Лекции по Линейной Алгебре

Дима Трушин

2023 — 2024

Содержание

1 Системы линейных уравнений

1.1 Системы линейных уравнений и связанная с ними терминология

Наша задачи 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):

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

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. Потому мы будем работать с числами из $\mathbb{R}$.

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

1.2 Матрицы связанные со СЛУ

Для каждой СЛУ введем следующие обозначения:

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

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 помногу 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.

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.

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

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 типа[1]:

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

$$\text{II тип: } \left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1n} & b_1 \\ a_{i1} & \ldots & a_{in} & b_i \\ a_{j1} & \ldots & a_{jn} & b_j \\ a_{m1} & \ldots & a_{mn} & b_m \end{array}\right) \mapsto \left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1n} & b_1 \\ a_{j1} & \ldots & a_{jn} & b_j \\ a_{i1} & \ldots & a_{in} & b_i \\ a_{m1} & \ldots & a_{mn} & b_m \end{array}\right)$$

$$\text{III тип: } \left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1n} & b_1 \\ a_{i1} & \ldots & a_{in} & b_i \\ a_{m1} & \ldots & a_{mn} & b_m \end{array}\right) \mapsto \left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1n} & b_1 \\ \lambda a_{i1} & \ldots & \lambda a_{in} & \lambda b_i \\ a_{m1} & \ldots & a_{mn} & b_m \end{array}\right) \quad \lambda \neq 0$$

Поясним словами, что делают преобразования:

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

[1] Стоит 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.[2] Разберем типичный ход алгоритма Гаусса на примере 3 уравнений и 4 неизвестных.[3]

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.

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

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

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

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & a_{22} & a_{23} & a_{24} & b_2 \\ 0 & 0 & a_{33} & a_{34} & b_3 \end{array}\right)$$

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 случаи[4]

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

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.

[2] Данный 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.

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

[4] Это 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.

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

$$\left(\begin{array}{cccc|c} 1 & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & 1 & a_{23} & a_{24} & b_2 \\ 0 & 0 & 1 & a_{34} & b_3 \end{array}\right) \quad \text{2-я строка } - a_{23} \cdot \text{3-я строка}$$

$$\left(\begin{array}{cccc|c} 1 & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & 1 & 0 & a_{24} & b_2 \\ 0 & 0 & 1 & a_{34} & b_3 \end{array}\right) \quad \text{1-я строка } - a_{13} \cdot \text{3-я строка}$$

$$\left(\begin{array}{cccc|c} 1 & a_{12} & 0 & a_{14} & b_1 \\ 0 & 1 & 0 & a_{24} & b_2 \\ 0 & 0 & 1 & a_{34} & b_3 \end{array}\right) \quad \text{1-я строка } - a_{12} \cdot \text{2-я строка}$$

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & a_{14} & b_1 \\ 0 & 1 & 0 & a_{24} & b_2 \\ 0 & 0 & 1 & a_{34} & b_3 \end{array}\right)$$

В специальных случаях приведенных выше, получим

$$\left(\begin{array}{cccc|c} 1 & 0 & a_{13} & 0 & b_1 \\ 0 & 1 & a_{23} & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \end{array}\right) \quad \left(\begin{array}{cccc|c} 1 & a_{12} & 0 & 0 & b_1 \\ 0 & 0 & 1 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \end{array}\right) \quad \left(\begin{array}{cccc|c} 1 & 0 & a_{13} & a_{14} & 0 \\ 0 & 1 & a_{23} & a_{24} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right) \quad \left(\begin{array}{cccc|c} 1 & 0 & a_{13} & a_{14} & b_1 \\ 0 & 1 & a_{23} & a_{24} & b_2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

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.

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

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

Получение решений В системе ниже, выберем переменную x_4 как параметр

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & a_{14} & b_1 \\ 0 & 1 & 0 & a_{24} & b_2 \\ 0 & 0 & 1 & a_{34} & b_3 \end{array}\right)$$

Тогда решения имеют вид[5]

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

[5]Операция 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:

$$\left(\begin{array}{cccc|c} 1 & 0 & a_{13} & 0 & b_1 \\ 0 & 1 & a_{23} & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \end{array}\right) \quad \text{Решения:} \quad \begin{pmatrix} x_1 \\ x_2 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} - x_3 \begin{pmatrix} a_{13} \\ a_{23} \\ 0 \end{pmatrix}$$

$$\left(\begin{array}{cccc|c} 1 & a_{12} & 0 & 0 & b_1 \\ 0 & 0 & 1 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \end{array}\right) \quad \text{Решения:} \quad \begin{pmatrix} x_1 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} - x_2 \begin{pmatrix} a_{12} \\ 0 \\ 0 \end{pmatrix}$$

$$\left(\begin{array}{cccc|c} 1 & 0 & a_{13} & a_{14} & 0 \\ 0 & 1 & a_{23} & a_{24} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right) \quad \text{Решения:} \quad \text{Нет решений, т.к. последнее уравнение } 0 = 1$$

$$\left(\begin{array}{cccc|c} 1 & 0 & a_{13} & a_{14} & b_1 \\ 0 & 1 & a_{23} & a_{24} & b_2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \quad \text{Решения:} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} - x_3 \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} - x_4 \begin{pmatrix} a_{14} \\ a_{24} \end{pmatrix}$$

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 $\mathrm{M}_{m\,n}(\mathbb{R})$. Множество квадратных матриц размера n будем обозначать $\mathrm{M}_n(\mathbb{R})$. 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.[6]

2.2 Операции над матрицами

Сложение Пусть $A, B \in \mathrm{M}_{m\,n}(\mathbb{R})$. 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.[7]

Умножение на скаляр Если $\lambda \in \mathbb{R}$ и $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, 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 $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ и $B \in \mathrm{M}_{n\,k}(\mathbb{R})$, то произведение $AB \in \mathrm{M}_{m\,k}(\mathbb{R})$ 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.

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

[6] Важно, 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.

[7] Можно 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.

$$\boxed{a_{i1} \quad \dots \quad a_{in}} \cdot \boxed{\begin{matrix} b_{1j} \\ \\ \vdots \\ \\ b_{nj} \end{matrix}} \mapsto \begin{matrix} \boxed{a_{i1} \quad \dots \quad a_{in}} \\ \boxed{b_{1j} \quad \dots \quad b_{nj}} \end{matrix} \mapsto$$

$$\mapsto \boxed{\begin{matrix} a_{i1} \\ \cdot \\ b_{1j} \end{matrix}} + \dots + \boxed{\begin{matrix} a_{in} \\ \cdot \\ b_{nj} \end{matrix}} = a_{i1}b_{1j} + \dots + a_{in}b_{nj}$$

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

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

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

1. Для любых матриц $A, B \in \mathrm{M}_n(\mathbb{R})$ верно $\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.
2. Для любой матрицы $A \in \mathrm{M}_n(\mathbb{R})$ и $\lambda \in \mathbb{R}$ выполнено $\operatorname{tr}(\lambda A) = \lambda \operatorname{tr}(A)$.
3. Для любых матриц $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ и $B \in \mathrm{M}_{n\,m}(\mathbb{R})$ выполнено $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

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:[8]

[8] Все 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 $A,B,C\in \mathrm{M}_{m\,n}(\mathbb{R})$

2. **Существование нейтрального элемента для сложения** Существует единственная матрица 0 обладающая следующим свойством $A+0=0+A=A$ для всех $A\in \mathrm{M}_{m\,n}(\mathbb{R})$. Такая матрица целиком заполнена нулями.

3. **Коммутативность сложения** $A+B=B+A$ для любых $A,B\in \mathrm{M}_{m\,n}(\mathbb{R})$.

4. **Наличие обратного по сложению** Для любой матрицы $A\in \mathrm{M}_{m\,n}(\mathbb{R})$ 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 $A\in \mathrm{M}_{m\,n}(\mathbb{R})$, $B\in \mathrm{M}_{n\,k}(\mathbb{R})$ и $C\in \mathrm{M}_{k\,t}(\mathbb{R})$ верно $(AB)C=A(BC)$.

6. **Существование нейтрального элемента для умножения** Для каждого k существует единственная матрица $E\in \mathrm{M}_k(\mathbb{R})$ такая, что для любой $A\in \mathrm{M}_{m\,n}(\mathbb{R})$ 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 $A,B\in \mathrm{M}_{m\,n}(\mathbb{R})$ и $C\in \mathrm{M}_{n\,k}(\mathbb{R})$ верно $(A+B)C=AC+BC$. Аналогично, для любых $A\in \mathrm{M}_{m\,n}(\mathbb{R})$ и $B,C\in \mathrm{M}_{n\,k}(\mathbb{R})$ верно $A(B+C)=AB+AC$.

8. **Умножение на числа ассоциативно** Для любых $\lambda,\mu\in\mathbb{R}$ и любой матрицы $A\in \mathrm{M}_{m\,n}(\mathbb{R})$ верно $\lambda(\mu A)=(\lambda\mu)A$. Аналогично для любого $\lambda\in\mathbb{R}$ и любых $A\in \mathrm{M}_{m\,n}(\mathbb{R})$ и $B\in \mathrm{M}_{n\,k}(\mathbb{R})$ 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 $A\in \mathrm{M}_{m\,n}(\mathbb{R})$ верно $(\lambda+\mu)A=\lambda A+\mu A$. Аналогично, для любого $\lambda\in\mathbb{R}$ и $A,B\in \mathrm{M}_{m\,n}(\mathbb{R})$ верно $\lambda(A+B)=\lambda A+\lambda B$.

10. **Умножение на скаляр нетривиально** Если $1\in\mathbb{R}$, то для любой матрицы $A\in \mathrm{M}_{m\,n}(\mathbb{R})$ верно $1A=A$.

11. **Умножение на скаляр согласовано с умножением матриц** Для любого $\lambda\in\mathbb{R}$ и любых $A\in \mathrm{M}_{m\,n}(\mathbb{R})$ и $B\in \mathrm{M}_{n\,k}(\mathbb{R})$ верно $\lambda(AB)=(\lambda A)B=A(\lambda B)$.

12. **Транспонирование согласовано с суммой** Для любых матриц $A,B\in \mathrm{M}_{m\,n}(\mathbb{R})$ верно $(A+B)^t=A^t+B^t$.

13. **Транспонирование согласовано с умножением на скаляр** Для любой матрицы $A\in \mathrm{M}_{m\,n}(\mathbb{R})$ и любого $\lambda\in\mathbb{R}$ верно $(\lambda A)^t=\lambda A^t$.

14. **Транспонирование согласовано с умножением** Для любых матриц $A,B\in \mathrm{M}_{m\,n}(\mathbb{R})$ верно $(AB)^t=B^tA^t$.

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 $A\in \mathrm{M}_{m\,k}(\mathbb{R})$ i-я строка полностью состоит из нулей и нам дана матрица $B\in \mathrm{M}_{k\,n}(\mathbb{R})$. Тогда в произведении AB i-я строка тоже будет нулевая. 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 смножителя (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 $B \in \mathrm{M}_{k\,n}(\mathbb{R})$ i-ый столбец полностью состоит из нулей и нам дана матрица $A \in \mathrm{M}_{m\,k}(\mathbb{R})$. 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 $z' - z \in E_0$. □

2.6 Дефекты матричных операций

Матрицы как новые числа Рассмотрим множество квадратных матриц с введенными выше операциями: $(\mathrm{M}_n(\mathbb{R}), +, -, \cdot, {}^t)$. 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, $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, $B \in \mathrm{M}_{n\,k}(\mathbb{R})$ и при этом $m \neq k$. В этом случае $AB \in \mathrm{M}_{m\,k}(\mathbb{R})$, однако, произведение BA просто не определено из-за отсутствия условия согласованности матриц. Если мы рассмотрим матрицы, которые можно перемножить в обоих порядках, то есть $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, $B \in \mathrm{M}_{n\,m}(\mathbb{R})$, то результаты будут $AB \in \mathrm{M}_m(\mathbb{R})$ и $BA \in \mathrm{M}_n(\mathbb{R})$. 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 $AB = 0$.[9] 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,

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

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.

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 $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, будем говорить, что $B \in \mathrm{M}_{n\,m}(\mathbb{R})$ является левым обратным к A, если $BA = E \in \mathrm{M}_n(\mathbb{R})$. Аналогично, $B \in \mathrm{M}_{n\,m}(\mathbb{R})$ – правый обратный к A, если $AB = E \in \mathrm{M}_m(\mathbb{R})$. Надо иметь в виду, что вообще говоря левые и правые обратные между собой никак не связаны и их может быть много. Например, пусть $A = (1, 0) \in \mathrm{M}_{1\,2}(\mathbb{R})$. 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 $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ называется обратимой, если к ней существует левый и правый обратный.[10]

Утверждение 1. *Пусть матрица $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ обратима. Тогда*

1. *Левый обратный и правый обратный единственны и совпадают друг с другом.*

[9] На 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 $\mathbb{R}^n$.

[10] Ниже мы покажем, что из двусторонней обратимости следует, что матрица A обязана быть квадратной.

2. *Матрица A обязательно квадратная, то есть $m = n$.*

Доказательство. (1) Пусть $L \in \mathrm{M}_{n\,m}(\mathbb{R})$ – произвольный левый обратный к A, а $R \in \mathrm{M}_{n\,m}(\mathbb{R})$ – 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 $B \in \mathrm{M}_{n\,m}(\mathbb{R})$ – двусторонний обратный к A, то есть $AB = E_m \in \mathrm{M}_m(\mathbb{R})$ и $BA = E_n \in \mathrm{M}_n(\mathbb{R})$.[11] Тогда по свойствам следа получим

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

□

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 $A, B \in \mathrm{M}_n(\mathbb{R})$ – 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 $(A^t)^{-1} = (A^{-1})^t$. □

Обратимые преобразования над СЛУ Пусть у нас есть $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ и $b \in \mathbb{R}^m$, которые задают систему линейных уравнений $Ax = b$, где $x \in \mathbb{R}^n$. Возьмем произвольную обратимую матрицу $C \in \mathrm{M}_m(\mathbb{R})$. 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.

[11] Тут E_n означает единичную матрицу размера n.

2.8 Матрицы элементарных преобразований

Тип I Пусть $S_{ij}(\lambda) \in \mathrm{M}_n(\mathbb{R})$ – 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:

$$\begin{array}{c} \\ i \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} & & j & \end{array} \\ \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \lambda & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix} \end{array}$$

Тогда прямая проверка показывает, умножение $A \in \mathrm{M}_{n\,m}(\mathbb{R})$ на $S_{ij}(\lambda)$ слева прибавляет j строку умноженную на λ к i строке матрицы A, а умножение $B \in \mathrm{M}_{m\,n}(\mathbb{R})$ 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 $T_{ij} \in \mathrm{M}_n(\mathbb{R})$ – матрица, полученная из единичной перестановкой i и j ($i \neq j$) столбцов (или что то же самое – строк). Эта матрица имеет следующий вид

$$\begin{array}{c} \\ i \\ \\ j \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} & & i & & j \end{array} \\ \begin{pmatrix} 1 & & & & \\ & 0 & & 1 & \\ & & \ddots & & \\ & 1 & & 0 & \\ & & & & 1 \end{pmatrix} \end{array}$$

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

Тип III Пусть $D_i(\lambda) \in \mathrm{M}_n(\mathbb{R})$ – 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

$$\begin{array}{c} \\ i \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} & & i & & \end{array} \\ \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} \end{array}$$

Тогда прямая проверка показывает, умножение $A \in \mathrm{M}_{n\,m}(\mathbb{R})$ на $D_i(\lambda)$ слева умножает i строку A на λ, а умножение $B \in \mathrm{M}_{m\,n}(\mathbb{R})$ 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 $A \in \mathrm{M}_n(\mathbb{R})$ – 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 $A = U_1 \cdot \ldots \cdot U_k$, 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)⇒(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)⇒(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)⇒(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)⇒(5) и (4)⇒(6) очевидно, так это переход от более сильного условия к более слабому.

(5)⇒(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)⇒(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. □

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

Определение 4. Пусть $A \in \mathrm{M}_n(\mathbb{R})$ – произвольная квадратная матрица. Будем говорить, что A невырождена[12], если удовлетворяет любому из перечисленных в предыдущем утверждении условий.

Делители нуля Пусть $A \in \mathrm{M}_n(\mathbb{R})$ – некоторая ненулевая матрица и пусть $B \in \mathrm{M}_{n\,m}(\mathbb{R})$. 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 $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ 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 $C \in \mathrm{M}_n(\mathbb{R})$ – обратимая.

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.

[12]Классически 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.

Насыщенность обратимых Я хочу продемонстрировать еще одно полезное следствие из Утверждения 3. Предположим у нас есть две матрицы $A, B \in \mathrm{M}_n(\mathbb{R})$. 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 $D \in \mathrm{M}_n(\mathbb{R})$ 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:

$$\begin{array}{c} \\ m \\ n \end{array}\!\!\begin{array}{c} \begin{array}{cc} k & s \end{array} \\ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ k \\ s \end{array}\!\!\begin{array}{c} \begin{array}{cc} u & v \end{array} \\ \begin{pmatrix} X & Y \\ W & Z \end{pmatrix} \end{array}$$

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

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

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 $\mathrm{M}_{n+1}(\mathbb{R})$ вида

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

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 $x_1, \ldots, x_m \in \mathbb{R}^n$ и $y_1, \ldots, y_m \in \mathbb{R}^n$ – столбцы. Составим из этих столбцов матрицы $X = (x_1|\ldots|x_m)$ и $Y = (y_1|\ldots|y_m)$.[13] Заметим, что $X, Y \in \mathrm{M}_{n\,m}(\mathbb{R})$. Тогда

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

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

2.11 Блочные элементарные преобразования

Преобразования первого типа Пусть у нас дана матрица

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

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

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

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

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

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

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

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

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

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.

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 $R \in \mathrm{M}_m(\mathbb{R})$ – 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 $T \in \mathrm{M}_k(\mathbb{R})$, 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 $Ax_1 = b_1$, …, $Ax_k = b_k$, где $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, $b_i \in \mathbb{R}^m$ и $x_i \in \mathbb{R}^n$. Определим матрицы $X = (x_1|\ldots|x_k) \in \mathrm{M}_{n\,k}(\mathbb{R})$ и $B = (b_1|\ldots|b_k) \in \mathrm{M}_{m\,k}(\mathbb{R})$ составленные из столбцов x_i и b_i соответственно. Тогда по формулам блочного умножения матриц

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

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

Решение матричных уравнений

Дано $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, $B \in \mathrm{M}_{m\,k}(\mathbb{R})$.

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

Алгоритм

1. Составить расширенную матрицу $(A|B)$. Например, если $A \in \mathrm{M}_{3\,3}(\mathbb{R})$, а $B \in \mathrm{M}_{3\,2}(\mathbb{R})$, 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 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 $A^tX^t = B^t$. Тогда его можно решать относительно X^t, а потом транспонировать ответ.

Нахождение обратной матрицы методом Гаусса

Дано Матрица $A \in \mathrm{M}_n(\mathbb{R})$.

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:
   (a) После приведения получили матрицу $(E|B)$. Тогда A обратима и $A^{-1} = B$.
   (b) 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 $AX = E$. Теперь мы видим, что у системы $X = B$ единственное решение B, а это значит, что и у системы $AX = E$ 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 $C \in \mathrm{M}_n(\mathbb{R})$ выполнено $D = CA$. Так как у матрицы D есть свободная позиция и она квадратная [14], то обязательно найдется нулевая строка. А раз так, то матрица D не может быть обратима справа. Действительно, тогда в произведении DR для любой $R \in \mathrm{M}_n(\mathbb{R})$ 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.

2.13 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 $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ и $B \in \mathrm{M}_{k\,n}(\mathbb{R})$ – матрицы в ступенчатом виде и системы $Ax = 0$ и $Bx = 0$ эквивалентны.*[15] *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 $1 \leqslant i \leqslant n$ введем следующее свойство

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

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 однозначно по свободным и мы получим какое-то решение x. По построению $x_i \neq 0$. И надо лишь проверить, что $x_{i+1} = \ldots = x_n = 0$. 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. □

Утверждение 6. *Пусть $S_1 \in \mathrm{M}_{m\,n}(\mathbb{R})$ и $S_2 \in \mathrm{M}_{k\,n}(\mathbb{R})$ – произвольные матрицы в улучшенном ступенчатом виде. Если $S_1x = 0$ эквивалентно $S_2x = 0$, то после удаления нулевых строк матрицы S_1 и S_2 совпадут.*

Доказательство. Из предыдущего утверждения мы уже знаем, что у систем $S_1x = 0$ и $S_2x = 0$ 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:

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

[14] Вот то место, где мы пользуемся квадратностью матрицы.
[15] То есть имеют одинаковое множество решений.

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

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

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 вида[16]

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

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 одинаковые. □

Из этого утверждения следует, что матрица улучшенного ступенчатого вида для любой матрицы $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ 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 $A, B \in \mathrm{M}_{m\,n}(\mathbb{R})$ и пусть $E_A, E_B \subseteq \mathbb{R}^n$ – множества решений систем $Ax = 0$ и $Bx = 0$, соответственно. Тогда следующее эквивалентно:*

1. *$E_A = E_B$, т.е. системы эквивалентны.*
2. *A приводится к B элементарными преобразованиями строк.*
3. *Существует обратимая $C \in \mathrm{M}_m(\mathbb{R})$ такая, что $B = CA$.*
4. *Матрица улучшенного ступенчатого вида для A совпадает с матрицей улучшенного ступенчатого вида для B.*

Доказательство. Мы все это уже доказали по сути, потому напомним, что откуда следует. (2)⇒(1) Так как элементарные преобразования меняют систему на эквивалентную. (1)⇒(4) Предыдущее утверждение. (4)⇒(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)⇔(3) следует из Утверждения 3 о том, что матрица обратима тогда и только тогда, когда она раскладывается в произведение элементарных. □

Смысл этого утверждения в следующем. Возьмем множество всех однородных систем фиксированного размера, которое описывается матрицами $\mathrm{M}_{m\,n}(\mathbb{R})$. Тогда на этом множестве есть отношение эквивалентности: системы эквиваленты 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).

[16] Тут красным обозначены значения главных переменных, а синим – свободных.

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.

2.14 Полиномиальное исчисление от матриц

Обозначим множество всех многочленов с вещественными коэффициентами через $\mathbb{R}[x]$. Формально это значит: $\mathbb{R}[x] = \{a_0 + a_1 x + \ldots + a_n x^n \mid n \in \mathbb{Z}_+,\, a_i \in \mathbb{R}\}$. Аналогично можно обозначать многочлены с рациональными, целыми, комплексными и т.д. коэффициентами.

Подстановка матриц в многочлены Пусть $p(x) = a_0 + a_1 x + \ldots a_n x^n$ – многочлен с вещественными коэффициентами, а $A \in \mathrm{M}_n(\mathbb{R})$. Тогда можно определить $f(A) = a_0 E + a_1 A^1 + \ldots + a_n A^n \in \mathrm{M}_n(\mathbb{R})$. Если определить $A^0 = E$, то формула становится более единообразной $f(A) = a_0 A^0 + a_1 A^1 + \ldots + a_n A^n$. 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 $A \in \mathrm{M}_n(\mathbb{R})$ 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* $C \in \mathrm{M}_n(\mathbb{R})$
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

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

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

Обнуляющий многочлен

Утверждение 8. *Пусть $A \in \mathrm{M}_n(\mathbb{R})$, тогда:*

1. *Существует многочлен $f \in \mathbb{R}[x]$ не равный тождественно нулю степени не больше n^2 такой, что $f(A) = 0$.*
2. *Если для какого-то многочлена $g \in \mathbb{R}[x]$ имеем $g(A) = 0$, а для $\lambda \in \mathbb{R}$ имеем $g(\lambda) \neq 0$, то $A - \lambda E$ является обратимой матрицей.*

Доказательство. (1) Давайте искать многочлен f с неопределенными коэффициентами в виде $f = a_0 + a_1 x + \ldots + a_{n^2} x^{n^2}$. Надо чтобы было выполнено равенство $a_0 E + a_1 A + \ldots + a_{n^2} A^{n^2} = 0$. Последнее равенство означает равенство матрицы слева нулевой матрице справа. Это условие задается равенством всех n^2 ячеек матриц: $(a_0 E + a_1 A + \ldots + a_{n^2} A^{n^2})_{ij} = 0$ для всех i, j. Каждое из этих условий является линейным уравнением вида $a_0 (E)_{ij} + a_1 (A)_{ij} + \ldots + a_{n^2} (A^{n^2})_{ij} = 0$. 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.

(2) 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 $A - \lambda E$. □

На самом деле можно показать, что найдется многочлен степени не больше n, зануляющий нашу матрицу. Однако, мы пока не в состоянии этого сделать.

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 $A \in \mathrm{M}_n(\mathbb{R})$ определим вещественный спектр матрицы A следующим образом:[17]

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

Утверждение 10. *Пусть $A \in \mathrm{M}_n(\mathbb{R})$ и пусть $f \in \mathbb{R}[x]$ такой, что $f(A) = 0$. Тогда $|\operatorname{spec}_{\mathbb{R}} A| \leqslant \deg f$. В частности спектр всегда конечен.*

Доказательство. Покажем, что любой элемент спектра является корнем f. Для этого достаточно показать двойственное утверждение, если λ не корень, то λ не в спектре. Но это в точности Утверждение 8 пункт (2). □

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.

Примеры

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

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

Так как диагональные матрицы складываются и умножаются поэлементно

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

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

[17] Аналогично определяются спектры в рациональном, комплексном и прочих случаях.

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 $\operatorname{spec}_{\mathbb{R}} A = \{\lambda_1, \dots, \lambda_n\}$. Так как многочлен f зануляет A, утверждение 8 пункт (2) 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 $A = \left(\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right) \in \mathrm{M}_2(\mathbb{R})$. Прямое вычисление показывает, что $A^2 = -E$, то есть многочлен $f(x) = x^2+1$ зануляет A. Покажем, что $\operatorname{spec}_{\mathbb{R}} A = \varnothing$. Действительно, по утверждению 8 пункт (2) 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 $\operatorname{spec}_{\mathbb{C}} A = \{i, -i\}$.

Минимальный многочлен Пусть $A \in \mathrm{M}_n(\mathbb{R})$ – 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 $A \in \mathrm{M}_n(\mathbb{R})$, тогда верны следующие утверждения:*

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

Доказательство. (1). По утверждению 8 пункт (1) 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, $r(A) = f_{min}(A) = f'_{min}(A) = 0$. А значит $r(x) = 0$.

(4). Мы уже знаем, что $\operatorname{spec}_{\mathbb{R}} A$ лежит среди корней f_{min} (утверждение 8 пункт (2)). Осталось показать обратное включение. Предположим обратное, что есть $\lambda \in \mathbb{R}$ такое, что $f_{min}(\lambda) = 0$, но $\lambda \notin \operatorname{spec}_{\mathbb{R}} A$. 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 $\lambda \notin \operatorname{spec}_{\mathbb{R}} A$, 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. □

Поиск минимального многочлена Пусть задана матрица $A \in \mathrm{M}_n(\mathbb{R})$. Тогда мы знаем, что найдется многочлен $f \in \mathbb{R}[x]$ такой, что $f(A) = 0$. Кроме того, я сообщил, что $\deg f \leqslant n$. Давайте обсудим, как найти подобный многочлен. Будем искать его с неопределенными коэффициентами $f(x) = a_0 + a_1 x + \ldots + a_n x^n$. Подставим в многочлен матрицу A и приравняем результат к нулю.

$$f(A) = a_0 E + a_1 A + \ldots + a_n A^n = 0$$

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

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

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 $A \in \mathrm{M}_n(\mathbb{R})$, 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

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

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 $\mathrm{spec}_{\mathbb{R}} A$ 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 $\mathrm{spec}_{\mathbb{C}} A$ 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 $A \in \mathrm{M}_n(\mathbb{R})$ 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 $f_{min} = a_0 + a_1 x + \ldots + a_m x^m$ для некоторой матрицы $A \in \mathrm{M}_n(\mathbb{R})$. Тогда

$$a_0 E + a_1 A + \ldots + a_m A^m = 0 \quad \Rightarrow \quad A(a_1 E + \ldots + a_m A^{m-1}) = -a_0 E \quad \Rightarrow \quad A\left(-\frac{a_1}{a_0}E - \ldots - \frac{a_m}{a_0}A^{m-1}\right) = E$$

То есть по определению

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

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.

2.15 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. $\mathbb{R}_+ = \{r \in \mathbb{R} \mid r \geqslant 0\}$.

Пусть задано отображение

$$\mathrm{M}_{m\,n}(\mathbb{R}) \to \mathbb{R}_+$$

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 $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, $|A| = 0$ тогда и только тогда, когда $A = 0$.
2. Для любой матрицы $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ и любого числа $\lambda \in \mathbb{R}$ выполнено $|\lambda A| = |\lambda| \cdot |A|$.[18]
3. Для любых двух матриц $A, B \in \mathrm{M}_{m\,n}(\mathbb{R})$ выполнено $|A + B| \leqslant |A| + |B|$.

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

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

Простые примеры

1. 1-норма на $\mathrm{M}_n(\mathbb{R})$:
$$|A|_1 = \sum_{ij} |a_{ij}|, \quad A \in \mathrm{M}_n(\mathbb{R})$$
2. ∞-норма на $\mathrm{M}_n(\mathbb{R})$:
$$|A|_\infty = \max_{ij} |a_{ij}|, \quad A \in \mathrm{M}_n(\mathbb{R})$$
3. Норма Фробениуса или 2-норма на $\mathrm{M}_n(\mathbb{R})$:
$$|A|_F = |A|_2 = \sqrt{\sum_{ij} |a_{ij}|^2}, \quad A \in \mathrm{M}_n(\mathbb{R})$$
4. p-норма на $\mathrm{M}_n(\mathbb{R})$:
$$|A|_p = \sqrt[p]{\sum_{ij} |a_{ij}|^p}$$

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

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

Индуцированная (согласованная) норма Пусть $|-|\colon \mathbb{R}^n \to \mathbb{R}_+$ – некоторая фиксированная норма. Определим индуцированную ей норму $\|-\|$ на $\mathrm{M}_n(\mathbb{R})$ следующим образом:[19]

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

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

Утверждение. *Для любой нормы $|-|\colon \mathbb{R}^n \to \mathbb{R}_+$ индуцированная ей норма $\|-\|\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}_+$ является субмультипликативной.*

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

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

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

Значит

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

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

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

□

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 $\mathrm{M}_n(\mathbb{R})$ 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 $\mathrm{M}_n(\mathbb{R})$ 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 $\mathrm{M}_n(\mathbb{R})$ уже считается более хитрым способом. Пусть $A \in \mathrm{M}_n(\mathbb{R})$. Если взять матрицу A^tA, то окажется, что ее спектр состоит целиком из неотрицательных вещественных чисел. Пусть σ_1 – максимальное такое число, тогда $\|A\|_2 = \sqrt{\sigma_1}$.[20]

2.16 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.

[19] Можно определить норму на прямоугольных матрицах, но тогда на до 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.

[20] К 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.

Сходимость Основная задача нормы – дать понятие о близости матриц друг к другу. А именно, если есть норма $|-|\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}_+$, то можно определить расстояние между матрицами $A, B \in \mathrm{M}_n(\mathbb{R})$ 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 $A_n \in \mathrm{M}_n(\mathbb{R})$, тогда скажем, что она сходится к матрице $A \in \mathrm{M}_n(\mathbb{R})$ и будем писать $A_n \to A$, $n \to \infty$ (или $\lim_n A_n = A$), если $\rho(A_n, A) \to 0$ как последовательность чисел при условии $n \to \infty$.

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 $|-|$ и $|-|'$ – две разные нормы на $\mathrm{M}_n(\mathbb{R})$. Будем говорить, что они эквивалентны, если существуют две положительные константы $c_1, c_2 \in \mathbb{R}$ такие, что $c_1|A| \leqslant |A|' \leqslant c_2|A|$ для всех матриц $A \in \mathrm{M}_n(\mathbb{R})$. Если две нормы эквивалентны, то несложно углядеть, что расстояние $\rho(A_n, A)$ в смысле нормы $|-|$ стремится к нулю тогда и только тогда, когда расстояние $\rho'(A_n, A)$ в смысле нормы $|-|'$ 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 $A \in \mathrm{M}_n(\mathbb{R})$, 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, $e^x = \sum_{k \geqslant 0} x^k/k!$. Тогда можно определить $e^A = \sum_{k \geqslant 0} A^k/k!$. 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 зануляющий 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 $\varphi\colon X \to Y$ – 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.[21]

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

Для любого множества X отображение $\mathrm{Id}\colon X \to X$ заданное по правилу $\mathrm{Id}(x) = x$ называется *тождественным*. Пусть $\varphi\colon X \to Y$ – некоторое отображение. Тогда $\psi\colon Y \to X$ называется *левым обратным* (соответственно *правым обратным*) к φ, если $\psi\varphi = \mathrm{Id}$ ($\varphi\psi = \mathrm{Id}$).[22] 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 $\varphi\colon X \to Y$ – некоторое отображение. Тогда*

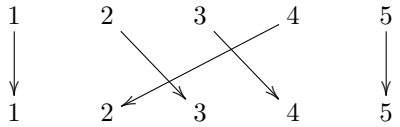
1. *φ инъективно тогда и только тогда, когда φ обладает левым обратным.*
2. *φ сюръективно тогда и только тогда, когда φ обладает правым обратным.*
3. *φ биективно тогда и только тогда, когда φ обратимо.*

3.2 Перестановки

Пусть $X_n = \{1, \ldots, n\}$ – конечное множество из n занумерованных элементов.[23] *Перестановкой* называется биективное отображение $\sigma\colon X_n \to X_n$. 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:

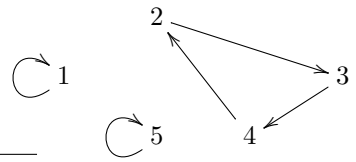
1. Задать стрелками соответствие на элементах



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



[21] В 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.

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

[23] Формально 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 $\sigma \in \mathrm{S}_n$ 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 виде[24]

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

Из однозначности табличной записи получаем следующее.

Утверждение. *Количество перестановок на n элементах есть $n!$, т.е. $|\,\mathrm{S}_n\,| = n!$.*

3.3 Операция на перестановках

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

$$X_n \xrightarrow{\tau} X_n \xrightarrow{\sigma} X_n, \quad X_n \xrightarrow{\sigma\tau} X_n$$

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 $\sigma, \tau, \rho \in \mathrm{S}_n$ – 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 $\sigma\tau \neq \tau\sigma$.[25]
- Тождественное отображение Id является нейтральным элементом для умножения перестановок в том смысле, что верно $\mathrm{Id}\,\sigma = \sigma\,\mathrm{Id} = \sigma$ 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 $\sigma\sigma^{-1} = \sigma^{-1}\sigma = \mathrm{Id}$. В табличной записи обратное отображение можно записать так

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

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

[25] Один 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. $\tau\colon \{1, \dots, n\} \to \{"1", \dots, "n"\}$. Потому, если мысленно убрать кавычки, то на переименование можно смотреть как на перестановку $\tau\colon X_n \to X_n$.

Пусть теперь у нас есть перестановка $\sigma\colon X_n \to X_n$. 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:

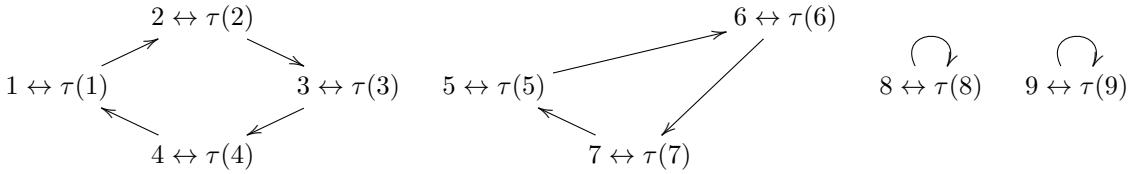
$$\begin{array}{ccc} \{1, \dots, n\} & \xrightarrow{\tau} & \{"1", \dots, "n"\} \\ {\scriptstyle \sigma_{\text{стар}}}\downarrow & & \downarrow{\scriptstyle \sigma_{\text{нов}}} \\ \{1, \dots, n\} & \xrightarrow{\tau} & \{"1", \dots, "n"\} \end{array}$$

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 $\sigma \in \mathrm{S}_n$ действует следующим образом: для некоторого множества $i_1, \dots, i_k$ $(k \geqslant 2)$ выполнено

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

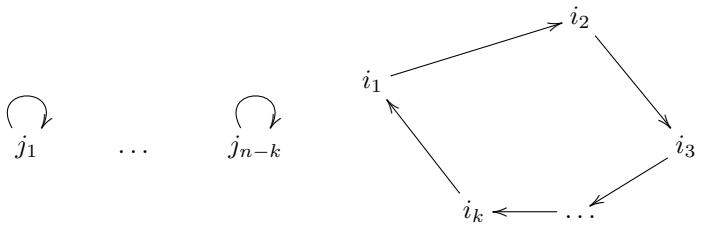
а все остальные элементы остаются на месте под действием σ. Тогда σ называется *циклом* длины k. Такая перестановка для краткости обозначается $(i_1, \dots, i_k)$. Заметим, что такая запись не единственная: например, можно сказать $\sigma = (i_2, \dots, i_k, i_1)$.[26] Стоит отметить, что если в определении 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

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

[26] Как 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 $(2,3)(1,2) = (3,2,1)$.[27, 28]

Утверждение 12. *Пусть $\rho = (i_1,\dots,i_k) \in \mathrm{S}_n$ – некоторый цикл длины k и $\tau \in \mathrm{S}_n$ – 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.

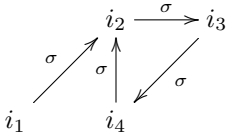
Второй способ – проверка в лоб. Надо проверить, что и левая и правая часть одинаково действуют на всех элементах вида $\tau(i)$. Возьмем элемент $\tau(i_1)$, тогда правая часть его переводит в $\tau(i_2)$. Посмотрим, что с ним делает левая часть. Вначале, мы переходим в i_1, потом в i_2, а потом в $\tau(i_2)$.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

Теперь мы готовы доказать структурный результат о перестановках.

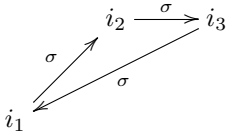
Утверждение 13. *Пусть $\sigma \in \mathrm{S}_n$ – произвольная перестановка. Тогда*

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 $\rho \in \mathrm{S}_n$ – произвольный цикл длины k, тогда его можно представить в виде $\rho = \tau_1 \dots \tau_{k-1}$, где τ_i – транспозиции.*[29]

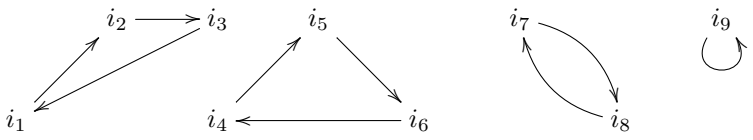
Доказательство. (1) Пусть $i_1 \in X_n$ – произвольный элемент. Подействуем на него σ, получим $i_2 = \sigma(i_1)$ и т.д. Так как X_n конечно, то мы в какой-то момент повторимся, например $i_5 = i_2$, как на рисунке ниже



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

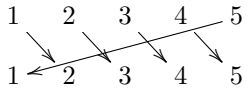


[27] Проверьте это.
[28] Зависимые циклы могут коммутировать, например $(1,2)$ коммутирует с $(1,2)$.
[29] Это представление уже не единственное.

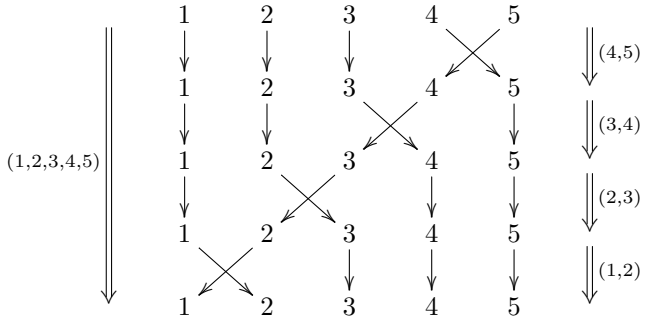
Значит перестановка выше раскладывается в циклы $\sigma = (i_1, i_2, i_3)(i_4, i_5, i_6)(i_7, i_8)$.[30]

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 $(1, 2, \dots, k) = (1, 2)(2, 3) \dots (k-2, k-1)(k-1, k)$.

□

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 12.

3.6 Знак перестановки

Рассмотрим произвольное отображение

$$\phi\colon \mathrm{S}_n \to \{\pm 1\}$$

удовлетворяющее следующим двум свойствам:

1. $\phi(\sigma\tau) = \phi(\sigma)\phi(\tau)$ для любых $\sigma, \tau \in \mathrm{S}_n$.[31]
2. $\phi \not\equiv 1$, 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.

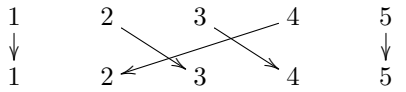
Утверждение 14. *Существует единственное отображение $\phi\colon \mathrm{S}_n \to \{\pm 1\}$ обладающее свойствами (1) и (2).*

В этом случае такое отображение обозначается $\mathrm{sgn}\colon \mathrm{S}_n \to \{\pm 1\}$ и называется знаком. Значение $\mathrm{sgn}(\sigma)$ называется знаком перестановки $\sigma \in \mathrm{S}_n$. Перестановка называется четной, если знак 1 и нечетной, если -1.

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

Пусть $\sigma \in \mathrm{S}_n$ – некоторая перестановка и $i, j \in X_n$ – НЕупорядоченная пара различных элементов.[33] 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 инверсия для всевозможных пар это $d(\sigma) = \sum_{i<j} d_{ij}(\sigma)$. Определим отображение $\mathrm{sgn}\colon \mathrm{S}_n \to \{\pm 1\}$ по правилу $\mathrm{sgn}(\sigma) = (-1)^{d(\sigma)}$. Для доказательства существования надо проверить, что sgn обладает указанными свойствами (1) и (2), то есть надо доказать следующее.

Утверждение. *Пусть $\sigma, \tau \in \mathrm{S}_n$ – произвольные перестановки, тогда*

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

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

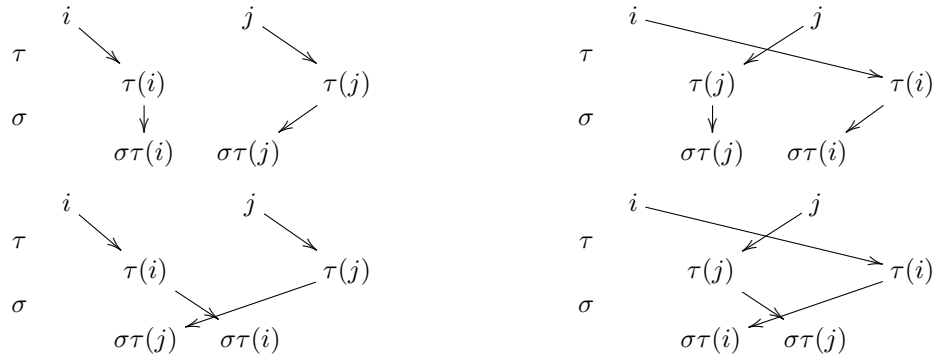
Для доказательства первого надо показать, что

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

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

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

[32] Оно же *число инверсий*.
[33] То 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 $\tau\colon X_n \to X_n$ – 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. □

Единственность

Утверждение. *Пусть $\phi\colon \mathrm{S}_n \to \{\pm 1\}$ обладает свойством (1). Тогда*

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

Доказательство. (1) Рассмотрим цепочку равенств

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

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

(2) Рассмотрим цепочку равенств

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

Значит число $\phi(\sigma^{-1})$ является обратным к $\phi(\sigma)$.[34]

(3) Заметим, что для любых различных $i,j \in X_n$ у нас обязательно существует перестановка $\tau \in \mathrm{S}_n$ такая, что $\tau(1) = i$ и $\tau(2) = j$.[35] Тогда по утверждению 12 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. □

Теперь давайте докажем единственность. Пусть у нас существуют два таких отображения $\phi,\psi\colon \mathrm{S}_n \to \{\pm 1\}$ удовлетворяющие свойствам (1) и (2). Давайте покажем, что $\phi(\sigma) = \psi(\sigma)$ для любой $\sigma \in \mathrm{S}_n$. Из утверждения 13 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 $\sigma \in \mathrm{S}_n$, 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.

3.7 Подсчет знака

Декремент *Декремент* перестановки $\sigma \in \mathrm{S}_n$ – это

$$\mathrm{dec}(\sigma) = n - \text{«количество нетривиальных циклов»} - \text{«количество неподвижных точек»}$$

Если рассматривать все неподвижные точки как тривиальные «циклы», то формула превращается в

$$\mathrm{dec}(\sigma) = n - \text{«количество циклов»}$$

Декремент можно описать еще так: каждая перестановка σ определяет граф на множестве вершин X_n, где (i,j) – ребро, если $\sigma(i) = j$. Тогда

$$\mathrm{dec}(\sigma) = \text{«количество вершин»} - \text{«количество компонент графа»}$$

[34] Так 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.

[35] Я оставляю это как упражнение.

Утверждение 15. *Пусть* $\sigma \in \mathrm{S}_n$, *тогда* $\operatorname{sgn}(\sigma) = (-1)^{\operatorname{dec}(\sigma)}$.

Доказательство. Действительно, разложим перестановку σ в произведение независимых циклов $\sigma = \rho_1 \dots \rho_k$. Пусть длины циклов $d_1, \dots, d_k$, соответственно. Тогда

$$\operatorname{sgn}(\sigma) = (-1)^{d_1 - 1} \dots (-1)^{d_k - 1} = (-1)^{\sum_i d_i - k}$$

Пусть s – количество неподвижных точек. Тогда

$$\operatorname{sgn}(\sigma) = (-1)^{\left(\sum_i d_i + s\right) - k - s} = (-1)^{n-k-s} = (-1)^{\operatorname{dec}(\sigma)}$$

□

При подсчете знака перестановки надо пользоваться декрементом. То есть надо разложить перестановку в произведение независимых циклов и сложить их длины без единицы. Например:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 8 & 2 & 3 & 7 & 1 & 5 & 9 & 6 \end{pmatrix}$$

Теперь видим, что

$$1 \to 4 \to 3 \to 2 \to 8 \to 9 \to 6 \to 1$$
$$5 \to 7 \to 5$$

Значит $\sigma = (1, 4, 3, 2, 8, 9, 6)(5, 7)$, а значит $\operatorname{dec}(\sigma) = 6 + 1 = 7$ и $\operatorname{sgn}(\sigma) = -1$.

3.8 Возведение в степень

Прежде всего сделаем два простых наблюдения:

1. Пусть $\sigma, \tau \in \mathrm{S}_n$ – две коммутирующие перестановки, тогда $(\sigma\tau)^m = \sigma^m \tau^m$.
2. Пусть $\rho \in \mathrm{S}_n$ – цикл длины d, тогда d совпадает с наименьшим натуральным числом k таким, что $\rho^k = \mathrm{Id}$.

Пусть теперь $\sigma \in \mathrm{S}_n$ – произвольная перестановка. Мы можем разложить ее в произведение независимых циклов $\sigma = \rho_1 \dots \rho_k$ с длинами $d_1, \dots, d_k$, соответственно. Тогда

$$\sigma^m = \rho_1^m \dots \rho_k^m = \rho_1^{m \pmod{d_1}} \dots \rho_k^{m \pmod{d_k}}$$

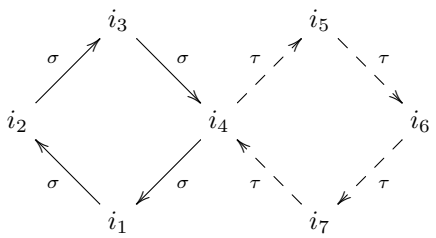
Таким образом, расчет произвольной степени перестановки σ сводится к возведению циклов в степень не большую их длины.

Оставим еще одно замечание в качестве упражнения. Если $\sigma = \rho_1 \dots \rho_k$ – разложение в произведение независимых циклов длин $d_1, \dots, d_k$, соответственно, то наименьшее натуральное r такое, что $\sigma^r = \mathrm{Id}$, равно наименьшему общему кратному чисел $d_1, \dots, d_k$.

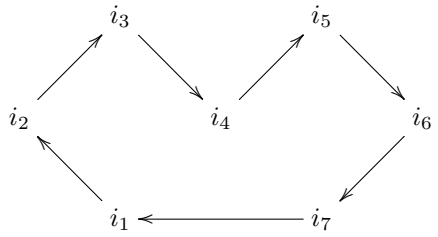
3.9 Произведение циклов

В этом разделе я приведу несколько примеров того, как перемножаются между собой зависимые циклы.

Два цикла Пусть циклы $\sigma, \tau \in \mathrm{S}_n$ пересекаются по одному элементу как на рисунке ниже

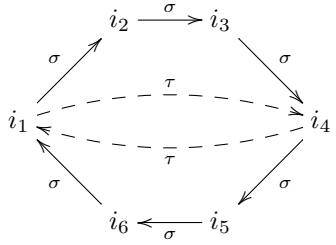


Надо найти произведение $\sigma\tau$. Нетрудно видеть, что результат имеет следующий вид:

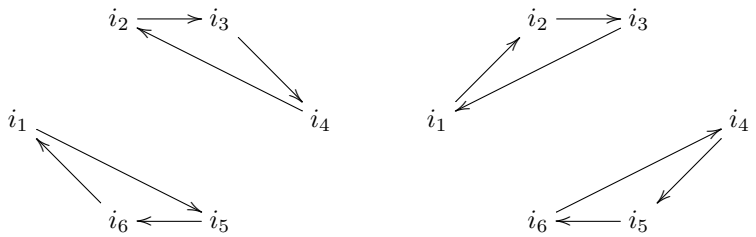


Таким образом мы получили формулу $(i_1, \ldots, i_k)(i_k, \ldots, i_n) = (i_1, \ldots, i_n)$.

Цикл и транспозиция Пусть $\sigma, \tau \in \mathrm{S}_n$, где σ – цикл, а τ – транспозиция, переставляющая два элемента цикла σ как на рисунке ниже.



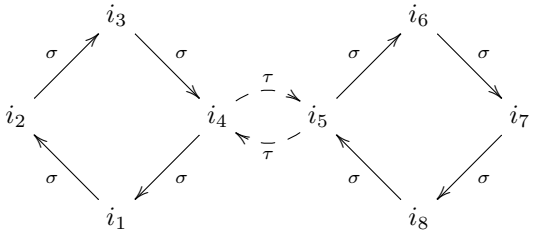
Вот так выглядят композиции для $\sigma\tau$ и $\tau\sigma$ соответственно



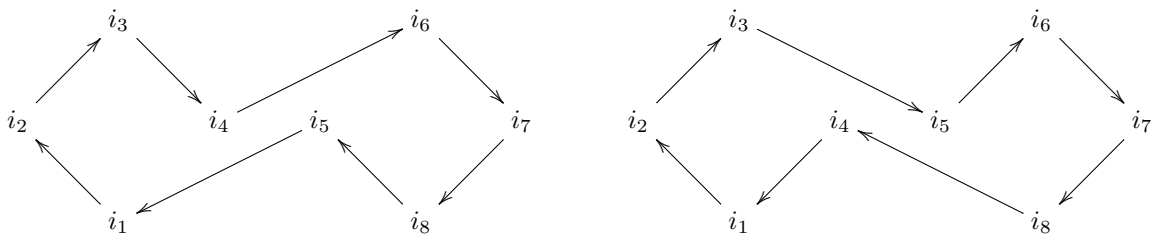
Таким образом общее правило выглядит так:

$$(i_1, \ldots, i_n)(i_1, i_k) = (i_1, i_{k+1}, \ldots, i_n)(i_2, \ldots, i_k)$$
$$(i_1, i_k)(i_1, \ldots, i_n) = (i_1, \ldots, i_{k-1})(i_k, \ldots, i_n)$$

Пара циклов и транспозиция Пусть $\sigma, \tau \in \mathrm{S}_n$, причем σ – произведение двух независимых циклов, а τ – транспозиция, переставляющая две вершины из разных циклов как на рисунке ниже.



Произведения $\sigma\tau$ и $\tau\sigma$ имеют вид



Таким образом общее правило выглядит так

$$(i_1, \ldots, i_k)(i_{k+1}, \ldots, i_n)(i_k, i_{k+1}) = (i_1, \ldots, i_k, i_{k+2}, \ldots, i_n, i_{k+1})$$
$$(i_k, i_{k+1})(i_1, \ldots, i_k)(i_{k+1}, \ldots, i_n) = (i_k, i_1, \ldots, i_{k-1}, i_{k+1}, \ldots, i_n)$$

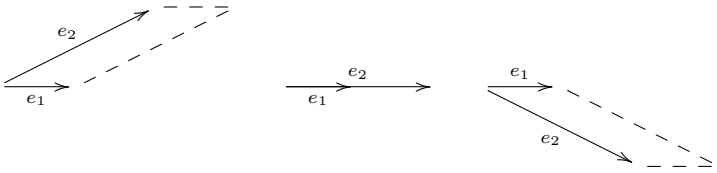
4 Определитель

4.1 Философия

Сейчас я хочу обсудить «ориентированный объем» на прямой, плоскости и в пространстве.

Прямая На прямой мы можем выбрать «положительное» направление. Обычно на рисунке выбирают слева направо. Тогда длина вектора, который смотрит слева направо, считается положительной, а справа налево – отрицательной.

Плоскость Здесь объем будет задаваться парой векторов, то есть некоторой квадратной матрицей размера 2, где вектора – это ее столбцы. Основная идея такая: пусть мы хотим посчитать площадь между двумя векторами на плоскости, точнее площадь параллелограмма натянутого на вектора e_1 и e_2 как на первом рисунке ниже.

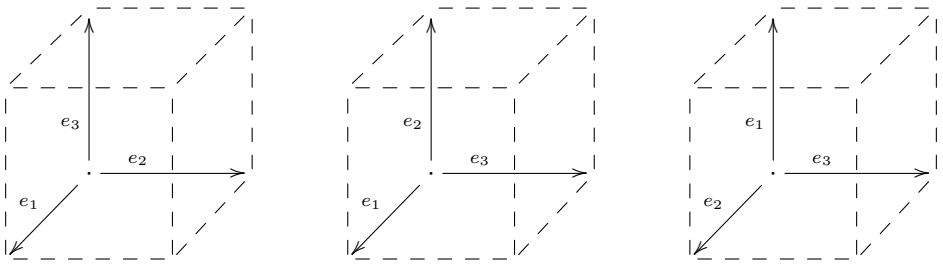


Давайте двигать вектор e_2 к вектору e_1. Тогда площадь будет уменьшаться и когда вектора совпадут, она будет равна нулю. Однако, если мы продолжим двигать вектор e_2, то площадь между векторами опять начнет расти и картинка в конце концов станет симметрична исходной, а полученный параллелограмм равен изначальному. Однако, эта ситуация отличается от предыдущей и вот как можно понять чем. Предположим, что между векторами была натянута хорошо сжимаемая ткань, одна сторона которой красная, другая – зеленая. Тогда в самом начале на нас смотрит красная сторона этой ткани, но как только e_2 прошел через e_1 на нас уже смотрит зеленая сторона. Мы бы хотели научиться отличать эти две ситуации с помощью знака, если на нас смотрит красная сторона – знак положительный, если зеленая – отрицательный.

Еще один способ думать про эту ситуацию. Представим, что плоскость – это наш стол, а параллелограмм вырезан из бумаги. Мы можем положить параллелограмм на стол двумя способами: лицевой стороной вверх или же вниз. В первом случае мы считаем площадь положительной, а во втором – отрицательной. Возможность определить лицевую сторону связана с тем, что мы знаем, где у стола верх, а где низ. Это возможно, потому что наша плоскость лежит в трехмерном пространстве и мы можем глядеть на нее извне. Однако, если бы мы жили на плоскости и у нас не было бы возможности выглянуть за ее пределы, то единственный способ установить «какой стороной вверх лежит параллелограмм» был бы с помощью порядка векторов.

Еще одно важное замечание. Если мы берем два одинаковых параллелограмма на нашем столе, которые лежат лицевой стороной вверх, то мы можем передвинуть один в другой, не отрывая его от стола. А вот если один из параллелограммов имеет положительный объем, а другой отрицательный, то нельзя перевести один в другой, не отрывая от стола. То есть, если вы живете на плоскости, то вам не получится переместить положительный параллелограмм в отрицательный, не сломав или не разобрав его.

Пространство В пространстве дело с ориентацией обстоит абсолютно аналогично. Мы хотим уже считать объемы параллелепипедов натянутых на три вектора. И мы так же хотим, чтобы эти объемы показывали «с какой стороны» мы смотрим на параллелепипед.



Здесь знак объема определяется по порядку векторов, как знак перестановки. На рисунке объемы первого и третьего положительные, а у второго отрицательный. Если вы сделаете модельки этих кубиков из подписан-

ных спичек, то третий кубик – это первый, но лежащий на другой грани. А вот второй кубик получить из первого вращениями не получится. Надо будет его разобрать и присобачить ребра по-другому.

Как и в случае с плоскостью, если бы мы могли выйти за пределы нашего трехмерного пространства, то у нас появилась бы лицевая и тыльная сторона, как у стола. И тогда первый и третий кубики лежали бы лицевой стороной вверх, а второй – вниз. Мы, конечно же, так сделать не сможем и никогда в жизни не увидим подобное, но думать про такое положение вещей по аналогии с плоскостью можем и эта интуиция бывает полезна.

Пояснение планов В текущей лекции я не собираюсь обсуждать объемы, а всего лишь хочу коснуться некоторой техники, которая используется для работы с ориентированными объемами. Чтобы начать честный рассказ про сами объемы (который обязательно будет, но позже), нам надо поговорить о том, что такое векторное пространство и как в абстрактном векторном пространстве мерить расстояния и углы. Потому, пока мы не покроем эти темы, всерьез говорить про настоящие объемы мы не сможем.

4.2 Три разных определения

Начнем с классического определения в виде явной формулы.

Определитель (I) Рассмотрим отображение $\det\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$, задаваемое следующей формулой: для любой матрицы $A \in \mathrm{M}_n(\mathbb{R})$ положим

$$\det A = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \dots a_{n\sigma(n)}$$

Данное отображение называется *определителем*, а его значение $\det A$ на матрице A называется определителем матрицы A.

Давайте неформально обсудим, как считается выражение для определителя. Как мы видим определитель состоит из суммы некоторых произведений. Каждое произведение имеет вид $a_{1\sigma(1)} \dots a_{n\sigma(n)}$ умноженное на $\operatorname{sgn}(\sigma)$. Здесь из каждой строки матрицы A[36] выбирается по одному элементу так, что никакие два элемента не лежат в одном столбце (это гарантированно тем, что σ – перестановка и потому $\sigma(i)$ не повторяются). Заметим, что слагаемых ровно столько, сколько перестановок – $n!$ штук. Из этих слагаемых половина идет со знаком плюс, а другая – со знаком минус.

Нормированные полилинейные кососимметрические отображения (II) Пусть $\phi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ – некоторое отображение и $A \in \mathrm{M}_n(\mathbb{R})$. Тогда про матрицу A можно думать, как про набор из n столбцов: $A = (A_1|\dots|A_n)$. Тогда функцию $\phi(A) = \phi(A_1, \dots, A_n)$ можно рассматривать как функцию от n столбцов.

В обозначениях выше рассмотрим отображения $\phi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$, удовлетворяющие следующим свойствам:

1. $\phi(A_1, \dots, A_i + A_i', \dots, A_n) = \phi(A_1, \dots, A_i, \dots, A_n) + \phi(A_1, \dots, A_i', \dots, A_n)$ для любого i.
2. $\phi(A_1, \dots, \lambda A_i, \dots, A_n) = \lambda\phi(A_1, \dots, A_i, \dots, A_n)$ для любого i и любого $\lambda \in \mathbb{R}$.
3. $\phi(A_1, \dots, A_i, \dots, A_j, \dots, A_n) = -\phi(A_1, \dots, A_j, \dots, A_i, \dots, A_n)$ для любых различных i и j.
4. $\phi(E) = 1$.

Первые два свойства вместе называются *полилинейностью* ϕ по столбцам, т.е. это уважение суммы и умножения на скаляр. Третье свойство называется *кососимметричностью* ϕ по столбцам. Последнее условие – это условие нормировки. Данный набор свойств можно заменить эквивалентным с переформулированным третьим свойством:

1. $\phi(A_1, \dots, A_i + A_i', \dots, A_n) = \phi(A_1, \dots, A_i, \dots, A_n) + \phi(A_1, \dots, A_i', \dots, A_n)$ для любого i.
2. $\phi(A_1, \dots, \lambda A_i, \dots, A_n) = \lambda\phi(A_1, \dots, A_i, \dots, A_n)$ для любого i и любого $\lambda \in \mathbb{R}$.
3. $\phi(A_1, \dots, A', \dots, A', \dots, A_n) = 0$, т.е. если есть два одинаковых столбца, то значение ϕ равно нулю.
4. $\phi(E) = 1$.

[36] Первый индекс – индекс строки.

Действительно, обозначим $\Phi(a,b) = \phi(A_1, \dots, a, \dots, b, \dots, A_n)$. Тогда Φ полилинейная функция двух аргументов.[37] И нам надо показать, что $\Phi(a,a) = 0$ для любого $a \in \mathbb{R}^n$ тогда и только тогда, когда $\Phi(a,b) = -\Phi(b,a)$ для любых $a, b \in \mathbb{R}^n$. Для $\Rightarrow$ подставим $b = a$, получим $\Phi(a,a) = -\Phi(a,a)$. Для обратного $\Leftarrow$ подставим $a+b$, получим $\Phi(a+b, a+b) = 0$. Раскроем скобки: $\Phi(a,a) + \Phi(a,b) + \Phi(b,a) + \Phi(b,b) = 0$. Откуда следует требуемое.

Пример На всякий случай поясню все свойства выше на примерах:

1. $\phi\begin{pmatrix}1 & 3\\ 2 & 7\end{pmatrix} = \phi\left(\begin{pmatrix}1\\2\end{pmatrix}\middle|\begin{pmatrix}3\\7\end{pmatrix}\right) = \phi\left(\begin{pmatrix}1\\2\end{pmatrix}\middle|\begin{pmatrix}1\\3\end{pmatrix} + \begin{pmatrix}2\\4\end{pmatrix}\right) = \phi\begin{pmatrix}1 & 1\\ 2 & 3\end{pmatrix} + \phi\begin{pmatrix}1 & 2\\ 2 & 4\end{pmatrix}$
2. $\phi\begin{pmatrix}1 & 3\\ 2 & 9\end{pmatrix} = 3\,\phi\begin{pmatrix}1 & 1\\ 2 & 3\end{pmatrix}$
3. $\phi\begin{pmatrix}1 & 3\\ 2 & 7\end{pmatrix} = -\phi\begin{pmatrix}3 & 1\\ 7 & 2\end{pmatrix}$

3' $\phi\begin{pmatrix}1 & 1\\ 2 & 2\end{pmatrix} = 0$

Везде далее будем упоминать отображения с такими свойствами, как отображения со свойством (II).

Нормированные полилинейные кососимметрические отображения (II') Аналогично (II) можно рассмотреть полилинейные кососимметрические отображения по строкам матрицы A вместо столбцов. Тогда можно рассматривать отображения $\phi'\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ с аналогами четырех свойств выше: полилинейность, кососимметричность, значение 1 на единичной матрице. Такие отображения мы будем называть, как отображения со свойствами (II').

Специальные мультипликативные отображения (III) Рассмотрим множество отображений $\psi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ удовлетворяющие следующими свойствам:

1. $\psi(AB) = \psi(A)\psi(B)$ для любых $A, B \in \mathrm{M}_n(\mathbb{R})$.
2. $\psi\begin{pmatrix}1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & d\end{pmatrix} = d$ для любого ненулевого $d \in \mathbb{R}$.

Всюду ниже будем упоминать отображения с такими свойствами, как отображения со свойством (I).[38]

План дальнейших действий Наша задача показать, что, во-первых, определитель обладает свойствами (II), (II') и (III), а, во-вторых, что кроме определителя никакое другое отображение не удовлетворяет этим свойствам. То есть все три определения между собой эквивалентны. Самое сложное будет показать, что (III) влечет остальные два определения. Это означает, что (III) легко проверять, но из него сложно выводить какие-либо свойства. Самые полезные с вычислительной точки зрения – определения (II) и (II').

4.3 Явные формулы для определителя

Подсчет в малых размерностях

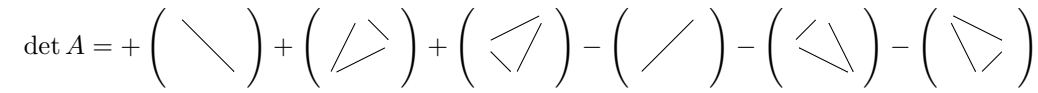
1. Если $A \in \mathrm{M}_1(\mathbb{R}) = \mathbb{R}$, то $\det A = A$.
2. Если $A \in \mathrm{M}_2(\mathbb{R})$ имеет вид $A = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$, то $\det A = ad - bc$. Графически: главная диагональ минус побочная.

[37] Такие отображения называются билинейными.

[38] Обратите внимание, что существует много отображений со свойством (1), не удовлетворяющих свойству (2). Действительно, если ψ – мультипликативное отображение, то есть удовлетворяет только свойству (1), то $\gamma_n(A) = \psi(A)^n$ – тоже мультипликативное отображение для любого натурального $n \in \mathbb{N}$. Кроме того, $\delta_\alpha(A) = |\psi(A)|^\alpha$ тоже является мультипликативным отображением для любого положительного $\alpha \in \mathbb{R}$.

3. Если $A \in \mathrm{M}_3(\mathbb{R})$ имеет вид $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, то определитель получается из 6 слагаемых три из них с $+$ три с $-$. Графически слагаемые можно изобразить так:

$$\det A = + \left(\quad \right) + \left(\quad \right) + \left(\quad \right) - \left(\quad \right) - \left(\quad \right) - \left(\quad \right)$$



Точная формула[39]

$$\det A = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}$$

Треугольные матрицы

Утверждение 16. *Для любых верхне и нижне треугольных матриц верны следующие формулы:*

$$\det \begin{pmatrix} \lambda_1 & \dots & * \\ & \ddots & \vdots \\ & & \lambda_n \end{pmatrix} = \lambda_1 \dots \lambda_n \quad \det \begin{pmatrix} \lambda_1 & & \\ \vdots & \ddots & \\ * & \dots & \lambda_n \end{pmatrix} = \lambda_1 \dots \lambda_n$$

В частности $\det E = 1$.

Доказательство. Я докажу утверждение для верхнетреугольных матриц, нижнетреугольный случай делается аналогично. Для доказательства надо посчитать определитель по определению и увидеть, что только одно слагаемое соответствующее тождественной перестановке является не нулем. Действительно, рассмотрим выражение $a_{1\sigma(1)} \dots a_{n\sigma(n)}$. Посмотрим, когда это выражение не ноль. Последний множитель $a_{n\sigma(n)}$ лежит в последней строке и должен быть не ноль. Для этого должно выполняться $\sigma(n) = n$. Теперь $a_{n-1\sigma(n-1)}$ должен быть не ноль. Так как $\sigma(n) = n$, то $\sigma(n-1) \neq n$. А значит, чтобы $a_{n-1\sigma(n-1)}$ был не ноль, остается только один случай $\sigma(n-1) = n-1$. Продолжая аналогично, мы видим, что $\sigma(i) = i$ для всех строк i. □

4.4 Свойства определителя

Определитель и транспонирование Прежде чем перейти к доказательству следующего утверждения сделаем одно полезное наблюдение. Если мы возьмем две произвольные перестановки $\sigma, \tau \in \mathrm{S}_n$ и матрицу $A \in \mathrm{M}_n(\mathbb{R})$, то выражения $a_{\tau(1)\sigma(\tau(1))} \dots a_{\tau(n)\sigma(\tau(n))}$ совпадает с выражением $a_{1\sigma(1)} \dots a_{n\sigma(n)}$ с точностью до перестановки сомножителей. Это делается методом пристального взгляда: замечаем что каждый сомножитель одного выражения ровно один раз встречается в другом и наоборот.

Утверждение 17. *Пусть* $A \in \mathrm{M}_n(\mathbb{R})$, *тогда* $\det A = \det A^t$.

Доказательство. Посчитаем по определению $\det A^t$, получим

$$\det A^t = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n} = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\sigma^{-1}(\sigma(1))} \dots a_{\sigma(n)\sigma^{-1}(\sigma(n))}$$

Теперь применим наше замечание перед доказательством:

$$a_{\sigma(1)\sigma^{-1}(\sigma(1))} \dots a_{\sigma(n)\sigma^{-1}(\sigma(n))} = a_{1\sigma^{-1}(1)} \dots a_{n\sigma^{-1}(n)}$$

Значит

$$\det A^t = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma^{-1}(1)} \dots a_{n\sigma^{-1}(n)}$$

Вспомним, что $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$. Следовательно:

$$\det A^t = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma^{-1}) a_{1\sigma^{-1}(1)} \dots a_{n\sigma^{-1}(n)}$$

[39] Для больших размерностей чем 3 на 3 явная формула не пригодна из-за слишком большого числа слагаемых. Даже с вычислительной точки зрения.

Теперь, если σ пробегает все перестановки, то σ^{-1} тоже пробегает все перестановки, так как отображение $\mathrm{S}_n \to \mathrm{S}_n$ по правилу $\sigma \mapsto \sigma^{-1}$ является биекцией.[40] То есть мы можем сделать замену $\tau = \sigma^{-1}$ и приходим к выражению

$$\det A^t = \sum_{\tau \in \mathrm{S}_n} \operatorname{sgn}(\tau) a_{1\tau(1)} \dots a_{n\tau(n)}$$

Последнее в точности совпадает с определением $\det A$. □

Отметим, что если мы доказали какое-то свойство определителя для столбцов, то это утверждение автоматически гарантирует, что такое же свойство выполнено и для строк. И наоборот, если что-то сделано для строк, то это автоматом следует для столбцов.

4.5 Полилинейность и кососимметричность определителя

Сейчас мы докажем, что определитель обладает всеми свойствами (II) и (II'). В силу утверждения 17 нам достаточно показать только (II).

Утверждение 18. *Отображение* $\det\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ *рассматриваемое как отображение столбцов матрицы является полилинейным и кососимметричным, т.е. удовлетворяет следующим свойствам:*

1. $\det(A_1, \dots, A_i + A_i', \dots, A_n) = \det(A_1, \dots, A_i, \dots, A_n) + \det(A_1, \dots, A_i', \dots, A_n)$ *для любого* i.
2. $\det(A_1, \dots, \lambda A_i, \dots, A_n) = \lambda \det(A_1, \dots, A_i, \dots, A_n)$ *для любого* i *и любого* $\lambda \in \mathbb{R}$.
3. $\det(A_1, \dots, A_i, \dots, A_j, \dots, A_n) = -\det(A_1, \dots, A_j, \dots, A_i, \dots, A_n)$ *для любых различных* i *и* j.
4. $\det E = 1$.

Доказательство. Мы знаем, что

$$\det A = \det A^t = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(i)i} \dots a_{\sigma(n)n}$$

Проверим свойство (1):

$$\det(A_1, \dots, A_i + A_i', \dots, A_n) = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots \left(a_{\sigma(i)i} + a'_{\sigma(i)i}\right) \dots a_{\sigma(n)n} =$$

$$= \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(i)i} \dots a_{\sigma(n)n} + \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots a'_{\sigma(i)i} \dots a_{\sigma(n)n} =$$

$$= \det(A_1, \dots, A_i, \dots, A_n) + \det(A_1, \dots, A_i', \dots, A_n)$$

Теперь свойство (2):

$$\det(A_1, \dots, \lambda A_i, \dots, A_n) = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots \left(\lambda a_{\sigma(i)i}\right) \dots a_{\sigma(n)n} = \lambda \det(A_1, \dots, A_i, \dots, A_n)$$

Для проверки свойства (3) введем следующее обозначение. Пусть $\tau \in \mathrm{S}_n$ обозначает транспозицию (i, j). Тогда посчитаем определитель с переставленными местами столбцами i и j:

$$\det(A_1, \dots, A_j, \dots, A_i, \dots, A_n) = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(i)j} \dots a_{\sigma(j)i} \dots a_{\sigma(n)n} =$$

$$= \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)\tau(1)} \dots a_{\sigma(i)\tau(i)} \dots a_{\sigma(j)\tau(j)} \dots a_{\sigma(n)\tau(n)} =$$

$$= \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(\tau^{-1}(1))1} \dots a_{\sigma(\tau^{-1}(n))n} = -\sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma\tau^{-1}) a_{\sigma(\tau^{-1}(1))1} \dots a_{\sigma(\tau^{-1}(n))n}$$

[40] Оно биекция, так как имеет обратное – оно само.

Здесь при переходе от второй строчки к третьей мы воспользовались замечанием перед утверждением 17. Так как отображение $\mathrm{S}_n \to \mathrm{S}_n$ по правилу $\sigma \mapsto \sigma\tau^{-1}$ является биекцией, то если σ пробегает все перестановки, то и $\sigma\tau^{-1}$ пробегает все перестановки. А значит, делая замену $\rho = \sigma\tau^{-1}$, получаем

$$-\sum_{\sigma\in \mathrm{S}_n} \operatorname{sgn}(\sigma\tau^{-1}) a_{\sigma(\tau^{-1}(1))1} \dots a_{\sigma(\tau^{-1}(n))n} = -\det(A_1, \dots, A_i, \dots, A_j, \dots A_n)$$

(4) Это непосредственно следует из определения, либо, если хотите, можно сослаться на утверждение 16. □

Утверждение 19. *Пусть* $\Phi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ *– полилинейное кососимметричное отображение по столбцам. И пусть матрица* A *имеет нулевой столбец, тогда* $\Phi(A) = 0$.[41]

Доказательство. Пусть $A = (A_1|\dots|0|\dots|A_n)$. Тогда

$$\Phi(A) = \Phi(A_1, \dots, 0, \dots, A_n) = \Phi(A_1, \dots, 0+0, \dots, A_n) = \Phi(A_1, \dots, 0, \dots, A_n) + \Phi(A_1, \dots, 0, \dots, A_n)$$

Теперь вычтем из обеих частей $\Phi(A_1, \dots, 0, \dots, A_n)$ и получим, что $\Phi(A_1, \dots, 0, \dots, A_n) = 0$. □

Утверждение 20. *Если* $A \in \mathrm{M}_n(\mathbb{R})$ *имеет нулевой столбец или нулевую строку, то* $\det A = 0$.

Доказательство. Так как определитель является полилинейной и кососимметричной функцией как по строкам так и по столбцам, то это утверждение следует из предыдущего. □

Определитель от элементарных матриц

Утверждение 21. *Верны следующие утверждения:*

1. $\det(S_{ij}(\lambda)) = 1$, *где* $S_{ij}(\lambda) \in \mathrm{M}_n(\mathbb{R})$ *– матрица элементарного преобразования первого типа.*
2. $\det(U_{ij}) = -1$, *где* $U_{ij} \in \mathrm{M}_n(\mathbb{R})$ *– матрица элементарного преобразования второго типа.*
3. $\det(D_i(\lambda)) = \lambda$, *где* $D_i(\lambda) \in \mathrm{M}_n(\mathbb{R})$ *– матрица элементарного преобразования третьего типа.*

Доказательство. (1) Является следствием для случая верхне- и нижнетреугольных матриц.

(2) Так как U_{ij} получается из единичной матрицы перестановкой i-го и j-го столбцов, то результат следует из кососимметричности определителя.

(3) Следует из полилинейности определителя – свойство (II) (2). □

4.6 Полилинейные кососимметрические отображения

Все утверждения в этом разделе доказываются для строк. Соответствующие утверждения для столбцов доказываются аналогично. Их формулировки и доказательства я оставляю в качестве упражнения.

Утверждение 22. *Пусть* $\phi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ *– полилинейное кососимметрическое отображение по строкам матриц, т.е. удовлетворяет следующим свойствам:*[42]

1. $\phi(A_1, \dots, A_i + A_i', \dots, A_n) = \phi(A_1, \dots, A_i, \dots, A_n) + \phi(A_1, \dots, A_i', \dots, A_n)$ *для любого* i.
2. $\phi(A_1, \dots, \lambda A_i, \dots, A_n) = \lambda\phi(A_1, \dots, A_i, \dots, A_n)$ *для любого* i *и любого* $\lambda \in \mathbb{R}$.
3. $\phi(A_1, \dots, A_i, \dots, A_j, \dots, A_n) = -\phi(A_1, \dots, A_j, \dots, A_i, \dots, A_n)$ *для любых различных* i *и* j.

Тогда $\phi(UA) = \det(U)\phi(A)$ *для любой матрицы* $A \in \mathrm{M}_n(\mathbb{R})$ *и любой элементарной матрицы* $U \in \mathrm{M}_n(\mathbb{R})$.

[41] Аналогичное утверждение выполнено для полилинейного и кососимметричного отображения по строкам, тогда в матрице A должна быть нулевая строка.

[42] Здесь через A_i обозначаются строки матрицы A идущие сверху вниз.

Доказательство. Случай $U = S_{ij}(\lambda)$.

$$\phi(S_{ij}(\lambda)A) = \phi(A_1, \dots, A_i + \lambda A_j, \dots, A_j, \dots, A_n) =$$
$$= \phi(A_1, \dots, A_i, \dots, A_j, \dots, A_n) + \lambda\phi(A_1, \dots, A_j, \dots, A_j, \dots, A_n) = \phi(A) = \det(S_{ij}(\lambda))\phi(A)$$

Случай $U = U_{ij}$.

$$\phi(U_{ij}A) = \phi(A_1, \dots, A_j, \dots, A_i, \dots, A_n) = -\phi(A_1, \dots, A_i, \dots, A_j, \dots, A_n) = -\phi(A) = \det(U_{ij})\phi(A)$$

Случай $U = D_i(\lambda)$.

$$\phi(D_i(\lambda)A) = \phi(A_1, \dots, \lambda A_i, \dots, A_n) = \lambda\phi(A_1, \dots, A_i, \dots, A_n) = \lambda\phi(A) = \det(D_i(\lambda))\phi(A)$$

□

Определитель и элементарные матрицы Заметим, что по утверждению 18, определитель тоже является полилинейной и кососимметрической функцией. Потому доказанное утверждение в частности означает, что $\det(UA) = \det(U)\det(A)$ для любой матрицы $A \in \mathrm{M}_n(\mathbb{R})$ и любой элементарной матрицы $U \in \mathrm{M}_n(\mathbb{R})$.

Подсчет определителя Предыдущее замечание позволяет дать эффективный способ вычисления определителя методом Гаусса. Мы берем матрицу A и приводим ее к ступенчатому виду, попутно запоминая как изменился определитель по сравнению с определителем изначальной матрицы. Если же мы будем использовать только элементарные преобразования первого типа, то определитель вовсе меняться не будет. Ступенчатый вид матрицы всегда верхнетреугольный. Там определитель считается как произведение диагональных элементов.

Следствия утверждения 22

Утверждение 23 (Единственность для полилинейных кососимметричных)**.** *Пусть* $\phi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ *– полилинейное кососимметрическое отображение по строкам матриц. Тогда* $\phi(X) = \det(X)\phi(E)$*. В частности, если* $\phi(E) = 1$*, то* $\phi = \det$*.*

Доказательство. Пусть $X \in \mathrm{M}_n(\mathbb{R})$ – произвольная матрица, тогда ее можно элементарными преобразованиями строк привести к улучшенному ступенчатому виду. Последнее означает, что $X = U_1 \dots U_k S$, где S – матрица улучшенного ступенчатого вида, а U_i – матрицы элементарных преобразований. Применим к этому равенству отдельно ϕ и отдельно det, получим

$$\phi(X) = \det(U_1)\dots\det(U_k)\phi(S)$$
$$\det(X) = \det(U_1)\dots\det(U_k)\det(S)$$

Теперь для матрицы S у нас есть два варианта: либо S единичная, либо содержит нулевую строку.

Пусть $S = E$, тогда

$$\phi(X) = \det(U_1)\dots\det(U_k)\phi(E)$$
$$\det(X) = \det(U_1)\dots\det(U_k)$$

Откуда и получаем требуемое $\phi(X) = \det(X)\phi(E)$.

Пусть теперь S имеет нулевую строку. Тогда утверждения 19 и 20 гарантируют, что $\Phi(S) = 0$ и $\det(S) = 0$. Что тоже влечет равенство $\phi(X) = \det(X)\phi(E)$. □

Утверждение 24 (Мультипликативность определителя)**.** *Пусть* $A, B \in \mathrm{M}_n(\mathbb{R})$ *– произвольные матрицы. Тогда* $\det(AB) = \det(A)\det(B)$*.*

Доказательство. Фиксируем матрицу $B \in \mathrm{M}_n(\mathbb{R})$ и рассмотрим отображение $\gamma\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ по правилу $A \mapsto \det(AB)$. Если $A_1, \dots, A_n$ – строки матрицы A, то $A_1B, \dots, A_nB$ – строки матрицы AB. Из этого легко видеть, что γ – полилинейна и кососимметрическая функция по строкам матрицы A. Значит по утверждению 23 $\gamma(A) = \det(A)\gamma(E)$. Но последнее равносильно $\det(AB) = \det(A)\det(B)$. □

Утверждение 25 (Определитель с углом нулей)**.** *Пусть* $A, \in \mathrm{M}_n(\mathbb{R})$ *и* $B \in \mathrm{M}_m(\mathbb{R})$*. Тогда*

$$\det \begin{pmatrix} A & * \\ 0 & B \end{pmatrix} = \det \begin{pmatrix} A & 0 \\ * & B \end{pmatrix} = \det(A) \det(B)$$

Доказательство. Рассмотрим функцию $\phi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ по правилу

$$\phi(X) = \det \begin{pmatrix} X & * \\ 0 & B \end{pmatrix}$$

Заметим, что эта функция является полилинейной и кососимметричной по столбцам матрицы X. В этом случае по утверждению 23 о единственности для полилинейных кососимметрических отображений она имеет вид $\phi(X) = \det(X)\phi(E)$, то есть

$$\det \begin{pmatrix} A & * \\ 0 & B \end{pmatrix} = \det A \det \begin{pmatrix} E & * \\ 0 & B \end{pmatrix}$$

Теперь рассмотрим функцию $\psi\colon \mathrm{M}_m(\mathbb{R}) \to \mathbb{R}$ по правилу

$$\psi(X) = \det \begin{pmatrix} E & * \\ 0 & X \end{pmatrix}$$

Заметим, что эта функция является полилинейной и кососимметричной по строкам матрицы X. В этом случае по утверждению 23 о единственности для полилинейных кососимметрических отображений она имеет вид $\psi(X) = \det(X)\psi(E)$, то есть

$$\det \begin{pmatrix} E & * \\ 0 & B \end{pmatrix} = \det B \det \begin{pmatrix} E & * \\ 0 & E \end{pmatrix}$$

Последний определитель равен 1, так как по утверждению 16 определитель верхнетреугольной матрицы равен произведению ее диагональных элементов. Теперь собираем вместе доказанные факты и получаем требуемый результат. □

Заметим, что таким образом мы можем считать определитель для блочно верхнетреугольных матриц и для блочно нижнетреугольных матриц с любым количеством блоков. Формулы тогда будут выглядеть так

$$\det \begin{pmatrix} A_1 & * & \dots & * \\ & A_2 & \dots & * \\ & & \ddots & \vdots \\ & & & A_k \end{pmatrix} = \det \begin{pmatrix} A_1 & & & \\ * & A_2 & & \\ \vdots & \vdots & \ddots & \\ * & * & \dots & A_k \end{pmatrix} = \det A_1 \dots \det A_k$$

где $A_i \in \mathrm{M}_{n_i}(\mathbb{R})$ – обязательно квадратные матрицы. Это правило является обобщением утверждения о вычислении определителя для треугольных матриц.

4.7 Мультипликативные отображения

Давайте подытожим, что мы показали. Утверждение 18 вместе с утверждением 17 объясняют почему определитель является полилинейной кососимметрической функцией как строк, так и столбцов. Далее утверждение 23 доказывает, что любая полилинейная кососимметричная функция по строкам, принимающая значение 1 на единичной матрице, должна быть определителем. С помощью утверждения 17 мы получаем аналогичный результат для столбцов. Таким образом мы показали эквивалентность подхода (I) подходам (II) и (II').

Теперь, утверждение 24 показывает, что определитель обязательно мультипликативен, а свойство (III) (2) следует из явных вычислений для элементарных матриц. Тем самым мы показали, что (I) и (II) влекут (III). Осталось показать, что (III) влечет (I), т.е. что определитель является единственной функцией с такими свойствами.

Утверждение 26. *Пусть* $\psi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ *– отображение, удовлетворяющее свойствам:*

1. $\psi(AB) = \psi(A)\psi(B)$ *для любых* $A, B \in \mathrm{M}_n(\mathbb{R})$.
2. $\psi \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & d \end{pmatrix} = d$ *для любого ненулевого* $d \in \mathbb{R}$.

Тогда $\psi = \det$.

Доказательство этого утверждения разобьем в несколько этапов. В начале докажем элементарные свойства мультипликативных отображений.

Утверждение 27. *Пусть* $\psi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ *отображение со свойством* $\psi(AB) = \psi(A)\psi(B)$ *для всех* $A, B \in \mathrm{M}_n(\mathbb{R})$. *Тогда*

1. *Если* $P \in \mathrm{M}_n(\mathbb{R})$ *такая, что* $P^2 = P$, *то* $\psi(P)$ *равно либо* 0, *либо* 1.
2. *В частности, значение* $\psi(0)$ *и* $\psi(E)$ *равно либо* 0, *либо* 1.
3. *Если* $\psi(E) = 0$, *то* $\psi(A) = 0$ *для любой матрицы* $A \in \mathrm{M}_n(\mathbb{R})$.
4. *Если* $\psi(0) = 1$, *то* $\psi(A) = 1$ *для любой матрицы* $A \in \mathrm{M}_n(\mathbb{R})$.
5. *Если* $\psi(E) = 1$, *то* $\psi(A^{-1}) = \psi(A)^{-1}$ *для любой обратимой матрицы* $A \in \mathrm{M}_n(\mathbb{R})$.

Доказательство. (1) Применим ψ к тождеству $P^2 = P$, получим $\psi(P) = \psi(PP) = \psi(P)\psi(P)$. То есть число $\psi(P)$ в квадрате равно самому себе. Значит либо $\psi(P) = 0$, либо $\psi(P) = 1$.

(2) Заметим, что $E^2 = E$ и $0^2 = 0$ и воспользуемся предыдущим пунктом.

(3) Применим ψ к тождеству $A = AE$, получим $\psi(A) = \psi(A)\psi(E) = 0$.

(4) Применим ψ к тождеству $0 = A0$, получим $\psi(0) = \psi(A)\psi(0)$. И так как $\psi(0) = 1$ по предположению, то $\psi(A) = 1$.

(5) Применим ψ к тождеству $AA^{-1} = E$, получим $1 = \psi(E) = \psi(AA^{-1}) = \psi(A)\psi(A^{-1})$. Значит число $\psi(A^{-1})$ является обратным к числу $\psi(A)$, что и требовалось показать. □

Утверждение 28. *Пусть* $\psi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ – *отображение, удовлетворяющее свойствам:*

1. $\psi(AB) = \psi(A)\psi(B)$ *для любых* $A, B \in \mathrm{M}_n(\mathbb{R})$.
2. $\psi(D_n(\lambda)) = \lambda$ *для любого ненулевого* $\lambda \in \mathbb{R}$.

Тогда

1. $\psi(S_{ij}(\lambda)) = 1 = \det(S_{ij}(\lambda))$.
2. $\psi(U_{ij}) = -1 = \det(U_{ij})$.
3. $\psi(D_i(\lambda)) = \lambda = \det(D_i(\lambda))$.

Доказательство. В начале заметим, что $\psi(E) = 1$. Потому что иначе $\psi(A) = 0$ для любой матрицы, что противоречит второму свойству. А раз $\psi(E) = 1$, то можно пользоваться пунктом (5) предыдущего утверждения.

(1) Для доказательства воспользуемся следующим замечанием: если $A, B \in \mathrm{M}_n(\mathbb{R})$ – произвольные обратимые матрицы, то $\psi(ABA^{-1}B^{-1}) = 1$. Действительно,

$$\psi(ABA^{-1}B^{-1}) = \psi(A)\psi(B)\psi(A)^{-1}\psi(B)^{-1} = \psi(A)\psi(A)^{-1}\psi(B)\psi(B)^{-1} = 1$$

Для доказательства нам достаточно представить $S_{ij}(\lambda)$ в таком виде. Давайте проверим, что

$$S_{ij}(\lambda) = D_i(2)S_{ij}(\lambda)D_i^{-1}(2)S_{ij}(\lambda)^{-1}$$

Это равенство проверяется непосредственно глядя на матрицы. Давайте для простоты проверим в случае 2 на 2, когда все наглядно:

$$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -\lambda \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2\lambda \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -\lambda \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$$

(3) Для доказательства этого пункта воспользуемся следующим наблюдением: если $A, B \in \mathrm{M}_n(\mathbb{R})$ причем A обратима, тогда $\psi(ABA^{-1}) = \psi(B)$. Действительно,

$$\psi(ABA^{-1}) = \psi(A)\psi(B)\psi(A)^{-1} = \psi(B)\psi(A)\psi(A)^{-1} = \psi(B)$$

Мы уже знаем, что $\psi(D_n(\lambda)) = \lambda$ по условию. Надо лишь доказать, что для всех i выполнено $\psi(D_i(\lambda)) = \lambda$. Для этого достаточно представить $D_i(\lambda) = AD_{i+1}(\lambda)A^{-1}$. Возьмем в качестве $A = U_{i,i+1}$ элементарную матрицу переставляющую i и $i+1$ строки. Тогда $A^{-1} = A$. Более того, легко видеть, что $D_i(\lambda) = U_{i,i+1}D_{i+1}(\lambda)U_{i,i+1}^{-1}$. Действительно, умножение на $U_{i,i+1}$ слева переставляет i и $i+1$ строки, а умножение на $U_{i,i+1}$ справа равносильно умножению на $U_{i,i+1}^{-1}$ и оно переставляет i и $i+1$ столбцы. Для наглядности двумерный случай:

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

(2) Здесь мы воспользуемся тем, что элементарные преобразования второго типа можно выразить через элементарные преобразования первого и третьего типа, а именно, давайте проверим, что

$$U_{ij} = D_i(-1)S_{ji}(1)S_{ij}(-1)S_{ji}(1)$$

Применив ψ к этому равенству и воспользовавшись предыдущими двумя пунктами мы получаем требуемое. Однако, остается законный вопрос: а как вообще можно догадаться до такого и проверить? Вот вам рассуждение приводящее к такому ответу. Давайте последовательно применять элементарные преобразования первого и третьего типа к единичной матрице, пока не получим из нее матрицу U_{ij}. Написанное равенство означает, что надо сделать так: (1) прибавить i строку к j, (2) вычесть j строку из i, (3) прибавить i строку к j, (4) умножить i строку на -1. Давайте для наглядности это проделаем на матрицах 2 на 2. Ниже мы последовательно умножаем матрицу с левой стороны слева на матрицу, написанную над стрелкой:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \xrightarrow{\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}} \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} \xrightarrow{\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \xrightarrow{\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

□

Утверждение 29. *Пусть $\psi\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ – отображение, удовлетворяющее свойствам:*

1. *$\psi(AB) = \psi(A)\psi(B)$ для любых $A, B \in \mathrm{M}_n(\mathbb{R})$.*
2. *$\psi(D_n(\lambda)) = \lambda$ для любого ненулевого $\lambda \in \mathbb{R}$.*

И пусть $P \in \mathrm{M}_n(\mathbb{R})$ матрица с нулевой строкой. Тогда $\psi(P) = 0$.

Доказательство. Пусть в матрице P нулевой является i-я строка. Тогда $D_i(\lambda)P = P$ при любом ненулевом $\lambda \in \mathbb{R}$. Применим к этому равенству ψ и получим

$$\lambda\psi(P) = \psi(D_i(\lambda))\psi(P) = \psi(P)$$

Выберем любое ненулевое число λ отличное от 1, тогда получим, что $\psi(P)$ обязано быть нулем. □

Доказательство Утверждения 26. В начале пусть $A \in \mathrm{M}_n(\mathbb{R})$ – невырожденная матрица. Тогда мы знаем, что она является произведением элементарных матриц $A = U_1 \dots U_k$. Применим ψ к этому равенству, получим $\psi(A) = \psi(U_1)\dots\psi(U_k)$. С другой стороны по утверждению 28 получаем $\psi(A) = \det(U_1)\dots\det(U_k)$. А из мультипликативности определителя следует, что правая часть равна $\det A$. То есть ψ совпадает с det на невырожденных матрицах.

Теперь покажем, что ψ совпадает с det на всех матрицах. Пусть $A \in \mathrm{M}_n(\mathbb{R})$ – вырожденная матрица. Тогда элементарными преобразованиями строк она приводится к ступенчатому виду, то есть A можно представить в виде TB, где T – обратимая, а B имеет улучшенный ступенчатый вид. Так как A вырождена, матрица B имеет нулевую строку. Теперь применим к равенству $A = TB$ отображение ψ и det. Получим

$$\psi(A) = \psi(T)\psi(B) \quad \text{и} \quad \det(A) = \det(T)\det(B)$$

Но мы знаем по утверждению 29, что $\psi(B) = 0$. Кроме того, мы знаем, что определитель от матриц с нулевой строкой тоже равен нулю по утверждению 20. Значит $\psi(A) = 0 = \det(A)$. □

4.8 Миноры и алгебраические дополнения

Определения Пусть $B \in \mathrm{M}_n(\mathbb{R})$ – некоторая матрица с b_{ij}. Рассмотрим матрицу $D_{ij} \in \mathrm{M}_{n-1}(\mathbb{R})$ полученную из B вычеркиванием i-ой строки и j-го столбца. Определитель матрицы D_{ij} обозначается M_{ij} и называется *минором* матрицы B или $i\,j$-минором для определенности. Число $A_{ij} = (-1)^{i+j} M_{ij}$ называется *алгебраическим дополнением* элемента b_{ij} или $i\,j$-алгебраическим дополнением матрицы B.

Покажем как все это выглядит на картинках. Если мы представим матрицу B в виде

$$B = \begin{pmatrix} X_{ij} & * & Y_{ij} \\ * \ \ldots & b_{ij} & \ldots \ * \\ Z_{ij} & * & W_{ij} \end{pmatrix}$$

Тогда

$$D_{ij} = \begin{pmatrix} X_{ij} & Y_{ij} \\ Z_{ij} & W_{ij} \end{pmatrix}, \quad M_{ij} = \det \begin{pmatrix} X_{ij} & Y_{ij} \\ Z_{ij} & W_{ij} \end{pmatrix} \quad \text{и} \quad A_{ij} = (-1)^{i+j} \det \begin{pmatrix} X_{ij} & Y_{ij} \\ Z_{ij} & W_{ij} \end{pmatrix}$$

Присоединенная матрица $\hat{B}$ для B определяется как

$$\hat{B} = \begin{pmatrix} A_{11} & A_{21} & \ldots & A_{n1} \\ A_{12} & A_{22} & \ldots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \ldots & A_{nn} \end{pmatrix}$$

То есть надо в матрице B каждый элемент b_{ij} заменить на его алгебраическое дополнение A_{ij}, а потом полученную матрицу транспонировать. Полезно держать перед глазами формулу для элемента присоединенной матрицы $\hat{B}_{ij} = A_{ji}$.

Формула разложения по строке

Утверждение 30. *Пусть $B \in \mathrm{M}_n(\mathbb{R})$ – произвольная матрица. Тогда*[43]

1. *Для любой строки i верно разложение*

$$\det B = \sum_{j=1}^{n} b_{ij} A_{ij}$$

2. *Для любого столбца j верно разложение*

$$\det B = \sum_{i=1}^{n} b_{ij} A_{ij}$$

Доказательство. Мы докажем формулу для строки, для столбца она получается аналогично либо применением транспонирования к матрице. Рассмотрим i-ю строку в матрице B

$$B = \begin{pmatrix} X_{ij} & * & Y_{ij} \\ b_{i1} \ \ldots & b_{ij} & \ldots \ b_{in} \\ Z_{ij} & * & W_{ij} \end{pmatrix}$$

Эту строку можно разложить в сумму следующих строк

$$(b_{i1}, \ldots, b_{in}) = \sum_{j=1}^{n} (0, \ldots, 0, b_{ij}, 0, \ldots, 0)$$

[43] Всюду в формулах A_{ij} обозначает алгебраическое дополнение.

Теперь вычислим определитель B пользуясь линейностью по i-ой строке

$$\det B = \sum_{j=1}^{n} \det \begin{pmatrix} & \boxed{\begin{matrix} * \\ \vdots \end{matrix}} & \\ \boxed{0 \quad \dots \quad b_{ij} \quad \dots \quad 0} & & \\ & \boxed{\begin{matrix} \vdots \\ * \end{matrix}} & \end{pmatrix} \begin{matrix} X_{ij} & Y_{ij} \\ & \\ Z_{ij} & W_{ij} \end{matrix}$$

Теперь отдельно посчитаем следующий определитель

$$\det \begin{pmatrix} & \boxed{\begin{matrix} * \\ \vdots \end{matrix}} & \\ \boxed{0 \quad \dots \quad b_{ij} \quad \dots \quad 0} & & \\ & \boxed{\begin{matrix} \vdots \\ * \end{matrix}} & \end{pmatrix} \begin{matrix} X_{ij} & Y_{ij} \\ & \\ Z_{ij} & W_{ij} \end{matrix} = (-1)^{j-1} \det \begin{pmatrix} \boxed{\begin{matrix} * \\ \vdots \end{matrix}} & & \\ \boxed{b_{ij} \quad \dots \quad 0} & & \\ \boxed{\begin{matrix} \vdots \\ * \end{matrix}} & & \end{pmatrix} \begin{matrix} X_{ij} & Y_{ij} \\ & \\ Z_{ij} & W_{ij} \end{matrix} = (-1)^{j-1}(-1)^{i-1} \det \begin{pmatrix} \boxed{b_{ij}} & \boxed{\dots \quad 0} \\ \boxed{\begin{matrix} \vdots \\ * \end{matrix}} & \begin{matrix} X_{ij} & Y_{ij} \\ Z_{ij} & W_{ij} \end{matrix} \end{pmatrix}$$

В первом равенстве мы переставили j-ый столбец $j-1$ раз, чтобы переместить его на место первого столбца. Во втором равенстве мы переставили i-ю строку $i-1$ раз, чтобы переставить ее на место первой строки. Последняя матрица является блочно нижнетреугольной, а следовательно, равенство можно продолжить так

$$(-1)^{i+j} b_{ij} \det \begin{pmatrix} X_{ij} & Y_{ij} \\ Z_{ij} & W_{ij} \end{pmatrix} = b_{ij}(-1)^{i+j} M_{ij} = b_{ij} A_{ij}$$

□

Явные формулы для обратной матрицы

Утверждение 31. *Для любой матрицы $B \in \mathrm{M}_n(\mathbb{R})$ верно*

$$\hat{B}B = B\hat{B} = \det(B)E$$

Доказательство. Нам надо отдельно доказать два равенства $\hat{B}B = \det(B)E$ и $B\hat{B} = \det(B)E$. Давайте докажем второе равенство, а первое показывается аналогично (или через трюк с транспонированием).

Для доказательства $B\hat{B} = \det(B)E$ нам надо показать две вещи: (1) все диагональные элементы матрицы $B\hat{B}$ равны $\det(B)$, (2) все внедиагональные элементы равны нулю.

(1) Рассмотрим i-ый диагональный элемент в матрице $B\hat{B}$:

$$(B\hat{B})_{ii} = \sum_{j=1}^{n} b_{ij}\hat{B}_{ji} = \sum_{j=1}^{n} b_{ij}A_{ij} = \det(B)$$

Последняя формула является разложением определителя $\det(B)$ по i-ой строке из утверждения 30.

(2) Рассмотрим элемент на позиции $i\,j$ для $i \neq j$:

$$(B\hat{B})_{ij} = \sum_{k=1}^{n} b_{ik}\hat{B}_{kj} = \sum_{k=1}^{n} b_{ik}A_{jk}$$

Нам надо показать, что последнее выражение равно нулю. Давайте рассмотрим матрицу B и заменим в ней j-ю строку на i-ю, все остальные оставим нетронутыми. Обозначим полученную матрицу через B'. Тогда

$$B' = \begin{pmatrix} * & \dots & * & \dots & * \\ \boxed{b_{i1} & \dots & b_{ik} & \dots & b_{in}} \\ * & \dots & * & \dots & * \\ \boxed{b_{i1} & \dots & b_{ik} & \dots & b_{in}} \\ * & \dots & * & \dots & * \end{pmatrix}$$

Давайте посчитаем определитель B' двумя способами. С одной стороны $\det(B') = 0$ так как в матрице есть две одинаковые строки. С другой стороны, давайте разложим определитель $\det(B')$ по j-ой строке

$$\det(B') = \sum_{k=1}^{n} b_{ik}A_{jk}$$

Что и требовалось доказать. □

В качестве непосредственного следствия этого утверждения получаем явные формулы обратной матрицы.[44]

Утверждение 32 (Явные формулы обратной матрицы)**.** *Пусть $B \in \mathrm{M}_n(\mathbb{R})$ – обратимая матрица, тогда*

$$B^{-1} = \frac{1}{\det(B)}\hat{B}$$

Заметим, что в случае матрицы 2 на 2 формулы принимают следующий вид

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

4.9 Формулы Крамера

Пусть $A \in \mathrm{M}_n(\mathbb{R})$ – произвольная матрица и $b \in \mathbb{R}^n$ – столбец. Рассмотрим систему линейных уравнений $Ax = b$. Давайте в матрице A i-ый столбец заменим на b, а остальные столбцы оставим как есть. Обозначим полученную матрицу через $\bar{A}_i$. Определим $\Delta = \det(A)$ и $\Delta_i = \det(\bar{A}_i)$.

Мы знаем, что данная система имеет единственное решение для любого b тогда и только тогда, когда матрица A обратима. Следующее утверждение дает явные формулы для координат решения системы в этом случае.

Утверждение 33 (Формулы Крамера)**.** *Пусть $A \in \mathrm{M}_n(\mathbb{R})$, $x, b \in \mathbb{R}^n$ и выполнено равенство $Ax = b$. Тогда $\Delta \cdot x_i = \Delta_i$ для любого i.*[45]

Доказательство. Рассмотрим матрицу A как строку из столбцов $A = (A_1|\ldots|A_n)$, где A_i – столбцы матрицы A. Тогда равенство $Ax = b$, пользуясь блочными формулами, можно переписать так $x_1A_1 + \ldots + x_nA_n = b$. Давайте посчитаем определитель $\bar{A}_i$, пользуясь последним равенством.

$$\det(\bar{A}_i) = \det(A_1|\ldots|b|\ldots|A_n) = \det(A_1|\ldots|\sum_{k=1}^{n} x_kA_k|\ldots|A_n) = \sum_{k=1}^{n} x_k \det(A_1|\ldots|\overset{i}{A_k}|\ldots|A_n)$$

В последней формуле, если $k \neq i$, то слагаемое имеет два одинаковых столбца A_i. Потому остается только одно слагаемое для $k = i$. Получаем

$$\det(\bar{A}_i) = x_i \det(A_1|\ldots|A_i|\ldots|A_n) = x_i \det(A)$$

Что и требовалось. □

Заметим, что если $\Delta = \det(A) \neq 0$, то имеется единственное решение системы $Ax = b$ для любой правой части b и координаты этого решения заданы по формулам $x_i = \frac{\Delta_i}{\Delta}$. Однако, если $\Delta = \det(A) = 0$, то либо решений бесконечное число, либо их вообще нет. В этом случае единственная информация из формул Крамера это: $\Delta_i = 0$.

4.10 Характеристический многочлен

Пусть $A \in \mathrm{M}_n(\mathbb{R})$ – произвольная квадратная матрица и $\lambda \in \mathbb{R}$. Рассмотрим функцию $\chi_A(\lambda) = \det(\lambda E - A)$.

Утверждение 34. *Пусть $A \in \mathrm{M}_n(\mathbb{R})$. Тогда верно*

1. *Функция $\chi_A(\lambda)$ является многочленом степени n со старшим коэффициентом 1.*
2. *Для произвольного числа λ верно, что $\lambda \in \mathrm{spec}_{\mathbb{R}} A$ тогда и только тогда, когда $\chi_A(\lambda) = 0$.*[46]

[44] Заметим, что для формулы требуется условие $\det(B) \neq 0$. Однако, матрица обратима тогда и только тогда, когда $\det(B) \neq 0$. Один из способов это показать – применить det к равенству $BB^{-1} = E$ и увидеть, что $\det(B)\det(B^{-1}) = 1$. А в обратную сторону – явные формулы.

[45] Здесь x_i – координаты вектора x.

[46] Аналогичное утверждение верно и для $\mathrm{spec}_{\mathbb{C}} A$.

Доказательство. (1) Давайте посмотрим на явную формулу определителя

$$\det B = \sum_{\sigma \in \mathrm{S}_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} \dots b_{n\sigma(n)}$$

Заметим, что данное выражение является многочленом от коэффициентов матрицы A, причем все его слагаемые имеют степень n. Теперь, когда мы считаем характеристический многочлен, мы находим $\det(\lambda E - A)$. То есть вместо b_{ii} мы должны подставить $\lambda - a_{ii}$, а вместо b_{ij} взять $-a_{ij}$ (при $i \neq j$). То есть мы в многочлен от многих переменных подставляем либо числа, либо линейный многочлен от λ. Понятно, что результатом будет многочлен от λ причем степени уж точно не больше n. Теперь давайте поймем какая будет у него степень и старший коэффициент.

$$\lambda E - A = \begin{pmatrix} \lambda - a_{11} & \dots & -a_{1n} \\ \vdots & \ddots & \vdots \\ -a_{n1} & \dots & \lambda - a_{nn} \end{pmatrix}$$

Ясно, что максимальная степень по λ может вылезти только из слагаемого являющегося произведением диагональных элементов – $(\lambda - a_{11}) \dots (\lambda - a_{nn})$. А его старший член λ^n. Вот и все.

(2) Вспомним, что $\lambda \in \operatorname{spec}_{\mathbb{R}} A$ тогда и только тогда, когда $A - \lambda E$ – необратимая матрица или что то же самое, $\lambda E - A$ – необратимая матрица. Матрица необратима тогда и только тогда, когда ее определитель ноль. Потому $\lambda \in \operatorname{spec}_{\mathbb{R}} A$ тогда и только тогда, когда $\det(\lambda E - A) = 0$, то есть $\chi_A(\lambda) = 0$. Что и требовалось. □

Определение 35. Для произвольной матрицы $A \in \mathrm{M}_n(\mathbb{R})$ многочлен $\chi_A(\lambda)$ называется *характеристическим многочленом* матрицы A.

Явные формулы для коэффициентов характеристического многочлена Вначале давайте введем некоторые обозначения. Пусть $A \in \mathrm{M}_n(\mathbb{R})$ – некоторая матрица. Рассмотрим произвольное k элементное подмножество в множестве чисел от 1 до n заданное в виде $i_1, \dots, i_k$[47] Вычеркнем из матрицы A столбцы и строки с этими номерами и обозначим полученную матрицу через $R_{i_1,\dots,i_k}$. Графически эта процедура выглядит так:

$$\begin{array}{c} \\ i_1 \\ \vdots \\ i_k \end{array} \begin{array}{c} \begin{array}{cccccc} & i_1 & \dots & i_k & & \end{array} \\ \left(\begin{array}{cccccc} R_{1\,1} & a_{1i_1} & \dots & a_{1i_k} & R_{1\,k+1} \\ a_{i_1 1} \ \dots & & & & \dots \ a_{i_1 n} \\ \vdots & & \ddots & & \vdots \\ a_{i_k 1} \ \dots & & & & \dots \ a_{i_k n} \\ R_{k+1\,1} & \vdots & \dots & \vdots & R_{k+1\,k+1} \\ & a_{ni_1} & & a_{ni_k} & \end{array} \right) \end{array} \mapsto R_{i_1,\dots,i_k} = \begin{pmatrix} R_{1\,1} & \dots & R_{1\,k+1} \\ \vdots & \ddots & \vdots \\ R_{k+1\,1} & \dots & R_{k+1\,k+1} \end{pmatrix} \in \mathrm{M}_{n-k}(\mathbb{R})$$

Пользуясь этими обозначениями покажем следующее.

Утверждение 36. *Пусть $A \in \mathrm{M}_n(\mathbb{R})$ и его характеристический многочлен имеет вид*

$$\chi_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$$

Тогда

1. *В обозначениях выше, для коэффициентов a_k верна следующая формула*[48]

$$a_k = (-1)^{n-k} \left(\sum_{i_1 < \dots < i_k} \det R_{i_1,\dots,i_k} \right)$$

[47] Здесь предполагается, что $i_1 < \dots < i_k$.

[48] Заметим, что эта формула также имеет смысл при $k = 0$ и при $k = n$. Если $k = 0$, то множество индексов пусто $\varnothing$ и $R_{\varnothing} = A$, потому формула превращается в равенство $a_0 = (-1)^n \det A$. При условии $k = n$, мы вычеркиваем все строки из матрицы и в этом случае $R_{1,\dots,n} \in \mathrm{M}_0(\mathbb{R})$. Такого объекта не существует, но мы можем для удобства считать, что в этом случае формула означает $\det R_{1,\dots,n} = 1$.

2. $a_0 = (-1)^n \det A$.

3. $a_{n-1} = -\operatorname{tr} A$.

Доказательство. (1) Введем обозначения для столбцов матрицы $A = (A_1|\dots|A_n)$ и пусть $e_i \in \mathbb{R}^n$ – столбец, у которого i-я координата равна 1, а все остальные 0. Нам надо посчитать $\det(\lambda E - A) = (-1)^n \det(A - \lambda E)$. Тогда,

$$\det(A - \lambda E) = \det(A_1 - \lambda e_1|\dots|A_n - \lambda e_n)$$

Теперь надо раскрыть последний определитель по полилинейности.[49] Всего у нас будет 2^n слагаемых, каждое из которых – это определитель матрицы состоящей из столбцов A_i или $-\lambda e_j$, стоящих вперемешку.

Давайте для определенности считать, что у нас $n = 5$, тогда мы считаем

$$\det(A_1 - \lambda e_1|A_2 - \lambda e_2|A_3 - \lambda e_3|A_4 - \lambda e_4|A_5 - \lambda A_5)$$

Среди слагаемых давайте посмотрим на слагаемое, содержащее 2 столбца матрицы A и 3 столбца вида $-\lambda e_i$, например, такое

$$\det(A_1| - \lambda e_2|A_3| - \lambda e_4| - \lambda e_5) = \det\begin{pmatrix} a_{11} & 0 & a_{13} & 0 & 0 \\ a_{21} & -\lambda & a_{23} & 0 & 0 \\ a_{31} & 0 & a_{33} & 0 & 0 \\ a_{41} & 0 & a_{43} & -\lambda & 0 \\ a_{51} & 0 & a_{53} & 0 & -\lambda \end{pmatrix}$$

Давайте последовательно разлагать этот определитель по 2-ому, 4-ому и 5-ому столбцам. Обратим внимание, что $-\lambda$ всегда будут стоять на диагонали, потому знаки всех алгебраических дополнений будут положительными:

$$\det\begin{pmatrix} a_{11} & 0 & a_{13} & 0 & 0 \\ a_{21} & -\lambda & a_{23} & 0 & 0 \\ a_{31} & 0 & a_{33} & 0 & 0 \\ a_{41} & 0 & a_{43} & -\lambda & 0 \\ a_{51} & 0 & a_{53} & 0 & -\lambda \end{pmatrix} = (-\lambda)\det\begin{pmatrix} a_{11} & a_{13} & 0 & 0 \\ a_{31} & a_{33} & 0 & 0 \\ a_{41} & a_{43} & -\lambda & 0 \\ a_{51} & a_{53} & 0 & -\lambda \end{pmatrix} = (-\lambda)^2\det\begin{pmatrix} a_{11} & a_{13} & 0 \\ a_{31} & a_{33} & 0 \\ a_{51} & a_{53} & -\lambda \end{pmatrix} = (-\lambda)^3\det\begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix}$$

В общем случае слагаемое с k столбцами вида $-\lambda e_i$ является определителем матрицы вида

$$\begin{array}{c} \\ i_1 \\ \vdots \\ i_k \\ \\ \end{array}\begin{array}{c} \begin{array}{ccccc} & & i_1 & \dots & i_k \end{array} \\ \begin{pmatrix} R_{1\,1} & \boxed{\begin{matrix} 0 \\ \vdots \end{matrix}} & \dots & \boxed{\begin{matrix} 0 \\ \vdots \end{matrix}} & R_{1\,k+1} \\ \boxed{\begin{matrix} a_{i_1 1} & \dots \end{matrix}} & \boxed{-\lambda} & & & \boxed{\begin{matrix} \dots & a_{i_1 n} \end{matrix}} \\ \vdots & & \ddots & & \vdots \\ \boxed{\begin{matrix} a_{i_k 1} & \dots \end{matrix}} & & & \boxed{-\lambda} & \boxed{\begin{matrix} \dots & a_{i_k n} \end{matrix}} \\ R_{k+1\,1} & \boxed{\begin{matrix} \vdots \\ 0 \end{matrix}} & \dots & \boxed{\begin{matrix} \vdots \\ 0 \end{matrix}} & R_{k+1\,k+1} \end{pmatrix} \end{array} = I_{i_1,\dots,i_k}$$

Раскладывая этот определитель по столбцам $i_1, \dots, i_k$ мы получаем

$$\det I_{i_1,\dots,i_k} = (-\lambda)^k \det R_{i_1,\dots,i_k}$$

Слагаемые при λ^k вылезут, когда ровно k столбцов имеют вид $-\lambda e_i$. Остается не забыть, что мы считали $(-1)^n\chi_A(\lambda)$.

(2) Свободный член многочлена $\chi_A(\lambda)$ всегда равен $\chi_A(0) = \det(0E - A) = \det(-A) = (-1)^n \det(A)$, что и требовалось.

(3) Для подсчета a_{n-1} воспользуемся формулой, получим[50]

$$a_{n-1} = (-1)^{n-(n-1)} \sum_{i=1}^{n} \det R_{1,\dots,\hat{i},\dots,n}$$

[49] Думать про это выражение надо так: надо мысленно заменить вертикальные черточки умножением и считать, что мы раскрываем скобки в произведении.

[50] Здесь $\hat{i}$ означает, что индекс i пропущен.

Но заметим, что $R_{1,\dots,\hat{i},\dots,n} = a_{ii}$, а значит предыдущее равенство превращается в

$$a_{n-1} = (-1)^{n-(n-1)} \sum_{i=1}^{n} a_{ii} = -\operatorname{tr} A$$

□

Примеры

1. Если $A \in \mathrm{M}_1(\mathbb{R})$, то есть $A = a \in \mathbb{R}$ – число, то $\chi_A(\lambda) = \lambda - a$.
2. Если $A \in \mathrm{M}_2(\mathbb{R})$, то $\chi_A(\lambda) = \lambda^2 - \operatorname{tr} A\lambda + \det A$.
3. Если $A \in \mathrm{M}_3(\mathbb{R})$, то $\chi_A(\lambda) = \lambda^3 - \operatorname{tr} A\lambda^2 + a_1\lambda - \det A$, где

$$a_1 = \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} + \det \begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix} + \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Стоит отметить, что считать характеристические многочлены от матриц большего размера через эти формулы практически не целесообразно. Максимальный разумный размер – матрица 4 на 4. Самый быстрый способ остается алгоритм Гаусса для подсчета определителя $\det(\lambda E - A)$ с символьными коэффициентами.

4.11 Теорема Гамильтона-Кэли

Многочлены с матричными коэффициентами Обозначим через $\mathrm{M}_n(\mathbb{R})[t]$ множество многочленов от переменной t имеющих матричные коэффициенты из $\mathrm{M}_n(\mathbb{R})$, т.е.

$$\mathrm{M}_n(\mathbb{R})[t] = \{A_0 + A_1 t + \ldots + A_k t^k \mid A_i \in \mathrm{M}_n(\mathbb{R})\}$$

здесь t – формальная переменная, которая представляет собой неизвестное число. Про эти многочлены надо думать как про картинки. Такие картинки можно складывать и умножать по формулам известным для многочленов с обычными числовыми коэффициентами:

- Сумма.

$$\left(\sum_i A_i t^i\right) + \left(\sum_j B_j t^j\right) = \sum_i (A_i + B_i) t^i$$

- Произведение.

$$\left(\sum_i A_i t^i\right)\left(\sum_j B_j t^j\right) = \sum_k \left(\sum_{s+t=k} A_s B_t\right) t^k$$

Надо лишь отметить, что в произведении нельзя переставлять местами A_s и B_t, так как матрицы вообще говоря не перестановочны.

Замечание Для тех, кто не хочет воспринимать многочлены как картинки, я хочу сформулировать важное замечание. Пусть $f, g \in \mathrm{M}_n(\mathbb{R})[t]$ – два многочлена с матричными коэффициентами. Предположим, что для любого $a \in \mathbb{R}$ имеет место равенство $f(a) = g(a)$, тогда многочлены f и g равны, то есть совпадают все их коэффициенты.

Действительно, если $f = \sum_k A_k t^k$ и $g = \sum_k B_k t^k$, то равенство $f(a) = g(a)$ означает, что матрицы $\sum_k A_k a^k$ и $\sum_k B_k a^k$ равны. Но это значит, что все их коэффициенты равны, то есть для любых i, j между 1 и n выполняется

$$\sum_k (A_k)_{i\,j} a^k = \sum_k (B_k)_{i\,j} a^k$$

Но это уже равенство обычных числовых многочленов для любого числа $a \in \mathbb{R}$. Отсюда следует, что коэффициенты этих многочленов равны, то есть $(A_k)_{i\,j} = (B_k)_{i\,j}$ для всех i и j. Но в этом случае матрицы A_k и B_k просто равны между собой, что и требовалось.

Таким образом, если вам не нравится работать с произвольным параметром-«картинкой» t или λ в выражениях ниже, вы можете объяснить многие свои сомнения через рассуждение выше. То есть вы можете предполагать, что мы доказываем равенство для произвольного $t \in \mathbb{R}$, а потом получаем равенство матричных коэффициентов в многочленах.

Подстановка матрицы в многочлен Теперь для произвольного многочлена $f \in \mathrm{M}_n(\mathbb{R})[t]$ и матрицы $D \in \mathrm{M}_n(\mathbb{R})$ определим подстановку матрицы D в многочлен f справа:

$$f(D) = A_0 + A_1 D + \ldots + A_k D^k$$

т.е. мы вместо t подставляем всюду матрицу D. Аналогично, можно определить левую подстановку:

$$(D)f = A_0 + DA_1 + \ldots + D^k A_k$$

Надо отметить, что вообще говоря $f(D) \neq (D)f$. Мы всегда будем пользоваться только правой подстановкой.

Свойства подстановки Пусть $f, g \in \mathrm{M}_n(\mathbb{R})[t]$ – два многочлена и $D \in \mathrm{M}_n(\mathbb{R})$ – некоторая матрица. Сделаем следующие замечания:

1. Всегда верно равенство
$$f(D) + g(D) = (f + g)(D)$$
2. Для произведения вообще говоря выполнено
$$f(D)g(D) \neq (fg)(D)$$
Действительно, возьмем $f(t) = t$, $g(t) = Bt$, тогда $(fg)(t) = Bt^2$. В этом случае $f(D)g(D) = DBD$, а $(fg)(D) = BDD$. Вообще говоря, имеем $DBD \neq BDD$ если матрицы B и D не коммутируют.
3. Если D коммутирует со всеми коэффициентами матрицы g, то верно равенство
$$f(D)g(D) = (fg)(D)$$
Это видно непосредственно из определения умножения и подстановки.

Теорема Теперь мы готовы к формулировке и доказательству полезного результата.

Утверждение 37 (Теорема Гамильтона-Кэли)**.** *Пусть* $A \in \mathrm{M}_n(\mathbb{R})$. *Тогда* $\chi_A(A) = 0$.

Прежде чем доказывать теорему, давайте объясним что в точности утверждает теорема и почему дурацкие доказательства не работают. Давайте поймем, что имеется в виду в утверждении. Сначала мы должны посчитать характеристический многочлен, то есть явно посчитать его коэффициенты и представить в следующем виде[51]

$$\chi_A(\lambda) = \det(\lambda E - A) = \sum_{k=0}^{n} a_k \lambda^k$$

И уже после такого вычисления, надо подставить вместо λ матрицу A и получить

$$\chi_A(A) = \sum_{k=0}^{n} a_k A^k = 0$$

Как раз не очевидно, что последнее равенство выполняется.

Надо обратить внимание, что если мы подставим матрицу A вместо λ в выражение $\det(\lambda E - A)$ мы действительно получим 0. Однако, равенство $\det(\lambda E - A) = \sum_{k=0}^{n} a_k \lambda^k$ верно только если λ является числом и потому вообще говоря не понятно почему это равенство должно сохраниться при $\lambda = A$.[52] Давайте продемонстрируем это на матрице 2 на 2. Пусть

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Тогда

$$\det(A - \lambda E) = \det\left(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} - \lambda E\right) = \det\begin{pmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{pmatrix}$$

[51] Тут коэффициенты a_k вычисляются по формулам из утверждения 36.

[52] На самом деле оно вообще говоря не будет верным для произвольных матриц потому что выражение слева всегда дает число, а выражение справа – матрицу.

Так вот, последнее равенство верно если λ является числом. Если же λ является матрицей, то оно непонятно, что значит. Можно понимать правую часть как блочную матрицу 2 на 2 из блоков 2 на 2 (т.е. всего 4 на 4), но тогда это просто не верное равенство. Это рассуждение можно докрутить до верного, но тогда в правой части надо использовать вместо определителя его более хитрую блочную версию. Подобное рассуждение растет из коммутативной алгебры, где доказательство естественным образом сводится к формулам Крамера, но для его освоения надо знать, что такое кольца и модули. Мы же пойдем чуть более простым путем.[53]

Доказательство. Рассмотрим матрицу $\lambda E - A$, где λ – неизвестное число. Введем следующее обозначение $R(\lambda) = \widehat{\lambda E - A}$.

Заметим, что каждый коэффициент $R(\lambda)$ является многочленом от λ, т.е. $R(\lambda) = (r_{ij}(\lambda))$ и $r_{ij}(\lambda)$ – многочлен. То есть $r_{ij}(\lambda) = \sum_k r_{ijk}\lambda^k$. Тогда $R(\lambda) = \sum_k R_k\lambda^k$, где $R_k = (r_{ijk})$. То есть $R(\lambda) \in \mathrm{M}_n(\mathbb{R})[\lambda]$. Для ясности, давайте проиллюстрируем сказанное на следующем примере.

$$\begin{pmatrix} 5-\lambda+2\lambda^2 & 3 \\ 4-\lambda & 2+\lambda \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ 4 & 2 \end{pmatrix} + \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix}\lambda + \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\lambda^2$$

Теперь применим формулы для перемножения матрицы с ее присоединенной из утверждения 31 для матрицы $\lambda E - A$, получим

$$(\lambda E - A)R(\lambda) = R(\lambda)(\lambda E - A) = \det(\lambda E - A)E = \chi_A(\lambda)E$$

Нас интересует только равенство

$$R(\lambda)(\lambda E - A) = \chi_A(\lambda)E$$

Тогда рассмотрим многочлены $f(\lambda) = R(\lambda)$, $g(\lambda) = \lambda E - A$. В этом случае $(fg)(\lambda) = \chi_A(\lambda)E$.[54] Возьмем в качестве матрицы D матрицу A. Заметим, что она коммутирует с коэффициентами g, потому что это E и $-A$. Значит верно равенство $f(D)g(D) = (fg)(D)$. Последнее означает

$$0 = R(A)(AE - A) = \chi_A(A)E = \chi_A(A)$$

Что и требовалось доказать. □

[53] Ну как сказать простым...

[54] Если вы в этом месте в ужасе от того, что я творю, я предлагаю заметить, что мы по сути доказали равенство значений многочленов с матричными коэффициентами при любом $\lambda \in \mathbb{R}$. А теперь можно воспользоваться замечанием, что отсюда следует совпадение многочленов как картинок, то есть все их матричные коэффициенты тоже совпадают.

5 Комплексные числа

5.1 Идея

Почему нам вдруг не хватает вещественных чисел? Давайте вспомним, а откуда получились вещественные числа? Для начала у нас есть натуральные числа: $\mathbb{N} = \{1, 2, 3, \ldots\}$. Которые мы беззаботно складывали и умножали. Но как только нам захотелось посчитать $5 - 8$, как нам понадобились другие числа – целые $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. И мы опять жили долго и счастливо, пока на не пришлось делить $2/3$ и тут пришлось построить рациональные числа $\mathbb{Q} = \{\frac{p}{q} \mid p, q \in \mathbb{Z},\ q \neq 0\}$. Вещественные нам пригодились, когда надо было решить уравнение $x^2 = 2$. Тогда пришлось добавить $\sqrt{2}$, а заодно и кучу других полезных чисел. Однако, этого опять оказалось мало и уравнение $x^2 + 1 = 0$ не решается в вещественных числах. При этом давайте заметим, что добавляя новые числа, операции над старыми мы не меняли. Мы добавили целые, рациональные, вещественные, а натуральные как складывались и умножались по старым правилам, так и продолжают складываться и умножаться.

А какие-же числа мы хотим получить в идеале. Прежде всего хочется решать уравнения вида $f(x) = 0$, где $f \in \mathbb{R}[x]$, всегда, когда это возможно. Например, если $f = 1$, то решить такое уравнение по понятным причинам не возможно, но вот если $\deg f > 0$, то очень хочется иметь решение. Новые числа должны содержать все вещественные как подмножество. Но кроме этого, мы хотим уметь делать все арифметические операции с новыми числами, да еще так, чтобы старые операции не изменились. И еще хочется по возможности быть экономными. Вдруг, можно построить много разных лишних чисел (например, когда нам понадобились рациональные числа, мы могли по наивности и безрассудству сразу же построить вещественные, но обошлись более экономным вариантом в виде рациональных чисел).

Для того, чтобы формализовать идеи выше, нам надо строго сказать, а какой математической структурой должны являться новые числа. Такими структурами являются поля. Потом надо объяснить как правильно обращаться с полями, что с ними можно делать, как их сравнивать между собой. И как только у нас появился зверинец полей, мы можем найти в нем поле комплексных чисел, как самое лучшее, которое только возможно среди тех, что удовлетворяют нашим запросам.

Сейчас нас ждет очередное абстрактное определение. Напомню, что оно всегда состоит из двух частей: в первой части сказано какие у нас данные, а во второй – каким аксиомам эти данные подчиняются.

Определение 38 (Поле)**.** Поле это следующий набор данных: $(F, +, \cdot)$, где

- F – некоторое множество. Элементы этого множества называются числами.
- $+\colon F \times F \to F$, $(x, y) \mapsto x + y$ – некоторая операция называемая сложением.
- $\cdot\colon F \times F \to F$, $(x, y) \mapsto xy$ – некоторая операция называемая умножением.

Эти данные должны подчиняться следующим десяти аксиомам:

1. **Ассоциативность сложения** Для любых элементов $x, y, z \in F$ выполнено $x + (y + z) = (x + y) + z$.
2. **Существования нейтрального по сложению** Существует такой элемент $0 \in F$ такой, что для любого $x \in F$ верно $x + 0 = 0 + x = x$. Такой элемент называется нулем.
3. **Существование обратного по сложению** Для любого $x \in F$ существует элемент $-x \in F$ такой, что $x + (-x) = (-x) + x = 0$. Такой элемент называется противоположным.
4. **Коммутативность сложения** Для любых элементов $x, y \in F$ верно $x + y = y + x$.
5. **Ассоциативность умножения** Для любых элементов $x, y, z \in F$ верно $x(yz) = (xy)z$.
6. **Существование нейтрального по умножению** Существует такой элемент $1 \in F$, что для любого $x \in F$, верно $x1 = 1x = x$. Такой элемент называется единицей.
7. **Существование обратного по умножению** Для любого элемента $x \in F \setminus \{0\}$ существует элемент $x^{-1} \in F$ такой, что $xx^{-1} = x^{-1}x = 1$. Такой элемент называется обратным к x.
8. **Коммутативность умножения** Для любых элементов $x, y \in F$ верно $xy = yx$.
9. **Дистрибутивность** Для любых элементов $x, y, z \in F$ верно $x(y + z) = xy + xz$ и $(x + y)z = xz + yz$.
10. **Нетривиальность** $0 \neq 1$.

Замечания Давайте сделаем несколько полезных замечаний.

1. Аксиомы сгруппированы следующим образом: (1–4) аксиомы на сложение, (5–8) аксиомы на умножение, (9) связь между сложением и умножением, (10) нетривиальность. Причем аксиомы (1–4) и (5–8) идут по одному и тому же шаблону: ассоциативность, нейтральный элемент, обратный, коммутативность. НО стоит отметить важную разницу между аксиомами (3) и (7). По сложению обратный должен быть для любого элемента, по умножению только для ненулевого. В частности, аксиому (7) нельзя сформулировать без аксиомы (2).

2. В аксиомах (2) и (6) не требуется единственность нуля и единицы. Однако, можно показать, что если ноль существует, то он обязательно единственный, аналогично с единицей. Действительно, если у нас есть два нуля 0_1 и 0_2, то рассмотрим их сумму 0_1+0_2. Так как 0_1 является нулем, то $0_1+0_2 = 0_2$. Так как 0_2 является нулем, то $0_1+0_2 = 0_1$. Значит оба нуля совпадают. Аналогично проверяется единственность единицы. Потому в силу однозначности эти элементы обозначаются 0 и 1.

3. В аксиомах (3) и (7) не требуется единственность обратного. Однако, можно показать, что для любого x существует единственный $-x$ и единственный x^{-1}. Действительно, если для элемента $x \in F$ есть два элемента $y, z \in F$ таких, что

   $$x + y = y + x = 0 \quad \text{и} \quad x + z = z + x = 0$$

   Тогда рассмотрим выражение

   $$(y + x) + z = y + (x + z)$$

   Его левая часть вычисляется в $0 + z = z$, а правая вычисляется в $y + 0 = y$. А значит y и z совпадают. То есть обратный по сложению будет один, аналогично с обратным по умножению. Именно по причине однозначности им даются такие имена. В частности однозначно определено число -1.

4. Мы привыкли к всяким замечательным свойствам, которым подчиняются числа 0 и 1. Например: $x0 = 0$ или $(-1)x = -x$ для любого x. Оказывается, что их можно доказать пользуясь аксиомами. Попробуйте сделать это.

5. Давайте рассмотрим множество $F = \{\cdot\}$ состоящее из одной точки. Тогда на таком множестве существует единственная операция, положим сложение и умножение равными ей. Тогда данный набор данных удовлетворяет всем аксиомам поля кроме последней. Здесь ноль равен единице и вообще все элементы равны друг другу и ничего кроме нуля (единицы) у нас нет. На самом деле, если ноль равен единице, то никаких других структур мы не построим. Действительно, предположим $(F, +, \cdot)$ удовлетворяет всем аксиомам с первой по девятую, а вместо десятой – ее отрицание $1 = 0$. Тогда для любого элемента $a \in F$ имеем $a = a \cdot 1 = a \cdot 0$. Давайте покажем, что $a \cdot 0 = 0$ для любого $a \in F$. Рассмотрим равенство $0 + 0 = 0$, которое следует из определения нуля. Умножим его на a, получим $a \cdot (0 + 0) = a \cdot 0$. Раскроем скобки и получим $a \cdot 0 + a \cdot 0 = a \cdot 0$. Теперь прибавим к обеим частям равенства элемент $-(a \cdot 0)$. Получим

   $$a \cdot 0 + a \cdot 0 + -(a \cdot 0) = a \cdot 0 + -(a \cdot 0)$$

   Что равносильно $a \cdot 0 + 0 = 0$, а значит $a \cdot 0 = 0$, что и требовалось. Потому последняя аксиома нужна для того, чтобы исключить именно этот дурацкий пример.

Примеры Как только вам скормили абстрактное определение, первым делом нужны примеры. Он помогут по-новому взглянуть на старых знакомых и разобраться с тем, а как вообще задавать эти самые новые объекты.

1. Рациональные и вещественные числа $\mathbb{Q}$ и $\mathbb{R}$ с обычными операциями.

2. Множество $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\} \subseteq \mathbb{R}$ с обычными операциями является примером поля, которое лежит между $\mathbb{Q}$ и $\mathbb{R}$.

3. Рациональные функции $\mathbb{R}(x) = \{\frac{f}{g} \mid f, g \in \mathbb{R}[x]\}$. Давайте думать про рациональные функции как про картинки. Тогда на них определены формальные операции сложения и умножения. Относительно этих операций они являются полем.

4. Теперь время экзотики $\mathbb{F}_2 = \{0, 1\}$, а операции берутся по модулю 2. Можно проверить, что и этот товарищ является полем. Это очень важное поле для computer science. Оно и его аналоги используются в теории кодирования, восстановления сигнала, архивирования и т.д.

Определение 39 (Подполе)**.** Пусть K и L – два поля причем $K \subseteq L$. Тогда K называется подполем в L, если операции сложения и умножения из L ограниченные на K дают сложение и умножение на K соответственно. Более подробно, пусть $+_L$ и $\cdot_L$ – сложение и умножение на L, а $+_K$ и $\cdot_K$ – сложение и умножение на K. Тогда для любых элементов $x, y \in K$ верно: $x +_L y = x +_K y$ и $x \cdot_L y = x \cdot_K y$.

По простому подполе – это подмножество чисел в нашем поле, которое само является полем относительно операций из большего поля. То есть нет никакой разницы какие операции использовать в подполе K: операции из K или операции из L, так как между ними нет разницы.

5.2 Абстрактное определение комплексных чисел

Определение 40. Пусть поле $\mathbb{C}$ обладает следующими свойствами:

1. $\mathbb{R} \subseteq \mathbb{C}$ – подполе, т.е. поле $\mathbb{C}$ содержит вещественные числа и операции сложения и умножения ограничиваются на $\mathbb{R}$ в обычные операции сложения и умножения.[55]

2. Для любого не константного многочлена $f \in \mathbb{C}[x]$ существует корень $\alpha \in \mathbb{C}$, т.е. $f(\alpha) = 0$.

3. Поле $\mathbb{C}$ является минимальным полем, удовлетворяющим предыдущим свойствам, т.е. для любого поля F такого, что $\mathbb{R} \subseteq F \subseteq \mathbb{C}$ если F обладает двумя предыдущими свойствами, то $F = \mathbb{C}$.

Тогда оно называется полем комплексных чисел.

Стоит сделать важное замечание. Из определения вообще говоря не следует, что подобное поле существует, но даже если оно и существует, то не понятно, вообще говоря, единственно ли оно. Как можно ожидать, такое поле обязательно существует и оказывается, что оно единственно в некотором естественном смысле. Потому идейно к определению выше надо относиться так: это набросок тех свойств, которые мы хотели бы получить от нашего поля, а дальше вся работа заключается в том, чтобы показать, во-первых, что таких свойств добиться можно и построить необходимое поле, а во-вторых, что как бы мы ни построили поле комплексных чисел, всегда получится одно и то же.

Как вы понимаете построить поле удовлетворяющее свойству (2) – не простая задача. Потому обычно поступают по-другому. Мы построим поле с более слабым свойством, что существует решение только у уравнения $x^2 + 1 = 0$. А уже потом покажем, что подобное поле удовлетворяет более сильному условию (2) из определения. Единственность мы с вами доказывать не будем в силу того, что эта тема будет затронута в курсе алгебры в общем виде.

5.3 Две модели комплексных чисел

В этом разделе я построю две модели комплексных чисел. Для того чтобы различать эти модели, я для начала буду обозначать их $\mathbb{C}_1$ и $\mathbb{C}_2$. Но как только мы поймем, что это одно и то же, мы будем опускать индекс и обозначать построенное поле через $\mathbb{C}$.

Символьная модель Пусть $\mathbb{C}_1$ – это множество картинок вида $a + bi$, где $a, b \in \mathbb{R}$ – вещественные числа, а i и $+$ – картинки. Множество мы определили, теперь надо определить операции сложения и умножения. Сумму картинок определим покомпонентно:

$$(a + bi) + (c + di) = (a + c) + (b + d)i, \quad \text{где} \quad a, b, c, d \in \mathbb{R}$$

Умножение определим исходя из соображений $i^2 = -1$. Тогда

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i, \quad \text{где} \quad a, b, c, d \in \mathbb{R}$$

На этом этапе необходимые данные для определения поля нами построены. Осталось дело за малым – проверить все 10 аксиом. Эту наинтереснейшую задачу я оставлю в качестве упражнения, но обязательно проверьте эти аксиомы.

[55] В терминологии ниже вложение $\mathbb{R} \to \mathbb{C}$ является гомоморфизмом полей.

Теперь $\mathbb{C}_1$ является полем. Вещественные числа в него вкладываются так: число $r \in \mathbb{R}$ идет в картинку $r+0i \in \mathbb{C}_1$. Теперь надо проверить, что сложить два вещественных числа – это все равно, что сложить их как два комплексных числа. Аналогично, умножить два вещественных числа – это все равно, что умножить их как два комплексных числа. Напоследок заметим, что комплексное число i выбрано так, чтобы оно являлось решением уравнения $x^2 + 1 = 0$.

Матричная модель Пусть $\mathbb{C}_2$ – это множество матриц вида $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right) \in \mathrm{M}_2(\mathbb{R})$. Множество построено, теперь дело за операциями. Придумывать их не надо, это будут обычные матричные сложение и умножение. Единственное, что надо проверить, это что сумма и произведение матриц из $\mathbb{C}_2$ остаются в $\mathbb{C}_2$.[56]

Как и с первой моделью, мы только что построили все необходимые данные для определения поля, теперь надо проверить аксиомы. И тут нам очень пригождаются матрицы. Почти все аксиомы будут автоматически следовать из соответствующих свойств матричных операций. Единственное, что надо проверить: коммутативность умножения и что любой ненулевой элемент обратим. Я сейчас опять поступлю не очень честно и попрошу жаждущего знаний читателя все проверить самостоятельно.[57]

Вещественные числа вкладываются в $\mathbb{C}_2$ в виде скалярных матриц, то есть $r \in \mathbb{R}$ идет в $rE \in \mathbb{C}_2$. Как мы знаем при этом операции матричного сложения и умножения превращаются в операции сложения и умножения вещественных чисел. Осталось найти решение уравнения $x^2 + 1 = 0$. Для этого заметим, что матрица $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ удовлетворяет этому уравнению.

Сравнение полей Для того чтобы сравнить различные поля и сказать, что они одинаковые или различные нам потребуется понятие изоморфизма полей.

Определение 41. Пусть F_1 и F_2 поля. Отображение $\varphi\colon F_1 \to F_2$ называется гомоморфизмом полей, если

1. $\varphi(x+y) = \varphi(x) + \varphi(y)$ для любых $x, y \in F_1$.
2. $\varphi(xy) = \varphi(x)\varphi(y)$ для любых $x, y \in F_1$.
3. $\varphi(1) = 1$.

Если φ является биекцией, то оно называется изоморфизмом.

Стоит отметить, что гомоморфизм полей всегда инъективен. Попробуйте доказать это. Кроме того, если отображение φ инъективно, то достаточно лишь проверить первые два свойства гомоморфизма, т.е. единица автоматически перейдет в единицу.[58]

Мы будем говорить что два поля изоморфны, если между ними существует изоморфизм. Про изоморфные поля надо думать, как про одинаковые поля. Действительно, что значит, что между множествами есть биекция. Это значит, что это на самом деле одно и то же множество, а биекция лишь переопределяет имена, которыми называются наши элементы. Изоморфизм кроме всего прочего сохраняет операции, это значит, что отождествив элементы наших полей, мы не различаем проделанных операций. Так как поле для нас – это множество с операциями, то значит мы не увидим никакой разницы, между полями, если в них одинаковые операции.

Задачи 42. Пусть $\varphi\colon K \to L$ – гомоморфизм полей. Покажите, что выполнены следующие вещи:

1. $\varphi(0) = 0$.
2. $\varphi(-x) = -\varphi(x)$ для любого $x \in K$.
3. Если в определении гомоморфизма оставить только свойства 1 и 2, то $\varphi(1)$ либо 0, либо 1. В частности, если φ инъективно, то $\varphi(1) = 1$.
4. $\varphi(x^{-1}) = \varphi(x)^{-1}$ для любого ненулевого $x \in K$.

[56] Это, пусть и легкое, упражнение мы оставляем на совести читателя.
[57] На самом деле я всего лишь полу-честен с вами ибо чуть ниже будут проведены все соответствующие проверки.
[58] Догадливый читатель уже сообразил, что в этом месте будет фраза: «Проверьте это».

Сравнение моделей комплексных чисел Прежде чем объяснить, что $\mathbb{C}_1$ и $\mathbb{C}_2$ – это одно и то же. Нам понадобится еще одно определение. Дело в том, что в наших полях лежат дополнительно вещественные числа и мы, когда будем сравнивать эти два поля, хотим чтобы это сравнение было согласовано в каком-то смысле с вещественными числами.

Определение 43. Пусть F_1 и F_2 – два поля, содержащие поле вещественных чисел $\mathbb{R}$, то есть $\mathbb{R} \subseteq F_1$ и $\mathbb{R} \subseteq F_2$ и операции с F_i ограничиваются на соответствующие операции на вещественных числах. Будем говорить, что $\varphi\colon F_1 \to F_2$ является изоморфизмом над $\mathbb{R}$, если

1. φ является изоморфизмом.
2. $\varphi(r) = r$ для любого вещественного числа $r \in \mathbb{R}$.

Давайте построим изоморфизм над $\mathbb{R}$ между $\mathbb{C}_1$ и $\mathbb{C}_2$. А именно: $\varphi\colon \mathbb{C}_1 \to \mathbb{C}_2$ будет действовать по правилу $a+bi \mapsto \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$. По построению очевидно, что данное отображение является биекцией. Кроме того, очевидно, что оно переводит сумму в сумму. Методом пристального взгляда проверяем, что φ сохраняет умножение. Вещественное число $r = r+0i$ переходит в матрицу $\left(\begin{smallmatrix} r & 0 \\ 0 & r \end{smallmatrix}\right) = rE$. С учетом нашего отождествления вещественных чисел с подмножествами в $\mathbb{C}_1$ и $\mathbb{C}_2$ последнее означает, что $\varphi(r) = r$ для любого $r \in \mathbb{R}$. То есть нет никакой разницы между этими двумя моделями. Причем на столько нет разницы, что при нашем отождествлении все новые числа в одной модели имеют ровно те же отношения со старыми числами, что и в другой (это по сути философия изоморфизма над $\mathbb{R}$). С этого момента мы будем обозначать любую из этих двух моделей через $\mathbb{C}$.

5.4 Простейшие свойства и операции

Комплексное сопряжение Определим следующую операцию $\bar{}\colon \mathbb{C} \to \mathbb{C}$ по правилу $z = a+bi \mapsto \bar{z} = a-bi$. На языке матричной модели эта операция соответствует транспонированию. Мы знаем, что транспонирование переводит сумму в сумму, а на произведении действует так $(AB)^t = B^tA^t$, но так как $\mathbb{C}_2$ коммутативно, то для матриц из $\mathbb{C}_2$ мы имеем $(AB)^t = A^tB^t$. Кроме того сопряжение биективно, как видно из построения и переводит вещественные числа в вещественные. Значит сопряжение является изоморфизмом $\mathbb{C}$ на $\mathbb{C}$ над $\mathbb{R}$.

Сделаем еще одно полезное замечание: на матричном языке сопряжение так же совпадает с вычислением присоединенной матрицы. Действительно,

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}^t = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} = \widehat{\begin{pmatrix} a & -b \\ b & a \end{pmatrix}}$$

У последнего замечания есть интересное философское следствие. Заметим, что сопряжение переводит i в $-i$. А так как оно является изоморфизмом, то это означает, что между i и $-i$ нет никакой разницы. То есть если мы внезапно обозначим $-i$ за j, то i превратится в $-j$ и все комплексные числа будут иметь вид $a+bj$ и в этой новой форме никто не догадается, что j это была $-i$, а не опечатка наборщика перепутавшего буквы i и j. То есть в поле комплексных чисел есть небольшая свобода выбора. Мы случайно выбрали один из корней уравнения $x^2+1=0$ за i и на самом деле нет никакой разницы какой из его корней мы так обозначим.

Вещественная и мнимая части Когда комплексное число записано в виде $z = a + bi$, где $a, b \in \mathbb{R}$, мы говорим, что это его алгебраическая форма. В таком случае число a называется его вещественной частью и обозначается $\operatorname{Re} z$, а b называется мнимой частью z и обозначается $\operatorname{Im} z$. Числа с нулевой мнимой частью – это вещественные числа, а числа с нулевой вещественной частью называются чисто мнимыми.

Заметим, что для любого числа $z \in \mathbb{C}$ верно

1. $z \in \mathbb{R}$ тогда и только тогда, когда $\bar{z} = z$.
2. $z \in i\mathbb{R}$ тогда и только тогда, когда $\bar{z} = -z$.

5.5 Геометрическая модель

Комплексные числа $\mathbb{C}$ можно отождествить с вещественной плоскостью $\mathbb{R}^2$, а именно $a + bi$ соответствует вектору на плоскости $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$. Таким образом про каждое комплексное число можно думать геометрически как про вектора. При этом сложение комплексных чисел соответствует сложению векторов на плоскости.

У каждого вектора есть длина $|z| = \sqrt{a^2+b^2}$ – эта величина называется модулем комплексного числа. В матричной модели у нас определитель, легко увидеть, что $\det z = |z|^2$. Перечислим свойства модуля в следующем утверждении.

Утверждение. *Модуль комплексного числа обладает следующими свойствами:*

1. $|-|\colon \mathbb{C} \to \mathbb{R}_+$ *является нормой, то есть*
   - $|z| \geqslant 0$ *для любого* $z \in \mathbb{C}$, *причем равенство нулю достигается тогда и только тогда, когда* $z = 0$.
   - $|\lambda z| = |\lambda||z|$ *для любого* $\lambda \in \mathbb{R}$ *и* $z \in \mathbb{C}$.
   - $|z + w| \leqslant |z| + |w|$ *для любых* $z, w \in \mathbb{C}$.
2. $z\bar{z} = |z|^2$ *для любого* $z \in \mathbb{C}$.
3. $|zw| = |z||w|$ *для любых* $z, w \in \mathbb{C}$.
4. $z^{-1} = \frac{\bar{z}}{|z|^2}$.

Доказательство. Проверку (1) я оставлю на совести читателя. (2) – это явная формула. (3) доказывается с использованием (2). А вот (4) – это явная формула для обратной матрицы, потому что в матричной модели $\bar{z}$ – это сопряженная матрица, а $|z|^2$ – это $\det(z)$. □

Тригонометрическая форма Пусть $z \in \mathbb{C}$ и пусть $z \neq 0$. Тогда мы можем сделать следующее

$$z = a + bi = \sqrt{a^2 + b^2}\left(\frac{a}{\sqrt{a^2 + b^2}} + \frac{b}{\sqrt{a^2 + b^2}}i\right)$$

У числа в скобках вещественная и мнимая часть после возведения в квадрат в сумме дают единицу, а значит они являются косинусом и синусом некоторого числа φ, а значит, z можно переписать в следующей форме

$$z = |z|(\cos\varphi + i\sin\varphi)$$

Такая запись комплексного числа называется тригонометрической. Число φ определено с точностью до $2\pi n$, $n \in \mathbb{Z}$ и называется аргументом комплексного числа. Геометрически φ – это угол между осью OX и вектором проходящим из нуля в z. Угол отсчитывается против часовой стрелки. Существует следующее удобное соглашение

$$e^{i\varphi} = \cos\varphi + i\sin\varphi$$

Оказывается, что при таком определении экспонента обладает всеми знакомыми нам свойствами.[59] В этом случае тригонометрическую форму можно записать так

$$z = |z|e^{i\varphi} = e^{\ln|z| + i\varphi}$$

Алгебраическая форма записи комплексного числа хорошо согласована со сложением, а тригонометрическая – с умножением, о чем говорит следующее.

Утверждение. *Для комплексных чисел в тригонометрической форме верны следующие формулы*

1. *Пусть* $z_1 = r_1(\cos\varphi + i\sin\varphi)$ *и* $z_2 = r_2(\cos\psi + i\sin\psi)$ *– два ненулевых комплексных числа, тогда*

$$z_1 z_2 = r_1 r_2(\cos(\varphi + \psi) + i\sin(\varphi + \psi))$$

2. *Пусть* $z_1 = r_1(\cos\varphi + i\sin\varphi)$ *и* $z_2 = r_2(\cos\psi + i\sin\psi)$ *– два ненулевых комплексных числа, тогда*

$$\frac{z_1}{z_2} = \frac{r_1}{r_2}(\cos(\varphi - \psi) + i\sin(\varphi - \psi))$$

3. **Формулы Муавра** *Пусть* $z = r(\cos\varphi + i\sin\varphi)$ *– ненулевое комплексное число, тогда*

$$z^n = r^n(\cos(n\varphi) + i\sin(n\varphi))$$

[59] Если заглянуть чуть глубже в большую науку, то окажется, что вас ждет некоторый набор чудес. Окажется, что в комплексном мире очень мало гладких функций и они очень жесткие. Это значит, что для любой вещественной гладкой функции $f\colon \mathbb{R} \to \mathbb{R}$ существует не более одной комплексной гладкой функции $\tilde{f}\colon \mathbb{C} \to \mathbb{C}$, продолжающей f, в том смысле, что $\tilde{f}(r) = f(r)$ для любой $r \in \mathbb{R}$. Потому как бы мы не продолжили нашу вещественную экспоненту в комплексный мир, все эти способы дают одно и то же.

Доказательство. 1) По определению

$$r_1(\cos\varphi + i\sin\varphi)r_2(\cos\psi + i\sin\psi) = r_1r_2(\cos\varphi\cos\psi - \sin\varphi\sin\psi + i(\sin\varphi\cos\psi + \sin\psi\cos\varphi)) =$$
$$r_1r_2(\cos(\varphi+\psi) + i\sin(\varphi+\psi))$$

2) Можно проверить двумя способами. Либо воспользоваться тем, что $1/z_2 = \bar{z}_2/|z_2|^2$, либо домножить требуемое равенство на z_2 и тогда проверка сводится к первому пункту.

3) Это непосредственное следствие первого пункта. □

5.6 Основная теорема алгебры

Определение 44. Поле F называется алгебраически замкнутым, если для любого многочлена $f \in F[x] \setminus F$ существует корень $\alpha \in F$, то есть $f(\alpha) = 0$.

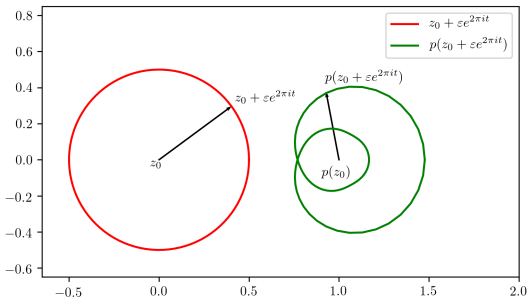
Утверждение 45 (Основная теорема алгебры)**.** *Поле $\mathbb{C}$ построенное в разделе 5.3 алгебраически замкнуто.*

План доказательства Давайте я расскажу идейный план доказательства теоремы, чтобы у вас было понимание, что мы на самом деле будем делать ниже.

1. Рассмотрим произвольный не константный многочлен $p \in \mathbb{C}[x]$ и определим отображение $|p|\colon \mathbb{C} \to \mathbb{R}$ по правилу $z \mapsto |p(z)|$. Первое что мы покажем, что это отображение достигает своего минимума в какой-то точке $z_0 \in \mathbb{C}$. Грубо говоря это будет следовать из того, что «на бесконечности» $|p(z)|$ стремится к бесконечности, а значит минимум может быть лишь «возле нуля». А «возле нуля», то есть в каком-то диске содержащем ноль, функция $|p(z)|$ непрерывна, а потому достигает минимума.

2. Так как отображение $|p(z)|$ принимает только неотрицательные значения, то наличие нуля у многочлена p равносильно тому, что в точке минимума z_0 значение $|p(z_0)|$ будет ноль. Теперь надо будет показать, что минимум не может быть положительным.

3. Предположим, что минимум $r = |p(z_0)| \neq 0$. Тогда значения $p(z)$ лежат вне диска радиуса r с центром в нуле или на его границе. Нам надо показать, что на самом деле если $r > 0$, то есть диск имеет внутренность, то у нас обязательно найдется значение p внутри диска, что будет противоречить тому, что z_0 – минимум $|p|$. Последнее делается так: давайте выберем «маленькую» окружность вокруг точки z_0 и пройдемся по ней против часовой стрелки. Тогда оказывается, что значения $p(z)$ пробегут по «маленькой почти окружности» вокруг $p(z_0)$ и может быть сделают более одного оборота. А это значит, что обойдя вокруг $p(z_0)$, мы обязательно попадем внутрь диска радиуса r.



Слева показана красным «маленькая» окружность вокруг z_0, а справа зеленым ее образ под действием p. Здесь зеленая «почти окружность» делает два оборота.

Теперь начнем реализовывать этот план доказательства. Начнем со следующего.

Утверждение 46. *Пусть $p \in \mathbb{C}[x]$ – произвольный многочлен отличный от константы. Тогда для любого $c > 0$ найдется $r > 0$, что $|p(z)| > c$ при $|z| > r$.*

Доказательство. Пусть $p(z) = a_0 + a_1z + \ldots + a_nz^n$ и $a_n \neq 0$. Тогда

$$p(z) = a_nz^n\left(1 + \frac{a_{n-1}}{a_nz} + \ldots + \frac{a_0}{a_nz^n}\right) = a_nz^n(1 + \omega(z))$$

Фиксируем произвольное положительное число $r > 1$ и рассмотрим $|z| > r$. Тогда

$$|\omega(z)| \leqslant \left|\frac{a_{n-1}}{a_n z}\right| + \ldots + \left|\frac{a_0}{a_n z^n}\right| \leqslant \left|\frac{a_{n-1}}{a_n}\right| \frac{1}{r} + \ldots + \left|\frac{a_0}{a_n}\right| \frac{1}{r^n} \leqslant \left(\left|\frac{a_{n-1}}{a_n}\right| + \ldots + \left|\frac{a_0}{a_n}\right|\right) \frac{1}{r}$$

Последнее выражение обозначим за $\delta(r)$, оно идет к нулю при $r \to \infty$. Давайте теперь оценим вне этого диска значение $|p(z)|$:

$$|p(z)| = |a_n z^n (1 + \omega(z))| = |a_n||z|^n |1 + \omega(z)| \geqslant |a_n||z|^n (1 - |\omega(z)|) \geqslant |a_n||r|^n (1 - \delta(r)) \to \infty, \text{ при } r \to \infty$$

То есть мы сможем найти r при котором вне диска $\{z \in \mathbb{C} \mid |z| \leqslant r\}$ будет выполняться $|p(z)| > c$. □

Утверждение 47. *Пусть $p \in \mathbb{C}[x]$ – произвольный многочлен, тогда отображение $|p|\colon \mathbb{C} \to \mathbb{R}$ заданное по правилу $z \mapsto |p(z)|$ достигает минимума, то есть найдется такая точка $z_0 \in \mathbb{C}$, что $|p(z_0)| \leqslant |p(z)|$ для любого $z \in \mathbb{C}$.*

Доказательство. Идея доказательства этого утверждения следующая. Пусть $c = |p(0)|$. Если это ноль, то мы нашли наш минимум. Пусть $c \neq 0$, тогда давайте найдем диск $D_r(0)$ с центром в нуле и радиуса r такой, что $|p(z)| > c$ для всех $z \notin D_r(0)$. Тогда, по утверждению 46 мы можем найти диск $D_r(0) = \{z \in \mathbb{C} \mid |z| \leqslant r\}$, вне которого $|p(z)| > c$. А значит, если мы найдем минимум для $|p(z)|$ на диске $D_r(0)$ он автоматически будет минимумом в $\mathbb{C}$. Действительно, внутри диска в этой точке мы будем принимать наименьшее значение, в частности значение будет не больше c. Но вне диска мы не можем принять значение меньше, так как там мы строго больше c.

Теперь надо найти минимум внутри диска $D_r(0)$. Давайте я приведу два доказательства: идейное и доказательство в лоб. Идейное доказательство такое. Функция $\phi\colon \mathbb{C} \to \mathbb{R}$ по правилу $z \mapsto |p(z)|$ есть композиция двух отображений: полиномиального $p\colon \mathbb{C} \to \mathbb{C}$ и модуля $|-|\colon \mathbb{C} \to \mathbb{R}$ по правилу $z \mapsto |z|$. Оба эти отображения непрерывны, а значит и отображение $z \mapsto |p(z)|$ тоже непрерывно. Кроме того, диск $D_r(0)$ является компактом, а любое непрерывное отображение на компакте достигает минимума.[60] □

Минимум на диске (по простому?) Чтобы найти минимум на диске, мне придется пользоваться фактами из математического анализа, ведь свойства комплексных чисел должны зависеть от особенностей их природы, которая не чисто алгебраическая. Нам понадобится следующий факт.[61]

Утверждение (БД)**.** *У любой последовательности на отрезке найдется сходящаяся подпоследовательность. То есть для любой $a_n \in [a, b]$ найдется подпоследовательность a_{n_k} такая, что существует*

$$\lim_{k\to\infty} a_{n_k} \in [a, b]$$

Рассмотрим функцию $f = |p|\colon D_r(0) \to \mathbb{R}$ по правилу $f(z) = |p(z)|$. Так как эта функция ограничена снизу нулем, то существует нижняя грань

$$a = \inf_{z \in D_r(0)} f(z)$$

По определению нижней грани, мы можем выбрать последовательность $z_n \in D_r(0)$ такую, что $f(z_n) \to a$. Такая последовательность обязательно имеет вид $z_n = a_n + ib_n$, где $a_n, b_n \in [-r, r]$ – последовательности вещественных чисел на отрезке. Из них мы по очереди можем выбрать сходящиеся подпоследовательности a_{n_k} и b_{n_k}, так что последовательность z_{n_k} сходится в $D_r(0)$ к какой-то точке z_0. А значит

$$a = \lim_{k\to\infty} f(z_{n_k}) = \lim_{k\to\infty} |p(z_{n_k})| = \left|\lim_{k\to\infty} p(z_{n_k})\right| = \left|p\left(\lim_{k\to\infty} z_{n_k}\right)\right| = |p(z_0)|$$

Давайте объясним все переходы. Первый – это определение нашей последовательности, мы по ней подбираемся к инфимуму. Второй – это непрерывность модуля, то есть для комплексного числа верно $\lim_{n\to\infty} |z_n| = |\lim_{n\to\infty} z_n|$. Действительно, ведь $|a + bi| = \sqrt{a^2 + b^2}$ и функция корня от суммы квадратов непрерывна, то есть предел в точке равен ее значению в точке. Третий переход следует из непрерывности многочлена $p(z)$. Действительно, такой многочлен – это сумма произведений мнимых и вещественных частей с коэффициентами, а в таких функциях мы тоже умеем переходить к пределу. Последнее равенство – это возможность взять предел у выбранной подпоследовательности. Таким образом в точке z_0 достигается нижняя грань на диске $D_r(0)$, а значит это точка минимума.

[60] Если вы сейчас пребываете в шоке, то это нормально. Сейчас я исправлюсь и напишу простое доказательство, но он будет несколько длиннее.

[61] Искренне надеюсь, что с ним вы знакомы. Его можно считать одной из аксиом вещественных чисел.

Минимум обязан быть нулем Теперь осталось доказать, что минимум для функции $|p(z)|$ (если p не константа) обязательно должен быть нулем. Точнее мы покажем, что если $|p(z_0)| \neq 0$, то обязательно найдется точка с еще меньшим модулем, то есть найдется $z_1 \in \mathbb{C}$, что $|p(z_1)| < |p(z_0)|$. Но в начале нам понадобится следующий пример.

Пример 48 (V.I.P. пример). Рассмотрим $p(z) = z^d$ и посмотрим на функцию $p\colon \mathbb{C} \to \mathbb{C}$ по правилу $z \mapsto p(z) = z^d$. Давайте рассмотрим окружность $z(t) = re^{2\pi it}$ для $t \in [0, 1]$. Когда t пробегает от 0 до 1, то $z(t) = re^{2\pi it}$ пробегает по окружности радиуса r один оборот против часовой стрелки. Давайте посмотрим на образ этой окружности под действием p, получим $p(z(t)) = r^d e^{2\pi idt}$. То есть теперь, когда t пробегает от 0 до 1 мы пробегаем окружность радиуса r^d но уже d раз против часовой стрелки (делаем d оборотов вместо одного).

Утверждение 49. *Пусть $p \in \mathbb{C}[x]$ – произвольный не константный многочлен и $z_0 \in \mathbb{C}$ такая точка, что $p(z_0) \neq 0$. Тогда найдется точка $z_1 \in \mathbb{C}$ такая, что $|p(z_1)| < |p(z_0)|$.*

Доказательство. Если точка z_0 не является точкой минимума для $|p|$, то нужная точка z_1 найдется по определению. То есть мы можем предположить, что z_0 – точка минимума. Давайте определим многочлен $g(z) = p(z_0 + z)$. Тогда у многочлена $g(z)$ то же множество значений, что и у $p(z)$, но у него точка 0 является точкой минимума. В свою очередь g представляется в виде

$$g(z) = a_0 + a_r z^r + a_{r+1} z^{r+1} + \ldots + a_n z^n$$

Здесь выше a_0 – это значение многочлена g в нуле, которое по условию не ноль. Число a_r – это первый ненулевой коэффициент после a_0 в многочлене g. Учтите, что r может быть 1, а может быть больше 1. Давайте перепишем многочлен g следующим образом:

$$g(z) = a_0 + a_r z^r \left(1 + \frac{a_{r+1}}{a_r} z + \ldots + \frac{a_n}{a_r} z^{n-r}\right) = a_0 + a_r z^r \left(1 + \omega(z)\right) = a_0 + g_0(z) + \omega(z) g_0(z)$$

где

$$\omega(z) = \frac{a_{r+1}}{a_r} z + \ldots + \frac{a_n}{a_r} z^{n-r} \quad \text{и} \quad g_0(z) = a_r z^r$$

Давайте в начале посмотрим на многочлен $h(z) = a_0 + a_r z^r = a_0 + g_0(z)$ и покажем что для него найдется точка z_1 такая, что $|h(z_1)| < |h(0)|$. Если $|z| = \delta$, то z^r описывает окружность радиуса δ^r вокруг нуля и делает r оборотов. Выражение $a_r z^r$ описывает окружность радиуса $R = |a_r|\delta^r$ вокруг нуля и делает r оборотов.[62] А значит $h(z) = a_0 + a_r z^r$ описывает окружность радиуса R вокруг точки a_0 и делает r оборотов. При малых δ эта окружность пересекается с радиус вектором a_0 в некоторой точке m. Тогда решая уравнение $m = a_0 + a_r z^r$,[63] мы найдем точку z_1 такую, что $h(z_1) = m$. Но по построению точка m ближе к нулю, чем a_0. Действительно, а это и значит, что

$$|h(z_1)| = |a_0| - R < |a_0| = |h(0)|$$

Теперь я хочу показать, что $|g(z_1)| < |g(0)|$ при малых δ. Главная идея заключается в том, что g и h имеют близкие значения если δ достаточно мало. Давайте для начала оценим $\omega(z)$, когда $|z| = \delta < 1$:

$$|\omega(z)| \leqslant \left|\frac{a_{r+1}}{a_r}\right| \delta + \ldots + \left|\frac{a_n}{a_r}\right| \delta^{n-r} \leqslant \left(\left|\frac{a_{r+1}}{a_r}\right| + \ldots + \left|\frac{a_n}{a_r}\right|\right) \delta = C\delta$$

Последнее неравенство следует из того, что $\delta < 1$. А через C мы обозначили полученную константу. Значит при $\delta < \varepsilon/C$ мы можем считать, что $|\omega(z)| < \varepsilon$.

Теперь посмотрим на значение $|g(z_1)|$:

$$|g(z_1)| \leqslant |h(z_1)| + |\omega(z_1)||a_r z^r| = |a_0| - R + |\omega(z_1)|R = |a_0| - (1 - |\omega(z_1)|)R < |a_0| - (1 - \varepsilon)R = |g(0)| - (1 - \varepsilon)R$$

Таким образом мы видим, что при малых δ значение $|g(z_1)|$ строго меньше $|g(0)|$. Что и требовалось. □

Доказательство основной теоремы алгебры. Давайте я проговорю строго доказательство утверждения 45 от начала и до конца пользуясь доказанными выше результатами.

Пусть $f \in \mathbb{C}[z] \setminus \mathbb{C}$ – произвольный неконстантный многочлен. Тогда по утверждению 47, отображение $|f|\colon \mathbb{C} \to \mathbb{R}_+$ достигает своего минимума. Пусть точка минимума z_0. Тогда $|f(z)| \geqslant |f(z_0)| \geqslant 0$ для всех $z \in \mathbb{C}$. Если $f(z_0) = 0$, то мы нашли корень у многочлена f. Потому можно считать, что $|f(z_0)| > 0$. Тогда по утверждению 49 найдется точка $z_1 \in \mathbb{C}$ такая, что $|f(z_1)| < |f(z_0)|$, что противоречит тому, что z_0 была точкой минимума. А значит такого быть не может, что $|f(z_0)| \neq 0$ и теорема доказана. □

[62] Начальная точка для окружности имеет аргумент равный аргументу a_r.

[63] Это мы можем сделать, так как тут задача про извлечение корня из числа $(m - a_0)/a_r$.

5.7 Многочлены

В этом разделе я хочу сказать пару слов про многочлены. Пусть F – некоторое поле. Тогда многочлен над полем F – это картинка вида $f = a_0 + a_1 x + \ldots + a_n x^n$, где $a_i \in F$. Формально, такая картинка – это конечная последовательность чисел $(a_0, \ldots, a_n)$, но психологически лучше и правильнее думать именно про картинки. Еще можно для краткости писать $f = \sum_{k \geqslant 0} a_k x^k$, подразумевая, что в этой сумме только конечное число ненулевых коэффициентов. Это удобное соображение позволяет удобно записать правила для сложения и умножения многочленов, которые определяются следующим образом

$$\Big(\sum_{k\geqslant 0} a_k x^k\Big) + \Big(\sum_{k\geqslant 0} b_k x^k\Big) = \sum_{k\geqslant 0} (a_k + b_k) x^k \quad \text{и} \quad \Big(\sum_{k\geqslant 0} a_k x^k\Big)\Big(\sum_{k\geqslant 0} b_k x^k\Big) = \sum_{k\geqslant 0}\Big(\sum_{m+n=k} a_m b_n\Big) x^k$$

Таким образом многочлен – это не функция, а картинка. Однако, каждый многочлен $f \in F[x]$ задает функцию $F \to F$ по правилу $x \mapsto f(x)$. Но в случае конечных полей (то есть полей из конечного числа элементов) разные многочлены могут давать одни и те же функции. Напомню, что степень многочлена f – это наибольший номер n, что коэффициент $a_n \neq 0$.[64]

Примеры

- Пример конечного поля.

  Пусть $p \in \mathbb{Z}$ – некоторое простое число. Обозначим через $\mathbb{Z}_p$ множество остатков по этому числу, то есть $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$. Введем на этом множестве операции сложения и умножения по модулю простого числа p, то есть

  $$a + b = a + b \pmod p \quad \text{и} \quad ab = ab \pmod p$$

  Тогда можно проверить, что Z_p является полем, где числа 0 и 1 являются нулем и единицей поля. Единственная аксиома, которая требует усилий – показать, что любой ненулевой элемент Z_p обратим. Давайте возьмем произвольный ненулевой элемент $a \in \mathbb{Z}_p$. Так как $a < p$ и p – простое число, то $(a, p) = 1$. По расширенному алгоритму евклида найдутся целые числа $u, v \in \mathbb{Z}$ такие, что $1 = ua + vp$. Рассмотрим это равенство по модулю простого числа p и получим, что $ua = 1 \pmod p$, а это и означает, что u является обратным к a по умножению.

- Пример, когда разные многочлены дают одну и ту же функцию.

  Рассмотрим многочлены $\mathbb{Z}_2[x]$. Тогда $\mathbb{Z}_2$ состоит только из 0 и 1. В этом случае все многочлены x^n задают одну и ту же функцию.

Определение 50. Пусть F – произвольное поле и $f \in F[x]$ – некоторый многочлен. Если число $a \in F$ является его корнем, то f делится на $x - a$, а значит представляется в виде $f(x) = (x-a)g(x)$ для некоторого $g \in F[x]$. Аналогично, если a является корнем g, то можно выделить $(x-a)$ и в g и так далее. В итоге можно найти разложение $f(x) = (x-a)^k g(x)$, где $g(a) \neq 0$. В этом случае говорят, что k – это кратность корня a в многочлене f. Корень кратности 1 называется простым.

Определение 51. Пусть $f \in F[x]$ имеет вид $f = a_0 + a_1 x + \ldots + a_n x^n$. Определим формальную производную следующим образом $f' = a_1 + 2a_2 x + \ldots + n a_n x^{n-1}$ или по-другому $f' = \sum_{k \geqslant 1} k a_k x^{k-1}$.

Несложно убедиться, что, определив таким образом производную, она удовлетворяет всем естественным свойствам, к которым мы привыкли в анализе. В качестве упражнения предлагается проверить следующее.

Утверждение 52. *Пусть F – произвольное поле. Для формальной производной выполнены следующие свойства:*

1. $(f+g)' = f' + g'$ *для любых* $f, g \in F[x]$.
2. $(\lambda f)' = \lambda f'$ *для любых* $\lambda \in F$ *и* $f \in F[x]$.
3. $(fg)' = f'g + fg'$ *для любых* $f, g \in F[x]$.
4. $f(g(x))' = f'(g(x))g'(x)$ *для любых* $f, g \in F[x]$.

[64] По-хорошему надо еще аккуратно определить степень нулевого многочлена. Но ее обычно определяют по ситуации так, как удобнее. Например можно положить -1 или $-\infty$ и есть еще пара способов. Но об этом можно особенно не запариваться.

С помощью формальной производной можно проверить кратность корня в произвольном многочлене. Для начала нам нужно следующее вспомогательное утверждение.

Утверждение 53. *Пусть F – произвольное поле, $f \in F[x]$ – некоторый многочлен и $a \in F$ – его корень кратности k. Тогда*

1. *Число a является корнем кратности хотя бы $k-1$ в многочлене f'.*
2. *Если число $k1 \neq 0$ в F, то a является корнем кратности в точности $k-1$ в многочлене f'.*

Доказательство. 1) По определению имеем $f = (x-a)^k g(x)$ причем $g(a) \neq 0$. Возьмем производную от f, получим

$$f' = k(x-a)^{k-1}g(x) + (x-a)^k g'(x) = (x-a)^{k-1}(kg(x) + (x-a)g'(x))$$

и мы видим, что у производной a имеет кратность хотя бы $k-1$.

2) Давайте поймем, когда кратность может вырасти. Только если множитель $(kg(x)+(x-a)g'(x))$ зануляется в a. Если подставить a, то получим $kg(a)$. Число $g(a) \neq 0$ по выбору, но если $k1 \neq 0$, то и их произведение не ноль в поле F, а это будет означать, что кратность корня в точности $k-1$. □

Примеры и замечания

1. Давайте продемонстрируем ситуацию, когда кратность корня может возрасти. Например, выберем $F = \mathbb{Z}_p$ и в качестве многочлена h рассмотрим $x^p - 1$. Тогда $h' = 0$. Теперь положим $f = xh(x) = x^{p+1} - x$. Тогда $f' = h(x) = x^p - 1$. С другой стороны $x^p - 1 = (x-1)^p$, а значит 1 имеет кратность p в многочлене f. Но и в многочлене f' 1 имеет кратность p.

2. Если для любого натурального числа $k \in \mathbb{N}$ в поле F выполнено, $k1 \neq 0$, то можно следующим образом проверить корень многочлена $f \in F[x]$ на кратность. Если $a \in F$ – некоторый корень. Надо посмотреть на $f'(a)$. Если это число ноль, то a корень кратности больше 1, а если не ноль, то кратности в точности 1.

3. Если F произвольное поле, то общий алгоритм проверки корня на простоту следующий. Надо взять многочлен $f \in F[x]$, для которого a является корнем. Посчитать производную f', потом посчитать нод $d(x) = (f, f')$. Если $d(a) = 0$, то a кратный корень, если $d(a) \neq 0$, то это корень кратности 1.[65]

[65] Я не буду останавливаться на доказательствах этих фактов. Все они вам встретятся в курсе алгебры.

6 Векторные пространства

6.1 Идея и определение

Идея Мы с вами до этого изучали много разных объектов, которые не сильно похожи друг на друга. Например, вектор-столбцы F^n, матрицы $\mathrm{M}_{m\,n}(F)$, функции $f\colon X \to F$, многочлены $F[x]$. Все эти товарищи нам постоянно встречаются и каждый раз приходится для каждого из них все доказывать заново и во время доказательств мы видим, что наши рассуждения повторяются. Это означает, что на самом деле у всех этих объектов есть некий общий интерфейс, через который мы на самом деле с ним работаем. Самое главное в этом интерфейсе то, что мы можем брать элементы из этих объектов, умножать эти элементы на числа и складывать между собой. Абстрактное векторное пространство как раз и формализует идею такого общего интерфейса, через который в множестве можно складывать элементы и умножать на числа.

У такого подхода есть несколько плюсов. Во-первых, формальное удобство: как только вы что-то сделали для абстрактного векторного пространства и увидели, что что-то конкретное является таковым, то все ваши достижения автоматом применимы в этой конкретной ситуации. Общий алгоритм для векторного пространства будет одинаково хорошо работать и для столбцов, и для матриц, и для функций и т.д. Во-вторых, есть менее очевидный бонус. Когда мы доказываем что-то про абстрактное векторное пространство, то про него надо думать как про F^n. Это поможет вам не потеряться в формализме и догадаться, что откуда берется. Неформально это означает, что если вы что-то умеете делать для F^n, то это автоматически верно для любого векторного пространства! Формально это не совсем правда, но в классе хороших пространств это так.[66] Тем не менее, даже в классе всех пространств, интуиция из F^n очень полезна.

Определение Следующее определение – это пример определения с контекстом. Это означает, что прежде, чем его дать, вы должны зафиксировать некоторую информацию, которая необходима для вашего определения и без этой информации оно – бессмысленный мусор. У определения векторного пространства в качестве такого контекста выступает некоторое поле F. Это значит, что пока вы не зафиксировали какое-то поле, вы не можете говорить о векторных пространствах над полем F, а «просто векторных пространств» без указания какого-либо поля не существует.

Определение 54. Пусть F – некоторое фиксированное поле. Тогда векторное пространство над полем F – это следующий набор данных $(V, +, \cdot)$, где

- V – множество. Элементы этого множества будут называться векторами.
- $+\colon V \times V \to V$ – бинарная операция, то есть правило, действующее так: $(v, u) \mapsto v + u$, где $u, v \in V$.
- $\cdot\colon F \times V \to V$ – бинарная операция, то есть правило, действующее так: $(\alpha, v) \mapsto \alpha v$, где $\alpha \in F$ и $v \in V$.

При этом эти данные удовлетворяют следующим 8 аксиомам:

1. **Ассоциативность сложения** Для любых векторов $u, v, w \in V$ верно $(u + v) + w = u + (v + w)$.
2. **Существование нулевого вектора** Существует такой вектор $0 \in V$, что для любого $v \in V$ выполнено $0 + v = v + 0 = v$.
3. **Существование противоположного вектора** Для любого вектора $v \in V$ существует вектор $-v \in V$ такой, что $v + (-v) = (-v) + v = 0$.
4. **Коммутативность сложения** Для любых векторов $u, v \in V$ верно $u + v = v + u$.
5. **Согласованность умножения со сложением векторов** Для любого числа $\alpha \in F$ и любых векторов $u, v \in V$ верно $\alpha(v + u) = \alpha v + \alpha u$.
6. **Согласованность умножения со сложением чисел** Для любых чисел $\alpha, \beta \in F$ и любого вектора $v \in V$ верно $(\alpha + \beta)v = \alpha v + \beta v$.
7. **Согласованность умножения с умножением чисел** Для любых чисел $\alpha, \beta \in F$ и любого вектора $v \in V$ верно $(\alpha\beta)v = \alpha(\beta v)$.
8. **Нетривиальность** Для любого $v \in V$ верно $1v = v$.[67]

[66] Под хорошими тут подразумеваются конечно мерные.
[67] Здесь $1 \in F$.

Примеры

1. Поле F (или кто больше привык к вещественным числам $\mathbb{R}$) является векторным пространством над F (соответственно над $\mathbb{R}$).

2. Более обще, множество вектор-столбцов F^n является векторным пространством над F.

3. Множество матриц $\mathrm{M}_{m\,n}(F)$ является векторным пространством над F.

4. Пусть X – произвольное множество, тогда множество функций $F(X,F) = \{f\colon X \to F\}$ является векторным пространством над F. Надо лишь объяснить как складывать функции и умножать на элементы F. Операции поточечные, пусть $f, g\colon X \to F$, тогда функция $(f+g)\colon X \to F$ действует по правилу $(f+g)(x) = f(x) + g(x)$. Если $\alpha \in F$, то функция $(\alpha f)\colon X \to F$ действует по правилу $(\alpha f)(x) = \alpha f(x)$.

5. Множество многочленов $F[x] = \{a_0 + a_1x + \ldots + a_nx^n \mid a_i \in F,\ n \in \mathbb{Z}_{\geqslant 0}\}$. Тут надо обратить внимание, что мы подразумеваем под многочленом. Для нас многочлен – это НЕ функция, многочлен – это картинка вида $a_0 + a_1x + \ldots + a_nx^n$.[68] Складываются и умножаются эти картинки по одинаковым правилам. Важно, что две такие картинки равны тогда и только тогда, когда у них равные коэффициенты. Множество всех многочленов $F[x]$ является векторным пространством над F.

Замечание Стоит отметить, что в обычных векторных пространствах мы привыкли к некоторым свойствам, которые бы хотелось иметь и в общем случае. Например, в F^n есть единственный нулевой вектор, а аксиомы в общем случае говорят, что нулевой вектор лишь существует. Однако, можно показать, что нулевой вектор автоматически единственный. Давайте перечислим некоторые непосредственные следствия из аксиом:

1. Нулевой вектор единственный. Действительно, если у нас два нуля 0_1 и 0_2, то рассмотрим $0_1 + 0_2$. Так как 0_1 является нулем, то по определению $0_1 + 0_2 = 0_2$. С другой стороны, так как 0_2 является нулем, то $0_1 + 0_2 = 0_1$. А это и означает, что оба нуля совпадают.

2. Для любого $v \in V$ существует единственный $-v$. Действительно, пусть $u, w \in V$ два вектора обратных к v. Это значит, что выполнены равенства

$$v + u = u + v = 0 \quad \text{и} \quad v + w = w + v = 0$$

Но тогда

$$w = 0 + w = (u + v) + w = u + (v + w) = u + 0 = u$$

3. Для любого вектора $v \in V$ имеем $0v = 0$.[69] Начнем с равенства $0 + 0 = 0$ для нуля из F. Раз два числа равны, то после умножения одного и того же вектора на них, результаты останутся одинаковыми. Значит для произвольного $v \in V$ имеем $(0 + 0)v = 0v$. Раскроем скобки и прибавим к обеим сторонам равенства элемент обратный к $0v$, получим

$$0v + 0v = 0v \quad \Rightarrow \quad 0v + 0v + (-(0v)) = 0v + (-(0v)) \quad \Rightarrow \quad 0v + 0 = 0 \quad \Rightarrow \quad 0v = 0$$

4. Для любого числа $\alpha \in F$ верно $\alpha 0 = 0$.[70] В этом случае доказательство аналогично предыдущему, надо лишь стартовать с равенства векторов, а не чисел $0 + 0 = 0$ в V. Умножим эти векторы на одно и то же число α, получим $\alpha(0 + 0) = \alpha 0$. Опять раскроем скобки и прибавим $-(\alpha 0)$, получим:

$$\alpha 0 + \alpha 0 = \alpha 0 \quad \Rightarrow \quad \alpha 0 + \alpha 0 + (-(\alpha 0)) = \alpha 0 + (-(\alpha 0)) \quad \Rightarrow \quad \alpha 0 + 0 = 0 \quad \Rightarrow \quad \alpha 0 = 0$$

5. Для любого вектора $v \in V$ верно $-v = (-1)v$. Для этого рассмотрим равенство $1 + (-1) = 0$. Умножим эти одинаковые числа на один и тот же вектор $v \in V$ и получим $(1 + (-1))v = 0$. Теперь надо раскрыть скобки и прибавить $-v$

$$1v + (-1)v = 0 \quad \Rightarrow \quad v + (-1)v = 0 \quad \Rightarrow \quad (-1)v = -v$$

[68] Для любителей формализма, можете считать, что многочлен – это конечная последовательность элементов F вида $(a_0, \ldots, a_n)$, но длина последовательности может быть любой, включая нулевую.
[69] Здесь слева 0 – это нулевой элемент поля, а справа 0 – это нулевой вектор.
[70] Здесь 0 – это нулевой вектор в обеих частях равенства.

6.2 Подпространства

Пусть V – векторное пространство над F. Тогда непустое подмножество $U \subseteq V$ называется подпространством, если на него можно ограничить операции $+$ и $\cdot$ и относительно них оно является векторным пространством. Давайте определим подпространство формально.

Определение 55. Пусть V – векторное пространство над полем F. Тогда подмножество $U \subseteq V$ называется подпространством, если

1. U не пусто.[71]
2. Для любых векторов $u, u' \in U$ верно, что $u + u' \in U$.
3. Для любого скаляра $\alpha \in F$ и вектора $u \in U$ верно, что $\alpha u \in U$.

Если U – подпространство в V, то на U можно корректно ограничить операции сложения и умножения на скаляр из исходного пространства V. Таким образом у нас получается набор данных $(U, +, \cdot)$ и теперь надо, чтобы выполнялись все аксиомы векторного пространства для них. Оказывается, что все аксиомы будут выполняться автоматически! Например, почему у нас будет $0 \in U$. Потому что если мы возьмем любой вектор $u \in U$, то $0 = 0u \in U$. Остальное я оставлю в качестве упражнения.

Примеры

1. Для любого векторного пространства V подмножества 0 и V всегда являются подпространствами.
2. Множество $\{y \in F^n \mid Ay = 0\} \subseteq F^n$, где $A \in \mathrm{M}_{m\,n}(F)$, является векторным подпространством в F^n.
3. Множество многочленов $\mathbb{R}[x]$ является подпространством в пространстве $F(\mathbb{R}, \mathbb{R})$ – всех функций на прямой.

Обратите внимание, что подпространство из второго примера кажется устроено сложнее, чем векторное пространство F^n, в котором оно лежит. Однако, окажется, что взаимодействие с ним через абстрактный интерфейс векторного пространства происходит точно так же. То есть на самом деле подпространство устроено не сложнее, чем исходное пространство. Об этом речь пойдет после того, как мы узнаем, что такое базисы и что значит, что какие-то векторные пространства одинаковые.

6.3 Линейные комбинации

Мотивация Пусть у нас есть векторное пространство V над полем F. Давайте поймем, а что вообще с ним можно делать? Во-первых, V – это множество. Значит из него можно брать элементы. Во-вторых, там есть операция умножения на числа, то есть любой вектор можно умножить на какое-то число. В-третьих, вектора можно складывать. Все это означает, что все что можно делать с векторным пространством, это набрать каких-то векторов из него $v_1, \ldots, v_n$ и написать выражение вида $\alpha_1 v_1 + \ldots + \alpha_n v_n$, для произвольных $\alpha_i \in F$. Это выражение будет задавать нам какой-то вектор из V. Как мы видим, особенно не разбежишься с разнообразием действий. Однако, важно, что с помощью подобных выражений можно вытащить абсолютно всю информацию из векторных пространств, которую только возможно. Именно поэтому все наше внимание будет посвящено выражениям такого вида, так как из них получится узнать все, что только можно про векторные пространства.

Линейные комбинации

Определение 56. Пусть V – некоторое векторное пространство над полем F и пусть $v_1, \ldots, v_n \in V$ – некоторый набор векторов. Тогда выражение вида $\alpha_1 v_1 + \ldots + \alpha_n v_n$, где $\alpha_i \in F$, называется линейной комбинацией $v_1, \ldots, v_n$. Линейная комбинация называется тривиальной, если все $\alpha_i = 0$. В противном случае она называется нетривиальной.

Определение 57. Вектора $v_1, \ldots, v_n \in V$ называются линейно зависимыми, если существует их нетривиальная линейная комбинация равная нулю, то есть для каких-то $\alpha_i \in F$ (так что хотя бы один не равен нулю) выражение $\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$. Подчеркнем, что вектора линейно независимы, если из равенства $\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$ следует, что все $\alpha_i = 0$.

[71] При наличии свойства (3) это свойство эквивалентно тому, что нулевой вектор V попадает в U. Действительно, если он попадает, то U не пусто. Наоборот, если $u \in U$ – какой-то вектор, то $0 = 0u \in U$ по третьему свойству.

Примеры

1. Вектор 0 всегда линейно зависим.
2. Вектор $v \in V$ линейно зависим тогда и только тогда, когда он равен нулю.
3. Вектора $v_1, v_2 \in V$ линейно зависимы тогда и только тогда, когда они пропорциональны (то есть один из них равен другому умноженному на элемент поля).

Заметим, что если множество векторов $v_1, \ldots, v_k$ линейно независимо, то и любое его подмножество тоже линейно независимо. Потому интересно не уменьшать, а увеличивать линейно независимые подмножества векторов. Линейно независимое множество векторов $v_1, \ldots, v_k$ называется максимальным, если при добавлении к нему любого вектора оно становится линейно зависимым.

Линейная оболочка

Определение 58. Пусть $E \subseteq V$ – некоторое подмножество в векторном пространстве V над полем F. Тогда обозначим через $\langle E \rangle$ множество всех линейных комбинаций векторов из E, то есть

$$\langle E \rangle = \{\alpha_1 v_1 + \ldots + \alpha_n v_n \mid \alpha_i \in F,\, v_i \in E,\, n \in \mathbb{N}\}$$

Сделаем важное замечание, если $E = \varnothing$ (пусто), то $\langle \varnothing \rangle$ полагаем равным нулевому подпространству (подпространству состоящему только из нуля). Это полезное и удобное соглашение можно понимать так: если берется линейная комбинация с нулевым числом слагаемых, то она равна нулю.

Заметим, что $\langle E \rangle$ является наименьшим векторным подпространством содержащим E. Потому, для любого подпространства $U \subseteq V$ верно $\langle U \rangle = U$.

Пример Полезно держать перед глазами следующий пример. Пусть $V = \mathbb{R}^3$ – пространство, $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$ – три вектора вдоль координатных осей. Тогда $\langle e_1 \rangle$, $\langle e_2 \rangle$ и $\langle e_3 \rangle$ – это в точности координатные оси. Подпространства $\langle e_1, e_2 \rangle$, $\langle e_1, e_3 \rangle$ и $\langle e_2, e_3 \rangle$ – это плоскости содержащие пары координатных осей, $\langle e_1, e_2, e_3 \rangle$ будет совпадать со всем пространством $\mathbb{R}^3$.

Порождающее подмножество

Определение 59. Пусть V – векторное пространство над полем K, тогда подмножество $E \subseteq V$ называется порождающим, если $\langle E \rangle = V$.

Другими словами, E является порождающим если любой вектор из V является линейной комбинацией векторов из E. Отметим, что V целиком всегда является порождающим. Если $E \subseteq E' \subseteq V$ и подмножество E является порождающим, то и E' тоже порождающее. Потому порождающее семейство всегда можно увеличить и это не интересно, интереснее попытаться его уменьшить и сделать более экономным. Порождающее множество E называется минимальным, если любое строго меньшее подмножество E уже не порождающее. Для этого достаточно проверить, что для любого $v \in E$ множество $E \setminus \{v\}$ уже не порождающее.

6.4 Базис

Подмножество $E \subseteq V$ называется линейно независимым, если любое конечное подмножество векторов E линейно независимо. Если $E' \subseteq E \subseteq V$ и E является линейно независимым, то E' тоже будет линейно независимым. Потому линейно независимое подмножество можно всегда уменьшать[72] и это не интересно, интереснее попытаться его увеличить и сделать наиболее большим. Линейно независимое подмножество E называется максимальным, если любое строго содержащее его подмножество является линейно зависимым. Для этого достаточно проверить, что для любого $v \in V \setminus E$ множество $E \cup \{v\}$ является линейно зависимым.

Утверждение 60. *Пусть V – некоторое векторное пространство над некоторым полем F. Тогда следующие условия на подмножество $E \subseteq V$ эквивалентны:*

1. *E – минимальное порождающее подмножество.*

[72] Вопрос линейной независимости пустого множества оставим на совести строгой аксиоматической теории множеств и не будем его касаться, чтобы не обжечься о всякий формальный геморрой.

2. *E – максимальное линейно независимое подмножество.*

3. *E – одновременно порождающее и линейно независимое подмножество.*

Доказательство. Будем доказывать по схеме (1)⇔(3)⇔(2).

(1)⇒(3). Пусть E – минимальное порождающее, нам надо показать, что оно будет линейно независимым. Предположим противное, пусть найдется вектора $v_1, \ldots, v_n \in E$ и числа $\alpha_1, \ldots, \alpha_n \in F$, так что не все из них равны нулю, что выполнено $\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$. Мы можем предположить, что $\alpha_1 \neq 0$. Тогда $v_1 = \beta_2 v_2 + \ldots + \beta_n v_n$ для некоторых $\beta_i \in F$. Давайте покажем, что тогда $E \setminus \{v_1\}$ тоже является порождающим, что будет противоречить минимальности E. Действительно, пусть $v \in V$ – произвольный вектор. Так как E – порождающее, то v выражается через вектора из E, $v = \sum_i \alpha_i' v_i'$. Если среди v_i' нет вектора v_1 то мы выразили v через $E \setminus \{v_1\}$, если есть то подставим вместо него выражение $\beta_2 v_2 + \ldots + \beta_n v_n$ и получим выражение v только через вектора из $E \setminus \{v_1\}$, что и требовалось.

(3)⇒(1). Пусть E одновременно порождающее и линейно независимое, нам надо показать, что оно минимальное порождающее. Достаточно проверить, что для любого $v \in E$ вектор v не лежит в $\langle E \setminus \{v\} \rangle$. Действительно, пусть лежит, тогда найдутся вектора $v_1, \ldots, v_n \in E \setminus \{v\}$ и числа $\alpha_i \in F$ такие, что $v = \alpha_1 v_1 + \ldots + \alpha_n v_n$, то тогда $(-1)v + \alpha_1 v_1 + \ldots + \alpha_n v_n = 0$ – нетривиальная линейная комбинация разных элементов E, что противоречит линейной независимости E.

(2)⇒(3). Пусть E – максимальное линейно независимое, нам надо показать, что оно будет порождающим. Нам надо показать, что $\langle E \rangle = V$. Пусть это не так, возьмем $v \in V \setminus \langle E \rangle$, тогда множество $E \cup \{v\}$ строго больше, а значит линейно зависимо. То есть для каких-то $v_1, \ldots, v_n \in E \cup \{v\}$ и чисел $\alpha_i \in F$ (так что не все из них нули) выполнено $\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$. Выкинув все нулевые слагаемые, можем считать, что на самом деле все α_i не равны нулю. Если среди v_i нет v, то значит все они из E. Тогда это означает, что E линейно зависимо, что неправда. Значит один из v_i – это v. Будем считать, что $v_1 = v$. Так как по нашему предположению все коэффициенты не нулевые, то $v = v_1$ выражается через остальные $v_2, \ldots, v_n$. Но это означает, что $v \in \langle E \rangle$, а это противоречит с выбором v.

(3)⇒(2). Пусть E одновременно порождающее и линейно независимое, нам надо показать, что оно максимальное линейно независимое. Для этого возьмем любой вектор $v \in V \setminus E$ и покажем, что $E \cup \{v\}$ линейно зависимо. Действительно, мы знаем, что E порождающее, значит v представляется в виде линейной комбинации векторов из E, то есть $v = \alpha_1 v_1 + \ldots + \alpha_n v_n$ для некоторых $v_i \in E$ и $\alpha_i \in F$. Но тогда $(-1)v + \alpha_1 v_1 + \ldots + \alpha_n v_n 0$ – нетривиальная линейная комбинация векторов из $E \cup \{v\}$, то есть последнее множество линейно зависимо, что и требовалось. □

Пусть V – векторное пространство над некоторым полем F, тогда подмножество $E \subseteq V$ удовлетворяющее одному из трех эквивалентных условий предыдущего утверждения называется базисом V.

Примеры

1. Пусть $V = F^n$, тогда вектора

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, v_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

   являются базисом. Очевидно, что эти вектора линейно независимы и любой вектор через них выражается.

2. Пусть $V = F[x]$ – множество многочленов, тогда в качестве базиса можно взять $E = \{1, x, x^2, \ldots, x^n, \ldots\}$ – множество всех степеней x. Заметим, что в данном случае базис получается бесконечным.

3. Пусть X – произвольное множество и $V = \{f \colon X \to F\}$ – множество всех функций на X со значениями в F. Тогда это векторное пространство над F с очень любопытным свойством.

   Ситуация с базисами тут устроена так. Чтобы работать с бесконечными множествами нам нужно использовать аккуратно определенную теорию множеств. Я не буду вдаваться в подробности, что там да как строится, но важно понимать, что в теории множеств вообще говоря не всякое утверждение является доказуемым или опровергаемым. Подобные утверждения можно включить в качестве дополнительных аксиом, а можно их отрицания использовать в качестве таких же законных аксиом и будут получаться

совершенно разные теории множеств. Есть такая популярная аксиома «аксиома выбора», которую очень любят включать в список стандартных.

Если вы используете аксиому выбора, то можно доказать, что всякое векторное пространство имеет базис. Если же вы не используете аксиому выбора, то нельзя ни доказать, ни опровергнуть существования базиса уже в пространстве V из этого примера. Оказывается, что факт существования базиса является более слабым утверждением, чем аксиома выбора. Кроме того, если базис существует по аксиоме выбора, то это значит, что не существует никакой процедуры, которая бы помогла вам описать этот базис, потому что существование подобной процедуры дало бы вам доказательство существования базиса без аксиомы выбора.

Замечания

- Пусть V – некоторое векторное пространство и $E' \subseteq V$ – произвольное линейно независимое подмножество. Тогда его всегда можно дополнить до базиса $E \supseteq E'$, потому что базис – это максимальное линейно независимое подмножество. В случае, если существует конечный базис, это просто. А если конечного не существует, то тут придется обращаться к аккуратной формулировке аксиоматики теории множеств.
- Пусть V – некоторое векторное пространство и $E'' \subseteq V$ – произвольное порождающее множество, тогда из него всегда можно выбрать базис $E \subseteq E''$. Как и в предыдущем случае, если существует конечный базис, то это просто. А если нет конечного базиса, то это требует аккуратной аксиоматики теории множеств.

6.5 Удобный формализм

Пусть V – некоторое векторное пространство над некоторым полем F. Возьмем некоторые вектора $v_1, \ldots, v_n \in V$ и набор чисел $x_1, \ldots, x_n \in F$. Тогда можно составить строку из векторов v_i и столбец из чисел x_i и перемножить в следующем порядке

$$\begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 v_1 + \ldots + x_n v_n$$

Таким образом мы можем записывать линейные комбинации с помощью матричных объектов, когда матрицы состоят не только из чисел, но и из векторов. Если при этом ввести обозначения

$$v = \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} \quad \text{и} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

то линейную комбинацию можно записать как vx. Если $w \in V$ – некоторый вектор, то тот факт, что он линейно выражается через v_i тогда записывается так $w = vx$ для некоторого $x \in F^n$. Пусть теперь у нас есть несколько векторов $w_1, \ldots, w_m \in V$ и каждый из них выражается через вектора $v_1, \ldots, v_n$, тогда

$$w_1 = \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} A_1, \ldots, w_m = \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} A_m$$

где $A_i \in F^n$. Тогда составим из A_i матрицу $A \in \mathrm{M}_{n\,m}(F)$ и получим запись

$$\begin{pmatrix} w_1 & w_2 & \ldots & w_m \end{pmatrix} = \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} A$$

6.6 Размерность

Наша задача сейчас показать, что в векторном пространстве любые два базиса имеют одинаковое количество элементов. Однако, обсуждать как сравнивать бесконечные множества между собой я не очень хочу, потому мы с этого момента ограничимся случаями конечных базисов. Для начала нам надо показать, что если векторное пространство имеет хотя бы один конечный базис, то все его базисы конечны и имеют одинаковое количество элементов.

Утверждение 61. *Пусть V – некоторое векторное пространство над полем F и пусть $\{e_1, \dots, e_n\} \subseteq V$ – базис V. Тогда если $E \subseteq V$ – некоторый базис V, то $|E| = n$.*

Доказательство. Нам достаточно показать, что $|E| \leqslant n$. Тогда базис E становится конечным и мы можем поменять местами два базиса и применить это же утверждение для доказательства обратного неравенства.

Предположим, что это не верно, тогда в E есть хотя бы $n+1$ элемент $v_1, \dots, v_{n+1}$. Так как $e_1, \dots, e_n$ – базис, то каждый v_i линейно выражается через этот базис. Значит можно найти матрицу $A \in \mathrm{M}_{n\,n+1}(F)$ такую, что

$$\begin{pmatrix} v_1 & v_2 & \dots & v_{n+1} \end{pmatrix} = \begin{pmatrix} e_1 & e_2 & \dots & e_n \end{pmatrix} A$$

Рассмотрим систему $Ax = 0$, где $x \in F^{n+1}$. В этой системе количество столбцов больше, чем количество строк. Значит обязательно существует ненулевое решение $x \in F^{n+1}$. Тогда умножим на него предыдущее равенство справа, получим

$$\begin{pmatrix} v_1 & v_2 & \dots & v_{n+1} \end{pmatrix} x = \begin{pmatrix} e_1 & e_2 & \dots & e_n \end{pmatrix} Ax = 0$$

То есть мы нашли нетривиальную линейную комбинацию векторов $v_1, \dots, v_{n+1}$. Но по определению E в нем не должно быть линейно зависимых векторов, противоречие. □

Пусть V – векторное пространство над полем F, тогда размерностью V называется число элементов в любом из его базисов.[73] Размерность V будем обозначать через $\dim V$ или $\dim_F V$, если надо подчеркнуть, какое поле F имеется в виду.

Утверждение 62. *Пусть $U \subseteq V$ – подпространство в векторном пространстве над полем F. Тогда*

1. $\dim U \leqslant \dim V$.
2. $\dim U = \dim V$ *тогда и только тогда, когда* $U = V$.

Доказательство. (1) В начале сделаем замечание. Пусть $E \subseteq V$ – какое-то линейно независимое подмножество V. Так как базис – это максимальное линейно независимое подмножество, то $|E| \leqslant \dim V$.

Пусть $E \subseteq U$ – базис U. Тогда E – линейно независимое подмножество U, а значит и V. Но тогда из замечания выше $|E| \leqslant \dim V$. А по определению $\dim U = |E|$.

(2) Теперь сделаем еще одно замечание. Пусть $E \subseteq V$ – некоторое линейно независимое подмножество V. Как понять, что оно максимальное? Достаточно, проверить, что в нем $\dim V$ элементов. Действительно, если бы при этом оно было не максимальным, то в максимальном было бы больше $\dim V$ элементов, что противоречит определению размерности.

Теперь пусть $E \subseteq U$ – базис U и пусть $\dim U = \dim V$. Мы хотим показать, что $U = V$. Тогда E – это линейно независимое подмножество в V и в нем $|E| = \dim U = \dim V$ элементов. Но тогда по замечанию выше оно является базисом в V. Так как E – базис в U, то $U = \langle E \rangle$, а так как E – базис в V, то $V = \langle E \rangle$, то есть $U = V$. Утверждение в обратную сторону очевидно. □

6.7 Конкретные векторные пространства

Пусть V – векторное пространство над некоторым полем F. Вообще говоря, в этом случае элементы V могут быть чем угодно (функции, вектор-столбцы, матрицы, отображения и т.д.), но если в нем можно выбрать конечный базис, то оно автоматически превратится в пространство F^n. Сейчас я хочу обсудить все этапы этого магического превращения.

Пусть $e_1, \dots, e_n \in V$ – некоторый базис пространства V. Тогда можно рассмотреть отображение

$$\begin{aligned} F^n &\to V \\ x &\mapsto ex \end{aligned}$$

где $e = (e_1, \dots, e_n)$, $x \in F^n$. Так как e порождает V, то это отображение сюръективно. С другой стороны из линейной независимости следует инъективность: если $ex = ey$ для $x, y \in F^n$, то $e(x-y) = 0$, а значит

[73] Корректность этого определения следует из предыдущей леммы в случае существования хотя бы одного конечного базиса. Однако, если бы мы были знакомы с теорией мощности для произвольных множеств, то мы бы показали, что количество элементов в базисе не зависит от базиса всегда. Потому можно говорить о размерности даже для бесконечно мерных пространств. Например, размерность многочленов $F[x]$ счетная, а для бесконечного множества X размерность пространства F^X совпадает с $|F^X|$, то есть она зависит от мощности поля.

$x - y = 0$. Таким образом мы получаем, что каждый вектор-столбец длины n однозначно соответствует некоторому вектору из V. Кроме того, если присмотреться внимательно, то мы увидим, что сложение столбцов соответствует сложению векторов и то же самое верно для умножения на скаляр. Таким образом, мы видим, что между этими пространствами нет никакой разницы. Изучать одно из них – это все равно, что изучать другое. По-другому, можно думать еще так: если вам дали произвольное конечномерное пространство, то всегда можно считать, что это F^n (для этого нужно всего лишь выбрать базис).

Координаты Если вектор $v \in V$ разложен по некоторому базису $e = (e_1, \dots, e_n)$ пространства V, то есть представлен в виде $v = ex$, где $x \in F^n$, то столбец $x = (x_1, \dots, x_n)$ называют координатами вектора v в базисе $e_1, \dots, e_n$.

Смена базиса Так как базис в пространстве выбирается не единственным образом, то в конструкции выше у нас есть некоторая свобода. Давайте проследим, что и как меняется при замене одного базиса другим.

Утверждение 63. *Пусть V – некоторое векторное пространство над полем F и пусть $e_1, \dots, e_n \in V$ – какой-нибудь базис этого пространства. Тогда*

1. *Для любой обратимой матрицы $C \in \mathrm{M}_n(F)$ набор векторов $(e_1, \dots, e_n)C$ тоже является базисом.*
2. *Если $f_1, \dots, f_n \in V$ – любой другой базис V, то найдется единственная обратимая матрица $C \in \mathrm{M}_n(F)$ такая, что $(f_1, \dots, f_n) = (e_1, \dots, e_n)C$.*

Доказательство. (1) Заметим, что набор $(e_1, \dots, e_n)C$ является линейно независимым и состоит из $n = \dim V$ элементов, а значит автоматом максимальный линейно независимый набор. Для проверки линейной независимости рассмотрим линейную комбинацию $(e_1, \dots, e_n)Cx = 0$, где $x \in F^n$. Тогда $Cx = 0$, так как e_i – базис. А следовательно $x = 0$, так как C обратима.

(2) Так как e_i – базис, то любой вектор однозначно раскладывается по этому базису, например, каждый f_i имеет представление $f_i = (e_1, \dots, e_n)C_i$, где $C_i \in F^n$. Тогда все эти равенства вместе можно записать так $(f_1, \dots, f_n) = (e_1, \dots, e_n)C$, где $C = (C_1 | \dots | C_n) \in \mathrm{M}_n(F)$ – квадратная матрица, составленная из столбцов C_i. То есть такая матрица найдется, а ее единственность следует из того, что любой вектор однозначно раскладывается по базису.

Теперь осталось доказать обратимость матрицы C. Применив то же самое рассуждение для базисов в обратном порядке, мы найдем матрицу $B \in \mathrm{M}_n(F)$ такую, что $(e_1, \dots, e_n) = (f_1, \dots, f_n)B$. Тогда получаем

$$(e_1, \dots, e_n) = (f_1, \dots, f_n)B = (e_1, \dots, e_n)CB$$

А значит $(e_1, \dots, e_n)(E - CB) = 0$. Из линейной независимости базиса следует, что $E = CB$. Аналогично доказывается $BC = E$,[74] то есть B является обратной к C, что и требовалось. □

Если $e_1, \dots, e_n$ и $f_1, \dots, f_n$ – два базиса пространства V, то матрица $C \in \mathrm{M}_n(F)$ такая, что $(f_1, \dots, f_n) = (e_1, \dots, e_n)C$ называется матрицей перехода от базиса e_i к базису f_i. Таким образом у нас есть, вообще говоря бесконечный, граф с вершинами пронумерованными базисами пространства V, а ребра соответствуют матрицам перехода C от базиса e_i к базису f_i, если $(f_1, \dots, f_n) = (e_1, \dots, e_n)C$. Предыдущее утверждение говорит, что этот граф связный и между любыми двумя вершинами есть ровно одно ребро.[75]

Смена координат Пусть теперь у нас $v \in V$ – некоторый вектор. Тогда мы его можем разложить по одному базису $v = ex$ с координатами $x \in F^n$ и по другому базису $v = fy$ с координатами $y \in F^n$. Пусть C – матрица перехода от базиса e к базису f, то есть $f = eC$. Тогда координаты x в старом базисе e связаны с новыми координатами в базисе f следующим образом: $x = Cy$. Действительно, с одной стороны $v = ex$, а с другой $v = fy = eCy$. Но так как разложение по базису однозначно, получаем, что $x = Cy$. Запоминать это правило надо так: если от базиса e к базису f мы перешли с помощью умножения справа на матрицу C, то на координатах у нас отображение в обратную сторону с помощью умножения на матрицу C слева (то есть тоже с другой стороны). Еще полезно держать перед глазами вот эту таблицу.

базис	новый $\xleftarrow{\cdot C}$ старый
координаты	новые $\xrightarrow{C\cdot}$ старые

[74] Либо можно воспользоваться утверждением 3.

[75] Кстати между вершиной и ей самой тоже есть ребро-петля, соответствующая единичной матрице.

6.8 Ранг

В начале обсудим общее понятие ранга системы векторов в произвольном векторном пространстве.

Утверждение 64. *Пусть V – некоторое векторное пространство над полем F. Пусть $S = (v_1, \dots, v_k)$ – система векторов из V.*[76] *Пусть $S' \subseteq S$ – максимальный линейно независимый поднабор в S.*[77, 78] *Тогда $|S'| = \dim_F \langle S \rangle$.*

Доказательство. Рассмотрим $\langle S \rangle \subseteq V$. Если мы покажем, что $\langle S' \rangle = \langle S \rangle$, то по определению S' будет базисом $\langle S \rangle$, как порождающее и линейно независимое. Для этого нам достаточно показать, что любой вектор из $S \setminus S'$ выражается через векторы из S'. Действительно, возьмем такой вектор $v \in S \setminus S'$, тогда набор составленный из S' и v уже будет линейно зависимым, то есть есть линейная комбинация вида $\sum_i \alpha_i v'_i + \alpha v = 0$, где $v'_i \in S'$ и $\alpha_i, \alpha \in F$. Коэффициент $\alpha \neq 0$ иначе это означало бы линейную зависимость S'. А значит можно выразить v через v'_i перенеся αv направо и разделив на $-\alpha$. □

В частности это утверждение делает корректным следующее.

Определение 65. Пусть V – векторное пространство и $S = (v_1, \dots, v_k)$ – набор векторов из V. Тогда рангом S называется размер максимального линейно независимого поднабора и обозначается $\operatorname{rk} S$.

Кроме корректности, утверждение выше говорит, что $\operatorname{rk}(v_1, \dots, v_k) = \dim_F \langle v_1, \dots, v_k \rangle$. К рангу надо относиться так – это дискретный аналог размерности. Векторное пространство – объект большой и сложный, базисов в нем много, потому его недостаток – с ним сложно работать. Конечный набор векторов – объект простой и понятный, с ним намного проще работать, чем целиком со всем пространством. Однако, главный недостаток – он недостаточно гибкий по сравнению с векторным пространством, если мы чуть-чуть поменяем вектора (например, прибавим один к другому) мы уже изменим набор, но не изменим векторного пространства. Как обычно, каким-то из этих понятий удобно пользоваться в одних ситуациях, а в каких-то ситуациях намного лучше подходит другое.

6.9 Матричный ранг

Пусть $A \in \mathrm{M}_{m\,n}(F)$ – некоторая матрица с коэффициентами в поле F. Я хочу определить понятие ранга матрицы A. В начале я дам пять разных определений для ранга и чтобы отличать их во время доказательств буду временно называть их по-разному. Сразу после мы докажем, что на самом деле все эти определения эквивалентны, а значит они определяют одну и ту же характеристику матрицы, которую мы и будем называть просто рангом.[79]

Определение 66. Пусть $A_1, \dots, A_n \in F^m$ – столбцы матрицы A, то есть $A = (A_1 | \dots | A_n)$. Тогда столбцовым рангом матрицы A называется ранг системы $(A_1, \dots, A_n)$, то есть $\operatorname{rk}_{\text{столб}} A = \operatorname{rk}(A_1, \dots, A_n)$.

Определение 67. Пусть $A_1, \dots, A_m \in F^n$ – строки матрицы A, то есть $A^t = (A_1 | \dots | A_m)$. Тогда строковым рангом матрицы A называется ранг системы $(A_1, \dots, A_m)$, то есть $\operatorname{rk}_{\text{стр}} A = \operatorname{rk}(A_1, \dots, A_m)$.

Определение 68. Факториальным рангом матрицы A называется следующее число[80]

$$\min\{k \mid A = BC, \text{ где } B \in \mathrm{M}_{m\,k}(F),\, C \in \mathrm{M}_{k\,n}(F)\}$$

то есть это минимальное число k такое, что матрица A представима в виде произведения матриц BC, где общая размерность для B и C, по которой они перемножаются, есть k.

[76] Формально $S \in V^k$, то есть это упорядоченный набор векторов, где векторы могут повторяться.

[77] Это означает, что элементы S' не повторяются, полученное множество является линейно независимым и к набору S' нельзя добавить ни один вектор из S, чтобы получился линейно независимый набор.

[78] Формально мы разбирались лишь со случаем множества векторов, но я не хочу разводить формальный геморрой на ровном месте и уверен, что каждый из вас сможет без труда распространить все необходимые определения и факты на наборы вместо множеств.

[79] Я предпочитаю не фиксировать какое-то из определений как основное. Вместо этого я предпочитаю в зависимости от задачи использовать любое из них в качестве ранга, но для такой гибкости нам придется сразу доказать их совпадение.

[80] У этого определения есть один дефект. Оно «плохо работает» в случае, когда ранг равен 0. Действительно, очень сложно понять, что такое матрицы с 0 столбцов или строк. И еще удивительнее, что должно быть их произведением. Можно копаться в основаниях теории множеств, а можно просто сказать: давайте в случае $A = 0$ просто аксиоматически положим, что факториальный ранг равен 0. Это выглядит как костыль, но на самом деле если бы вы погрузились в мучительный формализм строгой теории множеств, то получили бы в точности этот ответ.

Определение 69. Тензорным рангом матрицы A называется следующее число[81]

$$\min\{k \mid A = x_1 y_1^t + \ldots + x_k y_k^t, \text{ где } x_i \in F^m,\ y_i \in F^n\}$$

то есть это минимальное число k такое, что матрица A представима в виде суммы k «тощих» матриц вида xy^t, где $x \in F^m$ и $y \in F^n$.

Перед следующим определением нам нужна некоторая подготовка. Зафиксируем некоторый набор индексов для строк: $1 \leqslant i_1 < \ldots < i_k \leqslant m$, а так же некоторый набор индексов для столбцов в том же количестве $1 \leqslant j_1 < \ldots < j_k \leqslant n$. Тогда обозначим через $A^{j_1,\ldots,j_k}_{i_1,\ldots,i_k} \in \mathrm{M}_k(F)$ подматрицу образованную пересечением данных строк и столбцов. То есть формально ее элементы это $\bar{a}_{st} = a_{i_s j_t}$. Такую матрицу будем называть квадратной подматрицей матрицы A размера k.

Определение 70. Минорным рангом матрицы A называется размер максимальной невырожденной квадратной подматрицы, то есть минорный ранг A – это такое k, что существует невырожденная подматрица $A^{j_1,\ldots,j_k}_{i_1,\ldots,i_k} \in \mathrm{M}_k(F)$ такая, что любая квадратная подматрица ее содержащая уже вырождена.[82]

В начале сделаем очень полезное замечание.

Утверждение 71. *Для любой матрицы $A \in \mathrm{M}_{m\,n}(F)$ ее факториальный ранг равен тензорному.*

Доказательство. Пусть $B \in \mathrm{M}_{m\,k}(F)$ и $C \in \mathrm{M}_{k\,n}(F)$ – такие матрицы, что $A = BC$. Пусть $B = (B_1|\ldots|B_k)$ и $C^t = (C_1|\ldots|C_k)$. Тогда по блочным формулам $BC = B_1C_1^t + \ldots + B_kC_k^t$. Это доказывает, что тензорный ранг A не превосходит факториального. Наоборот, если задано разложение $BC = B_1C_1^t + \ldots + B_kC_k^t$, то определим матрицы $B = (B_1|\ldots|B_k)$ и $C^t = (C_1|\ldots|C_k)$ и получим, что $A = BC$. Это доказывает оценку рангов в другую сторону. □

Наша цель в этом разделе очень проста – показать, что все пять определений ранга совпадают между собой. Начнем со следующего.

Утверждение 72. *Пусть $A \in \mathrm{M}_{m\,n}(F)$ – произвольная матрица, тогда столбцовый, строковый, факториальный и тензорный ранги не меняются при домножении A слева или справа на невырожденную матрицу.*

Доказательство. (1) Столбцовый ранг. Пусть $A = (A_1|\ldots|A_n)$, где $A_i \in F^m$ и пусть $D \in \mathrm{M}_n(F)$ – обратимая матрица. Тогда $\langle(A_1,\ldots,A_n)\rangle = \langle(A_1,\ldots,A_n)D\rangle$ (включение $\supseteq$ очевидно, а обратное следует из обратимости D), а значит

$$\mathrm{rk}_{\text{столб}}(A) = \dim\langle(A_1,\ldots,A_n)\rangle = \dim\langle(A_1,\ldots,A_n)D\rangle = \mathrm{rk}_{\text{столб}}(AD)$$

Пусть теперь $C \in \mathrm{M}_m(F)$ – обратимая матрица. Тогда система $Ax = 0$ эквивалентна системе $CAx = 0$. Если какая-то линейная комбинация $x_1A_1 + \ldots + x_nA_n = 0$, то это значит, что $Ax = 0$, т.е. x является решением системы $Ax = 0$, а значит и решением системы $CAx = 0$, то есть столбцы матрицы CA удовлетворяют той же самой линейной комбинации, что и столбцы матрицы A. Это значит, что если какое-то множество столбцов в A было линейно независимо, то множество столбцов с теми же самыми номерами в CA тоже линейно независимо. И если какой-то столбец из A выражался через другие, то и в CA столбец с тем же номером будет выражаться с помощью той же самой линейной комбинации через другие. Потому максимальная линейно независимая система в A переходит в максимальную линейно независимую систему в CA.[83] Последнее по определению означает $\mathrm{rk}_{\text{столб}}\, A = \mathrm{rk}_{\text{столб}}(CA)$.

(2) Строковый ранг. Так как $\mathrm{rk}_{\text{столб}}\, A = \mathrm{rk}_{\text{стр}}(A^t)$, то этот случай следует из предыдущего.

(3) Факториальный ранг. Пусть $\mathrm{rk}_{\text{ф}}\, A = k$ и пусть $A = BC$ – разложение на котором достигается ранг A. Тогда $AD = B(CD)$ – некоторое разложение для AD с числом k, а значит по определению $\mathrm{rk}_{\text{ф}}(AD) \leqslant \mathrm{rk}_{\text{ф}}\, A$. Обратное неравенство следует из обратимости D, т.е. $AD = (AD)D^{-1}$. Домножение на матрицу слева делается аналогично.

(4) В силу замечания выше (утверждение 71) тензорный ранг и факториальный – это одно и то же, потому этот пункт следует из предыдущего. □

[81] В этом определении как и в предыдущем очень сложно понять, что значит сумма пустого числа слагаемых (опять случай $k = 0$). Опять же, это случается только в случае $A = 0$ и потому, в этом случае мы по определению скажем, что тензорный ранг равен 0. И точно так же как и выше (в случае факториального ранга) оказывается, что этот ответ можно вытащить из строгой теории множеств, но я не хочу играть в компилятор формализма.

[82] Обратите внимание, что нужно еще доказывать корректность этого определения, а именно, что число k не зависит от выбора максимальной невырожденной подматрицы. Это вообще говоря не очевидно.

[83] Еще один способ думать про это доказательство такое. Можно считать, что столбцы матрицы A – это векторы в F^m, а умножение на C слева – это замена координат в пространстве F^n, то есть мы ничего не делаем с нашими векторами, но меняем стандартный базис в F^n на какой-то другой. Потому столбцы CA – это координаты тех же самых векторов, что и исходные, только записанные в другом базисе. А раз это те же самые векторы, то у нас ничего не поменялось, кроме их « внешнего вида».

Утверждение 73. *Пусть $A \in \mathrm{M}_{m\,n}(F)$, тогда столбцовый, строковый, факториальный и тензорный ранги для нее совпадают.*

Доказательство. Домножив матрицу A слева и справа на обратимую, мы можем считать, что она имеет следующий вид $\left(\begin{smallmatrix} E & 0 \\ 0 & 0 \end{smallmatrix}\right)$. Из утверждения 72 следует, что достаточно доказать утверждение для последней матрицы.

Давайте посчитаем $\mathrm{rk}_{\text{столб}} \left(\begin{smallmatrix} E & 0 \\ 0 & 0 \end{smallmatrix}\right) = r$, где r – размер единичной матрицы E. Аналогично $\mathrm{rk}_{\text{стр}} \left(\begin{smallmatrix} E & 0 \\ 0 & 0 \end{smallmatrix}\right) = r$. Более того, для минорного ранга $\mathrm{rk}_{\text{м}} \left(\begin{smallmatrix} E & 0 \\ 0 & 0 \end{smallmatrix}\right) = r$, так как матрица E является невырожденной матрицей размера r, а все большие подматрицы вырождены, потому что имеют нулевую строку или столбец. То есть эти три ранга равны между собой.

Теперь осталось доказать, что факториальный ранг совпадает со столбцовым (строковым) рангом. Если $A = BC = (B_1|\ldots|B_k)C$ – равенство, на котором достигается факториальный ранг. Тогда столбцы матрицы A выражаются через столбцы матрицы B, то есть $\langle A_1, \ldots, A_m\rangle \subseteq \langle B_1, \ldots, B_k\rangle$, а значит

$$\mathrm{rk}_{\text{столб}}\, A = \dim\langle A_1, \ldots, A_m\rangle \leqslant \dim\langle B_1, \ldots, B_k\rangle \leqslant k = \mathrm{rk}_{\text{ф}}\, A$$

С другой стороны

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

где общая размерность матриц справа равна строковому рангу, а значит факториальный ранг не превосходит этой размерности. Что дает обратное неравенство. □

Утверждение 74. *Пусть $A \in \mathrm{M}_{m\,n}(F)$ – произвольная матрица, тогда ее минорный ранг корректно определен и совпадает со всеми остальными рангами.*

Доказательство. Давайте напомним, что минорным рангом мы назвали размер максимальной невырожденной квадратной подматрицы в матрице A, то есть мы ищем подматрицу $A^{j_1,\ldots,j_k}_{i_1,\ldots,i_k}$ натянутую на строки $i_1, \ldots, i_k$ и столбцы $j_1, \ldots, j_k$ такую, что она сама является невырожденной, а все квадратные подматрицы ее содержащие (если такие имеются) уже вырождены. Во-первых, не понятно, почему это число не зависит от выбора квадратной подматрицы, а во-вторых, почему оно совпадает со всеми остальными рангами. Мы поступим следующим образом. Выберем произвольную такую максимальную невырожденную квадратную подматрицу и докажем, что ее размер совпадает со столбцовым рангом. Тогда отсюда будет следовать, что ее размер не зависит от выбора подматрицы и что минорный ранг совпадает со всеми остальными.

Заметим, что если я переставлю строки и столбцы в матрице A, то ее квадратные подматрицы как-то переставятся местами. При этом если какая-то подматрица была максимальная невырожденная, то она останется максимальной невырожденной. Следовательно можно считать, что максимальная невырожденная квадратная подматрица натянута на первые k строк и столбцов и матрица A выглядит так

$$A = \begin{pmatrix} B & C \\ D & F \end{pmatrix} \quad \text{где } B \in \mathrm{M}_k(F) \text{ невырождена}$$

Наша задача показать, что $k = \mathrm{rk}_{\text{столб}}\, A$. Теперь будем делать элементарные преобразования строк и столбцов первого типа, когда разрешается прибавлять строку с номером от 1 до k к любой другой строке (аналогично со столбцами). Заметим, что при этом не меняется определитель подматрицы B и любой другой подматрицы содержащей матрицу B. Поэтому, при таких преобразованиях матрица B будет продолжать оставаться максимальной невырожденной матрицей.

С помощью указанных выше преобразований мы можем привести матрицу B к диагональному виду с ненулевыми числами на диагонали

$$A \mapsto \begin{pmatrix} b_1 & \ldots & 0 & * & \ldots & * \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & b_k & * & \ldots & * \\ * & \ldots & * & * & \ldots & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & \ldots & * & * & \ldots & * \end{pmatrix}$$

Теперь с помощью указанных элементарных преобразований строк можно занулить весь блок под левой верхней диагональной подматрицей, а с помощью указанных элементарных преобразований столбцов – весь

блок справа от нее.

$$\begin{pmatrix} b_1 & \dots & 0 & * & \dots & * \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & b_k & * & \dots & * \\ * & \dots & * & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & \dots & * & * & \dots & * \end{pmatrix} \mapsto \begin{pmatrix} b_1 & \dots & 0 & * & \dots & * \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & b_k & * & \dots & * \\ 0 & \dots & 0 & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & * & \dots & * \end{pmatrix} \mapsto \begin{pmatrix} b_1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & b_k & 0 & \dots & 0 \\ 0 & \dots & 0 & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & * & \dots & * \end{pmatrix}$$

При этом подматрица B в верхнем левом углу остается максимальной невырожденной подматрицей. В частности, если мы возьмем и добавим к ней i-ю строку и j-й столбец, то получится вырожденная матрица. С другой стороны эта матрица будет иметь вид

$$\begin{pmatrix} b_1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & b_k & 0 \\ 0 & \dots & 0 & d_{ij} \end{pmatrix}$$

Здесь через d_{ij} обозначен элемент на i-ой строке и j-ом столбце последней матрицы. Так как все b_i не равны нулю, то такая матрица может быть вырождена только если $d_{ij} = 0$. Применяя это рассуждение к произвольному столбцу и строке с номерами больше k, мы видим, что последняя матрица на самом деле имеет вид

$$A' = \begin{pmatrix} b_1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & b_k & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}$$

Теперь мы видим, что $k = \operatorname{rk}_{\text{столб}} A'$. Но так как мы перешли от матрицы A к матрице A' с помощью элементарных преобразований, то $\operatorname{rk}_{\text{столб}} A = \operatorname{rk}_{\text{столб}} A'$ по утверждению 72. □

Определение 75. Пусть $A \in \mathrm{M}_{m\,n}(F)$ – произвольная матрица. Тогда рангом A называется один из пяти рангов определенных выше и обозначается $\operatorname{rk} A$.

Теорема Кронекера-Капэлли Пользуясь определением ранга можно сделать следующее замечание.

Утверждение 76. *Пусть F – некоторое поле, $A \in \mathrm{M}_{m\,n}(F)$ и $b \in F^m$. Тогда система $Ax = b$ имеет решение тогда и только тогда, когда $\operatorname{rk}(A) = \operatorname{rk}(A|b)$.*

6.10 Подпространства в F^n

Давайте посмотрим как можно задавать подпространства в F^n. Существует два способа

Явный	**Неявный**
Если $v_1, \dots, v_k \in V$, тогда $U = \langle v_1, \dots, v_k \rangle$	Если $A \in \mathrm{M}_{m\,n}(F)$, тогда $U = \{y \in F^n \mid Ay = 0\}$

По-хорошему, хочется научиться пересчитывать векторное пространство заданное в одной из этих форм в другую. Мы разберем пока только одну из этих задач. А именно, пусть подпространство задано неявно в виде системы, то как найти его базис?

Если подпространство $U \subseteq F^n$ задано в виде $U = \{y \in F^n \mid Ay = 0\}$ для некоторой матрицы $A \in \mathrm{M}_{m\,n}(F)$, то любой базис пространства U будем называть фундаментальной системой решений (ФСР). Ниже мы разберем задачу построения какого-нибудь ФСР для однородной системы линейных уравнений.

Нахождение ФСР однородной СЛУ В начале мы приведем алгоритм находящий ФСР, а потом объясним почему он работает.

Дано Система однородных линейных уравнений $Ax = 0$, где $A \in \mathrm{M}_{m\,n}(F)$ и $x \in F^n$.

Задача Найти ФСР системы $Ax = 0$.

Алгоритм

1. Привести матрицу A элементарными преобразованиями строк к улучшенному ступенчатому виду. Например
$$A' = \begin{pmatrix} 1 & 0 & a_{31} & 0 & a_{51} \\ 0 & 1 & a_{32} & 0 & a_{52} \\ 0 & 0 & 0 & 1 & a_{53} \end{pmatrix}$$

2. Пусть $k_1, \ldots, k_r$ – позиции свободных переменных. Если положить одну из этих переменных равной 1, а все остальные нулями, то существует единственное решение, которое мы обозначим через u_i (всего r штук). Например, для матрицы A' выше свободные переменные имеют номера 3 и 5. Тогда вектора (записанные в строку)
$$u_1 = \begin{pmatrix} -a_{31} & -a_{32} & 1 & 0 & 0 \end{pmatrix}, u_2 = \begin{pmatrix} -a_{51} & -a_{52} & 0 & -a_{53} & 1 \end{pmatrix}$$
являются ФСР.

Доказательство корректности алгоритма поиска ФСР. Пусть в общем виде, ступенчатый вид матрицы A выглядит так
$$\begin{pmatrix} 1 & * & 0 & * & 0 & * & * & 0 & * \\ & & 1 & * & 0 & * & * & 0 & * \\ & & & & 1 & * & * & 0 & * \\ & & & & & & & 1 & * \end{pmatrix}$$
Тогда построенные вектора имеют вид
$$\begin{array}{c|ccccccccc} & & k_1 & & k_2 & & \ldots & \ldots & & k_r \\ u_1 & (* & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0) \\ u_2 & (* & 0 & * & 1 & 0 & 0 & 0 & 0 & 0) \\ \vdots & (* & 0 & * & 0 & * & 1 & 0 & 0 & 0) \\ \vdots & (* & 0 & * & 0 & * & 0 & 1 & 0 & 0) \\ u_r & (* & 0 & * & 0 & * & 0 & 0 & * & 1) \end{array}$$
В начале проверим, что u_i линейно независимы. Действительно, тогда линейная комбинация $\alpha_1 u_1 + \ldots + \alpha_r u_r$ имеет вид
$$\begin{pmatrix} * & \alpha_1 & * & \alpha_2 & * & \ldots & \ldots & * & \alpha_r \end{pmatrix}$$
Если эта линейная комбинация равна нулю, то значит и все α_i равны нулю.

Теперь пусть v – произвольное решение системы $Ax = 0$. Посмотрим на его координаты в свободных позициях
$$\begin{pmatrix} * & v_1 & * & v_2 & * & \ldots & \ldots & * & v_r \end{pmatrix}$$
Теперь рассмотрим вектор $w = v - v_1 u_1 - \ldots - v_r u_r$. С одной стороны это решение системы $Ax = 0$. С другой стороны у этого решения все свободные переменные равны нулю. А значит автоматически и все главные переменные равны нулю, что означает, что $w = 0$. То есть $v = v_1 u_1 + \ldots + v_r u_r$, что и требовалось. □

Замечание

- Обратите внимание, что ФСР – это любой базис в пространстве $\{y \in F^n \mid Ay = 0\}$, а не только тот, который построен по алгоритму.
- В алгоритме выше, мы могли бы вместо 1 и 0 расставить любой набор из r линейно независимых векторов длины r в позиции со свободными переменными. Это тоже дало бы базис. Однако, у построенного ФСР именно по алгоритму выше есть одно важное преимущество: в нем легко считать координаты. Действительно, для любого вектора из пространства решений его свободные переменные – это и есть координаты в построенном базисе.

7 Линейные отображения

7.1 Идея и определение

Пусть F – некоторое поле и нам даны два векторных пространства V и U. Самый первый вопрос, который встает глядя на них: а как их сравнить? Одно и то же ли это пространство? Для того, чтобы отвечать на подобные вопросы, нам и нужны линейные отображения. Давайте начнем с более простого вопроса: а как сравнивать множества? Множества сравниваются с помощью отображений. Всевозможные теоремы об эквивалентности разных подходов с помощью инъективных, сюръективных и биективных отображений в конце концов говорят, что два множества надо считать одинаковыми, если между ними есть биекция, то есть в них одинаковый запас элементов. Но теперь векторное пространство – это не просто множество, на нем еще есть и операции. Если на биекцию смотреть, как на способ переименовать элементы, то теперь мы хотим, чтобы при этом переименовании одни операции превращались в другие, то есть мы хотим, чтобы наше отображение было согласовано неким способом с операциями. Ну и как в случае множеств, мы не будем ограничивать себя только биекциями, ибо все остальные отображения тоже оказываются очень полезными для сравнения разных векторных пространств.

Определение 77. Пусть V и U – векторные пространства над некоторым полем F. Отображение $\varphi\colon V \to U$ называется линейным, если оно удовлетворяет двум свойствам:

1. $\varphi(v_1 + v_2) = \varphi(v_1) + \varphi(v_2)$ для всех $v_1, v_2 \in V$.
2. $\varphi(\lambda v) = \lambda\varphi(v)$ для всех $\lambda \in F$ и всех $v \in V$.

Примеры

1. $F^n \to F^m$ задано по правилу $x \mapsto Ax$, где $A \in \mathrm{M}_{m\,n}(F)$.
2. $V \to U$ задано по правилу $v \mapsto 0$ для любого $v \in V$. Такое отображение называется нулевым и обозначается 0.
3. $V \to V$ задано по правилу $v \mapsto v$. Такое отображение называется тождественным и обозначается Id.
4. Пусть $D[0,1]$ – множество дифференцируемых функций на отрезке $[0,1]$, $F[0,1]$ – множество всех функций на отрезке $[0,1]$. Тогда зададим линейное отображение $D[0,1] \to F[0,1]$ по правилу $f \mapsto f'$, то есть взятие производной является линейным отображением.
5. Пусть $L[0,1]$ – множество функций f на отрезке $[0,1]$ таких, что существует интеграл $\int\limits_0^1 |f(x)|\,dx$[84] и пусть $C[0,1]$ – множество непрерывных функций на отрезке $[0,1]$. Тогда рассмотрим следующее отображение $f \mapsto \int\limits_0^t f(x)\,dx$.
6. Пусть V – некоторое векторное пространство с базисом $e_1, \ldots, e_n$. Тогда отображение $F^n \to V$ по правилу $x \mapsto (e_1, \ldots, e_n)x$ является линейным.

Замечание Давайте посмотрим на примеры (1), (4) и (5). Они говорят, что в терминах линейных отображений можно говорить о системах линейных уравнений (1), о дифференциальных уравнениях (4) или об интегральных уравнениях (5). Пример (6) нам встречался, когда мы объясняли, что любое векторное пространство превращается в пространство столбцов, если в нем выбрать базис.

Определение 78. Пусть V и U – векторные пространства над некоторым полем F, тогда отображение $\varphi\colon V \to U$ называется изоморфизмом, если

1. φ линейно.
2. φ биективно.

[84] Для тех кто беспокоится о том, какой же тут берется интеграл, сообщим по честному, что на самом деле тут у нас функции интегрируемые по Лебегу на отрезке $[0,1]$.

Заметим, что если отображение $\varphi\colon V \to U$ изоморфзим, то существует обратное отображение (потому что φ биекция) и обратное так же является линейным.[85] Если между двумя векторными пространствами V и U существует изоморфизм, то говорят, что они изоморфны и пишут $V \simeq U$.

Множество всех линейных отображений из пространства V в пространство U обозначается через $\mathrm{Hom}_F(V,U)$. Формально

$$\mathrm{Hom}_F(V,U) = \{f\colon V \to U \mid f \text{ – линейное}\}$$

Множество линейных отображений из V в поле F называется двойственным пространством к V и обозначается через V^*. Формально

$$V^* = \{\xi\colon V \to F \mid \xi \text{ – линейное}\} = \mathrm{Hom}_F(V,F)$$

Если нет путаницы, то индекс, обозначающий поле F опускают и пишут просто $\mathrm{Hom}(V,U)$.[86]

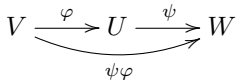
7.2 Операции на линейных отображениях

Давайте порадуем себя еще немного абстрактным формализмом и введем операции на линейных отображениях. Причем мы постараемся так, что множество всех линейных отображений между двумя векторными пространствами $\mathrm{Hom}_F(V,U)$ внезапно превратится в векторное пространство. А это не так уж и плохо, это значит, что постоянно оставаясь в рамках векторных пространств, мы сможем к новым конструкциям применять все те же методы, что мы изначально разрабатывали для произвольных векторных пространств.

Сумма Пусть $\varphi\colon V \to U$ и $\psi\colon V \to U$ два линейных отображения между векторными пространствами V и U. Тогда, чтобы определить отображение $(\varphi+\psi)\colon V \to U$, надо задать его действие на всех векторах из V и проверить, что полученное правило линейно. Зададим его так: для любого $v \in V$ положим $(\varphi+\psi)(v) = \varphi(v) + \psi(v)$.

Умножение на скаляр Пусть $\varphi\colon V \to U$ – линейное отображение между векторными пространствами V и U и пусть $\lambda \in F$ – произвольный элемент поля. Тогда определим $(\lambda\varphi)\colon V \to U$ как отображение задаваемое правилом $(\lambda\varphi)(v) = \lambda\varphi(v)$.

Композиция Пусть $\varphi\colon V \to U$ и $\psi\colon U \to W$ – два линейных отображения, а V, U и W – векторные пространства. Тогда теоретико множественная композиция отображений $\psi \circ \varphi\colon V \to W$ это отображение задаваемое по правилу $(\psi \circ \varphi)(v) = \psi(\varphi(v))$. Методом пристального взгляда на определение композиции и линейного отображения проверяется, что композиция тоже является линейным отображением. Теоретико множественная композиция тогда обозначается просто $\psi\varphi$ и называется композицией линейных отображений. Здесь полезно иметь перед глазами картинку:



Замечание Можно проверить методом пристального взгляда, что множество $\mathrm{Hom}_F(V,U)$ с операциями сложения и умножения на скаляр превращается в векторное пространство над полем F, а значит и двойственное пространство V^* тоже является векторным пространством над F.[87]

7.3 Построение линейных отображений

Мы хорошенько поиграли в абстракции, а теперь внимание главный вопрос дня: а как задавать линейные отображения между двумя векторыми пространствами V и U? И как следствие другой вопрос: а вообще существуют хоть какие-нибудь линейные отображения между V и U? К счастью на второй вопрос мы уже знаем ответ с помощью примера (2) выше. Нулевое отображение у нас есть всегда, но это совсем не интересно. На самом деле вопрос задания объектов нового вида – это один из самых важных вопросов. Когда мы проходили перестановки, нам нужен был способ с ними работать, работая с действительными числами, мы даже

[85] Предлагаю дотошным читателям проверить это заявление, а остальным пустить скупую слезу и, поверив мне на слово, двигаться дальше в мир неизведанного.

[86] Для интересующихся таким странным обозначением поясню, Hom происходит от слова homomorphism – гомоморфизм, которое является более общим определением и в случае векторных пространств дает в точности понятие линейного отображения.

[87] Глупо было бы ожидать другого результата.

не задумываемся над способом их задания, так как эти способы нам знакомы с детства, задавать функции многих из нас тоже научили в школе. Теперь пришло время научиться работать с линейными отображениями.

Утверждение 79. *Пусть V и U – векторные пространства и пусть $e_1, \ldots, e_n$ – базис пространства V. Тогда для любого набора $u_1, \ldots, u_n \in U$[88] существует единственное линейное отображение $\varphi\colon V \to U$ такое, что $\varphi(e_i) = u_i$.*

Доказательство. Давайте проверим единственность в начале. Пусть $\varphi, \psi\colon V \to U$ – два таких отображения, тогда для любого вектора $v \in V$ имеем $v = x_1e_1 + \ldots + x_ne_n$. А значит

$$\varphi(v) = \varphi(x_1e_1 + \ldots + x_ne_n) = x_1\varphi(e_1) + \ldots + x_n\varphi(e_n) = x_1u_1 + \ldots + x_nu_n$$

Аналогично $\psi(v)$ равно тому же самому. Потому $\varphi = \psi$.[89]

Теперь займемся вопросом существования. Для этого надо всего лишь проверить, что правило $\varphi(v) = x_1u_1 + \ldots + x_nu_n$ задает линейное отображение. Для этого надо проверить, что $\varphi(v_1 + v_2) = \varphi(v_1) + \varphi(v_2)$ и $\varphi(\lambda v) = \lambda\varphi(v)$. Я молча выпишу ниже две необходимые для проверки строчки и понадеюсь на вашу сознательность и сообразительность в завершении доказательства. Пусть $e = (e_1, \ldots, e_n)$ и $u = (u_1, \ldots, u_n)$, тогда для $v \in V$ найдется единственный $x \in F^n$, что $v = ex$. А значит выражение $ux \in U$ зависит только от v, а не от выбора координат. Потому можно положить $\varphi(v) = ux$. Теперь линейность следует из равенств

$$\varphi(v_1 + v_2) = \varphi(ex_1 + ex_2) = \varphi(e(x_1 + x_2)) = u(x_1 + x_2) = ux_1 + ux_2 = \varphi(v_1) + \varphi(v_2)$$
$$\varphi(\lambda v) = \varphi(\lambda ex) = \varphi(e(\lambda x)) = u(\lambda x) = \lambda ux = \lambda\varphi(v)$$

□

Таким образом, если у вас зафиксирован базис в пространстве, то чтобы построить линейное отображение достаточно отправить базисные векторы куда угодно и у вас автоматом будет только одно линейное отображение, действующее на базисе таким образом.

Классификация векторных пространств с точностью до изоморфизма Теперь мы можем дать критерий, когда два векторных пространства изоморфны.

Утверждение 80. *Пусть V и U – векторные пространства над полем F. Они изоморфны тогда и только тогда, когда $\dim_F V = \dim_F U$.*

Доказательство. Если векторные пространства V и U изоморфны, это значит, что существует изоморфизм $\varphi\colon V \xrightarrow{\sim} U$. Остается заметить, что при изоморфизме линейно независимое и порождающее множество переходит в линейно независимое и порождающее. А значит, базис переходит в базис. Следовательно пространства имеют одинаковую размерность.

Теперь пусть у пространств одинаковая размерность. Тогда пусть $e_1, \ldots, e_n$ – базис V и $f_1, \ldots, f_n$ – базис U. Обратим внимание, что их количество одинаковое, так как это и есть их размерности, которые совпадают между собой по условию. Тогда по предыдущему утверждению существует единственное отображение $\varphi\colon V \to U$ такое, что $e_i \mapsto f_i$. Аналогично существует обратное отображение $\psi\colon U \to V$ переводящее $f_i \mapsto e_i$. А следовательно композиции $\psi\varphi$ и $\varphi\psi$ оставляют на месте базисы пространств V и U соответственно. Последнее означает, что $\psi\varphi = \mathrm{Id}_V$ и $\varphi\psi = \mathrm{Id}_U$. То есть это обратимые отображения, то есть биекции. □

7.4 Линейные отображения и базис

Матрица линейного отображения Пусть в векторном пространстве V зафиксирован базис $e_1, \ldots, e_n$, а в пространстве U базис $f_1, \ldots, f_m$. Тогда любое линейное отображение $\varphi\colon V \to U$ задается набором $\varphi(e_1, \ldots, e_n) = (\varphi(e_1), \ldots, \varphi(e_n))$.[90] Любой вектор $\varphi(e_i)$ можно разложить по базису $f_1, \ldots, f_m$, то есть есть равенство $\varphi(e_i) = (f_1, \ldots, f_m)A_i$, где $A_i \in F^m$ – столбец коэффициентов образа e_i по $f_1, \ldots, f_m$. Тогда все эти равенства вместе можно записать так:

$$\varphi(e_1, \ldots, e_n) = (f_1, \ldots, f_m)A, \text{ где } A_\varphi = (A_1|\ldots|A_n) \in \mathrm{M}_{m\,n}(F)$$

[88] Векторам u_i разрешено повторяться.

[89] Чтобы проверить, что два отображения равны, как раз и надо проверить, что они совпадают на любом элементе, а мы ровно это и сделали.

[90] Это равенство можно понимать, как произведение матрицы 1 на 1 с элементом из $\mathrm{Hom}_F(V, U)$, с вектором размера 1 на n с элементами из пространства V.

Как только фиксированы базисы e_i и f_i, матрица A_φ определена однозначно. Эта матрица и называется матрицей линейного отображения φ в базисах e_i и f_i.[91] Еще раз повторим в кратких обозначениях, если $e = (e_1, \dots, e_n)$ и $f = (f_1, \dots, f_m)$ то, чтобы показать, что какая-то матрица A является матрицей линейного отображения $\varphi\colon V \to U$ в базисах e и f, то надо показать равенство $\varphi e = fA$.

Действие в координатах Пусть теперь $v \in V$. Тогда вектор v можно разложить по базису e в виде $v = ex$, для некоторого $x \in F^n$. Применим φ к v, получим

$$\varphi(v) = \varphi(ex) = (\varphi e)x = fA_\varphi x$$

То есть $\varphi(v)$ имеет координаты $A_\varphi x$. То есть, если отождествить пространство V с F^n посредством $v = ex \mapsto x$ и пространство U с F^m посредством $u = fy \mapsto y$, то наше линейное отображение φ превращается в отображение $F^n \to F^m$ по правилу $x \mapsto A_\varphi y$. Таким образом изучать линейное отображение между двумя конечномерными векторными пространствами – это все равно, что изучать отображение между пространствами столбцов, заданное умножением слева на матрицу.

Смена базиса Пусть теперь в пространстве V задано два базиса $e = (e_1, \dots, e_n)$ и $e' = (e'_1, \dots, e'_n)$. Аналогично, в пространстве U – два базиса $f = (f_1, \dots, f_m)$ и $f' = (f'_1, \dots, f'_m)$. Кроме того выполнены равенства $e' = eC$ для некоторой $C \in \mathrm{M}_n(F)$ и $f' = fD$ для некоторой $D \in \mathrm{M}_m(F)$, то есть D и C – матрицы перехода от нештрихованных базисов к штрихованным. Тогда в базисах e и f линейное отображение $\varphi\colon V \to U$ выглядит $\varphi e = fA_\varphi$, а в базисах e' и f' – $\varphi e' = f'A'_\varphi$. Давайте поймем какая связь между матрицами A_φ и A'_φ в терминах матриц перехода C и D.

$$\varphi e' = f'A'_\varphi \Rightarrow \varphi eC = fDA'_\varphi \Rightarrow \varphi e = fDA'_\varphi C^{-1}$$

В силу единственности матрицы линейного отображения

$$A_\varphi = DA'_\varphi C^{-1} \Rightarrow A'_\varphi = D^{-1}A_\varphi C$$

Последнее равенство показывает как меняется матрица линейного отображения, когда мы меняем базисы пространств. Запоминать можно так: когда мы хотим заменить старую матрицу A_φ на новую A'_φ надо справа домножить на матрицу перехода от e к e', а слева на обратную к матрице перехода от f к f'. Еще полезно перед глазами держать следующую картинку:

$$\begin{array}{lcc} \text{старые} & F^n \xrightarrow{\varphi} F^m & x = Cx' \longmapsto y = A_\varphi Cx' \\ & \uparrow \quad \downarrow & \uparrow \quad \downarrow \\ \text{новые} & F^n \xrightarrow{\varphi} F^m & x' \longmapsto y' = D^{-1}A_\varphi Cx' \end{array}$$

Связь матрицы отображения с операциями

Утверждение 81. *Пусть V и U – некоторые векторные пространства над полем F размерностей n и m соответственно и пусть $e = (e_1, \dots, e_n)$ и $f = (f_1, \dots, f_m)$ – базисы пространств V и U. Тогда отображение*

$$\begin{gathered} \mathrm{Hom}_F(V, U) \to \mathrm{M}_{m\,n}(F) \\ \varphi \mapsto A_\varphi \end{gathered}$$

является изоморфизмом векторных пространств, то есть выполнено

1. $A_{\varphi+\psi} = A_\varphi + A_\psi$.
2. $A_{\lambda\varphi} = \lambda A_\varphi$.

[91] Как обычно я злоупотребляю своей любовью к опусканию чрезмерных деталей в обозначениях. Надо помнить, что вообще говоря матрица A_φ не имеет значения, если не сказано в каких именно базисах она посчитана. Потому хорошо бы еще писать, что это $A_{\varphi\, e_i\, f_i}$, но на практике не удобно использовать такое громоздкое обозначение, да и не особенно много проку от него.

Доказательство. Биективность этого отображения мы уже знаем в других терминах: при зафиксированных базисах e и f в обоих пространствах для любого линейного отображения φ существует единственная матрица A_φ такая, что $\varphi e = fA_\varphi$.

Теперь надо проверить линейность этого отображения. По определению $(\varphi + \psi)e = fA_{\varphi+\psi}$. С другой стороны

$$(\varphi + \psi)e = \varphi e + \psi e = fA_\varphi + fA_\psi = f(A_\varphi + A_\psi)$$

Из единственности следует, что $A_{\varphi+\psi} = A_\varphi + A_\psi$. Аналогично, $(\lambda\varphi)e = fA_{\lambda\varphi}$ с одной стороны и

$$(\lambda\varphi)e = \lambda(\varphi e) = \lambda f A_\varphi = f(\lambda A_\varphi)$$

с другой. Потому из единственности получаем $A_{\lambda\varphi} = \lambda A_\varphi$. □

Таким образом изучать линейные отображения между двумя векторными пространствами – это все равно что изучать матрицы, причем с учетом операций на матрицах, а не просто как множество.

Утверждение 82. *Пусть V, U и W – векторные пространства размерностей m, n и k соответственно с базисами e, f и h соответственно. Пусть $\varphi\colon V \to U$ и $\psi\colon U \to W$, тогда $A_{\psi\varphi} = A_\psi A_\varphi$.*

Доказательство. Доказательство – прямая проверка в лоб по определению. С одной стороны $(\psi\varphi)e = hA_{\psi\varphi}$. С другой

$$(\psi\varphi)e = \psi(\varphi e) = \psi(fA_\varphi) = (\psi f)A_\varphi = hA_\psi A_\varphi$$

И результат следует из единственности. □

Все эти результаты вместе взятые означают, что выбрав базис, мы можем отождествить все конечномерные пространства с какими-то конкретными пространствами вида F^n, а линейные отображения между ними на матрицы соответствующих размеров. Тогда изучать конечно мерные векторные пространства – это то же самое, что изучать конкретные векторные пространства столбцов с матричными отображениями между ними. Потому, если что-то глобально доказано для вектор столбцов, оно автоматом есть и для конечно мерных векторных пространств. Более того, если какой-то факт о векторах бесконечномерного векторного пространства требует лишь конечного набора векторов из него, то они живут в конечномерном подпространстве, а значит подобные факты для бесконечно мерных пространств автоматически следуют из случая конечномерных, а значит и из случая вектор столбцов.

7.5 Ядро и образ

Определение 83. Пусть V и U – векторные пространства над некоторым полем F и $\varphi\colon V \to U$ – линейное отображение. Тогда ядром φ называется следующее множество

$$\ker\varphi = \{v \in V \mid \varphi(v) = 0\} = \varphi^{-1}(0)$$

– прообраз нуля, а образом φ называется

$$\operatorname{Im}\varphi = \{\varphi(v) \mid v \in V\} = \varphi(V)$$

– теоретико множественный образ отображения φ.

Отметим, что ядро и образ никогда не пусты, они всегда содержат 0. Кроме того, простая проверка показывает, что оба этих множества являются подпространствами, а именно: $\ker\varphi$ – подпространство в V, а $\operatorname{Im}\varphi$ – подпространство в U.

Пример Пусть $\varphi\colon F^n \to F^m$ – линейное отображение задаваемое матрицей $A \in \mathrm{M}_{m\,n}(F)$, и пусть $A = (A_1|\ldots|A_n)$, где $A_i \in F^m$ – столбцы матрицы A. Тогда

- $\operatorname{Im}\varphi = \langle A_1, \ldots, A_n\rangle$.
- $\ker\varphi = \{y \in F^n \mid Ay = 0\}$.

Второе получается просто по определению. Для того чтобы увидеть первое, поймем, что образ φ состоит из векторов вида $Ax = x_1A_1 + \ldots + x_nA_n$.

Утверждение 84. *Пусть V и U – векторные пространства над полем F и $\varphi\colon V \to U$ – линейное отображение. Тогда*

1. *φ сюръективно тогда и только тогда, когда $\operatorname{Im}\varphi = U$.*
2. *φ инъективно тогда и только тогда, когда $\ker\varphi = 0$.*
3. *$\dim_F \ker\varphi + \dim_F \operatorname{Im}\varphi = \dim_F V$.*

Доказательство. (1) Это просто переформулировка сюръективности на другом языке.

(2) Так как $\ker\varphi = \varphi^{-1}(0)$ и прообраз всегда содержит 0, то из инъективности вытекает, что $\ker\varphi = 0$. Наоборот, пусть $\varphi(v) = \varphi(v')$, тогда $\varphi(v) - \varphi(v') = 0$. А значит, $\varphi(v - v') = 0$. То есть $v - v'$ лежит в ядре, а значит равен 0, что и требовалось.

(3) Это самое интересное. Выберем базис ядра $\ker\varphi$, пусть это будет $e_1, \ldots, e_k$. Тогда дополним его до базиса всего пространства V с помощью векторов $e_{k+1}, \ldots, e_n$. Тогда образ φ является линейной оболочкой векторов $\varphi(e_1), \ldots, \varphi(e_n)$. Так как первые k из них лежат в ядре и идут в ноль, то $\operatorname{Im}\varphi = \langle\varphi(e_{k+1}), \ldots, \varphi(e_n)\rangle$. Потому если мы покажем, что $\varphi(e_{k+1}), \ldots, \varphi(e_n)$ является базисом $\operatorname{Im}\varphi$, то мы получим требуемое.

Для доказательства последнего достаточно показать линейную независимость $\varphi(e_{k+1}), \ldots, \varphi(e_n)$. Рассмотрим их какую-нибудь линейную комбинацию равную нулю

$$\lambda_{k+1}\varphi(e_{k+1}) + \ldots + \lambda_n\varphi(e_n) = \varphi(\lambda_{k+1}e_{k+1} + \ldots + \lambda_n e_n) = 0$$

Значит $\lambda_{k+1}e_{k+1} + \ldots + \lambda_n e_n \in \ker\varphi$. А все что лежит в ядре раскладывается по базису ядра $e_1, \ldots, e_k$, значит найдется выражение вида

$$\lambda_{k+1}e_{k+1} + \ldots + \lambda_n e_n = \lambda_1 e_1 + \ldots + \lambda_k e_k$$

Перенеся все в одну сторону, мы получим линейную комбинацию на $e_1, \ldots, e_n$, который по построению базис, значит, все λ_i равны нулю. Откуда следует требуемое. □

7.6 Оценки ранга суммы и произведения

Давайте начнем с простой оценки, которая не требует серьезных знаний.

Утверждение 85. *Пусть $A, B \in \mathrm{M}_{m\,n}(F)$. Тогда*

$$|\operatorname{rk} A - \operatorname{rk} B| \leqslant \operatorname{rk}(A + B) \leqslant \operatorname{rk} A + \operatorname{rk} B$$

Доказательство. Докажем сначала верхнюю оценку. Пусть $\operatorname{rk} A = r$ и $\operatorname{rk} B = s$, тогда по определению тензорного ранга существуют разложения

$$A = x_1y_1^t + \ldots + x_ry_r^t \quad \text{и} \quad B = u_1v_1^t + \ldots + u_sv_s^t, \quad x_i, u_i \in F^m,\ y_i, v_i \in F^n$$

Тогда

$$A + B = x_1y_1^t + \ldots + x_ry_r^t + u_1v_1^t + \ldots + u_sv_s^t$$

То есть мы получили какое-то разложение матрицы $A + B$ в сумму матриц ранга 1 из $r + s$ слагаемых. Но по определению тензорный ранг – это длина самого короткого разложения, значит $\operatorname{rk}(A+B) \leqslant r+s = \operatorname{rk} A+\operatorname{rk} B$.

А теперь выведем нижнюю оценку из верхней. Давайте для определенности считать, что $\operatorname{rk} A \geqslant \operatorname{rk} B$. Тогда нам надо доказать, что $\operatorname{rk} A - \operatorname{rk} B \leqslant \operatorname{rk}(A + B)$ или что то же самое, что $\operatorname{rk} A \leqslant \operatorname{rk}(A + B) + \operatorname{rk} B$. Но в этом случае

$$\operatorname{rk}(A) = \operatorname{rk}(A + B + (-B)) \leqslant \operatorname{rk}(A + B) + \operatorname{rk}(-B) = \operatorname{rk}(A + B) + \operatorname{rk}(B)$$

Здесь мы воспользовались верхней оценкой. □

А теперь давайте продемонстрируем, как можно применить векторные пространства и линейные отображения для доказательства более хитрых неравенств на ранги матриц.

Утверждение 86. *Пусть $A \in \mathrm{M}_{m\,k}(F)$ и $B \in \mathrm{M}_{k\,n}(F)$. Тогда*

$$\operatorname{rk} A + \operatorname{rk} B - k \leqslant \operatorname{rk}(AB) \leqslant \min(\operatorname{rk} A, \operatorname{rk} B)$$

Доказательство. Правая оценка – не самый большой сюрприз. Заметим, что столбцы AB есть линейная комбинация столбцов A, потому ранг AB не превосходит ранга A. С другой стороны, строки AB есть линейная комбинация строк B, потому ранг AB не превосходит ранга B.

Теперь перейдем к интересной части доказательства. Давайте заменим матрицы на линейные отображения следующим образом. Рассмотрим последовательность

$$F^n \xrightarrow{B} F^k \xrightarrow{A} F^m$$

Здесь линейные отображения я буду обозначать теми же самыми буквами, что и матрицы. Это окажется удобным и не приведет к путанице. В этом случае доказываемое неравенство превращается в такое:

$$\dim_F \operatorname{Im} A + \dim_F \operatorname{Im} B - \dim_F F^k \leqslant \dim_F \operatorname{Im} AB$$

Заметим, что $\operatorname{Im} AB = A(\operatorname{Im} B)$. Потому мы можем ограничить A со всего пространства F^k только на кусочек $\operatorname{Im} B$, то есть рассмотрим отображение

$$\operatorname{Im} B \xrightarrow{A|_{\operatorname{Im} B}} F^m$$

которое каждый вектор u переводит в Au, но только мы рассматриваем только те u, что лежат в $\operatorname{Im} B$.[92] Мы выбрали $A|_{\operatorname{Im} B}$ так, что выполнено равенство $\operatorname{Im} A|_{\operatorname{Im} B} = \operatorname{Im} AB$. Теперь применим утверждение 84 пункт (3) на связь размерности ядра и образа к $A|_{\operatorname{Im} B}$, получим

$$\dim_F \operatorname{Im} AB = \dim_F \operatorname{Im} A|_{\operatorname{Im} B} = \dim_F \operatorname{Im} B - \dim_F \ker A|_{\operatorname{Im} B}$$

Теперь наша задача оценить $\ker A|_{\operatorname{Im} B}$. По определению это векторы из $\operatorname{Im} B$, которые под действием A идут в ноль. То есть это $\operatorname{Im} B \cap \ker A$ по определению. В частности $\ker A|_{\operatorname{Im} B} \subseteq \ker A$, а значит можно продолжить равенство выше

$$\dim_F \operatorname{Im} B - \dim_F \ker A|_{\operatorname{Im} B} \geqslant \dim_F \operatorname{Im} B - \dim_F \ker A = \dim_F \operatorname{Im} B - (\dim_F F^k - \dim_F \operatorname{Im} A)$$

в последнем равенстве мы воспользовались утверждением 84 пункт (3) еще раз для оператора $A\colon F^k \to F^m$. Объединяя полученные равенства и оценку, мы приходим к требуемому результату. □

Доказательство оценки в этом случае получается технически несложным. Не надо рассматривать дурацкие линейные комбинации строк или столбцов от произведения матриц, которые не пойми как выражаются всякими непотребными формулами из исходных столбцов и строк матриц A и B. Это бонус абстрактного языка и правильного использования нужных объектов. Расплатой за это является идейная сложность. Тут надо сообразить и осознать, что мы вообще сделали, но как только вы с этим справитесь, то никаких проблем с доказательством у вас не будет.

7.7 Классификация для линейных отображений

Напомню, что если $\varphi\colon V \to U$ – линейное отображение между векторными пространствами над некоторым полем F. То после выбора базиса e в V и базиса f в U линейное отображение φ превращается в матрицу $A \in \mathrm{M}_{m\,n}(F)$, где $n = \dim_F V$ и $m = \dim_F U$. Если же мы выберем другие базисы e' и f' в пространствах V и U, соответственно, то φ превратится в матрицу A'. Если $e' = eC$ и $f' = fD$, где $C \in \mathrm{M}_n(F)$ и $D \in \mathrm{M}_m(F)$ – матрицы перехода к новым базисам, то $A' = D^{-1}AC$.

Утверждение 87. *Пусть V и U – векторные пространства над полем F размерностей n и m, соответственно, и пусть нам даны матрицы $A, B \in \mathrm{M}_{m\,n}(F)$. Тогда следующие условия эквивалентны:*

1. *Существует линейное отображение $\varphi\colon V \to U$ и базисы e и e' в V, f и f' в U такие, что A будет матрицей φ в базисах e и f, а B будет матрицей φ в базисах e' и f'.*
2. $\operatorname{rk} A = \operatorname{rk} B$.

[92] Такое линейное отображение $A|_{\operatorname{Im} B}$ называется ограничением A на $\operatorname{Im} B$.

Доказательство. (1)⇒(2). Здесь есть два доказательства: идейное и техническое. Я приведу оба. Давайте начнем с технического. Оно проще в понимании.

Техническое доказательство. Если такой φ и базисы существуют, то $B = D^{-1}AC$ для некоторых невырожденных матриц C и D подходящего размера. Тогда мы знаем по утверждению 72, что $\operatorname{rk} A = \operatorname{rk} B$, так как ранг не меняется при умножении на обратимую матрицу слева и справа.

Идейное доказательство. Если зафиксировать базисы e и f в пространствах V и U соответственно, то $\varphi\colon V \to U$ превращается в $A\colon F^n \to F^m$. Тогда образ φ совпадает с линейной оболочкой столбцов матрицы A. А значит $\operatorname{rk} A = \dim_F \operatorname{Im} \varphi$. Аналогично, $\operatorname{rk} B = \dim_F \operatorname{Im} \varphi$.

(2)⇒(1). Нам дано, что у матриц A и B равны ранги, а нам надо построить линейное отображение $\varphi\colon V \to U$ и еще пары базисов, чтобы в них матрицы φ совпали с A и B.

Так как $\operatorname{rk} A = \operatorname{rk} B$, мы можем найти обратимую матрицу $D \in \mathrm{M}_m(F)$ и обратимую матрицу $C \in \mathrm{M}_n(F)$ такие, что $B = D^{-1}AC$. Действительно, мы можем преобразованиями строк и столбцов матрицу A привести к виду $R = \left(\begin{smallmatrix} E & 0 \\ 0 & 0 \end{smallmatrix}\right)$, где E имеет размер $\operatorname{rk} A$. То есть $A = D_1 R C_1$. Аналогично, $B = D_2 R C_2$. Выразим R из первого равенства и подставим во второе. Получим требуемое.

Теперь выберем произвольный базис e в V и произвольный базис f в U. Чтобы задать линейное отображение из V в U нам надо отправить каждый базисный вектор из e куда-то в U (утверждение 79). Сделаем это так: $\varphi e = fA$. Тогда мы задали линейное отображение $\varphi\colon V \to U$ такое, что в базисах e и f он имеет матрицу A.

Далее положим $e' = eC$ и $f' = fD$. По утверждению 63 о классификации базисов, из обратимости C и D следует, что e' и f' – тоже базисы. Тогда оператор φ в этих базисах будет иметь матрицу $D^{-1}AC$, которая равна B по построению. Мы сделали все, что требовалось. □

8 Операции над подпространствами

До этого мы с вами работали с векторными пространствами «в терминах элементов», то есть основные модификации производились на языке векторов и вектора являлись основным объектом, к которому применялись операции. Настало время все изменить в своей жизни и сделать шаг вперед в прекрасный новый мир абстракций. Теперь мы хотим, чтобы основным объектом для нас было векторное пространство, потому мы хотим определить операции над самими векторными пространствами.

8.1 Сумма и пересечение

Пусть V – некоторое векторное пространство над полем F и пусть $U, W \subseteq V$ – некоторые его подпространства.

Пересечение Так как U и W являются подмножествами в V, то для них определено теоретико множественное пересечение $U \cap W$. Легкая проверка показывает, что такое пересечение обязательно является подпространством (оно в частности никогда не пусто, потому что 0 обязательно лежит в их пересечении).

Чтобы лучше себе представить пересечение, представьте себе трехмерное пространство $V = \mathbb{R}^3$ и в нем две различные плоскости $U, W \subseteq V$ проходящие через ноль. Тогда эти плоскости обязательно пересекаются по прямой $U \cap W$.

Пересечение подпространств U и W обладает следующим свойством: это наибольшее подпространство, которое одновременно лежит и в U и в W. Потому про него надо думать, как про НОД векторных подпространств.

Сумма Если мы возьмем объединение двух подпространств U и W, то это хозяйство уже не обязательно будет подпространством. Простейший пример $V = \mathbb{R}^2$ – плоскость, $U = \langle e_1 \rangle$ и $W = \langle e_2 \rangle$ – координатные прямые. Тогда объединение $U \cup W$ – это крест, состоящий из двух прямых. Но это не векторное подпространство, так как $e_1 + e_2$ там не лежит.

Потому мы определяем сумму подпространств $U + W = \{u + w \mid u \in U,\, w \in W\}$. То есть мы берем по вектору из каждого подпространства и рассматриваем все их возможные суммы. Легкая проверка показывает, что $U + W$ обязательно является подпространством в V.

Чтобы лучше себе представить сумму, давайте рассмотрим трехмерное пространство $V = \mathbb{R}^3$, а в качестве U и W – две произвольные прямые в нем проходящие через ноль. Тогда сумма U и W – это плоскость натянутая на эти две прямые.

Сумма подпространств U и W обладает следующим свойством: это наименьшее подпространство, которое одновременно содержит и U и W. Потому про него надо думать, как про НОК векторных подпространств.

Утверждение 88. *Пусть V – векторное пространство над полем F и $U, W \subseteq V$ – некоторые подпространства, тогда*

$$\dim_F(U + W) = \dim_F U + \dim_F W - \dim_F(U \cap W)$$

Доказательство. Пусть $e = (e_1, \ldots, e_r)$ – базис пересечения $U \cap W$. Теперь вспомним, что это линейно независимое подмножество во всем U и дополним его там до базиса U с помощью векторов $f = (f_1, \ldots, f_p)$. Аналогично, базис $U \cap W$ будет линейно независимым подмножеством в W, а значит его можно дополнить до базиса W векторами $h = (h_1, \ldots, h_t)$. Давайте покажем, что множество $e_1, \ldots, e_r, f_1, \ldots, f_p, h_1, \ldots, h_t$ является базисом подпространства $U + W$. Тогда из этого и будет следовать необходимое утверждение.

Любой вектор из $U + W$ имеет вид $u + w$. Тогда u раскладывается по e и f, а вектор w раскладывается через e и h. А значит система $e \cup f \cup h$ порождает $U + W$. Теперь надо показать, что она линейно независима. Рассмотрим произвольную линейную комбинацию

$$ex + fy + hz = 0, \text{ где } x \in F^r,\, y \in F^p,\, z \in F^t$$

Тогда

$$U \ni ex + fy = -hz \in W$$

а значит лежит в пересечении $U \cap W$. В частности $hz \in U \cap W$. Но все что попадает в пересечение раскладывается по e. Значит найдется какое-то $z' \in F^r$ такое, что $hz = ez'$. Но это означает, что линейная комбинация $h \cup e$ равна нулю, так как они образовывали базис W, то $z = 0$ и $z' = 0$. А значит $-hz = 0$. Что влечет $ex + fy = 0$. Так как $e \cup f$ образовывали базис U, последнее означает, что $x = 0$ и $y = 0$. То есть все коэффициенты линейной комбинации на $e \cup f \cup h$ равны нулю, что и требовалось. □

Бесконечные суммы и пересечения Если V – векторное пространство над полем F и в нем дано семейство подпространств $U_\alpha \subseteq V$, где $\alpha \in X$, то можно определить пересечение и сумму этого семейства. С пересечением все просто, это обычное теоретико-множественное пересечение $\bigcap_\alpha U_\alpha$ и как можно заметить, это подмножество является подпространством. Сумма же определяется следующим образом

$$\sum_{\alpha\in X} U_\alpha = \{u_{\alpha_1} + \ldots + u_{\alpha_k} \mid k \in \mathbb{Z}_{\geqslant 0},\ u_{\alpha_i} \in U_{\alpha_i}\}$$

То есть мы берем все возможные конечные суммы элементов из пространств U_α и складываем в один мешок. Как мы можем заметить, сумма таких выражений снова выражение такого вида и после умножения такого выражения на скаляр опять остается выражение такого вида. Значит, только что определенное подмножество является подпространством. В случае конечного числа слагаемых получаем следующее определение

$$U_1 + \ldots + U_k = \{u_1 + \ldots + u_k \mid u_i \in U_i\}$$

Определенные выше пересечение и сумма семейства U_α обладают следующим свойством:

1. Пересечение $\bigcap_\alpha U_\alpha$ является наибольшим подпространством, содержащимся во всех U_α. Здесь под наибольшим имеется в виду самое большое относительно порядка включения подпространств, то есть такое подпространство, что любое подпространство содержащееся во всех U_α в нем содержится. В частности пересечение можно определить, как объединение (или сумму) всех подпространств лежащих во всех U_α.

2. Сумма $\sum_\alpha U_\alpha$ является наименьшим подпространством, содержащим все U_α. Здесь под наименьшим имеется в виду самое маленькое относительно порядка включения подпространств, то есть такое подпространство, которое содержится в любом содержащем все U_α. В частности сумму можно определить как пересечение всех подпространств, содержащих все U_α.

3. Кроме того, отметим, что $\sum_\alpha U_\alpha$ совпадает с линейной оболочкой $\langle \bigcup_\alpha U_\alpha \rangle$. В частности для двух подпространств $U, W \subseteq V$ верно $U + W = \langle U, W\rangle = \langle U \cup W\rangle$. Таким образом к сумме можно относиться как к более удобному синтаксису задания линейных оболочек.

8.2 Прямые суммы

Если V – векторное пространство над полем F и $U_1, \ldots, U_k \subseteq V$ – его подпространства такие, что $U_1 + \ldots + U_k = V$. Тогда ясно, что можно не экономить на U_i и, заменив каждое из них на большее подпространство, равенство сохранится. Потому интересно в таких суммах выбирать U_i как можно меньше и быть экономными. Вопрос о том на сколько экономно и как можно и нужно выбирать U_i в таких ситуациях, мы и обсудим в этом разделе. Но в начале новая операция над пространствами.

Определение 89. Пусть V – векторное пространство и $U_1, \ldots, U_k \subseteq V$ – его некоторые подпространства. В этом случае $U_1, \ldots, U_k$ называются линейно независимыми, если для любых элементов $u_1 \in U_1, \ldots, u_k \in U_k$ условие $u_1 + \ldots + u_k = 0$ влечет $u_i = 0$ для всех i.

Определение 90. Если нам даны произвольные векторные пространства $V_1, \ldots, V_k$, то их декартово произведение $V_1 \times \ldots \times V_k$ обладает структурой векторного пространства. Действительно, его элементы – это наборы векторов $(v_1, \ldots, v_k)$. Тогда относительно поэлементных операций – это будет векторное пространство, то есть $(v_1, \ldots, v_k) + (u_1, \ldots, u_k) = (v_1 + u_1, \ldots, v_k + u_k)$ и $\lambda(v_1, \ldots, v_k) = (\lambda v_1, \ldots, \lambda v_k)$. Пространство $V_1 \times \ldots \times V_k$ называется внешней прямой суммой пространств $V_1, \ldots, V_k$.

Предположим, что у нас есть некоторое векторное пространство V и его подпространства $U_1, \ldots, U_k \subseteq V$. Тогда можно определить отображение $U_1 \times \ldots \times U_k \to V$ по правилу $(u_1, \ldots, u_k) \mapsto u_1 + \ldots + u_k$. Заметим, что это отображение будет линейным и образом этого отображения будет сумма подпространств $U_1 + \ldots + U_k$.

Утверждение 91. *Пусть V – векторное пространство над полем F и $U_1, \ldots, U_k \subseteq V$ – его подпространства такие, что $V = U_1 + \ldots + U_k$. Тогда следующие условия эквивалентны*

1. *$U_1, \ldots, U_k$ – линейно независимые подпространства.*

2. *Любой вектор $v \in V$ единственным образом представляется в виде суммы $v = u_1 + \ldots + u_k$, где $u_i \in U_i$.*

3. *Если e_i – базис U_i, то $\bigcup_{i=1}^k e_i$ – базис V.*

4. $\dim_F V = \sum_{i=1}^{k} \dim_F U_i$.

5. $U_i \cap (\sum_{j \neq i} U_j) = 0$ *для любого* i.

6. *Отображение* $U_1 \times \ldots \times U_k \to V$ *по правилу* $(u_1, \ldots, u_k) \mapsto u_1 + \ldots + u_k$ *является изоморфизмом.*

Доказательство. (1)⇔(2). (⇐) Если любой вектор единственным образом представляется в виде суммы u_i, то в частности это верно для 0, но он уже представляется в виде $0 = 0 + \ldots + 0$, а значит никаких других представлений у него нет. (⇒) Если $v = v_1 + \ldots + v_k$ и $v = u_1 + \ldots + u_k$, то, вычтя из одного другое, получим $0 = (v_1 - u_1) + \ldots + (v_k - u_k)$. А значит $u_i = v_i$, что и требовалось.

(1)⇔(3). (⇒) Достаточно показать, что $\bigcup_i e_i$ является линейно независимым. Пусть $\sum_i e_i x_i = 0$ – некоторая линейная комбинация, где $x_i \in F^{\dim_F U_i}$. Но тогда по (1) все $e_i x_i = 0$. А так как e_i линейно независимо, то $x_i = 0$, что и требовалось. (⇐) Пусть $u_1 + \ldots + u_k = 0$. Разложим каждый из них по базису $u_i = e_i x_i$. Тогда $\sum_i e_i x_i = 0$. Так как объединение всех e_i – базис, то $x_i = 0$, то есть $u_i = e_i x_i = 0$, что и требовалось.

(3)⇔(4). (⇒) Выполнено по определению, так как размерность – это количество векторов в любом базисе. (⇐) Если e_i – это базис U_i, то система $\bigcup_{i=1}^{k} e_i$ порождает V. Чтобы проверить, что это базис, достаточно показать, что в ней $\dim_F V$ элементов. Но так как эта система порождающая, то достаточно проверить, что в ней не больше $\dim_F V$ элементов. Теперь прямое вычисление показывает, что

$$|\bigcup_{i=1}^{k} e_i| \leqslant \sum_{i=1}^{k} |e_i| = \sum_{i=1}^{k} \dim_F U_i = \dim_F V$$

(1)⇔(5). (⇒) Пусть $u \in U_i \cap (\sum_{j \neq i} U_j)$, тогда $u = u_i$ и $u = \sum_{j \neq i} u_j$, а значит $u_i = \sum_{j \neq i} u_j$. Перенесем все в одну сторону, получим $\sum_{j \neq i} u_j - u_i = 0$. А значит все $u_i = 0$ по (1). А значит $u = u_i = 0$, что и требовалось. (⇐) Пусть $u_1 + \ldots + u_k = 0$ и какой-нибудь $u_i \neq 0$. Тогда $u_i = -\sum_{j \neq i} u_j$. Мы получили ненулевой вектор в пересечении $U_i \cap (\sum_{j \neq i} U_j)$, противоречие.

(2)⇔(6). Это переформулировка одного и того же свойства.

□

Определение 92. Пусть V – векторное пространство над полем F и $U_1, \ldots, U_k \subseteq V$ – его подпространства такие, что $V = U_1 + \ldots + U_k$ обладающие одним из эквивалентных свойств из предыдущего утверждения. Тогда сумма $V = U_1 + \ldots + U_k$ называется прямой и обозначается $V = U_1 \oplus \ldots \oplus U_k$. Тогда говорят, что V является внутренней прямой суммой подпространств $U_1, \ldots, U_k$.

Прямая сумма – это обычная сумма подпространств, только с условием, что эти подпространства удовлетворяют некоторому дополнительному свойству. Кроме того, свойство (6) из предыдущего утверждения означает, что прямая сумма векторных подпространств совпадает с внешней прямой суммой этих же подпространств рассматриваемых как абстрактные векторные пространства. Бонус от этого вот какой: с внешней прямой суммой работать очень просто, так как это наборы векторов, их легко складывать, умножать на числа и сравнивать друг с другом. А так как внутренняя прямая сумма от этой не отличается (с точностью до изоморфизма), то изучать одно это все равно, что изучать другое.

Примеры

1. Пусть V – векторное пространство с базисом $e_1, \ldots, e_n$. Положим $U_i = \langle e_i \rangle$, тогда $V = U_1 \oplus \ldots \oplus U_n$. То есть мы раскладываем все пространство в прямую сумму прямых натянутых на базисные векторы.

2. Условие (5) очень просто переформулируется в случае двух подпространств. Пусть $U, W \subseteq V$, тогда $V = U \oplus W$ тогда и только тогда, когда $V = U + W$ и $U \cap W = 0$.

3. Давайте посмотрим на предыдущий пример в конкретном случае. Пусть $V = \mathbb{R}^3$, U – некоторая плоскость проходящая через 0, а W – прямая проходящая через 0 и не содержащаяся в U. Тогда пересечение прямой и плоскости есть ноль, а наименьшее подпространство, которое их содержит – это все пространство V. Значит $V = U \oplus W$.

4. Условие (5) сильнее условия $U_i \cap U_j = 0$. Вот пример. Пусть $V = \mathbb{R}^2$, $U_1 = \langle e_1 \rangle$, $U_2 = \langle e_2 \rangle$, $U_3 = \langle e_1 + e_2 \rangle$. То есть я беру две координатные прямые на плоскости и прямую под углом 45 градусов через первый квадрант. Тогда все попарные пересечения прямых есть ноль. Но $U_i + U_j = V$ для любой пары прямых, потому условие (5) не выполнено.

9 Линейные операторы

В этом разделе я наконец-то вам начну рассказывать о самых важных объектах в линейной алгебре – линейных операторах.

9.1 Определение и базовые свойства

Определение 93. Пусть V – векторное пространство над полем F, тогда линейным оператором на V называется линейное отображение $\varphi\colon V \to V$.

Так как линейный оператор – это частный случай линейного отображения, то для него применимо все, о чем мы уже говорили в случае отображений. Про линейный оператор надо думать как про деформацию пространства V.

Примеры

1. $\mathrm{Id}\colon V \to V$, $v \mapsto v$. Тождественный линейный оператор, ничего не деформирует.
2. $0\colon V \to V$, $v \mapsto 0$. Нулевой линейный оператор, который все отправляет в ноль.
3. $A\colon \mathbb{R}^3 \to \mathbb{R}^3$, $x \mapsto Ax$, где $A \in \mathrm{M}_3(\mathbb{R})$ задана так
$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$$
– поворот на угол α вокруг оси $\langle e_1 \rangle$.
4. Пусть $V = U \oplus W$, тогда зададим $\pi\colon V \to V$ по правилу $v = u + w \mapsto u$. Так как разложение в прямой сумме однозначно, то это корректно задает линейный оператор, который называется проектором на U вдоль W. Обратите внимание, что $\ker\pi = W$ и $\operatorname{Im}\pi = U$. При этом для любого $u \in \operatorname{Im}\pi$ верно $\pi u = u$.

Утверждение 94. *Пусть V – векторное пространство над полем F и $\pi\colon V \to V$ – линейный оператор. Тогда следующие свойства эквивалентны:*

1. *Существуют подпространства $U, W \subseteq V$ такие, что $V = U \oplus W$ и π является проектором на U вдоль W.*
2. *$\pi^2 = \pi$.*

Доказательство. (1)⇒(2). Рассмотрим произвольный $v \in V$, тогда $\pi^2(v) = \pi(\pi(v))$. Но вектор $\pi(v)$ лежит в образе π, то есть в U. Как я уже отмечал в замечании выше, на векторах из образа проектор π действует тождественно, то есть $\pi(\pi(v)) = \pi(v)$, что и требовалось.

(2)⇒(1). Пусть $\pi^2 = \pi$. Для начала нам надо откуда-то взять подпространства U и W. Замечание выше подсказывает, что надо положить $U = \operatorname{Im}\pi$ и $W = \ker\pi$. Теперь надо показать две вещи: (1) V раскладывается в прямую сумму U и W, (2) действие π совпадает с действием проектора на U вдоль W.

Для (1) нам надо показать, что $U \cap W = 0$ и $U + W = V$. Начнем с пересечения. Пусть $v \in U \cap W$ – произвольный вектор. Тогда с одной стороны $v \in U = \operatorname{Im}\pi$, а значит $v = \pi(v')$ и $v' \in V$. С другой стороны, $v \in W = \ker\pi$, а значит $\pi(v) = 0$. Но тогда
$$0 = \pi(v) = \pi(\pi(v')) = \pi^2(v') = \pi(v') = v$$
Значит в пересечении лежит только нулевой вектор.

Теперь займемся суммой. Мы должны показать, что любой вектор из V представляется в виде суммы векторов из U и W. Пусть $v \in V$, рассмотрим следующее разложение
$$v = \pi(v) + (\mathrm{Id} - \pi)(v) = \pi(v) + (v - \pi(v))$$
Первый вектор $\pi(v)$ по определению попадает в $\operatorname{Im}\pi = U$. Проверим, что второй лежит в ядре:
$$\pi((\mathrm{Id} - \pi)(v)) = \pi(v - \pi(v)) = \pi(v) - \pi^2(v) = 0$$
Значит $V = U + W$.

Теперь мы знаем, что $V = U \oplus W = \operatorname{Im}\pi \oplus \ker\pi$. Давайте покажем, что π действует как проектор. Возьмем $v \in V$, тогда он представляется в виде $v = u + w$, где $u = \pi(v)$ и $w = v - \pi(v)$. Применим π к v и видим, что получаем u. По определению действие π совпадает с действием проектора на U вдоль W. □

Замечание

- Таким образом, если мы хотим разложить какое-то пространство V в прямую сумму подпространств, нам достаточно найти оператор на V, который в квадрате равен самому себе.
- Обратите внимание, что Id является по определению проектором на все пространство вдоль нулевого подпространства, а 0 является проектором на нулевое подпространство вдоль всего пространства. Эти операторы дают тривиальное разложение пространства V в прямую сумму $0 \oplus V$. Эти случаи надо иметь в виду.

9.2 Матрица линейного оператора

Пусть в векторном пространстве V задан некоторый базис $e_1, \ldots, e_n$ и пусть $\varphi\colon V \to V$ – линейный оператор. Так как у оператора пространство из которого он бьет и то в которое он бьет совпадают, то мы фиксируем всего лишь один базис (пространство-то у нас одно). Тогда по определению матрица линейного оператора φ – это такая матрица $A_\varphi \in \mathrm{M}_n(F)$, что выполнено $\varphi e = eA_\varphi$, где $e = (e_1, \ldots, e_n)$.

Пусть теперь у нас задан другой базис $e'_1, \ldots, e'_n$ в пространстве V с матрицей перехода $C \in \mathrm{M}_n(F)$, то есть $(e'_1, \ldots, e'_n) = (e_1, \ldots, e_n)C$. Пусть так же $e' = (e'_1, \ldots, e'_n)$. Тогда матрица φ в базисе e' пусть будет A'_φ, то есть $\varphi e' = e'A'_\varphi$. В этом случае связь между матрицами следующая $A'_\varphi = C^{-1}A_\varphi C$. То есть матрица A_φ сопряжена матрице A'_φ.[93]

Замечания

- Отметим, что матрица линейного оператора обязательно квадратная. Таким образом, изучение линейного отображения – это изучение прямоугольной матрицы, а изучение линейного оператора – это всегда изучение только квадратной матрицы.
- Если линейное отображение $\psi\colon V \to U$ бьет между двумя разными пространствами одинаковой размерности, то ему тоже соответствует квадратная матрица. Но принципиальная разница с линейным оператором заключается в том, что для линейного отображения мы можем независимо менять базисы в V и U, что соответствует замене $A'_\psi = C^{-1}A_\psi D$, а для линейного оператора, так как пространство одно и то же, базисы меняются одновременно, что соответствует $A'_\varphi = C^{-1}A_\varphi C$.
- Так как линейные операторы – это линейные отображения, то задавать их можно так же как и линейные отображения, например: либо с помощью образа базисных векторов, либо с помощью матрицы в фиксированном базисе.

9.3 Характеристики линейных операторов

В этом разделе я перечислю основные характеристики, которые можно определить для любого линейного оператора.

След Перед определением докажем техническое утверждение.

Утверждение. *Пусть V – векторное пространство над полем F и пусть $\varphi\colon V \to V$ – некоторый линейный оператор. Тогда число $\operatorname{tr}(A_\varphi)$ не зависит от базиса, в котором посчитана матрица A_φ.*

Доказательство. Действительно, пусть у нас есть два базиса $e' = (e'_1, \ldots, e'_n)$ и $e = (e_1, \ldots, e_n)$ связанные матрицей перехода $e' = eC$. Пусть $\varphi e = eA_\varphi$ и $\varphi e' = e'A'_\varphi$. Тогда как мы видели выше $A'_\varphi = C^{-1}A_\varphi C$. Тогда

$$\operatorname{tr}(A'_\varphi) = \operatorname{tr}(C^{-1}A_\varphi C) = \operatorname{tr}(A_\varphi CC^{-1}) = \operatorname{tr}(A_\varphi)$$

□

Положим по определению $\operatorname{tr}\varphi = \operatorname{tr} A_\varphi$ и будем называть это число следом оператора φ. Это определение корректно, так как данное число не зависит от базиса, в котором считается матрица оператора.[94]

[93] Напомню, что квадратные матрицы B и D называются сопряженными, если найдется обратимая матрица C такая, что $D = C^{-1}BC$.

[94] Тут нужно сделать важное замечание. Как мы видим след оператора определяется через его матрицу, но не зависит от матрицы, а зависит только от самого линейного оператора. Потому есть соблазн дать эквивалентное определение совсем не используя матрицу оператора. К сожалению так сделать невозможно. Одной из причин является отсутствие следа в бесконечно мерных векторных пространствах. Любые попытки дать «без координатное» определение следа на самом деле является лишь тщательной маскировкой его координатной природы.

Определитель Перед определением докажем техническое утверждение.

Утверждение. *Пусть V – векторное пространство над полем F и пусть $\varphi\colon V \to V$ – некоторый линейный оператор. Тогда число $\det(A_\varphi)$ не зависит от базиса, в котором посчитана матрица A_φ.*

Доказательство. Действительно, пусть у нас есть два базиса $e' = (e'_1, \ldots, e'_n)$ и $e = (e_1, \ldots, e_n)$ связанные матрицей перехода $e' = eC$. Пусть $\varphi e = eA_\varphi$ и $\varphi e' = e'A'_\varphi$. Тогда как мы видели выше $A'_\varphi = C^{-1}A_\varphi C$. Тогда

$$\det(A'_\varphi) = \det(C^{-1}A_\varphi C) = \det(C^{-1})\det(A_\varphi)\det(C) = \det(A_\varphi)$$

□

Положим по определению $\det\varphi = \det A_\varphi$ и будем называть это число определителем оператора φ. Это определение корректно, так как данное число не зависит от базиса, в котором считается матрица оператора.[95]

Характеристический многочлен Пусть опять $\varphi\colon V \to V$ – произвольный линейный оператор, тогда для любого $\lambda \in F$, $\lambda\,\mathrm{Id} - \varphi\colon V \to V$ – тоже линейный оператор. Тогда по предыдущему определению корректно определен определитель такого оператора, который мы обозначим так: $\chi_\varphi(\lambda) = \det(\lambda\,\mathrm{Id} - \varphi)$ и будем называть характеристическим многочленом оператора φ.

Пусть теперь в некотором базисе φ имеет матрицу A_φ. Тождественный оператор Id в любом базисе задается единичной матрицей. Тогда по предыдущему определению $\det(\lambda\,\mathrm{Id} - \varphi)$ совпадает с $\det(\lambda E - A_\varphi)$. То есть характеристический многочлен оператора – это характеристический многочлен любой из его матриц в каком-нибудь базисе (в силу корректности определения определителя оператора, все эти многочлены будут одинаковыми).

Есть другой способ смотреть на характеристический многочлен. Можно просто сказать, что для оператора φ его характеристический многочлен – это характеристический многочлен его матрицы A_φ и надо лишь показать, что он не зависит от базиса. Это делается следующей проверкой

$$\det(\lambda E - CA_\varphi C^{-1}) = \det(\lambda CC^{-1} - CA_\varphi C^{-1}) = \det(C(\lambda E - A_\varphi)C^{-1}) = \det(\lambda E - A_\varphi)$$

Спектр Как и выше $\varphi\colon V \to V$ – линейный оператор на векторном пространстве, тогда положим

$$\mathrm{spec}_F(\varphi) = \{\lambda \in F \mid \varphi - \lambda\,\mathrm{Id} \text{ не обратим}\}$$

И будем называть это множество спектром линейного оператора φ.[96] Пусть теперь A_φ – матрица линейного оператора в каком-нибудь базисе. Оператор обратим тогда и только тогда, когда обратима его матрица (потому что все операции над операторами превращаются в операции над матрицами). Потому условие $\varphi - \lambda\,\mathrm{Id}$ не обратим превращается в условие $A_\varphi - \lambda E$ не обратима. То есть спектр линейного оператора совпадает со спектром любой из его матриц в каком-нибудь базисе.

Минимальный многочлен Если у нас есть многочлен $f \in F[t]$ вида $f = a_0 + a_1 t + \ldots + a_n t^n$ и задан линейный оператор $\varphi\colon V \to V$, то можно определить оператор $f(\varphi)$ по правилу

$$f(\varphi) = a_0\,\mathrm{Id} + a_1\varphi + \ldots + a_n\varphi^n$$

Здесь степень φ^k – это композиция оператора φ с самим собой k раз, а сумма и умножение на коэффициенты из поля берутся поточечно.[97]

Если в результате подстановки оператора в многочлен мы получили нулевой оператор (тот который на всех векторах действует нулем), то мы говорим, что многочлен зануляет φ и пишем $f(\varphi) = 0$. Если в каком-то базисе $e_1, \ldots, e_m \in V$ оператор φ имеет матрицу $A \in \mathrm{M}_m(F)$, то $f(\varphi) = 0$ тогда и только тогда, когда $f(A) = 0$. Действительно, при переходе к базису $f(\varphi)$ имеет матрицу $f(A)$, а нулевой оператор соответствует нулевой матрице. Теперь мы можем определить минимальный многочлен оператора, как такой ненулевой многочлен $f_{\min\varphi} \in F[t]$, что

1. $f_{\min\varphi}(\varphi) = 0$.

[95] Здесь верно то же самое замечание, что и для следа. Определитель оператора нельзя определить без матрицы оператора, но в то же время он не зависит от матрицы, а зависит лишь от самого оператора.

[96] Заметим, что определение спектра дается без помощи матрицы линейного оператора.

[97] Смотри определения для линейных отображений в разделе 7.2.

2. $f_{\min \varphi}$ имеет наименьшую степень среди всех ненулевых многочленов зануляющих φ.

3. Старший коэффициент $f_{\min \varphi}$ равен единице.

В силу того, что для многочлена занулять оператор это тоже самое, что занулять его матрицу, то минимальный многочлен для линейного оператора совпадает с минимальным многочленом для его матрицы в любом базисе.

Ранг Мы знаем, что линейный оператор – это просто линейное отображение, но на одном пространстве. Для любого линейного отображения мы видели, что ранг его матрицы не меняется при смене базиса (см. утверждение 87). В частности ранг матрицы линейного оператора не меняется при смене базиса.

9.4 Обратимость оператора

Утверждение 95. *Пусть V – векторное пространство над некоторым полем F и $\varphi\colon V \to V$ – некоторый линейный оператор. Тогда следующие свойства эквивалентны:*

1. $\ker \varphi = 0$.

2. $\operatorname{Im} \varphi = V$.

3. *φ обратим.*

4. $\det \varphi \neq 0$.

Доказательство. Это утверждение является преформулировкой утверждения 3 на языке оператора. С другой стороны его можно получить из комбинации пунктов утверждения 84. □

9.5 Инвариантные подпространства

Пусть $U \subseteq V$ – подпространство в некотором векторном пространстве V над полем F и пусть $\varphi\colon V \to V$ – некоторый линейный оператор. Будем говорить, что векторное подпространство U является инвариантным относительно φ (или просто φ-инвариантным), если $\varphi(U) \subseteq U$.

Пример

- Рассмотрим пример поворота трехмерного пространства вокруг некоторой оси, а именно, пусть $A\colon \mathbb{R}^3 \to \mathbb{R}^3$, $x \mapsto Ax$, где $A \in \mathrm{M}_3(\mathbb{R})$ задана так

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$$

  В данном случае мы поворачиваем вокруг оси $\langle e_1 \rangle$. Заметим, что подпространство $\langle e_1 \rangle$ является инвариантным, любой вектор из этого подпространства остается неподвижным. Кроме того, подпространство $\langle e_2, e_3 \rangle$ – плоскость поворота, тоже является инвариантным относительно φ, любой вектор в ней поворачивается на угол α.

- Для любого оператора $\varphi\colon V \to V$ его ядро и образ являются инвариантными подпространствами.

Ограничение оператора

Определение 96. Пусть V – векторное пространство над полем F. Если $\varphi\colon V \to V$ – линейный оператор и $U \subseteq V$ – некоторое инвариантное подпространство, то тогда можно определить оператор $\varphi|_U\colon U \to U$, действующий по правилу $u \mapsto \varphi(u)$. Такой оператор называется ограничением φ на U.

Инвариантность в терминах матрицы Пусть $V = U \oplus W$ – прямая сумма подпространств. Выберем в U базис $e = (e_1, \ldots, e_n)$, а в W базис $f = (f_1, \ldots, f_m)$. Тогда $e \cup f$ является базисом V. Если $\varphi\colon V \to V$ – некоторый линейный оператор, то его можно записать в этом базисе в следующем блочном виде

$$\varphi(e, f) = (e, f) \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

Заметим, что при этом подпространство U будет φ-инвариантным тогда и только тогда, когда $C = 0$. Действительно, если U инвариантно, то $\varphi(U) \subseteq U$. С другой стороны $U = \langle e \rangle$. То есть U инвариантно тогда и только тогда, когда $\varphi(e) \subseteq U$. С другой стороны, по определению матрицы оператора $\varphi(e) = eA + fC$. Но $eA + fC$ лежит в $\langle e \rangle$ тогда и только тогда, когда $C = 0$. В этом случае определен оператор $\varphi|_U$ и матрица A будет матрицей этого оператора в базисе e.

Аналогично, подпространство W инвариантно тогда и только тогда, когда $B = 0$. Если же оба пространства инвариантны, то матрица φ является блочно диагональной. То есть отсюда мы видим геометрический смысл блочно верхнетреугольных и блочно диагональных матриц. Блочно верхнетреугольная означает наличие инвариантных подпространств натянутых на первый кусок базисных векторов. Блочно диагональный вид означает разложение пространства в прямую сумму инвариантных подпространств. Подобное разбиение в прямую сумму инвариантных позволяет сводить задачу про один оператор к задачам про оператор на пространстве меньшего размера. Это бывает полезно, если надо вести рассуждение индукцией по размерности подпространств.

Утверждение 97. *Пусть $\varphi, \psi\colon V \to V$ два коммутирующих линейных оператора. Тогда $\ker\varphi$ и $\operatorname{Im}\varphi$ являются ψ-инвариантными.*

Доказательство. Случай $\ker\varphi$. Мы должны показать, что $\psi(\ker\varphi) \subseteq \ker\varphi$. Возьмем произвольный вектор $v \in \ker\varphi$, нам надо показать, что $\psi(v) \in \ker\varphi$. То есть мы должны показать, что $\varphi(\psi(v)) = 0$. Но $\varphi\psi v = \psi\varphi v = \psi 0 = 0$.

Случай $\operatorname{Im}\varphi$. Мы должны показать, что $\psi(\operatorname{Im}\varphi) \subseteq \operatorname{Im}\varphi$. Возьмем произвольный вектор $v \in \operatorname{Im}\varphi$, нам надо показать, что $\psi(v) \in \operatorname{Im}\varphi$. Но условие $v \in \operatorname{Im}\varphi$ означает, что $v = \varphi(u)$ для некоторого $u \in V$. Но тогда $\psi v = \psi\varphi u = \varphi(\psi(u))$, что и требовалось. □

9.6 Собственные векторы и значения

Определение 98. Пусть V – некоторое векторное пространство над полем F и $\varphi\colon V \to V$ – линейный оператор. Вектор $v \in V$ называется собственным для φ, если найдется такое $\lambda \in F$, что $\varphi v = \lambda v$.

Замечания

- Вектор $v \in V$ является собственным тогда и только тогда, когда $\langle v \rangle$ является φ-инвариантным подпространством. Таким образом изучать собственные векторы – это то же самое, что изучать не более чем одномерные инвариантные подпространства.
- Вектор $0 \in V$ всегда является собственным для любого линейного оператора.

Определение 99. Пусть V – некоторое векторное пространство над полем F и $\varphi\colon V \to V$ – линейный оператор. Число $\lambda \in F$ называется собственным значением φ, если найдется ненулевой $v \in V$ такой, что $\varphi v = \lambda v$.

Замечания

- Важно отметить, что в определении требуется, чтобы $v \neq 0$. Это связано с тем, что вектор $0 \in V$ является собственным для любого λ, то есть всегда верно $\varphi 0 = \lambda 0$. И если не потребовать этого условия, то любое число удовлетворяет этому определению и в нем теряется смысл.Будь те внимательны.
- Популярная ошибка – считать, что 0 не может быть собственным значением. На самом деле, число 0 как может являться собственным значением, так и может не являться им. А именно, число 0 является собственным значением тогда и только тогда, когда $\ker\varphi \neq 0$. Потому что собственные векторы для значения 0 – это векторы $v \in V$ такие, что $\varphi(v) = 0v = 0$. Потому наличие ненулевого такого вектора означает, наличие ненулевого вектора в ядре, а это равносильно неинъективности, а значит и необратимости оператора (в силу утверждения 95).

Собственные и корневые подпространства

Определение 100. Пусть V – некоторое векторное пространство над полем F и $\varphi\colon V \to V$ – линейный оператор. Для любого числа $\lambda \in F$ определим собственное подпространство

$$V_\lambda = \{v \in V \mid \varphi v = \lambda v\}$$

Заметим, что такое подмножество обязательно является подпространством, например, потому что совпадает с $\ker(\varphi - \lambda \operatorname{Id})$. Действительно, $\varphi v = \lambda v$ тогда и только тогда, когда $\varphi v - \lambda v = 0$. Что равносильно тому, что $(\varphi - \lambda \operatorname{Id})v = 0$, что значит $v \in \ker(\varphi - \lambda \operatorname{Id})$.

Замечание Обратите внимание, что оператор φ на собственном подпространстве V_λ действует как скалярный оператор $\lambda \operatorname{Id}$, то есть все умножает на λ просто по определению $\varphi v = \lambda v$ для любого $v \in V_\lambda$. Тут Капитан Очевидность передает привет. Однако, не спешите, его помощник по имени Нетривиальное Следствие сейчас расскажет пару слов.

Давайте рассмотрим произвольный многочлен $f \in F[x]$, тогда определен оператор $f(\varphi)\colon V \to V$. Так вот, оператор $f(\varphi)$ на собственном подпространстве V_λ действует умножением на $f(\lambda)$, то есть $f(\varphi)v = f(\lambda)v$. Это очень простое наблюдение жутко полезно.

Утверждение 101. *Пусть V – некоторое векторное пространство над полем F и $\varphi\colon V \to V$ – линейный оператор. Тогда следующие условия равносильны:*

1. $V_\lambda \neq 0$.
2. $\lambda \in \operatorname{spec}_F \varphi$.
3. λ – *корень* $\chi_\varphi(t)$.
4. λ – *корень минимального многочлена для* φ.

Доказательство. Эквивалентность последних трех условий была доказана в утверждениях 11 и 34. Здесь эти условия приводятся, чтобы создать общую картину у читающего. Давайте проверим эквивалентность первого условия с оставшимися.

(1)$\Rightarrow$ Пусть $V_\lambda \neq 0$, тогда $\varphi v = \lambda v$ для некоторого ненулевого вектора. Значит $(\varphi - \lambda \operatorname{Id})v = 0$. А значит оператор $\varphi - \lambda \operatorname{Id}$ не обратим.

$\Rightarrow$(1) Пусть $\varphi - \lambda \operatorname{Id}$ не обратим. Тогда по одному из эквивалентных свойств обратимости оператора (утверждение 95), это означает, что $\varphi - \lambda \operatorname{Id}$ имеет не нулевое ядро. То есть есть ненулевой вектор $v \in V$ такой, что $(\varphi - \lambda \operatorname{Id})v = 0$. А это и значит, что $\varphi v = \lambda v$ после раскрытия скобок и переноса второго слагаемого в право. □

Определение 102. Пусть V – векторное пространство над полем F, $\varphi\colon V \to V$ – линейный оператор и λ – его собственное значение. Тогда кратность λ в характеристическом многочлене χ_φ называется кратностью собственного значения.

Почему это определение имеет смысл, вы увидите чуть позже, когда мы будем говорить про диагонализацию (утверждение 108).

Утверждение 103. *Пусть F – алгебраически замкнутое поле, V – векторное пространство над полем F и $\varphi\colon V \to V$ – линейный оператор. Тогда обязательно существует ненулевой собственный вектор $v \in V$ для некоторого $\lambda \in F$.*

Доказательство. Действительно, наличие такого вектора означает, что для некоторого $\lambda \in F$ пространство V_λ не нулевое. А это по предыдущему утверждению равносильно тому, что λ – корень характеристического многочлена для φ. Так как этот многочлен не константный (его степень равна размерности пространства),[98] а F – алгебраически замкнуто, то у нас обязательно существует корень $\lambda \in F$. А значит $V_\lambda \neq 0$ (по утверждению 101). □

Определение 104. Пусть V – некоторое векторное пространство над полем F и $\varphi\colon V \to V$ – линейный оператор. Для любого числа $\lambda \in F$ определим корневое подпространство

$$V^\lambda = \{v \in V \mid \exists n :\ (\varphi - \lambda \operatorname{Id})^n v = 0\}$$

[98] Мы скромно закроем глаза на случай $V = 0$, то есть пространство нульмерно. В этом случае большой вопрос, что считать спектром. Правильно полагать его пустым. Верность утверждения тогда зависит от аккуратности формулировки. Но не надо забивать себе этим голову, просто имейте в виду, что иногда этот случай нужен.

Замечания

- Заметим, что $V^\lambda = \bigcup_{n\geqslant 0} \ker(\varphi - \lambda \operatorname{Id})^n$. Каждое из ядер является подпространством. Однако в общем случае объединение подпространств не является подпространством. Но в данном случае $\ker(\varphi - \lambda \operatorname{Id})^k \subseteq \ker(\varphi - \lambda \operatorname{Id})^{k+1}$, то есть наши подпространства возрастают. Я оставлю в качестве упражнения убедиться, что при таком условии объединение обязательно будет подпространством.
- Кроме того, по определению $V_\lambda \subseteq V^\lambda$. При этом равенства в этом включении может не быть. Пусть

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

  Тогда $A\colon F^2 \to F^2$ – линейный оператор. При этом $A^2 = 0$. То есть $f(x) = x^2$ – зануляющий многочлен. Заметим, что он обязательно минимальный. А значит $\operatorname{spec}(A) = \{0\}$. Тогда $V_0 = \ker A$ и оно порождено вектором e_1. С другой стороны $F^2 = \ker A^2$, а потому $V^0 = F^2$.
- Обратите внимание, что $V_\lambda \neq 0$ тогда и только тогда, когда $V^\lambda \neq 0$. В одну сторону – это следует из вложения $V_\lambda \subseteq V^\lambda$. В другую сторону, если $v \in V^\lambda$ и $v \neq 0$, то найдем такое k, что $w = (\varphi - \lambda \operatorname{Id})^k v \neq 0$, а $(\varphi - \lambda \operatorname{Id})w = (\varphi - \lambda \operatorname{Id})^{k+1} v = 0$. Тогда $w \in V_\lambda$ и не нулевой.
- Подпространства V_λ и V^λ являются φ инвариантными для любого λ.

9.7 Лемма о стабилизации

Утверждение 105. *Пусть V – векторное пространство над полем F и $\varphi\colon V \to V$ – линейный оператор. Тогда*

1. *Найдется такое число $0 \leqslant k \leqslant \dim_F V$, что*

$$0 \subsetneq \ker\varphi \subsetneq \ker\varphi^2 \subsetneq \ldots \subsetneq \ker\varphi^k = \ker\varphi^{k+1} = \ldots$$

2. *Найдется такое число $0 \leqslant k \leqslant \dim_F V$, что*

$$V \supsetneq \operatorname{Im}\varphi \supsetneq \operatorname{Im}\varphi^2 \supsetneq \ldots \supsetneq \operatorname{Im}\varphi^k = \operatorname{Im}\varphi^{k+1} = \ldots$$

Давайте поясним, что утверждается. Мы говорим, что ядра оператора сначала строго растут, а начиная с какого-то момента обязательно становятся одинаковыми для всех последующих шагов. Аналогичное происходит с образами, только они сначала строго уменьшаются, а потом становятся одинаковыми. Стоит обратить внимание, что k может быть равным 0, это означает, что нет строгих включений и равенства начинаются с самого начала.

Доказательство. (1) Нам достаточно показать, что если в какой-то момент $\ker\varphi^m = \ker\varphi^{m+1}$, то $\ker\varphi^{m+1} = \ker\varphi^{m+2}$. Включение $\ker\varphi^{m+1} \subseteq \ker\varphi^{m+2}$ понятно из определения (если что-то зануляется φ^{m+1}, то оно зануляется и большей степенью φ^{m+2}). Надо показать обратное. Пусть $v \in \ker\varphi^{m+2}$, тогда $\varphi^{m+2}v = 0$. То есть $\varphi^{m+1}(\varphi v) = 0$. Это значит $\varphi v \in \ker\varphi^{m+1} = \ker\varphi^m$. Последнее означает, что $\varphi^m(\varphi v) = 0$, то есть $\varphi^{m+1}v = 0$. Значит $v \in \ker\varphi^{m+1}$, что и требовалось.

Теперь надо понять, что k не превосходит размерность V. Но это следует из того факта, что в цепочке

$$0 \subsetneq \ker\varphi \subsetneq \ker\varphi^2 \subsetneq \ldots$$

размерность подпространств каждый шаг растет хотя бы на единицу. Значит больше, чем $\dim_F V$ шагов у нас быть не может.

(2) Доказательство этого факта проходит аналогично. Либо можно воспользоваться соотношением между размерностями ядра и образа (утверждение 84 пункт (3)) и увидеть, что стабилизация у образов начинается на том же значении k, что и у ядер. □

Замечание В силу этого утверждения мы получаем, что $V^\lambda = \ker(\varphi - \lambda \operatorname{Id})^m$ для некоторого достаточно большого m. Понятно, что на самом деле, достаточно взять $m = \dim_F V$.[99]

[99] Если уж бы до конца честным, то можно еще сильнее уменьшить m. Тут достаточно взять кратность собственного значения в минимальном многочлене, я докажу это позже.

10 Классификационная задача для линейных операторов

Абстрактные объекты вроде векторных пространств и линейных операторов становятся более знакомыми после выбора базиса. Пространство превращаются в столбцы, а операторы в квадратные матрицы. Но так как базис выбирать можно по-разному, то и матрицы в такой ситуации получаются черт знает какими. Основной вопрос классификационной задачи: как понять по матрицам, что они задают один и тот же линейный оператор, но в разных базисах. Другой вопрос: к какому самому простому виду можно привести матрицу линейного оператора. Мы уже видели ответ в случае линейного отображения между разными пространствами, в этом случае все контролируется рангом матриц. Оказывается, что в случае линейного оператора ситуация сильно сложнее и зависит от выбора поля.

Пусть теперь $\varphi\colon V \to V$ – линейный оператор, т.е. линейное отображение из векторного пространства в себя. Тогда при выборе базиса e в V наш оператор превращается в матрицу $A \in \mathrm{M}_n(F)$, где $n = \dim_F V$. Если же мы выберем другой базис e' в V такой, что $e' = eC$ для некоторой обратимой $C \in \mathrm{M}_n(F)$. То матрица φ в базисе e' будет $C^{-1}AC$. Заметим сложность ситуации. Мы теперь не можем независимо домножать нашу матрицу с разных сторон на разные матрицы. Если думать в терминах элементарных преобразований, мы теперь должны неким сложным образом согласовывать преобразования строк и столбцов. Из-за этих ограничений кустарными методами (вроде подбора элементарных преобразований) для приведения матрицы в хороший вид нам обойтись не получится. Более того, степень «хорошести» нашей матрицы будет сильно зависеть от свойств поля F над которым определены наши векторные пространства. А так как элементарные преобразования ничего не знают про свойства поля, то это автоматически означает, что не мы такие неумелые, что не смогли воспользоваться элементарными преобразованиями, а этот метод в лоб просто не работает.

Для преодоления сложившихся трудностей в случае оператора приходится привлекать более продвинутую технику. К такой технике как раз и относятся собственные векторы и значения, собственные и корневые подпространства. Кульминацией для нас будет теорема о жордановой нормальной форме. Формулировать мы ее пока не будем, но обсудим некоторые стратегические соображения.

Для чего вообще меняется базис? Для того, чтобы сделать вид матрицы линейного оператора максимально простым. Тогда заменив его простой матрицей, его будет проще изучать. В идеале простой вид – это когда много нулей. Самый желанный для нас вид – диагональный. Было бы еще лучше, если бы можно было сделать матрицу диагональной с единицами и нулями на диагонали, но это совсем не возможно. Например, если оператор не вырожден, то его определитель не меняется, а он совпадает с произведением диагональных элементов матрицы. То есть скалярную матрицу λE никогда нельзя сделать единичной E путем замены базиса (это так же видно из формулы замены) если $\lambda \neq 1$.

Оказывается и диагональной можно сделать не всякую матрицу путем сопряжения, то есть не всякий линейный оператор приводится к диагональному виду в каком-то базисе. Потому один из первых вопросов, которым мы хотим заняться – это вопрос: когда линейный оператор задается диагональной матрицей в некотором базисе.

10.1 Диагонализуемость линейного оператора

Определение 106. Пусть $\varphi\colon V \to V$ – линейный оператор над некоторым полем F. Будем говорить, что φ диагонализуется или диагонализируемый, если в некотором базисе его матрица является диагональной.

Утверждение 107. *Пусть $\varphi\colon V \to V$ – линейный оператор в некотором векторном пространстве над полем F и пусть $\lambda_1, \ldots, \lambda_k \in F$ – различные числа. Тогда*

1. *Пространства $V_{\lambda_1}, \ldots, V_{\lambda_k}$ линейно независимы.*[100, 101]
2. *Пространства $V^{\lambda_1}, \ldots, V^{\lambda_k}$ линейно независимы.*

Доказательство. 1) В начале покажем случай собственных подпространств. Пусть $u_1 \in V_{\lambda_1}, \ldots, u_k \in V_{\lambda_k}$ – произвольные ненулевые векторы такие, что $u_1 + \ldots + u_k = 0$. Применим к этому равенству оператор $\varphi - \lambda_1 \operatorname{Id}$. Тогда u_1 занулится, а $(\varphi - \lambda_1 \operatorname{Id})u_i = (\lambda_i - \lambda_1)u_i$ будет ненулевым вектором из V_{λ_i} при $i \neq 1$. Обозначим эти векторы за $u'_2, \ldots, u'_k$. Тогда мы доказали, что если у нас дана сумма из k ненулевых векторов $u_1 + \ldots + u_k = 0$, то мы можем получить более короткую сумму из $k-1$ вектора $u'_2 + \ldots + u'_k = 0$.

2) Теперь давайте разберемся с корневыми. Пусть $v_1, \ldots, v_s$ – набор векторов такой, что $v_i \in V^{\lambda_i}$ и $v_1 + \ldots + v_s = 0$, где s – самое маленькое из возможных. Если $s = 1$, то имеем $v_1 = 0$ и доказывать нечего.

[100] Определение линейной независимости подпространств 89.

[101] Прошу обратить внимание, что линейно независимые подпространства могут быть нулевыми или часть из них может быть нулевыми.

Теперь считаем, что у нас $s > 1$. Так как v_s – корневой, то для некоторого m получаем $(\varphi - \lambda_s \operatorname{Id})^m v_s = 0$. Тогда получим

$$(\varphi - \lambda_s \operatorname{Id})^m v_1 + \ldots + (\varphi - \lambda_s \operatorname{Id})^m v_{s-1} = 0$$

Если мы покажем, что $(\varphi - \lambda_s \operatorname{Id})^m v_i$ лежит в V^{λ_i} и хотя бы одно из них не ноль, то мы придем к противоречию, так как получим более короткую сумму корневых векторов, дающую ноль.

Каждое подпространство V^{λ_i} является φ инвариантным. А значит и $\varphi - \lambda_s \operatorname{Id}$ инвариантным. А значит и $(\varphi - \lambda_s \operatorname{Id})^m$ инвариантным. Это показывает, что все слагаемые действительно остаются внутри соответствующего V^{λ_i}.

Теперь проверим, что хотя бы одно из слагаемых не равно нулю. Давайте покажем более сильное утверждение, если $v_i \neq 0$, то и $(\varphi - \lambda_s \operatorname{Id})^m v_i \neq 0$. По определению корневого пространства, мы можем найти такое d, что

$$(\varphi - \lambda_i \operatorname{Id})^d v_i = 0 \quad \text{и} \quad u_i = (\varphi - \lambda_i \operatorname{Id})^{d-1} v_i \neq 0$$

В частности $(\varphi - \lambda_i \operatorname{Id}) u_i = 0$, то есть u_i – собственный вектор. Чтобы показать, что $(\varphi - \lambda_s \operatorname{Id})^m v_i$ не нулевой, достаточно показать, что

$$(\varphi - \lambda_i \operatorname{Id})^{d-1} (\varphi - \lambda_s \operatorname{Id})^m v_i \neq 0$$

Действительно,

$$(\varphi - \lambda_i \operatorname{Id})^{d-1} (\varphi - \lambda_s \operatorname{Id})^m v_i = (\varphi - \lambda_s \operatorname{Id})^m (\varphi - \lambda_i \operatorname{Id})^{d-1} v_i = (\varphi - \lambda_s \operatorname{Id})^m u_i$$

Но по определению u_i – собственный вектор с собственным значением λ_i. Это значит, что умножение на φ совпадает с умножением на λ_i на векторе u_i. Значит

$$(\varphi - \lambda_s \operatorname{Id})^m u_i = (\lambda_i - \lambda_s)^m u_i \neq 0$$

□

Утверждение 108. *Пусть $\varphi \colon V \to V$ – некоторый линейный оператор в векторном пространстве над полем F. Тогда следующие утверждения эквивалентны:*

1. *Оператор φ диагонализуем.*
2. *Существует базис из собственных векторов.*
3. *Существует разложение $V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k}$.*
4. *(a) Характеристический многочлен раскладывается на линейные множители*

$$\chi_\varphi(t) = (t - \lambda_1)^{r_1} \ldots (t - \lambda_k)^{r_k}$$

   (b) для каждого i верно $\dim_F V_{\lambda_i} = r_i$.

Доказательство. (1)⇔(2). Пусть e – некоторый базис V. Тогда $\varphi e = eA$, где A – матрица φ в базисе e. По определению φ диагонализуем в базисе e тогда и только тогда, когда A диагональная. С другой стороны, все векторы в e собственные тогда и только тогда, когда A диагональная.

(2)⇒(3). Пусть e – базис из собственных векторов и $\varphi e = eA$, где A – диагональная с числами $\lambda_1, \ldots, \lambda_k$ на диагонали (эти числа могут повторяться). Для удобства переупорядочим вектора так, чтобы одинаковые числа λ_i шли по-порядку. Тогда базис e можно разделить на части $e = e_1 \sqcup \ldots \sqcup e_k$, где все векторы из e_i являются собственными с собственным значением λ_i. Значит $\langle v \mid v \in e_i \rangle \subseteq V_{\lambda_i}$. А значит

$$V = \langle e \rangle = \sum_i \langle e_i \rangle \subseteq \sum_i V_{\lambda_i} \subseteq V$$

То есть мы показали, что $V = \sum_i V_{\lambda_i}$ является суммой. С другой стороны, утверждение 107 гарантирует, что векторные подпространства V_{λ_i} линейно независимы, а значит сумма прямая (одно из эквивалентных определений по утверждению 91).

(3)⇒(2). Пусть $V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k}$ и пусть e_i – какой-нибудь базис V_{λ_i}. Тогда по одному из эквивалентных определений прямой суммы (утверждение 91) $e = e_1 \sqcup \ldots \sqcup e_k$ будет базисом для V. Тогда это и есть базис из собственных векторов.

(3)⇒(4). Выберем как в предыдущем пункте базис e_i в каждом слагаемом V_{λ_i}. Тогда $|e_i| = \dim_F V_{\lambda_i}$. Запишем матрицу нашего оператора в этом базисе в блочном виде

$$\varphi(e_1, \ldots, e_k) = (e_1, \ldots, e_k) \begin{pmatrix} \lambda_1 E & & \\ & \ddots & \\ & & \lambda_k E \end{pmatrix}$$

где размеры блоков равны в точности $r_i = |e_i| = \dim_F V_{\lambda_i}$. Тогда характеристический многочлен $\chi_\varphi(t) = (t-\lambda_1)^{r_1} \ldots (t-\lambda_k)^{r_k}$. Как мы видим многочлен разложился на линейные множители и кратности корней совпали с размерностями V_{λ_i}.

(4)⇒(3). Пусть $\lambda_1, \ldots, \lambda_k$ – все корни характеристического многочлена. Тогда по утверждению 107 сумма $V_{\lambda_1} + \ldots + V_{\lambda_k}$ всегда прямая. То есть мы имеем $V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k} \subseteq V$. И осталось лишь проверить равенство. Для этого посчитаем размерности. С одной стороны $\dim_F V = \deg \chi_\varphi$ по определению. С другой стороны размерность левой части есть

$$\sum_i \dim_F V_{\lambda_i} = \sum_i r_i = \deg \chi_\varphi$$

В первом равенстве мы воспользовались вторым условием, а во втором равенстве первым (если многочлен разложился на линейные множители, то сумма кратностей его корней равна степени). Значит обе размерности совпали и пространства оказались равны. □

Примеры

1. Рассмотрим оператор $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ по правилу $x \mapsto Ax$, где $A = \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$, $a, b \in \mathbb{R}$. Если $b \neq 0$ то этот оператор не диагонализуется, потому что его хар многочлен имеет только комплексные корни $a + bi$ и $a - bi$ и не имеет вещественных. Значит не выполняется пункт 4(a).

2. Теперь рассмотрим оператор заданный той же матрицей, но в случае комплексного векторного пространства $A\colon \mathbb{C}^2 \to \mathbb{C}^2$ по правилу $x \mapsto Ax$, где $A = \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$, $a, b \in \mathbb{R}$. Этот оператор диагонализуется и в некотором базисе записывается в виде $\left(\begin{smallmatrix} a+bi & 0 \\ 0 & a-bi \end{smallmatrix}\right)$.

3. Теперь рассмотрим оператор $A\colon \mathbb{C}^2 \to \mathbb{C}^2$ по правилу $x \mapsto Ax$, где $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$. Тогда его характеристический многочлен $\chi_A(t) = t^2$ раскладывается на линейные множители. Число $\lambda = 0$ является единственной точкой спектра, то есть это единственное собственное значение. Собственное подпространство V_λ для $\lambda = 0$ задается $\{x \in \mathbb{C}^2 \mid Ax = 0\}$, которое совпадает с $\langle e_1 \rangle$. То есть $\dim V_\lambda$ не равно кратности корня, то есть он не диагонализуется даже над $\mathbb{C}$.

Таким образом мы видим, что диагонализуемость зависит от поля. Часть причин недиагонализуемости – плохой выбор поля. В этом случае в утверждении 108 не выполняется условие 4(a). Такая проблема решается расширением поля до алгебраически замкнутого поля (так всегда можно сделать). Но последний пример показывает, что существуют операторы, которые не диагонализуются над любым полем. Это по-настоящему недиагонализуемые операторы. А действительно важное препятствие к диагонализуемости – это условие 4(b). Другими словами условие 4(b) означает, что размерность собственного подпространства должна совпасть с кратностью соответствующего собственного значения.

10.2 Свойства ограничения оператора

Теперь нам надо освоить несколько мелких технических утверждений связанных с поведением различных характеристик ограничения оператора на инвариантное подпространство.

Утверждение 109. *Пусть $\varphi\colon V \to V$ – линейный оператор и $U \subseteq V$ – инвариантное подпространство. Тогда*

1. *Если φ обратим, то $\varphi|_U$ обратим.*
2. $\operatorname{spec}_F \varphi|_U \subseteq \operatorname{spec}_F \varphi$.

Доказательство. (1) Если φ обратим, то $\ker \varphi = 0$, тогда $\ker \varphi|_U = \ker \varphi \cap U = 0$. А значит $\varphi|_U$ обратим по утверждению 95.

(2) Нам надо показать, что если $\lambda \notin \operatorname{spec}_F \varphi$, то $\lambda \notin \operatorname{spec}_F \varphi|_U$. То есть если $\varphi - \lambda \operatorname{Id}$ обратим, то и $\varphi|_U - \lambda \operatorname{Id}$ обратим. Но это следует из первого пункта и наблюдения

$$(\varphi - \lambda \operatorname{Id})|_U = \varphi|_U - (\lambda \operatorname{Id})|_U = \varphi|_U - \lambda \operatorname{Id}$$

□

Утверждение 110. *Пусть $V = U \oplus W$, $\varphi\colon V \to V$ – линейный оператор и подпространства U и W являются φ-инвариантными. Тогда*

1. $\operatorname{tr} \varphi = \operatorname{tr} \varphi|_U + \operatorname{tr} \varphi|_W$.
2. $\det \varphi = \det \varphi|_U \det \varphi|_W$.
3. $\chi_\varphi(t) = \chi_{\varphi|_U}(t)\chi_{\varphi|_W}(t)$.
4. $\operatorname{spec}_F \varphi = \operatorname{spec}_F \varphi|_U \cup \operatorname{spec}_F \varphi|_W$.

Доказательство. Пусть e – базис U и f – базис W. Тогда по одному из эквивалентных определений прямой суммы (утверждение 91) $e \cup f$ будет базисом V. Давайте запишем в этом базисе наш оператор φ:

$$\varphi(e, f) = (e, f) \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

Так как подпространство $U = \langle e \rangle$ φ-инвариантно, то есть $\varphi(U) \subseteq U$, то $\varphi(e)$ выражается только через e. Последнее означает, что $C = 0$. Аналогично, так как W φ-инвариантно, то $B = 0$. А значит

$$\varphi(e, f) = (e, f) \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}$$

При этом по определению A – это матрица $\varphi|_U$ в базисе e, а D – это матрица $\varphi|_W$ в базисе f. Тогда все четыре утверждения следуют из явного подсчета следа, определителя, характеристического многочлена и спектра для блочно диагональных матриц. □

10.3 Приведение к верхнетреугольному виду

Утверждение 111. *Пусть $\varphi\colon V \to V$ – линейный оператор в векторном пространстве размерности n над полем F и пусть $\lambda \in F$ – корень минимального многочлена для φ. Тогда существует базис, в котором матрица φ имеет следующий блочный вид*

$$A_\varphi = \begin{pmatrix} \lambda & * \\ 0 & B \end{pmatrix}, \text{ где } B \in \mathrm{M}_{n-1}(F)$$

Доказательство. Тут можно было бы воспользоваться утверждением 101. Тогда для корня λ подпространство V_λ не нулевое. А значит в нем есть искомый ненулевой вектор. Однако, полезно понимать, как найти собственный вектор с помощью минимального многочлена явно. Давайте проделаем это.

Пусть $f_{\min} = (t - \lambda)g(t)$. Тогда $g(\varphi) \neq 0$, то есть $g(\varphi)$ – ненулевой оператор. Последнее означает, что для какого-то вектора $v \in V$, $g(\varphi)v \neq 0$. Обозначим $u = g(\varphi)v$. Тогда это ненулевой вектор. С другой стороны

$$(\varphi - \lambda \operatorname{Id})u = (\varphi - \lambda \operatorname{Id})g(\varphi)v = f_{\min}(\varphi)v = 0$$

То есть u – ненулевой собственный вектор с собственным значением λ. Раз это ненулевой вектор, то множество $\{u\}$ линейно независимое. А значит его можно дополнить до базиса. Пусть это будет $u, u_2, \ldots, u_n$. Тогда в этом базисе матрица оператора φ будет иметь заявленный вид. Действительно, по определению

$$\varphi(u, u_2, \ldots, u_n) = (u, u_2, \ldots, u_n) \begin{pmatrix} a & * \\ w & B \end{pmatrix}$$

Тогда $\varphi u = au + (u_2, \ldots, u_n)w$. Но мы уже знаем, что $\varphi u = \lambda u$. То есть $a = \lambda$ и $w = 0$. □

Утверждение 112. *Пусть $\varphi\colon V \to V$ – линейный оператор в векторном пространстве над полем F и пусть минимальный многочлен f_{min} для φ раскладывается на линейные множители $(t - \lambda_1)^{k_1} \ldots (t - \lambda_r)^{k_r}$. Тогда существует базис в V такой, что матрица φ верхнетреугольная с числами $\lambda_1, \ldots, \lambda_r$ на диагонали (возможно с повторениями).*

Доказательство. Я не смогу вам дать доказательство полностью на языке операторов, так как мы не владеем некоторыми необходимыми техническими средствами в виде фактор пространств и фактор операторов. Потому надо будет переформулировать все в терминах матриц.

В начале выберем случайный базис в V. Тогда наш оператор превратится в $A\colon F^n \to F^n$. Нам надо найти такую обратимую матрицу $D \in \mathrm{M}_n(F)$, что $D^{-1}AD$ будет верхне треугольной с числами $\lambda_1, \ldots, \lambda_r$ на диагонали (может быть с повторениями).

Начнем. Так как минимальный многочлен раскладывается на линейные, то он имеет корень, например, выберем λ_1. Тогда предыдущее утверждение означает, что можно найти обратимую матрицу $C \in \mathrm{M}_n(F)$ такую, что

$$C^{-1}AC = \begin{pmatrix} \lambda_1 & * \\ 0 & B \end{pmatrix}$$

Заметим, что для произвольного многочлена f верно

$$f\begin{pmatrix} \lambda_1 & * \\ 0 & B \end{pmatrix} = \begin{pmatrix} f(\lambda_1) & * \\ 0 & f(B) \end{pmatrix}$$

А значит $f_{\min}$ зануляет блок B. То есть минимальный многочлен для B тоже раскладывается на линейные множители и его корни находятся среди корней $f_{\min}$, так как минимальный для B делит $f_{\min}$. А значит, для B по индукции найдется такая обратимая матрица $T \in \mathrm{M}_{n-1}(F)$, что $T^{-1}BT$ является верхне треугольной с числами $\lambda_1, \ldots, \lambda_r$ на диагонали (быть может с повторениями и пропусками). Тогда

$$\begin{pmatrix} 1 & \\ & T \end{pmatrix}^{-1} C^{-1}AC \begin{pmatrix} 1 & \\ & T \end{pmatrix} = \begin{pmatrix} 1 & \\ & T \end{pmatrix}^{-1} \begin{pmatrix} \lambda_1 & * \\ 0 & B \end{pmatrix} \begin{pmatrix} 1 & \\ & T \end{pmatrix} = \begin{pmatrix} \lambda_1 & * \\ 0 & T^{-1}BT \end{pmatrix}$$

Последняя матрица верхне треугольная с числами $\lambda_1, \ldots, \lambda_r$ на диагонали (быть может с повторениями). То есть мы доказали, что хотели с матрицей

$$D = C\begin{pmatrix} 1 & \\ & T \end{pmatrix}$$

□

Утверждение 113. *Пусть $\varphi\colon V \to V$ – некоторый линейный оператор на векторном пространстве V над полем F. Тогда $\chi_{\varphi|_{V^\lambda}}(t) = (t-\lambda)^{\dim V^\lambda}$.*

Доказательство. Оператор $\varphi|_{V^\lambda}$ зануляется многочленом вида $(t-\lambda)^d$. Значит его минимальный многочлен имеет вид $(t-\lambda)^k$. По утверждению 112 матрица оператора $\varphi|_{V^\lambda}$ приводится к верхнетреугольному виду с λ на диагонали. А значит $\chi_{\varphi|_{V^\lambda}}(t) = (t-\lambda)^{\dim V^\lambda}$. □

10.4 Идеальный спектр

Определение 114. Пусть $\varphi\colon V \to V$ – линейный оператор над произвольным полем F, тогда идеальный спектр φ это следующее множество:

$$\operatorname{spec}_F^I \varphi := \{p \in F[t] \mid p \text{ – неприводим со старшим коэффициентом 1 и } p(\varphi) \text{ необратим}\}$$

Пример Если рассмотреть линейный оператор $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ заданный матрицей $A = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, то его минимальный многочлен будет $f = x^2+1$. Таким образом вещественный спектр пуст $\operatorname{spec}_{\mathbb{R}} A = \varnothing$, так как f неприводим и не линеен, а значит не имеет корней. Однако идеальный спектр будет непустым $\operatorname{spec}_{\mathbb{R}}^I A = \{x^2+1\}$.

Утверждение 115. *Пусть $\varphi\colon V \to V$ – некоторый оператор над произвольным полем F и f_{min} – его минимальный зануляющий многочлен над F. Тогда*

1. *Для любого зануляющего многочлена f и любого $p \in \operatorname{spec}_F^I \varphi$ следует, что p делит f.*
2. *$p \in \operatorname{spec}_F^I \varphi$ тогда и только тогда, когда p делит f_{min}.*

Доказательство. (1) Для этого достаточно показать, что если p не делит f, то $p(\varphi)$ обратим. Действительно, так как p неприводим, это означает, что f и p взаимно просты. Тогда по расширенному алгоритму Евклида

мы знаем, что $1 = u(t)p(t) + v(t)f(t)$ для некоторых многочленов $u(t), v(t) \in F[t]$. Тогда подставив в последнее равенство φ мы видим

$$\mathrm{Id} = u(\varphi)p(\varphi) + v(\varphi)f(\varphi) = u(\varphi)p(\varphi)$$

То есть $u(\varphi)$ – обратный к $p(\varphi)$, что и требовалось.

(2) Пусть теперь $f_{\min}$ – минимальный многочлен и пусть p – неприводимый делитель $f_{\min}$, то есть $f_{\min} = ph$. Надо показать, что $p(\varphi)$ необратим.

Мы знаем, что $0 = f_{\min}(\varphi) = p(\varphi)h(\varphi)$. Предположим, что $p(\varphi)$ обратим. Тогда в равенстве $p(\varphi)h(\varphi) = 0$ можно сократить на $p(\varphi)$. Значит $h(\varphi) = 0$, что противоречит минимальности φ. □

Замечания

- Пусть $f_{\min}$ раскладывается на линейные множители

  $$f_{\min}(t) = (t - \lambda_1)^{k_1} \dots (t - \lambda_r)^{k_r}$$

  Тогда идеальный спектр φ – это в точности многочлены $\{t-\lambda_1, \dots, t-\lambda_r\}$. То есть каждый элемент идеального спектра однозначно соответствует элементу обычного спектра $\operatorname{spec}_F \varphi = \{\lambda_1, \dots, \lambda_r\}$. Потому идеальный спектр можно рассматривать как обобщение понятия спектра на случай, когда в минимальном многочлене есть нелинейные множители.

- Последнее утверждение можно рассматривать как обобщение утверждений 8 и 11 о том, что спектр лежит среди корней зануляющего многочлена и в точности совпадает с корнями минимального.

- Так как минимальный многочлен делит характеристический, то любой элемент идеального спектра является делителем характеристического многочлена.

  На самом деле можно показать, что элементы идеального спектра – это в точности делители характеристического многочлена. Но я не буду вас мучить доказательством этого утверждения нашими методами.

  Давайте я намекну про правильный способ. Пусть $\varphi\colon V \to V$ – некоторый оператор над полем F и пусть $f_{\min} = p_1^{k_1} \dots p_r^{k_r}$. Предположим, что многочлен χ_φ имеет неприводимый делитель p отличный от всех p_i. Пусть $\bar{F}$ – алгебраическое замыкание F. Тогда можно заменить оператор φ на его версию над $\bar{F}$, а именно $\varphi_{\bar{F}}\colon V_{\bar{F}} \to V_{\bar{F}}$, у которого будет тот же минимальный и характеристический многочлен. Например, это можно сделать, выбрав базис в V, оно превращается в F^n, потом взять $V_{\bar{F}} = \bar{F}^n$ и в нем задать $\varphi_{\bar{F}}$ той же матрицей, что и φ. Характеристический многочлен не изменится, потому что матрица та же самая, но надо пояснить, почему минимальный многочлен не изменится. В этом случае есть общая конструкция для $V_{\bar{F}}$, которая определяется так, что неизменность минимального многочлена будет очевидна (можно и руками показать, выбрав базис $\bar{F}$ как векторного пространства над F). После чего мы видим, что $f_{\min}$ и χ_φ имеют одни и те же корни в $\bar{F}$, то есть p имеет общий корень с каким-то p_i. Но это не возможно. Действительно, в силу их взаимной простоты, мы имеем $1 = u(t)p(t) + v(t)p_i(t)$. И если у них есть общий корень, то после подстановки его в равенсто, справа будет ноль, а слева – единица. На этом победа.[102]

Утверждение 116 (БД)**.** *Пусть $\varphi\colon V \to V$ – некоторый оператор над произвольным полем F. Тогда $p \in \operatorname{spec}_F^I \varphi$ тогда и только тогда, когда p делит χ_φ.*

10.5 Обобщение собственных и корневых подпространств

Определение 117. Пусть $\varphi\colon V \to V$ – линейный оператор над произвольным полем и $p \in \operatorname{spec}_F^I \varphi$. Тогда определим корневое подпространство как:

$$V^p = \{v \in V \mid \exists k\colon p^k(\varphi)v = 0\} = \bigcup_{k \geqslant 0} \ker p^k(\varphi)$$

и собственное подпространство

$$V_p = \{v \in V \mid p(\varphi)v = 0\} = \ker p(\varphi)$$

[102] В курсе алгебры вам расскажут как для любого поля F и любого многочлена $g \in F[t]$ построить большее поле $L \supseteq F$ такое, что в нем g раскладывается на линейные множители. Как мы видим из доказательства, этого нам достаточно. Остается аккуратно объяснить, почему не изменится минимальный многочлен и вы будете готовы доказать этот факт.

Замечания

- В силу леммы о стабилизации $V^p = \ker p^k(\varphi)$ для некоторого достаточно большого $k \leqslant \dim_F V$.
- Это определение является обобщением корневого и собственного подпространства на случай идеального спектра. Действительно, если $\lambda \in \operatorname{spec}_F \varphi$, то в идеальном спектре ему соответствует $p(t) = t - \lambda$. Тогда определения превращаются в те же самые, что были даны в предыдущих разделах.

10.6 Теоремы о разложении

Утверждение 118. *Пусть $\varphi\colon V \to V$ – линейный оператор над полем F и пусть $p, q \in F[t]$ – два взаимно простых многочлена. Тогда*

1. $\ker p(\varphi) \cap \ker q(\varphi) = 0$.
2. *Оператор $p(\varphi)|_{\ker q(\varphi)}$ существует и обратим.*

Доказательство. (1) Из того, что многочлены p и q взаимнопросты, по расширенному алгоритму Евклида, найдутся многочлены $u, v \in F[t]$ такие, что $1 = u(t)p(t) + v(t)q(t)$. Пусть теперь w – вектор из пересечения, тогда

$$w = u(\varphi)p(\varphi)w + v(\varphi)q(\varphi)w = 0$$

(2) Так как операторы $p(\varphi)$ и $q(\varphi)$ коммутируют, то ядро $q(\varphi)$ инвариантно относительно $p(\varphi)$ по утверждению 97. А значит существует оператор ограничения. Так как для операторов обратимость равносильна инъективности (утверждение 95), то нам достаточно показать, что $\ker q(\varphi)|_{\ker p(\varphi)} = 0$. Но $\ker q(\varphi)|_{\ker p(\varphi)} = \ker q(\varphi) \cap \ker p(\varphi) = 0$. □

Утверждение 119. *Пусть $\varphi\colon V \to V$ – линейный оператор над полем F, $f \in F[t]$ зануляющий многочлен такой, что $f = pq \in F[t]$, где $(p, q) = 1$. Тогда*

1. $\ker p(\varphi) = \operatorname{Im} q(\varphi)$.
2. $\ker q(\varphi) = \operatorname{Im} p(\varphi)$.
3. $V = \ker p(\varphi) \oplus \ker q(\varphi)$.
4. *Для любого φ инвариантного подпространства $U \subseteq V$ найдутся два инвариантных подпространства $W \subseteq \ker p(\varphi)$ и $E \subseteq \ker q(\varphi)$ такие, что $U = W \oplus E$. Более того подпространства W и E можно восстановить одним из двух способов:*
   (a) *$W = U \cap \ker p(\varphi)$ и $E = U \cap \ker q(\varphi)$.*
   (b) *$W = \pi_1(U)$, где $\pi_1\colon V \to V$ – проектор на $\ker p(\varphi)$ вдоль $\ker q(\varphi)$. И аналогично для $E = \pi_2(U)$, где $\pi_2\colon V \to V$ – проектор на $\ker q(\varphi)$ вдоль $\ker p(\varphi)$.*
5. *Если f – минимальный многочлен для φ, то многочлен p будет минимальным для оператора $\varphi|_{\ker p(\varphi)}$.*

Доказательство. В начале сделаем некие общие подготовительные работы. Из того, что многочлены p и q взаимнопросты, по расширенному алгоритму Евклида, найдутся многочлены $u, v \in F[t]$ такие, что $1 = a(t)p(t) + b(t)q(t)$. Подставим в это равенство и в f оператор φ, получим два равенства

$$\operatorname{Id} = a(\varphi)p(\varphi) + b(\varphi)q(\varphi)$$
$$0 = p(\varphi)q(\varphi)$$

(1) и (2). Так как утверждения (1) и (2) симметричны, то достаточно доказать одно из них.

В начале покажем, что $\operatorname{Im} q(\varphi) \subseteq \ker p(\varphi)$. Так как $p(\varphi)q(\varphi) = 0$, то для любого $v \in V$ верно, что $p(\varphi)q(\varphi)v = 0$, но это означает, что $q(\varphi)v \in \ker p(\varphi)$, то есть $\operatorname{Im} q(\varphi) \subseteq \ker p(\varphi)$.

Наоборот, возьмем $v \in \ker p(\varphi)$ и применим к нему первое операторное равенство, получим

$$v = a(\varphi)p(\varphi)v + b(\varphi)q(\varphi)v = b(\varphi)q(\varphi)v = q(\varphi)b(\varphi)v \in \operatorname{Im} q(\varphi)$$

(3) Из взаимной простоты p и q следует, что $\ker p(\varphi) \cap \ker q(\varphi) = 0$ по утверждению 118. Значит сумма этих подпространств прямая, то есть $\ker p(\varphi) \oplus \ker q(\varphi) \subseteq V$. Чтобы показать равенство, возьмем произвольный $v \in V$ и рассмотрим

$$v = a(\varphi)p(\varphi)v + b(\varphi)q(\varphi)v = p(\varphi)a(\varphi)v + q(\varphi)b(\varphi)v \in \operatorname{Im} p(\varphi) + \operatorname{Im} q(\varphi) = \ker q(\varphi) + \ker p(\varphi)$$

Здесь последнее равенство выполнено в силу предыдущих двух пунктов.

(4) Пусть $U \subseteq V$ инвариантное подпространство. Давайте определим подпространства W и E следующим образом:

$$W = \{q(\varphi)u \mid u \in U\} = q(\varphi)U \subseteq \operatorname{Im} q(\varphi) = \ker p(\varphi) \quad \text{и} \quad E = \{p(\varphi)u \mid u \in U\} = p(\varphi)U \subseteq \operatorname{Im} p(\varphi) = \ker q(\varphi)$$

В частности $W \cap E = 0$. Ясно, что построенные подпространства будут φ инвариантны. Теперь проверим, что $U = W + E$, для этого применим операторное равенство

$$\mathrm{Id} = a(\varphi)p(\varphi) + b(\varphi)q(\varphi)$$

к вектору $u \in U$ и получим

$$u = a(\varphi)p(\varphi)u + b(\varphi)q(\varphi)u$$

Но тогда

$$a(\varphi)p(\varphi)u = p(\varphi)(a(\varphi)u) \in p(\varphi)(U) = E$$

Аналогично проверяется, что второе слагаемое лежит в W, что и доказывает, что $U = W \oplus E$. Теперь покажем, почему слагаемые W и E восстанавливаются двумя способами. Если $U = W \oplus E \subseteq \ker p(\varphi) \oplus \ker q(\varphi)$, то легко видеть, что

$$U \cap \ker p(\varphi) = W = \pi_1(U) \quad \text{и} \quad U \cap \ker q(\varphi) = E = \pi_2(U)$$

(5) Теперь мы считаем, что $f = f_{\min}$ для φ на V. Заметим, что

$$p(\varphi|_{\ker p(\varphi)}) = p(\varphi)|_{\ker p(\varphi)} = 0$$

Значит p зануляет $\varphi|_{\ker p(\varphi)}$. Так как минимальный многочлен обязательно делит p, то нам надо показать, что никакой делитель p отличный от p не зануляет $\varphi|_{\ker p(\varphi)}$. Предположим противное, пусть $p_0|p$ и $p_0(\varphi|_{\ker p(\varphi)}) = 0$. Тогда рассмотрим многочлен $g = p_0 q$ и покажем, что $g(\varphi) = 0$. Так как $V = \ker p(\varphi) \oplus \ker q(\varphi)$, то нам достаточно показать, что $g(\varphi)$ действует нулем на любом векторе из $\ker p(\varphi)$ и на любом векторе из $\ker q(\varphi)$. Но на $\ker q(\varphi)$ нулем действует $q(\varphi)$, а $g(\varphi) = p_0(\varphi)q(\varphi)$. А на $\ker p(\varphi)$ оператор $p_0(\varphi)$ действует нулем по выбору p_0, а значит и $g(\varphi) = q(\varphi)p_0(\varphi)$ действует нулем. То есть многочлен g зануляет φ и имеет степень меньше, чем $f_{\min}$, противоречие. □

Утверждение 120. *Пусть $\varphi\colon V \to V$ – линейный оператор, f_{min} – его минимальный многочлен. Пусть*

$$f_{min} = p_1^{k_1} \dots p_r^{k_r}$$

разложение минимального в неприводимые многочлены. Тогда

1. *$V^{p_i} = \ker p_i^{k_i}(\varphi)$ причем k_i – минимальное такое k для которого выполнено равенство $V^{p_i} = \ker p_i^k(\varphi)$.*
2. *$V = V^{p_1} \oplus \dots \oplus V^{p_r}$.*
3. *Любое инвариантное подпространство $U \subseteq V$ имеет вид $U = U_1 \oplus \dots \oplus U_r$, где $U_i \subseteq V^{p_i}$ – произвольные инвариантные подпространства.*

Доказательство. 1) Рассмотрим разложение многочлена $f_{\min}$ на следующие множители

$$f_{\min} = \underbrace{p_1^{k_1}}_{p} \underbrace{p_2^{k_2} \dots p_r^{k_r}}_{q}$$

Тогда $V = \ker p(\varphi) \oplus \ker q(\varphi)$, то есть $V = \ker p_1^{k_1}(\varphi) \oplus \ker q(\varphi)$. По определению $\ker p_1^{k_1}(\varphi) \subseteq V^{p_1}$. Если $\ker p_1^{k_1}(\varphi) \neq V^{p_1}$, то V^{p_1} обязано пересекать $\ker q(\varphi)$ не по нулю. По лемме о стабилизации (утверждение 105) найдется такое N, что $V^{p_1} = \ker p_1^N(\varphi)$. Так как p_1 и q взаимнопросты, то из пункта (1) утверждения 118, следует $\ker p_1^N \cap \ker q(\varphi) = 0$, противоречие. Но теперь пункт (5) предыдущего утверждения 119 гласит, что так

как $f_{\min}$ был минимальным для φ на всем пространстве V, то $p_1^{k_1}$ является минимальным многочленом для φ ограниченным на V^{p_1}, то есть ни в какой меньшей степени k, чем k_1 мы не получим равенство $V^{p_1} = \ker p_1^k(\varphi)$.

2) Имея разложением

$$f_{\min} = \underbrace{p_1^{k_1}}_{p}\underbrace{p_2^{k_2}\dots p_r^{k_r}}_{q}$$

мы только что получили разложение

$$V = \ker p_1^{k_1}(\varphi) \oplus \ker q(\varphi)$$

Подпространство $\ker q(\varphi)$ является φ инвариантным, так как φ и $q(\varphi)$ коммутируют (утверждение 97). Пусть ψ – ограничение φ на $\ker q(\varphi)$. Так как мы рассматривали минимальный многочлен для φ на всем пространстве V, то q будет минимальным для ψ на $\ker q(\varphi)$ (пункт (5) утверждение 119). Тогда мы можем индукцией по количеству неприводимых получить разложение

$$\ker q(\varphi) = \ker p_2^{k_2}(\varphi) \oplus \ldots \oplus \ker p_r^{k_r}(\varphi)$$

Объединяя это с результатом первого пункта мы получаем разложение $V = V^{p_1} \oplus \ldots \oplus V^{p_r}$.

3) Это непосредственно следует из утверждения 119 пункт (4) индукцией по количеству прямых слагаемых. □

10.7 Геометрический смысл кратности корней минимального и характеристического многочлена

Утверждение 121. *Пусть $\varphi\colon V \to V$ – линейный оператор и $\lambda \in \operatorname{spec}_F \varphi$. Пусть число k выбрано так, что*

$$\ker(\varphi - \lambda \operatorname{Id})^{k-1} \neq \ker(\varphi - \lambda \operatorname{Id})^k = \ker(\varphi - \lambda \operatorname{Id})^{k+1}$$

Тогда

1. *Число k – это кратность λ в минимальном многочлене оператора φ.*
2. *Число $\dim \ker(\varphi - \lambda \operatorname{Id})^k = \dim V^\lambda$ – это кратность корня λ в характеристическом многочлене оператора φ.*

Доказательство. (1) Это частный случай утверждения 120 пункт (1) для $p_1(x) = x - \lambda$.

(2) Пусть $f_{\min} = (x-\lambda)^k g(x)$ – минимальный многочлен для φ на V. Тогда пользуясь утверждением 119 пункт (3) и предыдущим пунктом этого утверждения, мы видим, что $V = V^\lambda \oplus \ker g(\varphi)$. Теперь давайте посчитаем характеристический многочлен для φ. По утверждению 110 $\chi_\varphi(t)$ есть произведение $\chi_{\varphi|_{V^\lambda}}$ и $\chi_{\varphi|_{\ker g(\varphi)}}$. Утверждение 113 гласит, что $\chi_{\varphi|_{V^\lambda}}(t) = (t-\lambda)^{\dim V^\lambda}$. А утверждение 118 пункт (2), что $\varphi - \lambda \operatorname{Id}$ обратим на $\ker g(\varphi)$. Значит, оператор $\varphi|_{\ker g(\varphi)}$ не содержит λ в своем спектре, а значит λ не корень $\chi_{\varphi|_{\ker g(\varphi)}}$. Это завершает доказательство. □

10.8 Минимальные инвариантные

Пусть $\varphi\colon V \to V$ – некоторый оператор и $p \in \operatorname{spec}_F^I \varphi$. В этом случае подпространство $V_p = \ker p(\varphi)$ не нулевое. Более того, по определению оператор $p(\varphi)$ равен нулю на этом подпространстве, а значит p зануляет $\varphi|_{V_p}$. В частности минимальный многочлен $\varphi|_{V_p}$ должен делить p. Но так как p неприводим, то единственный вариант – минимальный многочлен совпадает с p. Аналогично, если $U \subseteq V_p$ произвольное ненулевое инвариантное подпространство, то минимальный многочлен φ_U будет p по тем же самым соображениям. Мы знаем, что в случае обычного спектра собственное подпространство V_λ состоит из инвариантных прямых, на которых φ действует растяжением в λ раз. В случае V_p все подпространство состоит из одинаковых инвариантных кусочков, которые уже не являются прямыми. Давайте опишем как эти инвариантные подпространства выглядят.

Утверждение 122. *Пусть $\varphi\colon V \to V$ – некоторый оператор и $p \in \operatorname{spec}_F^I \varphi$ и при этом $m = \deg p$. И пусть $v \in V_p$ – произвольный ненулевой вектор. Тогда $v, \varphi v, \ldots, \varphi^{m-1} v$ линейно независимы и их линейна оболочка будет минимальным инвариантным подпространством содержащем v.*

Доказательство. Давайте рассмотрим линейную оболочку

$$U = \langle v, \varphi v, \varphi^2 v, \ldots, \varphi^k v, \ldots \rangle$$

По построению ясно, что это инвариантное подпространство содержащее v. Кроме того, если какое-то инвариантное подпространство W содержит v, то оно обязано содержать φv. А значит обязано содержать $\varphi^2 v$ и так далее. То есть оно содержит U. Таким образом U является наименьшим инвариантным подпространством содержащим v.

Теперь рассмотрим элемент $p = t^m + a_{m-1}t^{m-1} + \ldots + a_1 t + a_0$. Подставим в него φ и получим ноль, то есть

$$0 = p(\varphi) = \varphi^m + a_{m-1}\varphi^{m-1} + \ldots + a_1\varphi + a_0 \operatorname{Id}$$

Теперь применим это равенство к вектору v, получим

$$\varphi^m v + a_{m-1}\varphi^{m-1}v + \ldots + a_1\varphi v + a_0 v = 0$$

А значит $\varphi^m v$ выражается через $\varphi^k v$, где $k < m$. Умножая это равенство на φ^N мы видим, что при $N \geqslant m$, $\varphi^N v$ выражается через $\varphi^k v$, где $k < N$. Следовательно подпространство U порождается векторами $v, \varphi v, \ldots, \varphi^{m-1}v$.

Теперь покажем, что эти векторы линейно независимы. Предположим противное, то есть найдется система коэффициентов $\alpha_0, \ldots, \alpha_{m-1}$ такая, что

$$\alpha_0 v + \alpha_1\varphi v + \ldots + \alpha_{m-1}\varphi^{m-1}v = 0$$

Если мы положим $q(t) = \alpha_0 + \alpha_1 t + \ldots + \alpha_{m-1}t^{m-1}$, то это означает, что $q(\varphi)v = 0$. Так как $q(\varphi)$ и φ^k коммутируют для любого k, то $q(\varphi)\varphi^k v = \varphi^k q(\varphi)v = 0$ для любого k. А это значит, что q зануляет $\varphi|_U$. Но по замечанию перед утверждением мы знаем, что минимальный многочлен для $\varphi|_U$ должен быть p. А значит p делит q и так как $\deg q < \det p$, такое возможно только если $q = 0$. То есть если все $\alpha_i = 0$, что и требовалось. □

Таким образом в случае $\lambda \in \operatorname{spec}_F \varphi$ ему соответствует линейный многочлен $t - \lambda \in \operatorname{spec}_F^I \varphi$. А значит линейному многочлену соответствуют одномерные инвариантные подпространства в V_λ. В случае если элемент идеального спектра p имеет степень больше единицы, то V_p содержит инвариантные размерности $\deg p$. На самом деле можно показать, что все такие инвариантные подпространства внутри V_p «одинаковые» в некотором смысле и все V_p есть их прямая сумма.

10.9 Структура векторного пространства с оператором

Изучение структуры матрицы линейного оператора в некотором базисе равносильна изучению инвариантных подпространств пространства V. Теорема о жордановой нормальной форме может рассматриваться таким образом как структурная теорема для векторного пространства с оператором. В общем виде теорема о жордановой нормальной форме доказывается в два шага: все пространство раскладывается в прямую сумму корневых, после чего задача сводится к нильпотентному оператору. Первый шаг мы уже на самом деле проделали в утверждении 120.

Утверждение 123. *Пусть $\varphi\colon V \to V$ – линейный оператор такой, что его характеристический (или минимальный) многочлен раскладывается на линейные множители. Тогда $V = V^{\lambda_1} \oplus \ldots \oplus V^{\lambda_r}$, где $\operatorname{spec}_F \varphi = \{\lambda_1, \ldots, \lambda_r\}$.*

Доказательство. Это частный случай утверждения 120. □

Замечание Давайте объясним, что мы доказали на данный момент. Пусть $\varphi\colon V \to V$ – некоторый линейный оператор, у которого характеристический многочлен раскладывается на линейные множители. Тогда $V = V^{\lambda_1} \oplus \ldots \oplus V^{\lambda_r}$. Выберем базис e_i в каждом V^{λ_i}. Тогда по одному из определений прямой суммы $e = e_1 \sqcup \ldots \sqcup e_r$ будет базисом V. Так как все V^{λ_i} инвариантны относительно φ, то когда мы запишем его матрицу в этом базисе, мы получим блочно диагональную матрицу вида

$$A_\varphi = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_r \end{pmatrix}$$

где A_i – это матрица ограничения $\varphi|_{V^{\lambda_i}}$, то есть ее размер равен $\dim V^{\lambda_i}$. Кроме того, из утверждения 112 следует, что в каждом V^{λ_i} можно найти такой базис, что матрица A_i будет верхне треугольной с числом λ_i на диагонали, то есть

$$A_i = \begin{pmatrix} \lambda_i & * & \dots & * \\ & \lambda_i & \dots & * \\ & & \ddots & \vdots \\ & & & \lambda_i \end{pmatrix}$$

Наша цель еще улучшить вид матриц A_i. Оказывается, почти все элементы верхнего блока можно сделать нулевыми. Что это в точности означает и как доказывается, вы узнаете в теореме о жордановой нормальной форме.

10.10 Отношение равенства по модулю подпространства

Определение 124. Пусть V – векторное пространство, а $U \subseteq V$ – подпространство. Тогда будем говорить, что векторы $v, w \in V$ равны по модулю U и писать $v = w \pmod U$, если $v - w \in U$.[103]

Например, если $V = \mathbb{R}^2$ – плоскость и $U = \langle e_1 \rangle$ – горизонтальная прямая, то все векторы лежащие на горизонтальных прямых между собой равны по модулю U. Например, $e_2 = e_2 + e_1 = e_2 - 3e_1 \pmod U$. Но $e_2 \neq 2e_2 \pmod U$.

Заметим, что обычное равенство – это равенство по модулю нулевого подпространства. С другой стороны, по модулю подпространства $U = V$ любые два вектора равны.

Определение 125. Пусть V – векторное пространство, $U \subseteq V$ – некоторое подпространство, и $v_1, \dots, v_n \in V$ – набор векторов.

1. Будем говорить, что $v_1, \dots, v_n$ линейно независимы по модулю U, если из равенства $\alpha_1 v_1 + \dots + \alpha_n v_n = 0 \pmod U$ следует, что $\alpha_1 = \dots = \alpha_n = 0$.
2. Будем говорить, что $v_1, \dots, v_n$ порождающие по модулю U, если для любого вектора $v \in V$ найдутся коэффициенты $\alpha_1, \dots, \alpha_n \in F$ такие, что $v = \alpha_1 v_1 + \dots + \alpha_n v_n \pmod U$.
3. Будем говорить, что $v_1, \dots, v_n$ являются базисом V по модулю U, если они одновременно линейно независимы и порождающие по модулю U.[104]

Утверждение 126. *Пусть V – векторное пространство, $U \subseteq V$ – подпространство, и $v_1, \dots, v_n \in V$ – набор векторов. Тогда*

1. *Векторы $v_1, \dots, v_n$ линейно независимы по модулю U тогда и только тогда, когда они линейно независимы и $\langle v_1, \dots, v_n \rangle \cap U = 0$.*
2. *Векторы $v_1, \dots, v_n$ порождающие по модулю U тогда и только тогда, когда $\langle v_1, \dots, v_n \rangle + U = V$.*
3. *Векторы $v_1, \dots, v_n$ являются базисом по модулю U тогда и только тогда, когда они линейно независимы и $\langle v_1, \dots, v_n \rangle \oplus U = V$.*

Доказательство. (1) $\Rightarrow$ Пусть $\alpha_1 v_1 + \dots + \alpha_n v_n = 0$, тогда $\alpha_1 v_1 + \dots + \alpha_n v_n \in U$. Последнее означает, что $\alpha_1 v_1 + \dots + \alpha_n v_n = 0 \pmod U$. А значит все $\alpha_i = 0$. Значит v_i линейно независимы. Теперь рассмотрим вектор $v \in \langle v_1, \dots, v_n \rangle \cap U$. Так как v лежит в первом подпространстве, то $v = \alpha_1 v_1 + \dots + \alpha_n v_n$. Так как он лежит в правом подпространстве, то $\alpha_1 v_1 + \dots + \alpha_n v_n = v \in U$. Значит $\alpha_1 v_1 + \dots + \alpha_n v_n = 0 \pmod U$. А следовательно все $\alpha_i = 0$. Но значит и $v = 0$, что и требовалось.

$\Leftarrow$ Пусть $\alpha_1 v_1 + \dots + \alpha_n v_n = 0 \pmod U$. Это значит, что $\alpha_1 v_1 + \dots + \alpha_n v_n \in U$. А значит $\alpha_1 v_1 + \dots + \alpha_n v_n \in \langle v_1, \dots, v_n \rangle \cap U = 0$. То есть $\alpha_1 v_1 + \dots + \alpha_n v_n = 0$. Но так как v_i линейно независимы, то $\alpha_i = 0$ для всех i.

(2) $\Rightarrow$ По определению, для любого $v \in V$ найдутся коэффициенты α_i такие, что $v = \alpha_1 v_1 + \dots + \alpha_n v_n \pmod U$. То есть $v - (\alpha_1 v_1 + \dots + \alpha_n v_n) = u \in U$. Значит, $v = \alpha_1 v_1 + \dots + \alpha_n v_n + u$, что и требовалось.

$\Leftarrow$ Пусть $v \in V = \langle v_1, \dots, v_n \rangle + U$. Тогда $v = \alpha_1 v_1 + \dots + \alpha_n v_n + u$. По определению это означает, что $v = \alpha_1 v_1 + \dots + \alpha_n v_n \pmod U$.

(3) Этот пункт получается из первых двух вместе взятых. □

[103] На самом деле равенство по модулю подпространства сводится к равенству в некотором новом пространстве, которое называется фактор пространством. Так как мы пока не знаем, что это такое, будем пользоваться лишь отношением равенства по модулю. Думать про него надо так же, как и про остатки в целых числах.

[104] Так как равенство по модулю сводится к равенству в некотором новом пространстве, то все факты про базис аналогичные обычным фактам, что мы доказывали будут верны. Вам же я предлагаю доказать их по аналогии в качестве упражнения.