МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Физический факультет

Кафедра высшей математики физического факультета

А. П. Ульянов

ГЕОМЕТРИЯ И АЛГЕБРА ДЛЯ СТУДЕНТОВ-ФИЗИКОВ

Семестр 1

Учебное пособие
по курсу линейной алгебры и геометрии

Версия от 30 января 2021 г.

Новосибирск
2007–2021

УДК: 510
ББК: В14я73-1 + В151.54я73-1
У 517

Ульянов А. П. Геометрия и алгебра для студентов-физиков: Учеб. пособие / Новосиб. гос. ун-т. Новосибирск, 2010, 2019. Ч. 1. 113?? с.

Пособие содержит краткий конспект лекций, прочитанных автором для студентов 1-го курса физического и геолого-геофизического факультетов НГУ в осеннем семестре 2019 года.

В пособие включены следующие темы: трёхмерная векторная алгебра, прямые и плоскости; матрицы, комплексные числа и многочлены; линии и поверхности второго порядка; системы линейных уравнений и начала линейной алгебры; определители и полиномиальная интерполяция; квадратичные формы.

Версия от 30 января 2021 г.

ISBN 978-5-94356-571-7

Содержание

Глава 1. Векторная алгебра и геометрия

1.1 Векторы в пространстве 4
1.2 Скалярное, смешанное и векторное произведения 7
1.3 Определители второго и третьего порядков 12
1.4 Отложенные геометрические доказательства 17
1.5 Задание прямой и плоскости 18
1.6 Взаимное расположение прямых и плоскостей 25

Глава 2. Матрицы, комплексные числа, многочлены

2.1 Введение в системы линейных уравнений 30
2.2 Смена координат и матричный язык 36
2.3 Комплексные числа 41
2.4 Основная теорема алгебры многочленов 47
2.5 Изображение функций комплексной переменной 50

Глава 3. Фигуры второго порядка

3.1 Эллипсы, параболы и гиперболы 52
3.2 Общее уравнение линии второго порядка 57
3.3 Поверхности второго порядка 61

Глава 4. Начала линейной алгебры

4.1 Системы линейных уравнений 1 67
4.2 Линейные пространства строк и столбцов 72
4.3 Системы линейных уравнений 2 76
4.4 Действия с подпространствами 79
4.5 Примеры линейных пространств 85

Глава 5. Определители

5.1 Комбинаторное строение определителей 87
5.2 Свойства определителей 91
5.3 Невырожденные матрицы 94
5.4 Полиномиальная интерполяция 96

Глава 6. Квадратичные формы

6.1 Введение и мотивация 102
6.2 Билинейные и квадратичные формы 106
6.3 Вещественные квадратичные формы 110

Глава 1. ВЕКТОРНАЯ АЛГЕБРА И ГЕОМЕТРИЯ

1.1. Векторы в пространстве

Лекция 1
02.09.19

Определение вектора

Знакомясь с векторами, мы опираемся на известные из элементарной геометрии понятия: пространство с содержащимися в нём точками, прямые, плоскости. Единицу измерения длин будем считать выбранной раз и навсегда.

Определение. **Вектором** называется упорядоченная пара точек. Первая точка есть **начало** вектора, а вторая – его **конец**. Если начало и конец вектора совпадают, то это **нулевой** вектор; в противном случае вектор можно представить направленным отрезком.

Расстояние между началом и концом вектора называют его **длиной**, **модулем**, либо **абсолютной величиной**.

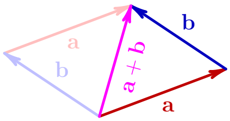
Пример. Многие физические величины являются векторами: перемещение, скорость, ускорение, сила...

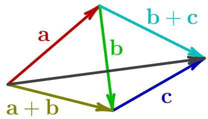
Векторы обозначают стрелочками над буквами ($\vec{a}$, на письме) либо жирным шрифтом ($\mathbf{a}$, в книгах и этом конспекте). Длину вектора $\mathbf{a}$ обозначают $|\mathbf{a}|$.

Начало вектора также называют его **точкой приложения**. Различают векторы **приложенные** и **свободные**. Свободными векторами пользуются, когда направление и длина существенны, а точка приложения безразлична. Это обычная ситуация в математике, где поэтому принято называть векторами именно свободные векторы, а при рассмотрении приложенного вектора — вводить уточнение оборотом «вектор, отложенный от точки». В этом курсе мы следуем математическому употреблению. Итак, вектор можно свободно переносить от одной точки приложения к другой, а полученный в результате переноса вектор считается *равным исходному*.

Операции с векторами

Сложение двух векторов $\mathbf{a}$ и $\mathbf{b}$ выполняется по **правилу треугольника**: откладываем $\mathbf{b}$ от конца $\mathbf{a}$, и тогда началом и концом $\mathbf{a}+\mathbf{b}$ будут начало $\mathbf{a}$ и конец $\mathbf{b}$. Иногда удобнее равносильное **правило параллелограмма**: $\mathbf{a}+\mathbf{b}$ есть диагональ параллелограмма со сторонами $\mathbf{a}$ и $\mathbf{b}$.





Утверждение. *Сложение векторов обладает алгебраическими свойствами, аналогичными свойствам обычного сложения чисел:*

(1) $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$ *для всех векторов* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$;
(2) $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$ *для всех векторов* $\mathbf{a}$, $\mathbf{b}$;
(3) *есть такой вектор* $\mathbf{0}$, *что* $\mathbf{a}+\mathbf{0}=\mathbf{a}$ *для каждого вектора* $\mathbf{a}$;
(4) *для каждого* $\mathbf{a}$ *есть такой вектор* $-\mathbf{a}$, *что* $\mathbf{a}+(-\mathbf{a})=\mathbf{0}$.

Доказательство. (1) **Ассоциативность** геометрически очевидна.
(2) **Коммутативность** геометрически очевидна.
(3) Это нулевой вектор.
(4) Вектор $-\mathbf{a}$ получается из $\mathbf{a}$ перестановкой начала и конца. □

Умножение вектора $\mathbf{a}$ на вещественное число λ выполняется умножением длины $\mathbf{a}$ на $|\lambda|$, а также изменением направления на противоположное в случае $\lambda<0$. Поэтому $1\mathbf{a}=\mathbf{a}$ и $(-1)\mathbf{a}=-\mathbf{a}$.

Утверждение. *Умножение векторов на числа обладает алгебраическими свойствами, аналогичными свойствам обычного умножения чисел, где все векторы и числа произвольны:*

(1) $1\mathbf{a}=\mathbf{a}$;
(2) $(\lambda\mu)\mathbf{a}=\lambda(\mu\mathbf{a})$;
(3) $(\lambda+\mu)\mathbf{a}=\lambda\mathbf{a}+\mu\mathbf{a}$;
(4) $\lambda(\mathbf{a}+\mathbf{b})=\lambda\mathbf{a}+\lambda\mathbf{b}$.

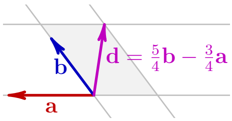
Доказательство. Всё проверяется непосредственно. □

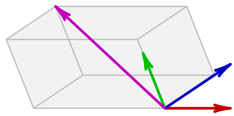
Базис и координаты

Два вектора, лежащие на одной прямой, то есть имеющие одинаковые или противоположные направления, называют **коллинеарными**. Три вектора, лежащие в одной плоскости, называют **компланарными**.

Утверждение. *Если векторы* $\mathbf{a}$, $\mathbf{b}$ *не коллинеарны, то каждый вектор* $\mathbf{d}$, *лежащий в одной плоскости с ними, представляется их комбинацией:* $\mathbf{d}=\alpha\mathbf{a}+\beta\mathbf{b}$.

Доказательство. Через начало и конец вектора $\mathbf{d}$ проведём по прямой, параллельной каждому из векторов $\mathbf{a}$ и $\mathbf{b}$. Стороны образованного параллелограмма равны $\alpha\mathbf{a}$ и $\beta\mathbf{b}$, где коэффициенты находятся как отношения длин с учётом направления. $\square$





Утверждение. *Если векторы* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ *не компланарны, то каждый вектор* $\mathbf{d}$ *представляется их комбинацией:* $\mathbf{d} = \alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}$.

Доказательство. Через начало и конец вектора $\mathbf{d}$ проведём по плоскости, параллельной каждой паре векторов из тройки $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$. Стороны образованного параллелепипеда равны $\alpha\mathbf{a}$, $\beta\mathbf{b}$ и $\gamma\mathbf{c}$. $\square$

Итак, выбрав три некомпланарных вектора, мы можем представить любой вектор пространства их комбинацией. Поэтому такую упорядоченную тройку $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ называют **базисом** пространства. Коэффициенты α, β, γ разложения $\mathbf{d} = \alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}$ по этому базису называют **координатами** вектора $\mathbf{d}$ относительно выбранного базиса.

Компланарность и линейная зависимость

Если векторы $\mathbf{a}$, $\mathbf{b}$ и $\mathbf{c}$ компланарны, то хотя бы один из них можно представить комбинацией двух других: $\mathbf{c} = \alpha\mathbf{a} + \beta\mathbf{b}$. Если при этом среди данных векторов нет коллинеарных, то каждый из них представляется комбинацией двух других.

Следующая равносильная формулировка избегает условностей и исключений: векторы $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ компланарны тогда и только тогда, когда найдутся такие числа α, β, γ, не все равные нулю, что комбинация $\alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}$ равна нулевому вектору. Такая формулировка легко распространяется на произвольное количество векторов и оказывается исключительно удобной при дальнейшем изучении линейной алгебры.

Определение. Множество $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ векторов называют **линейно независимым**, если

$$\alpha_1\mathbf{u}_1 + \ldots + \alpha_n\mathbf{u}_n = \mathbf{0} \implies \text{все } \alpha_i \text{ равны нулю.}$$

Линейно зависимым это множество называют в противном случае: если существуют такие коэффициенты $\alpha_1, \ldots, \alpha_n$, не все равные нулю, что

$$\alpha_1\mathbf{u}_1 + \ldots + \alpha_n\mathbf{u}_n = \mathbf{0}.$$

Ориентация пространства

Все базисы в пространстве делятся на два класса. Чтобы уловить различие между классами, представим, что первые два вектора, $\mathbf{e}_1$ и $\mathbf{e}_2$, уже выбраны. Проходящая через них плоскость делит пространство на две части, а третий вектор $\mathbf{e}_3$, отложенный от той же точки, что и первые два, должен лежать по одну сторону от этой плоскости. Повернём $\mathbf{e}_1$ в направлении $\mathbf{e}_2$ через меньший из образованных ими двух углов. С конца $\mathbf{e}_3$ это вращение будет казаться идущим либо по часовой стрелке, либо против. Базис $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ принято называть **левым** в первом случае и **правым** — во втором.

Для некоторых построений необходимо решить, базисы какого из двух классов считать положительными, а какого — отрицательными. Этот выбор называется выбором **ориентации** пространства, а когда он сделан, пространство называют **ориентированным**. Как правило, положительными считаются правые базисы.

В ориентированном пространстве понятие объёма дополняют более тонким понятием **ориентированного объёма**. Достаточно ввести его для параллелепипедов. Если базис положителен, то ориентированный объём построенного на нём параллелепипеда положителен и равен обычному объёму; для отрицательных базисов ориентированный объём отличается от обычного своим отрицательным знаком. Когда речь идёт о плоской фигуре, соответственно возникает **ориентированная площадь**.

1.2. Скалярное, смешанное и векторное произведения

В этом разделе геометрически определены названные операции и отмечены их простейшие свойства: поведение при перестановках аргументов, при умножении одного аргумента на число, а также **дистрибутивность** относительно сложения, ведущая к линейности по каждому аргументу.

Во всех трёх случаях несложно проверить геометрически первые два свойства. Доказательства дистрибутивности каждого произведения отсрочены в целях скорейшего достижения координатных формул для работы с этими произведениями.

Скалярное произведение

Определение. **Скалярным произведением** векторов $\mathbf{a}$ и $\mathbf{b}$, угол между которыми равен φ, называют число $|\mathbf{a}| \cdot |\mathbf{b}| \cos\varphi$.

длина!

В ходу несколько обозначений для скалярного произведения. Чаще всего это $\mathbf{a} \cdot \mathbf{b}$ либо $(\mathbf{a}, \mathbf{b})$.

Когда $\mathbf{a} = \mathbf{b}$, получаем связь скалярного квадрата вектора с его длиной: $|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}}$.

Утверждение. *Скалярное произведение обладает следующими свойствами, где все векторы и числа произвольны:*

(1) $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$*;*
(2) $(\lambda\mathbf{a}) \cdot \mathbf{b} = \lambda(\mathbf{a} \cdot \mathbf{b})$*;*
(3) $(\mathbf{a}_1 + \mathbf{a}_2) \cdot \mathbf{b} = \mathbf{a}_1 \cdot \mathbf{b} + \mathbf{a}_2 \cdot \mathbf{b}$*;*
(4) $(\lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2) \cdot \mathbf{b} = \lambda_1(\mathbf{a}_1 \cdot \mathbf{b}) + \lambda_2(\mathbf{a}_2 \cdot \mathbf{b})$*.*

Доказательство. (1) Формула для скалярного произведения не чувствительна к перестановке векторов.

(2) При умножении вектора $\mathbf{a}$ на $\lambda > 0$ на это же число умножится его длина. При умножении на $\lambda < 0$ изменится также направление, так что вместо угла φ будет угол $\pi - \varphi$. Поскольку $\cos(\pi - \varphi) = -\cos\varphi$, требуемое равенство сохранится.

(3) Доказательство дистрибутивности несколько длиннее. Отложим его до следующей лекции.

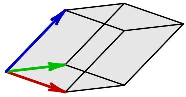
(4) Нужно применить (3) и (2). □

Последнее свойство в утверждении указывает на поведение скалярного произведения, когда в первый его аргумент подставлена линейная комбинация: оказывается возможным провести вычисления отдельно для каждого слагаемого, вынося коэффициенты. Это очень удобное и важное свойство называют **линейностью** скалярного произведения (по первому аргументу). Ввиду возможности переставлять аргументы, имеется линейность и по второму аргументу. Поэтому говорят, что скалярное произведение **билинейно**.

Смешанное произведение

Определение. **Смешанным произведением** трёх векторов $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ называют численное значение ориентированного объёма параллелепипеда на этих векторах. В компланарном случае оно равно нулю.

Смешанное произведение обозначают через $(\mathbf{a}, \mathbf{b}, \mathbf{c})$.



Утверждение. *Смешанное произведение обладает следующими свойствами, где все векторы и числа произвольны:*

(1) при перестановках меняется знак, то есть

$$(\mathbf{a}, \mathbf{b}, \mathbf{c}) = -(\mathbf{b}, \mathbf{a}, \mathbf{c}) = (\mathbf{b}, \mathbf{c}, \mathbf{a}) = -(\mathbf{c}, \mathbf{b}, \mathbf{a}) = (\mathbf{c}, \mathbf{a}, \mathbf{b}) = -(\mathbf{a}, \mathbf{c}, \mathbf{b});$$

(2) $(\lambda \mathbf{a}, \mathbf{b}, \mathbf{c}) = \lambda(\mathbf{a}, \mathbf{b}, \mathbf{c})$*;*

(3) $(\mathbf{a}_1 + \mathbf{a}_2, \mathbf{b}, \mathbf{c}) = (\mathbf{a}_1, \mathbf{b}, \mathbf{c}) + (\mathbf{a}_2, \mathbf{b}, \mathbf{c})$*;*

(4) $(\lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2, \mathbf{b}, \mathbf{c}) = \lambda_1(\mathbf{a}_1, \mathbf{b}, \mathbf{c}) + \lambda_2(\mathbf{a}_2, \mathbf{b}, \mathbf{c})$*.*

Доказательство. (1) Такие перестановки не меняют параллелепипед, но порядок векторов определяет «правость» или «левость» составленного из них базиса.

(2) При умножении одной стороны на λ объём параллелепипеда умножается на $|\lambda|$. Если $\lambda < 0$, то ориентация параллелепипеда меняется на противоположную.

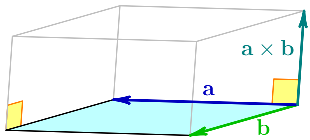
(3) Отложим доказательство и этой дистрибутивности.

(4) Нужно применить (3) и (2). □

Ввиду возможности переставлять аргументы, смешанное произведение линейно по всем трём аргументам (трилинейно).

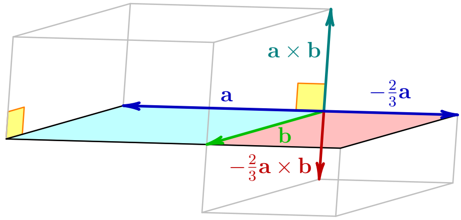
Векторное произведение

Определение. **Векторным произведением** векторов $\mathbf{a}$ и $\mathbf{b}$ называют вектор, обозначаемый через $\mathbf{a} \times \mathbf{b}$ либо $[\mathbf{a}, \mathbf{b}]$, длина и направление которого определены следующими правилами: если $\mathbf{a}$ и $\mathbf{b}$ коллинеарны, то $\mathbf{a} \times \mathbf{b} = \mathbf{0}$; иначе длина численно равна площади параллелограмма со сторонами $\mathbf{a}$, $\mathbf{b}$, то есть $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| \cdot |\mathbf{b}| \sin\varphi$, а направление перпендикулярно обоим векторам $\mathbf{a}$ и $\mathbf{b}$ и таково, что $\mathbf{a}$, $\mathbf{b}$, $\mathbf{a} \times \mathbf{b}$ является положительным базисом.



Утверждение. *Векторное произведение обладает следующими свойствами, где все векторы и числа произвольны:*

(1) $\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}$*;*
(2) $(\lambda\mathbf{a})\times\mathbf{b} = \lambda(\mathbf{a}\times\mathbf{b})$*;*
(3) $(\mathbf{a}_1+\mathbf{a}_2)\times\mathbf{b} = \mathbf{a}_1\times\mathbf{b}+\mathbf{a}_2\times\mathbf{b}$*;*
(4) $(\lambda_1\mathbf{a}_1+\lambda_2\mathbf{a}_2)\times\mathbf{b} = \lambda_1(\mathbf{a}_1\times\mathbf{b})+\lambda_2(\mathbf{a}_2\times\mathbf{b})$*.*



Доказательство. (1) Вектор $\mathbf{a}\times\mathbf{b}$ обладает всеми свойствами, характеризующими произведение $\mathbf{b}\times\mathbf{a}$, за исключением того, что $\mathbf{b},\mathbf{a},\mathbf{a}\times\mathbf{b}$ образуют отрицательный базис. Поэтому для $\mathbf{b}\times\mathbf{a}$ не остаётся другой возможности, кроме как равняться $-\mathbf{a}\times\mathbf{b}$.

(2) При $\lambda > 0$ наблюдаются лишь изменение длины и соответствующее изменение площади, а при $\lambda < 0$ наблюдается также изменение направлений первого сомножителя и всего произведения на противоположные.

(3) Отложим.

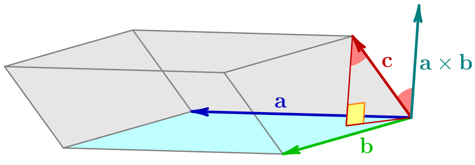
(4) Нужно применить (3) и (2). □

Утверждение. *Смешанное произведение выражается через скалярное и векторное:*

$$(\mathbf{a},\mathbf{b},\mathbf{c}) = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}.$$

рисунок?

Этим и объясняется термин «смешанное».



Доказательство. В некомпланарном случае, представив объём параллелепипеда как произведение площади основания на высоту:

$$|\mathbf{a}\times\mathbf{b}|\cdot|\mathbf{c}|\cos\theta,$$

где θ есть угол между $\mathbf{a}\times\mathbf{b}$ и $\mathbf{c}$, получаем совпадение абсолютных величин $|(\mathbf{a},\mathbf{b},\mathbf{c})| = |(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}|$. А знаки совпадут, потому что векторы

$\mathbf{a} \times \mathbf{b}$ и $\mathbf{c}$ направлены в одну сторону от плоскости, содержащей $\mathbf{a}$ и $\mathbf{b}$, тогда и только тогда, когда базис $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ положителен.

В компланарном случае получим $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \mathbf{0}$. □

Двойное векторное произведение

Легко убедиться, что $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ не обязано равняться $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$. Лекция 2 09.09.19

Утверждение. *Двойное векторное произведение можно выразить через скалярные по формуле*

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}),$$

известной как «бац минус цаб».

Доказательство. Когда мы выведем формулы для вычисления произведений в координатах, это равенство можно просто проверить, в чём и состоит стандартный способ его доказательства. Чисто геометрическое рассуждение будет дано ниже, но оно представляет скорее спортивный интерес. □

Следствие. *Векторное произведение удовлетворяет* **тождеству Якоби**

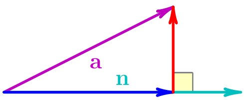
$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) + \mathbf{b} \times (\mathbf{c} \times \mathbf{a}) + \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) = \mathbf{0}.$$

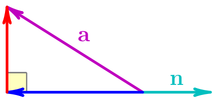
Доказательство. Шесть слагаемых, получаемых при раскрытии двойных векторных произведений, попарно сокращаются. □

Проекции вектора

Следствие. *Если* $|\mathbf{n}| = 1$, *то* $\mathbf{a} = (\mathbf{n} \cdot \mathbf{a})\mathbf{n} + (\mathbf{n} \times \mathbf{a}) \times \mathbf{n}$.

Доказательство. $(\mathbf{n} \times \mathbf{a}) \times \mathbf{n} = -\mathbf{n} \times (\mathbf{n} \times \mathbf{a}) = -\mathbf{n}(\mathbf{n} \cdot \mathbf{a}) + \mathbf{a}(\mathbf{n} \cdot \mathbf{n})$. □





Этим фактически указан способ разложить всякий вектор $\mathbf{a}$ в сумму $\mathbf{a} = \mathbf{a}^{\parallel \mathbf{n}} + \mathbf{a}^{\perp \mathbf{n}}$, где первое слагаемое, называемое **проекцией** $\mathbf{a}$ на $\mathbf{n}$, коллинеарно $\mathbf{n}$, а второе ему перпендикулярно. Из доказательства следует также забавная формула проектирования

$$\mathbf{a}^{\parallel \mathbf{n}} = \frac{\mathbf{a} \cdot \mathbf{n}}{\mathbf{n} \cdot \mathbf{n}} \mathbf{n}$$

на любой ненулевой вектор $\mathbf{n}$, не обязательно единичный. Дробь здесь представляет собой отношение скаляров, и сокращения в ней невозможны (вопреки желаниям некоторых студентов).

Лемма. *Если* $\mathbf{n} \neq \mathbf{0}$, *то* $\mathbf{a} \cdot \mathbf{n} = \mathbf{a}^{\|\mathbf{n}} \cdot \mathbf{n}$ *и* $\mathbf{a} \times \mathbf{n} = \mathbf{a}^{\perp\mathbf{n}} \times \mathbf{n}$ *для каждого вектора* $\mathbf{a}$.

Доказательство. Скалярные произведения равны $|\mathbf{a}||\mathbf{n}|\cos\varphi$. Векторные произведения дают равные и одинаково ориентированные площади параллелограмма и прямоугольника. □

Лемма. *Если* $\mathbf{n} \neq \mathbf{0}$, *то*

$$(\mathbf{a}+\mathbf{b})^{\|\mathbf{n}} = \mathbf{a}^{\|\mathbf{n}} + \mathbf{b}^{\|\mathbf{n}}, \quad (\mathbf{a}+\mathbf{b})^{\perp\mathbf{n}} = \mathbf{a}^{\perp\mathbf{n}} + \mathbf{b}^{\perp\mathbf{n}}$$

для любых векторов $\mathbf{a}$ *и* $\mathbf{b}$.

Доказательство. Рассмотрим треугольную призму. Коллинеарные $\mathbf{n}$ составляющие складываются вдоль её боковых рёбер, а перпендикулярные — в плоскостях её оснований. □

1.3. Определители второго и третьего порядков

Мотивирующие вычисления в координатах

Выберем в пространстве базис $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ и будем представлять каждый вектор $\mathbf{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$ тройкой его координат $[a_1, a_2, a_3]$, употребляя ту же букву, но «обезжиренную».

Утверждение. *Сложение векторов и умножение вектора на число выполняются покоординатно.*

Доказательство. Имеются в виду формулы

$$\lambda[a_1, a_2, a_3] = [\lambda a_1, \lambda a_2, \lambda a_3],$$
$$[a_1, a_2, a_3] + [b_1, b_2, b_3] = [a_1 + b_1, a_2 + b_2, a_3 + b_3],$$

всего лишь выражающие равенства

$$\lambda\mathbf{a} = (\lambda a_1)\mathbf{e}_1 + (\lambda a_2)\mathbf{e}_2 + (\lambda a_3)\mathbf{e}_3,$$
$$\mathbf{a}+\mathbf{b} = (a_1+b_1)\mathbf{e}_1 + (a_2+b_2)\mathbf{e}_2 + (a_3+b_3)\mathbf{e}_3$$

в координатах относительно базиса $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. □

Рассмотрим теперь скалярное произведение векторов **a** и **b**. Пользуясь билинейностью, раскрываем скобки:

$$\begin{aligned}\mathbf{a}\cdot\mathbf{b} &= (a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3)\cdot(b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + b_3\mathbf{e}_3)\\ &= a_1b_1\,\mathbf{e}_1\cdot\mathbf{e}_1 + a_1b_2\,\mathbf{e}_1\cdot\mathbf{e}_2 + a_1b_3\,\mathbf{e}_1\cdot\mathbf{e}_3\\ &+ a_2b_1\,\mathbf{e}_2\cdot\mathbf{e}_1 + a_2b_2\,\mathbf{e}_2\cdot\mathbf{e}_2 + a_2b_3\,\mathbf{e}_2\cdot\mathbf{e}_3\\ &+ a_3b_1\,\mathbf{e}_3\cdot\mathbf{e}_1 + a_3b_2\,\mathbf{e}_3\cdot\mathbf{e}_2 + a_3b_3\,\mathbf{e}_3\cdot\mathbf{e}_3.\end{aligned}$$

Видно, что для вычисления произвольных скалярных произведений в выбранном базисе достаточно знать скалярные произведения базисных векторов между собой. При этом особенно удобны базисы, удовлетворяющие условию $\mathbf{e}_i\cdot\mathbf{e}_j = \delta_{ij}$, где

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

есть **символ Кронекера**; проще говоря, в этом случае базисные векторы единичны и перпендикулярны друг другу. Такие базисы называют **ортонормированными**, сокращённо ОНБ. В ортонормированном базисе координатное вычисление скалярных произведений упрощается до

$$\mathbf{a}\cdot\mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3.$$

Аналогичным образом раскроем векторное произведение:

$$\begin{aligned}\mathbf{a}\times\mathbf{b} &= (a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3)\times(b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + b_3\mathbf{e}_3)\\ &= a_1b_1\,\mathbf{e}_1\times\mathbf{e}_1 + a_1b_2\,\mathbf{e}_1\times\mathbf{e}_2 + a_1b_3\,\mathbf{e}_1\times\mathbf{e}_3\\ &+ a_2b_1\,\mathbf{e}_2\times\mathbf{e}_1 + a_2b_2\,\mathbf{e}_2\times\mathbf{e}_2 + a_2b_3\,\mathbf{e}_2\times\mathbf{e}_3\\ &+ a_3b_1\,\mathbf{e}_3\times\mathbf{e}_1 + a_3b_2\,\mathbf{e}_3\times\mathbf{e}_2 + a_3b_3\,\mathbf{e}_3\times\mathbf{e}_3.\end{aligned}$$

Здесь по свойству $\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}$ три слагаемых равны нулю, а оставшиеся можно сгруппировать попарно:

$$\begin{aligned}\mathbf{a}\times\mathbf{b} &= (a_1b_2 - a_2b_1)\,\mathbf{e}_1\times\mathbf{e}_2\\ &+ (a_3b_1 - a_1b_3)\,\mathbf{e}_3\times\mathbf{e}_1\\ &+ (a_2b_3 - a_3b_2)\,\mathbf{e}_2\times\mathbf{e}_3.\end{aligned}$$

Обычно вычисление векторных произведений ведётся в правом ОНБ. В этом случае

$$\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3, \quad \mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2, \quad \mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1.$$

Взяв смешанное произведение трёх векторов, разложенных по базису, при раскрытии всех скобок мы получаем 27 слагаемых со смешанными произведениями базисных векторов, но из них 21 содержат повторения аргументов и потому равны нулю. В остальных слагаемых переставим аргументы, приводя их «в порядок» $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ с учётом изменения знаков:

$$\begin{aligned}(\mathbf{a}, \mathbf{b}, \mathbf{c}) &= a_1 b_2 c_3\,(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) + a_2 b_1 c_3\,(\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3)\\ &+ a_3 b_1 c_2\,(\mathbf{e}_3, \mathbf{e}_1, \mathbf{e}_2) + a_1 b_3 c_2\,(\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_2)\\ &+ a_2 b_3 c_1\,(\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1) + a_3 b_2 c_1\,(\mathbf{e}_3, \mathbf{e}_2, \mathbf{e}_1)\\ &= D\,(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3),\end{aligned}$$

где

$$D = a_1 b_2 c_3 - a_2 b_1 c_3 + a_3 b_1 c_2 - a_1 b_3 c_2 + a_2 b_3 c_1 - a_3 b_2 c_1.$$

Для правого ОНБ имеем $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = 1$ и получаем $(\mathbf{a}, \mathbf{b}, \mathbf{c}) = D$.

В целях более компактной записи координатных выражений векторного и смешанного произведений введём понятие **определителя**. Позже определители будут появляться во многих других ситуациях.

Полное раскрытие

Для непосредственных целей (мы же сейчас только в начале курса) нужны определители второго и третьего порядков, где порядок означает размер матрицы. Второй порядок:

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1;$$

третий порядок:

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 b_2 c_3 - a_2 b_1 c_3 + a_3 b_1 c_2 - a_1 b_3 c_2 + a_2 b_3 c_1 - a_3 b_2 c_1,$$

или в точности число D, возникшее в конце предыдущего пункта. Это формулы полного раскрытия определителя, с которыми выражения для смешанного и векторного произведений в правом ОНБ $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$

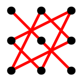
принимают вид

$$(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}; \qquad \mathbf{a} \times \mathbf{b} = \begin{vmatrix} a_1 & b_1 & \mathbf{e}_1 \\ a_2 & b_2 & \mathbf{e}_2 \\ a_3 & b_3 & \mathbf{e}_3 \end{vmatrix}.$$

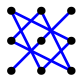
Второй определитель здесь *символический*, поскольку третий столбец заполнен векторами, а не числами.

Есть чёткие правила, по которым составлены формулы полного раскрытия. Не только в указанных случаях, но и для произвольного порядка n, раскрытие является суммой со знаками всевозможных слагаемых, равных произведению n компонент матрицы, взятых по одному из каждой строки и каждого столбца. Всего слагаемых $n!$. Детали общего правила выбора знаков мы разберём значительно позже, но для третьего (и только!) порядка отметим простую картинку:

$$(+): \quad \begin{matrix} a_1 & {\color{gray}b_1} & {\color{gray}c_1} \\ {\color{gray}a_2} & b_2 & {\color{gray}c_2} \\ {\color{gray}a_3} & {\color{gray}b_3} & c_3 \end{matrix} \qquad \begin{matrix} {\color{gray}a_1} & b_1 & {\color{gray}c_1} \\ {\color{gray}a_2} & {\color{gray}b_2} & c_2 \\ a_3 & {\color{gray}b_3} & {\color{gray}c_3} \end{matrix} \qquad \begin{matrix} {\color{gray}a_1} & {\color{gray}b_1} & c_1 \\ a_2 & {\color{gray}b_2} & {\color{gray}c_2} \\ {\color{gray}a_3} & b_3 & {\color{gray}c_3} \end{matrix}$$



$$(-): \quad \begin{matrix} {\color{gray}a_1} & b_1 & {\color{gray}c_1} \\ a_2 & {\color{gray}b_2} & {\color{gray}c_2} \\ {\color{gray}a_3} & {\color{gray}b_3} & c_3 \end{matrix} \qquad \begin{matrix} a_1 & {\color{gray}b_1} & {\color{gray}c_1} \\ {\color{gray}a_2} & {\color{gray}b_2} & c_2 \\ {\color{gray}a_3} & b_3 & {\color{gray}c_3} \end{matrix} \qquad \begin{matrix} {\color{gray}a_1} & {\color{gray}b_1} & c_1 \\ {\color{gray}a_2} & b_2 & {\color{gray}c_2} \\ a_3 & {\color{gray}b_3} & {\color{gray}c_3} \end{matrix}$$



Раскрытие по столбцу или строке

Есть много способов выразить определитель третьего порядка через определители второго порядка. Например, группируя слагаемые с одинаковыми c_i, получим

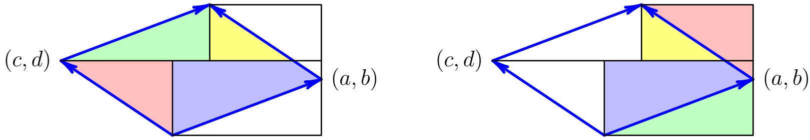
$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} c_1 - \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} c_2 + \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} c_3.$$

Здесь стоит обратить внимание на минус в правой части. Это формула раскрытия определителя третьего порядка по третьему столбцу. Группируя слагаемые иначе, можно получить формулы раскрытия по другим столбцам, а также по строкам.

Особенно употребительно приложение к векторному произведению:

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} a_1 & b_1 & \mathbf{e}_1 \\ a_2 & b_2 & \mathbf{e}_2 \\ a_3 & b_3 & \mathbf{e}_3 \end{vmatrix} = \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \mathbf{e}_3.$$

Отметим геометрический смысл этой формулы. Прежде всего, определитель второго порядка также выражает геометрическую величину: ориентированную площадь параллелограмма ($ad - bc$ на рисунке).



Выходит, каждый из трёх таких определителей в нашей формуле вычисляет ориентированную площадь проекции параллелограмма на векторах $\mathbf{a}$ и $\mathbf{b}$ на три координатные плоскости. Согласование ориентаций вносит минус перед вторым слагаемым.

Формула раскрытия определителя по столбцу или строке обобщается на определители любого порядка n. Множитель при каждом элементе раскрытия является определителем порядка $n-1$, получаемым из исходного вычёркиванием того столбца и той строки, где находится этот элемент. Ещё есть чередующиеся знаки плюс и минус; тут правило такое: при элементе в левом верхнем углу всегда нужен плюс, а в остальном это шахматный порядок (детали позже на семинарах).

Вычищение определителя

Вычислять значение определителя посредством его раскрытия редко бывает практичным само по себе, но становится гораздо проще и эффективнее, когда определитель удаётся предварительно «вычистить». Процесс вычищения основан на свойствах ориентированных объёмов. Прибавляя к одному из векторов другой, мы не изменим ориентированный объём:

$$(\mathbf{a}+\mathbf{c}, \mathbf{b}, \mathbf{c}) = (\mathbf{a}, \mathbf{b}, \mathbf{c}) + (\mathbf{c}, \mathbf{b}, \mathbf{c}) = (\mathbf{a}, \mathbf{b}, \mathbf{c}).$$

С тем же успехом можно прибавлять к $\mathbf{a}$ или к $\mathbf{b}$ вектор $\lambda\mathbf{c}$. При вычислении в координатах выгодно воспользоваться этим, чтобы заменить некоторые элементы в определителе нулями.

Примеры. Вычислим значение определителя:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 1 \\ 4 & 5 & 1 \\ 7 & 8 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 4 & 1 & 1 \\ 7 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 0 \\ 4 & 1 & 0 \\ 7 & 1 & 0 \end{vmatrix} = 0.$$

Здесь мы на первом шаге вычитаем второй столбец из третьего, а на втором — первый из второго. После этого мы видим два одинаковых столбца, что сразу указывает на зануление определителя ввиду его геометрического смысла; помимо того, можно совершить третье вычитание и получить нулевой столбец, раскрытие по которому даёт нуль в ответе.

Вычислим значение определителя:

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 0 & 2 & 6 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 2 & 6 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{vmatrix} = 2.$$

Для разнообразия, здесь применены операции со строками, а не со столбцами; для вычищения определителей они оказываются равноправны. У последнего определителя лишь одно из шести слагаемых полного раскрытия избегает нулевых сомножителей: оно составлено из элементов на **главной диагонали**, поэтому ответ легко считывается.

1.4. Отложенные геометрические доказательства

Дистрибутивности

Установим формулу $(\mathbf{a}+\mathbf{b})\cdot\mathbf{c} = \mathbf{a}\cdot\mathbf{c}+\mathbf{b}\cdot\mathbf{c}$. Если $\mathbf{c}=\mathbf{0}$, то обе части нулевые. Иначе, леммы позволяют заменить векторы $\mathbf{a}$ и $\mathbf{b}$ на их составляющие коллинеарные $\mathbf{c}$ и свести требуемое к равенству

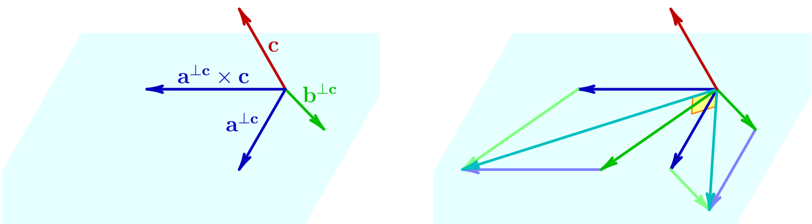
$$(\mathbf{a}+\mathbf{b})^{\parallel\mathbf{c}}\cdot\mathbf{c} = (\mathbf{a}^{\parallel\mathbf{c}}+\mathbf{b}^{\parallel\mathbf{c}})\cdot\mathbf{c} \overset{?}{=} \mathbf{a}^{\parallel\mathbf{c}}\cdot\mathbf{c}+\mathbf{b}^{\parallel\mathbf{c}}\cdot\mathbf{c},$$

проверяемому непосредственно, на одной прямой.

Установим формулу $(\mathbf{a}+\mathbf{b})\times\mathbf{c} = \mathbf{a}\times\mathbf{c}+\mathbf{b}\times\mathbf{c}$. Если $\mathbf{c}=\mathbf{0}$, то обе части нулевые. Иначе, леммы позволяют заменить векторы $\mathbf{a}$ и $\mathbf{b}$ на их составляющие перпендикулярные $\mathbf{c}$ и свести требуемое к равенству

$$(\mathbf{a}+\mathbf{b})^{\perp\mathbf{c}}\times\mathbf{c} = (\mathbf{a}^{\perp\mathbf{c}}+\mathbf{b}^{\perp\mathbf{c}})\times\mathbf{c} \overset{?}{=} \mathbf{a}^{\perp\mathbf{c}}\times\mathbf{c}+\mathbf{b}^{\perp\mathbf{c}}\times\mathbf{c}.$$

Параллелограмм на векторах $\mathbf{a}^{\perp\mathbf{c}}$ и $\mathbf{b}^{\perp\mathbf{c}}$ лежит в плоскости, перпендикулярной $\mathbf{c}$. При векторном умножении на $\mathbf{c}$ вся эта плоскость поворачивается на прямой угол и масштабируется с коэффициентом подобия $|\mathbf{c}|$.



Дистрибутивность смешанного произведения по каждому из аргументов сводится к дистрибутивности скалярного и векторного произведений. Например, по второму аргументу:

$$\begin{aligned}(\mathbf{a}, \mathbf{b}_1 + \mathbf{b}_2, \mathbf{c}) &= (\mathbf{a} \times (\mathbf{b}_1 + \mathbf{b}_2)) \cdot \mathbf{c} \\ &= (\mathbf{a} \times \mathbf{b}_1 + \mathbf{a} \times \mathbf{b}_2) \cdot \mathbf{c} \\ &= (\mathbf{a} \times \mathbf{b}_1) \cdot \mathbf{c} + (\mathbf{a} \times \mathbf{b}_2) \cdot \mathbf{c} \\ &= (\mathbf{a}, \mathbf{b}_1, \mathbf{c}) + (\mathbf{a}, \mathbf{b}_2, \mathbf{c}).\end{aligned}$$

Бац минус цаб

Установим теперь формулу $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$.

Предположим сначала, что $\mathbf{b} \times \mathbf{c} = \mathbf{0}$. Тогда сомножители коллинеарны: $\beta\mathbf{b} + \gamma\mathbf{c} = \mathbf{0}$. Выразив отсюда один из векторов и подставив в правую часть искомого равенства, найдём $\mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}) = \mathbf{0}$, так что в этом случае формула верна.

Предположим теперь, что $\mathbf{b} \times \mathbf{c} \neq \mathbf{0}$. Тогда по определению векторного произведения векторы $\mathbf{b} \times \mathbf{c}$, $\mathbf{b}$, $\mathbf{c}$ образуют базис пространства. Разложим $\mathbf{a} = \alpha\mathbf{b} \times \mathbf{c} + \beta\mathbf{b} + \gamma\mathbf{c}$ и проверим формулу отдельно для каждой компоненты, пользуясь дистрибутивностью всех слагаемых. Значит, эту проверку достаточно сделать в трёх «базисных» случаях: $\mathbf{a} = \mathbf{b} \times \mathbf{c}$, $\mathbf{a} = \mathbf{b}$ и $\mathbf{a} = \mathbf{c}$.

При $\mathbf{a} = \mathbf{b} \times \mathbf{c}$ все слагаемые равны нулю ввиду повторения векторов. При $\mathbf{a} = \mathbf{b}$ перепишем искомое равенство $\mathbf{b} \times (\mathbf{b} \times \mathbf{c}) \stackrel{?}{=} \mathbf{b}(\mathbf{b} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{b} \cdot \mathbf{b})$ в виде $(\mathbf{b} \cdot \mathbf{b})\mathbf{c} \stackrel{?}{=} (\mathbf{b} \cdot \mathbf{c})\mathbf{b} + (\mathbf{b} \times \mathbf{c}) \times \mathbf{b}$. Разделим на $\mathbf{b} \cdot \mathbf{b} \neq 0$, введём единичный вектор $\mathbf{n} = \mathbf{b}/|\mathbf{b}|$ и увидим уже давно знакомую формулу $\mathbf{c} = (\mathbf{n} \cdot \mathbf{c})\mathbf{n} + (\mathbf{n} \times \mathbf{c}) \times \mathbf{n}$ для разложения $\mathbf{c} = \mathbf{c}^{\parallel \mathbf{n}} + \mathbf{c}^{\perp \mathbf{n}}$. При $\mathbf{a} = \mathbf{c}$ проверка аналогична, только $\mathbf{b}$ и $\mathbf{c}$ меняются ролями.

1.5. Задание прямой и плоскости

Фигуры и уравнения

Лекция 2+ 09.09.19

Выберем в пространстве точку O в качестве начала отсчёта. Тогда каждой точке A пространства соответствует её **радиус-вектор** $\mathbf{r}(A)$: это вектор с началом O и концом A. Получаем взаимно-однозначное соответствие между точками и векторами. Подразумевая выбранное начало отсчёта, бывает удобно говорить о радиус-векторе, имея в виду соответствующую точку, либо наоборот.

Уравнениями геометрической фигуры в выбранной системе координат называют условия (в виде равенств) на координаты произвольной точки, которые выполнены тогда и только тогда, когда точка принадлежит фигуре. Система координат выбирается прежде всего, но мы не будем напоминать про этот этап в каждой ситуации.

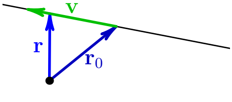
Уравнения прямой либо плоскости записывают в различных формах. Выделим четыре способа: параметрический, нормальный, общий, поточечный; при этом первые две формы используют радиус-векторы, а последние две — координаты. В векторных уравнениях обеих форм, будь то на плоскости или в пространстве, обозначим через $\mathbf{r}$ радиус-вектор произвольной (неизвестной) точки. Остальные участвующие в уравнениях векторы предполагаем известными.

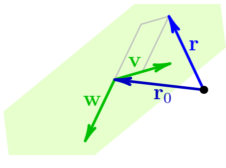
Параметрическое задание

Возьмём на данной прямой ℓ точку и обозначим её радиус-вектор через $\mathbf{r}_0$. Далее, возьмём параллельный ℓ вектор $\mathbf{v}$. Тогда для каждого вещественного числа t вектор $\mathbf{r}_0 + t\mathbf{v}$ есть радиус-вектор некоторой точки на прямой ℓ. Поэтому

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{v},$$

где t — переменный **параметр**, называют **параметрическим заданием** прямой. Постоянный вектор $\mathbf{v}$ называют **направляющим вектором** этой прямой, а точку (соответствующую) $\mathbf{r}_0$ — её **начальной точкой**. Когда t пробегает все вещественные числа, конец радиус-вектора $\mathbf{r}$ пробегает все точки прямой ℓ. Физически можно трактовать $\mathbf{r}_0$ и $\mathbf{v}$ как начальное положение и скорость «материальной точки». Всякая прямая имеет бесконечно много параметрических заданий с разными $\mathbf{r}_0$ и $\mathbf{v}$.





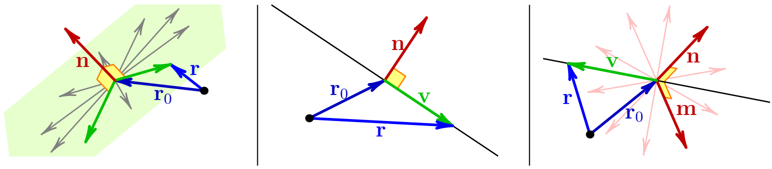
Возьмём на данной плоскости Π точку и обозначим её радиус-вектор через $\mathbf{r}_0$. Далее, возьмём в Π два неколлинеарных вектора $\mathbf{v}$, $\mathbf{w}$. Тогда для любых вещественных чисел t, s вектор $t\mathbf{v} + s\mathbf{w}$ лежит в этой же плоскости Π, причём все векторы в ней имеют такое представление, а $\mathbf{r}_0 + t\mathbf{v} + s\mathbf{w}$ есть радиус-вектор точки на Π. Поэтому

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{v} + s\mathbf{w},$$

где t и s — переменные параметры, называют параметрическим заданием плоскости. Постоянные векторы $\mathbf{v}$ и $\mathbf{w}$ называют направляющими векторами этой плоскости, а точку $\mathbf{r}_0$ — её начальной точкой. Когда t и s независимо пробегают все вещественные числа, конец радиус-вектора $\mathbf{r}$ пробегает все точки плоскости Π.

Нормальные уравнения

Нормалью к прямой или к плоскости называют вектор, ортогональный каждому вектору, лежащему в этой прямой/плоскости. **Нормальные уравнения** $\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0$ в случаях прямой на плоскости и плоскости в пространстве отражают тот факт, что одна точка $\mathbf{r}_0$ на прямой/плоскости и нормаль $\mathbf{n}$ к ней определяют прямую/плоскость. Поскольку здесь $\mathbf{n}$ и $\mathbf{r}_0$ известны, можно подставить их скалярное произведение в уравнение и получить его вариант $\mathbf{n} \cdot \mathbf{r} = D$.



Лекция 3
16.09.19

Для прямой в пространстве редко говорят о нормальном уравнении, хотя можно выбрать два *неколлинеарных* нормальных вектора $\mathbf{n}$, $\mathbf{m}$ и записать систему

$$\begin{cases} \mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0, \\ \mathbf{m} \cdot (\mathbf{r} - \mathbf{r}_0) = 0. \end{cases}$$

Она равносильна уравнению $\mathbf{v} \times (\mathbf{r} - \mathbf{r}_0) = \mathbf{0}$, в котором $\mathbf{v} = \mathbf{n} \times \mathbf{m}$ является направляющим вектором прямой. Здесь тоже можно подставить $\mathbf{a} = \mathbf{v} \times \mathbf{r}_0$ и получить вариант $\mathbf{v} \times \mathbf{r} = \mathbf{a}$. Однако одно уравнение с векторным произведением лишь на первый взгляд экономичнее системы из двух скалярных, ведь в координатах оно превращается в три уравнения, а значит, одно из них всегда оказывается лишним.

Общие уравнения

Теорема. *В координатах относительно произвольного базиса*

(1) в пространстве всякая плоскость задаётся уравнением вида

$$Ax + By + Cz + D = 0;$$

(2) на плоскости всякая прямая задаётся уравнением вида

$$Ax + By + D = 0;$$

(3) в пространстве всякая прямая задаётся системой вида

$$\begin{cases} A_1x + B_1y + C_1z + D_1 = 0, \\ A_2x + B_2y + C_2z + D_2 = 0. \end{cases}$$

Их называют **общими** уравнениями прямой или плоскости. За сам вид — без произведений и степеней переменных — такие уравнения называют **линейными**.

Доказательство. (1) Подставим разложение $\mathbf{r} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$ неизвестного радиус-вектора в нормальное уравнение $\mathbf{n}\cdot(\mathbf{r}-\mathbf{r}_0) = 0$ и сразу получим выражения для коэффициентов искомого общего уравнения:

$$(\mathbf{n}\cdot\mathbf{e}_1)x + (\mathbf{n}\cdot\mathbf{e}_2)y + (\mathbf{n}\cdot\mathbf{e}_3)z + (-\mathbf{n}\cdot\mathbf{r}_0) = 0.$$

(2) Аналогично первому случаю, но $\mathbf{r} = x\mathbf{e}_1 + y\mathbf{e}_2$.

(3) Аналогично первому случаю, но нормалей две. □

Теорема. *В координатах относительно произвольного базиса каждое линейное уравнение, в котором хотя бы один коэффициент при неизвестных ненулевой, задаёт плоскость в пространстве либо прямую на плоскости.*

В случае системы из двух линейных уравнений условие, при котором она задаёт прямую в пространстве, более сложное: это условие, что две плоскости пересекаются по прямой.

Доказательство. Теперь, наоборот, нужно по линейному уравнению $Ax + By + Cz + D = 0$ найти нормаль к гипотетически задаваемой им плоскости. Из-за произвольности базиса $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ ответ менее очевиден, чем был бы в ОНБ. Поэтому возьмём вектор

$$\mathbf{n} = \frac{A\mathbf{e}_2\times\mathbf{e}_3 + B\mathbf{e}_3\times\mathbf{e}_1 + C\mathbf{e}_1\times\mathbf{e}_2}{(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)}.$$

Для $\mathbf{r} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$ получим $\mathbf{n}\cdot\mathbf{r} = Ax + By + Cz$, ибо в числителе образуются 9 смешанных произведений, из которых 6 содержат повторяющиеся векторы, а остальные равны знаменателю. Далее, возьмём $\mathbf{r}_0 = -D\frac{\mathbf{n}}{\mathbf{n}\cdot\mathbf{n}}$; тогда $D = -\mathbf{n}\cdot\mathbf{r}_0$, так что исходное уравнение равносильно нормальному уравнению плоскости $\mathbf{n}\cdot(\mathbf{r}-\mathbf{r}_0) = 0$.

К уравнению $Ax + By + D = 0$ постараемся применить тот же трюк, считая, что оно имеет вид $Ax + By + Cz + D = 0$, но $C = 0$. Поскольку ситуация плоская и третьего базисного вектора нет, можно выйти в пространство и выбрать $\mathbf{e}_3$ ортогональным к $\mathbf{e}_1$ и $\mathbf{e}_2$. Тогда загадочная формула даст $\mathbf{n} = n_1\mathbf{e}_1 + n_2\mathbf{e}_2 + 0\mathbf{e}_3$ и о пространстве можно «забыть». Радиус-вектор начальной точки находится так же: $\mathbf{r}_0 = -D\frac{\mathbf{n}}{\mathbf{n}\cdot\mathbf{n}}$. □

Уравнение в отрезках

написать!

Таблица способов задания

Перед тем, как заняться уравнениями по точкам, соберём в таблицу наиболее употребительные способы задания прямых и плоскостей и некоторые дополнительные сведения:

(1) общее уравнение;
(2) нормальное уравнение;
(3) коразмерность;
(4) размерность;
(5) параметрическое уравнение;
(6) уравнение по точкам обычное;
(7) уравнение по точкам через определитель/-и;
(8) уравнение по точкам через ранг.

д строками
таблицы

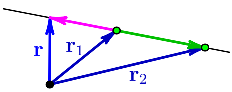
$\triangle$	$\Diamond$	∇	
плоскость в пространстве	прямая на плоскости	прямая в пространстве	
$Ax+By+Cz+D=0$	$Ax+By+C=0$	$\begin{cases} A_1x+B_1y+C_1z+D_1=0 \\ A_2x+B_2y+C_2z+D_2=0 \end{cases}$	(1)
$\mathbf{n}\cdot(\mathbf{r}-\mathbf{r}_0)=0$	$\mathbf{n}\cdot(\mathbf{r}-\mathbf{r}_0)=0$	$\begin{cases} \mathbf{n}\cdot(\mathbf{r}-\mathbf{r}_0)=0 \\ \mathbf{m}\cdot(\mathbf{r}-\mathbf{r}_0)=0 \end{cases}$	(2)
1	1	2	(3)
2	1	1	(4)
$\mathbf{r}=\mathbf{r}_0+t\mathbf{v}+s\mathbf{w}$	$\mathbf{r}=\mathbf{r}_0+t\mathbf{v}$	$\mathbf{r}=\mathbf{r}_0+t\mathbf{v}$	(5)
$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \end{vmatrix}=0$	$\frac{x-x_1}{x_2-x_1}=\frac{y-y_1}{y_2-y_1}$	$\frac{x-x_1}{x_2-x_1}=\frac{y-y_1}{y_2-y_1}=\frac{z-z_1}{z_2-z_1}$	(6)
$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix}=0$	$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix}=0$	$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix}=\begin{vmatrix} y & z & 1 \\ y_1 & z_1 & 1 \\ y_2 & z_2 & 1 \end{vmatrix}=0$	(7)
$\operatorname{rk}\begin{bmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{bmatrix}=3$	$\operatorname{rk}\begin{bmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{bmatrix}=2$	$\operatorname{rk}\begin{bmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \end{bmatrix}=2$	(8)

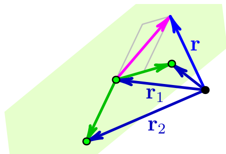
Значки ◊, △, и ▽ в таком контексте совершенно нестандартны, но в этой главе я нахожу удобным пользоваться ими для ссылок.

С видом уравнений, записанных в остальных строках таблицы, тесно связаны числа в строках (3) и (4): это соответственно **размерность** и **коразмерность** фигуры. Размерность равна количеству параметров, а коразмерность равна количеству общих уравнений и дополняет размерность до **объемлющей** размерности.

Уравнения по точкам

Через две различные фиксированные точки на плоскости или в пространстве, с радиус-векторами $\mathbf{r}_1$ и $\mathbf{r}_2$, проходит единственная прямая. **Уравнения по точкам** (6◊) и (6▽) этой прямой выражают коллинеарность векторов $\mathbf{r}-\mathbf{r}_1$ и $\mathbf{r}_2-\mathbf{r}_1$ как пропорциональность их координат. Аналогично, уравнение по точкам (6△) единственной плоскости, проходящей через три фиксированные точки в пространстве, не лежащие на одной прямой, выражает компланарность векторов $\mathbf{r}-\mathbf{r}_1$, $\mathbf{r}_2-\mathbf{r}_1$ и $\mathbf{r}_3-\mathbf{r}_1$ как зануление их смешанного произведения. Алгебраически эти условия в строке (6) таблицы кажутся разнородными, но, переформулировав их, можно сделать их похожими во всех трёх случаях. Этому посвящён (необязательный) последний пункт текущего раздела.





Смена способа задания

Каждый из перечисленных способов задания прямой или плоскости в определённых ситуациях лучше других, поэтому при решении задач регулярно требуется переходить от одного к другому. Для большинства переходов действия слабо зависят от варианта (△,◊,▽).

($1 \rightsquigarrow 2$) Каждая нормаль $\mathbf{n}$ выписывается из уравнения: $\mathbf{n} = (A, B)$ или $\mathbf{n} = (A, B, C)$. Чтобы угадать точку $\mathbf{r}_0$, можно подставить, например, $x = 0$, а также $y = 0$ в случае (1△). Если $A = 0$, следует занулить другую координату.

($1 \rightsquigarrow 6$) Угадывать можно так, как в ($1 \rightsquigarrow 2$), но используя две или три разные подстановки.

По		ищем			
		(1)	(2)	(5)	(6)
(1)	△	общее	списать $\mathbf{n}$ угадать $\mathbf{r}_0$	пройти через точки	угадать две или три точки
	◇				
	▽				
(2)	△	раскрыть скобки	нормальное	$\mathbf{v}, \mathbf{w} = (\times)\mathbf{n}$	пройти через параметры
	◇			$\mathbf{v} = (\times)\mathbf{n}$	
	▽			$\mathbf{v} = \mathbf{n} \times \mathbf{m}$	
(5)	△	пройти через нормальное	$\mathbf{n} = \mathbf{v} \times \mathbf{w}$	параметры	подставить значения параметров
	◇		$\mathbf{n} = (\times)\mathbf{v}$		
	▽		$\mathbf{n}, \mathbf{m} = (\times)\mathbf{v}$		
(6)	△	раскрыть всё	пройти через параметры	$\mathbf{r}_0 = \mathbf{r}_1$	по точкам
	◇			$\mathbf{v} = \mathbf{r}_2 - \mathbf{r}_1$	
	▽			$\mathbf{w} = \mathbf{r}_3 - \mathbf{r}_1$	

($2 \rightsquigarrow 5$) Обозначение $(\times)$ призвано напоминать простой способ нахождения векторов, ортогональных данному: если $\mathbf{n} = (a, b)$, то $\mathbf{v} = (b, -a)$; если $\mathbf{n} = (a, b, c)$, то $\mathbf{v} = (b, -a, 0)$ и $\mathbf{w} = (0, c, -b)$ при $b \neq 0$, а иначе, взять $(c, 0, -a)$ за $\mathbf{v}$ или $\mathbf{w}$.

($5 \rightsquigarrow 2$) Этот переход симметричен ($2 \rightsquigarrow 5$).

($6 \rightsquigarrow 1$) Раскрыть определитель и собрать коэффициенты.

Дополнение об уравнениях по точкам

Первым делом перепишем пропорцию (6◇) как

$$\begin{vmatrix} x - x_1 & y - y_1 \\ x_2 - x_1 & y_2 - y_1 \end{vmatrix} = 0,$$

что уже походит на выражение компланарности (6△): отличие лишь в порядке определителя. Далее, заметим совпадение (с точностью до знака) значений этого определителя и определителя (7◇):

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = \begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 - x_1 & y_2 - y_1 & 0 \end{vmatrix} = \begin{vmatrix} x & y & 1 \\ x_1 - x & y_1 - y & 0 \\ x_2 - x_1 & y_2 - y_1 & 0 \end{vmatrix}$$

Аналогичное явление заметно и в столбце (△), где появляется определитель *четвёртого* порядка. Да, он большой, зато по сравнению с (6△) элементы его завидно проще! Получаем задание прямой на плоскости и плоскости в пространстве уравнением «красивый определитель равен нулю». В случае прямой в пространстве из двух пропорций получаем два таких уравнения (7▽).

Однако можно пойти дальше. Равенство определителя квадратной матрицы нулю, как мы позже установим, эквивалентно тому, что одну из её строк можно выразить через другие, а это означает, что ранг этой матрицы меньше максимально возможного: «ранг падает». С другой стороны, ранг ни одной матрицы в строке (8) не может упасть более чем на единицу, ибо это исключено предположением **общего положения** зафиксированных точек: если их две, то они различны; если их три, то они не лежат на одной прямой.

В самом деле, для двух случаев прямых условия общего положения пары точек записываются с помощью рангов в виде

$$\operatorname{rk}\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{bmatrix} = 2, \qquad \operatorname{rk}\begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \end{bmatrix} = 2.$$

Эти равенства гарантируют, что ранги матриц в клетках (8◊) и (8▽) не меньше двух. Поэтому условие, что ранг такой матрицы не равен трём, то есть падает, задаёт искомую прямую.

Условие общего положения трёх точек в пространстве также записывается с помощью ранга в виде

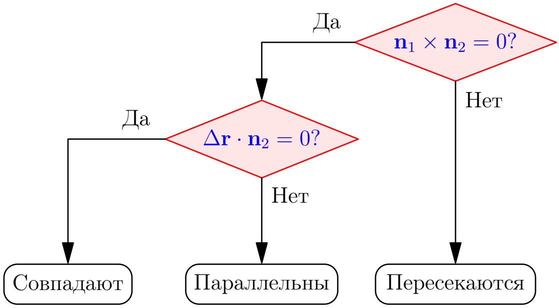
$$\operatorname{rk}\begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{bmatrix} = 3,$$

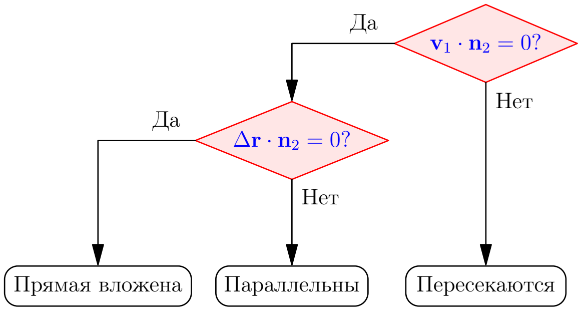
ибо в соответствии с уже обоснованным уравнением прямой (8▽) это означает, что первая точка не лежит на прямой, проходящей через вторую и третью. Итак, в строке (8) уравнения стали единообразны.

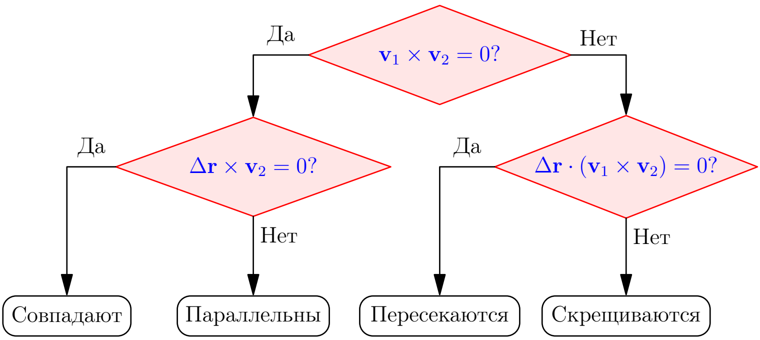
1.6. Взаимное расположение прямых и плоскостей

Варианты расположения пары фигур

Взаимное расположение двух прямых/плоскостей в пространстве, когда прямые заданы в параметрическом виде, а плоскости — в нормальном, можно определить по их направляющим векторам $\mathbf{v}_1$, $\mathbf{v}_2$, нормалям $\mathbf{n}_1$, $\mathbf{n}_2$ и вектору $\Delta\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ между их начальными точками. Всего здесь три случая: (1) две плоскости; (2) прямая и плоскость; (3) две прямых. В каждом из них укажем процедуру для выяснения расположения в виде алгоритма, а точнее его блок-схемы на следующей странице.


$\mathbf{n}_1 \times \mathbf{n}_2 = 0?$
Да
Нет
$\Delta\mathbf{r} \cdot \mathbf{n}_2 = 0?$
Да
Нет
Совпадают
Параллельны
Пересекаются


$\mathbf{v}_1 \cdot \mathbf{n}_2 = 0?$
Да
Нет
$\Delta\mathbf{r} \cdot \mathbf{n}_2 = 0?$
Да
Нет
Прямая вложена
Параллельны
Пересекаются


$\mathbf{v}_1 \times \mathbf{v}_2 = 0?$
Да
Нет
$\Delta\mathbf{r} \times \mathbf{v}_2 = 0?$
Да
Нет
$\Delta\mathbf{r} \cdot (\mathbf{v}_1 \times \mathbf{v}_2) = 0?$
Да
Нет
Совпадают
Параллельны
Пересекаются
Скрещиваются

Расстояния

Расстоянием между двумя фигурами называют кратчайшее расстояние от точки на одной фигуре до точки на другой. Расстояния между точками, прямыми и плоскостями находят через различные произведения; соберём все случаи в таблицу, а формулы для указанных в ней величин найдём позже. В этом разделе большинство формул содержат вектор $\Delta\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$.

Если фигуры параллельны, то для вычисления расстояния одна из них заменяется любой своей точкой; при этом в случае прямой и плоскости в пространстве заменяется прямая, потому что расстояния от её точек до плоскости одинаковы, но не наоборот. Без параллельности нетривиален только случай скрещивающихся прямых.

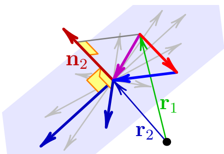
		если $\parallel$			если $\nparallel$		
	$\bullet$	$\triangle$	$\lozenge$	$\triangledown$	$\triangle$	$\lozenge$	$\triangledown$
$\bullet$	D_p	D_s	D_s	D_v			
$\triangle$	D_s	D_s		D_s	0		0
$\lozenge$	D_s		D_s			0	
$\triangledown$	D_v	D_s		D_v	0		D_t

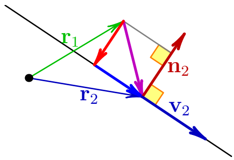
Найдём обещанные расстояния. Равенство $D_p = |\Delta\mathbf{r}|$ есть просто определение расстояния между точками, а формулы

$$D_s = \frac{|\mathbf{n}_2 \cdot \Delta\mathbf{r}|}{|\mathbf{n}_2|}, \qquad D_v = \frac{|\mathbf{v}_2 \times \Delta\mathbf{r}|}{|\mathbf{v}_2|}, \qquad D_t = \frac{|(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \Delta\mathbf{r}|}{|\mathbf{v}_1 \times \mathbf{v}_2|}.$$

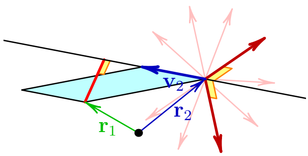
для расстояний D_s, D_v и D_t следуют из геометрических свойств скалярного, векторного и смешанного произведений.

(D_s) Это расстояние равно длине (норме) проекции $\Delta\mathbf{r}^{\parallel \mathbf{n}_2}$.

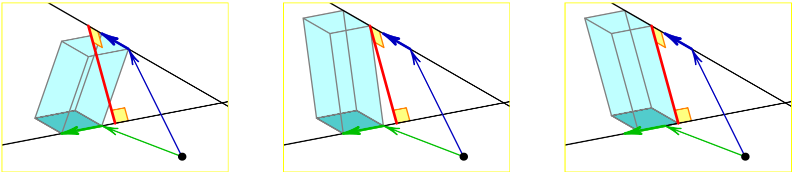




(D_v) Площадь параллелограмма делится на длину его основания.

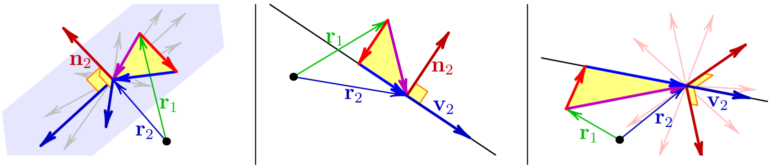


(D_t) Объём параллелепипеда делится на площадь его основания.



Проекции

Как узнать положение **проекций** — точек, реализующих кратчайшие расстояния? Привлекается проектирование на векторы, задающие данную прямую или плоскость. Поэтому возникают два случая: при размерности фигуры, равной 1, используем её направляющий вектор, а при коразмерности, равной 1 — нормаль к ней.



Положение проекции $\mathbf{r}_\perp$ точки $\mathbf{r}_1$ на прямую $\mathbf{r} = \mathbf{r}_2 + t\mathbf{v}_2$ как $(\Diamond)$ на плоскости, так и $(\triangledown)$ в пространстве, даёт формула

$$\mathbf{r}_2 - \mathbf{r}_\perp = (\mathbf{r}_2 - \mathbf{r}_1)^{\|\mathbf{v}_2} \implies \mathbf{r}_\perp = \mathbf{r}_2 - \Delta\mathbf{r}^{\|\mathbf{v}_2} = \mathbf{r}_2 - \frac{\Delta\mathbf{r}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2.$$

Положение проекции $\mathbf{r}_\perp$ точки $\mathbf{r}_1$ на плоскость $(\triangle)$ или прямую $(\Diamond)$ с нормальным уравнением $\mathbf{n}_2\cdot(\mathbf{r} - \mathbf{r}_2) = 0$ даёт формула

$$\mathbf{r}_\perp - \mathbf{r}_1 = (\mathbf{r}_2 - \mathbf{r}_1)^{\|\mathbf{n}_2} \implies \mathbf{r}_\perp = \mathbf{r}_1 + \Delta\mathbf{r}^{\|\mathbf{n}_2} = \mathbf{r}_1 + \frac{\Delta\mathbf{r}\cdot\mathbf{n}_2}{\mathbf{n}_2\cdot\mathbf{n}_2}\mathbf{n}_2.$$

В случае $(\Diamond)$ они дадут одинаковые ответы, ибо $\Delta\mathbf{r} = \Delta\mathbf{r}^{\|\mathbf{v}_2} + \Delta\mathbf{r}^{\|\mathbf{n}_2}$ потому что $\mathbf{v}_2$ и $\mathbf{n}_2$ ортогональны.

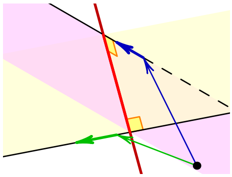
Общий перпендикуляр к скрещивающимся прямым

На каждой из данной пары скрещивающихся прямых $\mathbf{r} = \mathbf{r}_1 + t_1\mathbf{v}_1$ и $\mathbf{r} = \mathbf{r}_2 + t_2\mathbf{v}_2$ есть точка, ближайшая к другой прямой. Содержащую эти две точки прямую называют **общим перпендикуляром** к данным прямым,
чему перп?
а точки эти называют его **основаниями**. Нормальное уравнение общего

перпендикуляра можно найти, не зная самих оснований:

$$\begin{cases} (\mathbf{v}_1\times\mathbf{v}_2, \mathbf{v}_1, \mathbf{r} - \mathbf{r}_1) = 0, \\ (\mathbf{v}_1\times\mathbf{v}_2, \mathbf{v}_2, \mathbf{r} - \mathbf{r}_2) = 0. \end{cases}$$

Каждое из этих уравнений выражает через смешанное произведение условие компланарности и задаёт плоскость, содержащую одну из скрещивающихся прямых и искомый общий перпендикуляр. Тому, кто запомнил это свойство, не придётся запоминать сами формулы.



Подставляя в эти уравнения $\mathbf{r} = \mathbf{r}_1 + t_1\mathbf{v}_1$ и $\mathbf{r} = \mathbf{r}_2 + t_2\mathbf{v}_2$, сразу находим значения t_k, отвечающие положениям оснований:

$$t_1 = \frac{(\Delta\mathbf{r}\times\mathbf{v}_2)\cdot(\mathbf{v}_1\times\mathbf{v}_2)}{(\mathbf{v}_1\times\mathbf{v}_2)\cdot(\mathbf{v}_1\times\mathbf{v}_2)} \quad \text{и} \quad t_2 = -\frac{(\mathbf{v}_1\times\mathbf{v}_2)\cdot(\mathbf{v}_1\times\Delta\mathbf{r})}{(\mathbf{v}_1\times\mathbf{v}_2)\cdot(\mathbf{v}_1\times\mathbf{v}_2)}.$$

Здесь, как и прежде, $\Delta\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. Отметим, что для нахождения именно *расстояния* между скрещивающимися прямыми гораздо проще пользоваться указанной выше формулой (D_t), чем искать основания общего перпендикуляра.

другой способ!

Глава 2. МАТРИЦЫ, КОМПЛЕКСНЫЕ ЧИСЛА, МНОГОЧЛЕНЫ

2.1. Введение в системы линейных уравнений

Лекция 4
23.09.19

Решим несколько задач о пересечении прямых и плоскостей, заданных общими уравнениями. Координаты точки на одной фигуре должны удовлетворять её уравнению, а координаты точки на пересечении двух или более фигур должны удовлетворять уравнению каждой. В такой ситуации говорят о системе уравнений; мы уже видели это, задавая прямую в пространстве как пересечение двух плоскостей, что даёт систему двух линейных уравнений с тремя неизвестными.

Большинству студентов уже приходилось решать подобные системы, но небольшие; все знакомы с методом, пригодным для них. Этот раздел служит практическим введением в более мощный, даже универсальный **метод исключения неизвестных**.

Пример 1: единственное решение

Найдём пересечение двух прямых на плоскости. На алгебраическом языке — решим систему из двух линейных уравнений с двумя неизвестными:

$$\begin{cases} x_1 + 2x_2 = 3, \\ 3x_1 + 4x_2 = 5. \end{cases}$$

Если пара (x_1, x_2) удовлетворяет обоим уравнениям, то она удовлетворяет любой их линейной комбинации. Опираясь на это, следует искать комбинации, постепенно приближающие нас к ответу. Например, можно вычесть из второго уравнения первое, умноженное на 3, и получить (третье уравнение)

$$0x_1 - 2x_2 = -4.$$

Прибавляя его к первому, получим (четвёртое уравнение)

$$x_1 + 0x_2 = -1.$$

Поделив ещё третье уравнение на -2, считываем ответ:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

Здесь и далее наборы наших неизвестных, они же координаты искомых точек, располагаем в виде столбца по скрытой пока причине.

Пример 2: решение с одним параметром

Найдём пересечение двух плоскостей в трёхмерном пространстве. На алгебраическом языке — решим систему из двух линейных уравнений с тремя неизвестными

$$\begin{cases} x_1 + 2x_2 + 4x_3 = 4, \\ 3x_1 + 4x_2 + 6x_3 = 8. \end{cases}$$

Коэффициенты при x_1 и x_2 взяты такие же, как в первом примере. Такие же два действия как там дают третье уравнение

$$0x_1 - 2x_2 - 6x_3 = -4$$

и затем четвёртое уравнение

$$x_1 + 0x_2 - 2x_3 = 0.$$

Нужно ли нам столько уравнений? Нет! На каждом шаге прибавления или вычитания можно заменять то уравнение, куда прибавляют или вычитают, на новое (как обычно происходит в компьютере). Значит, после первого шага получим систему

$$\begin{cases} x_1 + 2x_2 + 4x_3 = 4, \\ 0x_1 - 2x_2 - 6x_3 = -4, \end{cases}$$

а после второго — систему

$$\begin{cases} x_1 + 0x_2 - 2x_3 = 0, \\ 0x_1 - 2x_2 - 6x_3 = -4. \end{cases}$$

Поделим ещё последнее уравнение на -2:

$$\begin{cases} x_1 + 0x_2 - 2x_3 = 0, \\ 0x_1 + x_2 + 3x_3 = 2. \end{cases}$$

Нам удаётся выразить x_1 и x_2 через x_3,

$$\begin{cases} x_1 = 0 + 2x_3, \\ x_2 = 2 - 3x_3, \end{cases}$$

но не удаётся определить единственные подходящие значения. Геометрическое объяснение этого в том, что две заданные плоскости пересекаются по прямой. Наша цель теперь задать эту прямую параметрическим уравнением $\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}$, только не в векторах, а в координатах. Параметром может служить $t = x_3$, поскольку остальные неизвестные выражены через него:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 + 2x_3 \\ 2 - 3x_3 \\ 0 + x_3 \end{bmatrix}.$$

Сюда искусственно добавлено третье уравнение, выполненное всегда. Записываем ответ, выделяя начальную точку и направляющий вектор найденной прямой:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} x_3.$$

Пример 3: решение с двумя параметрами

Здесь снова две плоскости:

$$\begin{cases} x_1 + 2x_2 + 4x_3 = 4, \\ 2x_1 + 4x_2 + 8x_3 = 8. \end{cases}$$

Вычитая из второго уравнения удвоенное первое, увидим $0 = 0$. Система избыточна, ибо два исходных уравнения задают одну и ту же плоскость; вычёркнем второе.

Получим параметрическое задание $\mathbf{r} = \mathbf{r}_0 + t\mathbf{v} + s\mathbf{w}$ этой плоскости, только в координатах. Выразим x_1 через x_2 и x_3:

$$x_1 = 4 - 2x_2 - 4x_3.$$

Как и выше, переменные в правой части берём в качестве параметров — тут их оказывается два. Добавляем уравнения $x_2 = x_2$ и $x_3 = x_3$, затем выделяем начальную точку и направляющие векторы:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 - 2x_2 - 4x_3 \\ 0 + x_2 + 0x_3 \\ 0 + 0x_2 + x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix} x_3.$$

Пример 4: избыточная система

Возьмём теперь две плоскости из Примера 2 и добавим третью:

$$\begin{cases} x_1 + 2x_2 + 4x_3 = 4, \\ 3x_1 + 4x_2 + 6x_3 = 8, \\ 2x_1 + 3x_2 + 5x_3 = 6. \end{cases}$$

Вычитая из третьего уравнения первое, получим (четвёртое)

$$x_1 + x_2 + x_3 = 2.$$

Вычитая из второго уравнения первое, получим (пятое)

$$2x_1 + 2x_2 + 2x_3 = 4.$$

Видно, что из двух последних одно лишнее. Исходная система тоже избыточна, но это проявляется только после её преобразований (хотя пример несложен, поэтому опытному глазу избыточность тут заметна сразу). Возможно, например, такое описание этой геометрической ситуации: третья плоскость пересекает первые две вдоль их общей прямой. Ответ совпадает с ответом в примере 2.

Пример 5: несовместная система

Возьмём теперь плоскости из Примера 4 и сдвинем третью:

$$\begin{cases} x_1 + 2x_2 + 4x_3 = 4, \\ 3x_1 + 4x_2 + 6x_3 = 8, \\ 2x_1 + 3x_2 + 5x_3 = 0. \end{cases}$$

Вычитая из третьего уравнения первое, получим (четвёртое)

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

Вычитая из второго уравнения первое, получим (пятое)

$$2x_1 + 2x_2 + 2x_3 = 4.$$

Вычитая из пятого уравнения удвоенное четвёртое, получим

$$0x_1 + 0x_2 + 0x_3 = 12.$$

Значит, решений нет. Тогда говорят, что исходная система **несовместна**. Геометрически — три плоскости не имеют общих точек; попарно же они пересекаются по трём параллельным прямым.

Пример 6: три плоскости в общем положении

Снова возьмём две плоскости из Примера 2 и добавим к ним третью:

$$\begin{cases} x_1 + 2x_2 + 4x_3 = 4, \\ 3x_1 + 4x_2 + 6x_3 = 8, \\ x_1 - x_2 - 4x_3 = -1. \end{cases}$$

Повторяя вычисления Примера 2, можно постепенно заменить первые два уравнения более простыми:

$$\begin{cases} x_1 + 0x_2 - 2x_3 = 0, \\ 0x_1 + x_2 + 3x_3 = 2, \\ x_1 - x_2 - 4x_3 = -1. \end{cases}$$

Продолжим упрощение. Из третьего уравнения вычтем первое:

$$0x_1 - x_2 - 2x_3 = -1;$$

затем к этому прибавим второе:

$$0x_1 + 0x_2 + x_3 = 1.$$

Наконец, с помощью полученного упростим остальные уравнения:

$$x_1 = 2,$$
$$x_2 = -1.$$

Ответ: система имеет единственное решение

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}.$$

Три заданные плоскости пересекаются в одной точке. Этот случай — *общее положение* трёх плоскостей.

Удаление букв

Выполнив даже относительно небольшое количество таких несложных, рутинных действий с линейными уравнениями, отметим два обстоятельства:

- все операции совершаются над числовыми коэффициентами, а «иксы» просто дописываются в нужных местах;
- чтобы понять каждый шаг, производимое действие требует комментария с указанием номеров участвующих в нём уравнений.

Укажем теперь приёмы иной записи процесса решения, позволяющие вести его намного эффективнее даже вручную. Провидчески, мы и так уже старались записывать каждую переменную в своей колонке, и мысленная связь переменной и колонки имеется. Давайте теперь вынесем переменные в заголовки колонок, а вместо уравнений будем писать строки коэффициентов. Для Примера 1 получим таблицу

x_1	x_2		
1	2	3	A_1
3	4	5	A_2

Компактные комментарии можно устроить, нумеруя строки таблицы и ссылаясь на операции с ними, будто с векторами. Здесь мы вычисляем последовательно $A_3 = A_2 - 3A_1$, $A_4 = A_1 + A_3$, $A_5 = A_3/(-2)$:

x_1	x_2		
1	2	3	A_1
3	4	5	A_2
0	−2	−4	$A_3 = A_2 - 3A_1$
1	0	−1	$A_4 = A_1 + A_3$
0	1	2	$A_5 = A_3/(-2)$.

Такая запись весьма экономична, и при появлении новых строк нет особой нужды удалять старые, ведь можно просто дописывать новые снизу. Здесь строки A_4 и A_5 интерпретируются как $x_1 = -1$ и $x_2 = 2$, то есть содержат искомый ответ.

Перепишем в таком виде вычисления в Примере 2:

x_1	x_2	x_3		
1	2	4	4	A_1
3	4	6	8	A_2
0	−2	−6	−4	$A_3 = A_2 - 3A_1$
1	0	−2	0	$A_4 = A_1 + A_3$
0	1	3	2	$A_5 = A_3/(-2)$.

Затем вычисления в Примере 4:

x_1	x_2	x_3		
1	2	4	4	A_1
3	4	6	8	A_2
2	3	5	6	A_3
1	1	1	2	$A_4 = A_3 - A_1$
2	2	2	4	$A_5 = A_2 - A_1$
0	0	0	0	$A_6 = A_5 - 2A_4$.

И наконец, вычисления в Примере 6:

x_1	x_2	x_3		
1	2	4	4	A_1
3	4	6	8	A_2
1	−1	−4	−1	A_3
0	−2	−6	−4	$A_4 = A_2 - 3A_1$
1	0	−2	0	$A_5 = A_1 + A_4$
0	1	3	2	$A_6 = A_4/(-2)$
0	−1	−2	−1	$A_7 = A_3 - A_5$
0	0	1	1	$A_8 = A_7 + A_2$
1	0	0	2	$A_9 = A_5 + 2A_8$
0	1	0	−1	$A_{10} = A_6 - 3A_8$.

Добавление крестиков

Как отмечено в Примере 4, существенных уравнений в системе может быть меньше их исходного общего количества. Описанный метод записи вычислений не позволяет следить за этой важной величиной, особенно когда операций со строками требуется больше, однако внести в него эту возможность нетрудно.

Достаточно помечать на каждом шаге выбывшую строку. Строка выбывает в трёх случаях: к ней прибавили другую строку (умножив ту на число); её умножили на число; она состоит только из нулей. При первых двух операциях число невыбывших строк сохраняется, а третья уменьшает это число на единицу.

Статичная демонстрация этого не очень эффективна, здесь нужно видеть сам процесс.

анимация!

2.2. Смена координат и матричный язык

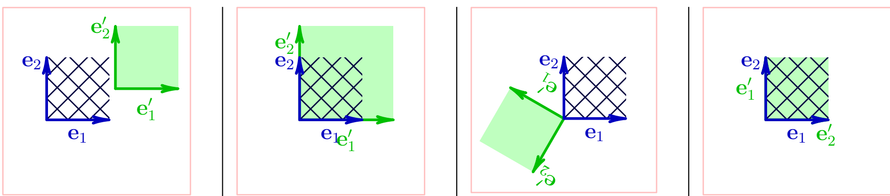
Простейшие типы преобразований

Лекция 4+ 23.09.19

Необходимо знать правила пересчёта координат точек и векторов из одной прямоугольной системы в другую. Две таких системы $\mathcal{K}$ и $\mathcal{K}'$ могут отличаться расположением начала, единицей длины, направлениями осей и ориентацией параллелограмма на ортах. Проще всего связать системы, отличающиеся лишь в одном из этих аспектов.

Переход $\mathcal{K} \rightsquigarrow \mathcal{K}'$	*Отличие* $\mathcal{K}'$ *от* $\mathcal{K}$
Перенос (реже сдвиг)	Расположение начала
Растяжение или сжатие	Единица длины
Поворот	Направления осей
Отражение	Ориентация

Выпишем правила пересчёта $(x, y) \mapsto (x', y')$ координат каждой точки в таких случаях. Координаты векторов при переносе неизменны, а при остальных переходах ведут себя идентично координатам точек.



перенос | растяжение | поворот | отражение

• При переносе, если начало системы $\mathcal{K}$ имеет координаты (u, v) в системе $\mathcal{K}'$:

$$x' = x + u,$$
$$y' = y + v;$$

• при растяжении или сжатии:

$$x' = cx,$$
$$y' = cy;$$

• при повороте осей на угол α против часовой стрелки:

$$x' = \quad x\cos\alpha + y\sin\alpha,$$
$$y' = -x\sin\alpha + y\cos\alpha;$$

• при отражении, меняющем оси местами:

$$x' = y,$$
$$y' = x.$$

Композиция линейных преобразований

В последних трёх случаях «старая» система $\mathcal{K}$ и «новая» система $\mathcal{K}'$ имеют общее начало, а три соответствующих правила можно объединить в одно:

$$x' = ax + by,$$
$$y' = cx + dy,$$

с подходящими значениями коэффициентов.

Поскольку системы $\mathcal{K}'$ и $\mathcal{K}$ совершенно равноправны, такую же форму должно иметь выражение старых координат через новые:

$$(\spadesuit) \qquad \begin{aligned} x &= a'x' + b'y', \\ y &= c'x' + d'y'. \end{aligned}$$

Далее мы будем изучать формулы перехода именно в этой форме.

Пусть теперь имеется три системы координат $\mathcal{K}$, $\mathcal{K}'$ и $\mathcal{K}''$ с общим началом. Цепочка из двух переходов $\mathcal{K} \rightsquigarrow \mathcal{K}' \rightsquigarrow \mathcal{K}''$ даёт тот же результат, что и непосредственный переход $\mathcal{K} \rightsquigarrow \mathcal{K}''$. Подставляя в выражения ($\spadesuit$) для перехода $\mathcal{K} \rightsquigarrow \mathcal{K}'$ аналогичные выражения для перехода $\mathcal{K}' \rightsquigarrow \mathcal{K}''$, мы получаем связь коэффициентов непосредственного перехода с коэффициентами двух шагов эквивалентной ему цепочки:

$$(\clubsuit) \qquad \begin{aligned} x &= (a'a'' + b'c'')x'' + (a'b'' + b'd'')y'', \\ y &= (c'a'' + d'c'')x'' + (c'b'' + d'd'')y''. \end{aligned}$$

Тупо запоминать подобные формулы опасно, но к счастью, не нужно, потому что для них есть более удобный язык.

Матрицы и операции над ними

Любой прямоугольный массив чисел называют **матрицей**. Первые характеристики матрицы — количество её **строк** и количество её **столбцов**. Выписывая всю матрицу или упоминая её отдельные элементы, номера строк и столбцов указывают двойными индексами, например:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}.$$

Две матрицы совпадающих размеров складывают, просто складывая элементы в соответствующих позициях:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{bmatrix}$$

Матрицы несовпадающих размеров без крайней нужды не складывают. Вычитание полностью аналогично сложению.

Более хитро устроено умножение матриц. Весь смысл хитрости в том, что формулы (♠) и (♣) переписываются в виде

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \end{bmatrix},$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \cdot \begin{bmatrix} a'' & b'' \\ c'' & d'' \end{bmatrix} \cdot \begin{bmatrix} x'' \\ y'' \end{bmatrix}.$$

Здесь координаты точек записаны столбцами, а информация о переходах, сохраняющих начало координат, собрана в квадратные **матрицы перехода**, которые от самих точек не зависят. Это позволяет изучать такие переходы цельно, абстрагируясь от конкретных точек. Композиция таких переходов соответствует произведению их матриц.

Выпишем правило перемножения в индексных обозначениях:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11}+a_{12}b_{21} & a_{11}b_{12}+a_{12}b_{22} \\ a_{21}b_{11}+a_{22}b_{21} & a_{21}b_{12}+a_{22}b_{22} \end{bmatrix}.$$

Инспектируя это правило, следует заключить, что в общем случае матричное произведение AB определено, только когда количество столбцов A равно количеству строк B. Тогда AB имеет столько же строк, сколько A, и столько же столбцов, сколько B. Итак,

$$(\text{матрица } m \times s) \cdot (\text{матрица } s \times n) = (\text{матрица } m \times n),$$

а элемент матрицы AB, стоящий в строке i и столбце j, вычисляется по формуле

$$(AB)_{ij} = \sum_{1\leqslant k\leqslant s} a_{ik}b_{kj}.$$

Если произведение AB определено, то вовсе не обязательно произведение BA также должно быть определено. Для квадратных матриц ситуация проще; тем не менее, сразу нужно запомнить, что и для них обычно $AB \neq BA$.

Пример. Легко проверить, что операции поворота и отражения прямоугольных систем координат не перестановочны. При повороте на угол φ вокруг нуля и при отражении относительно оси Ox координаты связаны соответственно правилами

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \end{bmatrix}, \qquad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \end{bmatrix}.$$

Перемножая матрицы переходов хотя бы для $\varphi = \frac{\pi}{2}$, получаем

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

Пример. Очень важную роль играет **единичная матрица**

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Она задаёт тождественное преобразование, сохраняющее систему координат и оставляющую все точки на месте. По отношению к умножению матриц, она ведёт себя аналогично единице: $AE = EA = A$.

Транспонирование

Матрицу с одинаковом количеством строк и столбцов называют **квадратной**; таковы все матрицы линейных преобразований, введённые выше. Произвольную матрицу обычно называют **прямоугольной** только с целью подчеркнуть, что она не обязательно квадратная. Наиболее часто встречающиеся частные случаи последних — самостоятельные строки (матрицы $1 \times n$) и столбцы (матрицы $n \times 1$).

В любой матрице набор всех позиций с совпадающими номерами строки и столбца называют **главной диагональю**. Она идёт из верхнего левого угла вправо и вниз, но попадает в правый нижний угол только в случае квадратных матриц.

Совершенно новая, притом весьма простая операция $A \mapsto A^\top$ на матрицах меняет роли строк и столбцов. Визуально это отражение относительно главной диагонали, но называется оно **транспонированием** (смутно «ставить поперёк» с латыни). Например,

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \implies A^\top = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

В индексных обозначениях пишут $(A^\top)_{ij} = (A)_{ji}$, то есть меняют местами индексы строки и столбца.

Упражнение. *Проверьте матричное тождество* $(AB)^\top = B^\top A^\top$.

Произвольные линейные и аффинные преобразования

Начав этот раздел с четырёх простейших типов преобразований координат на плоскости, мы пришли к их записи с помощью матриц. Необходимо добавить теперь некоторые дополнительные детали.

Дело в том, что 2×2 матриц общего вида больше, чем преобразований декартовых координат. Например, матрицы

$$A_1 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

переводят исходную прямоугольную декартову систему $\mathcal{K}$ в «запрещённые» системы координат. В первом случае меняется длина одного базисного вектора, но сохраняется длина другого. Во втором случае меняется ещё и угол между базисными векторами.

рисунок

Если в одних приложениях удобнее ограничиваться прямоугольными декартовыми координатами, то в других удобнее допускать переходы к системам координат с произвольным базисом. Сейчас мы не станем углубляться в эти вопросы, но позже они появятся вновь.

Активные и пассивные преобразования

Рассмотренные выше преобразования координат не затрагивают самих точек и векторов; все они относятся к задаче описания неподвижного объекта в различных системах координат. Такие преобразования называют **пассивными**.

Для описания физического объекта, движущегося в неподвижной системе координат, нужны **активные** преобразования. Формулы при этом получаются похожие. Деталь, которой они отличаются, зачастую ускользает за пределы внимания, что приводит к ошибкам. А именно, формулы активных и пассивных преобразований обратны друг другу. Например, при повороте фигуры вокруг начала координат на угол φ против часовой стрелки координаты всех её точек изменяются таким же образом, как если бы фигура оставалась на месте, а система координат повернулась на тот же угол по часовой стрелке.

2.3. Комплексные числа

Очень эффективный способ описания преобразований плоскости появляется с введением на радиус-векторах её точек особой операции умножения. Вместе с их сложением как векторов образуется замечательная структура **комплексных чисел**, во многом похожих на действительные, а кое в чём даже более удобных. Лекция 5 30.09.19

Алгебраическая форма и операции

Обозначим начало координат через 0, а концы орт — через 1 и i. Теперь точке (a, b) плоскости соответствует комплексное число $a + ib$;

$$\{\text{точки прямой}\} \leftrightarrow \mathbb{R}, \qquad \{\text{точки плоскости}\} \leftrightarrow \mathbb{C},$$

а сложение и вычитание выполняются по формуле

$$(a_1 + ib_1) \pm (a_2 + ib_2) = (a_1 \pm a_2) + i(b_1 \pm b_2).$$

Определим на них умножение — а позже и деление — так, чтобы:

- выполнялись привычные числовые законы ассоциативности, коммутативности, дистрибутивности (в школе их называют «сочетательный», «переместительный», «распределительный»), нуля и единицы;
- прямая, проходящая через точки 0 и 1, представляла $\mathbb{R} \subset \mathbb{C}$.

Ввиду требования дистрибутивности достаточно научиться перемножать орты, а свойства единицы

$$1 \cdot 1 = 1 \quad \text{и} \quad 1 \cdot i = i \cdot 1 = i$$

оставляют неясным лишь значение произведения $i^2 = i \cdot i$. Стоящую задачу решает выбор

$$i^2 = -1.$$

Любопытные могут самостоятельно исследовать, к чему ведут другие варианты — они лежат чуть в стороне от нашего сегодняшнего предмета. Итак, получаем правило умножения комплексных чисел:

$$(a_1 + ib_1)(a_2 + ib_2) = (a_1 a_2 - b_1 b_2) + i(a_2 b_1 + a_1 b_2).$$

Фактически, при умножении нужно просто раскрывать скобки, а затем заменять i^2 на -1.

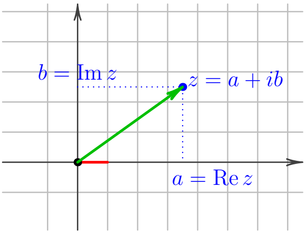
В поисках формулы для деления, рассмотрим сначала

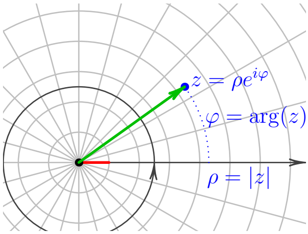
$$\frac{1}{a+ib} = \frac{a-ib}{(a+ib)(a-ib)} = \frac{a}{a^2+b^2} + i\frac{-b}{a^2+b^2}.$$

Поэтому, деление возможно на любое комплексное число $a+ib \neq 0$ и выполняется по формуле

$$\frac{a_1+ib_1}{a_2+ib_2} = \frac{a_1a_2+b_1b_2}{a_2^2+b_2^2} + i\frac{a_2b_1-a_1b_2}{a_2^2+b_2^2},$$

что весьма неудобно для запоминания. Проще помнить приём, применённый в предыдущем равенстве для избавления от i в знаменателе и называемый **домножением на сопряжённое**.





Показательная форма

Полярные координаты, связанные с прямоугольными формулами

$$\begin{cases} x = \rho\cos\varphi, \\ y = \rho\sin\varphi, \end{cases}$$

дают другую удобную форму записи комплексных чисел:

- прямоугольная система $\leftrightarrow$ **алгебраическая форма** $z = a+ib$;
- полярная система $\leftrightarrow$ **показательная форма** $z = \rho e^{i\varphi}$,

где $e^{i\varphi}$ из формулы Эйлера (Cotes, 1714; Euler, 1748)

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

пока можно считать просто удачным обозначением. Явная запись с косинусом и синусом, называемая **тригонометрической формой**, очевидно более громоздка, в особенности когда вместо φ подставляется очень большое выражение (что типично в физике). Поэтому рекомендуется привыкать именно к показательной форме.

$$\text{При этом называют} \left\{ \begin{array}{ll} a = \operatorname{Re} z & \textbf{действительной частью} \\ b = \operatorname{Im} z & \textbf{мнимой частью} \\ \rho = |z| & \textbf{модулем} \\ \varphi = \arg z & \textbf{аргументом} \end{array} \right\} \text{комплексного числа } z.$$

Сопряжение и свойства модуля

Часто встречается операция **комплексного сопряжения**:

$$z = a + ib = \rho e^{i\varphi} \mapsto \bar{z} = a - ib = \rho e^{-i\varphi}.$$

Физики любят писать z^* вместо $\bar{z}$. Геометрически это отражение относительно вещественной оси. Легко проверить, что сопряжение уважает арифметические операции:

$$\overline{z_1 \pm z_2} = \bar{z}_1 \pm \bar{z}_2, \quad \overline{z_1 z_2} = \bar{z}_1 \bar{z}_2, \quad \overline{z_1 / z_2} = \bar{z}_1 / \bar{z}_2.$$

Кроме того,

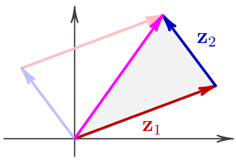
$$z + \bar{z} = 2a = 2\operatorname{Re} z \in \mathbb{R} \quad \text{и} \quad z \cdot \bar{z} = a^2 + b^2 = |z|^2 \in \mathbb{R}_{\geqslant 0}.$$

Отсюда следует **мультипликативность** модуля:

$$|z_1 z_2| = |z_1| \cdot |z_2|.$$

В отличие от $\mathbb{R}$, в $\mathbb{C}$ нет порядка: запись $z_1 < z_2$ не имеет полезного смысла для z_1, $z_2 \in \mathbb{C}$. Сравнивать комплексные числа можно только по модулю, причём выполнено **неравенство треугольника** **сдвинуть** z_2

$$|z_1 + z_2| \leqslant |z_1| + |z_2|.$$



Умножение и поворот

Оценим, насколько удачно обозначение $e^{i\varphi}$, вычисляя:

$$\begin{aligned} e^{i\varphi} e^{i\psi} &= (\cos\varphi + i \sin\varphi)(\cos\psi + i \sin\psi) \\ &= (\cos\varphi \cos\psi - \sin\varphi \sin\psi) + i(\sin\varphi \cos\psi + \cos\varphi \sin\psi) \\ &= \cos(\varphi + \psi) + i \sin(\varphi + \psi) \\ &= e^{i(\varphi+\psi)}. \end{aligned}$$

Поэтому показательная форма записи комплексных чисел удобна для умножения и деления:

$$\rho_1 e^{i\varphi_1} \cdot \rho_2 e^{i\varphi_2} = \rho_1\rho_2 e^{i(\varphi_1+\varphi_2)}; \quad \rho_1 e^{i\varphi_1}/\rho_2 e^{i\varphi_2} = (\rho_1/\rho_2) e^{i(\varphi_1-\varphi_2)}.$$

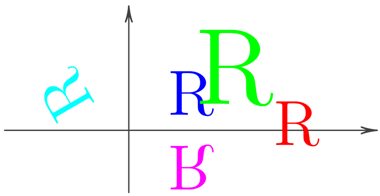
Теорема. *Умножение на комплексное число $w = e^{i\psi}$ является поворотом комплексной плоскости на угол $\psi = \arg w$ вокруг нуля.*

Первое доказательство. Умножим $z = x + iy$ на $w = \cos\psi + i\sin\psi$:

$$zw = (x\cos\psi - y\sin\psi) + i(x\sin\psi + y\cos\psi).$$

Теперь сравним с формулой поворота плоскости. Противоположность знаков при синусах объясняется тем, что здесь мы вертим сами *точки* плоскости, а в той формуле поворота — систему координат. □

Второе доказательство. Естественно, полярные координаты удобны для записи поворотов. Преобразование $z \mapsto ze^{i\psi}$ сохраняет начало координат (число 0) и все длины: $|ze^{i\psi}| = |z| \cdot |e^{i\psi}| = |z|$. Нужно также убедиться, что оно сохраняет ориентацию. □



Матричная форма

Другие простые действия с комплексными числами тоже дают преобразования плоскости уже известных нам типов. Соответствие между ними указано в таблице: сложение производит перенос; умножение на положительное вещественное число производит растяжение или сжатие; комплексное сопряжение производит отражение относительно вещественной оси.

Преобразование плоскости $\mathbb{C}$	*Формула*
Перенос	$z \mapsto z + w$, где $w \in \mathbb{C}$
Растяжение или сжатие	$z \mapsto z\,\rho$, где $\rho \in \mathbb{R}_{>0}$
Поворот	$z \mapsto z\,e^{i\psi}$, где $\psi \in \mathbb{R}$
Отражение мнимой оси	$z \mapsto \bar{z}$

Выпишем матрицы, задающие растяжение/сжатие и поворот:

$$\rho \leftrightarrow \begin{bmatrix} \rho & 0 \\ 0 & -\rho \end{bmatrix}; \quad e^{i\varphi} \leftrightarrow \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}.$$

В частности, умножение на мнимую единицу соответствует повороту на $90°$:

$$i = e^{i\pi/2} \leftrightarrow \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Так мы приходим к представлению любого комплексного числа матрицей специального вида:

$$\rho e^{i\varphi} \leftrightarrow \begin{bmatrix} \rho\cos\varphi & \rho\sin\varphi \\ -\rho\sin\varphi & \rho\cos\varphi \end{bmatrix}; \quad a+ib \leftrightarrow \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

Упражнение. *Какие операции с комплексными числами в матричной форме реализуются операциями с матрицами?*

Степени и корни

Правило умножения влечёт правило возведения в целую степень:

$$(\rho e^{i\varphi})^n = \rho^n e^{in\varphi} \text{ для } n \in \mathbb{Z}.$$

Формулой Муавра (de Moivre) называют более длинный вариант

$$[\rho(\cos\varphi + i\sin\varphi)]^n = \rho^n(\cos n\varphi + i\sin n\varphi).$$

Найдём теперь все значения выражения $\sqrt[n]{z}$ для положительного целого n. Из правила возведения в степень одно решение очевидно:

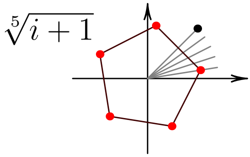
$$\sqrt[n]{\rho e^{i\varphi}} = \sqrt[n]{\rho}\, e^{i\varphi/n}.$$

Конечно, это не может дать всех решений. Даже среди вещественных чисел мы видим неоднозначность: например, $(-1)^2 = 1^2$, так что квадратных корней наверное два.

Хитрость, открывающая все решения, в том, что аргумент φ определён не однозначно, а лишь с точностью до $2\pi i$, поскольку $e^{i\varphi} = e^{i(\varphi+2\pi k)}$ для всех целых k. Следовательно, $\sqrt[n]{\rho e^{i\varphi}}$ есть *множество*

$$\left\{ \sqrt[n]{\rho}\, e^{i(\varphi+2\pi k)/n} \mid k = 0, 1, \ldots, n-1 \right\}.$$

Каждое ненулевое комплексное число имеет ровно n различных корней степени n. Это важное явление, подобного которому в $\mathbb{R}$ нет. На комплексной плоскости корни степени n из данного ненулевого $z \in \mathbb{C}$ расположены в вершинах правильного n-угольника с центром в нуле.



Интересные вещи выявляются при дотошном рассмотрении корней из единицы («унипотентов»). Произведение унипотентов степени n есть такой же унипотент, и даже произведение унипотентов неравных целых степеней — опять унипотент.

Комплексная экспонента

Вернёмся к формуле $e^{i\varphi}e^{i\psi} = e^{i(\varphi+\psi)}$. Функция $E(x)$, обладающая свойством $E(x_1+x_2) = E(x_1)E(x_2)$, обязательно имеет вид $E(x) = e^{\alpha x}$. При этом константу α можно определить из правила $\frac{d}{dx}(e^{\alpha x}) = \alpha e^{\alpha x}$. Применим это к функции $E(x) = \cos x + i \sin x$:

$$\tfrac{d}{dx}(\cos x + i\sin x) = (-\sin x + i\cos x) = i(\cos x + i\sin x).$$

Значит, в этом случае $E(x)$ является экспоненциальной функцией e^{ix}, что и оправдывает использование такого обозначения сразу, с предыдущего раздела.

К такому же выводу приводят разложения Тейлора:

$$\begin{aligned} e^{ix} &= 1 + (ix) + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \dots \\ &= 1 + ix - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} + \dots \\ &= \cos x + i \sin x. \end{aligned}$$

Подставляя в экспоненту значение аргумента π, выводим знаменитое тождество Эйлера:

$$e^{i\pi} + 1 = 0.$$

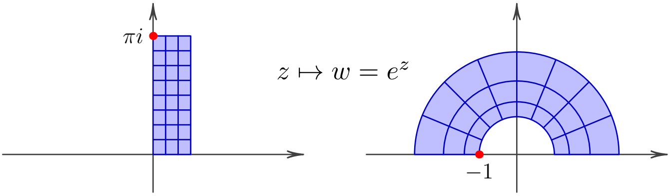
Отметим в качестве необязательного дополнения, что теперь легко определить функцию e^z комплексной переменной z, обладающую свойством

$$e^{z_1+z_2} = e^{z_1} \cdot e^{z_2}$$

для всех $z_i \in \mathbb{C}$. Достаточно положить для всех $a, b \in \mathbb{R}$

$$\exp(a+ib) = e^{a+ib} \leftrightharpoons e^a e^{ib} \leftrightharpoons e^a(\cos b + i\sin b).$$

Есть и другие, аналитические способы введения e^z, аналогичные случаю вещественной переменной: как предел последовательности, как сумма ряда; в результате получается одна и та же функция.



Определим также комплексные функции

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \qquad \operatorname{ch} z = \frac{e^z + e^{-z}}{2},$$
$$\sin z = \frac{e^{iz} - e^{-iz}}{2i}, \qquad \operatorname{sh} z = \frac{e^z - e^{-z}}{2}.$$

По сути, не только гиперболические, но и обычные косинус и синус оказываются всего лишь тенями комплексной экспоненты; её использование зримо упрощает вывод многих тригонометрических тождеств.

2.4. Основная теорема алгебры многочленов

Многочлены и операции над ними

Обозначим через $\mathbb{F}$ одну из числовых систем $\mathbb{Q}$, $\mathbb{R}$ или $\mathbb{C}$. **Многочленом** от буквы x с коэффициентами в $\mathbb{F}$ называют выражение вида

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

где все $a_k \in \mathbb{F}$. Если букве x не придавать никакого смысла, то её степени просто удобно организуют поведение коэффициентов при сложении и умножении многочленов по знакомым правилам: слагаемые с одинаковой степенью собираются вместе, а степени складываются при перемножении. Множество всех многочленов с коэффициентами в $\mathbb{F}$ обозначают через $\mathbb{F}[x]$. Индекс n последнего ненулевого (**старшего**) коэффициента многочлена f называют **степенью** f и пишут $\deg f = n$.

Операции деления на многочленах нет, потому что частным является, вообще говоря, дробь. Однако имеется деление с остатком.

Лемма. *Если даны многочлены f и $g \neq 0$, то однозначно находятся такие многочлены q и r, что $\deg r < \deg g$ и $f = qg + r$.*

Доказательство. По алгоритму деления «столбиком». □

Корни многочлена и их кратности

Вместо x можно подставлять различные выражения, и в частности, числа. **Корнем** многочлена $f(x)$ называют такое число c, что $f(c) = 0$. Многие задачи математики требуют отыскать корни многочлена или описать их совокупность.

Пример. Многочлен $x^2 + 1 \in \mathbb{R}[x]$ не имеет корней в $\mathbb{R}$, но имеет два корня $\pm i \in \mathbb{C}$.

Теорема (Bezout). *Равносильны утверждения:*

(1) число $c \in \mathbb{F}$ есть корень $f(x) \in \mathbb{F}[x]$;
(2) линейный многочлен $(x - c)$ делит $f(x)$ в $\mathbb{F}[x]$.

Доказательство. По делению с остатком, $f(x) = (x - c)q(x) + r(x)$ и $\deg r(x) < \deg(x - c) = 1$, поэтому $r(x) = \mathrm{const} = r(c) = f(c)$. □

Если $f(x)$ делится на $(x - c)^k$, но не на $(x - c)^{k+1}$, где $k \in \mathbb{Z}_{>0}$, то c называют корнем $f(x)$ **кратности** k.

Следствие. *Всякий многочлен степени n имеет не более n корней, учитывая кратности.*

Основная теорема

Теорема (Gauss, 1799; Argand, 1806). *Всякий многочлен $f(x) \in \mathbb{C}[x]$ степени n имеет в $\mathbb{C}$*

(1) хотя бы один корень;
(2) в точности n корней, учитывая кратности.

До сих пор, то есть уже более 200 лет, это утверждение называют «основная теорема алгебры». Более точным было бы говорить «фундаментальная» (как на Западе). Исторически, предмет алгебры состоял во многом в решении уравнений, а значит, именно в отыскании корней многочленов. В частности, этим и объясняется особо почётный статус основной теоремы алгебры многочленов.

Доказательство. (2) Прямо следует из (1) по теореме Безу.

(1) Для нашего курса это трудно; чисто алгебраического доказательства всё ещё нет, и вероятно, нет вообще. Простое доказательство даётся средствами комплексного анализа в курсах ТФКП. Тем не менее, идеи доказательства наглядно представлены ниже. □

Разложение многочлена на неприводимые

Следствие. *Всякий многочлен $f(x) \in \mathbb{C}[x]$ раскладывается на линейные множители:*

$$f(x) = a_n(x - x_1)(x - x_2) \cdots (x - x_n),$$

где x_i есть корни $f(x)$, повторяющиеся согласно кратностям.

Доказательство. Берём имеющийся по основной теореме корень x_1 и по теореме Безу выделяем множитель $x - x_1$. Второй множитель имеет меньшую степень $(n-1)$ и корень x_2. Так один за другим выделим все множители.

Более формально это рассуждение можно оформить (хотя нужды нет) в виде индукции по степени раскладываемого многочлена. □

Неприводимым называют многочлен без собственных делителей и оттого не разложимый на множители меньших степеней. В алгебре многочленов это аналог простого числа. Полезно иметь описание всех неприводимых многочленов в $\mathbb{F}[x]$. Решения этой задачи для различных $\mathbb{F}$ разительно отличаются, поскольку неприводимость конкретного многочлена сильно зависит от того, какие коэффициенты вообще допускаются при разложении; поэтому говорят о неприводимости, например, над $\mathbb{C}$, над $\mathbb{R}$, над $\mathbb{Q}$. Основная теорема алгебры многочленов, вместе с теоремой Безу, решает вопрос над $\mathbb{C}$ и над $\mathbb{R}$.

Следствие. *Все неприводимые над $\mathbb{C}$ многочлены в $\mathbb{C}[x]$ — линейные.*

Лемма. *Для всякого многочлена $f(x)$ с вещественными коэффициентами свойство $\overline{f(z)} = f(\bar{z})$ выполнено для всех $z \in \mathbb{C}$.*

Доказательство. Пользуемся тем, что сопряжение уважает операции:

$$\begin{aligned}
\overline{f(z)} &= \overline{a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n} \\
&= \overline{a_0} + \overline{a_1 z} + \overline{a_2 z^2} + \cdots + \overline{a_n z^n} \\
&= \overline{a_0} + \overline{a_1}\,\bar{z} + \overline{a_2}\,\overline{z^2} + \cdots + \overline{a_n}\,\overline{z^n} \\
&= a_0 + a_1 \bar{z} + a_2 \bar{z}^2 + \cdots + a_n \bar{z}^n \\
&= f(\bar{z}).
\end{aligned}$$

На последнем шаге $\overline{a_k} = a_k$, ибо $a_k \in \mathbb{R}$. □

Следствие. *Все неприводимые над $\mathbb{R}$ многочлены в $\mathbb{R}[x]$ — линейные или квадратичные с отрицательным дискриминантом.*

Доказательство. Разложим многочлен $f(x) \in \mathbb{R}[x]$ на линейные множители над $\mathbb{C}$. Вещественные корни приносят линейные множители над $\mathbb{R}$. Поскольку $\overline{f(z)} = f(\bar{z})$, все корни вне $\mathbb{R}$ появляются сопряжёнными парами z, $\bar{z}$, так что в комплексном разложении присутствует $(x-z)(x-\bar{z})$. Однако для каждого $z \in \mathbb{C} \setminus \mathbb{R}$ многочлен

$$(x-z)(x-\bar{z}) = x^2 - x(z+\bar{z}) + z\bar{z} = x^2 - 2x\operatorname{Re} z + |z|^2$$

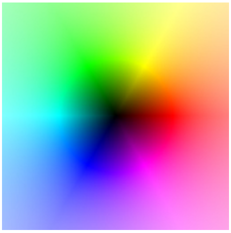
есть вещественный квадратичный с дискриминантом $-(2\operatorname{Im} z)^2$. □

Не известно никакого простого описания неприводимых многочленов над $\mathbb{Q}$; имеются таковые любой степени, например, $x^n - 2$.

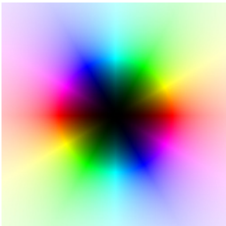
2.5. Изображение функций комплексной переменной

Желание нарисовать график даже простейшей комплекснозначной функции одной комплексной переменной наталкивается на серьёзное препятствие: необходимость рисовать в четырёх измерениях! Однако есть очень красивое решение этой проблемы.

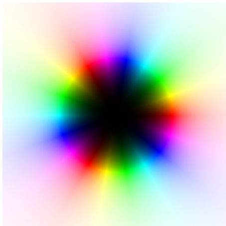
Два имеющихся у нас измерения отдадим независимой переменной, а значениям функции сопоставим вариации цвета. Именно, представим $|f(z)|$ и $\arg f(z)$ соответственно яркостью и оттенком точки z. Идею демонстрирует картинка для $f(z) = z$; на остальных картинках применено такое же соответствие между значением $f(z)$ и цветом.



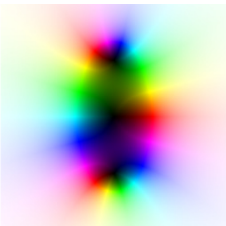
$f(z) = z$



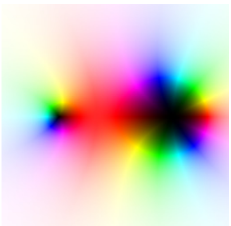
$f(z) = z^2$



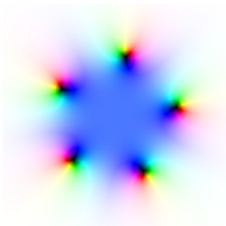
$f(z) = z^3$



$f(z) = z^3 + z$



$f(z) = (z+1)(z-1)^2$

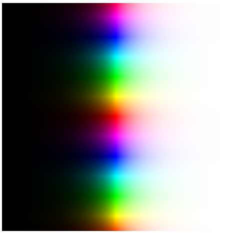


$f(z) = z^5 - i - 1$

Этот приём эффективен и для других функций, и здесь уже велик соблазн совершить познавательный экскурс в комплексный анализ. Но прежде сто́ит пояснить на картинках идею доказательства того, что всякий комплексный многочлен степени n имеет ровно n комплексных корней, с учётом кратностей.

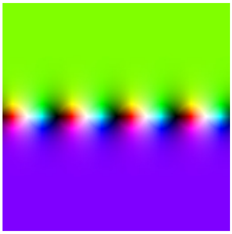
В комплексной плоскости с изображением многочлена $f(z)$ возьмём контур и обратим внимание на изменение $\arg f(z)$ при обходе контура, что легко отследить по изменению оттенка. Принято всегда обходить контур так, чтобы охватываемая им область была слева по ходу. Примеры показывают, что приращение $\arg f(z)$ при однократном обходе равно $2\pi i \cdot N$, где N есть сумма кратностей корней $f(z)$ внутри контура. Поскольку число корней конечно, в качестве контура можно взять окружность настолько большого радиуса, что по сравнению с ним все корни будут очень близки к нулю. Картинка $f(z)$, вмещающая контур, будет не отличима от картинки функции z^n, где $n = \deg f(z)$, а приращение $\arg f(z)$ будет равно $2\pi i \cdot n$. Значит, сумма кратностей всех комплексных корней многочлена всегда равна его степени.

пояснения к сло
картинкам?

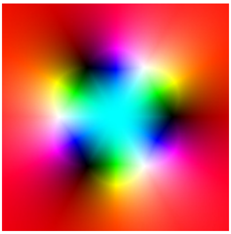


$f(z) = e^{3z}$

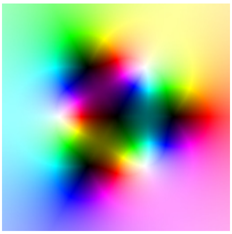
$f(z) = \cos 3z$



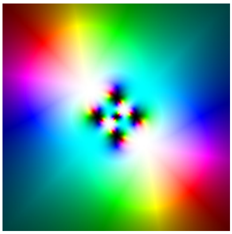
$f(z) = \operatorname{tg} 3z$



$f(z) = \dfrac{z^3 - 1}{z^3 + 1}$



$f(z) = z\dfrac{z^3 - 1}{z^3 + 1}$



тетраэдрическая дробь Клейна

Глава 3. ФИГУРЫ ВТОРОГО ПОРЯДКА

3.1. Эллипсы, параболы и гиперболы

Фигуры и уравнения

Лекция 6
07.10.19

Аналитическая геометрия — область геометрии, в которой геометрические фигуры изучаются с помощью линейной алгебры на основе применения координат. Приложения к геометрии более сложной алгебры относят в область **алгебраической геометрии**, а приложения математического анализа — в область **дифференциальной геометрии**. Эти предметы значительно превосходят аналитическую геометрию по своей трудности и объёму; достаточно сказать, что оба они в числе активно развивающихся областей современной математической науки.

Традиционно выделяют две задачи аналитической геометрии:

(1) фигура $\rightsquigarrow$ уравнения (**задание** фигур);
(2) уравнения $\rightsquigarrow$ фигура (**исследование** фигур).

Наши цели при изучении аналитической геометрии весьма скромны:

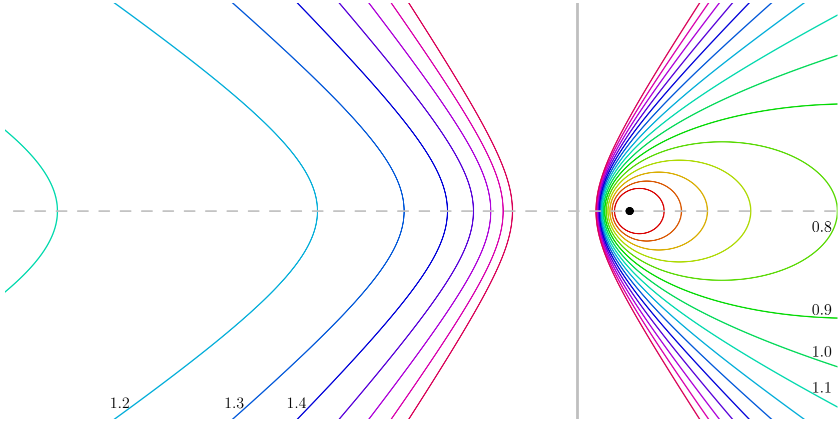
Уравнения	*Фигуры*	
	на плоскости	*в пространстве*
линейные	прямые	плоскости и прямые
квадратичные	коники	квадрики

Итак, линии и поверхности первого порядка мы изучали в предыдущей главе, а в этой займёмся линиями второго порядка, оставив поверхности на конец семестра. Важнейшие линии второго порядка — эллипсы (и окружности как их частный или предельный случай), параболы и гиперболы. Для получения полного списка к этим трём типам нужно добавить шесть различных типов распадения или вырождения, проявляющихся как пара прямых, прямая, точка или пустое множество.

Уравнение в полярных координатах

Зафиксируем на плоскости прямую δ (**директрису**) и не лежащую на ней точку F (**фокус**). Выберем число $e > 0$ (**эксцентриситет**) и рассмотрим геометрическое место точек X, расстояния от каждой из которых до δ и F связаны соотношением $d(X, F) = e \cdot d(X, \delta)$.

Из определения несложно вывести уравнение этой фигуры в полярных координатах (φ, ρ), помещая полюс в фокус и направляя ось прочь



от директрисы:

$$\rho = \frac{ep}{1 - e\cos\varphi}, \qquad (☼)$$

где p есть расстояние от фокуса до директрисы. Значение эксцентриситета определяет тип и форму полученной линии.

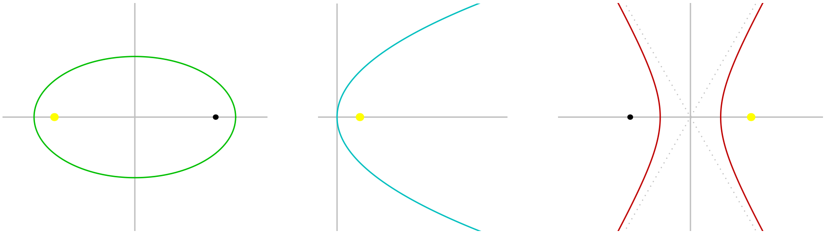
Пример. Именно этими формами (если пренебречь различными малыми возмущениями) обладают орбиты планет и траектории заряженных частиц в центрально-симметричном электрическом поле.

Канонические уравнения

В прямоугольной системе координат $x = \rho\cos\varphi$, $y = \rho\sin\varphi$ уравнение (☼) даст $x^2 + y^2 = e^2(x+p)^2$, но это не та запись, которую принято пропагандировать. Дальше обычно ведут анализ отдельно для каждого типа, переносом начала переходя к **каноническим** уравнениям. В них удобны новые параметры a и b, кроме случая параболы. Ось Oy канонической системы координат не совпадает с директрисой.

Фигура	Эллипс	Парабола	Гипербола
Эксцентриситет	$e < 1$	$e = 1$	$e > 1$
Каноническое уравнение	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$	$y^2 - 2px = 0$	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$

При стремлении эксцентриситета к нулю эллипс превращается в окружность; тогда в каноническом уравнении будет $a = b$. В полярных координатах уравнением окружности служит $\rho = \mathrm{const}$.

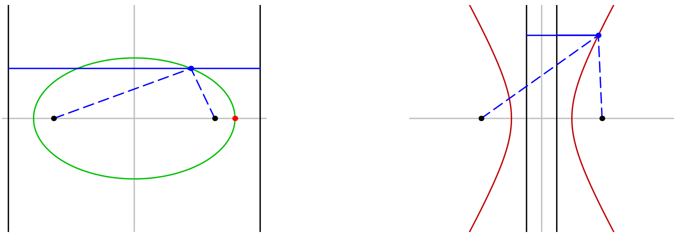


Фокус F лежит на оси Ox, а в случаях эллипса и гиперболы канонические уравнения неизменны при отражении относительно обеих осей. Следовательно, эти фигуры имеют и по второму фокусу.

Фокальные свойства

Теорема. *Для каждой точки P на эллипсе или гиперболе с фокусами F_1 и F_2 расстояния $d_i = d(P, F_i)$ до фокусов связаны следующими формулами:*

(1) *для эллипса $d_1 + d_2 = 2a$, независимо от точки P;*
(2) *для гиперболы $d_1 - d_2 = \pm 2a$, независимо от точки P, причём разные знаки соответствуют двум ветвям гиперболы.*

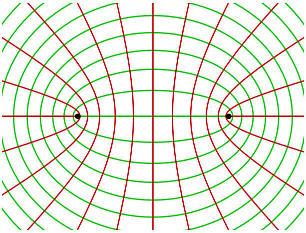


Доказательство. (1) Директрисы параллельны друг другу, поэтому сумма расстояний от P до них постоянна, а по изначальному определению эллипса она равна $(d_1 + d_2)/e$. Легко увидеть, что она же равна $2a$, т. е. **большой оси**, если выбрать P в вершине эллипса.

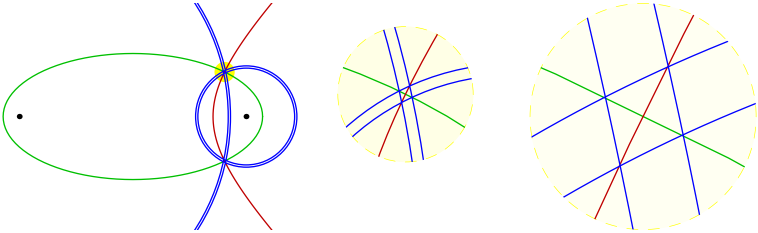
(2) Аналогично. Разность возникает, потому что обе директрисы лежат между двумя ветвями гиперболы. □

Ортогональность софокусных семейств

Теорема. *Эллипс и гипербола, у которых совпадают оба фокуса, пересекаются ортогонально.*



Доказательство. Обозначим через $d_i = d(P, F_i)$ расстояния от общей точки P эллипса и одной ветви гиперболы до их общих фокусов F_i. Рассмотрим в микроскоп вблизи точки P пары окружностей с центрами в фокусах и радиусами $d_i \pm \varepsilon$ для очень маленького $\varepsilon > 0$.



При достаточно большом увеличении эти окружности не отличимы от двух пар параллельных прямых, разделённых расстоянием 2ε, а точки их пересечения образуют крошечный ромбик. Эллипс же и гипербола, благодаря своим фокальным свойствам, при этом не отличимы от диагоналей ромбика. При стремлении $\varepsilon \to 0$ диагонали дают в пределе касательные к этим линиям. □

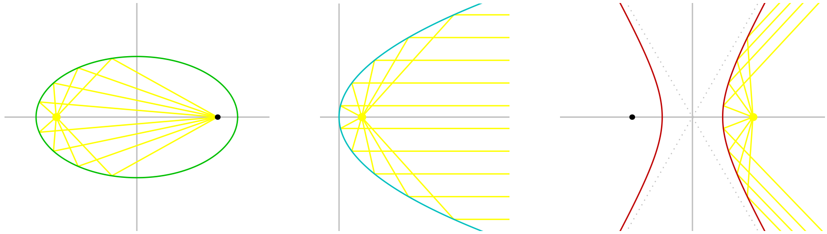
Оптические свойства

Теорема. *Поместим источник света в фокус эллипса, параболы, или гиперболы.*

(1) *Отразившись от эллипса, лучи пройдут через другой фокус.*
(2) *Отразившись от параболы, лучи станут параллельны её оси.*
(3) *Отразившись от гиперболы, лучи будут казаться выпущенными из другого фокуса.*

Доказательство. (1) Выберем на эллипсе точку P и проведём в ней касательную. Всякая точка $K \neq P$ касательной лежит снаружи эллипса, а потому — на софокусном эллипсе с большей большой полуосью, так что

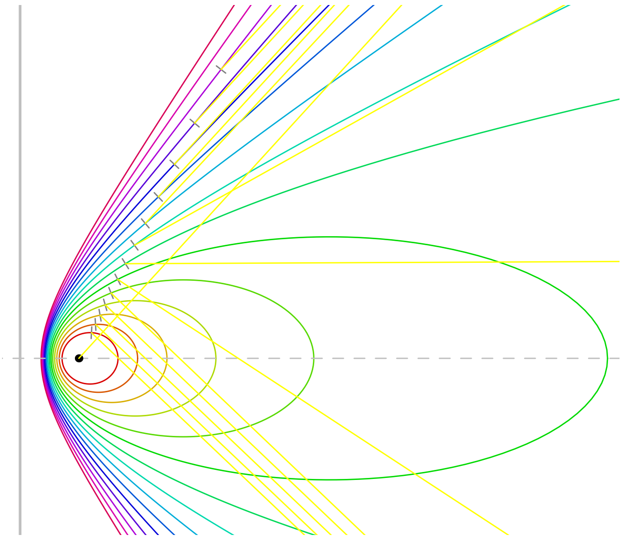
$$d(K, F_1) + d(K, F_2) > d(P, F_1) + d(P, F_2).$$



Теперь забудем про эллипс, оставив только его фокусы и касательную. Из полученного неравенства геометрически выводится (упражнение), что среди точек этой прямой именно в точке P реализуется оптический закон «угол падения равен углу отражения».

(3) Следствие ортогональности и оптического свойства эллипса.

(2) Вместе с параболой рассмотрим эллипсы и гиперболы (❂) с теми же фокусом и директрисой. Выпустим луч из фокуса. Если ещё до параболы он отразится от какого-то эллипса, ему суждено пересечь ось во втором фокусе. Если луч пройдёт сквозь прозрачную параболу, но отразится от какой-то гиперболы, то он продолжит удаляться от оси. Что остаётся делать лучу, если он отразится именно от параболы? □



Конические сечения

3.2. Общее уравнение линии второго порядка

Компактная запись общего уравнения

Определение. **Линией второго порядка** называют фигуру на плоскости $\mathbb{R}^2$, уравнение которой задаётся квадратичным многочленом:

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0,$$

или в индексных обозначениях,

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_1x + 2a_2y + a = 0.$$

Аналогично, **поверхностью второго порядка** называют фигуру в пространстве $\mathbb{R}^3$, уравнение которой задаётся квадратичным многочленом:

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz + 2a_1x + 2a_2y + 2a_3z + a = 0.$$

Выбор таких обозначений для коэффициентов, вместе с двойками, связан с тем, что эти многочлены можно переписать в виде матричных произведений. Для линии второго порядка получим общее уравнение

$$\left[\begin{array}{cc|c} x & y & 1 \end{array}\right] \cdot \left[\begin{array}{cc|c} a_{11} & a_{12} & a_1 \\ a_{12} & a_{22} & a_2 \\ \hline a_1 & a_2 & a \end{array}\right] \cdot \left[\begin{array}{c} x \\ y \\ \hline 1 \end{array}\right] = 0,$$

где матрицу коэффициентов можно выбрать *симметричной*: $A^\top = A$. Аналогично для поверхности, но с 4×4 матрицей коэффициентов.

Обозначим одной буквой каждый блок в предыдущей записи:

$$\left[\begin{array}{c|c} X^\top & 1 \end{array}\right] \cdot \left[\begin{array}{c|c} A_2 & A_1 \\ \hline A_1^\top & A_0 \end{array}\right] \cdot \left[\begin{array}{c} X \\ \hline 1 \end{array}\right] = 0.$$

Этот более компактный вариант пригоден и для линий, и для поверхностей, и даже для многомерных квадратичных гиперповерхностей. Раскрывая матричные произведения, приведём его к

$$X^\top A_2\, X + X^\top A_1 + A_1^\top X + A_0 = 0,$$

причём слагаемые $X^\top A_1$ и $A_1^\top X$ всегда совпадают. Наконец, если расширить столбец переменных, включив туда константу 1, то получится самая компактная запись общего уравнения квадрики:

$$X^\top A\, X = 0.$$

Задача классификации

Актуальна следующая геометрическая задача: *определить вид и расположение линии или поверхности второго порядка, заданной общим уравнением.* Для этого нужно по общему уравнению найти каноническое и связь канонической системы координат с исходной.

В зависимости от того, сколько информации о квадрике необходимо узнать, применяются различные методы. Наиболее резок контраст между теми, где система координат подвергается лишь переносам и поворотам, и теми, где вместе с переносами допускаются произвольные линейные преобразования. Для второго варианта достаточно помимо поворотов позволять растяжение вдоль одной из осей. Соответственно говорят о **метрической** и **аффинной** классификациях линий и поверхностей второго порядка.

При приведении к каноническому виду допускается только изменение направления осей и начала координат с сохранением всех углов и длин. Поэтому каноническая и исходная системы координат связаны поворотом и переносом, причём: при помощи поворота из уравнения всегда удаётся убрать произведения разных переменных; при помощи переноса удаётся сократить количество неквадратичных слагаемых до одного. Благодаря всегда присутствующей у этих линий осевой симметрии, трудности из-за ориентации в этих вопросах не возникают.

Метрическая классификация

Теорема. *Поворотом и переносом общее уравнение линии второго порядка приводится к одному из пяти видов:*

Класс	*Уравнение*
2+	$\lambda_1 x^2 + \lambda_2 y^2 + p = 0$
2	$\lambda_1 x^2 + \lambda_2 y^2 = 0$
2−	$\lambda_1 x^2 + 2py = 0$
1+	$\lambda_1 x^2 + p = 0$
1	$\lambda_1 x^2 = 0$

На практике можно сократить вычисления, делая перенос вперёд поворота и прибегая к дополнительным таблицам с различными случаями. Обоснование этого метода существенно длиннее.

Для поверхностей таблица уравнений содержит те же случаи, но появляются ещё три (самые интересные):

Класс	*Уравнение*
3+	$\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + p = 0$
3	$\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 = 0$
3−	$\lambda_1 x^2 + \lambda_2 y^2 + 2pz = 0$

Доказательство. Поворот плоскости всегда имеет вид

$$\begin{cases} x = x' \cos\varphi - y' \sin\varphi, \\ y = x' \sin\varphi + y' \cos\varphi. \end{cases}$$

Подставляя эту связь в уравнение, получим много ненужных слагаемых, но потребуем равенства нулю коэффициента при $x'y'$. Выходит

$$-2A\cos\varphi\sin\varphi + 2B(\cos^2\varphi - \sin^2\varphi) + 2C\sin\varphi\cos\varphi = 0.$$

Пользуясь формулами двойного угла, при условии $B \neq 0$ находим

$$\operatorname{ctg} 2\varphi = \frac{A - C}{2B}.$$

Если же $B = 0$, то поворот системы координат и не нужен.

Теперь уравнение линии упростилось до

$$A'(x')^2 + C'(y')^2 + 2D'x' + 2E'y' + F' = 0,$$

причём хотя бы один из коэффициентов при квадратах ненулевой: иначе уравнение стало бы линейным и обратный поворот никак не сделал бы из него исходное квадратичное. Применяя нижеследующую лемму, избавляемся хотя бы от одного из линейных слагаемых $2D'$ и $2E'$. Останется только то из них, которое является старшим по своей переменной; это случай $A' = 0$ либо $C' = 0$. Тогда вторым переносом можно избавиться от постоянного слагаемого. □

Лемма. *Для каждого многочлена $a(t')$ есть такой сдвиг $t' = t + s$, что в $a(t)$ коэффициент, следующий за старшим, равен нулю.*

Доказательство. Подстановкой найдём условие на сдвиг:

$$\begin{aligned} a(t+s) &= a_n(t+s)^n + a_{n-1}(t+s)^{n-1} + (\text{младшие по } t+s) \\ &= a_n t^n + (na_n s + a_{n-1})t^{n-1} + (\text{младшие по } t). \end{aligned}$$

Следовательно, $s = -a_{n-1}/na_n$. □

Полученные в теореме классы уравнений могут распадаться в зависимости от знаков коэффициентов на несколько типов, указанных в следующей таблице. Умножение уравнения на ненулевую константу не меняет фигуры, поэтому у большинства типов есть два варианта комбинации знаков коэффициентов, один из которых включён в таблицу.

Класс и тип линии		*Знаки*		
		λ_1	λ_2	p
2+	Мнимый эллипс	+	+	+
	Гипербола	−	+	+
	Эллипс	−	−	+
1+	Параллельные мнимые прямые	+	0	+
	Параллельные прямые	−	0	+
2−	Парабола	+	0	+
		−	0	+
2	Мнимые прямые, пересекающиеся в вещественной точке	+	+	0
	Пересекающиеся прямые	−	+	0
1	Двойная прямая	+	0	0

Уравнение касательной

Выведем теперь уравнение касательной к линии второго порядка, применив к её уравнению в форме $X^\top A\,X = 0$ общий рецепт решения такой задачи (**линеаризацию**) для линии C.

Лекция 7
14.10.19

рисунок?

- Возьмём две близкие точки $P = (x_0, y_0)$ и $Q = (x_0 + \delta u, y_0 + \delta v)$ на линии — так что $u = x - x_0$ и $v = y - y_0$, а δ мало;
- подставим координаты Q в уравнение;
- выделим три группы слагаемых смотря по степени вхождения δ;
- группа слагаемых, не содержащих δ, равна нулю, ибо $P \in C$;
- группу слагаемых, содержащих высокие степени δ, отбросим;
- оставшаяся группа после деления на δ и даёт искомое уравнение:

$$\begin{aligned} 0 &= (X_0 + \delta U)^\top A\,(X_0 + \delta U) \\ &= \big[X_0^\top A\,X_0\big] + \delta\big[U^\top A\,X_0 + X_0^\top A\,U\big] + \delta^2\big[U^\top A\,U\big] \\ &\approx \delta\big[(X - X_0)^\top A\,X_0 + X_0^\top A\,(X - X_0)\big] \\ &= \delta\big[X^\top A\,X_0 + X_0^\top A\,X\big] - 2\delta\big[X_0^\top A\,X_0\big] \\ &= 2\delta\big[X_0^\top A\,X\big]. \end{aligned}$$

Оставшиеся слагаемые совпадают ввиду симметричности матрицы коэффициентов, $A = A^\top$. Итак, уравнение касательной к линии второго порядка $X^\top A\,X = 0$ в точке X_0 есть $X_0^\top A\,X = 0$. Красиво!

К этому же результату приводит рассмотрение x и y как функций от t и дифференцирование исходного уравнения по t, которым пользуются чаще. Вот популярные ответы в канонической системе координат:

Фигура	Эллипс	Парабола	Гипербола
Каноническое уравнение	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$	$y^2 - 2px = 0$	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$
Уравнение касательной	$\frac{x_0x}{a^2} + \frac{y_0y}{b^2} = 1$	$y_0y - p(x + x_0) = 0$	$\frac{x_0x}{a^2} - \frac{y_0y}{b^2} = 1$

Поскольку наше вычисление в матричной форме не зависит от количества переменных, полученное уравнение $X_0^\top A\,X = 0$ годится и для касательной плоскости к поверхности второго порядка.

3.3. Поверхности второго порядка

Познакомимся с поверхностями второго порядка, или **квадриками** для краткости, анализируя по очереди все классы канонических уравнений, аналогичные уравнениям линий, в зависимости от знаков их коэффициентов. Во всех случаях очень полезно рассматривать сечения квадрики плоскостями, параллельными координатным.

Класс 3+: эллипсоиды и гиперболоиды

Уравнение $\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + p = 0$ попадает в один из четырёх типов. Это эллипсоиды и гиперболоиды. Уравнение

$$\tfrac{x^2}{a^2} + \tfrac{y^2}{b^2} + \tfrac{z^2}{c^2} = 1$$

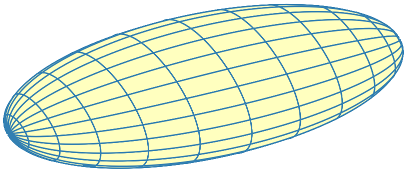
задаёт **эллипсоид**, причём можно считать, что $a \geqslant b \geqslant c$. Все его непустые и неодноточечные плоские сечения есть эллипсы. Различают четыре формы эллипсоидов:

- сфера при $a = b = c$;
- вытянутый эллипсоид вращения при $a > b = c$;
- сжатый эллипсоид вращения при $a = b > c$;
- трёхосный эллипсоид при $a > b > c$.

Вытянутый и сжатый эллипсоиды образованы вращением в пространстве эллипса

$$\tfrac{x^2}{a^2} + \tfrac{z^2}{c^2} = 1, \quad y = 0$$

вокруг его большой или малой оси соответственно.



Уравнение

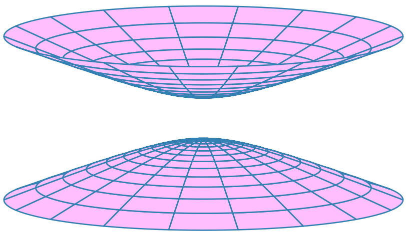
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = -1$$

задаёт «мнимый эллипсоид», но вещественных точек нет.

Уравнение

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1$$

задаёт **двуполостный гиперболоид**. Его сечения плоскостями $z = z_0$ пу-



сты при $|z_0| < c$, так что поверхность, как и плоская гипербола, состоит из двух частей, называемых **полостями**. При $|z_0| > c$ в сечениях будут эллипсы. Все сечения плоскостями $x = x_0$ и $y = y_0$ являются гиперболами. При $a = b$ двуполостный гиперболоид вращения образован вращением в пространстве гиперболы

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = -1, \quad y = 0$$

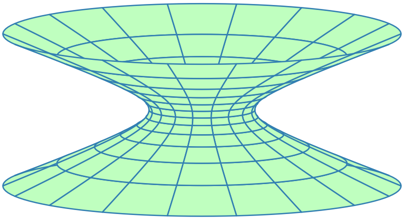
вокруг её действительной оси Oz. Дополнительное сжатие двуполостного гиперболоида вращения вдоль одной оси даёт двуполостный гиперболоид с $a \neq b$. Аналогично и в других случаях ниже.

Уравнение

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$$

задаёт **однополостный гиперболоид**. Все сечения плоскостями $z = z_0$ являются эллипсами, а плоскостями $x = x_0 \neq \pm a$ и $y = y_0 \neq \pm b$ — гиперболами, но парой пересекающихся прямых в оставшихся особых случаях. При $a = b$ однополостный гиперболоид вращения образован вращением в пространстве гиперболы

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1, \quad y = 0$$



вокруг её мнимой оси Oz. Интересно, что он также образован вращением любой из пары прямых

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 0, \quad \frac{y}{b} = 1$$

вокруг оси Oz.

Остальные варианты расстановки знаков в уравнении этого класса сводятся к уже рассмотренным умножением уравнения на -1. Таким же образом почти вдвое сокращается перечень типов в других классах.

Класс 3: конусы

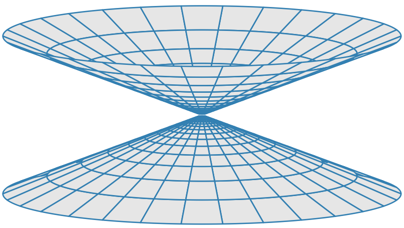
Уравнение $\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 = 0$ попадает в один из двух типов. Это конусы. Уравнение

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 0$$

задаёт «мнимый конус» с единственной вещественной точкой — вершиной в начале координат, а уравнение

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$$

задаёт **конус**. Все сечения конуса плоскостями $z = z_0$ при $z_0 \neq 0$ являются эллипсами, а плоскостями $x = x_0 \neq 0$ и $y = y_0 \neq 0$ — гиперболами. Сечения плоскостями $x = 0$ и $y = 0$ дают по паре пересекающихся в начале координат прямых. При $a = b$ конус образован вращением прямой



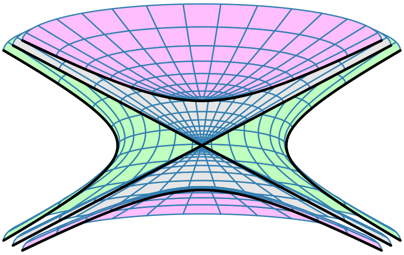
$$\frac{x}{a} - \frac{z}{c} = 0, \quad y = 0$$

вокруг оси Oz. Конус является объединением прямых (его **прямолинейных образующих**), проходящих через его вершину. Плоскость, параллельная прямолинейной образующей конуса и не содержащая её, пересекает конус по параболе.

Полезно рассмотреть конус совместно с двумя гиперболоидами

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = \pm 1.$$

Эту ситуацию описывают словами: данный конус является **асимптотическим конусом** обоих данных гиперболоидов.



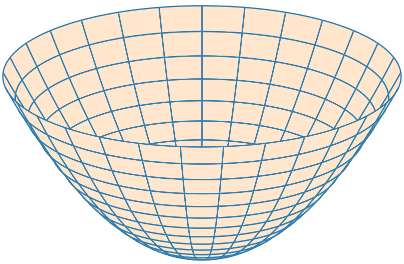
В качестве упражнения найдите плоскость, пересекающую гиперболоид по параболе.

Класс 3−: параболоиды

Уравнение $\lambda_1 x^2 + \lambda_2 y^2 + 2pz = 0$ попадает в один из двух типов. Это параболоиды. Уравнение

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = z$$

задаёт **эллиптический параболоид**. Сечения плоскостями $z = z_0 > 0$



являются эллипсами, а плоскостями $x = x_0$ и $y = y_0$ — параболами. При $a = b$ параболоид вращения образован вращением параболы

$$\frac{x^2}{a^2} = z, \quad y = 0$$

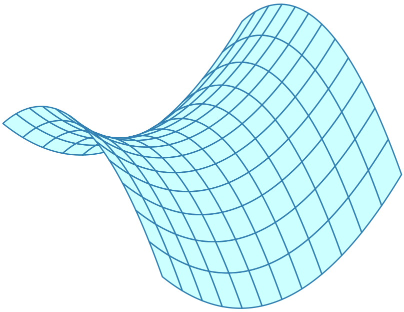
вокруг её оси Oz. Эллиптический параболоид также образован поступательным движением этой параболы так, что её вершина движется вдоль параболы

$$\frac{y^2}{b^2} = z, \quad x = 0.$$

Уравнение

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = z$$

задаёт **гиперболический параболоид**. Форма этой поверхности напоминает седло. Сечения плоскостями $z = z_0 \neq 0$ являются гиперболами,

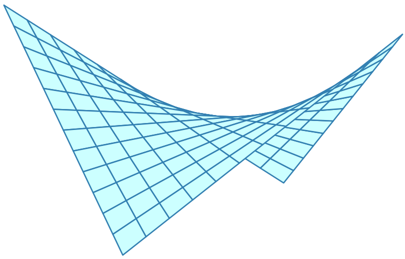


плоскостью $z = 0$ — парой прямых, пересекающихся в начале координат, а плоскостями $x = x_0$ и $y = y_0$ — параболами. Гиперболический параболоид образован поступательным движением параболы

$$\frac{x^2}{a^2} = z, \quad y = 0$$

так, что её вершина движется вдоль параболы

$$-\frac{y^2}{b^2} = z, \quad x = 0.$$



Интересно отметить, что гиперболический параболоид содержит два семейства прямых. Для простоты возьмём $a = b$ и тогда уравнения этих семейств будут

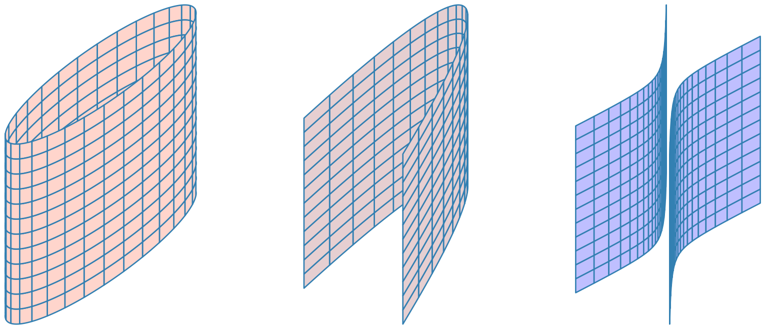
$$\begin{cases}\alpha(x-y)=\beta z,\\ \beta(x+y)=\alpha;\end{cases}\qquad \begin{cases}\alpha(x-y)=\beta,\\ \beta(x+y)=\alpha z.\end{cases}$$

Здесь α и β не равны нулю одновременно, а их отношение является параметром обоих семейств. Аналогично находятся два семейства прямых на однополостном гиперболоиде.

Остальные случаи расстановки знаков в уравнении этого класса сводятся к рассмотренным умножением уравнения на -1 и/или изменением направления оси Oz.

Низшие классы: цилиндры

Уравнения оставшихся классов не содержат третьей переменной. Задаваемая таким «неполным» уравнением квадрика получается из линии второго порядка в плоскости $z=0$ с тем же уравнением её поступательным движением вдоль оси Oz и называется **цилиндром**. Можно сказать и иначе: цилиндр заметается прямой, параллельной оси Oz и двигающейся через точки плоской линии.



Если линия есть эллипс, парабола, или гипербола, то заметённый цилиндр называют соответственно эллиптическим, параболическим, или гиперболическим. Есть ещё «мнимый эллиптический цилиндр» без вещественных точек.

В менее интересных случаях прочих типов линий второго порядка так получаются пары пересекающихся или параллельных плоскостей или «мнимых» плоскостей (причём две «мнимые» плоскости пересекаются по вещественной прямой), а также двойная плоскость.

Глава 4. НАЧАЛА ЛИНЕЙНОЙ АЛГЕБРЫ

4.1. Системы линейных уравнений 1

Начальная терминология

Займёмся общими системами m линейных уравнений

$$(\bigstar)\qquad \begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m, \end{cases}$$

с n неизвестными для произвольных m, $n \in \mathbb{Z}_{\geqslant 1}$. **Коэффициенты** a_{ij} и **свободные члены** b_j чаще всего являются вещественными числами. Встречаются задачи, где они комплексные, рациональные, либо ещё более хитрые, но поначалу это не существенно для нашей будущей теории. Поэтому будем обозначать через $\mathbb{F}$ **основную** числовую систему, в которой лежат значения всех известных и неизвестных букв.

Изучая линейные системы, удобно оставлять в тени неизвестные и выписывать лишь **основную** и **расширенную матрицы** системы:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots a_{1n} \\ \multicolumn{3}{c}{\dots\dots\dots\dots} \\ a_{m1} & a_{m2} & \dots a_{mn} \end{bmatrix}, \qquad [A|b] = \left[\begin{array}{ccc|c} a_{11} & a_{12} & \dots a_{1n} & b_1 \\ \multicolumn{4}{c}{\dots\dots\dots\dots\dots} \\ a_{m1} & a_{m2} & \dots a_{mn} & b_m \end{array}\right].$$

Решением системы $(\bigstar)$ называют список $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ элементов $\mathbb{F}$, подстановка которых вместо x_i превращает все уравнения системы $(\bigstar)$ в тождества. По количеству решений системы делятся на совместные и несовместные, на определённые и неопределённые:

<table>
<tr><th>Решения</th><th colspan="2">система</th></tr>
<tr><td>нет</td><td>несовместна</td><td rowspan="2">определена</td></tr>
<tr><td>одно</td><td rowspan="2">совместна</td></tr>
<tr><td>много</td><td>неопределена</td></tr>
</table>

Элементарные преобразования

Две линейные системы одинаковых размеров **эквивалентны**, если множества их решений одинаковы. Два простых преобразования переводят систему в ей эквивалентную:

(R1) умножение одного уравнения на ненулевой элемент $\alpha \in \mathbb{F}$;
(R2) прибавление к одному уравнению другого.

Комбинациями этих примитивов представимы преобразования:

(R2′) прибавление к одному уравнению другого, помноженного на любой элемент из $\mathbb{F}$;

(R3) перестановка пары уравнений местами.

Часто список **элементарных преобразований** составляют из преобразований (R1), (R2′) и (R3): это удобнее для применений на практике, в то время как простота (R1) и (R2) удобна в доказательствах.

Лемма. *Две линейные системы эквивалентны, если одна получается из другой конечной цепочкой элементарных преобразований.* □

Пример на метод исключения неизвестных

Чтобы перейти от данной системы к более простой, путём элементарных преобразований методично зануляют коэффициенты.

Пример. Для системы уравнений выписана расширенная матрица:

$$\begin{cases} -x_1 + x_3 - 3x_4 + 2x_5 = 1, \\ 2x_1 - x_3 + 5x_4 - x_5 = 3, \\ x_1 - 2x_3 + 4x_4 - 5x_5 = -6, \end{cases} \qquad \left[\begin{array}{rrrrr|r} -1 & 0 & 1 & -3 & 2 & 1 \\ 2 & 0 & -1 & 5 & -1 & 3 \\ 1 & 0 & -2 & 4 & -5 & -6 \end{array}\right].$$

Эта система элементарными преобразованиями…

$$\begin{cases} -x_1 + x_3 - 3x_4 + 2x_5 = 1, \\ 0x_1 + x_3 - x_4 + 3x_5 = 5, \\ x_1 - 2x_3 + 4x_4 - 5x_5 = -6, \end{cases} \qquad \left[\begin{array}{rrrrr|r} -1 & 0 & 1 & -3 & 2 & 1 \\ 0 & 0 & 1 & -1 & 3 & 5 \\ 1 & 0 & -2 & 4 & -5 & -6 \end{array}\right],$$

$$\begin{cases} -x_1 + x_3 - 3x_4 + 2x_5 = 1, \\ 0x_1 + x_3 - x_4 + 3x_5 = 5, \\ 0x_1 - x_3 + x_4 - 3x_5 = -5, \end{cases} \qquad \left[\begin{array}{rrrrr|r} -1 & 0 & 1 & -3 & 2 & 1 \\ 0 & 0 & 1 & -1 & 3 & 5 \\ 0 & 0 & -1 & 1 & -3 & -5 \end{array}\right],$$

$$\begin{cases} -x_1 + 0x_3 - 2x_4 - x_5 = -4, \\ 0x_1 + x_3 - x_4 + 3x_5 = 5, \\ 0x_1 + 0x_3 + 0x_4 + 0x_5 = 0, \end{cases} \qquad \left[\begin{array}{rrrrr|r} -1 & 0 & 0 & -2 & -1 & -4 \\ 0 & 0 & 1 & -1 & 3 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

приводится к виду

$$\begin{cases} x_1 + 0x_3 + 2x_4 + x_5 = 4, \\ 0x_1 + x_3 - x_4 + 3x_5 = 5, \\ 0x_1 + 0x_3 + 0x_4 + 0x_5 = 0, \end{cases} \qquad \left[\begin{array}{rrrrr|r} 1 & 0 & 0 & 2 & 1 & 4 \\ 0 & 0 & 1 & -1 & 3 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Общее решение исходной системы можно записать в виде

$$\begin{cases} x_1 = 4 - 2x_4 - x_5, \\ x_3 = 5 + x_4 - 3x_5, \\ x_2, x_4, x_5 \in \mathbb{R} \text{ — произвольные.} \end{cases}$$

Значит, система совместна и неопределена; при каждом наборе значений **параметров** x_2, x_4, x_5 имеется одно решение.

Ступенчатые матрицы

Матрицу коэффициентов назовём **ступенчатой**, когда:

(1) первый слева ненулевой элемент каждой строки есть единица, называемая **главной**;

(2) столбец, содержащий главную единицу, в остальном нулевой;

(3) главные единицы уходят направо и вниз:

$$\begin{bmatrix} \dots 0 & 1 & * & * \dots 0 & * & * \dots * & 0 & * \\ \dots 0 & 0 & 0 & 0 \dots 1 & * & * \dots * & 0 & * \\ \multicolumn{8}{c}{\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots} \\ \dots 0 & 0 & 0 & 0 \dots 0 & 0 & 0 \dots 0 & 1 & * \\ \dots 0 & 0 & 0 & 0 \dots 0 & 0 & 0 \dots 0 & 0 & 0 \end{bmatrix}.$$

Как и в рассмотренном выше примере, решения системы ступенчатого вида выписывают непосредственно по её расширенной матрице. При этом множества решений у любой пары различных систем ступенчатого вида обязательно получаются различны.

Общий алгоритм

Теорема. *Всякая система линейных уравнений эквивалентна (единственной) системе ступенчатого вида.*

Алгоритм. Доказательством существования служит конкретный метод приведения произвольной матрицы к ступенчатой. Результатом промежуточного этапа является матрица

$$\begin{array}{ccc} \begin{array}{c} \text{готовый} \\ \text{блок} \\ \hline \\ \text{нулевой} \\ \text{блок} \end{array} & \left[\begin{array}{ccc|ccccc} 1 & 0 & * & * & * & * & * & * \\ 0 & 1 & * & * & * & * & * & * \\ \hline 0 & 0 & 0 & * & * & * & * & * \\ 0 & 0 & 0 & * & * & * & * & * \\ 0 & 0 & 0 & * & * & * & * & * \end{array}\right] & \begin{array}{c} \text{пассивный} \\ \text{блок} \\ \hline \\ \text{активный} \\ \text{блок} \end{array} \end{array}$$

блочного вида. Её левый верхний блок есть ступенчатая матрица без нулевых строк, а правые блоки предстоит вычищать.

В активном блоке ищется первый ненулевой элемент:

$$\left[\begin{array}{ccc|ccccc} 1 & 0 & * & * & * & * & * & * \\ 0 & 1 & * & * & * & * & * & * \\ \hline 0 & 0 & 0 & 0 & \downarrow & * & * & * \\ 0 & 0 & 0 & \downarrow & \bullet & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \end{array}\right].$$

Если до его обнаружения в активном блоке найдены нулевые столбцы, то граница блоков (мысленно) передвигается вправо от них. Притом, если весь активный блок оказывается нулевым, то ступенчатый вид получен. Иначе, первый ненулевой элемент становится **ведущим**:

$$\left[\begin{array}{cccc|cccc} 1 & 0 & * & * & * & * & * & * \\ 0 & 1 & * & * & * & * & * & * \\ \hline 0 & 0 & 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & \bullet & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \end{array}\right].$$

Перестановками строк ведущий элемент помещаем в верхний левый угол активного блока; делением ведущей строки на ведущий элемент превращаем его в будущую главную единицу:

$$\left[\begin{array}{cccc|cccc} 1 & 0 & * & * & * & * & * & * \\ 0 & 1 & * & * & * & * & * & * \\ \hline 0 & 0 & 0 & 0 & 1 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \end{array}\right].$$

Вычитанием ведущей строки, помноженной на соответствующие элементы ведущего столбца, из остальных строк, зануляем все элементы этого столбца, кроме остающейся главной единицы:

$$\left[\begin{array}{cccc|cccc} 1 & 0 & * & * & * & * & * & * \\ 0 & 1 & * & * & * & * & * & * \\ \hline 0 & 0 & 0 & 0 & 1 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & * & * & * \end{array}\right].$$

Последнюю операцию проводим также и с верхними строками:

$$\left[\begin{array}{ccccc|ccc} 1 & 0 & * & * & 0 & * & * & * \\ 0 & 1 & * & * & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 1 & * & * & * \\ \hline 0 & 0 & 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & * & * & * \end{array}\right].$$

Сдвигая границу блоков вправо от ведущего столбца, получаем матрицу, пригодную для следующей итерации метода исключения. $\square$

Удивительно, что этот метод знали и применяли в Китае ещё во II веке до нашей эры, хотя европейская традиция приписывает его Гауссу (начало XIX века).

Замечание. В некоторых учебниках пропускают последний описанный шаг, а ступенчатый вид соответственно определяют, не требуя зануления элементов матрицы над главными единицами. Это допустимо для простых задач, поскольку обратную подстановку, дающую решение системы по ступенчатому виду её расширенной матрицы, легко выполнить и в таком ослабленном варианте.

Преимущество взятого здесь сильного варианта в том, что он позволяет утверждать единственность ступенчатого вида любой матрицы.

Следствие. (1) *Число параметров, описывающих множество решений совместной системы, равно числу неглавных столбцов.*

(2) *Система определена* $\iff$ *все столбцы главные.*

(3) *Система совместна* $\iff$ *каждой нулевой строке соответствует нулевой свободный член.*

Понятие о ранге

Наличие нулевых строк в расширенной ступенчатой матрице сигнализирует, что исходная система избыточна: лишние уравнения можно отбросить, не изменив множество решений. Какие именно уравнения излишни, сказать нелегко, но *количество* существенных уравнений видно: оно равно количеству главных единиц.

Рабочее определение. **Рангом** матрицы A назовём количество ненулевых строк в ступенчатом виде, к которому A приводится.

Рабочим это определение является сразу в двух смыслах:

(1) вскоре его заменят «настоящие» определения ранга, коих будет целых три;

(2) даже после грядущей замены, практически найти ранг матрицы обычно быстрее всего именно приведением её к ступенчатому виду.

Следствие (критерий совместности)**.** *Линейная система совместна* $\iff$ *ранги основной и расширенной матриц равны.*

Следствие. *Совместная система имеет единственное решение* $\iff$ *число неизвестных равно рангу системы.*

4.2. Линейные пространства строк и столбцов

Линейные пространства

Работая с векторами в фиксированной системе координат, мы оперируем над столбцами их координат; эти столбцы мы можем складывать и умножать на скаляры. Решая линейные системы элементарными преобразованиями, мы оперируем над строками расширенной матрицы; эти строки мы можем складывать и умножать на числа (тоже называемые скалярами). Дальнейшее изучение математики и физики постоянно будет сталкивать студента с объектами самой разной природы, которые можно складывать и умножать на скаляры. Как правило, при этом выполнены 8 привычных свойств сложения и умножения, перечисленных в самом первом разделе лекций.

Множество $\mathcal{L}$ с двумя такими операциями называют **линейным** или **векторным пространством**. Элементы любого линейного пространства называют векторами. Это довольно плохо, ибо вызывает конфликт с физическим пониманием вектора как направленного отрезка, но такое употребление в математике прижилось, а лучшего слова пока нет.

Познакомимся с основополагающими понятиями линейной алгебры на примере пространства вещественных строк длины n и пространства вещественных столбцов высоты n. Обозначим оба через $\mathbb{R}^n$, потому что часто нет разницы, о строках речь или о столбцах.

Комбинации и (не)зависимость

Зафиксируем множество векторов $\mathcal{X} = \{X_1, X_2, \ldots, X_k\} \subset \mathbb{R}^n$. Их **линейной комбинацией** с коэффициентами $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}$ называют

$$X = \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k.$$

Линейную комбинацию, в которой не все коэффициенты равны нулю, называют **нетривиальной**.

Понятия линейной (не)зависимости множества векторов линейного пространства вводят в точности так же, как и ранее для трёхмерного пространства. Множество $\{X_1, \ldots, X_k\}$ **линейно зависимо**, если найдутся такие коэффициенты $\alpha_1, \ldots, \alpha_k$, не все равные нулю, что линейная комбинация $\alpha_1 X_1 + \ldots + \alpha_k X_k$ равна нулевому вектору. Линейная зависимость множеств из одного, двух, или трёх векторов это: (1) равенство нулю; (2) коллинеарность; (3) компланарность.

В противном случае нулевой вектор нельзя представить нетривиальной линейной комбинацией векторов из данного множества, которое тогда называют **линейно независимым**. Часто используется равносильная переформулировка, говорящая, что нулевая линейная комбинация векторов линейно независимого множества $\{X_1, \ldots, X_k\}$ обязательно тривиальна:

$$\alpha_1 X_1 + \ldots + \alpha_k X_k = \mathbf{0} \implies \alpha_1 = 0 \wedge \ \ldots \wedge \ \alpha_k = 0.$$

Пример. В пространстве $\mathbb{R}^4$ возьмём строки

$$Z_1 = [1, -1, 0, 0], \quad Z_2 = [0, 1, -1, 0], \quad Z_3 = [0, 0, 1, -1], \quad Z_4 = [-1, 0, 0, 1].$$

Тогда $\{Z_1, Z_2, Z_3\}$ линейно независимо, а $\{Z_1, Z_2, Z_3, Z_4\}$ линейно зависимо: $Z_1 + Z_2 + Z_3 + Z_4 = 0$.

Линейная оболочка

Множество $\langle \mathcal{X} \rangle$ всевозможных линейных комбинаций векторов из $\mathcal{X}$ называют **линейной оболочкой** множества $\mathcal{X}$. Поскольку

$$X, Y \in \langle \mathcal{X} \rangle \implies \alpha X + \beta Y \in \langle \mathcal{X} \rangle,$$

каждая линейная оболочка сама является линейным пространством: в ней можно выполнять все действия с векторами, не выходя в объемлющее пространство $\mathbb{R}^n$. Можно сказать, что $\langle \mathcal{X} \rangle$ является **подпространством** в $\mathbb{R}^n$.

Пример. Перечислим все подпространства обычного трёхмерного пространства $\mathbb{R}^3$. Чтобы отождествлять точки с векторами, необходимо сперва выбрать начало координат $\mathbf{0}$.

(0) Линейная оболочка пустого множества векторов является наименьшим подпространством: $\langle \varnothing \rangle = \{\mathbf{0}\}$. Каждое подпространство его включает. Кроме того, $\langle \mathbf{0} \rangle = \{\mathbf{0}\}$.

(1) Линейная оболочка любого ненулевого вектора — подпространство. Это прямая, проходящая через $\mathbf{0}$.

(2) Линейная оболочка любой пары векторов — подпространство. Если векторы не коллинеарны, то это плоскость, проходящая через $\mathbf{0}$.

(3) Линейная оболочка любой тройки векторов — подпространство. Если векторы не компланарны, то получим всё исходное пространство. В других случаях получаются плоскости и прямые через $\mathbf{0}$.

Все бо́льшие множества векторов в $\mathbb{R}^3$ линейно зависимы и поэтому не дают линейных оболочек, отличных от уже перечисленных.

Появление и исчезновение зависимости

Лемма. *Для любого множества векторов* $\mathcal{X} = \{X_1, X_2, \ldots, X_k\}$ *произвольного линейного пространства:*

(1) *если* $\mathcal{Y} \subset \mathcal{X}$ *и* $\mathcal{Y}$ *линейно зависимо, то* $\mathcal{X}$ *линейно зависимо;*
(2) *если* $\mathcal{Y} \subset \mathcal{X}$ *и* $\mathcal{X}$ *линейно независимо, то* $\mathcal{Y}$ *линейно независимо;*
(3) *хотя бы один из* $X_i \in \mathcal{X}$ *есть линейная комбинация остальных* $\iff$ *множество* $\mathcal{X}$ *линейно зависимо;*
(4) *если* $\mathcal{X}$ *линейно независимо, то вектор* $Z \in \langle \mathcal{X} \rangle \iff$ *множество* $\mathcal{X} \cup \{Z\}$ *линейно зависимо.*

Доказательство. Это простое, но важное упражнение на логику.

(1) Зависимость не может исчезнуть при увеличении множества.

(2) Зависимость не может появиться при уменьшении множества.

(3) Вычитая вектор из его выражения через остальные, получим нулевую нетривиальную линейную комбинацию. Наоборот, имея оную, можно выразить хотя бы один из входящих векторов.

(4) Вывод слева направо верен для любого $\mathcal{X}$. Наоборот, ввиду линейной независимости $\mathcal{X}$, нулевая нетривиальная линейная комбинация векторов из $\mathcal{X} \cup \{Z\}$ невозможна без участия Z с ненулевым коэффициентом, а тогда Z выражается через векторы из $\mathcal{X}$. □

Базис (под)пространства

Определение. **Базисом** линейного подпространства $\mathcal{L} \subseteq \mathbb{R}^n$ называется такое (упорядоченное) множество векторов $\mathcal{X} = \{X_1, X_2, \ldots, X_k\}$, что:

(1) множество $\mathcal{X}$ линейно независимо и
(2) пространство $\mathcal{L}$ совпадает с линейной оболочкой $\langle \mathcal{X} \rangle$.

Иначе говоря, каждый вектор $X \in \mathcal{L}$ записывается единственным образом в виде линейной комбинации

$$X = \alpha_1 X_1 + \ldots + \alpha_k X_k.$$

При этом числа $\alpha_1, \ldots, \alpha_k$ называют **координатами** X в базисе $\mathcal{X}$.

Пример. **Стандартным базисом** $\mathbb{R}^n$ называют $\{E_{(1)}, \ldots, E_{(n)}\}$, где

$$E_{(i)} = [0, \ldots, 0, 1, 0, \ldots, 0] \quad \text{(единица написана в позиции } i\text{)}.$$

Пример. Любые три различных вектора из $\mathcal{Z} = \{Z_1, Z_2, Z_3, Z_4\}$, рассмотренного выше, образуют базис $\langle \mathcal{Z} \rangle$. Базисов $\langle \mathcal{Z} \rangle$ много: пригоден любой набор из трёх линейно независимых векторов из $\langle \mathcal{Z} \rangle$.

Лемма (Steinitz). *Если* $\{X_1, \ldots, X_r\}$ *есть базис линейного пространства* $\mathcal{L}$ *и множество* $\mathcal{Y} = \{Y_1, \ldots, Y_s\} \subset \mathcal{L}$ *линейно независимо, то:*

(1) *$s \leqslant r$;*
(2) *можно дополнить $\mathcal{Y}$ до базиса $\mathcal{L}$.*

Доказательство. (1) Составим линейную комбинацию

$$\alpha_1 Y_1 + \ldots + \alpha_s Y_s = \mathbf{0}$$

с неизвестными коэффициентами α_i. Распишем в ней все Y_i по базису,

$$Y_i = y_{1i} X_1 + \ldots + y_{ri} X_r,$$

и, собрав коэффициенты при $X_1, \ldots, X_r$, получим систему из r однородных линейных уравнений

$$y_{k1}\alpha_1 + \ldots + y_{ks}\alpha_s = 0, \quad k = 1, \ldots, r,$$

с s неизвестными α_i. Если $s > r$, она всегда имеет ненулевое решение, что противоречит линейной независимости $\{Y_1, \ldots, Y_s\}$.

(2) Выбросим из списка $Y_1, \ldots, Y_s, X_1, \ldots, X_r$ один за другим все векторы, линейно выражающиеся через предыдущие. Линейная оболочка списка при этом неизменна и равна $\mathcal{L}$, а множество оставшихся векторов линейно независимо и содержит $\mathcal{Y}$. □

Теорема. *Каждое подпространство $\mathcal{S} \subseteq \mathbb{R}^n$ имеет конечный базис.*

Доказательство. Будем наращивать линейно независимое множество векторов, используя утверждение (4) первой леммы. Начав с пустого, далее добавляем к имеющемуся списку $\{X_1, \ldots, X_k\}$ любой вектор X_{k+1} из $\mathcal{S} \setminus \langle X_1, \ldots, X_k \rangle$, пока таковые есть. По второй лемме с $\mathcal{L} = \mathbb{R}^n$ мы сможем так набрать не более n векторов. Значит, при некотором $r \leqslant n$ рост остановится и мы получим искомое $\mathcal{S} = \langle X_1, \ldots, X_r \rangle$. □

Размерность

Следствие. *Все базисы $\mathcal{S}$ состоят из одинакового числа векторов.*

Доказательство. Если в одном базисе s векторов, а в другом r, то по второй лемме $s \leqslant r \leqslant s$, где или первый базис есть $\mathcal{Y}$, или второй. □

Определение. Число векторов в базисе линейного пространства $\mathcal{L} \subseteq \mathbb{R}^n$ называют его **размерностью** и обозначают $\dim \mathcal{L}$. Кроме того, число векторов в базисе линейной оболочки множества $\mathcal{X}$ называют **рангом** $\mathcal{X}$.

Следствие. *Для любого пространства $\mathcal{L} \subseteq \mathbb{R}^n$:*

(1) *если $\mathcal{S} \subseteq \mathcal{L}$ есть подпространство, то $\dim \mathcal{S} \leqslant \dim \mathcal{L}$;*
(2) *если при том $\dim \mathcal{S} = \dim \mathcal{L}$, то $\mathcal{S} = \mathcal{L}$.*

Доказательство. (1) Базис $\mathcal{S}$ можно дополнить до базиса $\mathcal{L}$.

(2) Базис $\mathcal{S}$ уже является базисом $\mathcal{L}$. □

4.3. Системы линейных уравнений 2

Ранг матрицы

Определение. **Рангом по строкам/столбцам** матрицы A называют ранг множества её строк/столбцов. Эти числа временно обозначим через $\operatorname{rk}_\text{г} A$ и $\operatorname{rk}_\text{в} A$. **Пространством строк/столбцов** матрицы A называют линейную оболочку множества её строк/столбцов.

Лемма. *Элементарные преобразования строк матрицы не меняют пространство строк; аналогично для столбцов.*

Доказательство. Достаточно проверить (R1) и (R2). Преобразования обоих типов выдают линейные комбинации, то есть новые строки являются элементами исходного пространства; поэтому пространство строк не увеличивается. Преобразования обратимы, поэтому пространство строк не уменьшается. □

Лемма. *У всякой матрицы ступенчатого вида:*

(1) *все ненулевые строки линейно независимы;*
(2) *все столбцы являются линейными комбинациями главных;*
(3) *ранги по строкам и по столбцам равны.*

Доказательство. Упражнение. □

Лемма. *Если матрицы A и Z связаны конечной цепочкой элементарных преобразований строк, то их ранги по столбцам равны.*

Доказательство. Считаем, что первые $r = \operatorname{rk}_\text{в} A$ столбцов матрицы A составляют базис пространства её столбцов: иначе можно прибегнуть к перенумерации столбцов. Каждый столбец $A^{(k)}$ запишется какой-то линейной комбинацией

$$A^{(k)} = \alpha_{k1}A^{(1)} + \ldots + \alpha_{kr}A^{(r)}.$$

Любое элементарное преобразование строк сохраняет это соотношение между столбцами, поэтому выполнено равенство

$$Z^{(k)} = \alpha_{k1}Z^{(1)} + \ldots + \alpha_{kr}Z^{(r)}.$$

Значит, $\operatorname{rk}_\text{в} A \geqslant \operatorname{rk}_\text{в} Z$. Поскольку элементарные преобразования обратимы, аналогично выводится обратное неравенство. □

Теорема. *Ранги любой матрицы по строкам и по столбцам равны.*

Определение. Вот это число и называется **рангом** матрицы.

Доказательство. Возьмём произвольную прямоугольную матрицу A, а также её ступенчатый вид A', и проверим цепочку равенств

второе док-во

$$\operatorname{rk}_{\text{г}} A = \operatorname{rk}_{\text{г}} A' = \operatorname{rk}_{\text{в}} A' = \operatorname{rk}_{\text{в}} A.$$

Их поочерёдно обеспечивают три приведённых леммы. □

Критерий совместности

Теперь можно заново посмотреть на вопрос совместности системы линейных уравнений.

Теорема (критерий совместности)**.** *Система линейных уравнений совместна* $\iff$ *равны ранги её основной и расширенной матриц.*

Доказательство. Совместность линейной системы $AX = B$ эквивалентна тому, что B есть линейная комбинация столбцов матрицы A с какими-то неизвестными и разыскиваемыми коэффициентами x_i:

$$x_1 A^{(1)} + \ldots + x_n A^{(n)} = B.$$

По первой лемме о линейной (не)зависимости, это равносильно совпадению линейных оболочек $\langle\{\text{столбцы } A\}\rangle$ и $\langle\{\text{столбцы } A\} \cup \{B\}\rangle$, а тогда и ранги их равны. Обратно, если ранги равны, то совпадают и сами оболочки, так как вторая включает первую. □

Однородные системы

Однородной называют систему линейных уравнений, у которой столбец свободных членов нулевой: $AX = \mathbf{0}$. Если известно некоторое множество решений $\{X_1, \ldots, X_k\}$ однородной системы, то всякая линейная комбинация их также будет решением. Вообще, множество всех решений есть линейное пространство; оно называется **пространством решений** этой однородной системы.

Пример. Возьмём систему и приведём её к ступенчатой матрице:

$$\begin{cases} -x_1 + x_3 - 3x_4 + 2x_5 = 0, \\ 2x_1 - x_3 + 5x_4 - x_5 = 0, \\ x_1 - 2x_3 + 4x_4 - 5x_5 = 0; \end{cases} \qquad \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & -1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Все решения этой системы можно выписать, беря x_2, x_4 и x_5 в качестве параметров. Базис находим, поочерёдно подставляя единицу в каждый параметр и одновременно зануляя остальные:

$$X_1 = [0, 1, 0, 0, 0]^\top, \quad X_2 = [-2, 0, 1, 1, 0]^\top, \quad X_3 = [-1, 0, -3, 0, 1]^\top$$

составляют базис пространства решений, также называемый **фундаментальной системой решений**. Пространство решений задаётся в виде

$$\{\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 \mid \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}\};$$

эта запись даёт все решения исходной линейной системы и потому называется **общим решением** системы. Различных его записей такого вида бесконечно много, как и выборов фундаментальной системы решений.

Неоднородные системы

Множество $\mathcal{M}$ всех решений произвольной неоднородной линейной системы $AX = B$ не есть линейное пространство. Однако это подмножество пространства столбцов связано с пространством $\mathcal{L}$ решений **сопутствующей** однородной системы $AX = \mathbf{0}$. В самом деле,

$$AX = B \ \wedge \ AY = \mathbf{0} \quad \Longrightarrow \quad A(X + Y) = B;$$
$$AX = B \ \wedge \ AX' = B \ \Longrightarrow \ A(X - X') = \mathbf{0}.$$

Поэтому можно написать $\mathcal{M} = X + \mathcal{L}$, понимая под этим

$$\mathcal{M} = \{X + Y \mid Y \in \mathcal{L}\}.$$

Определение. **Линейным (под)многообразием** в линейном пространстве называется его подмножество вида $\mathcal{M} = \mathbf{v} + \mathcal{L}$ для каких-то фиксированных подпространства $\mathcal{L}$ и вектора $\mathbf{v}$.

Пример. Линейные подпространства и многообразия в обычном $\mathbb{R}^3$:

$\dim \mathcal{L}$	$\mathcal{L}$	$\mathcal{M}$
0	Нулевое подпространство $\{\mathbf{0}\}$	Точка
1	Прямая, содержащая $\mathbf{0}$	Прямая
2	Плоскость, содержащая $\mathbf{0}$	Плоскость
3	Само пространство $\mathbb{R}^3$	

Пример. Возьмём систему и приведём её к ступенчатой матрице:

$$\begin{cases} -x_1 + x_3 - 3x_4 + 2x_5 = 1, \\ 2x_1 - x_3 + 5x_4 - x_5 = 3, \\ x_1 - 2x_3 + 4x_4 - 5x_5 = -6; \end{cases} \qquad \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 2 & 1 & 4 \\ 0 & 0 & 1 & -1 & 3 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Многообразие её решений можно задать в виде

$$\{X_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 \mid \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}\},$$

где X_0 есть некоторое **частное** решение, скажем, $X_0 = [4, 0, 5, 0, 0]^\top$, а $\{X_1, X_2, X_3\}$ — фундаментальное решение сопутствующей однородной системы, рассмотренной выше. Частное решение проще всего выписать, положив все параметры равными нулю.

Альтернатива Фредгольма

Лемма. *Для всякой матрицы A верно равенство* $\operatorname{rk} A = \operatorname{rk} A^\top$.

Доказательство. Упражнение. □

Теорема. *Для каждой матрицы A верно только одно из следующих утверждений:*

(1) *либо система $AX = B$ совместна при любой правой части,*
(2) *либо система $A^\top Y = 0$ имеет ненулевое решение.*

Доказательство. Во второй системе столько же неизвестных, сколько уравнений в первой. Если ранг A равен этому числу, то первое утверждение истинно, а второе ложно. Если ранг меньше, то наоборот. □

4.4. Действия с подпространствами

Пересечение и сумма двух подпространств

При работе с множествами постоянно возникают пересечения и объединения подмножеств. Пытаясь наивно перенести эти операции на линейные подпространства, мы обнаруживаем, что

☺ пересечение $\mathcal{L}_1 \cap \mathcal{L}_2$ подпространств есть подпространство;
☹ объединение $\mathcal{L}_1 \cup \mathcal{L}_2$ подпространств не есть подпространство: достаточно посмотреть на две прямые на плоскости.

Однако представим пересечение и объединение подмножеств в терминах, перенос которых на подпространства оказывается дословным. Это *максимальность* пересечения и *минимальность* объединения:

$$\mathcal{S} \subseteq \mathcal{X}_1 \;\wedge\; \mathcal{S} \subseteq \mathcal{X}_2 \implies \mathcal{S} \subseteq \mathcal{X}_1 \cap \mathcal{X}_2,$$
$$\mathcal{S} \supseteq \mathcal{X}_1 \;\wedge\; \mathcal{S} \supseteq \mathcal{X}_2 \implies \mathcal{S} \supseteq \mathcal{X}_1 \cup \mathcal{X}_2.$$

Наименьшим линейным пространством, включающим два заданных подпространства, является линейная оболочка их объединения. Для него введём новый термин и новое обозначение:

$$\mathcal{L}_1 + \mathcal{L}_2 = \langle \mathcal{L}_1 \cup \mathcal{L}_2 \rangle$$

Определение. Линейную оболочку объединения двух подпространств $\mathcal{L}_1$ и $\mathcal{L}_2$ называют их **суммой** и обозначают через $\mathcal{L}_1 + \mathcal{L}_2$.

Формула размерностей Грассмана

Пример. Суммой двух непараллельных прямых является плоскость.

Пример. Суммой непараллельных прямой и плоскости является всё пространство $\mathbb{R}^3$.

Пример. Суммой двух непараллельных плоскостей в $\mathbb{R}^3$ является всё пространство $\mathbb{R}^3$.

Выпишем в этих трёх примерах размерности суммы, пересечения, а также исходных подпространств. Расположив их соответственно — например, $3 + 1 = 2 + 2$ в третьем случае,— увидим иллюстрацию следующей общей формулы.

Теорема (Grassmann, 1844)**.** *Формула размерностей*

$$\dim(\mathcal{L}_1 + \mathcal{L}_2) + \dim(\mathcal{L}_1 \cap \mathcal{L}_2) = \dim(\mathcal{L}_1) + \dim(\mathcal{L}_2)$$

верна для всякой пары конечномерных подпространств $\mathcal{L}_1$ и $\mathcal{L}_2$ любого линейного пространства.

Формула размерностей Грассмана аналогична формуле для числа элементов в объединении и пересечении конечных множеств:

$$^\#(\mathcal{X}_1 \cup \mathcal{X}_2) + {}^\#(\mathcal{X}_1 \cap \mathcal{X}_2) = {}^\#(\mathcal{X}_1) + {}^\#(\mathcal{X}_2).$$

Для её доказательства оформим подготовительные рассуждения в виде двух лемм, которые окажутся полезны и для следующей теоремы.

Доказательство

В первых двух примерах выше $\mathcal{L}_1 \cap \mathcal{L}_2 = \{\mathbf{0}\}$, а в третьем это не так: плоскости пересекаются вдоль прямой. Рассмотрим подробнее это условие на пересечение.

Лемма. *Для всякой пары подпространств $\mathcal{L}_1$ и $\mathcal{L}_2$ равносильны следующие условия:*

(1) $\mathcal{L}_1 \cap \mathcal{L}_2 = \{\mathbf{0}\}$;
(2) *нельзя представить* $\mathbf{0}$ *в виде суммы ненулевых* $\mathbf{v}_i \in \mathcal{L}_i$.

Расшифруем первое условие:

$$\mathbf{v}_1 = \mathbf{v}_2 \in \mathcal{L}_1 \cap \mathcal{L}_2 \quad \Longrightarrow \quad \mathbf{v}_1 = \mathbf{v}_2 = \mathbf{0}.$$

Зашифруем второе условие: для всех пар векторов $\mathbf{v}_i \in \mathcal{L}_i$ верен вывод

$$\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0} \quad \Longrightarrow \quad \mathbf{v}_1 = \mathbf{0} \ \wedge \ \mathbf{v}_2 = \mathbf{0}.$$

Можно поэтому сказать, что запись нулевого вектора в виде суммы $\mathbf{v}_1 + \mathbf{v}_2$ при условии $\mathbf{v}_k \in \mathcal{L}_k$ однозначна.

Доказательство. Рассуждение сводится к простому преобразованию:

$$\mathbf{a} = \mathbf{b} \Leftrightarrow \mathbf{a} - \mathbf{b} = \mathbf{0}.$$

(1 ⇒ 2) Если $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}$, то $\mathbf{v}_1 = -\mathbf{v}_2$, и (1) даёт $\mathbf{v}_1 = -\mathbf{v}_2 = \mathbf{0}$.
(2 ⇒ 1) Если $\mathbf{v}_1 = \mathbf{v}_2$, то $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$, и (2) даёт $\mathbf{v}_1 = -\mathbf{v}_2 = \mathbf{0}$. □

Лемма. *Выпишем два условия на пару множеств $\mathcal{X}$ и $\mathcal{Y}$ векторов одного линейного пространства:*

(A) *объединение $\mathcal{X} \cup \mathcal{Y}$ линейно независимо;*
(B) *пересечение линейных оболочек $\langle \mathcal{X} \rangle$ и $\langle \mathcal{Y} \rangle$ нулевое.*

Тогда:

(1) *вывод* (A ⇒ B) *верен для любых $\mathcal{X}$ и $\mathcal{Y}$;*
(2) *вывод* (B ⇒ A) *верен для любых линейно независимых $\mathcal{X}$ и $\mathcal{Y}$.*

Доказательство. Заменим условие (B) на равносильное ему второе условие предыдущей леммы, применённой к линейным оболочкам $\langle \mathcal{X} \rangle$ и $\langle \mathcal{Y} \rangle$. Затем, беря в $\langle \mathcal{X} \cup \mathcal{Y} \rangle = \langle \mathcal{X} \rangle + \langle \mathcal{Y} \rangle$ любой вектор

$$\mathbf{x} + \mathbf{y} = \Big(\sum \alpha_i X_i\Big) + \Big(\sum \beta_j Y_j\Big),$$

получим (A ⇒ B) непосредственно из определения:

$$(\mathbf{x} + \mathbf{y} = \mathbf{0}) \implies (\text{все } \alpha_i = 0 \text{ и все } \beta_j = 0) \implies (\mathbf{x} = \mathbf{0} \ \wedge \ \mathbf{y} = \mathbf{0}),$$

где первый переход обеспечен условием (A). Для (B ⇒ A) аналогично:

$$(\mathbf{x} + \mathbf{y} = \mathbf{0}) \implies (\mathbf{x} = \mathbf{0} \ \wedge \ \mathbf{y} = \mathbf{0}) \implies (\text{все } \alpha_i = 0 \text{ и все } \beta_j = 0).$$

Первый переход обеспечен условием (B) и предыдущей леммой, а вот далее нужна линейная независимость $\mathcal{X}$ и $\mathcal{Y}$ по отдельности. □

Доказательство теоремы. В пересечении $\mathcal{L}_1 \cap \mathcal{L}_2$ выберем базис $\mathcal{B}_0$. По лемме о дополнении до базиса, найдутся такие линейно независимые множества $\mathcal{B}_1$ и $\mathcal{B}_2$, что $\mathcal{B}_0 \cup \mathcal{B}_1$ и $\mathcal{B}_0 \cup \mathcal{B}_2$ есть базисы $\mathcal{L}_1$ и $\mathcal{L}_2$ соответственно:

$$\overbrace{\mathcal{B}_1 \cup \mathcal{B}_0}^{\mathcal{L}_1} \cup \mathcal{B}_2 = \mathcal{B}_1 \cup \underbrace{\mathcal{B}_0 \cup \mathcal{B}_2}_{\mathcal{L}_2}.$$

При этом $\langle \mathcal{B}_0 \rangle \cap \langle \mathcal{B}_1 \rangle = \{\mathbf{0}\}$ и $\langle \mathcal{B}_0 \rangle \cap \langle \mathcal{B}_2 \rangle = \{\mathbf{0}\}$ по выводу (1) последней леммы. Кроме того, $\langle \mathcal{B}_2 \rangle = \mathcal{L}_2 \cap \langle \mathcal{B}_2 \rangle$. Теперь вычисляем:

$$\langle \mathcal{B}_0 \cup \mathcal{B}_1 \rangle \cap \langle \mathcal{B}_2 \rangle = \mathcal{L}_1 \cap \mathcal{L}_2 \cap \langle \mathcal{B}_2 \rangle = \langle \mathcal{B}_0 \rangle \cap \langle \mathcal{B}_2 \rangle = \{\mathbf{0}\}.$$

По выводу (2) последней леммы $\mathcal{B} = \mathcal{B}_0 \cup \mathcal{B}_1 \cup \mathcal{B}_2$ линейно независимо. Поскольку $\langle \mathcal{B} \rangle = \mathcal{L}_1 + \mathcal{L}_2$, формула размерностей Грассмана проверяется простым подсчётом количества элементов базисов $\mathcal{B}_i$ и $\mathcal{B}$. $\square$

Сумма нескольких подпространств

Определение. Линейную оболочку объединения набора подпространств $\mathcal{L}_1, \dots, \mathcal{L}_k$ называют их **суммой** и обозначают через $\mathcal{L}_1 + \ldots + \mathcal{L}_k$:

$$\mathcal{L}_1 + \ldots + \mathcal{L}_k = \langle \mathcal{L}_1 \cup \ldots \cup \mathcal{L}_k \rangle = \{\mathbf{v}_1 + \ldots + \mathbf{v}_k \mid \mathbf{v}_i \in \mathcal{L}_i\}.$$

Как и для случая двух слагаемых, сумма является наименьшим пространством, включающим все $\mathcal{L}_i$.

Упражнение. *Начав с двух подпространств $\mathcal{L}_i$ и применяя операции пересечения и суммы многократно, нельзя получить новых подпространств, помимо $\mathcal{L}_1 \cap \mathcal{L}_2$ и $\mathcal{L}_1 + \mathcal{L}_2$.*

Начав с трёх подпространств $\mathcal{L}_i$ и применяя операции пересечения и суммы, сколько новых подпространств можно получить?

А с четырёх?

Прямые суммы

Изучать суммы многих подпространств, не накладывая условий на их пересечения, непросто из-за возникающей комбинаторики. Перейдём теперь к важнейшему случаю, обобщающему условия из первой леммы. Условий при этом дадим больше и сформулируем их иначе.

Теорема. *Следующие свойства суммы $\mathcal{L} = \mathcal{L}_1 + \ldots + \mathcal{L}_k$ конечного числа подпространств равносильны:*

(1) *запись каждого вектора $\mathbf{v} \in \mathcal{L}$ в виде суммы $\mathbf{v} = \mathbf{v}_1 + \ldots + \mathbf{v}_k$, где $\mathbf{v}_i \in \mathcal{L}_i$, однозначна;*

(2) *такая запись нулевого вектора однозначна, то есть*

$$\mathbf{v}_1 + \ldots + \mathbf{v}_k = \mathbf{0} \implies \mathbf{v}_1 = \mathbf{0} \wedge \ \ldots \wedge \ \mathbf{v}_k = \mathbf{0};$$

(3) *пересечение каждого $\mathcal{L}_i$ с суммой $\mathcal{R}_i$ всех остальных нулевое;*

(4) *объединение произвольных базисов всех $\mathcal{L}_i$ является базисом $\mathcal{L}$.*

Определение. Сумму $\mathcal{L} = \mathcal{L}_1 + \ldots + \mathcal{L}_k$ подпространств с такими свойствами называют **прямой суммой** и пишут $\mathcal{L} = \mathcal{L}_1 \oplus \ldots \oplus \mathcal{L}_k$.

Выразим второе свойство в теореме подобно лемме: нельзя представить $\mathbf{0}$ в виде суммы слагаемых $\mathbf{v}_i \in \mathcal{L}_i$, не все из которых нулевые. Не напоминает ли это одну из формулировок понятия линейной независимости? Действительно, если все исходные подпространства одномерны, так что $\mathcal{L}_i = \langle \mathbf{v}_i \rangle$, то все условия в теореме превращаются в обычную линейную независимость. Значит, прямоту суммы можно понимать как обобщение линейной независимости на уровень произвольных подпространств.

Доказательство. ($2 \Rightarrow 1$) Перегруппировка

$$\mathbf{v}_1 + \ldots + \mathbf{v}_k = \mathbf{v}'_1 + \ldots + \mathbf{v}'_k \iff (\mathbf{v}_1 - \mathbf{v}'_1) + \ldots + (\mathbf{v}_k - \mathbf{v}'_k) = \mathbf{0}$$

сводит общий случай (1) к его частному случаю (2), поскольку набор из k равенств $\mathbf{v}_i - \mathbf{v}'_i = \mathbf{0}$ как раз выражает искомую однозначность.

($3 \Rightarrow 2$) Из записи $\mathbf{v}_1 + \ldots + \mathbf{v}_k = \mathbf{0}$ выводим $\mathbf{v}_1 = \mathbf{0}$, применяя первую лемму к $\mathcal{L}_1$ и $\mathcal{R}_1$. Аналогично для остальных.

($1 \Rightarrow 3$) В пересечении $\mathcal{L}_1 \cap \mathcal{R}_1$ любой вектор $\mathbf{v}$ имеет две записи:

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{0}_2 + \ldots + \mathbf{0}_k,$$
$$\mathbf{v} = \mathbf{0}_1 + \mathbf{v}_2 + \ldots + \mathbf{v}_k,$$

где $\mathbf{v}_i, \mathbf{0}_i \in \mathcal{L}_i$. Однако выполнение первого свойства требует $\mathbf{v}_i = \mathbf{0}_i$, так что ненулевых векторов в $\mathcal{L}_1 \cap \mathcal{R}_1$ нет. Аналогично для остальных.

($4 \Rightarrow 1$) Запись каждого $\mathbf{v} \in \mathcal{L}$ линейной комбинацией фиксированных базисных векторов однозначна. Векторы $\mathbf{v}_i$ получаются собиранием слагаемых, лежащих в $\mathcal{L}_i$, и тоже находятся однозначно.

($2 \Rightarrow 4$) В каждом слагаемом $\mathcal{L}_i$ возьмём какой-нибудь базис $\mathcal{X}_i$ и рассмотрим $\mathcal{X} = \mathcal{X}_1 \cup \ldots \cup \mathcal{X}_k$. Проверим для $\mathcal{X}$ два определяющих свойства базиса: $\langle \mathcal{X} \rangle = \mathcal{L}_1 + \ldots + \mathcal{L}_k$ по определению суммы, а линейная независимость $\mathcal{X}$ следует в два шага. На первом шаге берём равную нулю линейную комбинацию векторов из $\mathcal{X}$. Группируя слагаемые из каждого $\mathcal{X}_i$ в вектор из $\mathcal{L}_i$ и называя его $\mathbf{v}_i$, получим запись нуля из второго свойства, поэтому все $\mathbf{v}_i = \mathbf{0}$. На втором шаге смотрим по отдельности на эти нулевые линейные комбинации базисных векторов из каждого $\mathcal{X}_i$ и заключаем, что все они тривиальны. □

Следствие. *Сумма всякого набора различных слагаемых $\mathcal{L}_i$ прямой суммы $\mathcal{L} = \mathcal{L}_1 \oplus \ldots \oplus \mathcal{L}_k$ также прямая.*

Доказательство. При ограничении любым набором слагаемых запись нулевого вектора в сумме по-прежнему однозначна. □

Размерность прямой суммы

pagebreak?

Теорема. *Следующие свойства суммы* $\mathcal{L} = \mathcal{L}_1 + \ldots + \mathcal{L}_k$ *конечного числа конечномерных подпространств эквивалентны:*

(1) $\mathcal{L} = \mathcal{L}_1 \oplus \ldots \oplus \mathcal{L}_k$;
(2) $\dim \mathcal{L} = \dim \mathcal{L}_1 + \ldots + \dim \mathcal{L}_k$.

Доказательство. ($1 \Rightarrow 2$) Индукция по числу слагаемых. База тривиальна: $k = 1$. При $k > 1$ сумма $\mathcal{R}_k = \mathcal{L}_1 \oplus \ldots \oplus \mathcal{L}_{k-1}$ является прямой по следствию, так что для неё утверждение (2) предполагаем выполненным. Сумма $\mathcal{L} = \mathcal{L}_k \oplus \mathcal{R}_k$ тоже прямая, ибо $\mathcal{L}_k \cap \mathcal{R}_k = \{\mathbf{0}\}$ по свойству (3) в предыдущей теореме. Шаг индукции обеспечивает формула размерностей Грассмана:

$$\begin{aligned}\dim \mathcal{L} &= \dim \mathcal{R}_k + \dim \mathcal{L}_k - \dim(\mathcal{L}_k \cap \mathcal{R}_k)\\ &= \big(\dim \mathcal{L}_1 + \ldots + \dim \mathcal{L}_{k-1}\big) + \dim \mathcal{L}_k.\end{aligned}$$

($2 \Rightarrow 1$) Объединение $\mathcal{X} = \mathcal{X}_1 \cup \ldots \cup \mathcal{X}_k$ базисов $\mathcal{X}_i$ слагаемых состоит из $\dim \mathcal{L}$ векторов и $\langle \mathcal{X} \rangle = \mathcal{L}$. Поэтому $\mathcal{X}$ обязано быть линейно независимым; это базис $\mathcal{L}$ и получено свойство (4) прямой суммы. □

Дополнение к подпространству

Следствие. *Для всякого подпространства* $\mathcal{U} \subseteq \mathcal{W}$ *найдётся такое подпространство* $\mathcal{V} \subseteq \mathcal{W}$*, что* $\mathcal{W} = \mathcal{U} \oplus \mathcal{V}$.

Доказательство. Базис $\mathcal{U}$ можно дополнить до базиса $\mathcal{W}$, при этом дополняющие векторы составят базис $\mathcal{V}$. □

Определение. Такое $\mathcal{V}$ называют **дополнением** к $\mathcal{U}$.

Если $\mathcal{U} \neq \mathcal{W}$, то дополнять базис $\mathcal{U}$ до базиса $\mathcal{W}$ можно различными способами. Поэтому, из $\mathcal{U} \oplus \mathcal{V}_1 = \mathcal{U} \oplus \mathcal{V}_2$ не следует, что $\mathcal{V}_1 = \mathcal{V}_2$.

Пример. Для подпространств

$$\begin{aligned}\mathcal{U} &= \{[x_1, \ldots, x_n] \mid x_1 = \ldots = x_n\},\\ \mathcal{V}_1 &= \{[x_1, \ldots, x_n] \mid x_1 + \ldots + x_n = 0\},\\ \mathcal{V}_2 &= \{[x_1, \ldots, x_n] \mid x_n = 0\}\end{aligned}$$

в $\mathbb{R}^n$ имеем $\mathcal{U} \oplus \mathcal{V}_1 = \mathcal{U} \oplus \mathcal{V}_2$ и $\mathcal{V}_1 \neq \mathcal{V}_2$.

Прямая сумма как конструкция

4.5. Примеры линейных пространств

Матрицы

Множество всех $m \times n$ матриц есть линейное пространство. Поэлементное сложение матриц одинаковых размеров уже встречалось; умножение матрицы на скаляр совершается также поэлементно:

$$A = [a_{ij}] \implies \lambda A = [\lambda a_{ji}].$$

Размерность равна mn, а базис можно составить из **матричных единиц** E_{ij}. У этих матриц лишь по одному ненулевому элементу: единица в позиции ij. Не путать матричные единицы и единичную матрицу!

Пространство $M_n(\mathbb{R})$ вещественных квадратных $n \times n$ матриц включает много интересных подпространств. Нам пригодятся приведённые в таблице самые важные из них, в которых матрицы имеют различные специальные виды. При том, можно заметить разложения

можно схемно и
выделяя клеточ

$$M_n(\mathbb{R}) = \mathcal{S} \oplus \mathcal{A} = \mathcal{N}_+ \oplus \mathcal{D} \oplus \mathcal{N}_- = \mathcal{K} \oplus \mathcal{O},$$

а также $\mathcal{U} = \mathcal{D} \oplus \mathcal{N}_+$ и $\mathcal{L} = \mathcal{D} \oplus \mathcal{N}_-$.

Обозначение и название		*Условие*	*Размерность*
$\mathcal{K}$	Скалярные	$A = \lambda E$	1
$\mathcal{D}$	Диагональные	$a_{ij} = 0$ для $i \neq j$	n
$\mathcal{U}$	Верхнетреугольные	$a_{ij} = 0$ для $i > j$	$n(n+1)/2$
$\mathcal{L}$	Нижнетреугольные	$a_{ij} = 0$ для $i < j$	$n(n+1)/2$
$\mathcal{N}_+$	Строго верхнетреугольные	$a_{ij} = 0$ для $i \geqslant j$	$n(n-1)/2$
$\mathcal{N}_-$	Строго нижнетреугольные	$a_{ij} = 0$ для $i \leqslant j$	$n(n-1)/2$
$\mathcal{S}$	Симметричные	$A^\top = A$	$n(n+1)/2$
$\mathcal{A}$	Кососимметричные	$A^\top = -A$	$n(n-1)/2$
$\mathcal{O}$	Бесследные	$a_{11} + \ldots + a_{nn} = 0$	$n^2 - 1$

Первые шесть указанных подпространств ещё и подкольца в $M_n(\mathbb{R})$. Остальные три тоже кольца, но с более хитрыми умножениями.

Дальнейшие примеры

Пример (*пространство решений линейного дифференциального уравнения*). Дифференциальное уравнение колебаний

$$\ddot{x}(t) + x(t) = 0$$

имеет два очевидных решения

$$x_1(t) = \cos t, \quad x_2(t) = \sin t.$$

Решением будет и любая линейная комбинация $\alpha_1 \cos t + \alpha_2 \sin t$, а других решений нет. Пространство решений двумерно; множество функций $\{\cos t, \sin t\}$ есть фундаментальная система решений.

Пример (*пространство многочленов*). На множестве многочленов $\mathbb{R}[x]$ есть операция умножения. Забыв о ней и сохранив, помимо сложения, лишь умножение многочленов на константы, мы можем трактовать $\mathbb{R}[x]$ как линейное пространство над $\mathbb{R}$. Оно не имеет конечного базиса, то есть оно бесконечномерно.

Если же степени многочленов ограничить фиксированным числом d, то получится пространство $\mathbb{R}[x]_{\leqslant d}$ размерности $d+1$. Самый простой базис его есть $\{1, x, x^2, \ldots, x^d\}$.

Пример (*пространство функций*). Возьмём отрезок $I \subset \mathbb{R}$ и обозначим через $\mathrm{Map}(I, \mathbb{R})$ множество всех функций на I с вещественными значениями. Функции можно складывать и умножать на число, можно строить их линейные комбинации. Значит, $\mathrm{Map}(I, \mathbb{R})$ есть линейное пространство. Оно бесконечномерно (если концы отрезка различны).

При рассмотрении «хороших» функций часто возникают значительно меньшие подпространства $\mathrm{Map}(I, \mathbb{R})$, хотя и бесконечномерные: все непрерывные функции, все непрерывно дифференцируемые функции, и много других, изучаемых в анализе.

Приведённых примеров должно быть достаточно, чтобы убедиться в исключительной употребимости и полезности понятия линейного пространства. Все определения и утверждения, данные выше для пространства $\mathbb{R}^n$, остаются применимы и верны для любого (конечномерного) линейного пространства.

Кроме того, можно заменить поле вещественных чисел любым полем $\mathbb{F}$ скаляров и изучать линейные пространства **над** $\mathbb{F}$. Важность случая комплексных линейных пространств ($\mathbb{F} = \mathbb{C}$) уже не должна вызывать удивления. При замене поля кольцом теория усложняется.

Глава 5. ОПРЕДЕЛИТЕЛИ

5.1. Комбинаторное строение определителей

Подходы к понятию определителя

Определители произвольного порядка формально вводят несколькими очень разными способами. Отметим основные:

(1) геометрически мотивированный;
(2) индуктивный;
(3) аксиоматический;
(4) комбинаторный.

При геометрическом подходе определителем $n \times n$ матрицы называют ориентированный объём параллелепипеда в $\mathbb{R}^n$, образованного её строками; мы уже видели это в размерностях 2 и 3. При индуктивном подходе определитель порядка n вводится через формулу раскрытия по строке или столбцу, аналогично данной в главе 1. При аксиоматическом подходе определитель возникает как единственная функция на наборе строк матрицы, имеющая три заданных свойства.

Мы выберем комбинаторный подход ради короткого и строгого определения, затем получим индуктивное свойство, кососимметричность и полилинейность, а также несколько простых свойств, являющихся их следствиями. Многомерная геометрия в этом курсе затронута позже, во втором семестре.

Комбинаторный подход

Комбинаторный подход основан на понятии перестановки.

Определение. Биективное отображение конечного множества Ω в себя называют **перестановкой** Ω. Обычно берут $\Omega = \{1, 2, \ldots, n\}$; множество его перестановок обозначают через $\mathbb{S}_n$ (или Σ_n).

Пример. Стандартный компактный способ записи перестановки:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 1 & 2 & 4 \end{pmatrix}, \quad \sigma \in \mathbb{S}_5,$$

указывает, что $\sigma(1) = 5$, $\sigma(2) = 3$, $\sigma(3) = 1$, $\sigma(4) = 2$ и $\sigma(5) = 4$. Будет полезно также изображать перестановку картинкой:

	1	2	3	4	5
1					x
2			x		
3	x				
4		x			
5				x	

Инверсии: $(1,2), (1,3), (1,4), (1,5),$
$(2,3), (2,4).$

Чётность: чётная, $\mathrm{inv}(\sigma) = 6$.

Определение. **Инверсией** перестановки $\sigma \in \mathbb{S}_n$ называют такую пару (i, j), что $i < j$ и $\sigma(i) > \sigma(j)$. На картинке это будет каждая пара крестиков, прямая через которую наклонена вправо.

Чётностью перестановки σ называют чётность количества $\mathrm{inv}(\sigma)$ всех её инверсий, а **знаком**, число $\mathrm{sgn}(\sigma) = (-1)^{\mathrm{inv}(\sigma)}$.

В доказательствах нам будет удобно говорить также о чётности картинки, имея в виду чётность изображаемой ею перестановки.

Перестановки дают возможность единообразно записать формулы полного раскрытия определителей второго и третьего порядков, указанные ещё в начале курса. Результат немедленно обобщается на произвольный порядок (Leibniz, 1683).

Определение. **Определителем** $n \times n$ матрицы $A = [a_{ij}]$ называют сумму

$$(\sharp) \qquad \det A \rightleftharpoons \sum_{\sigma \in \mathbb{S}_n} \mathrm{sgn}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \ldots a_{n,\sigma(n)}.$$

В правой части этой формулы стоит алгебраическая сумма $n!$ однотипных мономов. Например, нарисованная выше перестановка $\sigma \in \mathbb{S}_5$ вносит моном $a_{15}a_{23}a_{31}a_{42}a_{54}$, причём с плюсом, ибо она чётна:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 1 & 2 & 4 \end{pmatrix} \quad \rightsquigarrow\rightsquigarrow \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}.$$

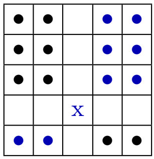
Вспомогательные свойства числа инверсий

Лемма. *Число инверсий картинки любой перестановки:*

(1) *неизменно при отражении относительно главной диагонали;*
(2) *меняет чётность при обмене двух строк или двух столбцов;*
(3) *содержащих крестик в строке i и столбце j, имеет такую же чётность как число $i + j$.*

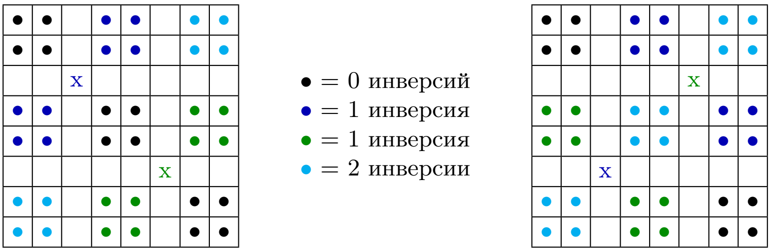
Доказательство. (1) Благодаря симметрии в определении инверсии.

(3) Встанем с компасом в строку i и столбец j. Севернее (выше) расположено $i-1$ крестиков, а западнее (левее) расположено $j-1$ крестиков. При том, если к северо-западу находится k крестиков, то к северо-востоку их $i-k-1$, а к юго-западу $j-k-1$. Значит, всего крестик в клетке (i, j) входит в $i+j-2(k+1)$ инверсионных пар.



(2) Прежде всего заметим, что если выбранная для обмена пара строк содержит инверсионную пару крестиков, то обмен делает её не инверсионной, и наоборот. Остаётся проверить, что чётность числа прочих инверсий не меняется.

Нужно рассмотреть все возможные случаи расположения одного неподвижного крестика относительно обменивающейся пары. Возьмём пару крестиков, не образующих инверсию, а в остальных клетках отметим разными кружками, сколько инверсий с обменивающейся парой даст крестик, помещённый туда.



Сравним с аналогичной картинкой после обмена строк. Всего видим девять случаев, но лишь в «центральной части» исследуемое число меняется, причём на 2. □

Схема трёх доказательств

Далее мы трижды проведём рассуждения по следующей схеме:

(א) Распишем определители в обеих частях доказываемого равенства по определению, как суммы мономов.

(ב) Убедимся, что списки мономов слева и справа одинаковы.

(ℶ) Сравним знаки каждого монома с двух сторон при помощи леммы о свойствах чётности перестановок.

Повторяющиеся при этом простые шаги упоминать явно не будем.

Лемма. *Для всякой квадратной матрицы,* $\det A = \det A^{\top}$.

Ввиду этой леммы все свойства определителей, формулируемые или доказываемые для строк, верны и для столбцов.

Доказательство. (ℷ) Транспонирование отражает картинку, сохраняя чётность. Поэтому мономы двух симметричных картинок, например,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \quad \text{и} \quad \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} & a_{51} \\ a_{12} & a_{22} & a_{32} & a_{42} & a_{52} \\ a_{13} & a_{23} & a_{33} & a_{43} & a_{53} \\ a_{14} & a_{24} & a_{34} & a_{44} & a_{54} \\ a_{15} & a_{25} & a_{35} & a_{45} & a_{55} \end{bmatrix},$$

входят с одинаковыми знаками в $\det A$ и $\det A^{\top}$. □

Миноры и раскрытие по строке либо столбцу

Определение. **Минором** порядка k произвольной (не обязательно квадратной) матрицы A называют определитель матрицы, составленной из элементов A, стоящих на пересечении каких-то выбранных в ней различных k строк и различных k столбцов:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}.$$

Примечательны миноры порядка $n-1$ квадратной матрицы размера $n \times n$. Обозначим через $M_{ij}(A)$ минор, получаемый вычёркиванием из A строки i и столбца j: Значение $(-1)^{i+j} M_{ij}(A)$ называют **алгебраическим дополнением** элемента a_{ij} матрицы A. Откуда такое название?

Лемма. *Зависимость определителя от клетки* (i, j) *имеет вид*

$$\det A = (-1)^{i+j} a_{ij} M_{ij}(A) + \sum (\textit{мономы без } a_{ij}).$$

Доказательство. (ℶ) Внимание сосредоточено на картинках перестановок $\sigma \in \mathbb{S}_n$ с крестиком в клетке (i, j). Удаление строки i и столбца j даёт картинки всех перестановок в $\mathbb{S}_{n-1}$, которые и нужны при раскрытии минора M_{ij}. Разберите случай $n = 3$ как упражнение.

($\gimel$) Третье утверждение в лемме об инверсиях и компенсирующий множитель $(-1)^{i+j}$ обеспечивают правильные знаки мономов. $\square$

Теорема (Leibniz, Laplace)**.** *Для каждой строки матрицы A определитель* $\det A$ *равен сумме произведений элементов в этой строке на их алгебраические дополнения; аналогично для каждого столбца:*

$$\det A = \sum_{1\leqslant j\leqslant n} (-1)^{i+j} a_{ij} M_{ij}(A) \quad \text{для всех } 1 \leqslant i \leqslant n;$$

$$\det A = \sum_{1\leqslant i\leqslant n} (-1)^{i+j} a_{ij} M_{ij}(A) \quad \text{для всех } 1 \leqslant j \leqslant n.$$

Доказательство. Достаточно применять лемму, двигаясь по строке i либо столбцу j. Каждый моном определителя попадёт ровно в одну группу слагаемых, выделяемых леммой для клеток, потому что в него входит ровно один элемент этой строки (либо столбца). $\square$

5.2. Свойства определителей

Треугольные матрицы

Определение. Квадратную матрицу $A = [a_{ij}]$, у которой $a_{ij} = 0$ при $i > j$, то есть все элементы ниже главной диагонали нулевые, называют **верхнетреугольной**. Аналогично определяют **нижнетреугольные** матрицы.

Следствие. *Определитель верхнетреугольной матрицы равен произведению элементов на её главной диагонали,* $\det A = a_{11}a_{22}\dots a_{nn}$; *также и для нижнетреугольной.*

Первое доказательство. Если в картинке нет крестиков ниже диагонали, то и выше диагонали их там тоже быть не может: нет места. Значит, единственный ненулевой моном суммы ($\sharp$) — диагональный. $\square$

Второе доказательство. По индукции, всё время раскрывая определитель по первому столбцу (для верхнетреугольной) или первой строке (для нижнетреугольной). $\square$

В частности, определитель единичной матрицы равен единице; назовём это свойством (D1).

Следствие. *Определители* **блочно-треугольных** *матриц*

$$\left[\begin{array}{c|c} A & B \\ \hline 0 & D \end{array}\right] \quad \text{и} \quad \left[\begin{array}{c|c} A & 0 \\ \hline C & D \end{array}\right],$$

где матрицы A и D квадратные, равны $\det A \cdot \det D$.

Доказательство. Скомбинировать идею первого доказательства предыдущего следствия и идею доказательства третьего утверждения в лемме об инверсиях. Ненулевые мономы большого определителя не затрагивают блоки B и C, а инверсий между блоками A и D нет, так что они раскрываются независимо и возникает произведение. □

Пример. Рассмотрим случай $n = 4$ с блоками 2×2. Здесь всего четыре потенциально ненулевых монома:

$$\begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ 0 & 0 & a_{33} & a_{43} \\ 0 & 0 & a_{34} & a_{44} \end{bmatrix}, \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ 0 & 0 & a_{33} & a_{43} \\ 0 & 0 & a_{34} & a_{44} \end{bmatrix},$$

$$\begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ 0 & 0 & a_{33} & a_{43} \\ 0 & 0 & a_{34} & a_{44} \end{bmatrix}, \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ 0 & 0 & a_{33} & a_{43} \\ 0 & 0 & a_{34} & a_{44} \end{bmatrix}.$$

Разность двух верхних мономов равна $(\det A)a_{33}a_{44}$, а разность двух нижних — $(\det A)a_{34}a_{43}$. Разность этих разностей равна $(\det A)(\det D)$.

Определитель как функция строк

Теорема. (D2) *Определитель матрицы меняет знак на противоположный при обмене местами любой пары её строк. Иначе говоря,* $\det$ *есть* **кососимметрическая** *функция строк матрицы.*

Доказательство. (ℶ) Картинки перестановок, дающие один и тот же моном до и после обмена строк, например,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \quad \text{и} \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix},$$

отличаются как во втором утверждении леммы об инверсиях. □

Теорема. (D3) *Определитель матрицы есть линейная функция элементов любой строки. Иначе говоря,* $\det$ *есть* **полилинейная** *функция строк матрицы.*

Доказательство. Утверждается, что если квадратные матрицы A, B и C отличаются лишь своими i-ми строками, равными соответственно $A_{(i)}$, $B_{(i)}$ и $C_{(i)} = \alpha A_{(i)} + \beta B_{(i)}$, то $\det C = \alpha \det A + \beta \det B$.

Для проверки этого заметим, что миноры $M_{ij}(A)$, $M_{ij}(B)$ и $M_{ij}(C)$ идентичны, подставим $c_{ij} = \alpha a_{ij} + \beta b_{ij}$ в формулу раскрытия $\det C$ по строке i и вынесем скаляры α и β из-под суммирования. $\square$

Теорема. *Если функция $\mathcal{D}\colon M_n(\mathbb{R}) \to \mathbb{R}$, как функция строк матрицы, кососимметрическая и полилинейная, то она отличается от определителя лишь постоянным множителем, равным своему значению на единичной матрице: $\mathcal{D}(A) = \mathcal{D}(E) \cdot \det A$ для всех A.*

Доказывать эту теорему мы не будем. Однако отсюда следует, что свойства (D1), (D2) и (D3) однозначно задают det как отображение $M_n(\mathbb{R}) \to \mathbb{R}$ и потому могут быть взяты в качестве абстрактного аксиоматического определения.

Несколько простых свойств

Полезны ещё несколько простых следствий основных свойств:

(D4) $\det(\lambda A) = \lambda^n \det A$ для всех $n \times n$ матриц A и скаляров λ;

(D5) определитель матрицы с нулевой строкой равен нулю;

(D6) определитель матрицы с двумя одинаковыми строками равен нулю;

(D7) определитель неизменен при элементарных преобразованиях строк типа (R2′).

Определитель произведения матриц

Теорема (Binet, Cauchy, 1812). *Для всех $n \times n$ матриц A и B,*

$$\det AB = \det A \cdot \det B.$$

Первое доказательство. Обозначим через $A_{(i)}$ строку i матрицы A и аналогично для других используемых матриц. Поскольку

$$A_{(i)} = a_{i1}E_{(1)} + \ldots + a_{in}E_{(n)} \quad \text{и} \quad (AB)_{(i)} = a_{i1}B_{(1)} + \ldots + a_{in}B_{(n)},$$

элементарными преобразованиями строк типа (R2′) левая матрица приводится к правой:

$$\left[\begin{array}{r|c} E & B \\ \hline -A & 0 \end{array}\right] \rightsquigarrow\rightsquigarrow\rightsquigarrow \left[\begin{array}{c|c} E & B \\ \hline 0 & AB \end{array}\right].$$

По свойству (D7) определители этих матриц равны; остаётся правильно их посчитать, используя свойства (D2) и (D4) и следствие про блочно-треугольные матрицы. $\square$

Второе доказательство. Зафиксируем B и определим отображение

$$\mathcal{D}_B \colon M_n(\mathbb{R}) \to \mathbb{R}, \quad A \mapsto \det AB.$$

Далее проверим, что это кососимметрическая и полилинейная функция строк матрицы A. Тогда по предыдущей теореме

$$\det AB = \mathcal{D}_B(A) = \mathcal{D}_B(E) \cdot \det A = \det B \cdot \det A. \qquad \square$$

Идея третьего доказательства. Возьмём в $\mathbb{R}^n$ линейно независимые векторы $X_1, \ldots, X_n$ и обозначим через $\operatorname{Vol}\{X_1, \ldots, X_n\}$ ориентированный объём n-мерного параллелепипеда на них. Оказывается, что

$$\frac{\operatorname{Vol}\{AX_1, \ldots, AX_n\}}{\operatorname{Vol}\{X_1, \ldots, X_n\}} = \det A.$$

Неосознанно, мы уже встречали это равенство при $n = 3$, вычисляя смешанное произведение в координатах. Определитель осмысливается как коэффициент искажения объёма, а при композиции преобразований такие коэффициенты умножаются. $\square$

5.3. Невырожденные матрицы

Критерий невырожденности

Теорема. *Равносильны следующие условия на $n \times n$ матрицу A:*

(1) $\det A \neq 0$;
(2) $\operatorname{rk} A = n$;
(3) *существует такая матрица A^{-1}, что $AA^{-1} = A^{-1}A = E$.*

Определение. Если эти условия выполнены, то матрицу A называют **невырожденной**, а матрицу A^{-1} — **обратной** к A.

Доказательство. ($1 \Rightarrow 2$) Если $\operatorname{rk} A < n$, то хотя бы одна из строк A есть линейная комбинация других. Тогда элементарными преобразованиями типа (R2′) можно получить нулевую строку, то есть $\det A = 0$ по свойствам (D5) и (D7).

($3 \Rightarrow 1$) Следует из теоремы об определителе произведения матриц.

($2 \Leftrightarrow 3$) Мы уже рассмотрели (на семинарах) способ нахождения A^{-1} как решения матричного уравнения $AX = E$; он работает тогда и только тогда, когда $\operatorname{rk} A$ максимально возможный. $\square$

Формула для обратной матрицы

Имеется явная формула для обратной матрицы. Обозначим через $A^\vee$ транспонированную матрицу алгебраических дополнений, то есть матрицу с элементами $a^\vee_{ij} = (-1)^{i+j} M_{ji}(A)$. Тогда

$$A^{-1} = (\det A)^{-1} A^\vee.$$

В самом деле, чтобы проверить, что $AA^\vee = E \det A$, вычислим произведение $AA^\vee$. Его элемент в строке i и столбце k равен

$$\sum_{1 \leqslant j \leqslant n} a_{ij} a^\vee_{jk} = \sum_{1 \leqslant j \leqslant n} (-1)^{k+j} a_{ij} M_{kj}(A).$$

Тут мы видим в точности правую часть формулы раскрытия определителя по k-й строке, примененную к матрице, полученной из A заменой строки k на копию строки i. При $i = k$ это будет сама матрица A и выражение сворачивается в $\det A$. При $i \neq k$ выражение сворачивается в определитель матрицы с двумя одинаковыми строками i и k; он равен нулю по свойству (D6).

Правило Крамера

Теперь мы можем получить известное ещё Лейбницу, хотя редко эффективное на практике уже при $n > 3$, «правило Крамера» для решения систем линейных уравнений с квадратной и невырожденной основной матрицей.

Следствие. *Если* $\det A \neq 0$, *то система* $AX = B$ *линейных уравнений в матричной форме записи, она же в форме*

$$A^{(1)} x_1 + \ldots + A^{(n)} x_n = B$$

линейной комбинации столбцов, имеет единственное решение

$$X = [\hat{x}_1, \ldots, \hat{x}_n]^\top, \qquad \hat{x}_j = D_j / \det A,$$

а D_j *есть определитель матрицы, полученной из* A *заменой столбца* $A^{(j)}$ *на столбец* B.

Доказательство. Поскольку основная матрица системы невырождена, единственность решения доказана ранее и его можно найти, вычислив $X = A^{-1}B$. Используя матрицу $A^\vee$, получаем

$$\hat{x}_j = (\det A)^{-1} \sum_{1 \leqslant i \leqslant n} a^\vee_{ji} b_i = (\det A)^{-1} \sum_{1 \leqslant i \leqslant n} (-1)^{i+j} b_i M_{ij}(A).$$

Раскрыв определитель D_j по столбцу j, лицезреем эту же сумму. $\square$

Ранг матрицы по минорам

Определение. **Рангом по минорам** произвольной (не обязательно квадратной) матрицы A называют наибольший порядок $\mathrm{rk}_{\text{м}} A$ её минора с отличным от нуля значением.

Теорема. *Ранг всякой матрицы A по минорам совпадает с её рангом по строкам/столбцам.*

Доказательство. Из матрицы A выберем $\mathrm{rk}\, A$ линейно независимых строк и составим из них матрицу A'. Из неё выберем $\mathrm{rk}\, A$ линейно независимых столбцов и составим из них матрицу A''. Она невырождена и является минором исходной матрицы. Поэтому $\mathrm{rk}\, A \leqslant \mathrm{rk}_{\text{м}} A$.

Найдём в A минор наибольшего порядка с отличным от нуля значением. Его строки линейно независимы, а тогда строки самой матрицы, его содержащие, тоже независимы. Поэтому $\mathrm{rk}\, A \geqslant \mathrm{rk}_{\text{м}} A$. □

Метод окаймляющих миноров

Определение. **Окаймляющим минором** данного минора M матрицы A называют каждый минор $\widetilde{M}$, вычёркивание из которого одной крайней строки и одного крайнего столбца даёт M.

Теорема. *Ранг A по минорам равен такому числу r, что у A имеется минор M порядка r с отличным от нуля значением, а значения всех миноров A, окаймляющих M, нулевые.*

Доказательство. Удалено за ненадобностью. □

Метод вычисления ранга матрицы пошаговым переходом от уже найденного ненулевого минора к окаймляющему называют **методом окаймляющих миноров**. Он особенно удобен в тех случаях, когда помимо ранга нужно узнать, какие именно строки и/или столбцы линейно независимы.

5.4. Полиномиальная интерполяция

Уравнение прямой

Чтобы задать тон всему разделу, начнём с простейшей задачи: на плоскости даны две различные точки с координатами (x_0, y_0) и (x_1, y_1), нужно найти уравнение проходящей через них прямой. Школьными методами можно выразить ответ в виде

$$y = y_0 + k \cdot (x - x_0), \text{ где } k = \frac{y_1 - y_0}{x_1 - x_0}. \tag{1}$$

Это, увы, некрасивый ответ. Тот факт, что при $x_0 = x_1$ он непригоден, нисколько не важен для дальнейшего. Но почему координаты заданных точек входят несимметрично? Преобразуем формулу к симметричному виду

$$y = y_0 \cdot \frac{x - x_0}{x_1 - x_0} + y_1 \cdot \frac{x - x_1}{x_0 - x_1}.$$

Лучше ли это? Не намного. Искомый красивый ответ выражается определителем:

$$\begin{vmatrix} 1 & x & y \\ 1 & x_0 & y_0 \\ 1 & x_1 & y_1 \end{vmatrix} = 0. \tag{2}$$

Раскрывая его, получим предыдущее равенство.

Мы проделывали этот эстетический фокус ещё в первой главе (там чуть отличались индексы и порядок столбцов). Далее мы разберём его обобщение. Основная идея раздела в том, что иногда не следует раскрывать подобные определители без необходимости и выдавать результаты раскрытия за готовые формулы. Гораздо лучше преподносить результаты в виде уравнений *красивый определитель равен нулю*.

Интерполяционный полином в форме Лагранжа

На плоскости возьмём три точки (x_k, y_k) с различными абсциссами и найдём полином наименьшей степени, график которого проходит через них. Его и называют интерполяционным полиномом. Поскольку условий три, то коэффициентов у полинома также должно быть три. Минимальная степень такого полинома — вторая, и среди квадратичных действительно удаётся найти решение.

Чтобы получить его, начнём с частного случая, когда одно из значений y_i равно 1, а остальные равны 0; например, $y_0 = 1$ для определённости. По теореме Безу, решение $P_2(x)$ делится нацело на $x - x_1$ и на $x - x_2$. Поэтому $P_2(x) = a(x - x_1)(x - x_2)$, где коэффициент a находится из условия $P_2(x_0) = 1$. Значит,

$$P_2(x) = \frac{x - x_1}{x_0 - x_1} \cdot \frac{x - x_2}{x_0 - x_2}.$$

Для произвольных y_0, y_1, y_2 отсюда получим решение

$$P_2(x) = y_0 \cdot \frac{x - x_1}{x_0 - x_1} \cdot \frac{x - x_2}{x_0 - x_2} + y_1 \cdot \frac{x - x_0}{x_1 - x_0} \cdot \frac{x - x_2}{x_1 - x_2} + y_2 \cdot \frac{x - x_0}{x_2 - x_0} \cdot \frac{x - x_1}{x_0 - x_1}.$$

Обобщение этой формулы на произвольное количество контрольных точек называют интерполяционным полиномом в форме Лагранжа. Однако, это некрасивый ответ!

По аналогии с (2), можно преобразовать равенство к виду

$$\begin{vmatrix} 1 & x & x^2 & y \\ 1 & x_0 & x_0^2 & y_0 \\ 1 & x_1 & x_1^2 & y_1 \\ 1 & x_2 & x_2^2 & y_2 \end{vmatrix} = 0. \tag{3}$$

Ответ сразу в таком виде можно получить иначе, в стиле аналитической геометрии: искомая парабола задаётся уравнением вида

$$A + Bx + Cx^2 + Dy = 0,$$

а такие задачи быстро решаются выписыванием определителя: мономы указывают на вид столбцов, а строки, кроме одной, соответствуют подстановкам контрольных точек. Это работы Крамера, откуда и пошло правило Крамера для решения линейных систем.

Конечно, при большем числе n контрольных точек стандартный вид

$$P_n(x) = \sum_{0\leqslant i\leqslant n} \left(y_i \prod_{j\neq i} \frac{x - x_j}{x_i - x_j} \right) \tag{4}$$

интерполяционного полинома в форме Лагранжа тоже является раскрытием определителя очень простого устройства, который читатель теперь легко напишет. При этом алгебраическое дополнение элемента y есть знаменитый определитель Вандермонда, отличный от нуля в точности тогда, когда все x_k различны.

Следующая задача: случай совпадающих абсцисс.

Полином Тэйлора

Вернёмся сперва к первой задаче с прямой, но изменим условие: теперь дана одна точка (x_0, y_0) и уклон k искомой прямой. Школьное уравнение такое же, а заданный коэффициент k можно интерпретировать как значение производной

$$y_0' = \frac{dy}{dx}(x_0),$$

результат предельного перехода в (1) при $(x_1, y_1) \to (x_0, y_0)$. Если же перейти к пределу в определителе, то мы увидим две одинаковые строки и бесполезное уравнение $0 = 0$. Чтобы избежать его, предварительно вычтем вторую строку из третьей, затем вынесем из разности множитель $x_1 - x_0$. После этого предельный переход даёт уравнение

$$\begin{vmatrix} 1 & x & y \\ 1 & x_0 & y_0 \\ 0 & 1 & y_0' \end{vmatrix} = 0 \quad \Leftrightarrow \quad y = y_0 + y_0' \cdot (x - x_0).$$

Назовём такие преобразования **конфлюэнтными пределами**, ибо в получаемых здесь и далее матрицах миноры, дополнительные к элементу y, называют конфлюэнтными матрицами Вандермонда. (Переводы «вырожденный», «совпадающий» несут нежелательные оттенки, поэтому оставим латынь.)

У нас получилось уравнение касательной к графику функции $y = y(x)$, правда, привычное лишь в правой форме. Касательная является линейным приближением; следующий шаг — квадратичное приближение. Контрольные данные тут состоят из точки (x_0, y_0), задающей значение $y_0 = y(x_0)$ функции, вместе со значениями y_0' и y_0'' её первой и второй производных в той же точке. Известный из анализа ответ — полином Тэйлора второй степени:

$$y = y_0 + y_0' \cdot (x - x_0) + y_0'' \cdot \tfrac{1}{2}(x - x_0)^2.$$

Можно проверить, что это выражение получается раскрытием определителя из уравнения

$$\begin{vmatrix} 1 & x & \frac{1}{2}x^2 & y \\ 1 & x_0 & \frac{1}{2}x_0^2 & y_0 \\ 0 & 1 & x_0 & y_0' \\ 0 & 0 & 1 & y_0'' \end{vmatrix} = 0. \tag{5}$$

Строки здесь являются значениями последовательных производных верхней строки в контрольной точке. Определитель можно найти из (3) предельными переходами с такими же трюками, как в предыдущем случае слипания двух контрольных точек, то есть конфлюэнтными пределами. Опустим сейчас эти детали, потому что мы ещё не рассмотрели промежуточную стадию тройного слипания, когда совпали только две точки.

В том же духе, полином Тэйлора

$$y = \sum_{0 \leqslant k \leqslant n} \left(y_0^{(k)} \cdot \frac{1}{k!}(x - x_0)^k \right)$$

более высокой степени n получается раскрытием аналогичного определителя порядка $n+2$. Если индексацию его столбцов начать с $j = 0$, а строк с $i = -1$, то элементы вне верхней строки и последнего столбца описываются общей формулой

$$a_{ij} = \left(\tfrac{d}{dx}\right)^i (x^j)\big|_{x=x_0},$$

но проще запомнить, что строки есть последовательные производные верхней строки.

Интерполяция Эрмита

Когда контрольных точек несколько и нужно учитывать значения производных, интерполяционный полином должен сочетать в себе черты обоих крайних своих частных случаев, то есть полинома Лагранжа и полинома Тэйлора. Эта тема не входит в начальные курсы анализа или алгебры, видимо, вследствие общепринятого мнения о громоздкости ответа, но при излагаемом здесь подходе она вполне доступна; кроме того, она пригодится во втором семестре для изучения функций от матриц.

Простейшая гибридная ситуация возникает с двумя контрольными точками (x_0, y_0) и (x_1, y_1), когда задано также значение производной $y_1' = y'(x_1)$. Возьмём определитель в (3), вставим половинки при квадратах и устремим $x_2 \to x_1$. Конфлюэнтный предел даёт уравнение

$$\begin{vmatrix} 1 & x & \frac{1}{2}x^2 & y \\ 1 & x_0 & \frac{1}{2}x_0^2 & y_0 \\ 1 & x_1 & \frac{1}{2}x_1^2 & y_1 \\ 0 & 1 & x_1 & y_1' \end{vmatrix} = 0. \tag{6}$$

При $x_0 \neq x_1$ отсюда можно выразить y, но не нужно: обобщать лучше именно в свёрнутом виде.

Если устремить $x_1 \to x_0$, то после необходимых конфлюэнтных преобразований определителя из (6) в пределе получим (5). Тут формулы на допредельной стадии (пока лишь для y_0'') посложнее уже обсуждавшихся и известны главным образом в разностном исчислении; однако они не нужны, поскольку общий результат предсказуем и ясен. Проверку в этом случае оставим читателю.

Теперь мы можем сразу писать ответы для более сложных случаев интерполяции Эрмита. Например, когда заданы контрольные точки (x_0, y_0) и (x_1, y_1) вместе со значениями y_0' и y_0'', получим уравнение

$$\begin{vmatrix} 1 & x & \frac{1}{2}x^2 & \frac{1}{6}x^3 & y \\ 1 & x_0 & \frac{1}{2}x_0^2 & \frac{1}{6}x_0^3 & y_0 \\ 0 & 1 & x_0 & \frac{1}{2}x_0^2 & y_0' \\ 0 & 0 & 1 & x_0 & y_0'' \\ 1 & x_1 & \frac{1}{2}x_1^2 & \frac{1}{6}x_1^3 & y_1 \end{vmatrix} = 0,$$

а когда в этих же точках заданы значения y_0' и y_1', получим уравнение

$$\begin{vmatrix} 1 & x & \frac{1}{2}x^2 & \frac{1}{6}x^3 & y \\ 1 & x_0 & \frac{1}{2}x_0^2 & \frac{1}{6}x_0^3 & y_0 \\ 0 & 1 & x_0 & \frac{1}{2}x_0^2 & y_0' \\ 1 & x_1 & \frac{1}{2}x_1^2 & \frac{1}{6}x_1^3 & y_1 \\ 0 & 1 & x_1 & \frac{1}{2}x_1^2 & y_1' \end{vmatrix} = 0.$$

Во всех случаях алгебраическое дополнение элемента y отлично от нуля, если все x_k различны.

Упражнение. *Найдите полином минимальной степени, требуя, чтобы в точках $x = 0$ и $x = 1$ его значения и первые производные были как у функции $\cos(\pi x)$.*

Глава 6. КВАДРАТИЧНЫЕ ФОРМЫ

6.1. Введение и мотивация

Векторы, координаты и однородные функции

В геометрии и опирающейся на неё физике регулярно встречаются разнообразные выражения, построенные из векторов, например: $\mathbf{r}$, $\mathbf{b}\cdot\mathbf{r}$, $\mathbf{r}\cdot\mathbf{r}$, $\mathbf{b}\times\mathbf{r}$, $\mathbf{r}\times(\mathbf{b}\times\mathbf{r})$. Выражаемая величина может быть скаляром, вектором, или даже чем-то более сложным и пока вам не знакомым (тензором). Характерной общей чертой многих таких выражений является однородность. В общем виде это свойство можно записать как $f(\lambda\mathbf{r}) = \lambda^k f(\mathbf{r})$, то есть при изменении длины переменного вектора $\mathbf{r}$ в λ раз зависящая от него величина изменяется в λ^k раз, и число k называют показателем однородности (или же степенью, порядком).

В этой главе мы изучим некоторые методы для работы с записью подобных величин в декартовых координатах. Здесь удобны индексные обозначения для координат. Выбрав базис $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, раскладываем по нему любой вектор как $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$. Будем также записывать набор координат в виде столбца: $X = [x_1, x_2, x_3]^\top$. Тогда, скажем, скалярное произведение выражается несколькими записями:

$$\mathbf{b}\cdot\mathbf{x} = b_1x_1 + b_2x_2 + b_3x_3 = \sum b_ix_i = B^\top X.$$

Тут даны полностью развёрнутое выражение и затем два его сокращения — с помощью суммирования и произведения матриц. Считая вектор $\mathbf{b}$ фиксированным, получим однородную функцию первой степени от вектора $\mathbf{x}$, или **линейную форму** от его координат. Всякая линейная форма имеет такой вид с соответствующим $\mathbf{b}$.

Скалярный квадрат переменного вектора, то есть квадрат его длины, является простейшим примером **квадратичной формы**:

$$\mathbf{x}\cdot\mathbf{x} = x_1^2 + x_2^2 + x_3^2 = \sum x_i^2 = X^\top X.$$

Теперь возьмём скалярное произведение двух переменных векторов:

$$\mathbf{x}\cdot\mathbf{y} = x_1y_1 + x_2y_2 + x_3y_3 = \sum x_iy_i = X^\top Y.$$

Эта величина является линейной формой как от (координат) $\mathbf{x}$, так и от (координат) $\mathbf{y}$. Поэтому её называют **билинейной формой**. Индексная и матричная записи в этом случае проще общего и не раскрывают его.

Привлекая постоянный вектор $\mathbf{b}$, можно соорудить другие билинейные формы. Векторное произведение $\mathbf{x}\times\mathbf{y}$ линейно по каждому аргументу, но не является билинейной формой, ибо результат не скаляр, а тоже вектор:

$$\mathbf{x} \times \mathbf{y} = (x_2y_3 - x_3y_2)\,\mathbf{e}_1 + (x_3y_1 - x_1y_3)\,\mathbf{e}_2 + (x_1y_2 - x_2y_1)\,\mathbf{e}_3.$$

Однако смешанное произведение $(\mathbf{b}, \mathbf{x}, \mathbf{y}) = \mathbf{b} \cdot (\mathbf{x} \times \mathbf{y})$ даёт пример билинейной формы от аргументов $\mathbf{x}$ и $\mathbf{y}$. В координатах имеем

$$\mathbf{b} \cdot (\mathbf{x} \times \mathbf{y}) = b_1x_2y_3 - b_1x_3y_2 + b_2x_3y_1 - b_2x_1y_3 + b_3x_1y_2 - b_3x_2y_1.$$

Трюк, позволяющий сокращать это выражение и предыдущее с помощью суммирования, вы увидите в физике многократно. Полезно сопоставить его с формулой полного раскрытия определителя. ε_{ijk}

Попробуем записать последнюю билинейную форму в матричном виде. В отличие от скалярного произведения выше, индексы координат двух переменных векторов меняются независимо друг от друга. Поэтому мы имеем дело с двойным суммированием и вводим новое обозначение для коэффициентов:

$$\sum_i \sum_j a_{ij}x_iy_j = X^\top A Y.$$

Из развёрнутого выражения считываем компоненты a_{ij} матрицы

$$A = \begin{bmatrix} 0 & b_3 & -b_2 \\ -b_3 & 0 & b_1 \\ b_2 & -b_1 & 0 \end{bmatrix}.$$

Можно заметить, что в матричном произведении X стоит левее A. Это лучше соответствует векторной записи $-\mathbf{x} \cdot (\mathbf{b} \times \mathbf{y})$, или даже $\mathbf{x} \cdot (-\mathbf{b} \times \mathbf{y})$, имеющим то же значение, что и $\mathbf{b} \cdot (\mathbf{x} \times \mathbf{y})$. Матрица A коэффициентов этой билинейной формы **кососимметрична**, поскольку $A^\top = -A$.

Отвлекаясь от рассматриваемого примера и конкретной матрицы, мы увидели тут общий вид билинейной формы, кратко записанный через суммирование и через матричное произведение. А выше, в случае стандартного скалярного произведения, матрица коэффициентов билинейной формы единична и тем самым скрыта: $X^\top E Y = X^\top Y$.

Векторное произведение как линейный оператор

Предыдущие рассуждения ведут к очень полезному наблюдению, притом уходя от основной темы главы. А именно, при фиксированном векторе $\mathbf{b}$ операция векторного произведения переводит произвольный вектор $\mathbf{y}$ в новый вектор $\mathbf{b} \times \mathbf{y}$, причём переводит линейно:

$$\mathbf{b} \times (\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{b} \times \mathbf{x} + \beta\mathbf{b} \times \mathbf{y}.$$

Подобные операции называют **линейными операторами**, часто просто операторами. Дальше в этом курсе им будет уделено специальное внимание. После выбора базиса пространства всякий линейный оператор представляется матрицей. Глядя на выписанную выше матрицу, можно понять, что оператор $\mathbf{y} \mapsto \mathbf{b} \times \mathbf{y}$ имеет матрицу

$$\hat{B} = \begin{bmatrix} 0 & -b_3 & b_2 \\ b_3 & 0 & -b_1 \\ -b_2 & b_1 & 0 \end{bmatrix}.$$

В физике вектор $\mathbf{b} \times \mathbf{y}$ называют (угловым) моментом вектора $\mathbf{y}$ относительно вектора $\mathbf{b}$, который должен быть ненулевым. Первыми в ходе обучения обычно возникают момент импульса и момент силы.

Тензор моментов инерции

При изучении вращательного движения моменты инерции играют роль, аналогичную массе для поступательного. Ситуация осложняется зависимостью от направления оси вращения; именно этим аспектом вопроса мы и займёмся.

Когда точечная масса m вращается вокруг оси, находящейся на расстоянии R, момент инерции равен $I = mR^2$. Для системы из нескольких масс соответственно имеем $I = \sum m_i R_i^2$. В более реалистичной постановке вращается твёрдое тело (причём вокруг своего центра масс), поэтому точечные массы нужно заменить элементом массы $dm = \rho\, dV$, а сумму — интегрированием по объёму, о чём говорится не только в курсе механики, но и в курсе основ математического анализа. Здесь нам важна алгебраическая сторона дела, проявляющаяся уже для точечной массы, а про интегралы можно вспомнить в конце вычисления.

Поэтому зафиксируем неподвижную систему отсчёта, поместим единичную массу в точку с радиус-вектором $\mathbf{b}$ и найдём момент инерции при вращении её вокруг прямой, проходящей через начало координат параллельно вектору $\mathbf{x}$. Расстояние от точки до оси вычислим по формуле из главы 2:

$$R = \frac{|\mathbf{b} \times \mathbf{x}|}{|\mathbf{x}|}.$$

Удобно избавиться от знаменателя, выбирая направляющий вектор $\mathbf{x}$ единичной длины. Тогда искомый момент инерции равен $(\mathbf{b}\times\mathbf{x})\cdot(\mathbf{b}\times\mathbf{x})$. Его зависимость от направления оси выражена квадратичной формой.

Выпишем её в координатах и затем выделим матрицу коэффициентов тремя способами. Первый — развёрнутое вычисление:

$$\begin{aligned}|\mathbf{b}\times\mathbf{x}|^2 &= (b_2x_3-b_3x_2)^2+(b_3x_1-b_1x_3)^2+(b_1x_2-b_2x_1)^2\\ &= b_2^2x_3^2+b_3^2x_2^2-2b_2b_3x_2x_3\\ &+ b_3^2x_1^2+b_1^2x_3^2-2b_3b_1x_3x_1\\ &+ b_1^2x_2^2+b_2^2x_1^2-2b_1b_2x_1x_2,\end{aligned}$$

откуда

$$|\mathbf{b}\times\mathbf{x}|^2 = X^\top A\,X,\ \text{где } A=\begin{bmatrix} b_2^2+b_3^2 & -b_1b_2 & -b_1b_3\\ -b_2b_1 & b_3^2+b_1^2 & -b_2b_3\\ -b_3b_1 & -b_3b_2 & b_1^2+b_2^2\end{bmatrix}.$$

В отличие от билинейных форм, запись квадратичной формы через матричное произведение неоднозначна; например, матрицы

$$\begin{bmatrix}0&2\\0&0\end{bmatrix},\qquad \begin{bmatrix}0&1\\1&0\end{bmatrix},\qquad \begin{bmatrix}0&0\\2&0\end{bmatrix}$$

соответствуют одной и той же форме $2x_1x_2$. Этого неудобства избегают, всегда выбирая симметричную матрицу, $A^\top = A$, которая находится однозначно.

Второй способ хитрее и короче. Из полученного выше представления о векторном произведении как операторе позаимствуем матричную запись $\mathbf{b}\times\mathbf{x} = \hat{B}X$. Тогда

$$(\mathbf{b}\times\mathbf{x})\cdot(\mathbf{b}\times\mathbf{x}) = (\hat{B}X)^\top(\hat{B}X) = X^\top\hat{B}^\top\hat{B}X.$$

Вычисляя произведение $\hat{B}^\top\hat{B}$, увидим в точности матрицу A, найденную по первому способу.

В третьем способе воспользуемся тождеством

$$(\mathbf{b}\times\mathbf{x})\cdot(\mathbf{b}\times\mathbf{x}) = (\mathbf{b}\cdot\mathbf{b})(\mathbf{x}\cdot\mathbf{x}) - (\mathbf{x}\cdot\mathbf{b})(\mathbf{b}\cdot\mathbf{x}).$$

Первое слагаемое правой части пропорционально квадратичной форме, выражающей квадрат длины, с коэффициентом $B^\top B = b_1^2+b_2^2+b_3^2$, поэтому соответствует скалярной матрице

$$(b_1^2+b_2^2+b_3^2)E = \begin{bmatrix} b_1^2+b_2^2+b_3^2 & 0 & 0\\ 0 & b_1^2+b_2^2+b_3^2 & 0\\ 0 & 0 & b_1^2+b_2^2+b_3^2\end{bmatrix}.$$

Второе слагаемое легко переписать в матричном виде как $X^\top BB^\top X$, где

$$BB^\top = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} = \begin{bmatrix} b_1b_1 & b_1b_2 & b_1b_3 \\ b_2b_1 & b_2b_2 & b_2b_3 \\ b_3b_1 & b_3b_2 & b_3b_3 \end{bmatrix}.$$

Сравнивая с матрицей A, найденной по первому способу, лучше понимаем её структуру.

Для твёрдого тела выражения с постоянными компонентами вектора $\mathbf{b}$ нужно заменить на интегралы по телу. Если начало координат находится в центре масс тела, то получаемую квадратичную форму называют тензором моментов инерции тела. Диагональные компоненты её/его матрицы I есть моменты инерции относительно координатных осей, а внедиагональные (с противоположным знаком) называют центробежными моментами. Например,

$$I_{11} = \iiint_T (t_2^2 + t_3^2)\rho(t_1, t_2, t_3)\, dV, \quad -I_{23} = \iiint_T t_2t_3\rho(t_1, t_2, t_3)\, dV.$$

Смысл слова «тензор» в том, что речь идёт о физической величине особого типа; для её координатных выражений в разных системах координат есть свои правила, уже знакомые в случае векторов и изучаемые в этой главе в случае билинейных и квадратичных форм.

6.2. Билинейные и квадратичные формы

От матрицы к форме

В этой главе мы будем изучать выражения

$$f(X, Y) = X^\top A\, Y,$$
$$q(X) = f(X, X) = X^\top A\, X$$

со столбцами $X = [x_1, \dots, x_n]^\top$ и $Y = [y_1, \dots, y_n]^\top$ из $\mathbb{R}^n$ в качестве независимых переменных и с $n \times n$ матрицей коэффициентов $A = [a_{ij}]$. Раскрыв матричное произведение, получим

$$f(X, Y) = X^\top A\, Y = \sum_{1 \leqslant i,j \leqslant n} a_{ij} x_i y_j.$$

Непосредственно проверяется, что

$$f(\alpha_1 X^{(1)} + \alpha_2 X^{(2)}, Y) = \alpha_1 f(X^{(1)}, Y) + \alpha_2 f(X^{(2)}, Y),$$
$$f(X, \beta_1 Y^{(1)} + \beta_2 Y^{(2)}) = \beta_1 f(X, Y^{(1)}) + \beta_2 f(X, Y^{(2)}).$$

Такая функция называется билинейной, как напоминает наш первый

Пример. Стандартное скалярное произведение на $\mathbb{R}^3$, вычисляемое по координатам в стандартном ОНБ по правилу

$$X^\top Y = x_1y_1 + x_2y_2 + x_3y_3.$$

Здесь матрицы A явно не видно: она единичная, а $X^\top EY = X^\top Y$.

Прежде в математике широко употреблялось слово «форма»; в этой тематике оно выжило и мы будем называть $f(X, Y)$ **билинейной формой**, а $q(X)$ **квадратичной формой**.

Пример. Интервал между событиями в пространстве Минковского вычисляется через квадратичную форму

$$q(X) = x_0^2 - x_1^2 - x_2^2 - x_3^2.$$

От формы к матрице

Утверждение. *Для всякой билинейной функции $f(X, Y)$ столбцов найдётся такая матрица A, что $f(X, Y) = X^\top AY$.*

Доказательство. Разложим столбцы по стандартому базису:

$$X = \sum_{1 \leqslant i \leqslant n} x_i E^{(i)}, \quad Y = \sum_{1 \leqslant j \leqslant n} y_j E^{(j)}.$$

Ввиду билинейности, как и в случае обычного скалярного произведения, $f(X, Y)$ выразится через координаты и значения $f(E^{(i)}, E^{(j)})$:

$$\begin{aligned} f(X,Y) &= f\left(\sum_{1 \leqslant i \leqslant n} x_i E^{(i)}, \sum_{1 \leqslant j \leqslant n} y_j E^{(j)}\right) \\ &= \sum_{1 \leqslant i \leqslant n} x_i f\left(E^{(i)}, \sum_{1 \leqslant j \leqslant n} y_j E^{(j)}\right) \\ &= \sum_{1 \leqslant i,j \leqslant n} x_i y_j f(E^{(i)}, E^{(j)}). \end{aligned}$$

Искомую матрицу A нужно составить из $a_{ij} = f(E^{(i)}, E^{(j)})$. □

Симметричные формы

Определение. Билинейная форма f называется **симметричной**, если

$$f(X, Y) = f(Y, X)$$

для всех столбцов X, Y.

Утверждение. *Билинейная форма симметрична тогда и только тогда, когда симметрична её матрица.*

Доказательство. Подставляя столбцы стандартного базиса в аргументы симметричной формы, получим симметричность её матрицы.

Наоборот, если $A^\top = A$, то

$$f(X,Y)^\top = (X^\top A\,Y)^\top = Y^\top A^\top X = Y^\top A\,X = f(Y,X).$$

Но $f(X,Y)$ есть не чувствующий транспонирования скаляр. □

Утверждение. *Всякая квадратичная форма задаётся симметричной матрицей.*

Доказательство. Возьмём квадратичную форму $q(X) = X^\top A\,X$ с произвольной матрицей и проверим, что $q(X) = X^\top B\,X$ с симметричной матрицей $B = \frac{1}{2}(A + A^\top)$. □

Таким образом, изучение квадратичных форм и изучение симметричных билинейных форм это практически одно и то же.

Инвариантные определения

Определяющее свойство билинейной формы — линейность по каждому из двух аргументов — легко переписать в вид, избегающий изначального выбора базиса пространства, а потому пригодный не только для пространства столбцов, но и для любого (конечномерного) линейного пространства. Достаточно поменять даже не буквы, а шрифт:

$$\begin{aligned} f(\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2, \mathbf{y}) &= \alpha_1 f(\mathbf{x}_1, \mathbf{y}) + \alpha_2 f(\mathbf{x}_2, \mathbf{y}), \\ f(\mathbf{x}, \beta_1\mathbf{y}_1 + \beta_2\mathbf{y}_2) &= \beta_1 f(\mathbf{x}, \mathbf{y}_1) + \beta_2 f(\mathbf{x}, \mathbf{y}_2). \end{aligned}$$

Переход к матричной записи $f(\mathbf{x},\mathbf{y}) = X^\top A\,Y$ в выбранном базисе $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ задаётся формулой $a_{ij} = f(\mathbf{e}_i, \mathbf{e}_j)$.

Поступить аналогично с квадратичными формами не удаётся. Формально они вводятся только посредством билинейных. При этом, если матричные виды $X^\top A\,Y$ и $X^\top A\,X$ явно указывают на близкое родство двух понятий, инвариантное определение квадратичной формы вынужденно идёт окольным путём. Это такая функция $\mathbf{x} \mapsto q(\mathbf{x})$, что функция

$$f(\mathbf{x},\mathbf{y}) = \frac{1}{2}\big(q(\mathbf{x}+\mathbf{y}) - q(\mathbf{x}) - q(\mathbf{y})\big)$$

является билинейной формой.

Поведение при смене базиса

Стандартный базис пространства столбцов может быть неудобным для работы с конкретной билинейной или квадратичной формой. Пример тесно связанного явления мы видели, занимаясь приведением линий второго порядка к каноническому виду.

При смене базиса матрица перехода S связывает столбцы старых и новых координат соотношениями $X = SX'$ и $Y = S\,Y'$. Подставляя эти выражения в билинейную форму от старых переменных, мы получим билинейную форму от новых переменных. Сравнивая записи в двух базисах:

$$X^\top A\,Y = (SX')^\top A\,(S\,Y') = (X')^\top S^\top A\,S\,Y' = (X')^\top A'\,Y',$$

приходим к закону изменения матрицы билинейной формы:

$$A' = S^\top A\,S.$$

Таков же закон изменения матрицы квадратичной формы.

Определение. Такую пару матриц A и A' называют **конгруэнтными**.

Следствие. *Ранги конгруэнтных матриц равны.*

Доказательство. Матрица перехода всегда обратима и умножение на неё не меняет ранг. □

Определение. Пару билинейных форм, получаемых одна из другой заменой переменных, называют **эквивалентными**.

В последующем полезно изменить точку зрения и считать эквивалентные билинейные формы разными *видами* одной и той же функции пары векторов, получающимися при выборе разных базисов в $\mathbb{R}^n$.

Канонический вид симметричной формы

Определение. Если матрица A симметричной билинейной формы f в каком-то базисе $\mathcal{B}$ диагональна, то $\mathcal{B}$ называют **каноническим базисом** для f и говорят, что в этом базисе f имеет **канонический вид**.

На практике квадратичные формы приводят к каноническому виду поочерёдным избавлением от недиагональных слагаемых ($i \neq j$):

$$\sum_{1\leqslant i,j\leqslant n} a_{ij}x_ix_j \rightsquigarrow \sum_{1\leqslant i\leqslant r} c_ix_i^2.$$

`Шаг 1:` если канонический вид ещё не получен, найти такую пару индексов $i < j$, что $a_{ij} \neq 0$.

`Шаг 2:` если $a_{ii} \neq 0$, прыгнуть на `шаг 4`.
`Шаг 3:` сделать замены $x_i = x'_i + x'_j$ и $x_j = x'_i - x'_j$, затем стереть штрихи (чтобы штрихи не накапливались).
`Шаг 4:` теперь квадратичная форма имеет вид

$$f = ax_i^2 + 2azx_i + g = a(x_i + z)^2 - az^2 + g,$$

причём $a \neq 0$, а z и g не зависят от x_i; подставить $x_i = x'_i - z$, затем стереть штрихи и повторить `шаг 1`.

В основе метода — выделение полных квадратов на шаге 4, а возникающее препятствие обходится на шаге 3. Чтобы в итоге получить не только канонический вид формы, но и соответствующий ей канонический базис, к этому рецепту нужно добавить ещё несложное средство отслеживания всех произведённых замен.

Теорема (Lagrange). *Для всякой квадратичной формы найдётся канонический базис.*

Доказательство. Метод выделения квадратов. □

6.3. Вещественные квадратичные формы

Сигнатура

Следствие. *Для всякой квадратичной формы q ранга r на пространстве $\mathbb{F}^n$, где $\mathbb{F} = \mathbb{R}$ или $\mathbb{C}$, найдётся базис, в котором*

$$q(\mathbf{x}) = c_1 x_1^2 + \ldots + c_r x_r^2, \qquad \text{где } c_i \in \mathbb{F}^\times = \mathbb{F} \setminus \{0\}.$$

В случае $\mathbb{F} = \mathbb{C}$ каждая квадратичная форма приводится к сумме квадратов, поскольку должной заменой все константы c_i упрятываются внутрь x_i. Случай $\mathbb{F} = \mathbb{R}$ сложнее: нельзя избавиться от минусов.

Следствие. *Для всякой вещественной квадратичной формы q найдётся базис, в котором она принимает* **нормальный вид**

$$q(\mathbf{x}) = x_1^2 + \ldots + x_p^2 - x_{p+1}^2 - \ldots - x_r^2.$$

виды форм на $\mathbb{R}^2$ и $\mathbb{R}^3$

Теорема (закон инерции). *Числа p и r, определяющие нормальный вид вещественной квадратичной формы q, зависят только от q, но не от выбора базиса, приводящего её к нормальному виду.*

Определение. Возникающие инварианты имеют особые названия:

r	**Индекс инерции** (ранг)
p	**Положительный** индекс инерции
$r-p$	**Отрицательный** индекс инерции
$2p-r$	**Сигнатура**

Часто сигнатурой также называют пару $(p, r-p)$, что информативнее.

Определение. В следующей таблице приведены названия важных специальных типов вещественных квадратичных форм на $\mathbb{R}^n$ с крайними значениями индексов. В нормальном виде таких форм априорное разнообразие $\{-1, 0, 1\}$ коэффициентов ещё более ограничено.

$r = n$	**Невырожденной**	
$p = n$	**Положительно определённой**	$\mathbf{x} \neq \mathbf{0} \implies q(\mathbf{x}) > 0$
$r - p = n$	**Отрицательно определённой**	$\mathbf{x} \neq \mathbf{0} \implies q(\mathbf{x}) < 0$
$r - p = 0$	**Неотрицательно полуопределённой**	$q(\mathbf{x}) \geqslant 0$
$p = 0$	**Неположительно полуопределённой**	$q(\mathbf{x}) \leqslant 0$

Доказательство закона инерции. Инвариантность ранга уже получена выше. Инвариантность положительного индекса инерции докажем от противного: допустим, что в базисе $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ форма имеет вид

$$q(\mathbf{x}) = x_1^2 + \ldots + x_s^2 - x_{s+1}^2 - \ldots - x_r^2,$$

а в базисе $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$ она же имеет вид

$$q(\mathbf{x}) = x_1^2 + \ldots + x_t^2 - x_{t+1}^2 - \ldots - x_r^2,$$

причём $s > t$. Тогда сужения q на линейные оболочки $\mathcal{L} = \langle \mathbf{v}_1, \ldots, \mathbf{v}_s \rangle$ и $\mathcal{L}' = \left\langle \mathbf{v}'_{t+1}, \ldots, \mathbf{v}'_n \right\rangle$ соответственно будут положительно определённая и неположительно полуопределённая квадратичные формы.

Если есть ненулевой вектор $\mathbf{x} \in \mathcal{L} \cap \mathcal{L}'$, то получается противоречие: $q(\mathbf{x})$ одновременно положительно и неположительно. Если же ненулевого общего вектора нет, то по определению никакие нетривиальные линейные комбинации $\alpha_1 \mathbf{v}_1 + \ldots + \alpha_s \mathbf{v}_s \in \mathcal{L}$ и $\alpha'_{t+1} \mathbf{v}'_{t+1} + \ldots + \alpha'_n \mathbf{v}'_n \in \mathcal{L}'$ не могут быть равны. Тогда множество $\{\mathbf{v}_1, \ldots, \mathbf{v}_s, \mathbf{v}'_{t+1}, \ldots, \mathbf{v}'_n\}$ из $s + n - t > n$ векторов линейно независимо, что невозможно. $\square$

Разные свойства(?)

Утверждение. *Положительная определённость квадратичной формы q сохраняется при сужении на любое подпространство.*

Доказательство. Сохраняется свойство $q(\mathbf{x}) > 0$ для всех $\mathbf{x} \neq \mathbf{0}$. □

Пример. Невырожденность формы может потеряться при сужении на подпространство. Например, сужение формы $x_1^2 - x_2^2$ (в стандартном базисе плоскости $\mathbb{R}^2$) на прямую $\langle [1,1]^\top \rangle$ тождественно нулевое.

Утверждение. *Для всякой симметричной билинейной формы f равносильны следующие свойства:*

(1) ассоциированная квадратичная форма вырождена;
(2) существует такой вектор $\mathbf{y} \neq \mathbf{0}$, что $f(\mathbf{x}, \mathbf{y}) = 0$ для всех векторов $\mathbf{x} \in \mathbb{R}^n$.

Доказательство. $(1 \Rightarrow 2)$ Годится вектор канонического базиса, соответствующий нулевому диагональному элементу матрицы формы.

$(2 \Rightarrow 1)$ Дополнив $\mathbf{y}$ произвольно до базиса всего пространства, мы представим форму f матрицей с нулевым столбцом. □

Лемма. *Определители любых двух конгруэнтных матриц имеют одинаковые знаки.*

Доказательство. Если $A' = S^\top A\, S$, то $\det A' = (\det S)^2 \det A$. □

Следствие. *Матрица положительно определённой формы в любом базисе имеет положительный определитель.*

Главные миноры, метод Якоби и критерий Сильвестра

Определение. **Главным минором** матрицы A называется минор $\Delta_k(A)$, составленный из её первых k строк и первых k столбцов.

Следствие. *Главный минор $\Delta_k(F)$ матрицы квадратичной формы в базисе $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ идентичен матрице сужения этой формы на подпространство $\langle \mathbf{v}_1, \ldots, \mathbf{v}_k \rangle$.*

С помощью главных миноров $\Delta_k(F)$ матрицы квадратичной формы часто удаётся быстро узнать её сигнатуру. Удобно не различать в обозначениях минор (матрицу) и его численное значение, а также полагать $\Delta_0 = 1$ и $\Delta_n = \det F$, причисляя их тоже к главным минорам.

Теорема. *Если все главные миноры матрицы вещественной квадратичной формы q имеют ненулевые значения $\Delta_0, \ldots, \Delta_n$, то количества сохранений и перемен знака в этой последовательности равны положительному и отрицательному индексам инерции формы q.*

Доказательство. Докажем даже, что такую форму можно привести к нормальному виду $\sum c_k z_k^2$, в котором знаки коэффициентов согласованы с переменами знаков главных миноров в том смысле, что

$$\operatorname{sgn} c_k = \operatorname{sgn}(\Delta_k/\Delta_{k-1})$$

для $1 \leqslant k \leqslant n$. Отсюда утверждение об индексах следует простым подсчётом количеств плюсов и минусов.

Воспользуемся индукцией по числу переменных. Приведём исходную форму к нормальному виду в три этапа. Сначала игнорируем слагаемые с последней переменной; все прочие составляют квадратичную форму от $n-1$ переменной с теми же главными минорами, кроме отсутствующего Δ_n. Предполагаем по индукции, что она приведена к согласованному нормальному виду в переменных y_k:

$$\left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} \pm1 & 0 & 0 & b_{14} \\ 0 & \pm1 & 0 & b_{24} \\ 0 & 0 & \pm1 & b_{34} \\ \hline b_{41} & b_{42} & b_{43} & b_{44} \end{array}\right].$$

Теперь задача весьма аналогична той, что мы решали, двигая линии второго порядка. Переходя к переменным

$$z_k = \begin{cases} y_k \pm b_{kn} y_n & \text{при } k < n, \\ y_k & \text{при } k = n, \end{cases}$$

получим канонический вид исходной формы в переменных z_k:

$$\left[\begin{array}{ccc|c} \pm1 & 0 & 0 & b_{14} \\ 0 & \pm1 & 0 & b_{24} \\ 0 & 0 & \pm1 & b_{34} \\ \hline b_{41} & b_{42} & b_{43} & b_{44} \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} \pm1 & 0 & 0 & 0 \\ 0 & \pm1 & 0 & 0 \\ 0 & 0 & \pm1 & 0 \\ \hline 0 & 0 & 0 & c_4 \end{array}\right].$$

Наконец, заменой последней переменной сведём c_n к ±1. По лемме о сохранении знака, определитель D полученной диагональной матрицы имеет тот же знак, что и Δ_n. Итак,

$$\operatorname{sgn} \Delta_n = \operatorname{sgn} D = \operatorname{sgn} \Delta_{n-1} \cdot \operatorname{sgn} c_n,$$

что и требовалось. □

Следствие (критерий Сильвестра)**.** *Для всякой вещественной квадратичной формы равносильны утверждения:*

(1) *она положительно определённая;*
(2) *в любом базисе все её главные миноры положительны.*