

Содержание

Введение	3
Теханимация и математика	3
Структура курса	5
О языке	6
Система координат. Наивный подход	7
Мотивировка	7
Координаты идеально	8
Координаты символично	9
FAQ	10
Функции	11
Функция как формула	11
Функция как соответствие	12
График функции	13
Кривые и поверхности как графики функций	14
Примеры функций и их графиков	15
Тригонометрия	17
Мотивация	17
Астрономия и хорды	18
Отношения в треугольнике	19
Эйлер и тригонометрия	20
Связь всего со всем	21
Мотивация	21
Отношения в прямоугольном треугольнике	21
Точка на окружности	22
Графики	23
Обратные тригонометрические функции	25
Зачем это все?	26
Криволинейные системы координат	28
Полярная система координат	28
Сферическая система координат	28
Цилиндрическая система координат	28
Векторы	29
История	29
Развилки в понимании	30
Наивное определение и базовые операции	31
Умножение векторов	33
Вектор и координатная запись	38
Аффинные преобразования	45
Аффинные системы координат	45
Трансформы на практике	53
Матрицы	57
Углы Эйлера	58

Откуда взялись кватернионы, и вообще?	62
Действительные числа	62
Комплексные числа	66
Sir W.R.Hamilton и кватернионы	68
Геометрия кватернионов	68

Введение

Теханимация и математика

Разберемся с тем, что является задачей теханиматора и какое место занимает математика в его работе.

Игра работает следующим образом:

- движок загружает данные о сцене, начинается бесконечный цикл вычисления фреймов
- на основе этих данных движок вычисляет на CPU данные для шейдеров (программ для вычисления цветов на GPU)
- на GPU вычисляются шейдеры, в итоге выдавая матрицу цветов с нужным разрешением, эту картинку мы видим на экране
- на следующем кадре на основе запрограммированной логики данные обновляются, все повторяется

С математической точки зрения, можно эту схему еще упростить до следующей

- на вход в движок подступает набор данных (сценарии, текстуры, анимации, акторы, их параметры и т.п.)
- на выходе получаем картинку (матрицу цветов)

То есть движок осуществляет некоторое сложное преобразование различных данных в достаточно понятный результат — цвета. Это преобразование можно назвать функцией (хотя дотошные математики скорее использовали бы слово "отображение"). Функция эта в таком виде — трудносязаемая вещь: мало кто может похвастаться знанием всех аспектов реализации движка и игровой логики. Однако, если декомпозировать эту функцию на множество подфункций, которыми занимаются более узкоспециализированные люди, то про них уже можно говорить более конкретно.

В случае с теханиматорами подфункция следующая. Каждый кадр в нетворк приходит набор параметров, которые преобразуются в изменение трансформации некоторых объектов, как правило, это трансформации твердого тела. В нашей ситуации это как правило кость.

Кость это вспомогательный объект в моделировании персонажей, вместе кости образуют скелет. Поза скелета, т.е. совокупность позиций его костей определяет геометрическую форму меша. Роль костей заключается именно в возможности удобной анимации персонажа: без костей пришлось бы анимировать отдельно каждую вершину меша.

* * *

Итак, сформулируем утверждение: задача теханиматора состоит в определении функции изменения положения геометрического объекта.

Тогда встают естественные вопросы:

- Как задать положение объекта?
- Как задать изменение положения объекта?

Для ответов на эти вопросы зададимся еще одним: что такое положение объекта?

Но, вообще, что такое объект с точки зрения геометрии? Это множество точек, определенным образом расположенных в пространстве. Это значит, что между этими точками есть некоторые отношения, например, в виде расстояний между ними.

В случае, если мы имеем дело с твердым телом, суть которого в том, что расстояния между всеми точками объекта сохраняются при любых его преобразованиях, то для описания движения всего объекта достаточно задать преобразование положения только для одной точки.

Движение объекта задается его поворотом и сдвигом относительно некоторой точки. Сдвиг определяется вектором, а поворот можно задать матрицей, углами Эйлера или кватернионами. Отсюда и возникает вся эта история с этими понятиями, которые находятся на стыке алгебры и геометрии.

С этого момента мы начнем разбираться с тем, откуда берутся векторы, матрицы, кватернионы, заглянем в их алгебраическую природу и постараемся полностью разобраться, как это все работает.

Изложение рассчитано на то, что читатель имеет школьное образование и некоторый опыт работы с теханимацией (то есть "знаю/слышал, что это есть и работает, но как — не понимаю"). Мы последовательно разберемся со всеми понятиями, при этом стараясь излишне не углубляться, но с другой стороны, будем придерживаться достаточно высокой степени строгости, чтобы каждый смог при желании получить стройное представление о математической стороне дела.

Структура курса

Задача курса состоит в том, чтобы разъяснить математическую сторону работы теханиматоров. Ключевые темы, которые мы ставим целью покрыть:

- работа с векторами (системы координат, скалярное и векторное произведения, проекции и т.д.)
- способы задания поворотов (матрицы, углы Эйлера, кватернионы)
- прочее: функции, тригонометрия, линейная интерполяция

* * *

Поэтому в первой части курса мы разберемся с темами, касающимися описание точек на плоскости и пространстве:

- Декартова система координат
- Функции
- Тригонометрия
- Полярная система координат, сферическая система координат, цилиндрическая система координат

затем введем понятие вектора и рассмотрим скалярное произведение векторов и его свойства:

- Вектор. Геометрическое представление, аналитико-геометрическое (координатное), алгебраическое (комплексные числа). Правило сложения векторов и умножения на скаляр.
- Скалярное произведение, угол между векторами, билинейность и симметричность. Длина вектора. Проекция вектора на прямую, на плоскость.

после этого можно будет изучать преобразование векторов с помощью матриц:

- Матрицы 2x2, 3x3. Умножение матрицы на вектор. Умножение матрицы на матрицу. Геометрический смысл.
- Скалярная матрица (масштабирующая). Матрица поворота. Обратная матрица. Как получить глобальные координаты, если есть локальные и наоборот.
- Определитель. Что такое ортогональные матрицы и почему мы всегда (наверно) используем именно их?
- Векторное произведение векторов, его геометрический смысл.

далее рассмотрим способы задания положения объекта:

- Твердое тело. Задание положения с помощью вектора и матрицы поворота.
- Задание положения с помощью вектора и углов Эйлера. Гимбал лок. Связь углов Эйлера с матрицей поворота.
- Линейная интерполяция. Сферическая интерполяция. (может и не здесь?)

и, наконец, поговорим о кватернионах:

- Числа. От единицы до кватернионов.
- Комплексные числа. Геометрическое представление двумерного вектора комплексным числом. Полярная форма. Фокусы с возведением в степень. Как задать поворот и сдвиг твердого (двумерного) тела с помощью комплексных чисел. Связь с двумерной матрицей поворота.
- Кватернионы. Сходства и различия с комплексными числами. Операции сложения и умножения. Модуль (длина) кватерниона. Связь с трехмерной матрицей поворота.

О языке

Математика — это искусство называть разные вещи одним и тем же именем.

— Jules Henri Poincaré

Любая область знаний это во многом язык: надо понимать слова и оперировать ими. Математика это демонстрирует очень хорошо, потому что она в принципе строится на абстракциях и на названиях этих абстракций. При этом каждое слово имеет четкое определение, которое очень желательно понимать, а это не всегда просто.

В нашем случае будет прекрасно видно, насколько язык нас способен запутывать, потому что мы будем постоянно находиться на стыке алгебры и геометрии, и часто одно слово будет иметь сразу два представления: алгебраическое и геометрическое. Здесь хочется сразу об этом предостеречь и дать пояснения.

Вообще, что такое алгебра и геометрия? Это не все математики понимают, а те, кто понимает, понимают по-разному. Но если углубиться в тему, то общие черты становятся вполне ясными.

До эпохи Возрождения, 15-16 века, многие математики считали, что единственную правильную математику это геометрия. Под этим термином понималось то, чем занимались древние греки, начиная со времен Пифагора. А пифагорейцы и их последователи считали, что числа выражают геометрические фигуры. И все рассуждения были через призму геометрических идеальных платоновских тел. Впрочем, можно сказать сильно проще: когда мы что-то рисуем, это геометрия, а числа просто описывают ее свойства.

При этом, конечно, была торговля, размены товарами, т.е. арифметика существовала и до Пифагора, еще в Египте и Вавилоне. И дальше уже индийцами и арабами развивались символические вычисления, решение уравнений, отрицательные числа и т.п. Это называлось алгеброй и считалось "трущими" европейскими математиками чем-то недостойным.

И вот, когда уже алгебра достигла результатов таких, что не считаться с ней было никак нельзя, два этих течения начали сливаться. Геометрия стала "цифровизироваться", потому что оказалось, что геометрические фигуры можно описывать алгебраическими уравнениями. Новую геометрию стали называть аналитической.

Процесс этот сейчас напоминает рефакторинг кода, когда был написан движок, но потом поняли, что надо менять концепцию и все переписывать под нее, но старое тоже нужно сохранить. Это может путать того, кто весь процесс не застал.

Так что нам полезно почаще себе задавать вопрос, о чём мы говорим, когда произносим те или иные слова: "точка", "система координат", "сдвиг" и т.д. Все эти слова

пришли из геометрической интуиции, но после "цифровизации" стали иметь строгие алгебраические определения.

Но есть и чисто алгебраические понятия, например, "матрица", "кватернион".

Система координат. Наивный подход

Мотивировка

Идея о том, что число что-либо говорит о геометрии, идет из древности — пифагорейцы считали, что все есть число. То есть все в мире можно как-то выразить числами, да и число в их учении понималось как геометрический объект, а не арифметический. Например, любое единое — это единица, отрезок — это двойка, треугольник — тройка, четверка это квадрат со стороной 2. Более сложные числа описывают все более замысловатые объекты. Тот же $\sqrt{2}$ это длина диагонали квадрата со стороной 1.

Эта концепция видеть за числами нечто большее, чем просто счеты, и лежит в основе математики как таковой. Далее этот мыслительный прием продлевается на все более высокие степени абстракции, крутится в играх разума так и сяк, в результате чего время от времени выкристализовываются удивительно красивые абстрактные объекты.

Что касается систем координат, то само по себе это изобретение древнее — всегда нужна была навигация. Как говорят специалисты, в "Географии" Птолемея (150 г. н.э.) приведено множество томов с координатами различных мест на карте Земли. Но то были координаты сферические — широта и долгота.

А вот привычная нам со школы и по работе с 3D-движками прямоугольная система координат с осями x, y, z и всеми атрибутами, сформировалась гораздо позже, между 16-18 веками. Идейно тут важен момент не придумывания типа системы координат, а именно понимание того, что система координат (любая) позволяет построить мостик между алгеброй и геометрией, т.е. между символными преобразованиями формул по определенным правилам и определением позиций точек в пространстве.

Считается, что первыми, кто применил эту идею, были Рене Декарт (René Descartes, 1596-1650) и Пьер Ферма (Pierre de Fermat, 1601-1665). Основным поворотным трудом, определяющим начало так называемой аналитической геометрии, считается "Геометрия" Декарта (1637) — приложение к философскому трактату "Рассуждение о методе, чтобы хорошо направлять свой разум и отыскивать истину в науках".

Правда, если открыть "Геометрию", то найти привычные оси координат не получится ни на одной странице. В чем же тогда дело?

Философия Декарта заключалась в том, что человеческие чувства могут подводить, на них опираться нельзя. Настоящие истины может производить только разум, производя чисто логические рассуждения. А геометрия в классическом ее понимании, в которую главную роль играют соотношения между длинами и углами, хитрые построения, чертежи (так называемая синтетическая геометрия, именно ее в основном мы и проходим в школе), представлялась Декарту скорее наукой чувственной, а значит, трудно проверяемой, ограниченной в своем потенциале.

При этом, уже неплохо развитая к тому времени алгебраическая техника, показывала альтернативный путь математики, когда утверждения носят чисто формальный характер. Его геометрический (вещественный) смысл совершенно может быть непонятен, но истинность его зависит лишь от корректности логических рассуждений, которую можно проверить.

Отсюда становится понятно, почему Декарту пришла в голову идея придать геометрии алгебраический характер. Это он и сделал в "Геометрии". Там он привел примеры решения не самых простых древних геометрических задач об описании кривых на плоскости в виде уравнений на неизвестные x и y .

Мы знаем из школы, например, что любая прямая на плоскости задается уравнением вида

$$y = kx + b,$$

а окружность радиуса R задается уравнением

$$x^2 + y^2 = R^2.$$

Декарт понял, что конические сечения (окружность, эллипс, гипербола, парабола), сильно волновавшие древних греков, задаются уравнениями второго порядка, т.е. следующего вида

$$ax^2 + bxy + cy^2 + dx + ey = 0,$$

и, более того, уравнение любого порядка от переменных x, y задает некоторую кривую на плоскости.

И это уже, пожалуй, первый очень заметный шаг, который демонстрирует мощь алгебраического подхода. Ведь до этого — удивительно об этом думать — считалось бессмысленным говорить о степенях выше третьей, так как непонятно было, что такое x^4 геометрически (x — отрезок длины x , xy — прямоугольник со сторонами x и y , xyz — параллелепипед со сторонами x, y, z). А теперь парадигма меняется: непонятно геометрически? Ну и ладно, зато это дает практический результат в той же геометрии кривых.

* * *

В общем, имя Декарта было необходимо выделить, чтобы показать переломный момент истории, который нас и привел к нынешнему образу жизни, включая все то, о чем мы планируем поговорить в курсе.

Говоря об истории, появляется желание перечислять все больше имен, мы этого делать не будем, но все же упомянем еще пару довольно известных личностей. Исаак Ньютона опирался на метод Декарта и Ферма, описывая траектории движения тел, подошел к понятию "функция" (вместе с Лейбницем) [6]. А уже во вполне современном виде о функциях, кривых, координатах писал Леонард Эйлер [8].

Координаты идейно

Смысл координат в том, чтобы закодировать некоторое множество объектов числами (такая цифровизация), но не абы как, а чтобы было удобно по этому множеству как бы перемещаться. В этом, кстати, этимологический смысл слова ордината — в ней должен быть заложен некий порядок.

Для нас привычно использовать координаты точек пространства, но можно же занумеровать что угодно: номер счета в банке, номер паспорта, сиденье в кинотеатре (причем тут уже две координаты: ряд и место), номер такта в музыкальном произведении (тоже могут быть дополнительные размерности, если это ноты оркестра) и т.д.

В случае геометрии, конечно, речь о неких абстрактных точках пространства. Можно представлять, что это пространство физического мира, который состоит из мельчайших частиц, причем они заполняют все так плотно, что пустоты между ними нигде не образуется.

Тогда удобно договориться о единице измерения, например, метр, и откладывать прямолинейные участки любой длины. Если мы работаем с шоссе между двумя городами, то достаточно одного числа, чтобы описать любую точку на этом шоссе. Тогда пространство называется одномерным.

Если же мы наблюдаем за футбольным матчем, то позиция игрока определяется уже двумя числами: относительно линий вдоль поля и поперек. Тут одним число никак не обойдешься, а двух хватает, поэтому это пространство считается двумерным.

Ну, а если нужно определить позицию мяча, который часто отрывается от поверхности земли на различную высоту, понадобится еще одна координата, собственно, для задания высоты полета. Тогда мы говорим о трехмерном пространстве.

Если описывается динамика точки, то появляется необходимость в четвертой координате — времени. Для описания пространства, в классическом понимании, больше размерностей не требуется. Хотя, конечно, благодаря развитию идеи Декарта, нас ничто не ограничивает изучать пространства любой размерности. Но мы в рамках этого курса об этом говорить не будем.

* * *

У координатного подхода большой плюс: точки закодированы числами, значит, можно решать любую задачу алгебраически с помощью уравнений. Но за это наступает расплата.

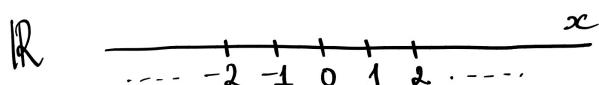
Закодировать точки можно поразному: во-первых координаты зависят от выбора начала отсчета, во-вторых, от выбора направлений осей и в-третьих, от единицы измерения. Это может приводить к дополнительным техническим трудностям пересчета из одной системы координат в другую (могут же два человека жить в разных городах и считать свой город центром), и вообще для решения задачи от выбора системы координат может зависеть многое: в одной системе будет простое решение, а в другой получится трактат на сто страниц.

Координаты символично

Здесь мы введем обозначения, которыми будем в дальнейшем пользоваться. Понятно, что все со школы умеют обозначать координаты точки, рисовать оси, подписывать буквы. Но на самом деле эти обозначения часто различаются то там, то сям. Поэтому лучше аккуратно проговорить один раз то, какими будут обозначения в этом курсе и почему они именно такие.

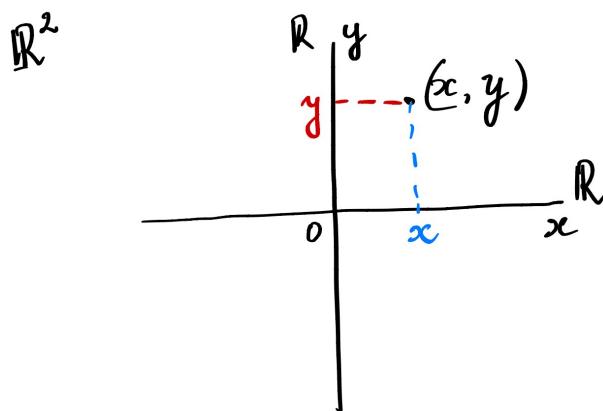
Про эволюцию понятия числа мы подробнее поговорим ближе к концу курса, когда замахнемся на кватернионы. Но множество всех чисел нам потребуется уже сейчас, оно в себя включает все целые, дробные и иррациональные числа, и обозначается буквой \mathbb{R} от слова Real. На русском эти числа называют вещественные (т.е. характеризует вещественное: масса, длина, угол, температура) или действительные (в пику комплексным) числа.

Множество \mathbb{R} часто представляется в виде прямой. Тут уже все и так закодировано. Координата x бегает по прямой. Начало отсчета — число 0, единица измерения — число 1, все готово.

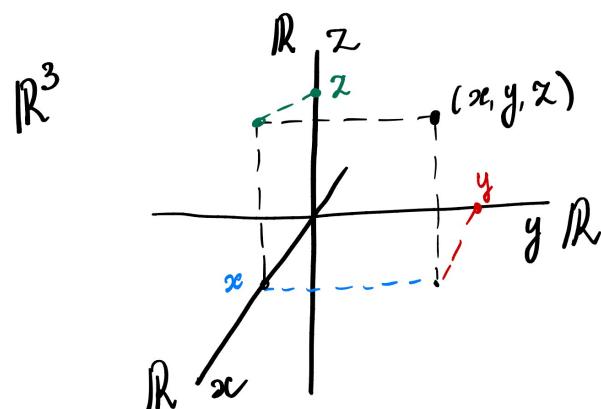


Для задания двумерного пространства нужно два числа. Множество всех пар чисел обозначается \mathbb{R}^2 , читается как "эр два". Координаты обозначаются как (x, y) .

Геометрически это представляется как две прямые, пересекающиеся в своих нулях под прямым углом.



Аналогично, для трехмерного пространства \mathbb{R}^3 добавляется еще одна прямая, перпендикулярная двум другим. Таким образом, координата точки имеет вид (x, y, z) .



FAQ

С чего это все-таки оси рисуются именно перпендикулярно друг другу?

Дело в том, что система координат и не обязана быть прямоугольной. Само по себе понятие угла на языке координат будет определено чуть позже, когда мы будем говорить о скалярном произведении векторов. А когда дойдем до матриц, будет ясно, как связаны различные системы координат между собой. Кстати, Декарт в "Геометрии" тоже не ограничивался прямоугольными координатами.

Но сейчас мы рассматриваем привычную нам модель описания пространства, где координата точки — это длина ее проекции на соответствующую координатную прямую (ось).

Зачем пишешь \mathbb{R} на концах?

Буквы \mathbb{R} рядом с осями я нарисовал, чтобы подчеркнуть, что мы как бы склонировали один и тот же объект (множество чисел) и геометрически их расположили в пространстве. Как правило, так не пишут, и я не буду в дальнейшем, ибо нет смысла. **Какая разница между черными x, y, z и цветными?** **Это одно и то же?**

Буквы x, y, z на "концах" просто указывают, какая переменная бегает по какой из прямых. Цветные x, y, z указывают конкретное значение этих переменных для данной точки. Часто, чтобы не путаться, конкретные значения обозначают с индексом, например, x_0, y_0, z_0 .

Почему координаты в круглых скобках? В школе/универе/видосах препод писал фигурные/квадратные.

Координаты самой точки указываются в круглых скобках через запятую. Круглые скобки — стандартное обозначение упорядоченного набора чисел. На языке программирования можно сказать, что круглые скобки это **Array**, а фигурные скобки — **Set**.

Где стрелочки?

Стрелки специально не указаны, потому что это задания направления, а про это, думается, лучше поговорить позже, когда обсудим векторы. Хочется, чтобы изложение было последовательным. Таким образом, обращаем внимание на то, что в принципе стрелки тут ни при чем. Важно то, как прямые расположены друг относительно друга.

Важно ли где какая координата? Почему x, y, z именно так нарисованы?

Да, есть понятие ориентации системы координат. По сути это значит, что направления осей имеют определенное расположение относительно друг друга.

В двумерном случае, традиционно, ось y смотрит вверх, а ось x — вправо. Если обеим осям поменять направление на противоположное, то ориентация не поменяется: ось y смотрит все равно влево относительно оси x . А вот если, например, поменять направление только оси x , то ориентация, т.е. взаимное расположение осей, изменится.

В трехмерном случае традиционной является такая ориентация, когда ось y смотрит влево относительно x (как и раньше), а ось z смотрит влево относительно y . Как бы слепили из двух плоскостей (x, y) и (y, z) одно трехмерное пространство.

Если говорить чуть более формально, то можем дать такое определение. Две системы координат называются одинаково ориентированными, если одну можно непрерывными деформациями перевести в другую. Вернемся к этому вопросу позже, когда дойдем до векторного произведения.

TODO: добавить картинки из Goodnotes.

Есть какой-то всемирный стандарт для этих записей?

Нет, в математике в принципе нет высеченных в камне стандартов. Например, в C++ можно открыть документацию к версии языка и точно узнать, что и как должно работать. В математике нет создателя/ей (человека по крайней мере), поэтому и документации такой нет. Есть только школы, традиции, которые часто схожи в разных сообществах, но далеко не всегда идентичны.

Так что здесь важно понимать суть. Мы в дальнейшем будем немного по-разному рисовать эти картинки, заодно станет понятно, что тут критично, а что нет.

Функции

Функция как формула

История развития понятия функции в общем-то очень переплетается с сюжетом про Декарта, Ньютона и Эйлера из раздела про системы координат.

Как уже говорилось, Декарт и Ферма начали описывать геометрические кривые алгебраическими уравнениями. Так как кривые рассматривались на плоскости, то уравнения имели две переменные x, y . Но что означает уравнение двух переменных?

Посмотрим на примере уравнения прямой:

$$ax + by + c = 0$$

Решения этого уравнения это все пары:

$$(x, -\frac{ax + c}{b})$$

Чтобы найти это решение мы *выразили* y через x :

$$y = -\frac{ax + c}{b}$$

Эйлер в "Анализе бесконечных" называл y функцией x , если y задан формулой, в которую входят только переменная x и константы, например,

$$y = 4x^5 - \sqrt{x} + \frac{x}{\sqrt{x+3}}.$$

Поэтому, по сути, в этом формульном смысле понятие функции витает и у Декарта, так как он прекрасно понимал, что уравнение задает зависимость между переменными. Но самого слова "функция" тогда еще не было. Ньютона, опять же, уже фактически использовал функциональную нотацию, указывая зависимость между величинами. Но как математический объект, определение функции дал Эйлер.

4. Функция переменного количества есть аналитическое выражение, составленное каким-либо образом из этого переменного количества и чисел или постоянных количеств¹).

Всякое аналитическое выражение, в котором, за исключением переменного количества z , все количества, составляющие это выражение, постоянны, будет функцией z ; так,

$$a + 3z, \quad az - 4z^2, \quad az + b\sqrt{a^2 - z^2}, \quad c^z$$

и т. д. будут функциями z .

Рис. 1: Определение функции из "Анализа бесконечных" Эйлера (1748) [8]

Функция как соответствие

Такое определение для обычного школьника было бы вполне привычным, но в дальнейшем представление о функции становилось более широким, чем просто формула.

Ключевая суть функции заключается в том, что она устанавливает взаимосвязь между переменными. Например, можно определить зависимость положения Солнца относительно времени на часах. Или величину зарплаты в зависимости от грейда. На уровне идеи формулы никакой нет. И хотя часто зависимость можно выразить формулой, есть риск, что все же от такого подхода мы что-то теряем.

И действительно, как сейчас известно, не любую функцию можно выразить формулой (как говорят, выразить "в радикалах"), да и не всегда это нужно.

Поэтому, где-то с работ Гаусса и Дирихле (18 век), начало формироваться представление о функции как о правиле, по которому одному или нескольким параметрам сопоставляется некий результат.

Как уже говорилось во введении, нетворк — это пример функции, которая по нетворк-параметрам преобразовывает трансформации анимационных объектов. И формулу для него на листочке записывать — идея страшная.

TODO: рисунок схематично иллюстрирующий нетворк.

§ 11.

Мы переходим теперь к исследованию вопроса, который интересен и сам по себе и кроме того играет важную роль в последующем изложении. Возьмем ряд натуральных чисел

$$1, 2, 3, 4, \dots, m$$

до какого-нибудь числа m и определим, сколько в этом ряду чисел, взаимно простых с числом m . Число таких чисел обозначается обычно через $\varphi(m)$, где буква φ имеет значение функционального символа¹⁾. Так как единица — число, взаимно простое с единицею, то

$$\varphi(1) = 1;$$

далее мы определяем непосредственно:

$$\varphi(2) = 1, \quad \varphi(3) = 2, \quad \varphi(4) = 2, \quad \varphi(5) = 4, \quad \dots$$

Рис. 2: Определение функции Эйлера в лекциях Дирихле [9]

Традиционно, функцию обозначают буквой f , а ее аргумент — буквой x . Значение функции f при заданном $x \in \mathbb{R}$ обозначается символом $f(x)$. Чтобы сразу указать, какие обозначения для конкретной функции будут использоваться, часто говорят "задана функция $f(x)$ ".

Есть мнение, что таких фраз нужно избегать, потому что это неграмотно: если написал $f(x)$, то это уже не функция, а значение функции в точке x . Тогда пишут либо просто f , либо $f(\cdot)$, если хотят подчеркнуть, что аргумент один.

Когда говорят о функции, важно понимать, какая у нее область определения, т.е. какие именно аргументы можно подставлять в функцию, чтобы получать корректное значение. Поэтому говорят "функция f , определенная на отрезке $[0, 1]$ " и т.п.

В программировании есть assert-ы. Это проверки, обычно в начале функции, что некоторое условие на входящие параметры выполнено. Если условие не выполнено, кидается ошибка. По сути, это и есть задание области определения функции.

TODO: добавить примеры?.

График функции

Функция — это понятие, делающее еще один шаг в сторону от геометрии к алгебре (привет Декарту) и далее к анализу (привет Эйлеру). Но интуиция и наглядность, присущая геометрии — никуда не денешься — вещь важная. Формула — хорошо, а формула с рисунком — еще лучше! Особенно, если ты физик, как Ньютон.

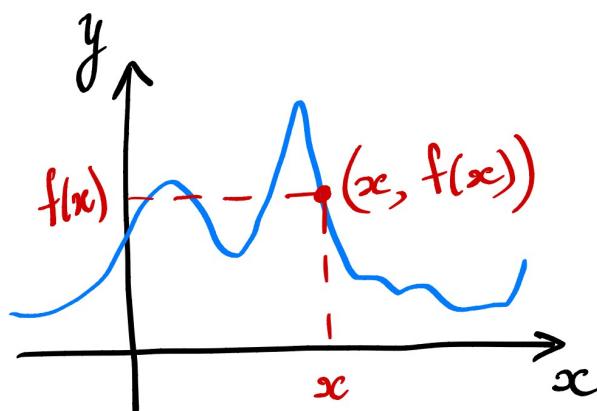
TODO: вставить картинку из Начал.

График функции является, наверное, главным виновником популярности именно прямоугольных осей координат. Для чего нужен график? Чтобы иллюстрировать то, как изменяется одна величина при изменении другой/других. И если функция числовая, т.е. принимает числовые значения (что вообще-то не обязательно), то логично изображать ее значения как бы столбиками, стоящими на аргументах функции. Все столбики, ясное дело, параллельны и стоят они все на одном уровне. Вот и получается сама собой прямоугольная система координат и привычный рисунок.

Опр. Графиком функции f , определенной на множестве X , называется множество пар $\{(x, f(x))\}_{x \in X}$. Простыми словами, взяли все значения переменной $x \in X$, посчитали от них значения $f(x)$ и записали все результаты в табличку:

x	x_1	x_2	x_3	x_4	\dots
$f(x)$	y_1	y_2	y_3	y_4	\dots

Для наглядного изображения графика функции удобнее всего нарисовать картинку, где все эти точки образуют гладкую кривую линию:



Так что, строго говоря, как и система координат, график не является "рисунком", это не геометрический объект, а множество пар аргумент+значение. Но это множество традиционно иллюстрируют в виде рисунка в прямоугольной системе координат.

N.B. Конечно, график далеко не обязан быть гладким и даже непрерывным. Отсюда возникают строгие понятия, характеризующие функцию, которые образуют теорию под названием анализ функций (по-народному "мата").

Кривые и поверхности как графики функций

Может возникнуть мысль, что график функции нужен просто ради наглядной картинки. Это отчасти так, но здесь опять проявляется связь алгебраического и геометрического. С одной стороны, по функции, негеометрическому объекту, можно построить график, который не только дает интуитивное графическое представление о ней, но и является уже описанием геометрического объекта: кривой, поверхности и т.п.

* * *

Например, мы знаем, что окружность радиуса 1 задается уравнением

$$x^2 + y^2 = 1.$$

Выразим y через x :

$$y^2 = 1 - x^2,$$

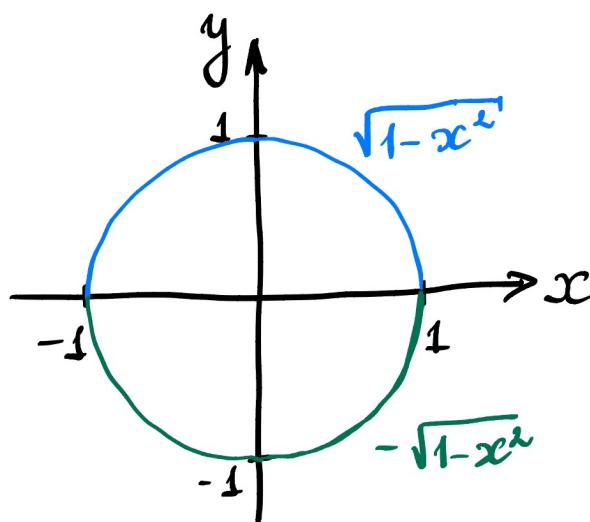
значит,

$$y = \sqrt{1 - x^2} \quad \text{или} \quad y = -\sqrt{1 - x^2}.$$

Здесь мы получили две функции $f_1(x) = \sqrt{1 - x^2}$ и $f_2(x) = -\sqrt{1 - x^2}$. Построим их графики:

Мы видим, что два этих графика в объединении дают геометрический объект — окружность. И так можно делать с любыми кривыми на плоскости, если они не слишком экзотические.

* * *



Аналогично, можно рассмотреть уравнение единичной сферы

$$x^2 + y^2 + z^2 = 1.$$

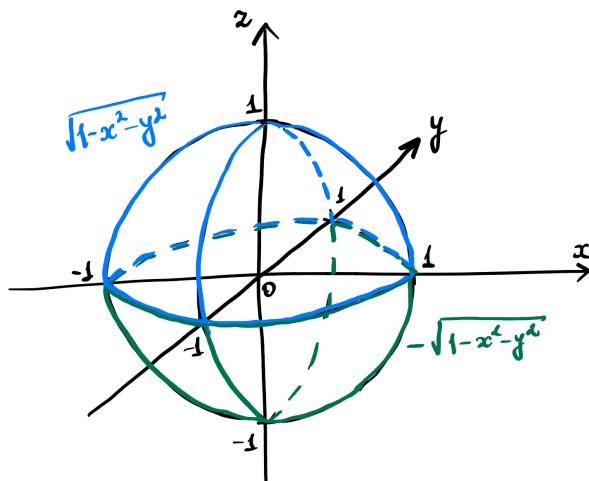
Выразим z^2 :

$$z^2 = 1 - x^2 - y^2$$

откуда

$$z = \sqrt{1 - x^2 - y^2} \quad \text{или} \quad z = -\sqrt{1 - x^2 - y^2}.$$

Графиками функций $f_1(x, y) = \sqrt{1 - x^2 - y^2}$ и $f_2(x, y) = -\sqrt{1 - x^2 - y^2}$ будут две полусфера.

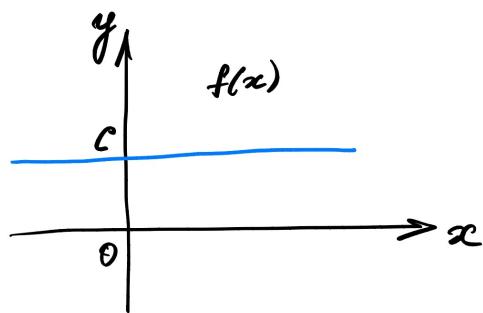


Примеры функций и их графиков

1 Константа

Функция, которая всюду принимает одно значение, называется константой.

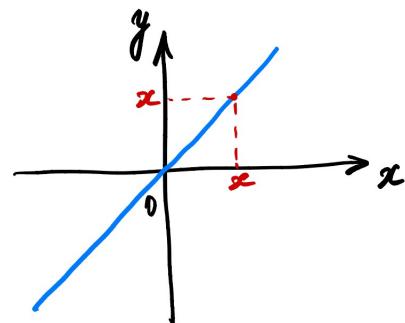
$$f(x) = C, \quad \text{где } C \in \mathbb{R} \text{ — число.}$$



2 Тождественная функция

Функция, которая возвращает то же значение, которое получила на вход, называется тождественной.

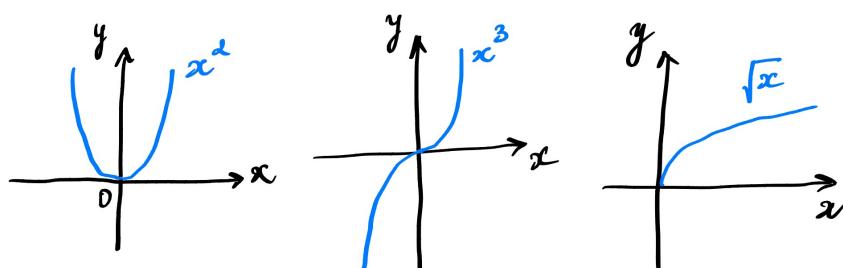
$$f(x) = x$$



3 Степенная функция

Степенная функция определяется формулой

$$f(x) = x^a, \quad a \in \mathbb{R}.$$



При разных значениях параметра a будут получаться различные функции:

$$a = 0 \Rightarrow f(x) = 1$$

$$a = 1 \Rightarrow f(x) = x$$

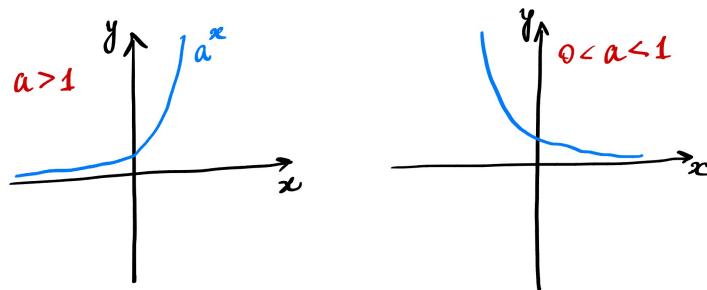
$$a = 2 \Rightarrow f(x) = x^2$$

$$a = \frac{1}{2} \Rightarrow f(x) = \sqrt{x}$$

4 Показательная функция

Показательная функция определяется формулой

$$f(x) = a^x, \quad a \geqslant 0.$$



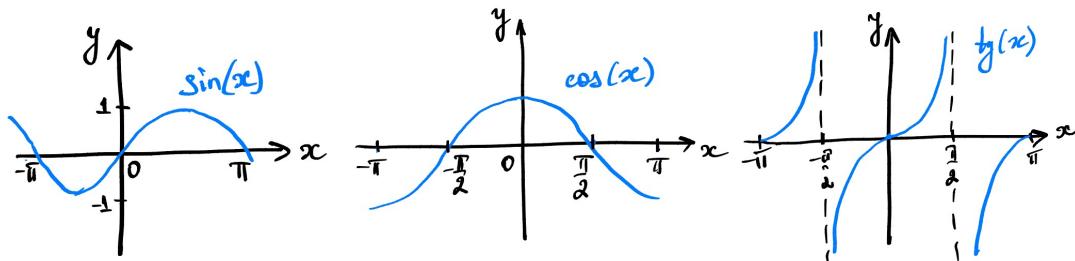
N.B. Частный случай $a = e$ дает функцию

$$f(x) = e^x,$$

которая называется экспонентой.

5 Тригонометрические функции

О тригонометрических функциях подробнее мы поговорим ниже, а пока просто посмотрим на их графики



Тригонометрия

Мотивация

В прошлом разделе мы поговорили о базовых функциях, которые еще называют элементарными в том смысле, что они как кирпичики – являются строительным материалом для более сложных функций. Но вот вопрос: все ли функции можно построить, имея только эти кирпичики?

Как было сказано, в современном понимании, термин "функция" означает соответствие между двумя множествами объектов. И в этом смысле, конечно, сложно говорить даже о том, что из себя представляет множество всех функций. Но мы не будем тягать непосильные гири, а продолжим говорить именно о числовых функциях, которые принимают на вход число и выдают тоже число. Что тогда?

Согласно Эйлеру, функция переменных x, y, z это аналитическое выражение, включающее в себя эти переменные и построенное путем арифметических операций, а также операции извлечения корня. И тогда, если принять позицию Эйлера, больше числовых функций нет.

Революцию в подходе к функциям после Эйлера совершил французский математик Жан Батист Фурье (1768–1830), который сделал несколько вещей. Во-первых,

он отметил, что методом Эйлера можно получить только непрерывные функции, а во-вторых, Фурье показал, что почти любую вразумительно выглядящую функцию можно представить в виде суммы синусов и косинусов. Такое представление называется тригонометрическим рядом. Этот поразительный вывод перевернул представление об устройстве функций и повлиял на все дальнейшее развитие математики и особенно сильно на ее приложения.

К Фурье мы еще вернемся в конце этого параграфа, а пока все же вернемся к тригонометрии.

Астрономия и хорды

1 Начало тригонометрии

Слово "тригонометрия" происходит от слов triangle и measure, т.е. получается, что тригонометрия – наука об измерениях треугольника. Однако, хотя этимология слов часто помогает понять суть предмета и его истоки, в данном случае, на мой взгляд, происходит противоположное действие. Дело в том, что само это название появилось лишь в конце 16-го века, тогда как сама наука уходит корнями в древние цивилизации!

Впрочем, в Египте действительно использовались расчеты сторон треугольников для сооружения пирамид. Но по-настоящему началом истории тригонометрии считается Древняя Греция, а "отцом" тригонометрии – Гиппарх (190-120 до н.э.), чье дело развили Птолемей (100-170 н.э.). Греки активнейшим образом изучали движение звезд, планет, Луны и Солнца на небесной сфере. Это было очень важным делом, ведь астрономические знания позволяли предсказывать многие ключевые события в религиозной жизни общества, такие как:

1. затмения, полнолуния, положение звезд и т.д.
2. календари для посевов
3. составление карт, навигация для передвижения и торговли

Естественно, что астрономия сводилась во многом к тому, чтобы вычислять расстояния между небесными светилами. А что это значит на нашем языке? Расстояние между точками на сфере это хорды. Так что в то время была скорее не тригонометрия, а наука об измерении длин хорд на сфере.

TODO: рисунок из википедии History of Trigonometry.

Чем связаны будущая тригонометрия и эти хорды? Тем, что для измерения длин в окружности автоматически возникает потребность в рассчете той величины, которая позже будет называться синусом.

Вот, посмотрим на картинку с окружностью, в которой из центра выпущены два радиуса под углом φ друг к другу. Какова длина получившейся хорды? Тот, кто часто пользуется тригонометрическими функциями быстро скажет, что

$$\text{chord } \varphi = 2r \sin \frac{\varphi}{2}$$

Позже, астрономия перетекла к индийцам, а затем к арабам, которые и придумали понятие синуса, как половина длины хорды, а потом и все остальные тригонометрические функции.

2 Тригонометрические функции на окружности

Мы увидели, что в измерении хорд уже вшито измерение синуса угла. Поэтому развитие науки об измерениях хорд неотделимо от истории тригонометрии. Но после расцвета древнегреческой математики ведущую роль в развитии математики захватили индийцы и арабы. Так, у индийцев появилось понятие ардха-джива (от ардха – "половина" и джива – "тетива"), которое означало половину хорды. Затем слово "ардха" отпало и осталось просто "джива", которое позже перетекло в европейские языки, как "синус" [5] (с. 199).

Так что под синусом изначально подразумевалась длина половины хорды. А когда определилось это понятие, то дальше пошло проще. Мы это увидим, если нарисуем синус на координатной плоскости следующим образом:

TODO картинка

Здесь синус это отрезок, полученный опусканием перпендикуляра на радиус, проведенный из центра в середину исходной дуги, то есть в наших терминах, на ось координат x . Но есть как бы родственная дуга, которая получается проведением перпендикуляра на ось координат y . Полученная отсюда полуходра получила логичное название косинус (**complement** к синусу).

Теперь проведем к точке начала дуги касательную и отметим ее пересечения с осями координат. Тогда два кусочка получившегося отрезка будут тангенсом и котангенсом заданной дуги (или угла, как кому нравится):

TODO: картинка.

Но и это не все. Отрезки от центра окружности до точек пересечения касательной с осями координат называются секансом и косекансом:

TODO: картинка.

Таким образом мы через измерений единичной окружности определили шесть основных тригонометрических функций: синус, косинус, тангенс, котангенс, секанс, косеканс.

В этом видео можно посмотреть все эти построения в более красивом анимированном исполнении, например здесь можно прямо регулировать угол и смотреть за изменением значений тригонометрических функций. Можете и сами сделать подобную анимацию.

Нам пришлось прибегнуть к словам "система координат" и "оси координат", хотя мы знаем, что они как концепция возникли у Декарта и Ферма лишь в 17-ом веке. Это было сделано для более сжатого изложения: нам важно понимать, откуда появился синус как отрезок в окружности и то, что из него уже выросло все остальное.

Отношения в треугольнике

Вы, желающие изучать великие и замечательные вещи, кто интересуется движением звезд, должны прочитать эти теоремы о треугольниках

— Regiomontanus "De triangulis omnimodis"

1 Понимание связи с треугольником

Термин "тригонометрия" возник в Европе. Назовем ключевые имена и события, которые сыграли определяющую роль в этом направлении.

Сначала отметим еще раз, что измерения синусов и прочих определенных выше длин, которые сейчас мы называем тригонометрическими функциями, и в Европе

продолжают жить в контексте прежде всего астрономии. По большому счету вся современная математика, можно сказать, вдохновлена этой наукой.

Первой ключевой фигурой выделяется европейский математик немецкого происхождения Региомонтан, который в своем труде *"De triangulis omnimodis"* ("О треугольниках всех видов", 1464) [4] собрал воедино и структурировал все сведения о достижениях индийских и арабских ученых, которые помогали решать геометрические задачи в астрономии. И, что крайне важно, уже по названию видно, что Региомонтан поставил в этой дисциплине в центр именно треугольник.

Второй ключевой фигурой мы назовем австрийского математика Ретика, ученика Коперника. Ретик занимался составлением таблиц значений указанных выше шести тригонометрических функций с точностью до 10 знаков после запятой. Результаты были опубликованы в трактате *"Canon of the Science of Triangles"* в 1551 году, а позже в 1596 году ученик Ретика Валентин Отто опубликовал завершенную версию, которая заняла около 1500 страниц. Причем точность таблиц была такой, что ее хватало для астрономических вычислений до начала 20-го века.

И наконец третьей ключевой фигурой является еще один немецкий астроном Питиск, который собственно и ввел впервые термин "тригонометрия", назвав свою работу по плоской и сферической тригонометрии *"Trigonometria: sive de solutione triangolorum tractatus brevis et percipicuus"* (Тригонометрия, или краткий и ясный трактат о решении треугольников), 1595.

Здесь же вспомним героя рассказа раздела об элементарных функциях Джона Непера, который первым придумал логарифмы. Там не было сказано, но на самом деле он вычислял логарифмы не любых чисел, а именно тригонометрических функций. То есть составлял таблицы функций типа $\log(\sin x)$. Нам-то понятно, что если умеешь вычислять логарифмы любых чисел, то умеешь считать то же и для значений синуса, которые являются просто отрезком $[-1, 1]$, но Непер не сразу это понял. Говоря о логарифмах, мы отмечали, что Непер преследовал цель ускорить процесс перемножения чисел, переводя умножение в сложение. Так вот на самом деле, ему нужно было ускорять вычисления именно для астрономических вычислений, поэтому тригонометрические функции его и волновали в первую очередь.

2 Тригонометрия 8-го класса

TODO

3 Поиск проекции

TODO

Эйлер и тригонометрия

1 Определение тригонометрических функций

С развитием понятия TODO

2 Графики тригонометрических функций

TODO

3 Обратные тригонометрические функции

TODO

Связь всего со всем

1 Эйлер и экспонента

TODO

2 Гиперболические тригонометрические функции

TODO

3 Пружина с дампингом

TODO

4 Фурье

TODO

Мотивация

Как говорилось выше, геометрия изучает метрические свойства фигур в плоскости или пространстве. Под фигурами можно понимать то, что можно изобразить, нарисовать. Например, прямые, кривые, окружности, эллипсы, сферы, параллелепипеды, сечения одних фигур другими и т.д.

Естественным историческим ходом развития мысли в геометрии выделялись более узкие направления. Так, известно, что в Египте строились пирамиды (3000 лет до н.э.), где требовались расчеты длин и углов в треугольнике для того чтобы понять сколько материалов нужно для выстраивания пирамиды определенных размеров.

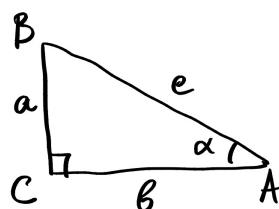
Затем греческие астрономы (190-120 лет до н.э.) стали замечать, что Солнце каждый день встает немного смещаясь в восток, и скоро поняли, что Солнце движется относительно Земли по окружности, а полный круг делает за год. Отсюда пошло разделение полного оборота вокруг оси на 360 градусов. Встал вопрос, а какого радиуса эта окружность, т.е. на каком расстоянии Солнце находится от Земли. Выяснилось, что эта задача решается через построения прямоугольных треугольников и поиска в них длин сторон и углов.

Таким образом, стала выкристаллизовываться тригонометрия – геометрия треугольников. Впрочем, такой взгляд скорее вреден, поскольку сейчас известно о приложениях тригонометрии в совершенно далеких друг от друга областях, от астрономии, механики, электричества до методов сжатия изображений.

Об истории тригонометрии и о ее красивых приложениях есть 40-минутная лекция канадского историка математики Glen Van Brummelen The Story of Trigonometry: Revolutions in the Heavens, and on the Earth. Для демонстрации он там использует сайт MyPhysicsLab.

Отношения в прямоугольном треугольнике

Изобразим прямоугольный треугольник $\triangle ABC$ с прямым углом $\angle C$ и сторонами a, b, c . Обозначим угол $\angle A = \alpha$.



Тогда по определению синус (sine), косинус (cosine), тангенс (tangent) и котангенс (cotangent) угла α есть отношения сторон:

$$\begin{aligned}\sin \alpha &= \frac{a}{c} \\ \cos \alpha &= \frac{b}{c} \\ \operatorname{tg} \alpha &= \frac{a}{b} \\ \operatorname{ctg} \alpha &= \frac{b}{a} = \frac{1}{\operatorname{tg} \alpha}\end{aligned}$$

Отметим свойства, следующие из этого определения

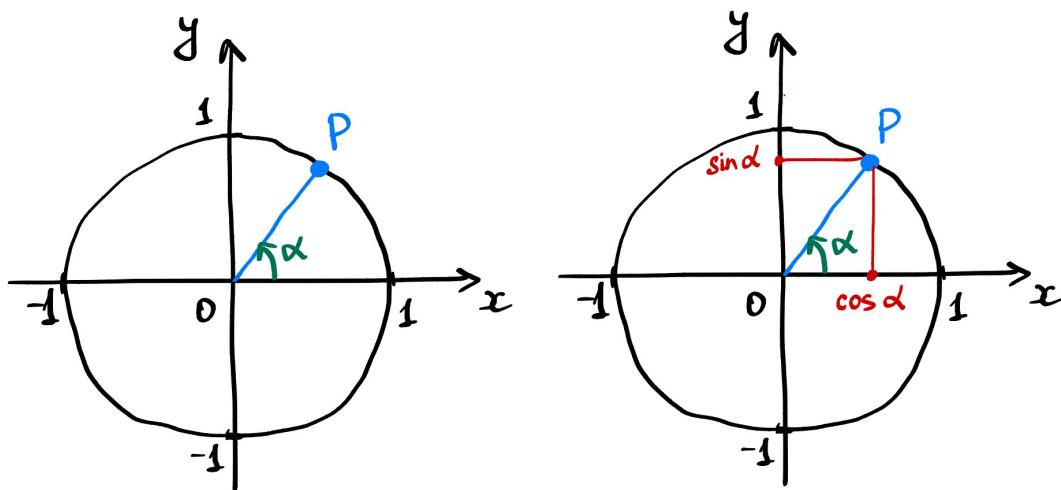
- $0 \leq \sin \alpha \leq 1, \quad 0 \leq \cos \alpha \leq 1$
- $\sin^2 \alpha + \cos^2 \alpha = \frac{a^2}{c^2} + \frac{b^2}{c^2} = \frac{a^2+b^2}{c^2} = 1$ (теорема Пифагора)
- $b = c \cdot \cos \alpha$ – проекция AB на прямую AC

Точка на окружности

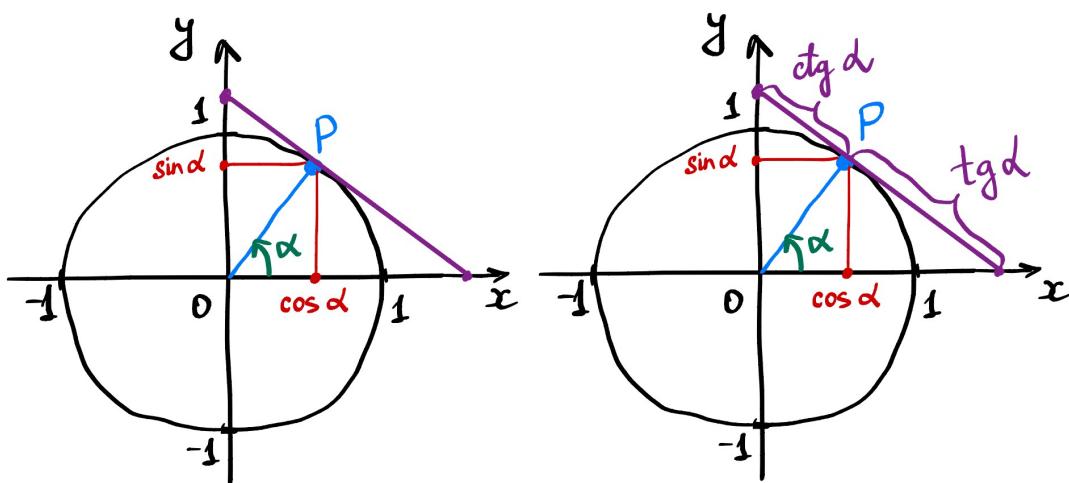
Теперь совершим эволюционный переход от египтян и вавилонян к грекам, и определим тригонометрические функции, связав их с геометрией окружности единичного радиуса.

Итак, рассмотрим окружность радиуса 1 и поместим ее в начало прямоугольной системы координат. Затем возьмем произвольную точку P на этой окружности, соединим ее с центром окружности и обозначим получившийся угол за α .

Тогда по определению синусом угла α называется координата y точки P , а косинусом – координата x .



Но это еще не все. Проведем касательную к точке P и продлим ее до пересечения с осями системы координат. Тогда тангенсом называется длина отрезка от точки P до пересечения с осью абсцисс, а котангенсом называется длина отрезка от точки P до пересечения с осью ординат.



Вот так с одной стороны определение получается более сложным: используется окружность, координаты, проводится касательная. А с другой стороны, такой взгляд будто открывает глубину понятий. Слово синус в греческом языке означает "хорда". И действительно, мы видим по рисунку, что синус это половина длины хорды, а косинус (ко-синус!) половина длины смежной хорды.

Тангенс означает "касание". И мы видим, что тангенс это просто длина части касательной. А котангенс (ко-тангенс!) – другая часть этой же касательной.

Рекомендую посмотреть об этом видео Beautiful Trigonometry от Numberphile.

* * *

Теперь отметим, что при определении тригонометрических функций через окружность, они могут принимать и отрицательные значения. Так,

$$\begin{aligned} -1 \leq \sin \alpha \leq 1 \\ -1 \leq \cos \alpha \leq 1 \end{aligned}$$

Сразу же видны и многие другие важные свойства, которые в школе часто пытаются зазубрить:

- периодичность: $\sin(\alpha) = \sin(\alpha + 2\pi)$
- нечетность синуса: $\sin(-\alpha) = -\sin \alpha$
- четность косинуса: $\cos(-\alpha) = \cos \alpha$
- $\sin(\alpha + \pi) = -\sin \alpha$
- $\sin(\alpha + \frac{\pi}{2}) = -\cos \alpha$

и так далее. Учить тяжело и не нужно, но понимать легко и полезно.

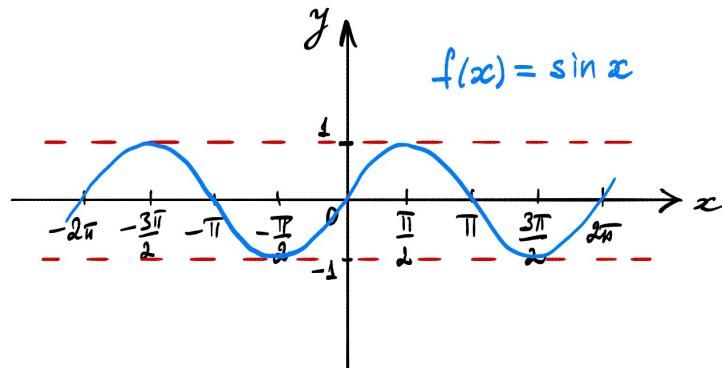
Графики

Итак, определения синуса, косинуса, тангенса и котангенса через единичную окружность задают числовые функции

$$\sin(x), \quad \cos(x), \quad \operatorname{tg}(x), \quad \operatorname{ctg}(x),$$

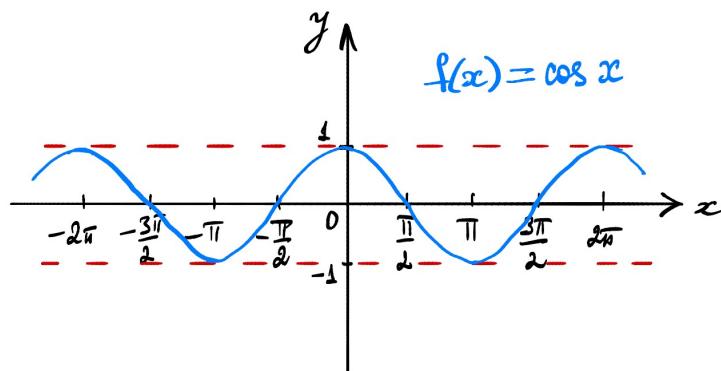
причем если $\sin(x)$ и $\cos(x)$ определены на всей числовой прямой, то $\operatorname{tg}(x)$ и $\operatorname{ctg}(x)$ имеют неопределенности.

Нарисуем график синуса.

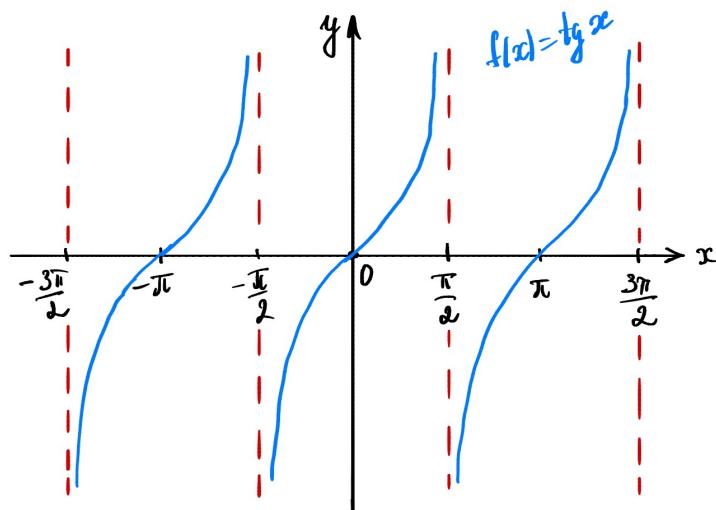


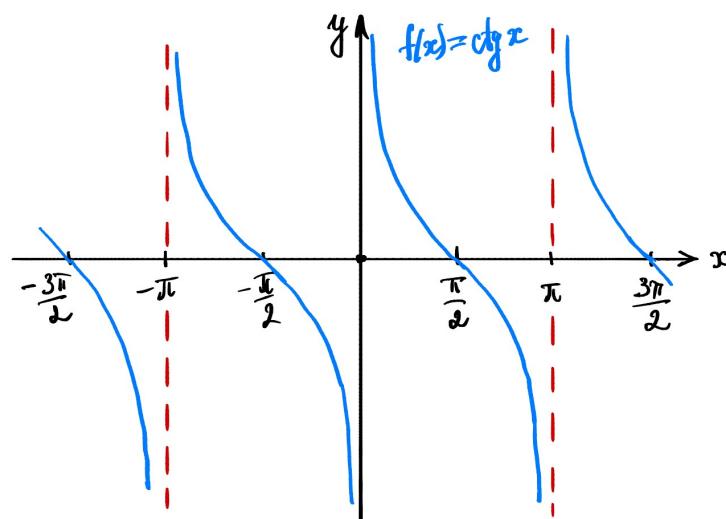
Полученная кривая напоминает волну. Она постоянно колеблется между максимальным и минимальным значениями, причем, подходя к ним, плавно замедляется, а отдаляясь, также плавно ускоряется.

График косинуса ничем не отличается от синуса, кроме сдвига на $\pi/2$ по горизонтали. Это видно из построения косинуса на окружности.



Тангенс и котангенс имеют также схожие между собой графики, однако отличные от графиков синуса и косинуса.



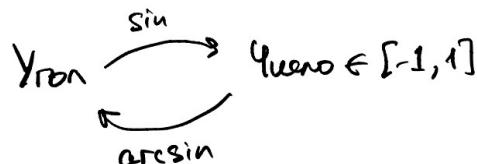


Обратные тригонометрические функции

Чтобы получить по значению тригонометрической функции ее аргумент, т.е. угол, нужно знать соответствующую обратную тригонометрическую функцию. Например, мы знаем все стороны в прямоугольном треугольнике, а значит легко можем найти любые соотношения между ними, в том числе значения любых тригонометрических функций углов треугольника. Обратные тригонометрические функции дают возможность узнать и сами углы.

Такие функции называются также как и прямые тригонометрические функции, только с приставкой **arc** (дуга):

$$\arcsin(x), \quad \arccos(x), \quad \operatorname{arctg}(x), \quad \operatorname{arcctg}(x).$$



Тут есть, правда, одна проблема. Если снова взглянуть на график, например, синуса, то мы увидим, что каждое его значение повторяется бесконечное число раз. Это понятно, ведь функция периодическая. Отсюда возникает неоднозначность обращения функции. Допустим, мы хотим понять, чему равен $\arcsin(0)$. Тогда мы получаем значения $0, \pm\pi, \pm 2\pi, \dots$ и что же делать? Здесь, как правило, с неоднозначностью борются очень просто: волевым усилием решается, что берется то значение, которое лежит в промежутке $[-\pi/2, \pi/2]$.

Тогда график арксинуса будет выглядеть таким образом:

TODO: график арксинуса.

График арккосинуса

TODO: график

График арктангенса

TODO: график

График арккотангенса

TODO: график

Зачем это все?

Музыка, Фурье, формула Эйлера, механика и т.д.

1 Формула Тейлора

Итак, мы теперь знаем графики тригонометрических функций и их значения при любых аргументах. Ведь так? Конечно, нет! Ведь графики мы нарисовали очень приблизительно, еще и воспользовавшись знаниями, которые не обсуждались в этом изложении. Например, мы считаем, что известны табличные значения рассмотренных тригонометрических функций типа $\sin(\pi/6) = 1/2$ и т.д.

Но даже знание нескольких табличных значений не дает построить точно график функции на всей области определения. Действительно, вот например, чему будет равно $\sin(1)$? Очевидно, что вычислять такое нужно, ведь угол может быть любым, но как это делать?

В древности для этих целей стали искать способ найти хотя бы приближенное значение. Из идеи о том, что синус это длина хорды единичной окружности, которую образует некоторый угол, геометрическими соображениями начали приближать это значение. И получилось следующее

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots,$$

где $n! = 1 \cdot 2 \cdots n$ – факториал числа n .

Для косинуса получилась такая формула:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$$

Сейчас эти формулы называются рядами Тейлора в честь английского математика Брука Тейлора (1685–1731), который жил уже во времена Ньютона и Лейбница, создателей дифференциального исчисления. Тейлор показал, что любую функцию, у которой есть производная, можно разложить в подобный степенной ряд. А именно, если есть функция f , то ее можно записать в виде

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \dots$$

О рядах Тейлора есть замечательное видео на канале 3Blue1Brown – Taylor series | Chapter 11, Essence of calculus.

2 Формула Эйлера

Современники Тейлора не слишком оценили этот результат, но немного позже Эйлер, Лагранж и другие математики поняли насколько мощный это инструмент, ведь теперь можно, не задумываясь о геометрической составляющей, выписывать ряды самых непонятных функций.

Так, теперь очень легко выписать разложение в ряд Тейлора для экспоненты e^x :

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \dots$$

Хочется спросить: ничего не напоминает? Действительно, разложения для синуса и косинуса очень похожи на разложение экспоненты. Кажется, будто нужно просто сложить синус и косинус, и получится экспонента. Посмотрим:

$$\sin(x) + \cos(x) = 1 + x - \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} - \frac{x^6}{6!} - \frac{x^7}{7!} + \dots$$

Эх, почти! Вот и Эйлер также подумал. Но не остановился на этом.

Что происходит в полученной сумме синуса и косинуса? Слагаемые те же, что в экспоненте, но знаки другие, они чередуются через два: два плюса, два минуса, потом все повторяется. Эйлер заметил, что похожее чередование происходит, если возводить в степень мнимую единицу:

i^0	i^1	i^2	i^3	i^4	i^5	i^6	\dots
1	i	-1	$-i$	1	i	-1	\dots

Снова два плюса, два минуса, два плюса, два минуса, \dots

Тогда Эйлер подставил в формулу для экспоненты аргумент ix вместо x и получил:

$$e^{ix} = 1 + ix - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - i\frac{x^7}{7!} + \dots$$

и теперь становится очевидно, что

$$e^{ix} = \cos x + i \sin x.$$

А если подставить здесь $x = \pi$, то получим знаменитую формулу Эйлера:

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

одну из самых красивых формул в математике, соединяющую в себе самые важные константы: 0, 1, π , e .

3 Движение груза на пружине с дампингом

Уравнение движения груза на пружине с дампингом:

$$mx'' + bx' + kx = 0,$$

здесь $x(t)$ – смещение груза, m – масса груза, b – дампинг, k – жесткость пружины (stiffness).

Положим для простоты записи $m = 1$. Тогда уравнение примет вид

$$x'' + bx' + kx = 0.$$

Это линейное однородное дифференциальное уравнение второго порядка. Для его решения найдем корни характеристического многочлена

$$\lambda^2 + b\lambda + k = 0.$$

Это обычное квадратное уравнение, его корни легко находятся через дискриминант:

$$\lambda_{1,2} = \frac{-b \pm \sqrt{b^2 - 4k}}{2}.$$

Тогда, согласно правилу решения уравнений такого рода, все решения описываются так:

$$x(t) = C_1 e^{-\lambda_1 t} + C_2 e^{-\lambda_2 t},$$

в случае если $\lambda_{1,2}$ – действительные, и

$$x(t) = e^{-\alpha t} (C_1 \cos(\beta t) + C_2 \sin(\beta t)),$$

если $\lambda_{1,2} = \alpha \pm i\beta$ имеют мнимую часть.

Характеристические числа $\lambda_{1,2}$ будут мнимыми в случае, если

$$b^2 < 4k,$$

тогда движение будет осциллирующим, а дампинг будет лишь оказывать затухающее действие, постепенно уменьшая амплитуду колебаний. Например, при $k = 3$ дампинг не должен превышать значения $2\sqrt{3} \approx 3.5$. А если превысит, то закон движения перестанет носить осциллирующий характер и будет вести себя экспоненциально!

При этом если дампинг равен нулю, т.е. $b = 0$, то движение будет максимально простым:

$$x(t) = C_1 \cos(\sqrt{k}t) + C_2 \sin(\sqrt{k}t).$$

и если начальная скорость равна нулю, т.е. $x'(0) = 0$, то находим, что $C_2 = 0$. А если начальное смещение равно L , то $C_1 = L$, т.е.

$$x(t) = L \cos(\sqrt{k}t).$$

А это просто колебательное движение с амплитудой L и частотой $\frac{\sqrt{k}}{2\pi}$.

Визуализацию этой модели можно посмотреть на сайте MyPhysicsLab.

4 Фурье

Криволинейные системы координат

*Решил ты вырваться за предел
Осточертевших квадратных форм.
— В.А. Лифшиц "Квадраты"*

Пока что мы продолжаем жить в геометрическом мире, поэтому слово криволинейное можно понимать вполне конкретно — что-то не прямолинейное. А прямолинейно — ну ясно же, что такое! Вон, палка прямая валяется.

Система координат, в сущности, это способ закодировать объекты удобным образом. Понятно, что делать это можно по-разному. Как известно, с древности навигация осуществлялась по звездам, города кодировались двумя числами: долготой и широтой. Поэтому сферическая геометрия была развита очень хорошо.

Полярная система координат

Сферическая система координат

Цилиндрическая система координат

Векторы

История

Смело предположу, что большинству из нас понятие вектора кажется довольно простым. С другой стороны, как мы увидели выше, тригонометрия и анализ функций – очень даже крепкие орешки, на понимание которых даже поверхностно нужно потратить некоторое количество усилий, а глубина этих наук очевидна так, словно стоишь на вершине каньона и смотришь вниз.

А что же векторы? Из школы многие помнят, что вектор имеет начало и конец. Чтобы получить вектор \overrightarrow{AB} нужно как бы из конца B вычесть начало A . Складывать и умножать на число тоже понятно как. Скалярное и векторное произведения? Не так очевидно, но формулы есть: одна с косинусом, другая с синусом. И вроде бы и все ладно.

И, скорее всего, вам этих знаний хватит для 99% задач, связанных с векторами. Но есть неявные, но весомые причины разобраться в этой теме чуть глубже, которые можно увидеть, проанализировав историю становления векторов в математике. Она нас недурно удивит.

Первое удивление состоит в том, что понятие "вектор" в математике появилось лишь в 1843 году. Меньше двухсот лет назад! К этому времени Декарт и Ферма открыли аналитическую геометрию, указав метод решения геометрических задач на языке уравнений, Ньютон и Лейбниц построили дифференциальное и интегральное исчисление, Бернулли доказал первый вариант Закона Больших Чисел в теории вероятностей, Эйлер нашел выражение комплексной экспоненты через синус и косинус

$$e^{ix} = \cos x + i \sin x,$$

и многое-многое другое произошло к середине 19-го века.

Второе удивление еще менее правдоподобное. Ввел понятия "вектор", а также "скалярное" и "векторное" произведения ирландский математик и физик Уильям Роэн Хэмилтон, когда – внимание! – открыл кватернионы.

Вообще, исторически было бы более правильно сначала рассказывать о кватернионах, а только потом о векторах. Но мы, учитывая в большей степени техническую сложность тем, все же пойдем своим путем и о кватернионах подробно поговорим в соответствующей части курса.

Дело в том, что в 19-м веке было сразу множество, порядка 5-10, попыток создать нечто вроде современного векторного метода, то есть идея витала в воздухе и как бы даже перезрела. Еще Лейбниц в письмах к Гюйгенсу писал, что математике не хватает инструментов, более полно описывающих положение динамических объектов: позицию описать можем, закон движения можем, а скорость – нет. Физики все больше нуждались в математическом языке, который бы описывал величины, имеющие не только численное значение, но и направление. Но долгое время ничего не получалось.

Мы не будем подробно окунаться в эту длинную историю с кучей имен и связанных с ними драм, это всегда можно сделать, открыв замечательную книгу *Crowe – A History of Vector Analysis* [3]. Но сам факт, что векторы приобрели первое очертание лишь в 1844, после чего потребовалось еще полвека, чтобы осознать это и выстроить систему, поражает. Так и хочется понять: а в чем такая трудность-то была?

Ответить на этот вопрос можно одним словом – "абстракция". Абстракция это когда вместо конкретного объекта видишь в нем как бы его суть. Самая первая абстракция это число: 5 яблок, 5 дней или 5 пальцев – не важно, это все – пять. Но понять это – совсем не просто.

А вектор это абстракция уже над понятием числа. Абстракция в том смысле, что над вектором можно производить те же действия, что и над числами: складывать, умножать, вычитать, делить. Среди векторов есть ноль и один: нулевой и единичный векторы. Единственное, чем вектор отличается от числа, это наличием направления. Вот и вся суть векторов.

Осознание этого и в наше время вызовет затруднение, а в начале 19-го века не было никаких аналогов чисел. Было очевидно, что складывать, умножать, вычитать и делить можно только числа. А все эти разговоры про складывание и умножение неких векторов воспринималось как заумная философия. Позже из этого осознания выросла вся современная высшая алгебра, которая как раз занимается изучением абстрактных структур, на которых заданы операции сложения и умножения. Если интересно, почитайте статью о теории групп.

Итак, из всей приведенной истории можно сделать выводы

1. В случае, если у вас возникает трудность при понимании каких-либо утверждений, связанных с векторами, убедитесь, что вы понимаете идею об абстракции над числами.
2. Векторы являются основой линейных преобразований (в том числе поворотов и сдвигов), поэтому в случае попадания ваших интересов в 0.01%, вам потребуется более свободное владение этой абстракцией.
3. Как было сказано, векторы впервые появились в математике как следствие открытия кватернионов, поэтому кватернионы понять без понимания векторов не получится.
4. Векторный анализ это в первую очередь язык: обозначения, терминология. Основная проблема языков часто не в том, что они сложны, а в том, что они непривычны.

Это было историческо-философское вступление, а дальше мы как раз начнем изучать основы векторного языка.

Развилки в понимании

Вектор это нечто, имеющее величину и направление:

If anything has magnitude and direction, its magnitude and direction taken together constitute what is called a vector.

(Gibbs, Elements of Vector Analysis, 1881 [2])

Из этого определения можно попытаться что-то выудить. Раз есть направление, значит, есть какое-то пространство, в котором можно куда-то направляться. Если есть величина, то есть чем измерять. В принципе, это все.

Дальше понятие вектора приобретает различный смысл в зависимости от того, в какой контекст его погрузить. Есть три основные развилки, которые определяют дальнейшую работу с векторами.

Первое это физика. Под векторами понимаются силы, скорость, магнитное поле и т.п. В этой ситуации векторы описывают действия над объектами физического мира.

Второе это геометрия. Здесь вектор это отрезок между двумя точками пространства, имеющий начало и конец. Вектор указывает на то, как из одной точки попасть в другую. Или же можно понимать вектор как перенос одной точки в определенном направлении на определенное расстояние.

И третье — алгебра. Здесь векторы это нечто, что можно складывать и масштабировать, получая через эти две операции новые векторы. Слова из геометрии типа "направление", "отрезок" и т.п. здесь уже нелегитимны. Главное — операции.

* * *

В каждой из этих трех концепций будут свои развики, некоторые из которых мы затронем.

Сначала мы вспомним геометрическое представление, операции сложения и расстояния. Поговорим о проекциях и в этом контексте воспользуемся тригонометрией.

Затем перейдем к алгебраическому подходу, где воспользуемся координатным представлением вектора и увидим, как координатные операции связаны с геометрическим представлением. Поговорим о скалярном и векторном произведении, а также о действиях матриц на векторы. Это будет основная часть.

А когда будем говорить про комплексные числа и кватернионы, взглянем на векторы совсем иначе.

Наивное определение и базовые операции

1 Определение

Начнем с наглядного определения, которое легко воспринимается интуитивно, а затем скорректируем его так, чтобы все было стройно с математической точки зрения. Итак, дадим следующее определение:

Опр. Две точки A и B , называемые началом и концом, называются вектором \overrightarrow{AB} .

Под точкой мы понимаем неопределяемое понятие в смысле аксиом Евклида. Важно, что здесь мы не прибегаем к координатам.

* * *

Теперь введем понятие длины вектора. Строго говоря, длину нельзя задать численно, не определившись с единицами измерения. Правда, можно говорить о соотношениях длин ("длины равны", "одна длина больше другой в λ раз", и т.д.), чего нам в этой части будет достаточно. Но в текущем наивном изложении можно смело довериться интуиции. Нам сейчас важна исключительно терминология и обозначения, которые в дальнейшем будут часто использоваться.

Опр. Длина отрезка AB называется длиной или величиной (magnitude) вектора \overrightarrow{AB} и обозначается $|\overrightarrow{AB}|$.

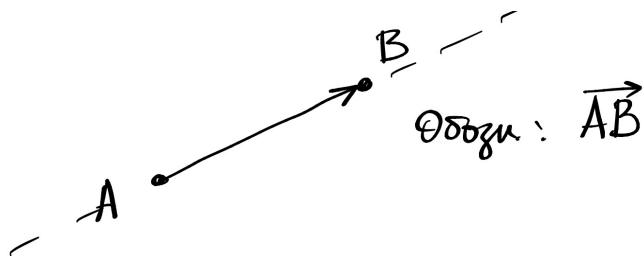


Рис. 3: Вектор как две точки

2 Операции

Вся суть работы с векторами заключается в операциях над ними. Как глаголы в речи, так операции в математике нужны для описания и анализа взаимодействия между объектами.

Подумаем, какие операции и как можно определить на векторах. Начнем с простого. Если два вектора имеют одно направление, то естественно, что их "сумма" должна иметь то же направление, а величина будет равна сумме величин этих векторов.

В общем случае (когда направления могут различаться) сумма векторов определяется следующим образом.

Опр. Сложение векторов, имеющих общее начало, осуществляется по правилу параллелограмма. Обозначается

$$\overrightarrow{AB} + \overrightarrow{AC} = \overrightarrow{AD}.$$

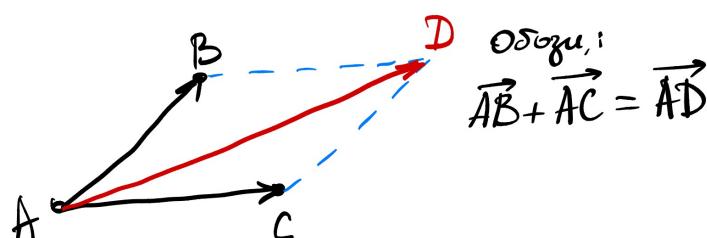


Рис. 4: Правило параллелограмма

Правило параллелограмма уже совсем не так естественно, однако оказывается, оно оправдывается физическими свойствами векторных величин (таких как скорость или сила).

Геометрически оно тоже понятно. Если воспринимать вектор как сдвиг точки, то сдвиг точки A сначала на вектор \overrightarrow{AB} , а затем на вектор \overrightarrow{AC} даст суммарный сдвиг в точку D .

* * *

Помимо сложения векторов, напрашивается еще одна операция – масштабирование вектора.

Опр. Умножение вектора \overrightarrow{AB} на положительное число λ (скаляр) дает новый вектор с тем же началом и направлением, но длиной $\lambda |\overrightarrow{AB}|$. Если $\lambda < 0$, то направление меняется на противоположное.

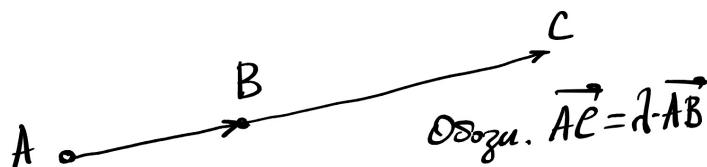


Рис. 5: Умножение вектора на скаляр

3 Как же сложение векторов с разными началами?

Прежде, чем углубиться в более сложные операции над векторами, такие как скалярное и векторное произведения, нужно решить одну важную проблему. В текущем определении сложения векторов по правилу параллелограмма требуется, чтобы

векторы имели общее начало. Иначе операция неопределена. Этую проблему можно решить, слегка пожертвовав наглядностью определения.

Вернемся к абстрактному определению Гиббса: вектором является все, что имеет направление и величину. Здесь ничего не сказано про то, что вектор должен иметь начало и конец. Эти понятия нам были нужны для наглядности, и они однозначно задают величину и направление. Но если этот вектор сдвинуть в другое место, он будет иметь те же величину и направление, но другое начало и конец, т.е. в нашем определении это новый вектор.

Поэтому нужно перестать различать векторы, полученные сдвигом какого-то одного вектора. Надо как бы их все склеить в один. Тогда такой вектор называется свободным. И тогда нет смысла указывать начало такого вектора, поэтому свободные векторы обозначаются маленькими буквами, как правило используют буквы $\mathbf{u}, \mathbf{v}, \mathbf{w}$.

N.B. Строго говоря, для определения свободного вектора нужно задать отношение эквивалентности на множестве всех векторов, имеющих начало и конец. Два вектора \overrightarrow{AB} и \overrightarrow{CD} эквивалентны, если \overrightarrow{CD} получается сдвигом из \overrightarrow{AB} . Тогда получившиеся классы эквивалентности образуют множество свободных векторов. Но если эти слова пугают, просто игнорируйте их, они просто формализуют приведенное выше объяснение.

Итак, свободный вектор можно отложить от любой точки. И вот на таких векторах можно задать сложение:

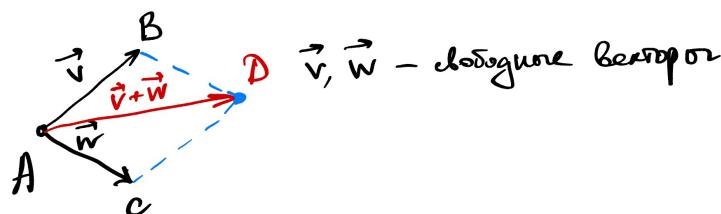
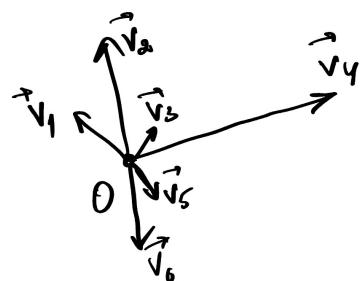


Рис. 6: Сложение свободных векторов

Можно считать, что все свободные векторы отложены от одной точки:



В дальнейшем мы будем называть свободные векторы просто векторами. В частности, те векторы, с которыми вы работаете, всегда являются свободными.

Умножение векторов

В этом разделе мы геометрически определим скалярное и векторное произведения на векторах. При этом я постараюсь смотивировать каждый шаг для тех, кто задается вопросами "а почему умножение именно такое?", "а почему это именно умножение?", "а зачем это вообще было нужно?". Если вам лично это не интересно, то можете

смело пропускать большую часть текста и вернуться к нему при необходимости в дальнейшем.

1 Векторная алгебра

В алгебре как части математики есть отдельный термин для алгебраического объекта под названием "алгебра". Такая получается перегрузка термина. Но в каком-то смысле она оправдана тем, что алгебра как часть математики по сути изначально занимается тем, что в абстрактной форме изучает арифметические операции над численными объектами, а алгебра как объект это множество, наделенное такими операциями:

1. сложение двух элементов множества
2. умножение элемента множества на скаляр
3. умножение двух элементов множества

А чистые алгебраисты – ярые любители таких абстракций, видят множество – сразу пытаются изобрести на нем операции. Поэтому им не нужны никакие стимулы, чтобы просто задаться вопросом "вот сложение на векторах есть, а вдруг можно и умножение определить?". Другими словами, они хотят придумать алгебру на векторах, или векторную алгебру. Побудем и мы в этом разделе такими алгебраистами.

Итак, мы имеем множество векторов, на котором уже задали сложение и умножение на скаляр. Для построения алгебры не хватает только умножения вектора на вектор. Но что такое в принципе умножение и как оно должно работать с векторами, совсем не очевидно.

2 Что такое умножение?

Продолжим погружаться в сознание алгебраиста и все же разберемся, что можно называть умножением. Логично, что раз исходно мы научились умножать числа, то и умножение других объектов должно быть сходно по своим свойствам лбычному численному умножению. А свойства у умножения чисел вот какие:

1. $a \cdot b = b \cdot a$ (перестановочность или, иначе, коммутативность)
2. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ (ассоциативность, т.е. нет разницы где поставлены скобки)
3. $a \cdot (b + c) = a \cdot b + a \cdot c$ (дистрибутивность, т.е. правило раскрытия скобок)

Первые два свойства как бы само собой разумеются, но, что важно, они присущи и операции сложения чисел. А вот дистрибутивность (третье свойство) как раз отличает умножение от сложения и указывает на взаимодействие этих двух операций.

Таким образом, умножение вектора на вектор должно, в идеале, также удовлетворять этим свойствам.

Для определения подобной операции есть два основных подхода, основанные на идеях проекции и ориентированной площади. В результате получаются две операции: скалярное и векторное умножение.

3 Скалярное произведение

Скалярное произведение (scalar/dot/inner product) основано на идеи проекции одного вектора на другой, а вернее на его направление:

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}| \cdot \underbrace{(\text{проекция } \mathbf{w} \text{ на } \mathbf{v})}_{|\mathbf{w}| \cos \varphi} = |\mathbf{v}| \cdot |\mathbf{w}| \cos \varphi$$

Здесь мы обозначили операцию умножения обычной точкой. Обычно вводят другую нотацию: (\mathbf{v}, \mathbf{w}) или $\langle \mathbf{v}, \mathbf{w} \rangle$ и т.п., но мы пока обойдемся без них.

Итак, можно было бы просто сказать, что

$$\mathbf{v} \cdot \mathbf{w} = \underbrace{(\text{проекция } \mathbf{w} \text{ на } \mathbf{v})}_{|\mathbf{w}| \cos \varphi} = |\mathbf{w}| \cos \varphi,$$

но тогда было бы не выполнялось равенство $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$.

Упр. Проверить, что при текущем определении свойство перестановочности выполняется, т.е. $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$.

Поэтому полезно понимать, что определение скалярного произведения таково из-за алгебраических соображений о свойствах умножения. Однако идея проекции при этом сохраняется, что дает в дальнейшем использовать эту операцию в геометрических расчетах. Отсюда же возникает необходимость в нормализации вектора. Об этом скажем чуть ниже.

А сейчас вернемся к теории и посмотрим на другие два свойства умножения. Ассоциативности у скалярного произведения нет. Если векторы \mathbf{v}_1 и \mathbf{v}_3 имеют разное направление, то очевидно, что

$$\mathbf{v}_1 \cdot \underbrace{(\mathbf{v}_2 \cdot \mathbf{v}_3)}_{\text{число}} \neq \underbrace{(\mathbf{v}_1 \cdot \mathbf{v}_2)}_{\text{число}} \cdot \mathbf{v}_3,$$

так как слева вектор получен масштабированием \mathbf{v}_1 , а справа – масштабированием \mathbf{v}_3 .

А вот дистрибутивность будет выполняться

$$\mathbf{v}_1 \cdot (\mathbf{v}_2 + \mathbf{v}_3) = \mathbf{v}_1 \cdot \mathbf{v}_2 + \mathbf{v}_1 \cdot \mathbf{v}_3,$$

что все-таки легализует слово "произведение" для этой операции.

Упр. Проверить это свойство самостоятельно.

Упр. Раскрыть скобки в выражении

$$(\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3) \cdot (\mathbf{v}_4 + \mathbf{v}_5).$$

Почувствуйте, что это делается так же, как с числами (спасибо дистрибутивности!).

Полезные свойства скалярного произведения:

- Если $\mathbf{v} \perp \mathbf{w}$, то $\varphi = \frac{\pi}{2}$, значит

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}| |\mathbf{w}| \cos \underbrace{\frac{\pi}{2}}_{=0} = 0.$$

- Скалярный квадрат дает квадрат длины вектора

$$\mathbf{v} \cdot \mathbf{v} = |\mathbf{v}| |\mathbf{v}| \underbrace{\cos 0}_{=1} = |\mathbf{v}|^2.$$

Поэтому $|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.

3. Знак скалярного произведения говорит о тупизне угла:

$$\mathbf{v} \cdot \mathbf{w} > 0 \Rightarrow \cos \varphi > 0 \Rightarrow -\frac{\pi}{2} < \varphi < \frac{\pi}{2} \text{ (острый угол)}$$

$$\mathbf{v} \cdot \mathbf{w} < 0 \Rightarrow \cos \varphi < 0 \Rightarrow \frac{\pi}{2} < \varphi < \frac{3\pi}{2} \text{ (тупой угол)}$$

4 Нахождение проекции на направление

Отдельно скажем про нахождение проекции вектора и нормализацию, так как это часто требуется в работе с графикой.

Как изначально было сказано, скалярное произведение основано на идеи проекции одного из векторов на направление другого вектора. Дальше мы подкорректировали формулу для выполнения свойства перестановочности и получили итоговое определение:

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}| |\mathbf{w}| \cos \varphi.$$

Если бы вектор \mathbf{w} имел длину $|\mathbf{w}| = 1$, то мы бы получили

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}| |\mathbf{w}| \cos \varphi = |\mathbf{v}| \cos \varphi,$$

т.е. буквально проекцию \mathbf{v} на направление \mathbf{w} . А так как направление от масштабирования не меняется, то это нужная нам изначально проекция.

Итак, если нужно найти проекцию на направление вектора \mathbf{w} и при этом $|\mathbf{w}| \neq 1$, то нужно взять вектор длины 1 с тем же направлением, а именно, подойдет вектор

$$\mathbf{w}' = \frac{1}{|\mathbf{w}|} \mathbf{w}.$$

Он имеет то же направление, что и \mathbf{w} , но длина $|\mathbf{w}'| = 1$. И тогда

$$(\text{длина проекции } \mathbf{v} \text{ на } \mathbf{w}) = \mathbf{v} \cdot \mathbf{w}'.$$

Говорят, что вектор \mathbf{w}' нормированный, или что он получен из вектора \mathbf{w} нормализацией.

Вопрос. Что геометрически означает, что $(\mathbf{v}, \mathbf{w}) = 0$?

Вопрос. Как геометрически выглядит множество решений уравнения $(\mathbf{v}, \mathbf{x}) = 0$?

5 Векторное произведение

В скалярном произведении, как было замечено, есть слабое место, а именно отсутствие ассоциативности. Вместе с этим вызывает напряжение то, что результатом перемножения двух векторов получается не третий вектор, а число. Такое алгебраистам тоже обычно не нравится чисто эстетически, а физиков беспокоит то, что в результате перемножения теряется природа "направленности" величины.

Тут начинается головоломка, как же тогда определить произведение, чтобы в результате получался вектор, и выполнялись все свойства умножения. Самая наглядная идея, которая приводит к нужному результату – идея ориентированной площади.

В чем смысл? Исходно есть два вектора \mathbf{v} и \mathbf{w} . На них можно, как говорят, натянуть параллелограмм. Но чтобы сохранить векторную составляющую (направление), выбирается ориентация этого параллелограмма. Можно представлять это, например, как направление обхода периметра.

Тут интуиция ломает мозг, потому что теперь вместо числа на выходе получается вообще странный объект, который является чем угодно, но не вектором. На таких

ориентированных параллелограммах нужно бы опять определять сложение, умножение на скаляр и т.д. А потом и умножать их тоже захочется... В общем, пахнет жареным.

Все это можно сделать, и называется это *внешней алгеброй*. Но мы, конечно, этого делать не будем.

Нас спасет то, что вместо ориентированного параллелограмма мы возьмем вектор $\mathbf{v} \times \mathbf{w}$ с длиной, равной площади этого параллелограмма, т.е.

$$|\mathbf{v} \times \mathbf{w}| = S_{\text{parallelogram}(\mathbf{v}, \mathbf{w})} = |\mathbf{v}| |\mathbf{w}| \sin \varphi,$$

а направление его будет перпендикулярно плоскости векторов \mathbf{v} и \mathbf{w} .

Но перпендикуляр к плоскости можно задать двумя способами. Стандартно вектор $\mathbf{v} \times \mathbf{w}$ задается так, что если смотреть из его конца, то кратчайший поворот от \mathbf{v} к \mathbf{w} происходит против часовой стрелки. Это также называется *правилом правой руки*.

Такая операция $\mathbf{v} \times \mathbf{w}$ называется *векторным произведением векторов \mathbf{v} и \mathbf{w}* (vector/cross product).

Что насчет свойств умножения? Для векторного произведения верны следующие утверждения:

1. $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$ (антиперестановочность или антикоммутативность)
2. $\mathbf{v}_1 \times (\mathbf{v}_2 \times \mathbf{v}_3) = (\mathbf{v}_1 \times \mathbf{v}_2) \times \mathbf{v}_3$
3. $\mathbf{v}_1 \times (\mathbf{v}_2 + \mathbf{v}_3) = \mathbf{v}_1 \times \mathbf{v}_2 + \mathbf{v}_1 \times \mathbf{v}_3$

Упр. Доказать это.

Из-за учета ориентации возникает антиперестановочность: изменение порядка умножения влечет изменение знака результирующего вектора. Это особенность векторной алгебры, от которой никуда не деться. В остальном все свойства совпадают с умножением чисел.

Отметим важные на практике факты:

1. Если векторы \mathbf{v} и \mathbf{w} параллельны, то $\mathbf{v} \times \mathbf{w} = \mathbf{0}$, т.к. угол между ними равен либо 0 , либо π .
2. Векторное произведение дает на выходе вектор, перпендикулярный плоскости, образованной векторами \mathbf{v} и \mathbf{w} , т.е. перпендикулярный обоим векторам:

$$\mathbf{v}_1 \times \mathbf{v}_2 \perp \mathbf{v}_1, \mathbf{v}_2,$$

или то же в терминах скалярного произведения:

$$(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_1 = 0 \quad \text{и} \quad (\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_2 = 0.$$

Вектор и координатная запись

В этом разделе мы разберем координатный подход в работе с векторами. Хотя слово "координата" применимо скорее к точкам, чем к векторам, но разница здесь достаточно тонкая. Я постараюсь быть аккуратным в терминах в этой части, чтобы отделить мух от котлет. А в последующем, если понимать, о чём идет речь, то можно будет снова употреблять эти термины так, как привыкли.

1 Базис и компоненты вектора

Векторы можно складывать, а можно наоборот раскладывать в сумму других векторов. Это бывает крайне удобно во многих ситуациях, в том числе практических.

Например, объект движется с некоторой скоростью, а нас интересует только составляющая его скорости в заданном направлении. Допустим, хотим узнать, движется ли объект вертикально, тогда нужно просто взять проекцию вектора скорости на вертикальное направление. Если проекция нулевая, то не движется.

А может, нам нужно проанализировать по отдельности проекции на разные направления. Тогда надо искать их.

Разложение вектора на сумму более простых упрощает его анализ, как разбор слова на морфемы, разбор здания на кирпичики или разложение цвета на RGB-палитру, примеры можно придумывать долго. Вот набор таких примитивов среди векторов, из которых можно построить любой другой вектор, называется векторным базисом.

* * *

Базисом называется любой набор векторов, через который можно выразить любой другой вектор, причем единственным способом. Обычно базисные векторы обозначают буквами $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$.

В одномерном случае достаточно одного вектора \mathbf{e} , из которого масштабированием можно получить любой другой вектор: все вектора будут иметь вид

$$\mathbf{v} = \alpha \mathbf{e}.$$

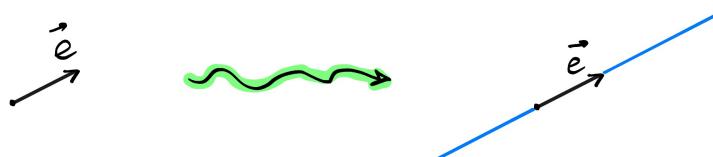


Рис. 7: Один (любой) вектор порождает "прямую" – все векторы того же направления.

В двумерном случае нужно взять два вектора \mathbf{e}_1 и \mathbf{e}_2 с различными направлениями, а любой вектор будет их комбинацией

$$\mathbf{v} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2.$$

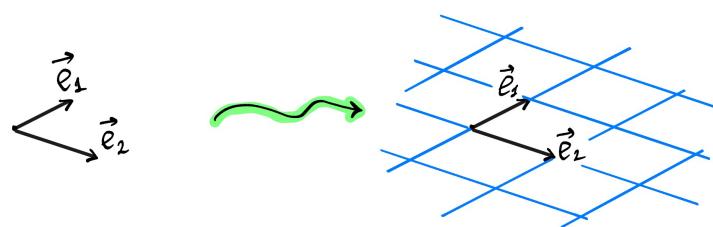


Рис. 8: Базис в двумерном векторном пространстве.

Наконец, для трехмерного случая нужен базис из трех векторов $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, который порождает векторы

$$\mathbf{v} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \alpha_3 \mathbf{e}_3.$$

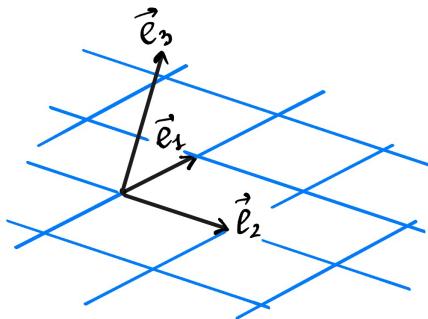


Рис. 9: Базис в трехмерном векторном пространстве.

Часто это же пишут не в виде суммы векторов, а в координатном виде

$$\mathbf{v} = (\alpha_1, \alpha_2, \alpha_3),$$

а числа $\alpha_1, \alpha_2, \alpha_3$ называют компонентами вектора \mathbf{v} в базисе $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$.

Например, базисные векторы имеют следующий вид в координатной записи

$$\begin{aligned}\mathbf{e}_1 &= (1, 0, 0), \\ \mathbf{e}_2 &= (0, 1, 0), \\ \mathbf{e}_3 &= (0, 0, 1).\end{aligned}$$

Наоборот, если дан вектор, например, $\mathbf{v} = (1, -1, 3)$, то это значит, что его можно представить в виде комбинации базисных векторов:

$$\mathbf{v} = 1 \cdot \mathbf{e}_1 + (-1) \cdot \mathbf{e}_2 + 3 \cdot \mathbf{e}_3.$$

N.B. Ясно, что выбрать три вектора, которые бы не лежали в одной плоскости, можно многими способами, поэтому и базисов бесконечно много. Однако дальше мы увидим, что выбор базиса может упростить или усложнить расчеты, поэтому важно выбрать правильный базис.

Вопрос. Что означает геометрически, если только одна из компонент вектора не равна нулю? Например, $(0, \alpha_2, 0)$.

Вопрос. Что означает геометрически, если одна из компонент вектора равна нулю? Например, $(0, \alpha_2, \alpha_3)$.

2 Сложение векторов в координатах

Давайте сложим два вектора \mathbf{v} и \mathbf{w} с компонентами

$$\begin{aligned}\mathbf{v} &= (\alpha_1, \alpha_2, \alpha_3) \\ \mathbf{w} &= (\beta_1, \beta_2, \beta_3).\end{aligned}$$

Запишем в виде разложения по базису и сложим:

$$\begin{aligned}\mathbf{v} + \mathbf{w} &= (\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \alpha_3 \mathbf{e}_3) + (\beta_1 \mathbf{e}_1 + \beta_2 \mathbf{e}_2 + \beta_3 \mathbf{e}_3) = \\ &= (\alpha_1 + \beta_1) \mathbf{e}_1 + (\alpha_2 + \beta_2) \mathbf{e}_2 + (\alpha_3 + \beta_3) \mathbf{e}_3.\end{aligned}$$

Таким образом, суммарный вектор $\mathbf{v} + \mathbf{w}$ имеет компоненты

$$\mathbf{v} + \mathbf{w} = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \alpha_3 + \beta_3).$$

То есть для суммирования векторов, нужно просто сложить их покомпонентно. Можно то же самое переписать в таком виде:

$$(\alpha_1, \alpha_2, \alpha_3) + (\beta_1, \beta_2, \beta_3) = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \alpha_3 + \beta_3),$$

и тогда можно забыть о геометрии и просто смотреть на вектор как на набор из трех чисел, которые складываются по этому правилу.

Упр. Нарисуйте какой-нибудь базис $\mathbf{e}_1, \mathbf{e}_2$ в двумерном пространстве. А затем изобразите в это базисе вектор $(-2, 4)$.

Упр. Сложите три вектора $\mathbf{u} = (1, -1, 3)$, $\mathbf{v} = (0, 0, -5)$, $\mathbf{w} = (-2, 2, 3)$. Изобразите результат в каком-нибудь случайно нарисованном базисе $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$.

3 Умножение вектора на скаляр в координатах

Возьмем вектор $\mathbf{v} = (\alpha_1, \alpha_2, \alpha_3)$, умножим его на скаляр λ и найдем его компоненты. Для этого произведем операции над вектором \mathbf{v} в разложении по базису:

$$\begin{aligned}\lambda \mathbf{v} &= \lambda(\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \alpha_3 \mathbf{e}_3) = \\ &= \lambda\alpha_1 \mathbf{e}_1 + \lambda\alpha_2 \mathbf{e}_2 + \lambda\alpha_3 \mathbf{e}_3 = \\ &= (\lambda\alpha_1, \lambda\alpha_2, \lambda\alpha_3).\end{aligned}$$

Таким образом, при умножении вектора на скаляр, каждая его компонента умножается на этот скаляр.

Упр. Даны векторы $\mathbf{v} = (-2, 2)$, $\mathbf{w} = (-3, 2)$. Найти компоненты вектора

$$\mathbf{u} = -4\mathbf{v} + 2\mathbf{w}.$$

4 Скалярное произведение в координатах

Попробуем теперь скалярно перемножить два вектора. Для простоты возьмем двумерный случай:

$$\begin{aligned}\mathbf{v} &= \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 \\ \mathbf{w} &= \beta_1 \mathbf{e}_1 + \beta_2 \mathbf{e}_2\end{aligned}$$

тогда согласно свойствам перестановочности и дистрибутивности, раскрываем скобки, словно работаем с числами:

$$\begin{aligned}\mathbf{v} \cdot \mathbf{w} &= (\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2) \cdot (\beta_1 \mathbf{e}_1 + \beta_2 \mathbf{e}_2) = \\ &= \alpha_1 \beta_1 \mathbf{e}_1 \cdot \mathbf{e}_1 + \alpha_1 \beta_2 \mathbf{e}_1 \cdot \mathbf{e}_2 + \alpha_2 \beta_1 \mathbf{e}_2 \cdot \mathbf{e}_1 + \alpha_2 \beta_2 \mathbf{e}_2 \cdot \mathbf{e}_2 = \\ &= \alpha_1 \beta_1 |\mathbf{e}_1|^2 + (\alpha_1 \beta_2 + \alpha_2 \beta_1) \mathbf{e}_1 \cdot \mathbf{e}_2 + \alpha_2 \beta_2 |\mathbf{e}_2|^2.\end{aligned}$$

Так как базис выбирать можно каким угодно (он все равно порождает одно и то же трехмерное пространство), то глядя на эту формулу, можно понять, какой базис самый удобный.

Во-первых, хочется, чтобы среднего слагаемого не было, т.е. чтобы $\mathbf{e}_1 \cdot \mathbf{e}_2 = 0$. Геометрически это значит, что базисные векторы должны быть перпендикулярны. Тогда получим

$$\mathbf{v} \cdot \mathbf{w} = \alpha_1 \beta_1 |\mathbf{e}_1|^2 + \alpha_2 \beta_2 |\mathbf{e}_2|^2.$$

Во-вторых, хочется, чтобы не было множителей $|\mathbf{e}_1|^2$ и $|\mathbf{e}_2|^2$. То есть чтобы они были равны единице. Геометрически это значит, что базисные векторы должны быть единичной длины. Тогда

$$\mathbf{v} \cdot \mathbf{w} = \alpha_1\beta_1 + \alpha_2\beta_2.$$

Это и есть наша привычная формула скалярного произведения. Если обозначить компоненты привычными буквами x, y , а базисные векторы $\mathbf{e}_x, \mathbf{e}_y$:

$$\begin{aligned}\mathbf{v} &= x_1\mathbf{e}_x + y_1\mathbf{e}_y = (x_1, y_1) \\ \mathbf{w} &= x_2\mathbf{e}_x + y_2\mathbf{e}_y = (x_2, y_2),\end{aligned}$$

то получим формулу

$$\mathbf{v} \cdot \mathbf{w} = x_1x_2 + y_1y_2.$$

Таким образом, формула скалярного произведения зависит от того, по какому базису разложили векторы! Самый простой вид эта формула имеет в *ортонормированном* базисе: когда векторы базиса попарно перпендикулярны (ортогональны) и нормированы.

В трехмерном случае, соответственно, получим формулу

$$\mathbf{v} \cdot \mathbf{w} = x_1x_2 + y_1y_2 + z_1z_2.$$

N.B. В дальнейшем у нас будут только ортонормированные базисы.

Вопрос. Дан вектор $\mathbf{v} = (x, y, z)$. Чему равны скалярные произведения с базисными векторами

$$\mathbf{v} \cdot \mathbf{e}_x, \quad \mathbf{v} \cdot \mathbf{e}_y, \quad \mathbf{v} \cdot \mathbf{e}_z?$$

Упр. Найти скалярное произведение векторов

$$\begin{aligned}\mathbf{v} &= (5, -2, 4), \\ \mathbf{w} &= (-3, 1, 7).\end{aligned}$$

Упр. Для вектора $\mathbf{v} = (1, 2, 3)$ найти какой-нибудь вектор, перпендикулярный ему.

5 Длина вектора и угол между векторами

В прошлом разделе о векторах, мы получили формулу длины вектора

$$|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

Если вектор \mathbf{v} имеет компоненты $\mathbf{v} = (x, y, z)$, то по полученной формуле для скалярного произведения имеем

$$|\mathbf{v}| = \sqrt{x^2 + y^2 + z^2}.$$

Также из геометрического определения скалярного произведения

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}| |\mathbf{w}| \cos \varphi$$

следует, что угол φ выражается следующим образом

$$\varphi = \arccos \frac{\mathbf{v} \cdot \mathbf{w}}{|\mathbf{v}| |\mathbf{w}|} = \arccos \frac{v_x w_x + v_y w_y + v_z w_z}{\sqrt{v_x^2 + v_y^2 + v_z^2} \cdot \sqrt{w_x^2 + w_y^2 + w_z^2}},$$

где $\mathbf{v} = (v_x, v_y, v_z)$, $\mathbf{w} = (w_x, w_y, w_z)$.

Обычно подобные страшные формулы нет надобности использовать. Это бич координатной записи: возникает громоздкость формул из-за множества букв и индексов.

Упр. Даны векторы

$$\mathbf{v} = (5, -2, 4), \\ \mathbf{w} = (-3, 1, 7).$$

Найдите их длины и угол между ними.

6 Нормализация вектора

Нормализация вектора \mathbf{v} это процесс приведения его к вектору \mathbf{v}' , который имеет то же направление, но при этом $|\mathbf{v}'| = 1$.

Для того, чтобы отнормировать вектор, достаточно его умножить на величину, обратную к его длине (можно лихо сказать "поделить на его длину", но формально все же операции деления не вводилось):

$$\mathbf{v}' = \frac{1}{|\mathbf{v}|} \mathbf{v}.$$

Нормализация пригождается в разнообразных ситуациях. Посмотрим на несколько таких кейсов.

Во-первых, мы уже видели выше, что базисные векторы оказалось удобно взять именно отнормированными, иначе бы скалярное произведение считалось менее удобно.

Во-вторых, если векторы \mathbf{v}, \mathbf{w} нормированы, то формула для угла становится проще:

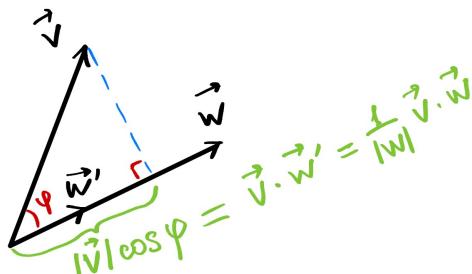
$$\varphi = \arccos(\mathbf{v} \cdot \mathbf{w}).$$

Наконец, если нужно найти длину проекции вектора \mathbf{v} на направление вектора \mathbf{w} , т.е. найти $|\mathbf{v}| \cos \varphi$, то можно сделать так

$$|\mathbf{v}| \cos \varphi = \frac{\mathbf{v} \cdot \mathbf{w}}{|\mathbf{w}|},$$

а можно вначале отнормировать \mathbf{w} , получив единичный вектор \mathbf{w}' и найти скалярное произведение

$$|\mathbf{v}| \cos \varphi = \mathbf{v} \cdot \mathbf{w}'.$$



Упр. Нормализовать вектор $(1, 2, 3)$.

Вопрос. Что геометрически представляет из себя множество решений уравнения

$$|\mathbf{v}| = 1?$$

Указание. Напишите это уравнение в координатном виде (по определению $|\mathbf{v}|$). Что задает это уравнение?

7 Векторное произведение в координатах

Здесь пришло время еще несколько уточнить насчет выбора базиса. Дело в том, что ортонормированный базис отличаться с точки зрения векторного произведения или, как говорят, *ориентацией*.

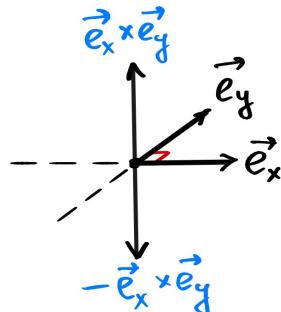
Допустим у нас есть два ортонормированных вектора $\mathbf{e}_x, \mathbf{e}_y$. Как выбрать третий вектор \mathbf{e}_z ? Есть два способа, которые на языке векторных произведений выглядят так:

$$\mathbf{e}_z = \pm \mathbf{e}_x \times \mathbf{e}_y.$$

Оба вектора будут единичными, оба перпендикулярны плоскости векторов \mathbf{e}_x и \mathbf{e}_y . В случае, когда $\mathbf{e}_z = \mathbf{e}_x \times \mathbf{e}_y$ говорят, что базис положительно ориентирован, а если $\mathbf{e}_z = -\mathbf{e}_x \times \mathbf{e}_y$ – отрицательно.

Для стандартного базиса традиционно выбирается положительная ориентация, т.е.

$$\mathbf{e}_z = \mathbf{e}_x \times \mathbf{e}_y.$$



Вопрос. Чему равняется $\mathbf{e}_x \times \mathbf{e}_y, \mathbf{e}_x \times \mathbf{e}_z, \mathbf{e}_y \times \mathbf{e}_z$?

Вопрос. Чему равняется $\mathbf{e}_x \times \mathbf{e}_z, \mathbf{e}_z \times \mathbf{e}_y$?

Вопрос. Чему равняется $\mathbf{e}_x \times \mathbf{e}_x, \mathbf{e}_y \times \mathbf{e}_y, \mathbf{e}_z \times \mathbf{e}_z$?

Вопрос. Что геометрически из себя представляет множество решений векторного уравнения $\mathbf{v} \times \mathbf{x} = 0$?

Зная ответы на эти вопросы, мы спокойно можем найти координатную формулу для векторного произведения. Для этого нужно проделать аналогичную процедуру, которую делали для скалярного произведения.

Итак, возьмем два вектора

$$\begin{aligned}\mathbf{v} &= x_1 \mathbf{e}_x + y_1 \mathbf{e}_y + z_1 \mathbf{e}_z = (x_1, y_1, z_1), \\ \mathbf{w} &= x_2 \mathbf{e}_x + y_2 \mathbf{e}_y + z_2 \mathbf{e}_z = (x_2, y_2, z_2),\end{aligned}$$

и векторно перемножим их

$$\begin{aligned}
 \mathbf{v} \times \mathbf{w} &= (x_1 \mathbf{e}_x + y_1 \mathbf{e}_y + z_1 \mathbf{e}_z) \times (x_2 \mathbf{e}_x + y_2 \mathbf{e}_y + z_2 \mathbf{e}_z) = \\
 &= x_1 x_2 \underbrace{(\mathbf{e}_x \times \mathbf{e}_x)}_{=0} + x_1 y_2 (\mathbf{e}_x \times \mathbf{e}_y) + x_1 z_2 (\mathbf{e}_x \times \mathbf{e}_z) + \\
 &\quad + y_1 x_2 \underbrace{(\mathbf{e}_y \times \mathbf{e}_x)}_{=-(\mathbf{e}_x \times \mathbf{e}_y)} + y_1 y_2 \underbrace{(\mathbf{e}_y \times \mathbf{e}_y)}_{=0} + y_1 z_2 (\mathbf{e}_y \times \mathbf{e}_z) + \\
 &\quad + z_1 x_2 \underbrace{(\mathbf{e}_z \times \mathbf{e}_x)}_{=-\mathbf{e}_x \times \mathbf{e}_z} + z_1 y_2 \underbrace{(\mathbf{e}_z \times \mathbf{e}_y)}_{=-\mathbf{e}_y \times \mathbf{e}_z} + z_1 z_2 \underbrace{(\mathbf{e}_z \times \mathbf{e}_z)}_{=0} = \\
 &= (x_1 y_2 - y_1 x_2) \underbrace{\mathbf{e}_x \times \mathbf{e}_y}_{=\mathbf{e}_z} + (x_1 z_2 - z_1 x_2) \underbrace{\mathbf{e}_x \times \mathbf{e}_z}_{=-\mathbf{e}_y} + (y_1 z_2 - z_1 y_1) \underbrace{\mathbf{e}_y \times \mathbf{e}_z}_{=\mathbf{e}_x} = \\
 &= (x_1 y_2 - y_1 x_2) \mathbf{e}_z - (x_1 z_2 - z_1 x_2) \mathbf{e}_y + (y_1 z_2 - z_1 y_1) \mathbf{e}_x.
 \end{aligned}$$

То есть в координатной записи получим

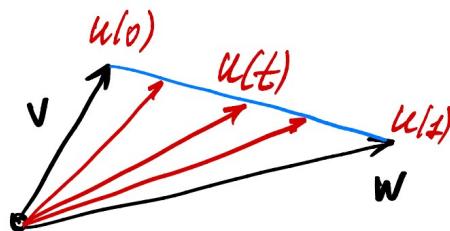
$$\mathbf{v} \times \mathbf{w} = (y_1 z_2 - z_1 y_1, z_1 x_2 - x_1 z_2, x_1 y_2 - y_1 x_2).$$

8 Линейная интерполяция между двумя векторами

Рассмотрим два вектора \mathbf{v}, \mathbf{w} . Задача заключается в том, чтобы задать переход от \mathbf{v} к \mathbf{w} по прямой линии.

Возьмем вектор, который имеет направление от конца \mathbf{v} к концу \mathbf{w} . Этим вектором является вектор $\mathbf{w} - \mathbf{v}$. Его длина это длина между \mathbf{v} и \mathbf{w} . Поэтому можем задать вектор $\mathbf{u}(t)$, зависящий от параметра $t \in [0, 1]$ следующим образом:

$$\mathbf{u}(t) = \mathbf{v} + t(\mathbf{w} - \mathbf{v}).$$



Здесь с ростом t происходит движение по отрезку между концами \mathbf{v} и \mathbf{w} . Эта формула может быть переписана иначе:

$$\mathbf{u}(t) = (1 - t)\mathbf{v} + t\mathbf{w}.$$

Аффинные преобразования

Вектор как абстракция может иметь самые разные интерпретации. Об этом мы упоминали в начале предыдущей главы. Впрочем, мы не говорили, что вектором может считаться и функция. Тогда возникают бесконечномерные пространства, так как ни один конечный набор функций не порождает все остальные функции, т.е. базис должен состоять из бесконечного числа функций. Грубо говоря, бесконечность базиса – основное отличие двух области математики: линейной алгебры и функционального анализа.

У нас же есть конкретная задача: описывать преобразование геометрических фигур. Для этого нужно описывать точки в трехмерном пространстве и их преобразования, так что спокойно – нам бесконечность не светит.

В этой главе мы совместим понятия точек и векторов в одну структуру, которую в линейной алгебре называют аффинным пространством (*affine space*), а преобразования (*transformations*) в нем – аффинными. Термин несколько страшный, но это название буквально того, с чем мы работаем в 3D-редакторах.¹

Переходя от разговора о векторах к разговору об аффинных пространствах, отметим идейно, что такое вектор, а что такое точка.

Точка это неделимая часть пространства, из точек складываются множества точек – геометрические фигуры. Грубо скажем, что точка это материальная неделимая частица. Ее нельзя ”сложить с другой точкой”, ”умножить на число”, зато ее можно наблюдать и двигать.

Вектор – инструмент *воздействия* на точку, к которой он приложен. А именно, его действие – перенос точки в заданном направлении на заданную дистанцию в другую точку. Вектор не имеет материального окраса, его нельзя увидеть или поймать в микроскоп или телескоп. Но зато действия над векторами имеют алгебраическую структуру: их можно складывать, умножать на число, задавать скалярное и векторное произведение и т.д.

При этом часто мы, говоря о векторах, думаем о точках, и наоборот. Это происходит потому, что точки часто отождествляются с векторами, а точнее с радиус-векторами. Далее об этом будет сказано подробнее.

Аффинные системы координат

1 Аффинное пространство

Аффинным пространством (A, V) называется совокупность множества всех точек A и множества всех векторов V . Причем задана операция сложения точки и вектора, в результате которой получается новая точка. Другими словами, есть операция *сдвига* (*translation*):

$$P + \mathbf{v}, \quad \text{где } P \in A - \text{точка}, \mathbf{v} \in V - \text{вектор}.$$

Множество A можно представлять себе просто как абстрактное пространство точек (прямая, плоскость, 3D-пространство). Здесь мы еще не имеем права говорить о координатах, так как для этого нужно задать систему координат.

Аналогично V – абстрактное множество всех векторов, на котором, как раньше, можно задавать всевозможные базисы.

¹Ну, ладно, не совсем – мы обычно работаем с *движениями* (сдвиги и повороты), это подмножество аффинных преобразований, которые допускают еще и растяжения в произвольных направлениях.

Сразу посмотрим на ситуацию, когда к точке A прибавили вектор \mathbf{v} и получили точку B :

$$A + \mathbf{v} = B.$$

В такой ситуации вектор \mathbf{v} можно обозначить \overrightarrow{AB} , тогда имеем

$$A + \overrightarrow{AB} = B.$$

Теперь, если чисто формально, не задумываясь об осмысленности, перебросить букву A направо, то получим запись

$$\overrightarrow{AB} = B - A.$$

Получается, чтобы найти вектор, который переносит точку A в точку B , нужно в каком-то смысле из точки B "вычесть" точку A . Об интерпретации этой записи скажем чуть ниже.

N.B. Отметим, что здесь запись \overrightarrow{AB} означает не больше и не меньше, чем вектор \mathbf{v} , такой, что

$$A + \mathbf{v} = B.$$

Да, интуитивно с ним можно работать, как с тем геометрическим вектором с началом A и концом B , который мы определяли в самом начале, но сейчас мы говорим о векторах в другой, алгебраической, парадигме.

2 Системы координат

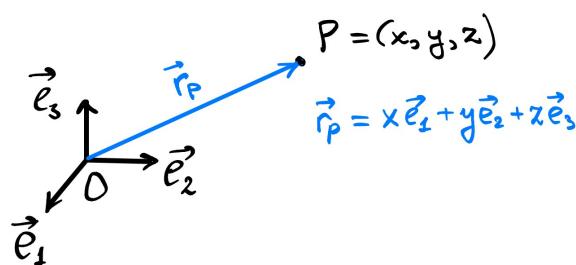
Аффинной системой координат $(O, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ называется совокупность точки $O \in A$ и базиса $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ на V . Точка P в этом случае называется *началом координат (origin)* или *точкой отсчета*².

Другими словами, система координат определяется базисом, который закрепили в выбранной точке пространства.

Радиус-вектором точки P в системе координат $(O, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ называется такой вектор $\mathbf{r}_P \in V$, что

$$O + \mathbf{r}_P = P.$$

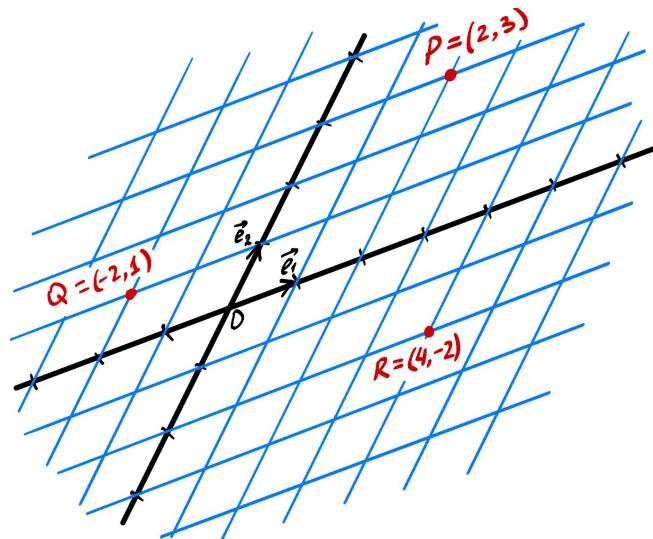
Иначе вектор \mathbf{r}_P можно обозначить \overrightarrow{OP} .



Координатами точки P в системе координат $(O, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ называются координаты (вернее, компоненты разложения) ее радиус-вектора \overrightarrow{OP} .

Образно говоря, координаты точки это маршрут, по которому до нее можно добраться из начала координат, используя базисные векторы в качестве транспорта.

²Вообще, говорят, что начало координат и точка отсчета – не совсем синонимы.

Рис. 10: Координатная сетка с отмеченными на ней точками P, Q, R .

3 Координаты вектора \overrightarrow{AB}

Вернемся к ситуации

$$A + \overrightarrow{AB} = B.$$

Перепишем это выражение иначе через радиус векторы точек A, B . А именно, так как

$$\begin{aligned} A &= O + \overrightarrow{OA}, \\ B &= O + \overrightarrow{OB}, \end{aligned}$$

то получаем

$$O + \overrightarrow{OA} + \overrightarrow{AB} = O + \overrightarrow{OB}.$$

Теперь вычтем из обеих частей вектор \overrightarrow{OA} :

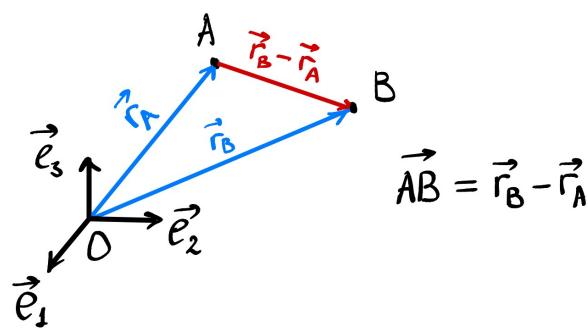
$$O + \overrightarrow{AB} = O + (\overrightarrow{OB} - \overrightarrow{OA}),$$

отсюда получаем, что

$$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = \mathbf{r}_B - \mathbf{r}_A.$$

Таким образом, для того, чтобы найти координаты вектора \overrightarrow{AB} , соединяющего точки A и B , нужно вычесть координаты точки A из координат точки B :

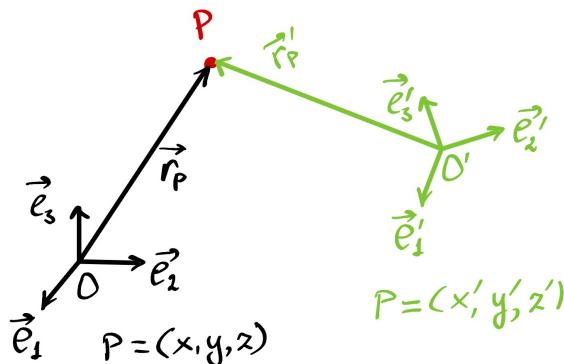
$$\overrightarrow{AB} = (x_B - x_A, y_B - y_A, z_B - z_A).$$



Вопрос. Зададим точки $A = (1, -5, 7)$, $B = (-3, -8, 19)$ в какой-то системе координат. Найти координаты вектора \vec{AB} в той же системе координат.

4 Преобразование системы координат

Если сменить точку отсчета или векторный базис, то получим новую систему координат $(O', \mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3)$, в которой точки будут иметь новые координаты.



Отметим, что здесь мы не можем получить криволинейных координат. Этим аффинное пространство отличается от пространства в общем виде.

Также важно помнить, что смена системы координат никак не меняет само аффинное пространство: точки остаются те же самые на тех же "местах", векторы тоже остаются неизменными. Меняется только "точка зрения" на них, способ их выразить числами.

В следующих пунктах мы разберем по-отдельности два необходимых нам способа преобразования системы координат: сдвиг и поворот. А в конце мы их объединим в одно преобразование, которое мы обычно в 3D-движках называем *трансформами*³:

$$\text{Трансформ с.к.} = \text{Сдвиг начала} + \text{Трансформ базиса}$$

Любой сдвиг задается вектором. С базисом все значительно сложнее.

Преобразование базиса (Basis Transform) отвечает сразу за многие действия, которые можно классифицировать так:

1. поворот (Rotation),
2. масштаб (Scale),
3. скосы (Shear),
4. отражения (Reflection).

Любое преобразование базиса представимо единым объектом – матрицей размера 3x3 (таблица чисел из трех столбцов и трех строк). В сущности, матрица – аналог координат для более сложных преобразований, чем для сдвига. И она так же, как координаты вектора, меняется при смене базиса.

О матрицах можно выстроить целую алгебраическую теорию. Нам это будет не нужно, так как вообще-то для нас не так важно, как реализуется само преобразование объекта на уровне математики. Мы лишь затронем по касательной эту тему, чтобы при необходимости, вы могли понять, о чём идет речь.

5 Сдвиг системы координат

³Тут я хочу соединить язык, принятый в 3D-движках, и принятый в русскоязычной математике.

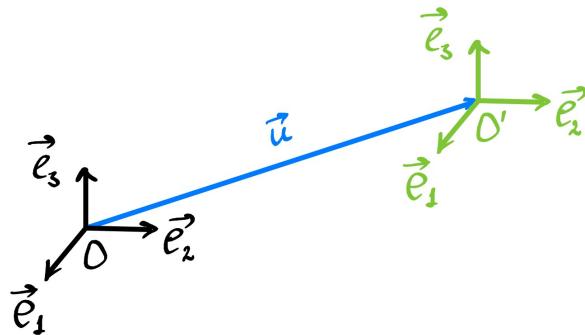
Начнем с самой простого преобразования системы координат – сдвига.

Пусть у нас есть система координат $(O, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, будем называть ее старой, а координаты точек будем обозначать (x, y, z) .

Сдвинем начало координат O на вектор \mathbf{u} , получим новую точку O' :

$$O' = O + \mathbf{u}.$$

Базис векторов никак не изменился. Полученную систему координат $(O', \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ будем называть новой системой координат, а координаты точек в ней будем обозначать (x', y', z') .



Н.В. Часто даже говорят "переход от координат (x, y, z) к (x', y', z') " или "...в координатах (x', y') ", имея в виду не координаты конкретной точки, а сами системы координат $(O, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ и $(O', \mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3)$. Это делается для лаконичности речи или письма.

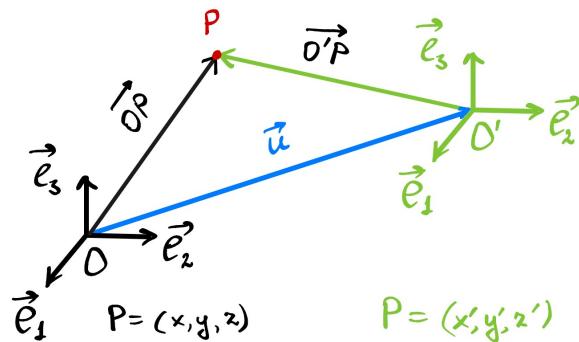
Нам надо научиться переводить старые координаты точки в новые и наоборот. Принцип, к которому мы приедем здесь, будет работать и с поворотами, а значит, и с матрицами и кватернионами.

* * *

Итак, возьмем произвольную точку P и допустим, что нам известны ее координаты в старой системе координат:

$$P = (x, y, z).$$

Найдем координаты (x', y', z') в новой системе координат.



Для этого, по определению, нужно найти координаты радиус-вектора $\overrightarrow{O'P}$ точки P в базисе новой системы координат.

Так как базисы совпадают, то любой вектор будет иметь одинаковые координаты в обеих системах координат, но зато сами по себе радиус-векторы точки P будут различаться! А значит, будут различаться и их координаты.

Радиус-вектор точки P в новой системе координат выражается так:

$$\overrightarrow{O'P} = \overrightarrow{O'O} + \overrightarrow{OP} = \overrightarrow{OP} - \mathbf{u}.$$

Здесь мы воспользовались тем, что $\overrightarrow{O'O} = -\overrightarrow{OO'} = -\mathbf{u}$.

Таким образом,

$$(x', y', z') = (x, y, z) - \mathbf{u},$$

т.е. для того, чтобы найти новые координаты, нужно из старых вычесть координаты вектора сдвига:

$$(x', y', z') = (x - x_{\mathbf{u}}, y - y_{\mathbf{u}}, z - z_{\mathbf{u}}).$$

Упр. Сдвиг системы координат задан вектором $\mathbf{u} = (4, -5, 6)$. В старой системе координат точка P имеет координаты $P = (10, -4, 3)$. Какие координаты точки P имеет в новой системе координат?

* * *

Теперь обратно, зная новые координаты (x', y', z') хотим найти старые координаты (x, y, z) . Тут можно повторить те же рассуждения про радиус-векторы и получить, что

$$\overrightarrow{OP} = \overrightarrow{O'P} + \mathbf{u}.$$

А можно посмотреть на полученную формулу для координат:

$$(x', y', z') = (x, y, z) - \mathbf{u},$$

и просто выразить (x, y, z) :

$$(x, y, z) = (x', y', z') + \mathbf{u}.$$

Все просто.

6 Общий принцип

Отвлечемся от координатных формул и вникнем в суть. Для этого обозначим через T функцию сдвига точки на вектор \mathbf{u} :

$$T(P) = P + \mathbf{u},$$

а через T^{-1} – обратное действие, т.е. сдвиг точки на $-\mathbf{u}$:

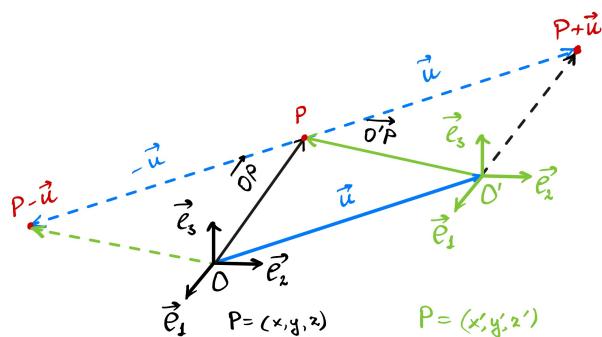
$$T^{-1}(P) = P - \mathbf{u}.$$

Тогда наши формулы преобразования координат запишутся вот как:

$$\begin{aligned} (x', y', z') &= T^{-1}(x, y, z), \\ (x, y, z) &= T(x', y', z'). \end{aligned}$$

Переведем на русский язык: для *координат* точки P сдвинуть систему координат на вектор \mathbf{u} это то же самое, что сдвинуть саму точку P на $-\mathbf{u}$ в старой системе координат.

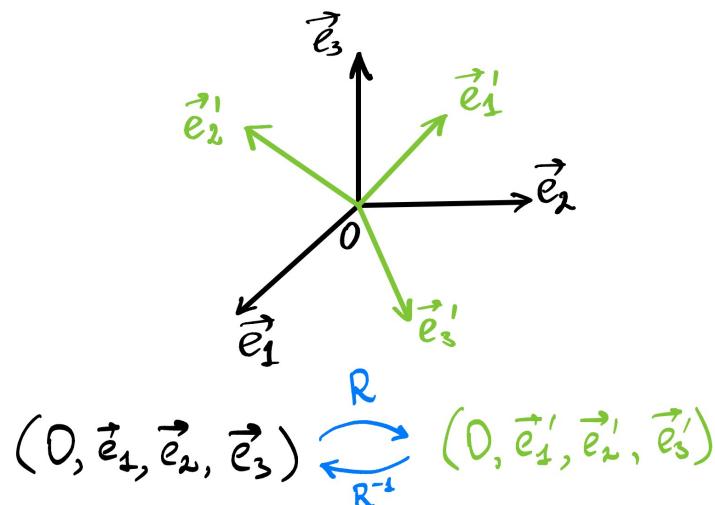
Если вы едете на поезде, вам все равно, поезд едет вперед, или земля под ним – назад.



7 Поворот системы координат

Второе преобразование системы координат, которое постоянно используется, это поворот.

Снова пусть $(O, \vec{e}_1, \vec{e}_2, \vec{e}_3)$ – исходная система координат, в которой задана точка $P = (x, y, z)$. Для поворота системы координат нужно повернуть базис, а начало координат оставить на месте, поэтому новая система координат будет иметь вид $(O, \vec{e}'_1, \vec{e}'_2, \vec{e}'_3)$.



Поставим тот же вопрос: как связаны новые (x', y', z') и старые (x, y, z) координаты точки P ?

Обозначим через R функцию поворота, которая поворачивает базисные векторы именно так, как нам нужно:

$$R(\vec{e}_1) = \vec{e}'_1,$$

$$R(\vec{e}_2) = \vec{e}'_2,$$

$$R(\vec{e}_3) = \vec{e}'_3.$$

А через R^{-1} обозначим обратный поворот, возвращающий базис на место.

Про функцию R стоит думать, как о функции движка `Rotate(...)`, о реализации которой мы ничего обычно не знаем, но понимаем, что она делает геометрически.

Теперь вользуемся той же логикой, что и со сдвигами: для координат точки P поворот исходной системы координат равносителен тому, что повернули саму точку в противоположном направлении вокруг точки O . Поэтому можно сразу сказать, что

$$(x', y', z') = R^{-1}(x, y, z),$$

и наоборот,

$$(x, y, z) = R(x', y', z').$$

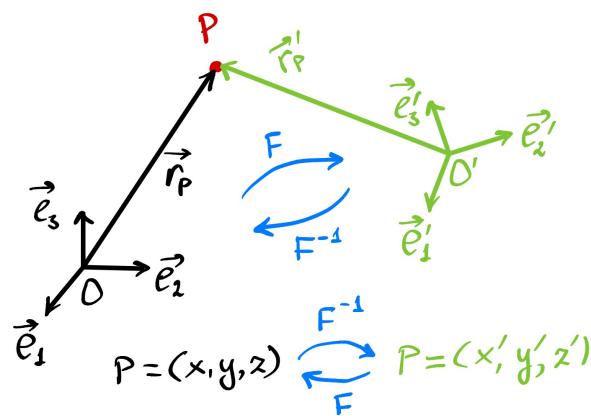
Если вам надо проехать на машине по кольцу, вам не важно, вы проедете сами, или кольцо повернется под вами в противоположном направлении.

8 Общая схема преобразований координат

Объединим поворот⁴ и сдвиг в одно преобразование, которое обозначим буквой F , под которым можно понимать привычный трансформ объекта.

Тогда продолжаем логику: для координаты точки P преобразование F исходной системы координат равносильно преобразованию F^{-1} самой точки P :

$$(x', y', z') = F^{-1}(x, y, z), \\ (x, y, z) = F(x', y', z').$$



⁴В общем случае – любое преобразование базиса.

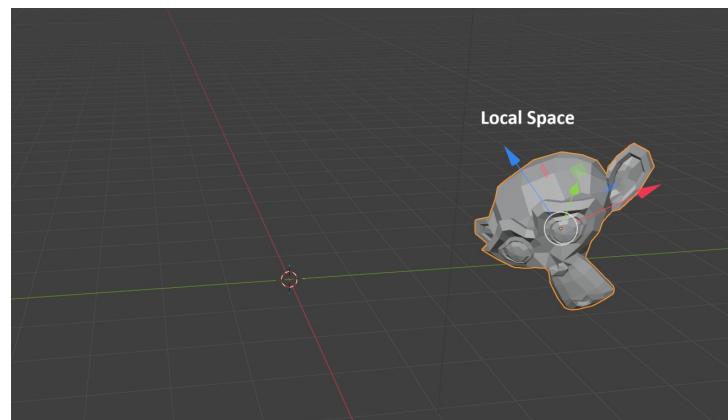
Трансформы на практике

В этом разделе мы рассмотрим практические задачи:

1. Переводить локальные трансформы в глобальные и наоборот.
2. Находить трансформ объекта относительно любого другого объекта.
3. Пересчитывать координаты точек/векторов относительно разных объектов.

Перед тем как разобрать эти ситуации нужно прояснить, что понимается под трансформами и системами координат на практике.

В любом 3D-редакторе создание геометрического объекта влечет создание его *локальной* системы координат, или *Local Space*, которое изображается тремя стрелочками, которые перемещаются и врачаются вместе с объектом (если выбран Local Space во вьюпорте). Стрелочки это базис, а точка, из которой они нарисованы – начало координат. Еще говорят, что система координат “вморожена” в объект.



Точки меша (геометрии объекта) заданы в локальной системе координат объекта, поэтому когда преобразовывается система координат, вместе с ней автоматически преобразовывается вся геометрия. Значит, для описания текущего положения точек геометрии достаточно знать только одно – преобразование F глобальной системы координат в локальную систему объекта. Это преобразование мы часто называем *трансформом объекта*.

Парентинг объекта B на объект A означает, что трансформ объекта B теперь переводит не глобальную систему координат в локальную систему объекта B , а локальную систему A в локальную систему B . Иначе это можно понимать так, что теперь для объекта B глобальную систему как бы подменили локальной системой A .

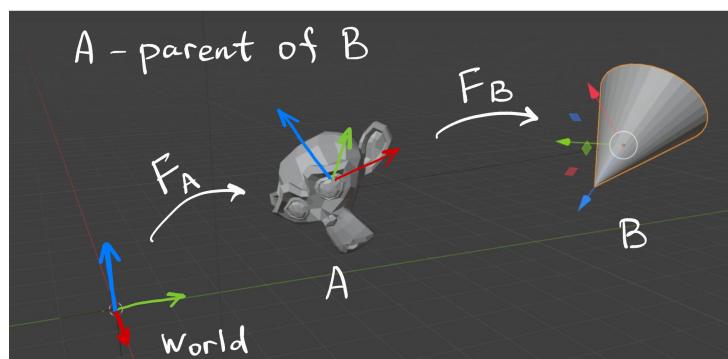


Рис. 11: Объект A является парентом объекта B .

Таким образом, если задан какой-то трансформ объекта, важно то, **относительно какого объекта он задан**. Другими словами, из какой системы координат задано преобразование к координатам данного объекта: из глобальной или из локальной системы какого-то другого объекта.

N.B. Принципиально, глобальная система координат (Global или World Space) отличается от локальных только тем, что ее программно запрещено парентить на другие системы координат. Но технически можно сымитировать новую "глобальную систему": создать объект и запарантить на него все остальные объекты.

N.B. Если говорится, что задан локальный трансформ, то всегда нужно понимать, локальный относительно какого объекта. Это должно быть либо указано явно, либо ясно из контекста. Обычно под локальным трансформом имеется в виду преобразование относительно родителя.

1 Из локальных в глобальные

Задача. Пусть объект A является парентом объекта B , трансформ объекта A обозначим через F_A , а трансформ объекта B относительно родителя через F_B . Найдем трансформ объекта B относительно World Space, который обозначим F_B^W .

Решение. Это просто: нужно применить сначала преобразование F_A , а затем F_B . Если умножаются матрицы или кватернионы, то обычно слева стоит множитель, который отвечает за последнее преобразование, а справа – за первое, то есть получается надо "умножать" F_B на F_A :

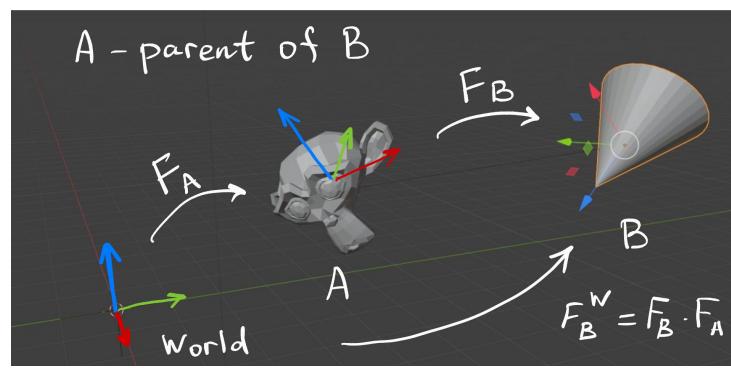
$$F_B^W = F_B \cdot F_A.$$

Не вдаваясь в детали, объясняется именно такой порядок следующим образом. Если помнить, что все F -ки это функции от векторов, то функция F_B^W должна действовать так:

$$F_B^W(\mathbf{v}) = F_B(F_A(\mathbf{v})),$$

сначала применили F_A , затем F_B .

Порядок умножения важен, так как надеть майку, а потом рубашку – не то же самое, что надеть рубашку, а потом майку.



2 Из глобальных в локальные

Задача. Пусть объект A является парентом объекта B , трансформ объекта A обозначим через F_A , а трансформ объекта B относительно World – через F_B^W . Найдем трансформ объекта B относительно Local Space (= Parent Space), который обозначим F_B .

Решение. Перефразируем задачу: нужно найти преобразование, переводящее систему координат объекта A в систему координат объекта B . Для этого просто идем по известным нам стрелочкам: нужно сделать преобразование F_A^{-1} и потом F_B^W , то есть искомое преобразование:

$$F_B = F_B^W \cdot F_A^{-1}.$$

3 Из локальных в локальные

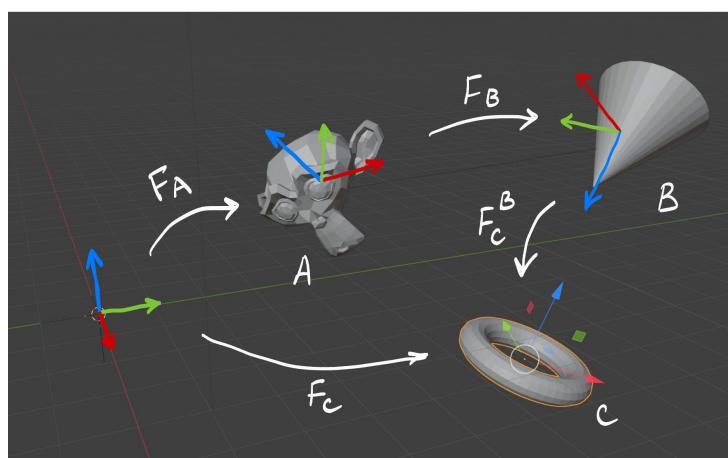
Добавим в схему еще объект C . Обозначим его глобальный трансформ через F_C , а локальный относительно объекта B через F_C^B .

Допустим известен локальный трансформ F_C^B , а надо найти глобальный F_C . Найдим:

$$F_C = F_C^B \cdot F_B \cdot F_A.$$

А теперь обратная ситуация: известен глобальный трансформ F_C , а надо найти локальный F_C^B . Тогда, обращая стрелочки, получаем

$$F_C^B = F_C \cdot F_A^{-1} \cdot F_B^{-1}.$$



4 Нахождение координат точки или вектора

Задача. Те же три объекта. Допустим в системе координат объекта B задана точка P координатами (x^B, y^B, z^B) . Надо найти координаты (x^C, y^C, z^C) точки P относительно системы координат объекта C .

Решение. Мы знаем, что

$$(x^C, y^C, z^C) = (F_C^B)^{-1}(x^B, y^B, z^B),$$

где F_C^B – трансформ C относительно B , который уже найден:

$$F_C^B = F_C \cdot F_A^{-1} \cdot F_B^{-1}.$$

Осталось найти его обратный, т.е. по сути F_B^C – трансформ B относительно C . Можно снова идти по стрелочкам⁵:

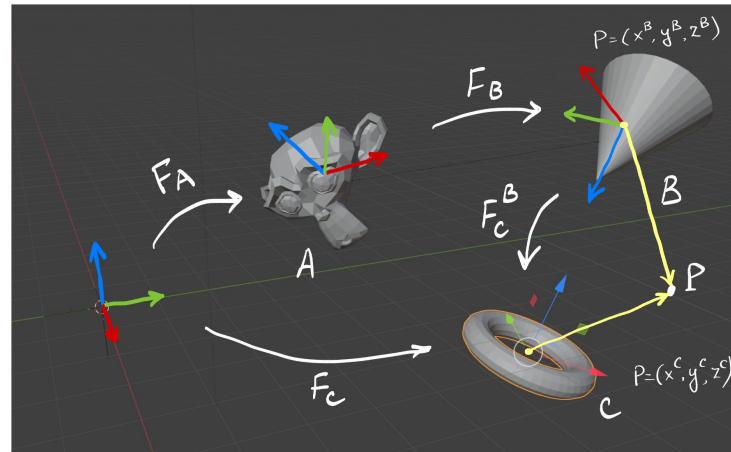
$$(F_C^B)^{-1} = F_B^C = F_B \cdot F_A \cdot F_C^{-1}.$$

Все готово.

⁵Либо воспользоваться правилом $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.

Чтобы найти координаты точки P относительно заданной системы координат, нужно найти преобразование F из исходной системы координат в заданную, и применить обратное к нему преобразование F^{-1} к исходным координатам. ^a

^aЛибо сразу искать преобразование из заданной в исходную, тогда не придется инвертировать трансформ. Это все равно, результат будет тот же. Главное понять, что происходит.



Матрицы

Morpheus: *The Matrix is everywhere. It is all around us. Even now, in this very room. You can see it when you look out your window or when you turn on your television. You can feel it when you go to work... when you go to church... when you pay your taxes. It is the world that has been pulled over your eyes to blind you from the truth.*

Neo: *What truth?*

Morpheus: *That you are a slave, Neo. Like everyone else you were born into bondage. Into a prison that you cannot taste or see or touch. A prison for your mind.*

— The Matrix

Этот раздел мы включили для демонстрации общего представления о том, что такое матрицы. Дело в том, что матрицы проникли во множество вычислительных алгоритмов. Попробую объяснить, почему.

Дело в том, что движение объектов описывается как правило сложными нелинейными функциями и уравнениями. Основная идея математического анализа состоит в том, что в малых интервалах нелинейные процессы можно очень точно приблизить линейными. В частности, дифференциал функции это ее линейная часть, или иначе, линейное приближение.

Другими словами, если вы нарисуете график любой (непрерывной) функции, а затем очень сильно его приблизите, то увидете просто прямую. В этом смысле равенство $\sin x = x$ верно в окрестности⁶ точки $x = 0$. В этом смысле говорят, что $f(x) = x$ — линейная составляющая синуса или дифференциал функции $\sin x$ в точке $x = 0$.

Не углубляясь далее, просто скажем, что основная идея анализа функций сводится к изучению линейных операций. А основным языком изучения таких операций является язык матриц.

1 Что такое матрица?

Матрицей размера $n \times m$ называется таблица чисел из n строк и m столбцов.

2 Умножение матрицы на матрицу

3 Умножение матрицы на вектор

4 Матрица преобразования базиса

5 Матрица масштабирования

6 Матрица поворота 2x2

7 Матрица поворота 3x3

⁶“Бесконечно малой” окрестности.

Углы Эйлера

Пришло время взглянуться внимательнее в вопрос описания поворотов в трехмерном пространстве. К этой задаче можно подходить с разных сторон: с практической стороны для механических поворотов твердых тел, с теоретической стороны для алгебраического и геометрического описания вращений.

1 Зачем?

Вспомним, что нам уже хорошо известно из рассказанного на курсе.

Мы поняли, что любое преобразование системы координат в аффинном пространстве (место совместного обитания точек и векторов) раскладывается на два действия: сдвиг и поворот базиса. Как задавать сдвиги – ясно, с помощью векторов. Для определения поворота у нас тоже есть инструмент – матрица.

Матрица описывает поворот хитрым образом, совсем не используя понятия угла. Дело в том, что матрицами задаются вообще все преобразования базиса, а не только поворот. Делается это следующим образом: по столбцам матрицы записываются координаты новых базисных векторов (относительно старого базиса).

Теоретически, матрица – очень мощный инструмент для анализа в том числе вращений. Но при этом есть ощущение, что использовать космический шаттл для поездки в магазин – идея не самая лучшая, а где-то даже вредная.

Интуитивно, поворот должен описываться в терминах углов, а не координат векторов. Углы Эйлера делают именно это: тремя углами позволяют повернуть тело в любое положение.

2 Что такое углы Эйлера?

Углами Эйлера называются три числа (φ, ψ, θ) , которые задают **алгоритм**, по которому можно тремя **последовательными** поворотами вокруг **некоторых осей** из старого базиса получить новый заданный базис.

Н.В. В этой части мы будем говорить только об ортонормированных базисах $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$. Давайте здесь обозначать системы координат $(O, \mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ через $Oxyz$, а направления векторов $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ будем называть осями x, y, z . Эйлер про аффинные пространства и базисы еще не знал :)

Данное определение хоть и неконкретное, но ёмкое. Конкетику мы обсудим очень скоро. Дело в том, что углов Эйлера бывает много разных.

Обратим внимание на выделенные жирным слова в определении углов Эйлера. Их смысл будет раскрываться в нашем рассказе.

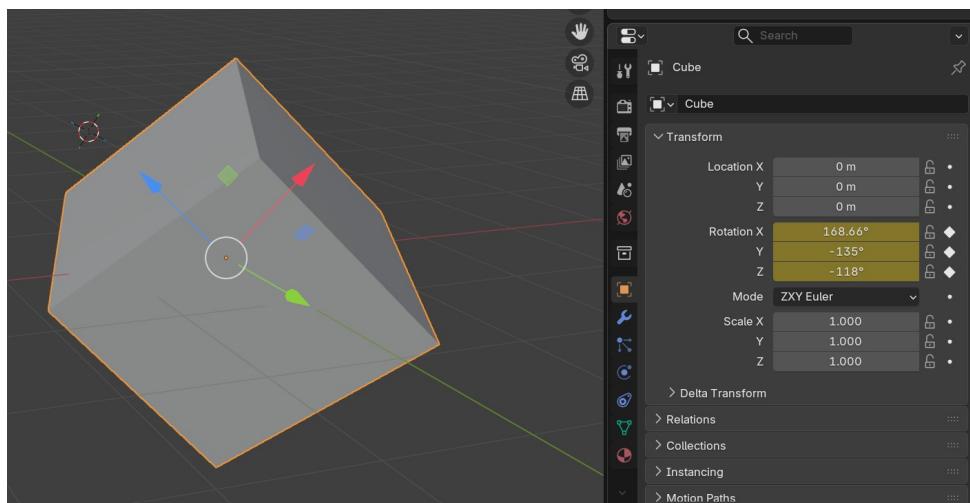
Во-первых, углы Эйлера задают алгоритм, т.е. инструкцию "делай раз, делай два, делай три". Цель этих действий – перевести тело в заданное положение.

Во-вторых, последовательность поворотов меняет инструкцию. Поэтому, если речь идет об углах Эйлера, нужно всегда понимать, какой порядок поворотов подразумевается. Символично это обычно записывается как XYZ , ZXY , ZXZ и т.п. Что означают эти аббревиатуры, мы скажем отдельно.

В-третьих, все три поворота выполняются вокруг осей. Важно понимать, что это оси не глобальные и не локальные. Оси вращения опять же зависят от типа углов Эйлера: XYZ и т.д. Как определяются оси, мы также обсудим подробно.

3 Углы Эйлера в 3D-редакторе

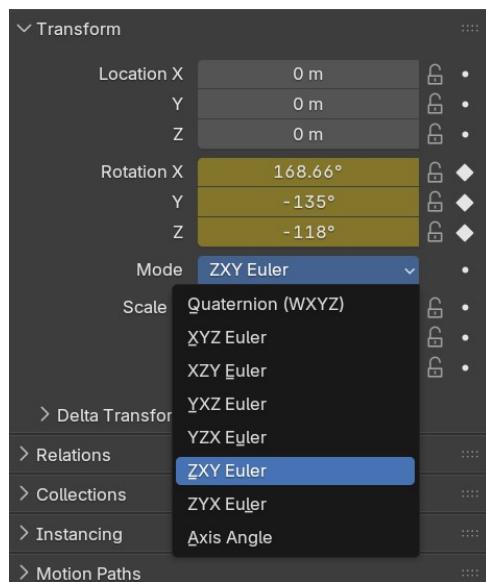
Посмотрим, как работают углы Эйлера в 3D-редакторах. Судя по наблюдениям, во всех редакторах эта функция работает одинаково.



Итак, в редакторе имеется панелька для заданием трансформа, где можно настраивать Location, Rotation, Scale. Сегодня нас интересуют вращения.

Для задания поворота объекта реализовано несколько способов:

1. Кватернионы
2. Углы Эйлера в шести вариациях
3. Поворот вокруг заданной оси



Углы Эйлера являются наиболее понятным и удобным для обычного пользователя вариантом вращения объекта, так как они задают вращение относительно локальных координатных осей объекта. При этом поворот задается обычными углами в градусах или радианах – все просто.

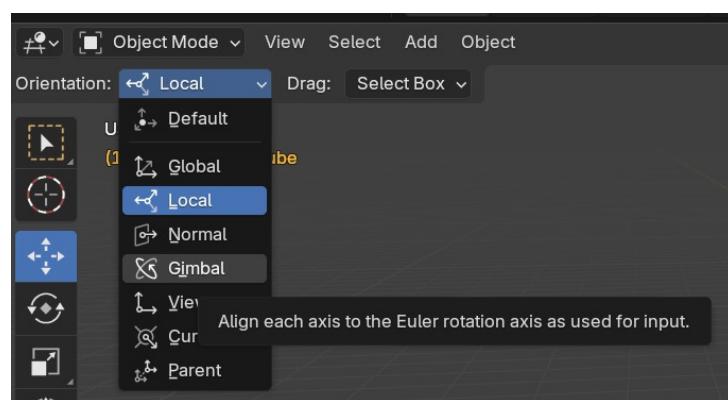
* * *

Обозначение ZXY указывает на порядок применения вращений: сначала вокруг оси Y , затем X , и наконец Z . То бишь сперва налево. По крайней мере такая нотация используется в Blender и Maya.

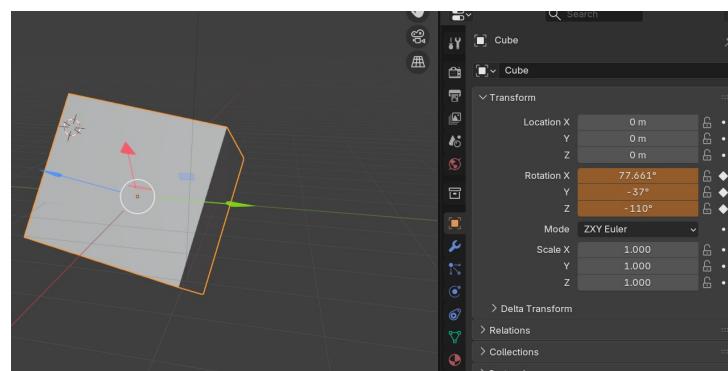
Проверить легко: включите отображение локальных осей и поверните объект в рандомное положение (например, как на скриншоте), а затем подергайте ползунок первой оси (в нашем случае оси Z). Если объект вращается строго вокруг соответствующей локальной оси, значит, вращение вокруг нее происходит последним.

* * *

Если же менять ползунки, соответствующие первому и второму повороту, то можно заметить, что объект будет вращаться совсем не всегда вокруг указанных локальных осей. Чтобы увидеть те оси, вокруг которых будет происходить вращение, в редакторах есть специальный режим отображения осей под названием Gimbal:



Тогда во виджопорте будут видны оси, которые совсем не обязательно будут ортогональными:



На приведенном скриншоте оси Z и Y образуют угол, близкий к 180 градусам.

В режиме Gimbal отображаются не локальные оси объекта, а те оси, вокруг которых происходит вращение при изменении соответствующего угла Эйлера.

В самой экстремальной ситуации две оси из трех лягут на одну прямую, и тогда эти две оси будут производить одно и то же вращение. Эта ситуация называется Gimbal Lock. Мы разберем все это подробно отдельно.

4 Как определяются углы Эйлера для конкретного поворота?

5 Вокруг каких осей происходит вращение?

Для того, чтобы понять, откуда берутся эти оси в режиме Gimbal, нужно просто пошагово нарисовать тот алгоритм, который дают углы Эйлера. Предположим, что мы находимся в порядке ZXY и заданы углы Эйлера (φ, ψ, θ) .

Шаг 1. Повернуть исходные оси $Oxyz$ вокруг оси y на угол ψ . Ось y останется на месте, а две другие оси повернутся: $x \mapsto x'$, $z \mapsto z'$. Получим новые оси $Ox'y'z'$.

Шаг 2. Повернуть исходные оси $Ox'y'z'$ вокруг оси x' на угол φ . Ось x' останется на месте, а две другие оси повернутся: $y \mapsto y''$, $z' \mapsto z''$. Получим новые оси $Ox'y''z''$.

Шаг 3. Повернуть исходные оси $Ox'y''z''$ вокруг оси z'' на угол θ . Ось z'' останется на месте, а две другие оси повернутся: $x' \mapsto x'''$, $y'' \mapsto y'''$. Получим новые оси $Ox'''y'''z''$.

Итого, три шага:

$$Oxyz \mapsto Ox'y'z' \mapsto Ox'y''z'' \mapsto Ox'''y'''z''.$$

В режиме Gimbal отображаются те оси, вокруг которых происходило вращение. В нашем случае это оси y, x', z'' . Аналогично для любого другого порядка углов Эйлера.

6 Gimbal Lock

7 Проблема углов Эйлера

8 Эйлер придумал фигню?

9 Как минимизировать проблему?

10 Какая связь с матрицами?

11 Углы Эйлера как координаты в группе вращений SO_3

Откуда взялись кватернионы, и вообще?

*Минус на минус – всегда только плюс
Отчего так бывает, сказать не берусь.*

— У.Г. Оден

Мы уже знакомы с двумя способами задавать вращения пространства и, в частности, твердого тела, матрицами и углами Эйлера. Наконец, пришла пора поговорить о третьем способе – кватернионах.

Можно заметить, что материалы о кватернионах с трудом воспринимаются большинством интересующихся этой темой людей. И в лучшем случае человек что-то примерно поймет, чего ему хватит для работы.

Но думать, что студентам-математикам в этом вопросе сильно проще, ошибочно. По моему опыту, о кватернионах говорится совсем немного лишь на лекциях по алгебре, причем без погружения в их геометрические свойства. В стандартных учебниках и курсах ничего интересного нам о кватернионах просто не пишут. Так что, думаю, проблема с материалами для прикладников может быть связана еще и с тем, что сами математики в подавляющей массе толком не разбираются в этой теме.

Еще одна причина, по которой кватернионы используются как правило без сколько-нибудь понимания того, как они работают, банальна: кватернионы – это сложная штука, абстрактная алгебраическая структура, которая родилась в контексте другой абстрактной алгебраической структуры – комплексных чисел, для описания трехмерной геометрии!

Кватернионы напрашиваются на то, чтобы изучать их в историческом контексте. В 19 веке в математике и физике произошло так много всего, и так много оказалось переплетено, что совсем в отрыве от истории теряется очень многое.

О кватернионах нам потребуется хорошо понять две вещи:

1. Почему кватернионы это числа, и откуда они появились?
2. Как кватернионы задают вращение и какое именно?

Поэтому мы разделим рассказ о кватернионах на две части, посвященные этим вопросам.

Действительные числа

Числа, которые для нас максимально привычны, называются *действительными*. Множество всех таких чисел обычно представляют либо в виде числовой прямой, каждой точке которой соответствует некоторое число, либо в виде десятичной дроби. Многие люди даже не подозревают, что есть какие-то еще числа.

Для тех, кто со школы забыл слово "действительные", заметим, что с точки зрения программирования, аналогом действительных чисел выступает тип **float**, который позволяет записывать десятичную дробь с большой точностью.

Цель данной секции – увидеть, что действительные числа не такие простые, как могут казаться по привычке. Да, интуитивно все просто. Но если всмотреться, картина оказывается совсем не плоской. Понимание этого настроит нас на более адекватное восприятие и кватернионов, с которыми эта глубина гораздо более явно выражена.

Для этого быстро пройдемся по подмножествам множества всех действительных чисел и проследим логику, которая делает неизбежным их выделение.

1 Ясли (натуральные числа)

Вспомним поступательно, какие числа нам известны, и начнем с натуральных чисел.

$$\mathbb{N} = \{1, 2, 3, 4, 5, \dots, 100, 101, \dots, 10^{100}, \dots\}$$

Натуральными эти числа называются из-за того, что они наиболее естественные (natural). Есть мнение, что математика это суть наука об абстракциях, а самая базовая абстракция – натуральные числа.

Греки говорили: "все есть число", и понимали числа как описание вещества. Смотря на натуральные числа с их точки зрения, единица означает нечто целое, а произвольное число n – количество единиц.

[2] Детский сад (нуль и дроби)

[3] Первый класс (отрицательные числа)

* * *

«Я знаю людей, которые никак не могут понять, что если из нуля вычесть четыре, то будет нуль.»

— Паскаль —

* * *

«Меньшее число не может относиться к большему, как большее к меньшему.»

— Арно (1612-1697) об $-1 : 1 = 1 : -1$ ("Мысли") —

* * *

«Алгебраические правила действий над отрицательными числами ныне общеприняты, и все признают их точными независимо от того, что бы мы не думали о природе этих чисел.»

— Д'Аламбер ("Отрицательное") —

* * *

«Вычесть $(-b)$ это все равно, что прибавить b , так как погасить долг означает поднести дар.»

— Эйлер —

* * *

В современных обозначениях, совокупность натуральных чисел, взятых со знаками плюс и минус, и нуля называется множеством целых чисел \mathbb{Z} (от нем. Zahlen – "числа"):

$$\mathbb{Z} = \{\dots, -100, \dots, -2, -1, 0, 1, 2, \dots, 100, \dots\}.$$

Раз выделили целые числа, то в противовес определяются и дробные, которые представляют из себя отношение двух целых чисел (ratio или quotient). Множество таких чисел называется множеством рациональных чисел и обозначается \mathbb{Q} :

$$\mathbb{Q} = \left\{ \frac{1}{2}, \frac{4}{3}, -\frac{100}{13}, \dots \right\} = \left\{ \frac{m}{n} \mid m, n \in \mathbb{Z}, n \neq 0 \right\}.$$

4 Средняя школа (иррациональные числа)

Древние греки доказали, что длина диагонали квадрата со стороной равной единице, не представляется в виде рациональной дроби.

Согласно теореме Пифагора, длина такой диагонали равна

$$\sqrt{1^2 + 1^2} = \sqrt{2}$$

Легко доказать от противного, что нет целых чисел m, n , которые давали бы дробь

$$\frac{m}{n} = \sqrt{2}.$$

▷ TODO

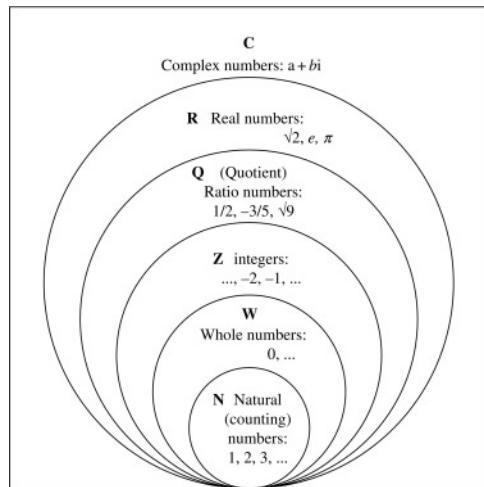
До этого момента считалось, что кроме рациональных чисел больше ничего нет. Интуитивно это вполне понятно: деля на все более мелкие кусочки целое, можно получить получить любую точность количества вещества. Однако, интуиция подвела, это оказалось неверно. Нелогичные, а иначе, иррациональные числа существуют, и получить такое число оказалось очень просто.

О слове "существует" существует непрекращающийся философский спор: $\sqrt{2}$, конечно, не рациональное число, но может все-таки тогда его и числом нельзя называть? Мейнстрим сошелся на том, что все же если понимать под $\sqrt{2}$ длину диагонали квадрата, то логично считать его числом. Причем, хоть $\sqrt{2}$ абсолютно непонятно, как вычислять, дураку очевидно (например, приложив метр), что это число больше единицы, но меньше двойки.

5 Действительные или вещественные числа

Таким образом, возникло разделение действительных чисел \mathbb{R} на множество рациональных и иррациональных чисел. Специального обозначения для иррациональных чисел нет. А обозначение \mathbb{R} для множества действительных чисел объясняется просто: от англ. **Real Numbers**. Можете догадаться, что должны быть и **Irreal Numbers**. Такие и правда есть, только называются они комплексными числами, о них – в следующем разделе.

Еще действительные числа называют *вещественными*, потому что этими числами измеряется длина, площадь, скорость, температура, вес и т.д. То есть параметры вещества.



6 Старшая школа (трансцендентные числа)

Время вспомнить два иррациональных числа, которые имеют свои личные обозначения: π и e .

Число π равно отношению длины окружности к ее диаметру. Можно долго разглагольствовать, но, говоря коротко, это число каким-то образом вшито в суть геометрии окружности. Поэтому всюду, где есть тригонометрия, нас преследует число π .

Про число e мы говорили на в разделе о функциях, когда обсуждали, что среди всех показательных функций a^x каким-то образом выделяется одна e^x , которую называли просто экспонентой. Если про число π можно сказать четко, что оно связано с окружностями и колебательными движениями, то число e ведет себя несколько иначе. Оно не имеет такого геометрического характера, зато возникает как снег на голову то тут, то там.

Не углубляясь в детали, скажу по опыту, что число e часто появляется в описании траекторий движений (из-за дифференциальных свойств экспоненты), и в вероятностных задачах.

Например, есть такая забавная задача про русскую рулетку. В комнате N человек, у каждого заряженный револьвер. По звонку каждый стреляет в случайного человека. Какой шанс выжить?

Оказывается, что если количество людей N большое, то этот шанс приближается к $\frac{1}{e}$. То есть приблизительно 37%.

* * *

Числа π и e вызывают ощущение совсем не обычных чисел. Обычное число означает количество, длину и т.д. А эти числа как будто несут какую-то свою информацию, можно сказать информацию о природе.

Но это мистическая сторона дела, которая объясняет термин "трансцендентный". Конечно, математики так легко словами не разбрасываются, чтобы каким-то двум числам присвоить особый статус по сути за то, что они непонятные.

Строго говоря, трансцендентными называются такие числа, которые не являются корнями многочлена с целыми коэффициентами.

Например, $x = 1$ является корнем уравнения

$$x - 1 = 0.$$

Дальше, отрицательное число $x = -5$ является корнем уравнения

$$x + 5 = 0.$$

Любое дробное число тоже является корнем такого уравнения, например, $x = \frac{4}{9}$ является корнем

$$x - \frac{4}{9} = 0.$$

И квадратные корни из натуральных чисел нас не пугают. Тот же $\sqrt{2}$ является корнем уравнения

$$x^2 = 2.$$

А вот числа π и e таким образом не получить, о чем говорит теорема Линденмана, которая доказывается довольно сложно.

Комплексные числа

1 Что это такое?

Комплексные числа (Complex numbers) возникли очень давно, причем также из решений уравнений. Индийцам и арабам помогло здесь то, что они не придавали числам такого вещественного значения, как древние греки, и развивали формальные методы решения уравнений без лишних сдерживающих воображение мыслей.

Так были разработаны методы решения квадратных уравнений. Из школы известен железобетонный метод – через дискриминант.

Решения уравнения $ax^2 + bx + c = 0$ находятся из формулы

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

И тогда есть два случая, когда подкоренное выражение, то бишь дискриминант

$$D = b^2 - 4ac$$

больше или равен нулю, либо меньше нуля. В школе считалось, что если дискриминант меньше нуля, то решений нет.

Например, квадратное уравнение

$$x^2 + 1 = 0,$$

таким образом, не имеет решения, ведь решение должно быть таким, что $x^2 = -1$, т.е. $x = \sqrt{-1}$.

Говоря простыми словами, для решения таких уравнений нам не хватает умения каким-то образом извлекать квадратный корень из -1 . Среди действительных чисел действительно такого быть не может. Ведь любое действительное число в квадрате дает положительное число, а это загадочное $\sqrt{-1}$ в квадрате дает отрицательное число $(\sqrt{-1})^2 = -1$.

* * *

При этом попытки все же как-то работать с такими непонятными формулами, которые не имеют понятного геометрического и вещественного значения, оказалось полезно. См. например сюжет про историю формулы Кардано решения кубических уравнений. Мы не имеем времени погружаться в эту тему, поэтому пойдем дальше.

* * *

«Квадратные корни из отрицательных чисел не равны 0, не меньше 0 и не больше 0. Отсюда ясно, что квадратные корни из отрицательных чисел не могут находиться среди возможных (вещественных) чисел. Следовательно, нам не остается ничего другого, как признать их невозможными числами. Это приводит нас к понятию чисел, по своей природе невозможных и обычно называемых мнимыми или воображаемыми, потому что они существуют только в нашем воображении.»

— Эйлер —

* * *

«Дух божий нашел тончайшую отдушину в этом чуде анализа, уроде из мира идей, двойственной сущности, находящейся между бытием и небытием, которую мы называем мнимым корнем из отрицательной единицы.»

— Лейбниц —

2 Как определяются алгебраически

Комплексным числом называется запись

$$a + bi,$$

где a и b – обычные действительные числа, а символ i называется *мнимой единицей* (от англ. *imaginary*) для которого определено правило умножения на себя таким образом:

$$i \cdot i = i^2 = -1.$$

Сложение и умножение комплексных чисел определяется тем же способом, как для обычных чисел.

Пример. Возьмем два комплексных числа $z = 1 + 3i$, $w = -4 + 7i$. Складывая их, получаем

$$z + w = 1 + 3i - 4 + 7i = -3 + 10i,$$

а перемножая,

$$z \cdot w = (1 + 3i)(-4 + 7i) = -4 + 7i - 12i + 21i^2 = -4 - 5i - 21 = -25 - 5i.$$

N.B. Множество комплексных чисел является расширением множества действительных чисел. Это значит, что любое действительное число является комплексным: если a – действительное число, то оно имеет комплексное представление $a + 0i$.

Давайте сделаем важные выводы:

1. комплексные числа – чисто алгебраическая придумка, которая позволяет играться с формулами даже тогда, когда они перестают иметь значение из-за извлечения корней из отрицательных чисел
2. комплексные числа называются числами, потому что во-первых, они обобщают понятие обычного действительного числа, во-вторых, над ними все также можно производить те же операции сложения и умножения (вспомните лекцию о векторах и их умножении)
3. априори, совершенно не понятно, почему эта выдуманная конструкция может быть где-либо полезна, помимо развития фантазии

3 Геометрическая интерпретация

Не смотря на то, что о комплексных числах знали очень давно, споры о них ходили до 19 века. Даже Эйлер, который владел комплексными числами, относился к ним с той же чисто формальной точки зрения.

Перевернуло отношение к комплексным числам открытие, что на комплексное число $x + iy$ можно смотреть как на точку на плоскости с координатами (x, y) .

Тогда сложение таких точек (вообще, векторов, но в начале 19 века такого слова еще не было) с координатами $z = (x_1, y_1)$ и $w = (x_2, y_2)$ на языке комплексных чисел будет:

$$z = x_1 + iy_1 \quad \text{и} \quad w = x_2 + iy_2$$

чье сложение дает число

$$z + w = (x_1 + x_2) + i(y_1 + y_2),$$

то есть точку с координатами $(x_1 + x_2, y_1 + y_2)$. Буквально то, что мы делали с векторами.

Таким образом, на комплексное число можно смотреть по сути как на иное представление вектора.

[4] Тригонометрическая форма

[5] Комплексная экспонента и поворот

Sir W.R.Hamilton и кватернионы

[1] Немного истории

[2] Определение кватернионов

[3] Вектор и умножение кватернионов

Геометрия кватернионов

[1] С алгеброй понятно, а что с геометрией?

[2] И что с этим делать?

[3] Кватернионы и повороты трехмерного пространства

[4] Что еще нужно знать?

Список литературы

- [1] Cayley A. *A Memoir on the Theory of Matrices*. 1857. URL: <https://scispace.com/pdf/a-memoir-on-the-theory-of-matrices-14gzduwllk.pdf>.
- [2] Gibbs J.W. *Elements of Vector Analysis: Arranged for the use of Students in Physics*. 1881-4.
- [3] Crowe M.J. *A History of Vector Analysis: The Evolution of the Idea of a Vectorial System*. Notre Dame: Univercity of Notre Dame Press, 1967.
- [4] Regiomontanus. *De triangulis omnimodis*. 1464.
- [5] Юшкевич А.П. *История математики с древнейших времен до начала XIX столетия, в трех томах, том 1*. Наука, 1970.
- [6] Ньютона И. *Математические работы*. Объединенное научно-техническое издательство НКТИ СССР, главная редакция технико-теоретической литературы (1937), 1637.
- [7] Птолемей К. *Руководство по географии*. Государственное издательство географической литературы, Москва, 1953.
- [8] Эйлер Л. *Введение в анализ бесконечных, Том 1,2*. Государственное издательство физико-математической литературы, Москва (1961), 1748.
- [9] Дирихле П.Г. *Лекции по теории чисел*. Государственное издательство географической литературы, Москва, 1953.
- [10] Декарт Р. *Рассуждение о методе*. Академия наук СССР (1953), 1637.
- [11] Декарт Р. и Ферма П. *"Геометрия" с приложением избранных работ П. Ферма и переписки Декарта*. Государственное объединенное научно-техническое издательство НКТП СССР, редакция технико-теоретической литературы (1938), 1637.