

Лекции по Линейной Алгебре

Дима Трушин

2024 — 2025

Содержание

1	Системы линейных уравнений	2
1.1	Системы линейных уравнений и связанная с ними терминология	2
1.2	Матрицы связанные со СЛУ	2
1.3	Элементарные преобразования	3
1.4	Алгоритм Гаусса	4
1.5	Описание множества решений	5
1.6	Наблюдения из алгоритма	6
2	Матрицы	7
2.1	Определение матриц	7
2.2	Операции над матрицами	7
2.3	Специальные виды матриц	8
2.4	Свойства операций	8
2.5	Связь с системами линейных уравнений	10
2.6	Дефекты матричных операций	10
2.7	Деление	11
2.8	Матрицы элементарных преобразований	13
2.9	Невырожденные матрицы	14
2.10	Блочное умножение матриц	15
2.11	Блочные элементарные преобразования	16
2.12	Массовое решение систем	17
2.13	Нахождение обратной матрицы методом Гаусса	18
2.14	Классификация СЛУ	19
2.15	Полиномиальное исчисление от матриц	21
2.16	Спектр	22
2.17	Матричные нормы	26
2.18	Обзор применения матричных норм	28
3	Перестановки	30
3.1	Отображения множеств	30
3.2	Перестановки	30
3.3	Операция на перестановках	31
3.4	Переименование элементов	32
3.5	Циклы	32
3.6	Знак перестановки	34

1 Системы линейных уравнений

1.1 Системы линейных уравнений и связанная с ними терминология

Наша задача научиться решать Системы Линейных Уравнений (СЛУ), то есть находить все их решения или доказывать, что решений нет. Общий вид СЛУ и ее однородная версия (ОСЛУ):

$$\begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = b_1 \\ \dots \\ a_{m1}x_1 + \dots + a_{mn}x_n = b_m \end{cases} \quad \begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = 0 \\ \dots \\ a_{m1}x_1 + \dots + a_{mn}x_n = 0 \end{cases}$$

Коэффициенты Где живут коэффициенты a_{ij} и b_j ? Варианты:

- Вещественные числа \mathbb{R}
- Комплексные числа \mathbb{C}
- Рациональные числа \mathbb{Q}

Для решения СЛУ **НЕ** имеет значения откуда берутся коэффициенты, так как решения будут лежать там же. Главное, чтобы для коэффициентов были определены операции сложения, вычитания, умножения и деления.¹ Потому мы будем работать с числами из \mathbb{R} . А вот целые числа \mathbb{Z} уже не подходят, так как в них нельзя делить.

Решение Решением системы линейных уравнений называется набор чисел (c_1, \dots, c_n) , $c_i \in \mathbb{R}$ такой, что при подстановке c_i вместо x_i , все уравнения системы превращаются в верные равенства. Введем обозначение $\mathbb{R}^n = \mathbb{R} \times \dots \times \mathbb{R} = \{(c_1, \dots, c_n) \mid c_i \in \mathbb{R}\}$. То есть элемент \mathbb{R}^n – это набор из n вещественных чисел. Потому любое решение $c = (c_1, \dots, c_n)$ является элементом \mathbb{R}^n .

1.2 Матрицы связанные со СЛУ

Для каждой СЛУ введем следующие обозначения:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad (A|b) = \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{array} \right)$$

Названия:

- A – матрица системы
- b – вектор правой части
- $(A|b)$ – расширенная матрица системы
- x – вектор решений

Будем кратко записывать СЛУ и ее однородную версию так: $Ax = b$ и $Ax = 0$. Также для краткости будем обозначать системы буквами Σ .

При решении системы линейных уравнений приходится много раз переписывать кучу данных. Для того чтобы сократить эти записи, целесообразно сократить количество записываемой на бумаге информации. Расширенная матрица системы $(A|b)$ является необходимым минимумом такой информации. Потому сейчас к такой записи можно относиться как к удобному способу компактно записать систему.

Количество решений Случай одного уравнения и одной неизвестной $ax = b$, где $a, b \in \mathbb{R}$:

- При $a \neq 0$ – одно решение $x = b/a$.
- При $a = 0$, $b \neq 0$ – нет решений.
- При $a = 0$, $b = 0$ – любое число является решением, т.е. бесконечное число решений.

¹Позже мы введем понятие поля и допустимыми будут коэффициенты из произвольного поля.

Что значит решить систему Обычно принято говорить следующие слова: решить систему значит описать множество ее решений, то есть либо доказать, что система не имеет решений вовсе, либо описать все наборы, которые являются решениями. Если система не имеет решений, она называется *несовместной*, в противном случае – *совместной*. Однако, в этой фразе не понятно, что имеется в виду. Действительно, если я запишу систему

$$\begin{cases} x + 2y + 3z = 0 \\ 4x + 5y + 6z = 0 \\ 7x + 8y + 9z = 0 \end{cases}$$

То я могу описать множество решений словами: это все наборы (x, y, z) подходящие под систему. Это звучит как жульничество, тем не менее в этом описании есть содержательная мысль, давайте я ее озвучу явно. Приведенная выше система позволяет нам быстро решать такую задачу: проверить лежит ли данный набор среди решений. То есть мы эффективно умеем тестировать вопрос на принадлежность множеству решений. Однако, такое описание не позволяет решать другую важную задачу: генерировать все решения по одному разу. То есть, глядя на эту систему, вы вряд ли сходу предъявите набор отличный от $(0, 0, 0)$ в качестве решения. Более того, тут даже не понятно, есть ли какие-то другие наборы или нет. Потому мы хотим найти другое описание множества решений, которое бы позволяло быстро решать обе задачи: и задачу принадлежности и задачу генерации. Оказывается, что это тоже делается с помощью систем, только в очень специальном виде.

Эквивалентные системы Пусть даны две системы линейных уравнений с одинаковым числом неизвестных (но быть может разным числом уравнений) Σ_1 и Σ_2 . Будем говорить, что эти системы эквивалентны и писать $\Sigma_1 \sim \Sigma_2$, если множества решений этих систем совпадают. Если $E_i \subseteq \mathbb{R}^n$ – множество решений i -ой системы, то системы эквивалентны, если $E_1 = E_2$.

Вот полезный пример эквивалентных систем:

$$\begin{cases} x + y = 1 \\ x - y = 0 \end{cases} \sim \begin{cases} 2x = 1 \\ 2y = 1 \end{cases}$$

Как решать систему Пусть нам надо решить систему Σ . Идея состоит в том, чтобы постепенно менять ее на эквивалентную до тех пор, пока она не упростится до такого состояния, что все ее решения легко описать.

$$\Sigma = \Sigma_1 \mapsto \Sigma_2 \mapsto \dots \mapsto \Sigma_n \leftarrow \text{легко решается}$$

Теперь надо объяснить две вещи: (1) какого сорта преобразования над системами мы будем делать и (2) к какому замечательному виду мы их приводим и как в нем выглядят все решения. Ответам на эти два вопроса и будет посвящена оставшаяся часть лекции.

1.3 Элементарные преобразования

Мы разделим все преобразования на три типа²:

$$\begin{aligned} \text{I тип: } & \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ a_{i1} & \dots & a_{in} & b_i \\ a_{j1} & \dots & a_{jn} & b_j \\ a_{m1} & \dots & a_{mn} & b_m \end{array} \right) \mapsto \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ a_{i1} & \dots & a_{in} & b_i \\ a_{j1} + \lambda a_{i1} & \dots & a_{jn} + \lambda a_{in} & b_j + \lambda b_i \\ a_{m1} & \dots & a_{mn} & b_m \end{array} \right) \quad i \neq j \\ \\ \text{II тип: } & \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ a_{i1} & \dots & a_{in} & b_i \\ a_{j1} & \dots & a_{jn} & b_j \\ a_{m1} & \dots & a_{mn} & b_m \end{array} \right) \mapsto \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ a_{j1} & \dots & a_{jn} & b_j \\ a_{i1} & \dots & a_{in} & b_i \\ a_{m1} & \dots & a_{mn} & b_m \end{array} \right) \\ \\ \text{III тип: } & \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ a_{i1} & \dots & a_{in} & b_i \\ a_{m1} & \dots & a_{mn} & b_m \end{array} \right) \mapsto \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \lambda a_{i1} & \dots & \lambda a_{in} & \lambda b_i \\ a_{m1} & \dots & a_{mn} & b_m \end{array} \right) \quad \lambda \neq 0 \end{aligned}$$

Поясним словами, что делают преобразования:

1. Прибавляем к j -ой строке i -ю, умноженную на константу $\lambda \in \mathbb{R}$.

²Стоит отметить, что нумерация типов преобразования не является общепринятой и отличается от учебника к учебнику.

2. Меняем местами i -ю и j -ю строки.
3. Умножаем i -ю строку на ненулевую константу $\lambda \neq 0, \lambda \in \mathbb{R}$.

1.4 Алгоритм Гаусса

Этот метод заключается в приведении СЛУ к некоторому «ступенчатому виду», где множество решений очевидно.³ Разберем типичный ход алгоритма Гаусса на примере 3 уравнений и 4 неизвестных.⁴

Прямой ход алгоритма Гаусса Идея прямого хода алгоритма в следующем. Мы смотрим на левый верхний элемент в матрице и пытаемся с помощью него обнулить все элементы под ним. Как только обнулили, забываем про первую строку и столбец и повторяем процедуру с оставшейся подматрицей.

$$\begin{aligned} & \left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & b_1 \\ a_{21} & a_{22} & a_{23} & a_{24} & b_2 \\ a_{31} & a_{32} & a_{33} & a_{34} & b_3 \end{array} \right) \quad \begin{array}{l} \text{2-я строка} \\ \text{3-я строка} \end{array} \begin{array}{l} - \frac{a_{21}}{a_{11}} \cdot \text{1-я строка} \\ - \frac{a_{31}}{a_{11}} \cdot \text{1-я строка} \end{array} \\ & \left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & a_{22} & a_{23} & a_{24} & b_2 \\ a_{31} & a_{32} & a_{33} & a_{34} & b_3 \end{array} \right) \quad \begin{array}{l} \text{3-я строка} \\ \text{3-я строка} \end{array} \begin{array}{l} - \frac{a_{31}}{a_{11}} \cdot \text{1-я строка} \\ - \frac{a_{32}}{a_{22}} \cdot \text{2-я строка} \end{array} \\ & \left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & a_{22} & a_{23} & a_{24} & b_2 \\ 0 & a_{32} & a_{33} & a_{34} & b_3 \end{array} \right) \quad \begin{array}{l} \text{3-я строка} \\ \text{3-я строка} \end{array} \begin{array}{l} - \frac{a_{31}}{a_{11}} \cdot \text{1-я строка} \\ - \frac{a_{32}}{a_{22}} \cdot \text{2-я строка} \end{array} \\ & \left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & a_{22} & a_{23} & a_{24} & b_2 \\ 0 & 0 & a_{33} & a_{34} & b_3 \end{array} \right) \end{aligned}$$

В рассуждениях выше, мы пользовались тем, что угловые элементы a_{11} и a_{22} не нули. Но вообще говоря так могло не получиться. Например на третьем шаге могла быть одна из следующих ситуаций (здесь мы смотрим на элемент a_{22}):

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & 0 & a_{23} & a_{24} & b_2 \\ 0 & a_{32} & a_{33} & a_{34} & b_3 \end{array} \right) \quad \text{или} \quad \left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & 0 & a_{23} & a_{24} & b_2 \\ 0 & 0 & a_{33} & a_{34} & b_3 \end{array} \right)$$

В первом случае $a_{22} = 0$, но при этом есть какой-то ненулевой элемент под ним. В такой ситуации надо переставить вторую строчку с другой строкой, где второй элемент не ноль. В примере надо переставить вторую и третью строчку местами. Во втором случае $a_{22} = 0$ и все элементы под ним. В такой ситуации надо пропустить второй столбец и перейти к третьему и мы смотрим на элемент a_{23} . После чего повторяем алгоритм Гаусса с подматрицей.

В результате прямого хода алгоритма из-за обнуления коэффициентов могут возникнуть следующие случаи⁵

$$\left(\begin{array}{cccc|c} \underline{a_{11}} & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & \underline{a_{22}} & a_{23} & a_{24} & b_2 \\ 0 & 0 & 0 & \underline{a_{34}} & b_3 \end{array} \right) \quad \left(\begin{array}{cccc|c} \underline{a_{11}} & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & 0 & \underline{a_{23}} & a_{24} & b_2 \\ 0 & 0 & 0 & \underline{a_{34}} & b_3 \end{array} \right) \quad \left(\begin{array}{cccc|c} \underline{a_{11}} & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & \underline{a_{22}} & a_{23} & a_{24} & b_2 \\ 0 & 0 & 0 & 0 & \underline{b_3} \end{array} \right) \quad \left(\begin{array}{cccc|c} \underline{a_{11}} & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & \underline{a_{22}} & a_{23} & a_{24} & b_2 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

Главные и неглавные переменные Подчеркнутые элементы считаются не равными нулю. Это первые ненулевые коэффициенты в строке, они называются лидерами. В ступенчатом виде все переменные делятся на главные и свободные. Соответствующие лидерам переменные называются главными. Остальные переменные называются свободными.

³Данный метод является самым быстрым возможным как для написания программ, так и для ручного вычисления. При вычислениях руками, однако, полезно местами пользоваться «локальными оптимизациями», то есть если вы видите, что какая-то хитрая комбинация строк сильно упростит вид системы, то сделайте ее.

⁴При переходе от одной матрицы к другой я новым коэффициентам даю старые имена, чтобы не захламлять текст новыми обозначениями.

⁵Это не полный список всех случаев.

Обратный ход алгоритма Гаусса Задача обратного хода алгоритма в том, чтобы сделать все лидирующие коэффициенты единицами, а все коэффициенты над ними обнулить. Обратный ход осуществляется снизу вверх. Разберем типичный обратный ход алгоритма Гаусса. Подчеркнутые элементы считаются не равными нулю.

$$\begin{pmatrix} \underline{a_{11}} & a_{12} & a_{13} & a_{14} & | & b_1 \\ 0 & \underline{a_{22}} & a_{23} & a_{24} & | & b_2 \\ 0 & 0 & \underline{a_{33}} & a_{34} & | & b_3 \end{pmatrix} \quad \text{разделить } i\text{-ю строку на } a_{ii}$$

$$\begin{pmatrix} 1 & a_{12} & a_{13} & a_{14} & | & b_1 \\ 0 & 1 & a_{23} & a_{24} & | & b_2 \\ 0 & 0 & 1 & a_{34} & | & b_3 \end{pmatrix} \quad \text{2-я строка} - a_{23} \cdot \text{3-я строка}$$

$$\begin{pmatrix} 1 & a_{12} & a_{13} & a_{14} & | & b_1 \\ 0 & 1 & 0 & a_{24} & | & b_2 \\ 0 & 0 & 1 & a_{34} & | & b_3 \end{pmatrix} \quad \text{1-я строка} - a_{13} \cdot \text{3-я строка}$$

$$\begin{pmatrix} 1 & a_{12} & 0 & a_{14} & | & b_1 \\ 0 & 1 & 0 & a_{24} & | & b_2 \\ 0 & 0 & 1 & a_{34} & | & b_3 \end{pmatrix} \quad \text{1-я строка} - a_{12} \cdot \text{2-я строка}$$

$$\begin{pmatrix} 1 & 0 & 0 & a_{14} & | & b_1 \\ 0 & 1 & 0 & a_{24} & | & b_2 \\ 0 & 0 & 1 & a_{34} & | & b_3 \end{pmatrix}$$

В специальных случаях приведенных выше, получим

$$\begin{pmatrix} 1 & 0 & a_{13} & 0 & | & b_1 \\ 0 & 1 & a_{23} & 0 & | & b_2 \\ 0 & 0 & 0 & 1 & | & b_3 \end{pmatrix} \quad \begin{pmatrix} 1 & a_{12} & 0 & 0 & | & b_1 \\ 0 & 0 & 1 & 0 & | & b_2 \\ 0 & 0 & 0 & 1 & | & b_3 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & a_{13} & a_{14} & | & 0 \\ 0 & 1 & a_{23} & a_{24} & | & 0 \\ 0 & 0 & 0 & 0 & | & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & a_{13} & a_{14} & | & b_1 \\ 0 & 1 & a_{23} & a_{24} & | & b_2 \\ 0 & 0 & 0 & 0 & | & 0 \end{pmatrix}$$

Полученный в результате обратного хода вид расширенной матрицы называется улучшенным ступенчатым видом, т.е. это ступенчатый вид, где все лидирующие коэффициенты – единицы, и все коэффициенты над ними равны нулю.

1.5 Описание множества решений

Удобный формализм Пока мы подробно не говорили о матрицах, введем некие удобные обозначения, которые упростят запись решений СЛУ.

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n \text{ и } b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{R}^n. \text{ Тогда } a + b = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix} \in \mathbb{R}^n \text{ и } \lambda a = \begin{pmatrix} \lambda a_1 \\ \vdots \\ \lambda a_n \end{pmatrix} \in \mathbb{R}^n \text{ для любого } \lambda \in \mathbb{R}.$$

Получение решений В системе ниже, выберем переменную x_4 как параметр

$$\begin{pmatrix} 1 & 0 & 0 & a_{14} & | & b_1 \\ 0 & 1 & 0 & a_{24} & | & b_2 \\ 0 & 0 & 1 & a_{34} & | & b_3 \end{pmatrix}$$

Тогда решения имеют вид⁶

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} - x_4 \begin{pmatrix} a_{14} \\ a_{24} \\ a_{34} \end{pmatrix}$$

⁶Операция умножения матрицы на число покомпонентная (умножаем каждый элемент на число). Сумма и разность двух матриц покомпонентная (складываем или вычитаем числа на одних и тех же позициях).

Специальные случаи:

$$\begin{array}{l}
 \left(\begin{array}{cccc|c} 1 & 0 & a_{13} & 0 & b_1 \\ 0 & 1 & a_{23} & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \end{array} \right) \text{ Решения: } \begin{pmatrix} x_1 \\ x_2 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} - x_3 \begin{pmatrix} a_{13} \\ a_{23} \\ 0 \end{pmatrix} \\
 \left(\begin{array}{cccc|c} 1 & a_{12} & 0 & 0 & b_1 \\ 0 & 0 & 1 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \end{array} \right) \text{ Решения: } \begin{pmatrix} x_1 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} - x_2 \begin{pmatrix} a_{12} \\ 0 \\ 0 \end{pmatrix} \\
 \left(\begin{array}{cccc|c} 1 & 0 & a_{13} & a_{14} & 0 \\ 0 & 1 & a_{23} & a_{24} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right) \text{ Решения: Нет решений, т.к. последнее уравнение } 0 = 1 \\
 \left(\begin{array}{cccc|c} 1 & 0 & a_{13} & a_{14} & b_1 \\ 0 & 1 & a_{23} & a_{24} & b_2 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right) \text{ Решения: } \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} - x_3 \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} - x_4 \begin{pmatrix} a_{14} \\ a_{24} \end{pmatrix}
 \end{array}$$

1.6 Наблюдения из алгоритма

Количество решений в ступенчатом виде Если во время прямого хода алгоритма Гаусса в расширенной матрице системы вам встретилась строка вида $(0 \dots 0 \mid b)$, где b – произвольное ненулевое число, то данная система решений не имеет. В этом случае нет необходимости переходить к обратному ходу. Если же таких строк не встретилось, то система обязательно имеет решения. При этом, если есть свободные переменные, то решений бесконечное число, а если их нет, то решение единственное.

Широкие системы Пусть вам дана однородная система из m уравнений с n неизвестными и $m < n$, тогда утверждается, что она обязательно имеет хотя бы одно ненулевое решение, то есть такое решение, где хотя бы одна переменная отличная от нуля. Действительно, так как у однородной системы всегда есть решение, то нам достаточно доказать, что у системы в улучшенном ступенчатом виде есть свободные переменные. Если это так, то придавая свободным переменным ненулевые значения, мы получим ненулевое решение. Чтобы понять, что есть свободные переменные, давайте оценим сверху количество главных. Главных переменных не больше чем строк, а строк строго меньше, чем всех переменных, вот и все.

Технические рекомендации Работая с целочисленными матрицами, старайтесь во время прямого хода алгоритма Гаусса не выходить за рамки целых чисел.

- Используйте элементарные преобразования I типа только с целым параметром.
- Полезно не злоупотреблять умножением на ненулевое целое, умножайте только на ± 1 . Иначе придется работать с большими числами.

На этапе обратного хода алгоритма Гаусса избавиться от деления уже не возможно.

2 Матрицы

2.1 Определение матриц

Матрица – это прямоугольная таблица чисел

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}, \text{ где } a_{ij} \in \mathbb{R}$$

Множество всех матриц с m строками и n столбцами обозначается $M_{m \times n}(\mathbb{R})$. Множество квадратных матриц размера n будем обозначать $M_n(\mathbb{R})$. Матрицы с одним столбцом или одной строкой называются векторами (вектор-столбцами и вектор-строками соответственно). Множество всех векторов с n координатами обозначается через \mathbb{R}^n . Мы по умолчанию считаем, что наши вектора – вектор-столбцы.⁷

2.2 Операции над матрицами

Сложение Пусть $A, B \in M_{m \times n}(\mathbb{R})$. Тогда сумма $A + B$ определяется покомпонентно, т.е. $C = A + B$, то $c_{ij} = a_{ij} + b_{ij}$ или

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{pmatrix}$$

Складывать можно только матрицы одинакового размера.⁸

Умножение на скаляр Если $\lambda \in \mathbb{R}$ и $A \in M_{m \times n}(\mathbb{R})$, то λA определяется так: $\lambda A = C$, где $c_{ij} = \lambda a_{ij}$ или

$$\lambda \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \dots & \lambda a_{1n} \\ \vdots & \ddots & \vdots \\ \lambda a_{m1} & \dots & \lambda a_{mn} \end{pmatrix}$$

Умножение матриц Пусть $A \in M_{m \times n}(\mathbb{R})$ и $B \in M_{n \times k}(\mathbb{R})$, то произведение $AB \in M_{m \times k}(\mathbb{R})$ определяется так: $AB = C$, где $c_{ij} = \sum_{t=1}^n a_{it}b_{tj}$ или

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nk} \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^n a_{1t}b_{t1} & \dots & \sum_{t=1}^n a_{1t}b_{tk} \\ \vdots & \ddots & \vdots \\ \sum_{t=1}^n a_{mt}b_{t1} & \dots & \sum_{t=1}^n a_{mt}b_{tk} \end{pmatrix}$$

Умножение матриц полезно как-то визуализировать у себя в голове. Предположим мы хотим посчитать коэффициент c_{ij} . Тогда надо из матрицы A взять i -ю строку (она имеет длину n), а из матрицы B взять j -ый столбец (он тоже имеет длину n). Тогда их надо скалярно перемножить и результат подставить в c_{ij} , как показано ниже на картинке.

$$\begin{pmatrix} a_{11} & \dots & \dots & a_{1n} \\ \vdots & \ddots & & \vdots \\ \boxed{a_{i1} \quad \dots \quad a_{in}} \\ \vdots & & \ddots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & \boxed{b_{1j}} & \dots & b_{1k} \\ \vdots & \ddots & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & \ddots & \vdots \\ b_{n1} & \dots & \boxed{b_{nj}} & \dots & b_{nk} \end{pmatrix} = \begin{pmatrix} * & \dots & & \dots & * \\ \vdots & \ddots & & & \vdots \\ \vdots & & \boxed{\sum_{t=1}^n a_{it}b_{tj}} & & \vdots \\ \vdots & & \vdots & \ddots & \vdots \\ * & \dots & & \dots & * \end{pmatrix}$$

⁷Важно, directX и OpenGL используют вектор-строки! Потому часть инженерной литературы на английском, связанной с трехмерной графикой, оперирует со строками. Это важно учитывать, так как нужно вносить поправки в соответствующие формулы.

⁸Можно по аналогии определить и вычитание матриц, но в этом нет необходимости. Например, потому что вычитание можно определить как $A + (-1)B$, где $(-1)B$ – умножение на скаляр. Либо можно определить аксиоматически, как это сделано ниже в следующем разделе.

При этом умножение строки на столбец можно себе представлять так: мы приставляем строку к столбцу, потом поэлементно перемножаем их, а затем складываем все произведения вместе.

$$\begin{array}{c}
 \boxed{a_{i1} \quad \dots \quad a_{in}} \cdot \boxed{\begin{array}{c} b_{1j} \\ \vdots \\ b_{nj} \end{array}} \mapsto \begin{array}{c} \boxed{a_{i1} \quad \dots \quad a_{in}} \\ \boxed{b_{1j} \quad \dots \quad b_{nj}} \end{array} \mapsto \\
 \mapsto \boxed{\begin{array}{c} a_{i1} \\ \vdots \\ b_{1j} \end{array}} + \dots + \boxed{\begin{array}{c} a_{in} \\ \vdots \\ b_{nj} \end{array}} = a_{i1}b_{1j} + \dots + a_{in}b_{nj}
 \end{array}$$

Транспонирование Пусть A – матрица вида

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Определим транспонированную матрицу $A^t = (a'_{ij})$ так: $a'_{ij} = a_{ji}$. Наглядно, транспонированная матрица для приведенных выше

$$\begin{pmatrix} a_{11} & \dots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{mn} \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix} \quad \text{или} \quad (x_1 \quad x_2 \quad x_3)$$

След матрицы Пусть $A \in M_n(\mathbb{R})$, тогда определим след матрицы A , как сумму ее диагональных элементов: $\text{tr } A = \sum_{i=1}^n a_{ii}$. Давайте отметим следующие свойства следа:

1. Для любых матриц $A, B \in M_n(\mathbb{R})$ верно $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$.
2. Для любой матрицы $A \in M_n(\mathbb{R})$ и $\lambda \in \mathbb{R}$ выполнено $\text{tr}(\lambda A) = \lambda \text{tr}(A)$.
3. Для любых матриц $A \in M_{mn}(\mathbb{R})$ и $B \in M_{nm}(\mathbb{R})$ выполнено $\text{tr}(AB) = \text{tr}(BA)$.

Все эти свойства проверяются непосредственным вычислением по определению.

2.3 Специальные виды матриц

Ниже мы перечислим названия некоторых специальных классов матриц:

- $A = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix}$ – диагональная матрица. Все ненулевые элементы стоят на главной диагонали, то есть в позиции, где номер строки равен номеру столбца.
- $A = \begin{pmatrix} \lambda & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda \end{pmatrix}$ – скалярная матрица. Диагональная матрица с одинаковыми элементами на диагонали.

2.4 Свойства операций

Все операции на матрицах обладают «естественными свойствами» и согласованы друг с другом. Вот перечень базовых свойств операций над матрицами:⁹

⁹Все эти свойства объединяет то, что они являются аксиомами в различных определениях для алгебраических структур. Позже мы столкнемся с такими структурами.

1. **Ассоциативность сложения** $(A + B) + C = A + (B + C)$ для любых $A, B, C \in M_{m \times n}(\mathbb{R})$
2. **Существование нейтрального элемента для сложения** Существует единственная матрица 0 обладающая следующим свойством $A + 0 = 0 + A = A$ для всех $A \in M_{m \times n}(\mathbb{R})$. Такая матрица целиком заполнена нулями.
3. **Коммутативность сложения** $A + B = B + A$ для любых $A, B \in M_{m \times n}(\mathbb{R})$.
4. **Наличие обратного по сложению** Для любой матрицы $A \in M_{m \times n}(\mathbb{R})$ существует матрица $-A$ такая, что $A + (-A) = (-A) + A = 0$. Такая матрица единственная и состоит из элементов $-a_{ij}$.
5. **Ассоциативность умножения** Для любых матриц $A \in M_{m \times n}(\mathbb{R})$, $B \in M_{n \times k}(\mathbb{R})$ и $C \in M_{k \times t}(\mathbb{R})$ верно $(AB)C = A(BC)$.
6. **Существование нейтрального элемента для умножения** Для каждого k существует единственная матрица $E \in M_k(\mathbb{R})$ такая, что для любой $A \in M_{m \times n}(\mathbb{R})$ верно $EA = AE = A$. У такой матрицы $E_{ii} = 1$, а $E_{ij} = 0$. Когда нет путаницы, матрицу E обозначают через 1 .
7. **Дистрибутивность умножения относительно сложения** Для любых матриц $A, B \in M_{m \times n}(\mathbb{R})$ и $C \in M_{n \times k}(\mathbb{R})$ верно $(A + B)C = AC + BC$. Аналогично, для любых $A \in M_{m \times n}(\mathbb{R})$ и $B, C \in M_{n \times k}(\mathbb{R})$ верно $A(B + C) = AB + AC$.
8. **Умножение на числа ассоциативно** Для любых $\lambda, \mu \in \mathbb{R}$ и любой матрицы $A \in M_{m \times n}(\mathbb{R})$ верно $\lambda(\mu A) = (\lambda\mu)A$. Аналогично для любого $\lambda \in \mathbb{R}$ и любых $A \in M_{m \times n}(\mathbb{R})$ и $B \in M_{n \times k}(\mathbb{R})$ верно $\lambda(AB) = (\lambda A)B$.
9. **Умножение на числа дистрибутивно относительно сложения матриц и сложения чисел** Для любых $\lambda, \mu \in \mathbb{R}$ и $A \in M_{m \times n}(\mathbb{R})$ верно $(\lambda + \mu)A = \lambda A + \mu A$. Аналогично, для любого $\lambda \in \mathbb{R}$ и $A, B \in M_{m \times n}(\mathbb{R})$ верно $\lambda(A + B) = \lambda A + \lambda B$.
10. **Умножение на скаляр нетривиально** Если $1 \in \mathbb{R}$, то для любой матрицы $A \in M_{m \times n}(\mathbb{R})$ верно $1A = A$.
11. **Умножение на скаляр согласовано с умножением матриц** Для любого $\lambda \in \mathbb{R}$ и любых $A \in M_{m \times n}(\mathbb{R})$ и $B \in M_{n \times k}(\mathbb{R})$ верно $\lambda(AB) = (\lambda A)B = A(\lambda B)$.
12. **Транспонирование согласовано с суммой** Для любых матриц $A, B \in M_{m \times n}(\mathbb{R})$ верно $(A + B)^t = A^t + B^t$.
13. **Транспонирование согласовано с умножением на скаляр** Для любой матрицы $A \in M_{m \times n}(\mathbb{R})$ и любого $\lambda \in \mathbb{R}$ верно $(\lambda A)^t = \lambda A^t$.
14. **Транспонирование согласовано с умножением** Для любых матриц $A, B \in M_{m \times n}(\mathbb{R})$ верно $(AB)^t = B^t A^t$.

К этим свойствам надо относиться так. Доказывая что-то про матрицы, можно лезть внутрь определений операций над ними, а можно пользоваться свойствами операций. Так вот, список выше – это минимальный набор свойств операций, из которых можно вытащить базовую информацию про эти операции и при этом не лезть внутрь определений.

Нулевые строки и столбцы Пусть в матрице $A \in M_{m \times k}(\mathbb{R})$ i -я строка полностью состоит из нулей и нам дана матрица $B \in M_{k \times n}(\mathbb{R})$. Тогда в произведении AB i -я строка тоже будет нулевой. Изобразим это ниже графически

$$AB = \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ 0 & 0 & \dots & 0 \\ * & * & \dots & * \end{pmatrix} \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \end{pmatrix} = \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ 0 & 0 & \dots & 0 \\ * & * & \dots & * \end{pmatrix}$$

Действительно, i -я строка произведения зависит от i -ой строки левого множителя (матрицы A) и всех столбцов B . Но умножая нулевую строку A на что угодно, получим нули в i -ой строке результата. Аналогичное

утверждение верно для столбцов в матрице B , а именно. Пусть в матрице $B \in M_{k \times n}(\mathbb{R})$ i -ый столбец полностью состоит из нулей и нам дана матрица $A \in M_{m \times k}(\mathbb{R})$. Тогда в произведении AB i -ый столбец тоже будет нулевой.

$$AB = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ \vdots & \vdots & \vdots & \vdots \\ * & * & * & * \end{pmatrix} \begin{pmatrix} * & * & 0 & * \\ * & * & 0 & * \\ \vdots & \vdots & \vdots & \vdots \\ * & * & 0 & * \end{pmatrix} = \begin{pmatrix} * & * & 0 & * \\ * & * & 0 & * \\ \vdots & \vdots & \vdots & \vdots \\ * & * & 0 & * \end{pmatrix}$$

2.5 Связь с системами линейных уравнений

Пусть нам дана система линейных уравнений соответствующая матрицам

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad (A|b) = \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{array} \right)$$

Мы кратко записывали такую систему $Ax = b$, а ее однородную версию через $Ax = 0$. Но теперь, когда мы знаем умножение матриц, видно, что Ax – это произведение матрицы A , на вектор неизвестных x .

Главный бонус от матриц и операций над ними заключается вот в чем. У нас исходно была большая и неуклюжая система линейных уравнений, в которой участвовали очень знакомые и простые для использования числа. Теперь же мы заменили много линейных уравнений с кучей неизвестных на одно линейное матричное уравнение $Ax = b$. Однако, теперь вместо приятных в использовании чисел у нас встретились более сложные объекты – матрицы. Потому к матрицам надо относиться как к более продвинутой версии чисел.

Линейная структура Пусть у нас дана система $Ax = b$ как выше. Тогда $y \in \mathbb{R}^n$ является решением этой системы, если выполнено матричное равенство $Ay = b$. Аналогично и для однородной системы. Теперь заметим следующее:

1. Если $y_1, y_2 \in \mathbb{R}^n$ – решения системы $Ax = 0$, то $y_1 + y_2$ тоже является решением системы $Ax = 0$. Действительно, надо показать, что $A(y_1 + y_2) = 0$. Но $A(y_1 + y_2) = Ay_1 + Ay_2 = 0 + 0 = 0$.
2. Если $y \in \mathbb{R}^n$ – решение системы $Ax = 0$ и $\lambda \in \mathbb{R}$, то λy – тоже решение $Ax = 0$. Действительно, $A(\lambda y) = \lambda Ay = 0$.

Теперь сравним решения систем $Ax = b$ и $Ax = 0$. Прежде всего заметим, что однородная система всегда имеет решение $x = 0$. И вообще говоря, может так оказаться, что $Ax = b$ не имеет решений. Например, $(A|b) = (0|1)$. Однако, если $Ax = b$ совместна, то обе системы имеют «одинаковое число» решений.

Утверждение. Пусть система $Ax = b$ имеет хотя бы одно решение $z \in \mathbb{R}^n$ и пусть $E_b \subseteq \mathbb{R}^n$ – множество решений $Ax = b$ и $E_0 \subseteq \mathbb{R}^n$ – множество решений $Ax = 0$. Тогда $E_b = z + E_0 = \{z + y \mid y \in E_0\}$.

Доказательство. Для доказательства $z + E_0 \subseteq E_b$ надо заметить, что если $y \in E_0$, то $z + y \in E_b$. Для обратного включения проверяется, что если $z' \in E_b$, то $z' - z \in E_0$. \square

2.6 Дефекты матричных операций

Матрицы как новые числа Рассмотрим множество квадратных матриц с введенными выше операциями: $(M_n(\mathbb{R}), +, -, \cdot, {}^t)$. Про это множество стоит думать как про новый вид чисел со своими операциями. Принципиальное отличие – нельзя делить на любую ненулевую матрицу, как это можно было делать с числами. Однако, это не единственное отличие.

Аномалии матричных операций Матричные операции обладают несколькими аномалиями по сравнению со свойствами операций над обычными числами.

1. Существование вычитания следует из «хорошести» операции сложения. Она позволяет определить вычитание без проблем. Однако, операция умножения уже хуже, чем на обычных числах, потому не получится просто так определить на матрицах операцию деления. Про деление и как его можно было бы обобщать мы поговорим ниже.

2. Давайте обсудим порядок матриц в произведении. Тут может быть несколько проблем. Может так оказаться, что матрицы можно перемножить в одном порядке, но нельзя в другом. Например, $A \in M_{m \times n}(\mathbb{R})$, $B \in M_{n \times k}(\mathbb{R})$ и при этом $m \neq k$. В этом случае $AB \in M_{m \times k}(\mathbb{R})$, однако, произведение BA просто не определено из-за отсутствия условия согласованности матриц. Если мы рассмотрим матрицы, которые можно перемножить в обоих порядках, то есть $A \in M_{m \times n}(\mathbb{R})$, $B \in M_{n \times m}(\mathbb{R})$, то результаты будут $AB \in M_m(\mathbb{R})$ и $BA \in M_n(\mathbb{R})$. А значит при разных m и n это будут матрицы разного размера и не могут быть одинаковыми. Таким образом, чтобы AB могло совпасть с BA нам надо, чтобы матрицы A и B были квадратными. Но даже в этом случае умножение матриц НЕ коммутативно. Действительно, вот простой пример в малой размерности

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{но} \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

3. В матрицах есть «делители нуля», т.е. существуют две ненулевые матрицы A и B такие, что $AB = 0$.¹⁰
Пример:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = 0$$

4. В матрицах есть «нильпотенты», то есть можно найти такую ненулевую матрицу A , что $A^n = 0$. Пример,

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = 0$$

Замечание Пример с делителями нуля выше является частным случаем некоторого более общего явления. Если есть две диагональные матрицы, то они складываются и умножаются покомпонентно, то есть

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} + \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{pmatrix} = \begin{pmatrix} \lambda_1 + \mu_1 & & \\ & \ddots & \\ & & \lambda_n + \mu_n \end{pmatrix}$$

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \cdot \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{pmatrix} = \begin{pmatrix} \lambda_1 \mu_1 & & \\ & \ddots & \\ & & \lambda_n \mu_n \end{pmatrix}$$

Чтобы получить правильную интуицию про такие матрицы, про них надо думать как про функции на конечном множестве. То есть матрица с диагональю из чисел $\lambda_1, \dots, \lambda_n$ это функция $f: \{1, \dots, n\} \rightarrow \mathbb{R}$ такая, что $f(k) = \lambda_k$. Тогда операции сложения и умножения функций превращаются в операции сложения и умножения диагональных матриц. И пример с делителями нуля выше, это пример двух индикаторов непересекающихся множеств. То есть если мы рассмотрим множество $\{1, 2\}$ и две функции

$$\chi_1(x) = \begin{cases} 1, & x = 1 \\ 0, & x = 2 \end{cases} \quad \text{и} \quad \chi_2(x) = \begin{cases} 0, & x = 1 \\ 1, & x = 2 \end{cases}$$

При этом для индикаторов подмножеств $A, B \subseteq X$ имеем $\chi_A \cdot \chi_B = \chi_{A \cap B}$. И так как χ_1 – индикатор $\{1\}$, а χ_2 – индикатор $\{2\}$, то $\chi_1 \cdot \chi_2 = 0$. Потому если вы можете пронаблюдать какой-то эффект на функциях, то скорее всего мы можете пронаблюдать его уже на диагональных матрицах. Это полезно иметь в виду.

2.7 Деление

Деление и некоммутативность Теперь мы хотим научиться делить матрицы друг на друга. Прежде чем это делать, обратим внимание, что умножение в матрицах у нас не коммутативно. Это значит умножать можно на матрицу слева или справа. А значит и делить надо уметь и слева и справа, то есть вместо одной операции деления, надо иметь целых две операции деления. Кроме того, надо будет установить, как эти операции связаны друг с другом, например, разделить сначала справа, а потом слева или наоборот сначала слева, а потом справа дают один и тот же результат или нет? Возникает куча сложностей работы с двумя операциями деления. Таким путем пойти можно, и есть тексты, которые делают именно так, но это не очень удобно. Вместо этого предлагается пойти другим путем, но для этого надо переосмыслить понятие деления.

¹⁰На самом деле, это очень «хорошая» аномалия, так как она связана с тем, что ОСЛУ имеют решения. Действительно, вопрос решения ОСЛУ $Ax = 0$ – это в точности вопрос существования правых делителей нуля для A в множестве \mathbb{R}^n .

Что значит деление в числах? Предположим, что у нас есть два числа $a, b \in \mathbb{R}$. Тогда деление $a/b = a \cdot b^{-1}$ – это просто умножение на обратный элемент, а обратный элемент b^{-1} определяется свойством $bb^{-1} = 1$. Данное наблюдение дает ключ к распространению деления и обращения на случай матриц. А именно, вместо деления, мы будем рассматривать обратные матрицы и умножение на них. И вот сразу преимущество такого подхода. Нам не надо вводить два вида деления: левое и правое. Вместо этого, намного проще изучать обратные матрицы и умножать на них слева и справа с помощью обычного умножения.

Односторонняя обратимость Пусть $A \in M_{m \times n}(\mathbb{R})$, будем говорить, что $B \in M_{n \times m}(\mathbb{R})$ является левым обратным к A , если $BA = E \in M_n(\mathbb{R})$. Аналогично, $B \in M_{n \times m}(\mathbb{R})$ – правый обратный к A , если $AB = E \in M_m(\mathbb{R})$. Надо иметь в виду, что вообще говоря левые и правые обратные между собой никак не связаны и их может быть много. Например, пусть $A = (1, 0) \in M_{1 \times 2}(\mathbb{R})$. Тогда у такой матрицы нет левого обратного, а любая матрица вида $(1, a)^t$ является правым обратным. Если для матрицы A существует левый обратный, то она называется обратимой слева. Аналогично, при существовании правого обратного – обратимой справа.

Обратимые матрицы Матрица $A \in M_{m \times n}(\mathbb{R})$ называется обратимой, если к ней существует левый и правый обратный.¹¹

Утверждение 1. Пусть матрица $A \in M_{m \times n}(\mathbb{R})$ обратима. Тогда

1. Левый обратный и правый обратный единственны и совпадают друг с другом.
2. Матрица A обязательно квадратная, то есть $m = n$.

Доказательство. (1) Пусть $L \in M_{n \times m}(\mathbb{R})$ – произвольный левый обратный к A , а $R \in M_{n \times m}(\mathbb{R})$ – произвольный правый обратный. Тогда рассмотрим выражение LAR , расставляя по разному скобки имеем:

$$R = ER = (LA)R = L(AR) = LE = L$$

Теперь, если L и L' – два разных левых обратных. Зафиксируем произвольный правый обратный R . Из выше сказанного следует, что $L = R$ и $L' = R$. Значит все левые обратные равны между собой. Аналогично для правых.

(2) Теперь покажем, что двусторонний обратный есть только у квадратных матриц. Пусть $B \in M_{n \times m}(\mathbb{R})$ – двусторонний обратный к A , то есть $AB = E_m \in M_m(\mathbb{R})$ и $BA = E_n \in M_n(\mathbb{R})$.¹² Тогда по свойствам следа получим

$$m = \text{tr}(E_m) = \text{tr}(AB) = \text{tr}(BA) = \text{tr}(E_n) = n$$

□

Значит, если матрица A обратима, то она как минимум квадратная и существует единственная матрица B , удовлетворяющая свойствам $AB = BA = E$. Такую матрицу B обозначают A^{-1} и называют обратной к матрице A .

Утверждение 2. Пусть $A, B \in M_n(\mathbb{R})$ – обратимые матрицы. Тогда

1. AB тоже обратима и при этом $(AB)^{-1} = B^{-1}A^{-1}$.
2. A^t также будет обратима и $(A^t)^{-1} = (A^{-1})^t$ и обозначается A^{-t} .

Доказательство. 1) Действительно, надо проверить, что для AB существует двусторонняя обратная. Заметим, что $B^{-1}A^{-1}$ является таковой:

$$ABB^{-1}A^{-1} = E \quad \text{и} \quad B^{-1}A^{-1}AB = E$$

В частности, последнее означает, что $(AB)^{-1} = B^{-1}A^{-1}$.

2) Пусть матрица A обратима, тогда

$$AA^{-1} = E \quad \text{и} \quad A^{-1}A = E$$

Транспонируем оба равенства, получим

$$(A^{-1})^t A^t = E \quad \text{и} \quad A^t (A^{-1})^t = E$$

Это означает, что A^t обратима и при этом $(A^t)^{-1} = (A^{-1})^t$.

□

¹¹Ниже мы покажем, что из двусторонней обратимости следует, что матрица A обязана быть квадратной.

¹²Тут E_n означает единичную матрицу размера n .

Обратимые преобразования над СЛУ Пусть у нас есть $A \in M_{m \times n}(\mathbb{R})$ и $b \in \mathbb{R}^m$, которые задают систему линейных уравнений $Ax = b$, где $x \in \mathbb{R}^n$. Возьмем произвольную обратимую матрицу $C \in M_m(\mathbb{R})$. Тогда система $Ax = b$ эквивалентна системе $CAx = Cb$. Действительно, если для некоторого $y \in \mathbb{R}^n$ имеем $Ay = b$, то, умножая обе части на C слева, получим $CAy = Cb$, значит y решение второй системы. Наоборот, пусть $CAy = Cb$, тогда, умножая обе части на C^{-1} слева, получим $Ay = b$, значит y решение первой системы.

Сказанное выше значит, что мы можем менять СЛУ на эквивалентные с помощью умножения слева на любую обратимую матрицу. Мы уже знаем, что есть другая процедура преобразования СЛУ с таким же свойством – применение элементарных преобразований. Возникает резонный вопрос: какая процедура лучше? Оказывается, что между ними нет разницы в том смысле, что умножение на обратимую матрицу всегда совпадает с некоторой последовательностью элементарных преобразований и наоборот любое элементарное преобразование можно выразить с помощью умножения на обратимую матрицу. Этому свойству и будет посвящен остаток лекции.

2.8 Матрицы элементарных преобразований

Тип I Пусть $S_{ij}(\lambda) \in M_n(\mathbb{R})$ – матрица, полученная из единичной вписыванием в ячейку $i \ j$ числа λ (при этом $i \neq j$, то есть ячейка берется не на диагонали). Эта матрица имеет следующий вид:

$$i \quad \begin{matrix} & & j & \\ \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \lambda & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix} \end{matrix}$$

Тогда прямая проверка показывает, умножение $A \in M_{n \times m}(\mathbb{R})$ на $S_{ij}(\lambda)$ слева прибавляет j строку умноженную на λ к i строке матрицы A , а умножение $B \in M_{m \times n}(\mathbb{R})$ на $S_{ij}(\lambda)$ справа прибавляет i столбец умноженный на λ к j столбцу матрицы B . Заметим, что $S_{ij}(\lambda)^{-1} = S_{ij}(-\lambda)$.

Тип II Пусть $T_{ij} \in M_n(\mathbb{R})$ – матрица, полученная из единичной перестановкой i и j ($i \neq j$) столбцов (или что то же самое – строк). Эта матрица имеет следующий вид

$$\begin{matrix} & i & & j & \\ i & \begin{pmatrix} 1 & & & \\ & 0 & & 1 \\ & & \ddots & \\ & 1 & & 0 \end{pmatrix} & \\ j & & & & 1 \end{matrix}$$

Тогда прямая проверка показывает, умножение $A \in M_{n \times m}(\mathbb{R})$ на T_{ij} слева переставляет i и j строки матрицы A , а умножение $B \in M_{m \times n}(\mathbb{R})$ на T_{ij} справа переставляет i и j столбцы матрицы B . Заметим, что $T_{ij}^{-1} = T_{ij}$.

Тип III Пусть $D_i(\lambda) \in M_n(\mathbb{R})$ – матрица, полученная из единичной умножением i строки на $\lambda \in \mathbb{R} \setminus 0$ (или что то же самое – столбца). Эта матрица имеет следующий вид

$$i \quad \begin{matrix} & & i & \\ \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \lambda & \\ & & & \ddots \\ & & & & 1 \end{pmatrix} \end{matrix}$$

Тогда прямая проверка показывает, умножение $A \in M_{n \times m}(\mathbb{R})$ на $D_i(\lambda)$ слева умножает i строку A на λ , а умножение $B \in M_{m \times n}(\mathbb{R})$ на $D_i(\lambda)$ справа умножает i столбец матрицы B на λ . Заметим, что $D_i(\lambda)^{-1} = D_i(\lambda^{-1})$.

2.9 Невырожденные матрицы

Начнем с полезного утверждения.

Утверждение 3. Пусть $A \in M_n(\mathbb{R})$ – произвольная квадратная матрица. Тогда следующие условия эквивалентны:

1. Система $Ax = 0$ имеет только нулевое решение.
2. Система $A^t y = 0$ имеет только нулевое решение.
3. Матрица A представляется в виде $A = U_1 \cdot \dots \cdot U_k$, где U_i – матрицы элементарных преобразований.
4. Матрица A обратима.
5. Матрица A обратима слева, т.е. существует L такая, что $LA = E$.
6. Матрица A обратима справа, т.е. существует R такая, что $AR = E$.

Доказательство Утверждения 3. (1) \Rightarrow (3). Приведем A к улучшенному ступенчатому виду с помощью Гаусса. Так как $Ax = 0$ имеет только нулевое решение, то ступенчатый вид – это единичная матрица E . Пусть S_1, \dots, S_k – матрицы элементарных преобразований, которые мы совершили во время Гаусса. Это значит, что мы произвели следующие манипуляции

$$A \mapsto S_1 A \mapsto S_2 S_1 A \mapsto \dots \mapsto (S_k \dots S_1 A) = E$$

То есть $A = S_1^{-1} \dots S_k^{-1}$. Заметим, что S_i^{-1} – это матрица обратного элементарного преобразования к S_i . Обозначим $U_i = S_i^{-1}$ и получим требуемое.

(2) \Rightarrow (3). Проведем предыдущее рассуждение для матрицы A^t вместо A . Получим, что $A^t = U_1 \dots U_k$. Тогда $A = U_k^t \dots U_1^t$. Теперь осталось заметить, что U_i^t тоже является матрицей элементарного преобразования.

(3) \Rightarrow (4). Мы имеем $A = U_1 \dots U_k$, причем каждая из U_i обратима. Так как произведение обратимых обратимо, то A также обратима.

(4) \Rightarrow (5) и (4) \Rightarrow (6) очевидно, так это переход от более сильного условия к более слабому.

(5) \Rightarrow (1). Пусть A обратима слева и нам надо решить систему $Ax = 0$. Умножим ее слева на левый обратный к A , получим $x = 0$, что и требовалось.

(6) \Rightarrow (2). Пусть A обратима справа и нам надо решить систему $A^t y = 0$. Умножим эту систему слева на R^t , где R – правый обратный к A . Тогда $R^t A^t x = 0$. Но $R^t A^t x = (AR)^t x = Ex = x = 0$, что и требовалось. \square

В силу этого утверждения, мы не будем различать невырожденные и обратимые матрицы между собой.

Определение 4. Пусть $A \in M_n(\mathbb{R})$ – произвольная квадратная матрица. Будем говорить, что A невырождена¹³, если удовлетворяет любому из перечисленных в предыдущем утверждении условий.

Делители нуля Пусть $A \in M_n(\mathbb{R})$ – некоторая ненулевая матрица и пусть $B \in M_{nm}(\mathbb{R})$. Матрица B называется правым делителем нуля для A , если $AB = 0$. Условие (1) предыдущего утверждения эквивалентно отсутствию правых делителей нуля. Условие (1) не сильнее, значит надо показать, что оно влечет отсутствие делителей нуля. Если B – правый делитель нуля для A , то любой столбец b матрицы B удовлетворяет условию $Ab = 0$, а значит нулевой.

Аналогично определяются левые делители нуля для A и показывается, что их отсутствие равносильно условию (2) предыдущего результата.

Элементарные преобразования и обратимость Пусть $A \in M_{mn}(\mathbb{R})$ и $b \in \mathbb{R}^m$. Тогда у нас есть две процедуры преобразования СЛУ $Ax = b$:

1. Применение элементарных преобразований к строкам системы.
2. Умножение обеих частей равенства на обратимую матрицу: $Ax = b$ меняем на $CAx = Cb$, где $C \in M_n(\mathbb{R})$ – обратимая.

¹³Классически невырожденные матрицы определяются совсем по-другому, однако, все эти определения между собой эквивалентны. Будьте готовы к тому, что в литературе вы увидите совсем другое определение.

Так как любое элементарное преобразование сводится к умножению слева на обратимую матрицу, то мы видим, что первый вид модификации систем является частным случаем второго. В обратную сторону, из доказанного утверждения следует, что любая обратимая матрица может быть расписана как произведение матриц элементарных преобразований. Значит, умножить на обратимую матрицу слева – это все равно что сделать последовательность элементарных преобразований.

Главный плюс элементарных преобразований – у них простые матрицы, а минус – их нужно много, очень много, чтобы преобразовать одну систему в другую. С обратимыми матрицами все наоборот: сами матрицы устроены непонятно как, но зато нужно всего одно умножение матриц, чтобы перевести систему из одной в другую. Именно на это надо обращать внимание при выборе подхода по преобразованию систем.

Насыщенность обратимых Я хочу продемонстрировать еще одно полезное следствие из Утверждения 3. Предположим у нас есть две матрицы $A, B \in M_n(\mathbb{R})$. Тогда AB обратима тогда и только тогда, когда A и B обратимы. Действительно, справа налево мы уже знаем, обратимость обеих матриц A и B влечет обратимость произведения, мы даже знаем, что при этом $(AB)^{-1} = B^{-1}A^{-1}$. Надо лишь показать в обратную сторону. Предположим, что AB обратима, это значит, что для некоторой матрицы $D \in M_n(\mathbb{R})$ выполнено

$$ABD = E \quad \text{и} \quad DAB = E$$

Тогда первое равенство говорит, что BD является правым обратным к A . А в силу эквивалентности пунктов (4) и (6) Утверждения 3 это означает, что A обратима. Аналогично, DA является левым обратным к B и в силу эквивалентности пунктов (4) и (5) Утверждения 3, матрица B обратима. Так что произведение матриц обратимо тогда и только тогда, когда каждый сомножитель обратим.

2.10 Блочное умножение матриц

Формулы блочного умножения Пусть даны две матрицы, которые разбиты на блоки как показано ниже:

$$\begin{matrix} & \begin{matrix} k & s \end{matrix} \\ \begin{matrix} m \\ n \end{matrix} & \begin{pmatrix} A & B \\ C & D \end{pmatrix} \end{matrix} \quad \begin{matrix} & \begin{matrix} u & v \end{matrix} \\ \begin{matrix} k \\ s \end{matrix} & \begin{pmatrix} X & Y \\ W & Z \end{pmatrix} \end{matrix}$$

Числа m, n, k, s, u, v – размеры соответствующих блоков. Наша цель понять, что эти матрицы можно перемножать блочно. А именно, увидеть, что результат умножения этих матриц имеет вид

$$\begin{matrix} & \begin{matrix} u & v \end{matrix} \\ \begin{matrix} m \\ n \end{matrix} & \begin{pmatrix} AX + BW & AY + BZ \\ CX + DW & CY + DZ \end{pmatrix} \end{matrix}$$

Делается это таким трюком. В начале заметим, что

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ C & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}$$

После чего методом «пристального взгляда» перемножаем матрицы с большим количеством нулей (попробуйте проделать это!).

На этот факт можно смотреть вот как. Матрица – это прямоугольная таблица заполненная числами. А можно составлять прямоугольные таблицы, заполненные другими объектами, например матрицами. Тогда они складываются и перемножаются так же как и обычные матрицы из чисел. Единственное, надо учесть, что в блочном умножении есть разница между $AX + BW$ и $XA + BW$, так как A, B, X и W не числа, а матрицы, то их нельзя переставлять местами, порядок теперь важен.

Вот полезный пример. Пусть дана матрица из $M_{n+1}(\mathbb{R})$ вида

$$\begin{pmatrix} A & v \\ 0 & \lambda \end{pmatrix}, \quad \text{где} \quad A \in M_n(\mathbb{R}), \quad v \in \mathbb{R}^n, \quad \lambda \in \mathbb{R}$$

Тогда

$$\begin{pmatrix} A & v \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} A & v \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} A^2 & Av + v\lambda \\ 0 & \lambda^2 \end{pmatrix} = \begin{pmatrix} A^2 & Av + \lambda v \\ 0 & \lambda^2 \end{pmatrix} = \begin{pmatrix} A^2 & (A + \lambda E)v \\ 0 & \lambda^2 \end{pmatrix}$$

Предпоследнее равенство верно, так как не важно, с какой стороны умножать v на скаляр λ .

Вот еще один полезный пример блочного умножения. Пусть $x_1, \dots, x_m \in \mathbb{R}^n$ и $y_1, \dots, y_m \in \mathbb{R}^n$ – столбцы. Составим из этих столбцов матрицы $X = (x_1 | \dots | x_m)$ и $Y = (y_1 | \dots | y_m)$.¹⁴ Заметим, что $X, Y \in M_{n \times m}(\mathbb{R})$. Тогда

$$XY^t = (x_1 | \dots | x_m)(y_1 | \dots | y_m)^t = \sum_{i=1}^m x_i y_i^t$$

2.11 Блочные элементарные преобразования

Преобразования первого типа Пусть у нас дана матрица

$$\begin{matrix} & k & s \\ m & \begin{pmatrix} A & B \end{pmatrix} \\ n & \begin{pmatrix} C & D \end{pmatrix} \end{matrix}$$

Я хочу взять первую «строку» из матриц (A, B) умножить ее на некую матрицу R слева и прибавить результат к «строке» (C, D) . Для этого матрица R должна иметь n строк и m столбцов. То есть процедура будет выглядеть следующим образом

$$\begin{matrix} & k & s \\ m & \begin{pmatrix} A & B \end{pmatrix} \\ n & \begin{pmatrix} C & D \end{pmatrix} \end{matrix} \mapsto \begin{matrix} & k & s \\ m & \begin{pmatrix} A & B \\ C + RA & D + RB \end{pmatrix} \\ n & \end{matrix}$$

Оказывается, что такая процедура является умножением на обратимую матрицу слева, а именно

$$\begin{matrix} m & \begin{pmatrix} E & 0 \\ R & E \end{pmatrix} \\ n & \end{matrix} \begin{matrix} k & s \\ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \\ n & \end{matrix} = \begin{matrix} k & s \\ \begin{pmatrix} A & B \\ C + RA & D + RB \end{pmatrix} \\ n & \end{matrix}$$

Заметим, что

$$\begin{pmatrix} E & 0 \\ R & E \end{pmatrix}^{-1} = \begin{pmatrix} E & 0 \\ -R & E \end{pmatrix}$$

В частности из этого наблюдения следует, что блочные элементарные преобразования строк не меняют множества решений соответствующей системы.

Аналогично можно делать блочные элементарные преобразования столбцов. А именно

$$\begin{matrix} & k & s \\ m & \begin{pmatrix} A & B \end{pmatrix} \\ n & \begin{pmatrix} C & D \end{pmatrix} \end{matrix} \mapsto \begin{matrix} & k & s \\ m & \begin{pmatrix} A & B + AT \\ C & D + CT \end{pmatrix} \\ n & \end{matrix}$$

где T матрица с k строками и s столбцами. Как и в случае преобразований со строками, эта процедура сводится к операции умножения на обратимую матрицу справа

$$\begin{matrix} & k & s \\ m & \begin{pmatrix} A & B \\ C & D \end{pmatrix} \\ n & \end{matrix} \begin{matrix} k & s \\ \begin{pmatrix} E & T \\ 0 & E \end{pmatrix} \\ s & \end{matrix} = \begin{matrix} & k & s \\ m & \begin{pmatrix} A & B + AT \\ C & D + CT \end{pmatrix} \\ n & \end{matrix}$$

Как и раньше

$$\begin{pmatrix} E & T \\ 0 & E \end{pmatrix}^{-1} = \begin{pmatrix} E & -T \\ 0 & E \end{pmatrix}$$

Замечание Обратите внимание, что при блочных преобразованиях строк умножение на матрицу-коэффициент R происходит слева, а при преобразованиях столбцов умножение на матрицу-коэффициент T происходит справа.

¹⁴Данная запись означает, что мы берем столбцы x_i и записываем их подряд в одну большую таблицу.

Преобразования второго типа Преобразование вида

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \mapsto \begin{pmatrix} C & D \\ A & B \end{pmatrix}$$

сводится к умножению на обратимую блочную матрицу слева

$$\begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} C & D \\ A & B \end{pmatrix}$$

А преобразование

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \mapsto \begin{pmatrix} B & A \\ D & C \end{pmatrix}$$

сводится к умножению на обратимую блочную матрицу справа

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix} = \begin{pmatrix} B & A \\ D & C \end{pmatrix}$$

При этом

$$\begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}$$

Преобразования третьего типа Если $R \in M_m(\mathbb{R})$ – обратимая матрица, то

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \mapsto \begin{pmatrix} RA & RB \\ C & D \end{pmatrix}$$

является преобразованием умножения на обратимую матрицу слева, а именно

$$\begin{pmatrix} R & 0 \\ 0 & E \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} RA & RB \\ C & D \end{pmatrix}$$

при этом

$$\begin{pmatrix} R & 0 \\ 0 & E \end{pmatrix}^{-1} = \begin{pmatrix} R^{-1} & 0 \\ 0 & E \end{pmatrix}$$

Аналогично, для обратимой матрицы $T \in M_k(\mathbb{R})$, преобразование

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \mapsto \begin{pmatrix} AT & B \\ CT & D \end{pmatrix}$$

является преобразованием умножения на обратимую матрицу справа, а именно

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} T & 0 \\ 0 & E \end{pmatrix} = \begin{pmatrix} AT & B \\ CT & D \end{pmatrix}$$

Как и раньше, при работе со строками умножение на матрицу-коэффициент происходит слева, а при работе со столбцами – справа.

2.12 Массовое решение систем

Пусть нам надо решить сразу несколько систем $Ax_1 = b_1, \dots, Ax_k = b_k$, где $A \in M_{m \times n}(\mathbb{R})$, $b_i \in \mathbb{R}^m$ и $x_i \in \mathbb{R}^n$. Определим матрицы $X = (x_1 | \dots | x_k) \in M_{n \times k}(\mathbb{R})$ и $B = (b_1 | \dots | b_k) \in M_{m \times k}(\mathbb{R})$ составленные из столбцов x_i и b_i соответственно. Тогда по формулам блочного умножения матриц

$$AX = A(x_1 | \dots | x_k) = (Ax_1 | \dots | Ax_k) = (b_1 | \dots | b_k) = B$$

То есть массовое решение системы уравнений равносильно решению матричного уравнения $AX = B$.

Решение матричных уравнений

Дано $A \in M_{m \times n}(\mathbb{R}), B \in M_{m \times k}(\mathbb{R})$.

Задача Найти $X \in M_{n \times k}(\mathbb{R})$ такую, что $AX = B$.

Алгоритм

1. Составить расширенную матрицу $(A|B)$. Например, если $A \in M_{3 \times 3}(\mathbb{R})$, а $B \in M_{3 \times 2}(\mathbb{R})$, то получим

$$(A|B) = \left(\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & b_{11} & b_{12} \\ a_{21} & a_{22} & a_{23} & b_{21} & b_{22} \\ a_{31} & a_{32} & a_{33} & b_{31} & b_{32} \end{array} \right)$$

2. Привести расширенную матрицу $(A|B)$ к улучшенному ступенчатому виду. В примере выше, может получиться

$$\left(\begin{array}{ccc|cc} 1 & a_{12} & 0 & b_{11} & 0 \\ 0 & 0 & 1 & b_{21} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right) \text{ или } \left(\begin{array}{ccc|cc} 1 & 0 & a_{13} & b_{11} & b_{12} \\ 0 & 1 & a_{23} & b_{21} & b_{22} \\ 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

3. Для каждого столбца матрицы X выразить его главные переменные через свободные и записать ответ в виде матрицы. Если для какого-то столбца решений нет, то нет решений и у матричного уравнения $AX = B$. В примере выше, в первом случае нет решения для второго столбца, потому что решений нет в этом случае. Во втором случае,

$$X = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -a_{13} \\ -a_{23} \\ 1 \end{pmatrix} \begin{pmatrix} t & u \end{pmatrix}, \text{ где } t, u \in \mathbb{R}$$

Если нужно решить матричное уравнение $XA = B$ для матриц соответствующего размера, то можно его транспонировать и свести задачу к рассмотренной. А именно, это уравнение равносильно уравнению $A^t X^t = B^t$. Тогда его можно решать относительно X^t , а потом транспонировать ответ.

2.13 Нахождение обратной матрицы методом Гаусса

Дано Матрица $A \in M_n(\mathbb{R})$.

Задача Понять, обратима ли матрица A , и если она обратима, то найти ее обратную A^{-1} .

Алгоритм

1. Нам надо по сути решить систему $AX = E$, где E – единичная матрица. Потому составим расширенную матрицу системы $(A|E)$.
2. Приведем эту матрицу к улучшенному ступенчатому виду.
3. В результате возможны 2 случая:
 - (а) После приведения получили матрицу $(E|B)$. Тогда A обратима и $A^{-1} = B$.
 - (б) После приведения получили матрицу $(D|B)$ и у матрицы D есть свободные позиции. Тогда матрица A не обратима.

Заметим, что если в процессе алгоритма мы слева от черты в расширенной матрице нашли свободную переменную, то на этом можно остановиться – матрица A необратима.

Корректность алгоритма Давайте я поясню почему алгоритм работает корректно. Пусть у нас есть система $AX = B$ с краткой записью $(A|B)$. Если мы применим элементарное преобразование строк к краткой записи, то это будет означать умножение на матрицу элементарного преобразования слева, то есть при переходе $(A|B) \mapsto (UA|UB)$ мы меняем систему $AX = B$ на $UAX = UB$. А значит, если матрица X была решением $AX = B$, то мы имеем верное равенство двух матриц $AX = B$. Если две одинаковые матрицы слева домножить на одну и ту же матрицу, то результат получится равным, то есть отсюда следует, что $UAX = UB$. То есть любое решение системы $AX = B$ превращается в решение системы $UAX = UB$. Так как матрица элементарного преобразования U обратима, то мы можем домножить второе на U^{-1} , а значит работает рассуждение в обратную сторону и все решения второй являются решениями первой.

Теперь мы знаем, что меняя по алгоритму систему, мы не меняем множество решений. Кроме того, по алгоритму, у нас в результате работы бывают две ситуации, либо мы приходим к ситуации $(E|B)$ либо к $(D|B)$ и в D есть свободная позиция. Давайте разберем их отдельно.

1. Пусть мы привели систему к виду $(E|B)$. Эта запись соответствует системе $EX = B$, то есть $X = B$. Более того, полученная система эквивалентна исходной $AX = B$. Теперь мы видим, что у системы $X = B$ единственное решение B , а это значит, что и у системы $AX = B$ единственное решение B (так как они эквивалентны). А значит в этом случае B – это правая обратная к A , а следовательно и просто обратная.
2. Теперь предположим, что мы получим $(D|B)$, где у D есть свободная переменная. Так как мы переходили от $(A|E)$ к $(D|B)$ элементарными преобразованиями строк, то для некоторой обратимой матрицы $C \in M_n(\mathbb{R})$ выполнено $D = CA$. Так как у матрицы D есть свободная позиция и она квадратная ¹⁵, то обязательно найдется нулевая строка. А раз так, то матрица D не может быть обратима справа. Действительно, тогда в произведении DR для любой $R \in M_n(\mathbb{R})$ будет иметь нулевую строку там же, где нулевая строка у D . А значит, не может быть E . Раз матрица D не обратима, то и матрица A не обратима, иначе D была бы обратима как произведение обратимых матриц.

2.14 Классификация СЛУ

Единственность улучшенного ступенчатого вида Давайте вначале ответим на очень важный вопрос: а единственный ли у матрицы улучшенный ступенчатый вид? Очевидно, что ступенчатый вид не единственный. Однако, улучшенный ступенчатый вид окажется однозначно определенным. Это означает, что у ступенчатого вида однозначно определена его форма (количество и длины ступенек). В частности у любой СЛУ однозначно определены главные и свободные переменные. Все это не бросается сразу в глаза и требует доказательства. Давайте начнем со сравнения ступенчатых видов.

Утверждение 5. Пусть $A \in M_{m,n}(\mathbb{R})$ и $B \in M_{k,n}(\mathbb{R})$ – матрицы в ступенчатом виде и системы $Ax = 0$ и $Bx = 0$ эквивалентны.¹⁶ Тогда набор главных и свободных переменных у этих систем совпадает.

Доказательство. Пусть $E_A, E_B \subseteq \mathbb{R}^n$ – множества решений систем $Ax = 0$ и $Bx = 0$, соответственно. Теперь для системы $Ax = 0$ и номера $1 \leq i \leq n$ введем следующее свойство

$$P_i(A) = \text{«Для любого решения } x \in E_A \text{ из условия } x_{i+1} = \dots = x_n = 0 \text{ следует, что } x_i = 0\text{.»}$$

Я утверждаю, что свойство $P_i(A)$ выполнено тогда и только тогда, когда x_i является главной переменной для $Ax = 0$. Если это так, то отсюда следует наше утверждение. Действительно, так как системы $Ax = 0$ и $Bx = 0$ имеют одинаковое множество решений, то и свойство $P_i(A)$ выполняется тогда и только тогда, когда выполняется свойство $P_i(B)$. Значит, множество главных переменных совпадает, и как следствие совпадает множество свободных переменных.

Теперь докажем требуемое перебором. Пусть x_i является главной переменной для $Ax = 0$. Рассмотрим множество

$$E_A^0 = \{x \in E_A \mid x_k = 0 \text{ при } k > i\}$$

Но в ступенчатом виде переменная x_i выражается через переменные с более высоким индексом, раз все переменные правее равны нулю, то и $x_i = 0$. То есть для главной переменной свойство выполняется.

Теперь покажем, что если x_i свободная, то можно предъявить решение $x \in E_A^0$ такое, что $x_i \neq 0$. Действительно, зададим все свободные переменные кроме x_i нулями, а $x_i = 1$. Тогда главные восстанавливаются

¹⁵Вот то место, где мы пользуемся квадратностью матрицы.

¹⁶То есть имеют одинаковое множество решений.

однозначно по свободным и мы получим какое-то решение x . По построению $x_i \neq 0$. И надо лишь проверить, что $x_{i+1} = \dots = x_n = 0$. Действительно, все свободные переменные правее x_i заданы нулями. Но все главные переменные правее x_i зависят от еще более правых переменных, то есть они тоже должны быть нулями, что и требовалось. \square

Утверждение 6. Пусть $S_1 \in M_{m,n}(\mathbb{R})$ и $S_2 \in M_{k,n}(\mathbb{R})$ – произвольные матрицы в улучшенном ступенчатом виде. Если $S_1 x = 0$ эквивалентно $S_2 x = 0$, то после удаления нулевых строк матрицы S_1 и S_2 совпадут.

Доказательство. Из предыдущего утверждения мы уже знаем, что у систем $S_1 x = 0$ и $S_2 x = 0$ одинаковый набор главных и свободных переменных. То есть у них одинаковая ступенчатая форма и надо лишь показать, что ненулевые числа в ней одинаковые. Пусть для определенности матрицы S_1 и S_2 имеют следующий вид:

$$S_1 = \begin{pmatrix} 1 & * & 0 & * & 0 & 0 & * & * & * \\ & & 1 & * & 0 & 0 & * & * & * \\ & & & & 1 & 0 & * & * & * \\ & & & & & & 1 & * & * \end{pmatrix} \quad S_2 = \begin{pmatrix} 1 & \bullet & 0 & \bullet & 0 & 0 & \bullet & \bullet & \bullet \\ & & 1 & \bullet & 0 & 0 & \bullet & \bullet & \bullet \\ & & & & 1 & 0 & \bullet & \bullet & \bullet \\ & & & & & & 1 & \bullet & \bullet \end{pmatrix}$$

Теперь для каждого свободного столбца покажем, что они совпадают. Рассмотрим произвольный i -ый свободный столбец

$$S_1 = \begin{pmatrix} 1 & * & 0 & * & 0 & 0 & a_1 & * & * \\ & & 1 & * & 0 & 0 & a_2 & * & * \\ & & & & 1 & 0 & a_3 & * & * \\ & & & & & & 1 & a_4 & * \end{pmatrix} \quad S_2 = \begin{pmatrix} 1 & \bullet & 0 & \bullet & 0 & 0 & b_1 & \bullet & \bullet \\ & & 1 & \bullet & 0 & 0 & b_2 & \bullet & \bullet \\ & & & & 1 & 0 & b_3 & \bullet & \bullet \\ & & & & & & 1 & b_4 & \bullet \end{pmatrix}$$

Давайте построим решение системы $S_1 x = 0$, где все свободные переменные кроме x_i равны нулю, а $x_i = 1$. Тогда получится одно единственное решение вида¹⁷

$$(-a_1, 0, -a_2, 0, -a_3, -a_4, 1, 0, 0)$$

С другой стороны, давайте построим решение системы $S_2 x = 0$, где все свободные переменные кроме x_i равны нулю, а $x_i = 1$. Тогда получится одно единственное решение вида

$$(-b_1, 0, -b_2, 0, -b_3, -b_4, 1, 0, 0)$$

Так как эти системы эквивалентны, то оба решения являются решениями, например, $S_1 x = 0$. А раз это решения одной системы с одинаковыми значениями свободных переменных, то и значения главных будут совпадать. А это и значит, что i -ые столбцы в системах S_1 и S_2 будут одинаковыми. \square

Из этого утверждения следует, что матрица улучшенного ступенчатого вида для любой матрицы $A \in M_{m,n}(\mathbb{R})$ определена однозначно. Так как если матрица A приводится к двум разным ступенчатым видам, то их однородные системы эквивалентны, а значит они совпадают. Потому, говоря о матрице A , можно говорить и о ее улучшенном ступенчатом виде без какой-либо неоднозначности.

Классификация

Утверждение 7. Пусть $A, B \in M_{m,n}(\mathbb{R})$. Тогда следующее эквивалентно:

1. Системы $Ax = 0$ и $Bx = 0$ эквивалентны.
2. A приводится к B элементарными преобразованиями строк.
3. Существует обратимая $C \in M_m(\mathbb{R})$ такая, что $B = CA$.
4. Матрица улучшенного ступенчатого вида для A совпадает с матрицей улучшенного ступенчатого вида для B .

¹⁷Тут красным обозначены значения главных переменных, а синим – свободных.

Доказательство. Мы все это уже доказали по сути, потому напомним, что откуда следует. $(2) \Rightarrow (1)$ Так как элементарные преобразования меняют систему на эквивалентную. $(1) \Rightarrow (4)$ Предыдущее утверждение. $(4) \Rightarrow (2)$ Если матрицы A и B приводятся элементарными преобразованиями к одной и той же матрице (улучшенного ступенчатого вида), то они переводятся и друг в друга. Эквивалентность $(2) \Leftrightarrow (3)$ следует из Утверждения 3 о том, что матрица обратима тогда и только тогда, когда она раскладывается в произведение элементарных. \square

Смысл этого утверждения в следующем. Возьмем множество всех однородных систем фиксированного размера, которое описывается матрицами $M_{m \times n}(\mathbb{R})$. Тогда на этом множестве есть отношение эквивалентности: системы эквивалентны если они имеют одинаковое множество решений. Это полезное свойство, потому что нам не важно какую из систем решать среди эквивалентных. Однако, это свойство сложно проверяется. С другой стороны, у нас есть процедура изменения системы (элементарные преобразования), которая меняет системы на заведомо эквивалентные. Сделаем следующие замечания:

1. Утверждается, что эта процедура эффективная в том смысле, что если уж какие-то системы были эквивалентны, то мы обязательно от одной к другой сможем перейти элементарными преобразованиями.
2. Все то же самое верно и для второй процедуры – умножение на обратимую матрицу слева (потому что это по сути та же самая процедура).
3. Утверждается, что в каждом классе эквивалентных систем мы можем найти одну единственную матрицу улучшенного ступенчатого вида. То есть классов попарно неэквивалентных систем ровно столько же, сколько матриц улучшенного ступенчатого вида.
4. Последнее означает, что свойства системы с произвольной матрицей точно такие же, как у какой-то системы в улучшенном ступенчатом виде. Потому в абстрактных задачах про системы можно всегда предполагать, что система уже имеет улучшенный ступенчатый вид.

2.15 Полиномиальное исчисление от матриц

Обозначим множество всех многочленов с вещественными коэффициентами через $\mathbb{R}[x]$. Формально это значит: $\mathbb{R}[x] = \{a_0 + a_1x + \dots + a_nx^n \mid n \in \mathbb{Z}_+, a_i \in \mathbb{R}\}$. Аналогично можно обозначать многочлены с рациональными, целыми, комплексными и т.д. коэффициентами.

Подстановка матриц в многочлены Пусть $p(x) = a_0 + a_1x + \dots + a_nx^n$ – многочлен с вещественными коэффициентами, а $A \in M_n(\mathbb{R})$. Тогда можно определить $f(A) = a_0E + a_1A^1 + \dots + a_nA^n \in M_n(\mathbb{R})$. Если определить $A^0 = E$, то формула становится более единообразной $f(A) = a_0A^0 + a_1A^1 + \dots + a_nA^n$. Однако, психологически проще думать так: вместо x подставляем A , а свободный член отождествляем со скалярными матрицами. Отметим, что если два многочлена равны, то и их значения на матрице A тоже равны.

Утверждение. Пусть $A \in M_n(\mathbb{R})$ и $f, g \in \mathbb{R}[x]$ – два произвольных многочлена, тогда:

1. $(f + g)(A) = f(A) + g(A)$.
2. $(fg)(A) = f(A)g(A)$.
3. $f(\lambda E) = f(\lambda)E$.
4. $f(C^{-1}AC) = C^{-1}f(A)C$ для любой обратимой $C \in M_n(\mathbb{R})$
5. Матрицы $f(A)$ и $g(A)$ коммутируют между собой.

Доказательство. Все это делается прямой проверкой по определению. Давайте объясним свойства (2) и (4).

(2) Пусть

$$f = \sum_{k=0}^n a_k x^k \text{ и } g = \sum_{k=0}^m b_k x^k$$

тогда

$$fg = \sum_{k=0}^{n+m} \left(\sum_{s+t=k} a_s b_t \right) x^k$$

Потому надо проверить равенство:

$$\left(\sum_{k=0}^n a_k A^k\right) \left(\sum_{k=0}^m b_k A^k\right) = \sum_{k=0}^{n+m} \left(\sum_{s+t=k} a_s b_t\right) A^k$$

которое следует из перестановочности A со своими степенями и коэффициентами.

(4) Заметим, что

$$(C^{-1}AC)^n = C^{-1}ACC^{-1}AC \dots C^{-1}AC = C^{-1}A^nC$$

Осталось воспользоваться дистрибутивностью умножения, т.е. $C^{-1}(A+B)C = C^{-1}AC + C^{-1}BC$. \square

Обнуляющий многочлен

Утверждение 8. Пусть $A \in M_n(\mathbb{R})$, тогда: Существует многочлен $f \in \mathbb{R}[x]$ не равный тождественно нулю степени не больше n^2 такой, что $f(A) = 0$.

Доказательство. Давайте искать многочлен f с неопределенными коэффициентами в виде $f = a_0 + a_1x + \dots + a_{n^2}x^{n^2}$. Надо чтобы было выполнено равенство $a_0E + a_1A + \dots + a_{n^2}A^{n^2} = 0$. Последнее равенство означает равенство матрицы слева нулевой матрице справа. Это условие задается равенством всех n^2 ячеек матриц: $(a_0E + a_1A + \dots + a_{n^2}A^{n^2})_{ij} = 0$ для всех i, j . Каждое из этих условий является линейным уравнением вида $a_0(E)_{ij} + a_1(A)_{ij} + \dots + a_{n^2}(A^{n^2})_{ij} = 0$. То есть у нас есть система с n^2 уравнениями и $n^2 + 1$ неизвестной. А значит при приведении этой системы к ступенчатому виду у нас обязательно будет свободная переменная, а значит мы сможем найти ненулевое решение. \square

На самом деле можно показать, что найдется многочлен степени не больше n , зануляющий нашу матрицу. Однако, мы пока не в состоянии этого сделать. Усиленный результат будет показан позже с помощью определителей. Соответствующее утверждение называется теоремой Гамильтона-Кэли.

2.16 Спектр

Теперь я хочу обсудить очень полезную матричную характеристику – спектр. В дальнейшем окажется, что это так называемые «собственные значения» матрицы. Определение ниже позволяет избежать сложного геометрического определения и использовать «собственные значения» гораздо раньше в теории, чем это обычно делается.

Определение 9. Пусть $A \in M_n(\mathbb{R})$ определим вещественный спектр матрицы A следующим образом:¹⁸

$$\text{спес}_{\mathbb{R}} A = \{\lambda \in \mathbb{R} \mid A - \lambda E \text{ не обратима}\}$$

Утверждение 10. Пусть $A \in M_n(\mathbb{R})$ и пусть $g \in \mathbb{R}[x]$ такой, что $g(A) = 0$. Тогда $\text{спес}_{\mathbb{R}} A$ лежит среди вещественных корней многочлена g .¹⁹ В частности $|\text{спес}_{\mathbb{R}} A| \leq \deg g$ и спектр всегда конечен.

Доказательство. Чтобы показать, что любой элемент спектра является корнем g , мы покажем обратное. А именно, что если число не корень g , то оно не лежит в спектре. Пусть $\lambda \in \mathbb{R}$ – число не являющееся корнем g , то есть $g(\lambda) \neq 0$. Нам надо показать, что λ не в спектре A , то есть, что $A - \lambda E$ обратима. Для этого разделим многочлен g на $x - \lambda$ с остатком, получим $g(x) = h(x)(x - \lambda) + g(\lambda)$. Теперь в левую и правую часть равенства подставим A . Получим

$$0 = g(A) = h(A)(A - \lambda E) + g(\lambda)E$$

Перенесем $g(\lambda)E$ в другую сторону и поделим на $-g(\lambda)$, получим

$$E = -\frac{1}{g(\lambda)}h(A)(A - \lambda E)$$

То есть $-\frac{1}{g(\lambda)}h(A)$ является обратным к $A - \lambda E$, что и требовалось. \square

Так как у нас для любой матрицы найдется многочлен степени n^2 ее зануляющий, то спектр всегда конечен и его размер не превосходит n^2 . Как говорилось выше, на самом деле, можно найти многочлен степени n , потому спектр всегда не превосходит по мощности n .

¹⁸Аналогично определяются спектры в рациональном, комплексном и прочих случаях.

¹⁹Если мы вместо вещественного спектра берем любой другой спектр (комплексный или рациональный или еще какой), то соответствующее утверждение остается верным. А именно, комплексная часть спектра будет содержаться среди комплексных корней многочлена, рациональная часть – среди рациональных корней и т.д.

Примеры

1. Пусть $A \in M_n(\mathbb{R})$ – диагональная матрица с числами $\lambda_1, \dots, \lambda_n$ на диагонали, т.е.

$$A = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

Так как диагональные матрицы складываются и умножаются поэлементно

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} + \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{pmatrix} = \begin{pmatrix} \lambda_1 + \mu_1 & & \\ & \ddots & \\ & & \lambda_n + \mu_n \end{pmatrix}$$

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{pmatrix} = \begin{pmatrix} \lambda_1 \mu_1 & & \\ & \ddots & \\ & & \lambda_n \mu_n \end{pmatrix}$$

То для любого многочлена $f \in \mathbb{R}[x]$ верно

$$f \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix}$$

То есть многочлен f зануляет A тогда и только тогда, когда он зануляет все λ_i . Например, в качестве такого многочлена подойдет $f(x) = (x - \lambda_1) \dots (x - \lambda_n)$.

Давайте покажем, что $\text{спес}_{\mathbb{R}} A = \{\lambda_1, \dots, \lambda_n\}$. Так как многочлен f зануляет A , утверждение 10 влечет, что спектр содержится среди его корней. Значит, надо показать, что $A - \lambda_i E$ необратим для любого i . Последнее легко видеть, так как $A - \lambda_i$ содержит 0 на i -ом месте на диагонали.

2. Пусть матрица A имеет верхнетреугольный или нижнетреугольный вид

$$A = \begin{pmatrix} \lambda_1 & \dots & * \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix} \quad \text{или} \quad A = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ * & \dots & \lambda_n \end{pmatrix}$$

Тогда $\text{спес}_{\mathbb{R}} A = \{\lambda_1, \dots, \lambda_n\}$. Для этого достаточно показать, что верхнетреугольная матрица обратима тогда и только тогда, когда на диагонали только ненулевые элементы. Пусть все диагональные элементы ненулевые. Тогда составим уравнение для поиска обратной $AX = E$ и будем решать его обратным ходом Гаусса.

$$\left(\begin{array}{ccc|ccc} \lambda_1 & \dots & * & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n & 0 & \dots & 1 \end{array} \right) \mapsto \left(\begin{array}{ccc|ccc} 1 & \dots & * & \frac{1}{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & 0 & \dots & \frac{1}{\lambda_n} \end{array} \right) \mapsto \left(\begin{array}{ccc|ccc} 1 & \dots & 0 & \frac{1}{\lambda_1} & \dots & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & 0 & \dots & \frac{1}{\lambda_n} \end{array} \right)$$

И мы видим, что обратная матрица тоже верхнетреугольная и на диагонали у нее стоят обратные элементы.

В другую сторону, пусть на диагонали матрицы A встретился ноль. Если ноль в последней строке, то матрица содержит нулевую строку и значит не обратима. Значит можем считать, что ноль стоит где-то выше. Сделаем наблюдение, что если матрица A' получена из матрицы A элементарными преобразованиями строк, то A обратима тогда и только тогда, когда A' обратима. Это непосредственно следует из того, что $A' = UA$ для обратимой матрицы U . Значит нам достаточно элементарными преобразованиями строк добиться нулевой строки в матрице A . Рассмотрим самый нижний ноль на диагонали A .

$$\begin{pmatrix} \lambda_1 & \dots & \dots & * & * \\ & \ddots & \dots & * & * \\ & & 0 & * & * \\ & & & \lambda_{n-1} & * \\ & & & & \lambda_n \end{pmatrix}$$

Вычтем с подходящим коэффициентом следующую строку из строки с нулем на диагонали и занулим элемент справа от нуля.

$$\begin{pmatrix} \lambda_1 & \dots & \dots & * & * \\ & \ddots & \dots & * & * \\ & & 0 & 0 & * \\ & & & \lambda_{n-1} & * \\ & & & & \lambda_n \end{pmatrix}$$

Потом вычтем следующую строку и так далее и получим

$$\begin{pmatrix} \lambda_1 & \dots & \dots & * & * \\ & \ddots & \dots & * & * \\ & & 0 & 0 & 0 \\ & & & \lambda_{n-1} & * \\ & & & & \lambda_n \end{pmatrix}$$

Это доказывает необратимость матрицы в случае нуля на диагонали. Абсолютно аналогичные рассуждения работают с нижнетреугольными матрицами.

3. Пусть $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$. Прямое вычисление показывает, что $A^2 = -E$, то есть многочлен $f(x) = x^2 + 1$ зануляет A . Покажем, что $\text{спес}_{\mathbb{R}} A = \emptyset$. Действительно, по утверждению 10 спектр должен содержаться среди корней многочлена $f(x) = x^2 + 1$. Однако, этот многочлен не имеет вещественных корней. Этот пример объясняет, почему вещественных чисел иногда не достаточно и мы хотим работать с комплексными числами. Например, в комплексном случае $\text{спес}_{\mathbb{C}} A = \{i, -i\}$.
4. Обратите внимание, что матрица A является корнем линейного многочлена $x - \lambda$ тогда и только тогда, когда $A = \lambda E$. В частности, не скалярная матрица зануляется многочленом степени 2 и выше.

Минимальный многочлен Пусть $A \in M_n(\mathbb{R})$ – некоторая матрица. Рассмотрим множество всех ненулевых многочленов зануляющих A . Формально мы смотрим на множество

$$M = \{f \in \mathbb{R}[x] \mid f(A) = 0, f \neq 0\}$$

Пусть $f_{\min} \in M$ – многочлен самой маленькой степени со старшим коэффициентом 1. Тогда он называется минимальным многочленом матрицы A .

Утверждение 11. Пусть $A \in M_n(\mathbb{R})$, тогда верны следующие утверждения:

1. Минимальный многочлен f_{\min} существует.
2. Минимальный многочлен делит любой другой многочлен зануляющий A .
3. Минимальный многочлен единственный.
4. $\lambda \in \text{спес}_{\mathbb{R}} A$ тогда и только тогда, когда $f_{\min}(\lambda) = 0$.

Доказательство. (1). По утверждению 8 у нас всегда найдется многочлен зануляющий A , а значит M не пусто. Так как степень не может убывать бесконечно, то мы обязательно найдем многочлен самой маленькой степени, который зануляет A . Осталось разделить его на старший коэффициент.

(2). Пусть $f \in M$ – произвольный многочлен, а f_{\min} – какой-то минимальный. Тогда разделим f на f_{\min} с остатком, получим

$$f(x) = h(x)f_{\min}(x) + r(x)$$

где $\deg r < \deg f_{\min}$. Подставим в это равенство матрицу A , получим

$$0 = f(A) = h(A)f_{\min}(A) + r(A) = r(A)$$

Значит мы нашли многочлен r , который зануляет A и меньше f_{\min} по степени. Такое может быть только если $r(x) = 0$.

(3). Пусть f_{\min} и f'_{\min} – два минимальных многочлена матрицы A . Тогда у них по определению одинаковая степень. Рассмотрим $r(x) = f_{\min}(x) - f'_{\min}(x)$. Многочлен $r(x)$ степени строго меньше, так как оба минимальных имеют старший коэффициент 1. Кроме того, $r(A) = f_{\min}(A) - f'_{\min}(A) = 0 - 0 = 0$. А значит $r(x) = 0$.

(4). Мы уже знаем, что $\text{спес}_{\mathbb{R}} A$ лежит среди корней f_{\min} (утверждение 10). Осталось показать обратное включение. Предположим обратное, что есть $\lambda \in \mathbb{R}$ такое, что $f_{\min}(\lambda) = 0$, но $\lambda \notin \text{спес}_{\mathbb{R}} A$. Тогда $f_{\min}(x) = (x - \lambda)h(x)$. Подставим в это равенство матрицу A и получим

$$0 = f_{\min}(A) = (A - \lambda E)h(A)$$

Так как $\lambda \notin \text{спес}_{\mathbb{R}} A$, то матрица $A - \lambda E$ обратима, а значит на нее можно сократить, то есть $h(A) = 0$ и степень h строго меньше степени f_{\min} , хотя сам h – ненулевой многочлен. Последнее противоречит с нашим предположением о том, что f_{\min} минимальный. \square

Поиск минимального многочлена Пусть задана матрица $A \in M_n(\mathbb{R})$. Тогда мы знаем, что найдется многочлен $f \in \mathbb{R}[x]$ такой, что $f(A) = 0$. Кроме того, я сообщил, что $\deg f \leq n$. Давайте обсудим, как найти подобный многочлен. Будем искать его с неопределенными коэффициентами $f(x) = a_0 + a_1x + \dots + a_nx^n$. Подставим в многочлен матрицу A и приравняем результат к нулю.

$$f(A) = a_0E + a_1A + \dots + a_nA^n = 0$$

Тогда то, что написано, является системой из n^2 уравнений, а именно

$$\{1 \leq i, j \leq n\} E_{ij}a_0 + A_{ij}a_1 + \dots + (A^n)_{ij}a_n = 0$$

Здесь через B_{ij} обозначены коэффициенты матрицы B , например, E_{ij} – это ij -ый коэффициент единичной матрицы, а $(A^n)_{ij}$ – ij -ый коэффициент матрицы A^n .

Теперь нас интересует ненулевое решение этой системы, у которого как можно больше нулей справа. Давайте поясню. Такое решение отвечает зануляющему многочлену. Мы хотим выбрать такой многочлен как можно меньшей степени. То есть мы хотим по возможности занулить a_n , потом a_{n-1} , потом a_{n-2} и так далее, пока находится ненулевое решение. Предположим, что мы привели систему к ступенчатому виду и a_k – самая левая свободная переменная. Я утверждаю, что k и будет степенью минимального многочлена, а чтобы его найти надо положить $a_k = 1$, и все остальные свободные переменные равными нулю.

Действительно, если мы сделали, как описано, то все главные переменные правее a_k тоже равны нулю, ибо они зависят от свободных переменных, стоящих правее, а они в нашем случае нулевые. То есть a_k будет старший ненулевой коэффициент в искомом многочлене, а значит k будет его степенью. Почему нельзя найти меньше. Чтобы найти меньше надо занулить еще и a_k . То есть все свободные переменные в этом случае будут нулевыми, а тогда и все главные будут нулевыми, а это даст нулевое решение, что противоречит нашим намерениям найти ненулевой многочлен.

Вычленение из какого-то зануляющего Предположим, что вы угадали какой-нибудь зануляющий многочлен для вашей матрицы $A \in M_n(\mathbb{R})$, а именно, нашли какой-то $f \in \mathbb{R}[x]$ такой, что $f(A) = 0$. Тогда можно попытаться найти минимальный многочлен среди делителей многочлена f . Эта процедура требует уметь искать эти самые делители. Но в некоторых ситуациях эта процедура тоже бывает полезна. Например, в случае большой блочной матрицы A бывает проще найти зануляющий многочлен.

Замечание о спектре Можно показать, что любой вещественный многочлен $f \in \mathbb{R}[x]$ единственным образом разваливается в произведение

$$f(x) = (x - \lambda_1) \dots (x - \lambda_k) q_1(x) \dots q_r(x)$$

где числа $\lambda_i \in \mathbb{R}$ могут повторяться, а $q_i(x)$ – многочлены второй степени с отрицательным дискриминантом (то есть без вещественных корней).

Пусть теперь f_{\min} – минимальный многочлен некоторой матрицы A . Разложим его подобным образом. Тогда мы видим из предыдущего утверждения, что $\text{спес}_{\mathbb{R}} A$ помнит информацию только о первой половине сомножителей и теряет информацию о квадратичных многочленах. Однако, если бы мы рассмотрели f_{\min} как многочлен с комплексными коэффициентами, то мы бы могли доразложить все $q_i(x)$ на линейные множители и $\text{спес}_{\mathbb{C}} A$ помнит информацию о всех сомножителях f_{\min} . Еще надо понимать, что каждое $x - \lambda$ может несколько раз участвовать в разложении f_{\min} , но спектр не помнит это количество, он лишь знает был ли там данный $x - \lambda$ или нет.

Замечание об арифметических свойствах матриц Если вы работаете с матрицами, то готовьтесь к тому, чтобы думать про них как про более сложную версию чисел. А значит, вы будете писать с ними различного рода алгебраические выражения. Например, для какой-нибудь матрицы $A \in M_n(\mathbb{R})$ можно написать $A^3 + 2A - 3E$. И предположим вы хотите упростить это выражение как-нибудь, не зная как именно выглядит ваша матрица A . Единственное, что вам поможет в этом случае – зануляющий многочлен. Пусть, например, $f(x) = x^2 - 3$ зануляет A . Это значит, что $A^2 = 3E$. Тогда выражение выше можно упростить так

$$A^3 + 2A - 3E = 3A + 2A - 3E = 5A - 3E$$

Роль минимального многочлена заключается в том, что это «самый лучший» многочлен, который помнит как можно больше соотношений на матрицу A , чтобы можно было упрощать выражения. Более того, минимальный многочлен автоматически говорит, когда можно делить на выражение от матрицы, а когда нет. Например, на $A - E$ поделить можно, так как 1 не является корнем f , с другой стороны на матрицы $A \pm \sqrt{3}E$ делить нельзя.

Обратимость и минимальный многочлен Обратимость матрицы по определению равносильна тому, что в ее спектре нет нуля, а это то же самое, что у минимального многочлена свободный член отличен от нуля. В этом случае мы можем явно выразить обратную матрицу через исходную. Действительно, пусть $f_{\min} = a_0 + a_1x + \dots + a_mx^m$ для некоторой матрицы $A \in M_n(\mathbb{R})$. Тогда

$$a_0E + a_1A + \dots + a_mA^m = 0 \Rightarrow A(a_1E + \dots + a_mA^{m-1}) = -a_0E \Rightarrow A\left(-\frac{a_1}{a_0}E - \dots - \frac{a_m}{a_0}A^{m-1}\right) = E$$

То есть по определению

$$A^{-1} = -\frac{a_1}{a_0}E - \dots - \frac{a_m}{a_0}A^{m-1}$$

Обратите внимание, что данная формула работает при условии, что $a_0 \neq 0$. Эта процедура похожа на процедуру избавления от иррациональности в знаменателе дробей или избавления от мнимой части в знаменателе в комплексных дробях. Это неспроста, это в точности тот же самый метод.

Рассуждения выше позволяют добавить еще пару условий к шести эквивалентным утверждениям невырожденности (см. утверждение 3).

Утверждение 12. Пусть $A \in M_n(\mathbb{R})$ – произвольная квадратная матрица. Тогда следующие условия эквивалентны:

1. Система $Ax = 0$ имеет только нулевое решение.
2. Система $A^t y = 0$ имеет только нулевое решение.
3. Матрица A представляется в виде $A = U_1 \cdot \dots \cdot U_k$, где U_i – матрицы элементарных преобразований.
4. Матрица A обратима.
5. Матрица A обратима слева, т.е. существует L такая, что $LA = E$.
6. Матрица A обратима справа, т.е. существует R такая, что $AR = E$.
7. $0 \notin \text{спес}_{\mathbb{R}} A$.
8. $f_{\min A}(0) \neq 0$.

2.17 Матричные нормы

Здесь нас ждет пример первого абстрактного определения. Любое такое определение устроено одинаково, оно состоит из двух частей: первая часть содержит данные, а вторая аксиомы на них.

Нормы Будем через \mathbb{R}_+ обозначать множество неотрицательных вещественных чисел, т.е. $\mathbb{R}_+ = \{r \in \mathbb{R} \mid r \geq 0\}$.

Пусть задано отображение

$$M_{m \times n}(\mathbb{R}) \rightarrow \mathbb{R}_+$$

т.е. это правило, которое по матрице A выдает некоторое неотрицательное вещественное число, которое будет обозначаться $|A| \in \mathbb{R}_+$. Такое отображение называется нормой, если выполнены следующие аксиомы

1. Для любой матрицы $A \in M_{m \times n}(\mathbb{R})$, $|A| = 0$ тогда и только тогда, когда $A = 0$.
2. Для любой матрицы $A \in M_{m \times n}(\mathbb{R})$ и любого числа $\lambda \in \mathbb{R}$ выполнено $|\lambda A| = |\lambda| \cdot |A|$.²⁰
3. Для любых двух матриц $A, B \in M_{m \times n}(\mathbb{R})$ выполнено $|A + B| \leq |A| + |B|$.

Стоит отметить, что \mathbb{R}^n можно отождествить с матрицами $M_{n \times 1}(\mathbb{R})$. Потому определение выше дает понятие нормы на векторах из \mathbb{R}^n .

Субмультипликативность Пусть на квадратных матрицах $M_n(\mathbb{R})$ задана некоторая норма. Тогда она называется субмультипликативной, если выполнено следующее свойство: для любых матриц $A, B \in M_n(\mathbb{R})$ выполнено $|AB| \leq |A| \cdot |B|$.

Простые примеры

1. 1-норма на $M_n(\mathbb{R})$:

$$|A|_1 = \sum_{ij} |a_{ij}|, \quad A \in M_n(\mathbb{R})$$

2. ∞ -норма на $M_n(\mathbb{R})$:

$$|A|_\infty = \max_{ij} |a_{ij}|, \quad A \in M_n(\mathbb{R})$$

3. Норма Фробениуса или 2-норма на $M_n(\mathbb{R})$:

$$|A|_F = |A|_2 = \sqrt{\sum_{ij} |a_{ij}|^2}, \quad A \in M_n(\mathbb{R})$$

4. p -норма на $M_n(\mathbb{R})$:

$$|A|_p = \sqrt[p]{\sum_{ij} |a_{ij}|^p}$$

Выясните в качестве упражнения, какие из этих норм являются субмультипликативными.

Индукцированная (согласованная) норма Пусть $|\cdot|: \mathbb{R}^n \rightarrow \mathbb{R}_+$ – некоторая фиксированная норма. Определим индуцированную ей норму $\|-\|$ на $M_n(\mathbb{R})$ следующим образом:²¹

$$\|A\| = \sup_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{|Ax|}{|x|}, \quad A \in M_n(\mathbb{R})$$

Методом пристального взгляда мы замечаем, что отображение $\|-\|: M_n(\mathbb{R}) \rightarrow \mathbb{R}_+$ удовлетворяет первым трем аксиомам нормы, а значит действительно является матричной нормой. Более того, верно следующее.

Утверждение. Для любой нормы $|\cdot|: \mathbb{R}^n \rightarrow \mathbb{R}_+$ индуцированная ей норма $\|-\|: M_n(\mathbb{R}) \rightarrow \mathbb{R}_+$ является субмультипликативной.

²⁰Здесь $|\lambda|$ означает модуль числа, а $|A|$ – норма от матрицы.

²¹Можно определить норму на прямоугольных матрицах, но тогда надо иметь две нормы одну на \mathbb{R}^n , а другую на \mathbb{R}^m . В этом случае индуцированная норма зависит от двух норм, одна фигурирует в знаменателе, другая в числителе.

Доказательство. Из определения индуцированной нормы следует, что $|Ax|/|x| \leq \|A\|$ для любого ненулевого $x \in \mathbb{R}^n$. Ясно, что тогда для любого $x \in \mathbb{R}^n$, верно $|Ax| \leq \|A\| \cdot |x|$.

Пусть теперь $A, B \in M_n(\mathbb{R})$ и нам надо показать, что $\|AB\| \leq \|A\| \cdot \|B\|$. Рассмотрим произвольный $x \in \mathbb{R}^n$, тогда

$$|ABx| = |A(Bx)| \leq \|A\| \cdot |Bx| \leq \|A\| \cdot \|B\| \cdot |x|$$

Значит

$$\frac{|ABx|}{|x|} \leq \|A\| \cdot \|B\|$$

для любого ненулевого $x \in \mathbb{R}^n$. А значит, можно перейти к супремуму по таким x , и следовательно

$$\|AB\| = \sup_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{|ABx|}{|x|} \leq \|A\| \cdot \|B\|$$

□

Примеры индуцированных норм Индуцированные нормы хороши тем, что они субмультипликативны. Однако, обычно для них не существует явных формул для вычисления. Ниже мы приведем несколько случаев, когда такие формулы все же возможны. Все примеры будут даны без доказательств.

1. Пусть на \mathbb{R}^n дана 1-норма $|x| = \sum_i |x_i|$. Тогда индуцированная норма $\|-\|_1$ на матрицах $M_n(\mathbb{R})$ будет задаваться по формуле

$$\|A\|_1 = \max_j \sum_i |a_{ij}|$$

2. Пусть теперь на \mathbb{R}^n дана ∞ -норма $|x| = \max_i |x_i|$. Тогда индуцированная норма $\|-\|_\infty$ на матрицах $M_n(\mathbb{R})$ будет задаваться по формуле

$$\|A\|_\infty = \max_i \sum_j |a_{ij}|$$

3. И наконец, пусть на \mathbb{R}^n дана 2-норма $|x| = \sqrt{\sum_i |x_i|^2}$. Тогда индуцированная норма $\|-\|_2$ на матрицах $M_n(\mathbb{R})$ уже считается более хитрым способом. Пусть $A \in M_n(\mathbb{R})$. Если взять матрицу $A^t A$, то окажется, что ее спектр состоит целиком из неотрицательных вещественных чисел. Пусть σ_1 – максимальное такое число, тогда $\|A\|_2 = \sqrt{\sigma_1}$.²²

2.18 Обзор применения матричных норм

Для простоты изложения, я буду рассматривать лишь квадратные матрицы ниже. Хотя какие-то вопросы и можно формулировать и для прямоугольных матриц, это не сделает материал более интересным.

Сходимость Основная задача нормы – дать понятие о близости матриц друг к другу. А именно, если есть норма $|\cdot|: M_n(\mathbb{R}) \rightarrow \mathbb{R}_+$, то можно определить расстояние между матрицами $A, B \in M_n(\mathbb{R})$ следующим образом $\rho(A, B) = |A - B|$. А как только у нас есть понятие расстояния между объектами, мы можем ввести понятие предела и сходимости, а именно: пусть задана последовательность матриц $A_n \in M_n(\mathbb{R})$, тогда скажем, что она сходится к матрице $A \in M_n(\mathbb{R})$ и будем писать $A_n \rightarrow A$, $n \rightarrow \infty$ (или $\lim_n A_n = A$), если $\rho(A_n, A) \rightarrow 0$ как последовательность чисел при условии $n \rightarrow \infty$.

Эквивалентность норм Тут встает законный вопрос: у нас есть много различных норм на матрицах, а потому много расстояний, а значит получается огромное количество разных сходимостей (по одной на каждый вид нормы). Оказывается, что все возможные нормы на матрицах дают расстояния приводящие к одинаковому определению предела. Ключом к пониманию этого явления является определение эквивалентности норм. Пусть $|\cdot|$ и $|\cdot|'$ – две разные нормы на $M_n(\mathbb{R})$. Будем говорить, что они эквивалентны, если существуют две положительные константы $c_1, c_2 \in \mathbb{R}$ такие, что $c_1|A| \leq |A'| \leq c_2|A|$ для всех матриц $A \in M_n(\mathbb{R})$. Если две нормы эквивалентны, то несложно угледеть, что расстояние $\rho(A_n, A)$ в смысле нормы $|\cdot|$ стремится к нулю

²²К этому явлению надо относиться так: есть спектр – объект из мира алгебры и есть норма – объект из мира анализа. Оказывается, что между анализом и алгеброй есть мостик через спектр и индуцированную 2-норму. Это позволяет задачи про спектр изучать аналитическими методами и наоборот задачи про сходимости изучать алгебраическими.

тогда и только тогда, когда расстояние $\rho'(A_n, A)$ в смысле нормы $|\cdot|'$ стремится к нулю. А значит эквивалентные нормы дают одну и ту же сходимость. Второй ключевой факт – все матричные нормы между собой эквивалентны. Это не очень сложный результат и по сути связан с тем, что единичный куб в \mathbb{R}^n является компактным множеством.

Анализ для матриц Как только у нас есть понятие предела для матриц, мы можем с помощью него развивать анализ аналогичный анализу для обычных чисел. Например, можно определить хорошо известные гладкие функции от матриц. Скажем, пусть $A \in M_n(\mathbb{R})$, тогда можно сказать, что значит e^A , $\ln A$, $\sin A$ или $\cos A$. Конечно, $\ln A$ будет существовать не для всех матриц A , так же как и обычный логарифм существует только для положительных чисел. Знакомые тождества вроде $e^{\ln A} = A$ и $\ln e^A = A$ будут оставаться справедливыми. Свойства $e^{A+B} = e^A e^B$ будет верным, в случае если A и B коммутируют.

Одним из простейших подходов к определению таких функций – использование степенных рядов. Например, $e^x = \sum_{k \geq 0} x^k/k!$. Тогда можно определить $e^A = \sum_{k \geq 0} A^k/k!$. Доказательство свойств экспоненты тогда сводится к игре в раскрытие скобок со степенными рядами. И в этой игре нам важно, чтобы символы были перестановочны между собой, потому какие-то свойства могут нарушиться, если исходные матрицы не коммутируют.

Еще стоит отметить такой момент. Так как для любой матрицы A существует минимальный многочлен ее аннулирующий f_{min} , то, оказывается, что любую гладкую функцию от A можно приблизить многочленом. А именно, если φ – некоторая гладкая функция, то для любой матрицы A найдется такой многочлен f (зависящий от A) степени меньше, чем $\deg f_{min}$, что $\varphi(A) = f(A)$. Более того, существует общая алгоритмическая процедура по нахождению такого многочлена f . Эта процедура является эффективным способом вычисления гладких функций от матриц.

3 Перестановки

3.1 Отображения множеств

Пусть X, Y – некоторые множества, а $\varphi: X \rightarrow Y$ – отображение. Тогда φ называется *инъективным*, если оно «не склеивает точки», т.е. для любых $x, y \in X$ из условия $x \neq y$ следует $\varphi(x) \neq \varphi(y)$. Отображение φ называется *сюръективным*, если в любой элемент что-то переходит, т.е. для любого $y \in Y$ существует $x \in X$ такой, что $\varphi(x) = y$. Отображение φ называется *биективным*, если оно одновременно инъективно и сюръективно.²³

Свойства отображения можно подчеркивать видом стрелки. Например, инъективное отображение обычно обозначается $\varphi: X \hookrightarrow Y$, сюръективное – $\varphi: X \twoheadrightarrow Y$, а биективное – $\varphi: X \xrightarrow{\sim} Y$.

Для любого множества X отображение $\text{Id}: X \rightarrow X$ заданное по правилу $\text{Id}(x) = x$ называется *тождественным*. Пусть $\varphi: X \rightarrow Y$ – некоторое отображение. Тогда $\psi: Y \rightarrow X$ называется *левым обратным* (соответственно *правым обратным*) к φ , если $\psi\varphi = \text{Id}$ ($\varphi\psi = \text{Id}$).²⁴ Левых и правых обратных для φ может быть много. Однако, если есть оба обратных и ψ_1 – левый обратный, а ψ_2 – правый обратный, то они совпадают, так как $\psi_1 = \psi_1(\varphi\psi_2) = (\psi_1\varphi)\psi_2 = \psi_2$. А следовательно совпадают все левые обратные со всеми правыми и такой единственный элемент называют *обратным* и обозначают φ^{-1} , а φ называют *обратимым*. Легко проверить следующее.

Утверждение. Пусть $\varphi: X \rightarrow Y$ – некоторое отображение. Тогда

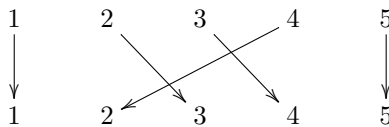
1. φ инъективно тогда и только тогда, когда φ обладает левым обратным.
2. φ сюръективно тогда и только тогда, когда φ обладает правым обратным.
3. φ биективно тогда и только тогда, когда φ обратимо.

3.2 Перестановки

Пусть $X_n = \{1, \dots, n\}$ – конечное множество из n занумерованных элементов.²⁵ Перестановкой называется биективное отображение $\sigma: X_n \rightarrow X_n$. Множество всех перестановок на n элементном множестве будем обозначать через S_n .

Как задавать перестановки Как только вам встречается новый объект, первый важный вопрос – а как подобные объекты вообще задавать? Для перестановок есть три способа:

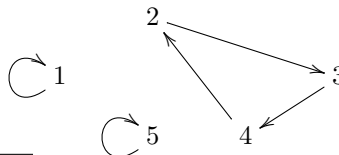
1. Задать стрелками соответствие на элементах



2. С помощью таблицы значений (графика). Здесь под каждым элементом пишется его образ:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 2 & 5 \end{pmatrix}$$

3. Графически в виде действия на элементах



²³В теории множеств, множества – это мешки с элементами, а отображения «сравнивают» эти мешки между собой. Биекция, между множествами говорит, что это по сути одно и то же множество, но по-разному заданное. Потому на биекцию между X и Y можно смотреть не как на отображение между разными множествами, а как на правило «переименовывающее» элементы на одном и том же множестве.

²⁴Легко проверить, что существование левого обратного никак не связано с существованием правого обратного и наоборот.

²⁵Формально говоря, это множество из n элементов и фиксированный линейный порядок на нем.

Все эти виды записи однозначно задают перестановку. Самым популярным методом в литературе является второй способ. В общем виде для перестановки $\sigma \in S_n$ табличная запись выглядит следующим образом:

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$

Заметим, что, если записать элементы $1, \dots, n$ в другом порядке, скажем, i_1, \dots, i_n , то перестановка σ запишется в виде²⁶

$$\begin{pmatrix} i_1 & i_2 & \dots & i_n \\ \sigma(i_1) & \sigma(i_2) & \dots & \sigma(i_n) \end{pmatrix}$$

Из однозначности табличной записи получаем следующее.

Утверждение. *Количество перестановок на n элементах есть $n!$, т.е. $|S_n| = n!$.*

3.3 Операция на перестановках

Так как перестановки являются отображениями, а на отображениях есть операция композиции, то и на перестановках появляется операция. Пусть $\sigma, \tau \in S_n$ – две произвольные перестановки, определим $\sigma\tau$ как композицию, т.е. $\sigma\tau(k) = \sigma(\tau(k))$. На языке диаграмм

$$\begin{array}{ccccc} X_n & \xrightarrow{\tau} & X_n & \xrightarrow{\sigma} & X_n \\ & \searrow \sigma\tau & & \nearrow & \end{array}$$

Важно Обратите внимание, что перестановки применяются к элементам справа налево. Это связано с тем, что они являются отображениями, а когда вы считаете композицию отображений, то вы сначала применяете к аргументу самое правое, потом следующее за ним и так далее.

Давайте посмотрим как выглядит произведение двух перестановок в табличной записи. Пусть даны перестановки

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} \text{ и } \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$$

То перестановки $\sigma\tau$ и $\tau\sigma$ имеют вид

$$\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix} \text{ и } \tau\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$$

Свойства умножения

- Если $\sigma, \tau, \rho \in S_n$ – произвольные перестановки, то как легко видеть по определению $(\sigma\tau)\rho = \sigma(\tau\rho)$. То есть в выражениях составленных из перестановок и произведений не важно в каком порядке расставлять скобки. Потому скобки обычно опускаются.
- Умножение перестановок не коммутативно, то есть вообще говоря $\sigma\tau \neq \tau\sigma$.²⁷
- Тожественное отображение Id является нейтральным элементом для умножения перестановок в том смысле, что верно $\text{Id}\sigma = \sigma\text{Id} = \sigma$ для любой перестановки σ . В табличной записи Id имеет вид

$$\begin{pmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{pmatrix}$$

- Обратное отображение к σ будем обозначать через σ^{-1} . Оно будет обратным элементом относительно операции в том смысле, что $\sigma\sigma^{-1} = \sigma^{-1}\sigma = \text{Id}$. В табличной записи обратное отображение можно записать так

$$\begin{pmatrix} \sigma(1) & \sigma(2) & \dots & \sigma(n) \\ 1 & 2 & \dots & n \end{pmatrix}$$

²⁶Заметим, что в этой записи можно произвольным образом перемешивать столбцы, это никак не изменит задаваемую перестановку.

²⁷Один пример мы уже видели, еще один будет в разделе «Циклические перестановки».

3.4 Переименование элементов

В нашем определении перестановка – это биекция на множестве X_n . Однако, элементы X_n имеют конкретные имена – это числа от 1 до n . А что произойдет, если мы сменим имена элементов? Как изменится табличная запись перестановки?

В начале надо понять, что значит переименование элементов. Во-первых, у нас есть запас старых имен $\{1, \dots, n\}$, во-вторых, у нас должен быть список новых имен, скажем, $\{“1”, \dots, “n”\}$ и, в-третьих, у нас должно быть соответствие, которое по старым именам строит новые, т.е. $\tau: \{1, \dots, n\} \rightarrow \{“1”, \dots, “n”\}$. Потому, если мысленно убрать кавычки, то на переименование можно смотреть как на перестановку $\tau: X_n \rightarrow X_n$.

Пусть теперь у нас есть перестановка $\sigma: X_n \rightarrow X_n$. Ее можно записать в табличном виде в старых и новых именах. Чтобы различать эти таблицы мы будем использовать обозначения $\sigma_{\text{стар}}$ и $\sigma_{\text{нов}}$ для них соответственно. Тогда мы можем записать связь между ними с помощью следующей диаграммы:

$$\begin{array}{ccc} \{1, \dots, n\} & \xrightarrow{\tau} & \{“1”, \dots, “n”\} \\ \sigma_{\text{стар}} \downarrow & & \downarrow \sigma_{\text{нов}} \\ \{1, \dots, n\} & \xrightarrow{\tau} & \{“1”, \dots, “n”\} \end{array}$$

Если вспомнить, что $\{“1”, \dots, “n”\} = \{\tau(1), \dots, \tau(n)\}$, то действие $\sigma_{\text{нов}}$ в новых именах устроено так: мы берем произвольный элемент с новым именем $\tau(k)$, находим его старое имя – k , на старом имени можем подействовать $\sigma_{\text{стар}}$, которое есть $\sigma(k)$, а теперь надо найти новое имя для образа, что есть $\tau(\sigma(k))$.

Подытожим, что $\sigma_{\text{нов}} = \tau \sigma_{\text{стар}} \tau^{-1}$. В табличной записи перестановки выглядят так

$$\sigma_{\text{стар}} = \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix} \quad \sigma_{\text{нов}} = \begin{pmatrix} \tau(1) & \tau(2) & \dots & \tau(n) \\ \tau(\sigma(1)) & \tau(\sigma(2)) & \dots & \tau(\sigma(n)) \end{pmatrix}$$

Хорошо еще иметь перед глазами следующую картинку:



Здесь в вершинах подписаны и старые и новые имена, а перестановка одна и та же.

3.5 Циклы

Пусть $\sigma \in S_n$ действует следующим образом: для некоторого множества i_1, \dots, i_k ($k \geq 2$) выполнено

$$\sigma(i_1) = i_2, \dots, \sigma(i_{k-1}) = i_k, \sigma(i_k) = i_1,$$

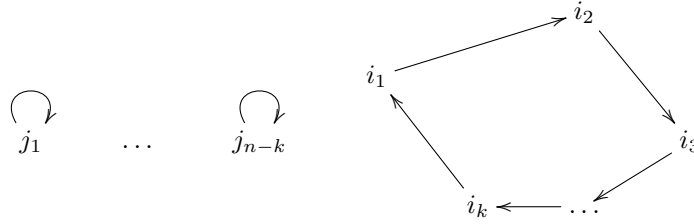
а все остальные элементы остаются на месте под действием σ . Тогда σ называется *циклом* длины k . Такая перестановка для краткости обозначается (i_1, \dots, i_k) . Заметим, что такая запись не единственная: например, можно сказать $\sigma = (i_2, \dots, i_k, i_1)$.²⁸ Стоит отметить, что если в определении выше выбрать $k = 1$, то перестановка обозначаемая (i_1) совпадает с тождественной перестановкой. Потому циклов длины 1 просто не существует. Однако, в некоторых случаях сама запись (i_1) является удобным обозначением для единообразия в формулах. Потому такие «циклы» принято называть тривиальными (подразумевая не цикл, а обозначение), а настоящие циклы – нетривиальными.

Таблицей цикл задается следующим образом

$$\begin{pmatrix} i_1 & \dots & i_{k-1} & i_k & j_1 & \dots & j_{n-k} \\ i_2 & \dots & i_k & i_1 & j_1 & \dots & j_{n-k} \end{pmatrix}$$

²⁸Как легко видеть, другой неоднозначности в записи цикла нет.

где $\{1, \dots, n\} = \{i_1, \dots, i_k\} \sqcup \{j_1, \dots, j_{n-k}\}$. Графически этот цикл выглядит так



Цикл длины 2 называется *транспозицией*, т.е. транспозиция (i, j) – это перестановка двух элементов i и j . Два цикла (i_1, \dots, i_k) и (j_1, \dots, j_m) называются *независимыми*, если множества $\{i_1, \dots, i_k\}$ и $\{j_1, \dots, j_m\}$ не пересекаются, т.е. множества действительно перемещаемых элементов не пересекаются. Заметим, что независимые циклы коммутируют друг с другом, а зависимые вообще говоря нет, как показывает следующий пример: $(1, 2)(2, 3) = (1, 2, 3)$, а $(2, 3)(1, 2) = (3, 2, 1)$.^{29, 30}

Утверждение 13. Пусть $\rho = (i_1, \dots, i_k) \in S_n$ – некоторый цикл длины k и $\tau \in S_n$ – произвольная перестановка, тогда

$$\tau(i_1, \dots, i_k)\tau^{-1} = (\tau(i_1), \dots, \tau(i_k))$$

Доказательство. Есть два способа понять это равенство. Первый – посмотреть на τ как на переименование элементов. Тогда справа написан цикл по элементам с новыми именами, а слева – правило переименования.

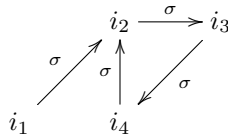
Второй способ – проверка в лоб. Надо проверить, что и левая и правая часть одинаково действуют на всех элементах вида $\tau(i)$. Возьмем элемент $\tau(i_1)$, тогда правая часть его переводит в $\tau(i_2)$. Посмотрим, что с ним делает левая часть. Вначале, мы переходим в i_1 , потом в i_2 , а потом в $\tau(i_2)$. Получили то же самое. Аналогично проверяется, что $\tau(i)$ остается на месте, если i не совпадает ни с одним из i_s . \square

Теперь мы готовы доказать структурный результат о перестановках.

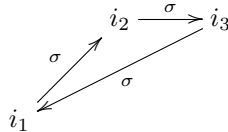
Утверждение 14. Пусть $\sigma \in S_n$ – произвольная перестановка. Тогда

1. Перестановку σ можно представить в виде $\sigma = \rho_1 \dots \rho_k$, где ρ_i – независимые циклы. Причем это представление единственное с точностью до перестановки сомножителей.
2. Пусть $\rho \in S_n$ – произвольный цикл длины k , тогда его можно представить в виде $\rho = \tau_1 \dots \tau_{k-1}$, где τ_i – транспозиции.³¹

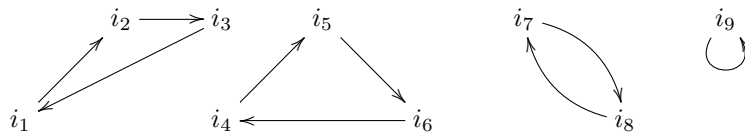
Доказательство. (1) Пусть $i_1 \in X_n$ – произвольный элемент. Подействуем на него σ , получим $i_2 = \sigma(i_1)$ и т.д. Так как X_n конечно, то мы в какой-то момент повторимся, например $i_5 = i_2$, как на рисунке ниже



На этой картинке видно, что $\sigma(i_1) = \sigma(i_4)$, но σ инъективно, потому $i_1 = i_4$. То есть правильная картинка следующая



Далее возьмем элемент, который не попал на этот цикл и повторим рассуждение для него. Так найдем другой цикл и т.д. В итоге картинка будет приблизительно такая



²⁹Проверьте это.

³⁰Зависимые циклы могут коммутировать, например $(1, 2)$ коммутирует с $(1, 2)$.

³¹Это представление уже не единственное.

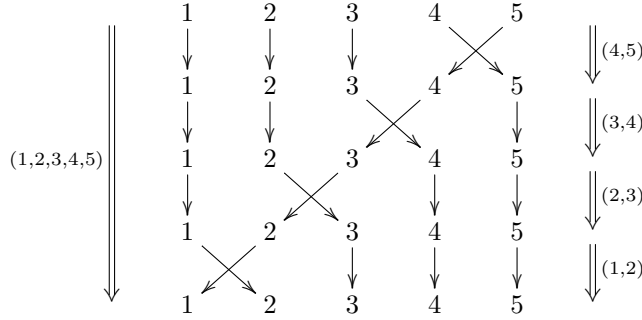
Значит перестановка выше раскладывается в циклы $\sigma = (i_1, i_2, i_3)(i_4, i_5, i_6)(i_7, i_8)$.³²

Единственность такого разложения следует из метода пристального взгляда на картинку и наше рассуждение. Если нужно формальное объяснение, то нужно делать так. Пусть $\sigma = \rho_1 \dots \rho_k$ и пусть $\rho_1 = (i_1, \dots, i_s)$. Подействуем σ на элемент i_1 . Так как циклы справа независимы, то только ρ_1 действует на i_1 и значит $\sigma(i_1) = \rho_1 \dots \rho_k(i_1) = i_2$. То есть i_2 однозначно определено. Продолжая в том же духе, мы видим, что все циклы однозначно определяются через σ .

(2) Пусть цикл σ действует по правилу как на картинке ниже



Чтобы получить цикл длины k нам необходимо применить $k - 1$ транспозицию. То есть в нашем примере надо применить 4. Сделаем это следующим образом



То есть в общем случае $(1, 2, \dots, k) = (1, 2)(2, 3) \dots (k - 2, k - 1)(k - 1, k)$. \square

Давайте поймем, почему представление во втором случае не единственное. Рассмотрим перестановку $(1, 2)(2, 3)$. Тогда

$$(1, 2)(2, 3) = (1, 2)(2, 3)(1, 2)^{-1}(1, 2) = (1, 3)(1, 2)$$

здесь в первом равенстве мы поделили и домножили на $(1, 2)$, а во втором воспользовались утверждением 13.

3.6 Знак перестановки

Рассмотрим произвольное отображение

$$\phi: S_n \rightarrow \{\pm 1\}$$

удовлетворяющее следующим двум свойствам:

1. $\phi(\sigma\tau) = \phi(\sigma)\phi(\tau)$ для любых $\sigma, \tau \in S_n$.³³
2. $\phi \neq 1$, т.е. ϕ не равно тождественно 1.

Заметим, что несложно найти отображение удовлетворяющее только первому свойству, например, $\phi(\sigma) = 1$ для любого σ , что не интересно. Наша основная задача доказать следующее.

Утверждение 15. *Существует единственное отображение $\phi: S_n \rightarrow \{\pm 1\}$ обладающее свойствами (1) и (2).*

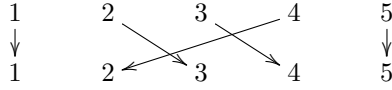
В этом случае такое отображение обозначается $\text{sgn}: S_n \rightarrow \{\pm 1\}$ и называется знаком. Значение $\text{sgn}(\sigma)$ называется знаком перестановки $\sigma \in S_n$. Перестановка называется четной, если знак 1 и нечетной, если -1 .

³²Цикл (i_9) здесь не используется, так как он совпадает с тождественной перестановкой Id , как и любой другой цикл длины 1.

³³Здесь справа стоит произведение чисел вида 1 или -1 .

Существование Обычно знак перестановки σ определяют в виде $(-1)^{d(\sigma)}$, где $d(\sigma)$ – некоторая целочисленная характеристика перестановки σ . Классическим определением является *число беспорядков*.³⁴

Пусть $\sigma \in S_n$ – некоторая перестановка и $i, j \in X_n$ – НЕупорядоченная пара различных элементов.³⁵ Тогда эта пара называется *инверсией*, если « σ меняет характер монотонности», т.е. $i < j$ влечет $\sigma(i) > \sigma(j)$, а $i > j$ влечет $\sigma(i) < \sigma(j)$. Если использовать запись перестановки в виде



то инверсия соответствует пересечению стрелок. Определим число $d_{ij}(\sigma) = 1$, если пара i, j образует инверсию и 0, если не образуют. Тогда число всех инверсий для всевозможных пар это $d(\sigma) = \sum_{i < j} d_{ij}(\sigma)$. Определим отображение $\text{sgn}: S_n \rightarrow \{\pm 1\}$ по правилу $\text{sgn}(\sigma) = (-1)^{d(\sigma)}$. Для доказательства существования надо проверить, что sgn обладает указанными свойствами (1) и (2), то есть надо доказать следующее.

Утверждение. Пусть $\sigma, \tau \in S_n$ – произвольные перестановки, тогда

$$\text{sgn}(\sigma\tau) = \text{sgn}(\sigma)\text{sgn}(\tau) \quad \text{и} \quad \text{sgn}(1, 2) = -1$$

Доказательство. Второе утверждение очевидно, в перестановке $(1, 2)$ всего одна инверсия, а значит $\text{sgn}(1, 2) = -1$.

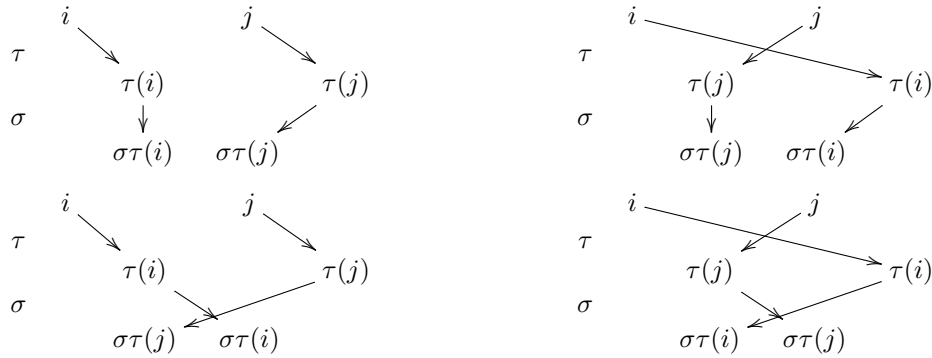
Для доказательства первого надо показать, что

$$d(\sigma) + d(\tau) = d(\sigma\tau) \pmod{2}$$

Давайте фиксируем пару i, j и докажем следующее равенство

$$d_{ij}(\tau) + d_{\tau(i)\tau(j)}(\sigma) = d_{ij}(\sigma\tau) \pmod{2}$$

Возможны следующие 4 случая:



Занесем результаты в таблицу

$d_{ij}(\tau)$	0	1	0	1
$d_{\tau(i)\tau(j)}(\sigma)$	0	0	1	1
$d_{ij} + d_{\tau(i)\tau(j)}(\sigma)$	0	1	1	2
$d_{ij}(\sigma\tau)$	0	1	1	0

Что доказывает равенство

$$d_{ij}(\tau) + d_{\tau(i)\tau(j)}(\sigma) = d_{ij}(\sigma\tau) \pmod{2}$$

Теперь сложим его для всех пар $i < j$. Получим

$$\sum_{i < j} d_{ij}(\tau) + \sum_{i < j} d_{\tau(i)\tau(j)}(\sigma) = \sum_{i < j} d_{ij}(\sigma\tau) \pmod{2}$$

³⁴Оно же *число инверсий*.

³⁵То есть пару i, j и j, i мы считаем одной и той же.

Откуда

$$d(\tau) + \sum_{i < j} d_{\tau(i)\tau(j)}(\sigma) = d(\sigma\tau) \pmod{2}$$

Так как $\tau: X_n \rightarrow X_n$ – биекция, то если (i, j) пробегает все разные пары, то и $(\tau(i), \tau(j))$ пробегает все разные пары. Значит оставшаяся сумма равна $d(\sigma)$, что завершает доказательство. \square

Единственность

Утверждение. Пусть $\phi: S_n \rightarrow \{\pm 1\}$ обладает свойством (1). Тогда

1. $\phi(\text{Id}) = 1$
2. $\phi(\sigma^{-1}) = \phi(\sigma)^{-1}$
3. Значение ϕ совпадает на всех транспозициях.

Доказательство. (1) Рассмотрим цепочку равенств

$$\phi(\text{Id}) = \phi(\text{Id}^2) = \phi(\text{Id})\phi(\text{Id})$$

Так как это числовое равенство (все числа ± 1), то можно сократить на $\phi(\text{Id})$ и получим требуемое.

(2) Рассмотрим цепочку равенств

$$1 = \phi(\text{Id}) = \phi(\sigma\sigma^{-1}) = \phi(\sigma)\phi(\sigma^{-1})$$

Значит число $\phi(\sigma^{-1})$ является обратным к $\phi(\sigma)$.³⁶

(3) Заметим, что для любых различных $i, j \in X_n$ у нас обязательно существует перестановка $\tau \in S_n$ такая, что $\tau(1) = i$ и $\tau(2) = j$.³⁷ Тогда по утверждению 13 получаем $(i, j) = \tau(1, 2)\tau^{-1}$. А значит

$$\phi(i, j) = \phi(\tau(1, 2)\tau^{-1}) = \phi(\tau)\phi(1, 2)\phi(\tau^{-1}) = \phi(1, 2)\phi(\tau)\phi(\tau^{-1}) = \phi(1, 2)$$

В предпоследнем равенстве мы воспользовались тем, что числа можно переставлять. Следовательно, значение на любой транспозиции равно значению на фиксированной транспозиции $(1, 2)$. То есть значение на всех транспозициях одинаковое. \square

Теперь давайте докажем единственность. Пусть у нас существуют два таких отображения $\phi, \psi: S_n \rightarrow \{\pm 1\}$ удовлетворяющие свойствам (1) и (2). Давайте покажем, что $\phi(\sigma) = \psi(\sigma)$ для любой $\sigma \in S_n$. Из утверждения 14 следует, что σ представляется в виде $\sigma = \tau_1 \dots \tau_r$, где τ_i – транспозиции.

Значение ϕ одно и то же на всех транспозициях: либо 1, либо -1 . Предположим, что значение равно 1. Тогда $\phi(\sigma) = \phi(\tau_1 \dots \tau_r) = \phi(\tau_1) \dots \phi(\tau_r) = 1$ для всех $\sigma \in S_n$, что противоречит свойству (2). А значит $\phi(\tau) = -1$ для любой транспозиции τ . Аналогично, $\psi(\tau) = -1$ для любой транспозиции τ . А следовательно

$$\phi(\sigma) = \phi(\tau_1 \dots \tau_r) = \phi(\tau_1) \dots \phi(\tau_r) = (-1)^r = \psi(\tau_1) \dots \psi(\tau_r) = \psi(\tau_1 \dots \tau_r) = \psi(\sigma)$$

То есть, на самом деле, все определяется значением на транспозиции.

³⁶Так как все наши числа ± 1 , то можно было бы сказать $\phi(\sigma^{-1}) = \phi(\sigma)$. Но в указанной форме равенство лучше запоминается и встретится вам еще не раз.

³⁷Я оставляю это как упражнение.