII веке. В 1750 году Г. Крамер (1704–1752) опубликовал свои труды по детерминантам квадратных матриц и предложил алгоритм нахождения обратной матрицы, известный как правило Крамера. Гаусс в 1809 году опубликовал работу, посвященную движению небесных тел, в которой был изложен метод для решения линейных систем, известный как метод исключения.

В 40-х годах XX века с появлением компьютеров сильно возрос интерес к численным методам. Тогда же началось активное исследование существующих методов для их реализации на ЭВМ и предпринимались активные попытки увеличить их точность.

Вплоть до 80-х годов решение вычислительных задач было ограничено ресурсами ЭВМ, поэтому особое значение придавалось экономичности алгоритмов.

В настоящее время ограничения по оперативной памяти и быстродействию ЭВМ потеряли актуальность в связи с появлением относительно дешевых мини- и суперкомпьютеров.

Целью статьи является краткое изложение в идейном плане некоторых прямых и итерационных методов решения линейных систем. Для облегчения чтения статьи в приложении даются краткие сведения по векторному и матричному анализу, необходимые для понимания работы.

УРАВНЕНИЕ С ОДНИМ НЕИЗВЕСТНЫМ

Уравнение вида

$$ax - b = 0,$$

где a и b – выражения, зависящие только от параметров, а x – неизвестное, называется линейным относительно x . Это уравнение приводится к виду $ax = b$ и при $a \neq 0$ имеет единственное решение $x = b/a$ при допустимом множестве значений параметров.

При $a = 0$ и $b = 0$ имеем: x – любое число, $x \in \mathbf{R}$.

При $b \neq 0$ и $a = 0$ решений нет, $x \in \emptyset$.

Пример. $(k^2 - 1)x = 2k^2 + k - 3$. Это уравнение линейно относительно x и имеет смысл при любых действительных значениях параметра k . Оно приводится к виду

$$(k - 1)(k + 1)x = (2k + 3)(k - 1).$$

При $k = 1$ его решением является любое действительное число, $x \in \mathbf{R}$. При $k = -1$ уравнение имеет вид $0 \cdot x = -2$, то есть не имеет решений; $x \in \emptyset$.

Уравнение имеет единственное решение при $k \neq \pm 1$:

$$x = \frac{2k + 3}{k + 1}.$$

ЛИНЕЙНЫЕ СИСТЕМЫ ДВУХ УРАВНЕНИЙ С ДВУМЯ НЕИЗВЕСТНЫМИ

Линейной системой двух уравнений с двумя неизвестными называется система вида

$$\begin{cases} a_1x + b_1y = c_1, & a_1^2 + b_1^2 \neq 0, \\ a_2x + b_2y = c_2, & a_2^2 + b_2^2 \neq 0. \end{cases} \quad (1)$$

Для решения таких систем применяются метод подстановки, метод линейного преобразования и правило Крамера [1–5].

Метод подстановки основан на равносильности систем уравнения

$$\begin{cases} x = R(y), \\ F(x, y) = 0 \end{cases} \quad \text{и} \quad \begin{cases} x = R(y), \\ F(R(y), y) = 0. \end{cases}$$

(Системы уравнений равносильны, если множества их решений совпадают.)

Пример 1. Решить систему уравнений

$$\begin{cases} 2x + y = 4, \\ 5x - 3y = -1. \end{cases}$$

Данная система равносильна системе

$$\begin{cases} y = 4 - 2x, \\ 5x - 3y = -1. \end{cases}$$

Отсюда следует, что исходная система равносильна системе

$$\begin{cases} y = 4 - 2x, \\ 5x - 3(4 - 2x) = -1. \end{cases}$$

После тождественных преобразований получаем $x = 1$, $y = 2$. Следовательно, исходная система имеет единственное решение $(1; 2)$.

Метод линейного преобразования системы основан на утверждении о равносильности систем уравнений

$$\begin{cases} F_1(x, y) = 0, \\ F_2(x, y) = 0 \end{cases} \quad \text{и} \quad \begin{cases} \alpha F_1(x, y) + \beta F_2(x, y) = 0, \\ F_2(x, y) = 0, \quad \alpha \neq 0, \quad \beta \neq 0. \end{cases}$$

Пример 2. Решить систему уравнений

$$\begin{cases} 2x + y = 3, \\ 3x - 2y = 1. \end{cases}$$

Умножив первое уравнение на 2, получим систему

$$\begin{cases} 4x + 2y = 6, \\ 3x - 2y = 1. \end{cases}$$

Складывая первое и второе уравнения, получим систему, равносильную исходной:

$$\begin{cases} 7x = 7, \\ 3x - 2y = 1. \end{cases}$$

Отсюда получаем, что исходная система имеет единственное решение $(1; 1)$.

Правило Крамера для нахождения решения системы (1) основано на вычислении соответствующих определителей [4]

$$x = \frac{\Delta_1}{\Delta}, \quad y = \frac{\Delta_2}{\Delta}, \quad \Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1, \quad (2)$$

$$\Delta_1 = \begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix} = c_1b_2 - c_2b_1,$$

$$\Delta_2 = \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} = a_1c_2 - a_2c_1.$$

Здесь предполагается, что $\Delta \neq 0$. В этом случае система (1) имеет единственное решение.

Пример 3. Решить систему уравнений

$$\begin{cases} 7x + 4y = 18, \\ 5x + 3y = 13. \end{cases}$$

По правилу Крамера имеем

$$\Delta = \begin{vmatrix} 7 & 4 \\ 5 & 3 \end{vmatrix} = 21 - 20 = 1,$$

$$\Delta_1 = \begin{vmatrix} 18 & 4 \\ 13 & 3 \end{vmatrix} = 54 - 52 = 2, \quad \Delta_2 = \begin{vmatrix} 7 & 18 \\ 5 & 13 \end{vmatrix} = 91 - 90 = 1,$$

$$x = \frac{2}{1} = 2, \quad y = \frac{1}{1} = 1.$$

В общем случае при решении двух линейных уравнений (1) возможны три ситуации:

• система имеет единственное решение; это имеет место при

$$\frac{a_1}{a_2} \neq \frac{b_1}{b_2}, \quad \Delta \neq 0;$$

• система имеет бесконечно много решений. В этом случае

$$\frac{a_1}{a_2} = \frac{b_1}{b_2} = \frac{c_1}{c_2}, \quad \Delta = 0, \quad \Delta_1 = 0, \quad \Delta_2 = 0;$$

• система не имеет решений. Указанный случай имеет место при

$$\frac{a_1}{a_2} = \frac{b_1}{b_2} \neq \frac{c_1}{c_2}, \quad \Delta = 0, \quad \Delta_1 \neq 0; \quad \Delta = 0, \quad \Delta_2 \neq 0.$$

Заметим, что приведенные пропорции имеют смысл и в том случае, когда некоторые из знаменателей равны 0.

Пример 4. Решить систему уравнений

$$\begin{cases} x + y = 3, \\ 3x + 3y = 9. \end{cases}$$

Разделив второе уравнение на 3, получим систему, равносильную исходной:

$$\begin{cases} x + y = 3, \\ x + y = 3. \end{cases}$$

Полученная система состоит из двух одинаковых уравнений. Положим $x = p$, где $p \in \mathbf{R}$. Тогда $y = 3 - p$. Следовательно, решениями исходной системы являются все пары чисел вида $(p; 3 - p)$, $p \in \mathbf{R}$.

Пример 5. Решить систему уравнений

$$\begin{cases} x + y = 2, \\ 2x + 2y = 6. \end{cases}$$

Равносильная система имеет вид

$$\begin{cases} x + y = 2, \\ x + y = 3. \end{cases}$$

Эта система противоречива, то есть исходная система решений не имеет.

Ответ: $x \in \emptyset, y \in \emptyset$.

СИСТЕМА ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

Пусть дана система m линейных уравнений с n неизвестными

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \dots & \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m \end{aligned} \quad (3)$$

Систему (3) можно записать в матрично-векторной форме

$$AX = B, \quad (4)$$

где A — матрица коэффициентов, содержащая m строк и n столбцов:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}. \quad (5)$$

Здесь B — заданный m -компонентный вектор, X — искомый n -компонентный вектор;

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}. \quad (6)$$

Формы записи (3), (4) определяют операцию умножения матрицы A (5) на вектор X (6).

ФОРМУЛЫ КРАМЕРА

Пусть число неизвестных равно числу уравнений, то есть $n = m$. Далее предположим, что система (3) невырождена, то есть ее определитель (детерминант) отличен от нуля: $\Delta = \det A \neq 0$. Тогда она имеет единственное решение, которое можно вычислить по формулам Крамера

$$x_i = \frac{\Delta_i}{\Delta}, \quad i = 1, 2, \dots, n. \quad (7)$$

Здесь $\Delta_i = \det A_i$, матрица A_i получается из матрицы A заменой i -го столбца столбцом правых частей B :

$$A_i = \begin{pmatrix} a_{11} & \dots & a_{1i-1} & b_1 & a_{1i+1} & \dots & a_{1n} \\ a_{21} & \dots & a_{2i-1} & b_2 & a_{2i+1} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{ni-1} & b_n & a_{ni+1} & \dots & a_{nn} \end{pmatrix}.$$

Хотя формулы Крамера дают решение системы (3) в явном виде, они неэффективны с вычислительной точки зрения [1], за исключением специальных случаев.

Пример. Если вычислять определитель непосредственно по его определению, то нужно выполнить $(n-1)n!$ умножений и $n!$ сложений. Положим $n = 15$ и возьмем ЭВМ с производительностью 10^6 умножений в секунду. Тогда вычисление только одного определителя потребует $14 \cdot 15! \approx 1,8 \cdot 10^{13}$ умножений, в результате чего на вычисление решения уйдет около 10 лет непрерывной работы ЭВМ. Вместе с тем для решения такой системы аналогичной

ЭВМ потребуется примерно 0,002 секунд с использованием метода Гаусса.

МЕТОД ГАУССА

Метод последовательного исключения неизвестных Гаусса является одним из наиболее универсальных и эффективных методов решения линейных систем. (Карл Фридрих Гаусс (1777–1855) – немецкий математик и физик, работы которого оказали большое влияние на дальнейшее развитие высшей алгебры, геометрии, теории чисел, теории электричества и магнетизма.) Этот метод известен в различных вариантах уже более 2000 лет. Он относится к числу прямых методов.

Процесс решения по методу Гаусса состоит из двух этапов, называемых прямым и обратным ходом. На первом этапе система (3) приводится к треугольному виду; на втором (обратный ход) идет последовательное определение неизвестных из указанной треугольной системы.

Для реализации метода Гаусса требуется примерно $(2/3)n^3$ арифметических операций, причем основное число этих действий совершается на этапе прямого хода [1]. В качестве примера, иллюстрирующего метод Гаусса, рассмотрим следующую систему [6].

Пример 1.

$$\begin{aligned} 1,2357x_1 + 2,1742x_2 - 5,4834x_3 &= -2,0735, \\ 6,0696x_1 - 6,2163x_2 - 4,6921x_3 &= -4,8388, \\ 3,4873x_1 + 6,1365x_2 - 4,7483x_3 &= 4,8755. \end{aligned} \quad (8)$$

Все результаты расчетов представлены в виде чисел с плавающей запятой с пятью значащими цифрами.

Первый шаг гауссова исключения сводится сначала к умножению первого уравнения на $(-4,9119)$, а затем на $(-2,8221)$. Полученные уравнения складываются со вторым и третьим уравнениями системы (8). В результате получаем равносильную систему

$$\begin{aligned} 1,2357x_1 + 2,1742x_2 - 5,4834x_3 &= -2,0735, \\ -16,895x_2 + 22,242x_3 &= 5,3462, \\ 0,0007x_2 + 10,727x_3 &= 10,727. \end{aligned}$$

Чтобы сделать второй шаг, умножим второе уравнение на $\frac{0,0007}{16,895}$ и сложим с третьим уравнением. С учетом этого получим систему треугольного вида

$$\begin{aligned} 1,2357x_1 + 2,1742x_2 - 5,4834x_3 &= -2,0735, \\ -16,895x_2 + 22,242x_3 &= 5,3462, \\ 10,728x_3 &= 10,727. \end{aligned} \quad (9)$$

Решение указанной системы дает следующие значения неизвестных:

$$x_1 = 0,99968, \quad x_2 = 0,99994, \quad x_3 = 0,9991. \quad (10)$$

Этот результат близок к точному решению системы (8)

$$x_1 = x_2 = x_3 = 1.$$

Отметим, что первый этап метода Гаусса может быть использован для вычисления определителя системы (8). Действительно, прямой ход метода Гаусса основан на том, что многократно выполняется операция сложения одной из строк матрицы (5) с другой строкой, взятой с некоторым множителем. Известно [1, 4], что такая операция не меняет определителя. Определитель треугольной матрицы равен произведению ее диагональных элементов. В нашем случае для системы (8)

$$\det R = 1,2357 \cdot (-16,895) \cdot 10,728 = -223,97,$$

где R – треугольная матрица системы (9).

Подчеркнем, что метод Гаусса также устанавливает факт отсутствия решения системы (3) (несовместность), а также неединственность.

Пример 2. Решить систему уравнений методом Гаусса:

$$\begin{aligned} x_1 + x_2 + x_3 &= 3, \\ 2x_1 + 3x_2 + 4x_3 &= 9, \\ 2x_1 + 2x_2 + 2x_3 &= 6. \end{aligned}$$

В результате применения метода Гаусса к данной системе получаем следующий результат:

$$\begin{aligned} x_1 + x_2 + x_3 &= 3, \\ x_2 + 2x_3 &= 3, \\ 0 \cdot x_3 &= 0. \end{aligned}$$

Отсюда видно, что указанная система имеет неединственное решение.

Пример 3. Решить систему уравнений методом Гаусса:

$$\begin{aligned} x_1 + x_2 + x_3 &= 3, \\ 2x_1 + 3x_2 + 4x_3 &= 9, \\ 2x_1 + 2x_2 + 2x_3 &= 7. \end{aligned}$$

Используя процедуру Гаусса, получаем следующую систему:

$$\begin{aligned} x_1 + x_2 + x_3 &= 3, \\ x_2 + 2x_3 &= 3, \\ 0 \cdot x_3 &= 1. \end{aligned}$$

Таким образом, данная система не имеет решений, то есть несовместна:

$$x_1, x_2, x_3 \in \emptyset.$$

УМЕНЬШЕНИЕ ОШИБОК ОКРУГЛЕНИЯ

Поменяем местами в системе (8) второе и третье уравнения. Полученную систему решим методом Гаусса. В результате после исключения из третьего уравнения x_2 получим систему треугольного вида

$$\begin{aligned} 1,2357x_1 + 2,1742x_2 - 5,4834x_3 &= -2,0735, \\ 0,0007x_2 + 10,727x_3 &= 10,727, \\ 258\,930x_3 &= 258\,910. \end{aligned}$$

Определяя из нее последовательно x_3 , x_2 , x_1 , найдем решение

$$x_1 = 2,9021, \quad x_2 = 1,4286, \quad x_3 = 0,9992. \quad (11)$$

При осуществлении идеального вычислительного процесса без ошибок округления ответы (10), (11) должны совпадать с точным решением. В одном случае получается вполне разумный ответ (10), в другом (11) в ответе возникает большая ошибка. Причина такого различия связана с малым значением коэффициента второго уравнения при x_2 , что, в свою очередь, приводит к резкому увеличению коэффициента в третьем уравнении.

Чтобы избежать нежелательного роста элементов матрицы во время прямого хода, обычно на каждом шаге производят перестановку строк матрицы, выбирая максимальный первый коэффициент. Существуют и другие более сложные и надежные схемы исключения, однако мы не будем останавливаться на их описании.

Любая специальная схема для повышения точности требует дополнительных вычислительных ресурсов. Но в конечном счете это себя оправдывает благодаря уменьшению ошибок округления.

ИТЕРАЦИОННЫЕ МЕТОДЫ

Рассмотрим итерационные методы решения систем линейных уравнений. Вообще говоря, эти методы не дают возможности найти точное решение за конечное число итераций даже при отсутствии ошибок округления. Суть метода состоит в построении векторной последовательности $\{\mathbf{X}_k\}$, такой, что

$$\lim_{k \rightarrow \infty} \mathbf{X}_k = \mathbf{X},$$

где \mathbf{X} – точное решение.

Указанные методы широко применяются при численном решении дифференциальных уравнений с частными производными. Матрицы их дискретных аналогов имеют большое число нулевых элементов. Итерационные методы в отличие от прямых методов сохраняют структуру исходной матрицы (не увеличивают число ненулевых элементов). Эффективность итерационных методов определяется скоростью сходимости последовательных приближений \mathbf{X}_k к решению \mathbf{X} .

Первым шагом в итерационном методе является преобразование исходной системы (4) к виду

$$A_1\mathbf{X} = A_2\mathbf{X} + \mathbf{B}_1. \quad (12)$$

Здесь матрицы A_1 , A_2 и вектор \mathbf{B}_1 определяются по матрице A и вектору \mathbf{B} . Системы (4) и (12) эквивалентные, то есть их решения совпадают.

Вторым шагом является расстановка индексов или номеров приближений в формуле (12) и задание нулевого приближения. Например,

$$A_1\mathbf{X}_{k+1} = A_2\mathbf{X}_k + \mathbf{B}_1, \quad k = 0, 1, 2, \dots, \quad (13)$$

где \mathbf{X}_0 – заданный вектор (нулевое приближение).

Третьим шагом итерационного метода является обоснование сходимости последовательных приближений $\{\mathbf{X}_k\}$ (13) к точному решению \mathbf{X} системы (4) и оценка погрешности k -го приближения

$$\|\mathbf{X}_k - \mathbf{X}\| \leq \varepsilon. \quad (14)$$

Оценка такого типа при заданном ε (точность решения) позволяет закончить итерационный процесс (13).

Здесь также предполагается, что система (13) решается относительно \mathbf{X}_{k+1} гораздо легче, чем исходная система (4) относительно \mathbf{X} .

Разные итерационные методы различаются первыми двумя шагами, а выбор конкретного метода производится на основании оценки (14).

МЕТОД ЯКОБИ

Метод Якоби проиллюстрируем на следующей линейной системе (Карл Густав Якоб Якоби (1804–1851) – немецкий математик):

$$\begin{aligned} 3x_1 + 4x_2 - x_3 &= 7, \\ 2x_1 + 6x_2 + 3x_3 &= -2, \\ -x_1 + x_2 + 4x_3 &= 4. \end{aligned}$$

Перепишем указанную матрицу в виде

$$\begin{aligned} x_1 &= \frac{1}{3}[-4x_2 + x_3] + \frac{7}{3}, \\ x_2 &= \frac{1}{6}[-2x_1 - 3x_3] - \frac{1}{3}, \\ x_3 &= \frac{1}{4}[x_1 - x_2] + 1, \end{aligned}$$

или в матрично-векторных обозначениях

$$\mathbf{X} = A_2\mathbf{X} + \mathbf{B}_1,$$

где

$$A_2 = \begin{pmatrix} 0 & \frac{4}{3} & \frac{1}{3} \\ -\frac{1}{3} & 0 & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{4} & 0 \end{pmatrix}, \quad \mathbf{B}_1 = \begin{pmatrix} \frac{7}{3} \\ -\frac{1}{3} \\ 1 \end{pmatrix}.$$

Отсюда следует, что мы решили j -е уравнение явно относительно j -го неизвестного и перенесли все остальное в правую часть. Теперь метод Якоби состоит в определении начального \mathbf{X}_0 приближения и применении итерации типа (13)

$$\mathbf{X}_{k+1} = A_2 \mathbf{X}_k + \mathbf{B}_1, \quad k = 0, 1, 2, \dots \quad (15)$$

Этот метод зависит от нумерации уравнений и неизвестных.

Справедлива следующая

Теорема [4]. Пусть выполнено условие $\|A_2\| < 1$ для процесса (15). Тогда:

1) решение \mathbf{X} системы (4) существует и единственно;

2) метод Якоби сходится при произвольном начальном приближении \mathbf{X}_0 и справедлива оценка погрешности

$$\|\mathbf{X}_k - \mathbf{X}\| \leq \|A_2\|^k \|\mathbf{X}_0 - \mathbf{X}\|.$$

МЕТОД ГАУССА–ЗЕЙДЕЛЯ

Вторым старым итерационным методом является метод Гаусса–Зейделя (Людвиг Зейдель (1821–1896) – немецкий астроном и математик). Отметим, что этот метод не был известен Зейделю и презирился Гауссом как бесполезный; таковы капризы исторической точности в науке. Изложим идею метода для предыдущей системы. Для этой цели переписем ее в виде

$$3x_1 = 7 - 4x_2 + x_3,$$

$$2x_1 + 6x_2 = -2 - 3x_3,$$

$$-x_1 + x_2 + 4x_3 = 4.$$

Здесь в j -м уравнении мы перенесли в правую часть все члены, содержащие x_i , для $i > j$. Эта запись может быть представлена в виде

$$(L + D)\mathbf{X} = -U\mathbf{X} + \mathbf{B}_1,$$

где

$$D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \quad L = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ -1 & 1 & 0 \end{pmatrix},$$

$$U = \begin{pmatrix} 0 & 4 & -1 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{B}_1 = \begin{pmatrix} 7 \\ -2 \\ 4 \end{pmatrix}.$$

В принятых обозначениях D означает диагональ матрицы A , U – ее верхнюю треугольную часть и L – ее нижнюю треугольную часть. Итеративный процесс в методе Гаусса–Зейделя строится по формуле

$$(L + D)\mathbf{X}_{k+1} = -U\mathbf{X}_k + \mathbf{B}_1, \quad k = 0, 1, 2, \dots,$$

после выбора соответствующего начального приближения \mathbf{X}_0 . Метод Гаусса–Зейделя можно рассматривать как модификацию метода Якоби. Основная идея модификации состоит в том, что новые значения \mathbf{X}_k используются здесь сразу же по мере получения, в то время как в методе Якоби они не используются до следующей итерации. Приведем достаточное условие сходимости метода.

Теорема [1, 4]. Пусть $\|A_2\| < 1$, где $A_2 = -(L + D)^{-1}U$, $(L + D)^{-1}$ – матрица, обратная к $(L + D)$. Тогда при любом выборе начального приближения \mathbf{X}_0 метод Гаусса–Зейделя сходится со скоростью геометрической прогрессии, знаменатель которой $q \leq \|A_2\|$, и верна оценка погрешности $\|\mathbf{X}_k - \mathbf{X}\| \leq q^k \|\mathbf{X}_0 - \mathbf{X}\|$.

ОБУСЛОВЛЕННОСТЬ МАТРИЦ

Матрица A и правая часть \mathbf{B} системы (4) во многих случаях задаются приближенно. Причины погрешностей могут быть самые разные – от ошибок округления при вводе чисел в ЭВМ до ошибок измерения, если система связана с обработкой экспериментальных данных. Ошибки вносит также процесс вычислений. Возникает вопрос, как все это влияет на точность получаемого решения. Для ответа следует познакомиться с особой характеристикой матриц, которую называют обусловленностью [6].

Рассмотрим уравнение $\mathbf{y} = A\mathbf{x}$, где матрица A ставит в соответствие любому вектору \mathbf{x} вектор \mathbf{y} . Считая вектор \mathbf{x} принадлежащим единичной сфере

$$\|\mathbf{x}\| = \left[\sum_{i=1}^n x_i^2 \right]^{1/2} = 1, \quad (16)$$

подсчитаем длину вектора $\mathbf{y} = A\mathbf{x}$:

$$\|\mathbf{y}\| = \left[\sum_{i=1}^n y_i^2 \right]^{1/2} = \left[\sum_{i=1}^n \left(\sum_{j=1}^n a_{ij} x_j \right)^2 \right]^{1/2} = \varphi(\mathbf{x}). \quad (17)$$

Функция $\varphi(x)$ непрерывна на ограниченном замкнутом множестве (16) и поэтому достигает на нем своего максимума и минимума. Тогда

$$M = \max_{\|x\|=1} \|Ax\| = \max_{\|x\|=1} \varphi(x), \quad (18)$$

$$m = \min_{\|x\|=1} \|Ax\| = \min_{\|x\|=1} \varphi(x). \quad (19)$$

Заметим, что $m \geq 0$; $m = 0$ тогда и только тогда, когда матрица A вырождена ($\det A = 0$) и однородное уравнение $Ax = 0$ имеет нетривиальное решение ($x \neq 0$).

Отношение $\mu = \frac{M}{m} \geq 1$ называется обусловленностью матрицы A .

Из (18), (19) следует двойное неравенство

$$m \leq \|Ax\| \leq M, \quad \|x\| = 1. \quad (20)$$

Для произвольного вектора x в силу линейности $y = Ax$ неравенство (20) принимает вид

$$m\|x\| \leq \|Ax\| \leq M\|x\|. \quad (21)$$

С помощью этого неравенства исследуем устойчивость решения системы (4) по отношению к малым изменениям входящих в нее величин. Для простейшего случая, когда матрица A задана точно, а правая часть — приблизительно, можно путем несложных вычислений получить оценку

$$\frac{\|\Delta x\|}{\|x\|} \leq \mu \frac{\|\Delta B\|}{\|B\|}, \quad A\Delta x = \Delta B. \quad (22)$$

Величина $\|\Delta B\|/\|B\|$ характеризует относительное возмущение правой части, а $\|\Delta x\|/\|x\|$ — относительную ошибку в решении, вызванную этим возмущением. Обусловленность матрицы играет в (22) роль множителя, определяющего максимально возможное увеличение ошибки. Если μ близко к единице, то система хорошо обусловлена: для такой системы относительная ошибка в решении сравнима с относительной погрешностью задания правой части. По мере увеличения чувствительность решения к погрешности в правой части возрастает — система становится плохо обусловленной.

Пример. Рассмотрим систему

$$x_1 + 0 \cdot x_2 = 1, \quad \det A = 0,01 \neq 0, \quad (23)$$

$$x_1 + 0,01x_2 = 1, \quad x_1 = 1, \quad x_2 = 0.$$

Изменим в системе (23) правую часть:

$$x_1 + 0 \cdot x_2 = 1, \quad x_1 = 1, \quad x_2 = 1,$$

$$x_1 + 0,01x_2 = 1,01.$$

Видно, что небольшое возмущение правой части системы (23) привело к существенному изменению решения. Это означает, что матрица A системы (23) плохо обусловлена. Расчеты дают $\mu(A) \approx 200$.

Разобранный пример показывает, какие трудности могут возникать при решении реальных систем с приближенно заданной правой частью при больших μ .

АНАЛИЗ ОШИБОК

Основная цель анализа ошибок состоит в оценке $\|X - X_*\|$, где X — истинное решение, а X_* — вычисленное решение. К сожалению, практические оценки $\|X - X_*\|$ получить очень трудно, а в большинстве случаев невозможно. Перечислим возможные вопросы при анализе влияния ошибок и неопределенных данных на ответ задачи.

1. Какова максимальная ошибка, которая может иметь место?
2. Какие ошибки на самом деле были?
3. Каких ошибок следует ожидать?
4. Насколько задача чувствительна к ошибкам?
5. Насколько близка фактически решенная задача к решаемой задаче?

Общий эффект ошибок обычно учитывают изучением воздействия ошибок входных данных, метода и округлений на получаемый результат X , и пытаются оценить некоторую меру близости X_* к истинному решению X . Такой метод исследования называют прямым анализом ошибок. Особенно трудным является прямой анализ вычислительной погрешности. Отсюда следует вывод, что первые четыре вопроса либо не имеют ответов, либо ответы на них дают мало информации о действительной величине ошибки [1, 3, 5].

На пятый вопрос может быть дан удовлетворительный ответ с помощью обратного анализа ошибок. Согласно этому анализу, приближенное решение какой-либо задачи, например системы линейных уравнений (4), является точным решением некоторой близкой задачи, в нашем примере системы линейных уравнений с немного измененными коэффициентами. Вместо того чтобы оценивать отличие приближенного решения от точного, можно оценить отличие коэффициентов исходной и измененной системы. Обратный анализ ошибок округления реализуется намного проще, чем прямой.

Пример. Пусть коэффициенты системы линейных уравнений получены в результате физических измерений и известны с некоторой ошибкой. Если обратный анализ показывает, что влияние ошибок округления равносильно искажению коэффициентов, не превосходящему ошибок измерения, то можно считать, что вычисления проводятся с достаточной точностью. При этом не следует интересоваться отличием вычисленного решения от точного решения системы (4), которая сама определена недостаточно точно.

ПОЧТИ-ВЫРОЖДЕННЫЕ МАТРИЦЫ

Возмущение коэффициентов системы линейных уравнений может не просто немного исказить ее решение, а иметь более серьезные последствия. Например, система линейных уравнений [7]

$$x + 0,99y = 1,01,$$

$$x + 1,01y = 0,99$$

имеет единственное решение, но за счет изменения коэффициентов и свободных членов на 1% можно получить как противоречивую систему, так и систему, имеющую бесконечно много решений.

В приведенном примере детерминант матрицы системы мал по сравнению с ее коэффициентами. Однако можно привести пример матрицы, детерминант которой не мал, но обращается в нуль при малом возмущении элементов матрицы. Рассмотрим диагональную матрицу пятого порядка, у которой все элементы на диагонали равны 10, кроме последнего, равного 10^{-4} . Детерминант такой матрицы равен 1, но может обратиться в 0 изменением одного элемента на 10^{-4} . Следует отметить, что такая матрица является плохо обусловленной и ее число обусловленности μ равно 10^4 .

Сейчас важно подчеркнуть принципиальную сторону этого явления. Если числа не могут быть заданы точно, то стирается четкая грань между вырожденными и невырожденными матрицами. Появляется класс почти-вырожденных матриц, границы которого зависят от принятой меры точности в конкретном исследовании. В пределах заданной точности можно получить как вырожденную, так и невырожденную матрицу, изменяя ее элементы.

НОРМАЛЬНАЯ СИСТЕМА

В практических задачах часто бывает необходимо удовлетворить некоторому числу противоречивых требований. Если задача сводится к системе линейных уравнений, то такая система, вообще говоря, несовместна. В этом случае задачу можно решить путем выбора некоторого компромисса — все требования могут быть удовлетворены не полностью, а лишь до некоторой степени.

Рассмотрим систему (4) с матрицей A для произвольных размеров $m \times n$, то есть никаких условий на m и n не накладывается.

Невязкой называется вектор, который дает значение вектора X при подстановке в систему (4)

$$R = B - AX.$$

Решение системы — это вектор X , дающий нулевую невязку.

Если система (4) несовместна, то обычно стараются найти такой X , который дает невязку с минимальной нормой. Под нормой вектора R понимается его евклидова норма

$$\|R\| = [r_1^2 + r_2^2 + \dots + r_n^2]^{1/2}.$$

Если такой X существует, то его считают обобщенным решением системы (4). Очевидно, если система совместна, ее решение будет также и обобщенным решением.

Таким образом, рассматривается задача о минимуме нормы невязки

$$\|R\|^2 = (B - AX)^T(B - AX).$$

Решение указанной задачи приводит к уравнению

$$A^TAX = A^TB, \quad (24)$$

где значком T отмечена транспонированная матрица.

Система линейных уравнений (24) называется нормальной системой по отношению к системе (4). Она является следствием системы (4) и всегда совместна независимо от совместности системы (4).

Справедливо следующее утверждение.

Предложение [7]. *Нормальная система имеет единственное решение тогда и только тогда, когда система $AY = 0$ имеет только нулевое решение, то есть столбцы (строки) матрицы линейно независимы.*

Если решение нормальной системы неединственно, то выбирается решение с минимальной нормой.

Определение. **Нормальным псевдорешением системы (24)** называется вектор X_0 с минимальной нормой среди всех векторов X , дающих минимальную по норме невязку при подстановке в эту систему.

Имеет место

Теорема [7]. *Каждая система линейных уравнений имеет одно и только одно нормальное псевдорешение.*

МЕТОДЫ РЕГУЛЯРИЗАЦИИ

Предыдущие рассмотрения указывают необходимость выделения класса задач, для которых малые возмущения данных задач приводят к малому изменению решения. Класс таких задач называется корректным.

Приведем понятие корректности задачи по Адамару (Жак Адамар (1865–1963) — французский математик).

Задача называется корректной, если выполнены следующие три требования:

- 1) ее решение существует при любых входных данных задачи;
- 2) это решение единственно;
- 3) решение устойчиво по отношению к малым возмущениям входных данных.

В том случае, когда хотя бы одно из этих требований не выполнено, задача называется некорректной. Идеи Адамара были развиты И.Г. Петровским (Иван Георгиевич Петровский (1901–1973) — российский математик).

называется матрица C , элементы которой определяются по формуле

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, r.$$

Произведение AB определено только в том случае, когда число столбцов матрицы A равно числу строк матрицы B .

Пример 1.

$$A = \begin{pmatrix} 3 & 0 \\ 1 & 1 \\ 5 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 7 \\ 6 & 8 \end{pmatrix},$$

$$AB = \begin{pmatrix} 3 \cdot 4 + 0 \cdot 6 & 3 \cdot 7 + 0 \cdot 8 \\ 1 \cdot 4 + 1 \cdot 6 & 1 \cdot 7 + 1 \cdot 8 \\ 5 \cdot 4 + 2 \cdot 6 & 5 \cdot 7 + 2 \cdot 8 \end{pmatrix} = \begin{pmatrix} 12 & 21 \\ 10 & 15 \\ 32 & 51 \end{pmatrix}.$$

Заметим, что $AB \neq BA$ (если они имеют смысл), то есть произведение некоммутативно.

Транспонирование – замена строк в матрице столбцами.

Пример 2.

$$A = \begin{pmatrix} 2 & 5 \\ 3 & 6 \\ 4 & 8 \end{pmatrix}, \quad A^T = \begin{pmatrix} 2 & 3 & 4 \\ 5 & 6 & 8 \end{pmatrix}.$$

Умножение матрицы на вектор. Такое умножение является частным случаем умножения двух матриц, где вектор рассматривается как матрица размером $1 \times n$.

Пример 3.

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 1 \\ 5 & 2 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad A\mathbf{X} = \begin{pmatrix} 3x_1 + x_2 \\ x_1 + x_2 \\ 5x_1 + 2x_2 \end{pmatrix}.$$

Единичная матрица. Единичная матрица обозначается через E . По определению,

$$E = \{\delta_{ij}\}, \quad \delta_{ij} = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j, \end{cases}$$

где δ_{ij} называется символом Кронекера.

Пример 4.

$$E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Обратная матрица. Для любого действительного числа $a \neq 0$ существует число a^{-1} (называемое обратным для a), такое, что $a^{-1}a = aa^{-1} = 1$. Обратная матрица B для A определяется соотношением $AB = BA = E$. Такая матрица B всегда существует и единственна для квадратной матрицы A и такой, что $\det A \neq 0$. Обозначается символом A^{-1} , то есть $B = A^{-1}$.

Пример 5.

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

Существуют несколько способов вычисления обратной матрицы. Соответствующие алгоритмы можно найти в литературе [1–5].

ЛИТЕРАТУРА

1. Амосов А.А., Дубинский Ю.А., Копченова Н.В. Вычислительные методы для инженеров. М.: Высш. шк., 1994. 544 с.
2. Калиткин Н.Н. Численные методы. М.: Наука, 1978. 512 с.
3. Райс Дж. Матричные вычисления и математическое обеспечение. М.: Мир, 1984. 264 с.
4. Демидович Б.П., Марон И.А. Основы вычислительной математики. М.: Наука, 1970. 664 с.
5. Воробьева Г.Н., Данилова А.Н. Практикум по численным методам. М.: Высш. шк., 1979. 184 с.
6. Тихонов А.Н., Костомаров Д.П. Вводные лекции по прикладной математике. М.: Наука, 1984. 190 с.
7. Беклемишев Д.В. Дополнительные главы линейной алгебры. М.: Наука, 1983. 335 с.
8. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1986. 320 с.
9. Годунов С.К. Решение систем линейных уравнений. Новосибирск: Наука, 1980.

* * *

Василий Васильевич Дикусар, доктор физико-математических наук, профессор кафедры высшей математики МФТИ, ведущий научный сотрудник Вычислительного центра Российской академии наук. Область научных интересов: теория управления, численные методы, финансовая математика. Автор более 80 научных статей, двух изобретений и трех монографий.