Математика и физика для экономистов

И.Г.Царев

Оглавление

Введение **i**

I Анализ категорий **1**

1 Предварительные сведения **3**
1.1 Множества 4
1.2 Подмножества (фигуры) в пространстве 10
1.3 Отображения 15
1.4 Координаты пространства 21
1.5 Элементарные функции 24
1.6 Функции и непрерывность 29
1.7 Комплексные числа и функции 36

2 Линейная алгебра **45**
2.1 Векторная алгебра 45
2.2 Базисное представление векторов 51
2.3 Матрица и определитель 56
2.4 Евклидова и эрмитова структуры 69
2.5 Преобразование пространства 76
2.6 Многомерные параллелепипеды 84
2.7 Внешние алгебраические формы 89
2.8 Тензоры 97

3 Дифференцирование **101**
3.1 Дифференцирование скалярной функции 105
3.2 Различные виды производных 111
3.3 Дифференциальные формы 119
3.4 Многообразие 130
3.5 Производная по направлению 138
3.6 Дифференциальная геометрия 144
3.7 Преобразования производных 154

4 Интегрирование **163**
4.1 Интеграл . . . 164
4.2 Теорема Стокса . . . 171
4.3 Интегрирование в E^3 . . . 177
4.4 Несобственные интегралы . . . 184
4.5 Скалярный и векторный потенциалы . . . 187
4.6 Дельта-функция Дирака (δ-функция) . . . 193

5 Фазовое пространство **199**
5.1 Однопараметрический диффеоморфизм . . . 200
5.2 Общее решение линейного уравнения . . . 207
5.3 Первые интегралы . . . 216
5.4 Симплектическое многообразие . . . 219
5.5 Производящая функция . . . 227
5.6 Функциональный анализ . . . 232
5.7 Теорема Лиувилля–Арнольда . . . 236

6 Функция вероятности **243**
6.1 Предварительные определения . . . 243
6.2 Операции с вероятностями . . . 245
6.3 Биноминальное распределение . . . 248
6.4 Случайная величина . . . 251
6.5 Непрерывно распределенные величины . . . 254

II Анализ систем **257**

7 Механическая система **261**
7.1 Механика Гамильтона и Лагранжа . . . 261
7.2 Механика Ньютона . . . 266
7.3 Вращение механической системы . . . 271
7.4 Принцип Д'Аламбера . . . 277
7.5 Механические инварианты . . . 281
7.6 Релятивистская механика . . . 283

8 Электродинамическая система **295**
8.1 Первоначальные определения . . . 295
8.2 Потенциалы поля . . . 302
8.3 Уравнения Лагранжа в электродинамике . . . 304
8.4 Уравнения Максвелла . . . 309
8.5 Энергия электромагнитного поля . . . 314
8.6 Стационарное поле . . . 318
8.7 Граница электромагнитных сред . . . 322

9 Термодинамическая система **325**
9.1 Представляющий ансамбль 325
9.2 Термодинамические параметры 328
9.3 Термодинамические потенциалы 333
9.4 Законы термодинамики 338
9.5 Принцип Ле-Шателье–Брауна 345
9.6 Сложная и открытая системы 349
9.7 Фазовые переходы 358
9.8 Кривая фазового перехода 363

10 Экономическая система **371**
10.1 Экономические термины 371
10.2 Аналогии экономической системы 379
10.3 Функция полезности 387
10.4 Производственная функция 391
10.5 Формула Блэка–Шоулза–Мертона 398
10.6 Распределение дохода 402

Заключение **409**

Введение

При написании этой книги во главу угла ставился прикладной аспект. Книга не претендует на математическую строгость, ее назначение, наряду с упрощенным изложением необходимых математических фактов, продемонстрировать начинающему исследователю некоторые «наглядные» и «интуитивные» методы, которые можно применять при исследовании физических, а также экономических явлений.

Под наглядными математическими методами понимаются, в данном случае, геометрические построения. Используемые математические объекты также рассматриваются в основном как геометрические объекты (фигуры) или инварианты inv. *Инвариант* (от лат. invarians, род. падеж invariantis — неизменяющийся) — термин, обозначающий нечто неизменяемое. В геометрии инвариант означает независимость объекта от выбранной системы координат в пространстве, а в физике — неизменность некоторой величины (энергии, импульса, момента импульса) со временем в каком-либо физическом процессе или независимость некоторой величины (скорости света) от условий наблюдения.

Геометрические представления естественно приводят к развитию определенных понятий (пространство, вектор, многообразие, расслоенное пространство и т.д.) и определенного математического аппарата (векторный анализ, анализ дифференциальных форм и т.д.).

Этот подход, по мнению автора, облегчает понимание математического аппарата. Многие современные математические теории возникли из проблем геометрии и механики и лишь впоследствии приняли тот аксиоматически-абстрактный вид, который так затрудняет их изучение. Поэтому приводимые определения могут показаться наивными, по сравнению с абстрактными конструкциями современной математики, основанной на теории множеств. Абстрактные определения возникают при попытках обобщить «наивные» понятия, сохраняя их основные свойства. Однако установлено, что эти попытки не приводят к реальному расширению круга рассматриваемых объектов и значит нецелесообразны.

В связи с тем, что многие определения математики сформулированы достаточно абстрактно, перед каждой главой проводится краткий обзор её содержания, с целью создать у читателя наглядное и связное представление об излагаемом материале.

Математический аппарат строится в настоящей книге с самого начала,

так что у читателя не предполагается предварительных знаний и использования дополнительной литературы. Все необходимые понятия и определения приводятся в тексте. Аналогичный подход использован также и во второй части книги, посвящённой физике и экономике.

Эксплуатация тесной взаимосвязи между геометрическими и аналитическими понятиями и идеями позволяет показать, что свойства различных систем в целом основаны на геометрических свойствах пространства, и тем самым продемонстрировать прямую связь между различными физическими дисциплинами, например, такими как механика, электродинамика и термодинамика.

С другой стороны, в используемом математическом аппарате отсутствуют какие-либо изначально постулированные ограничения и требования. Это позволяет предположить, что законы природы, установленные в физике, имеют общий характер, и, следовательно, проводя аналогии между некоторыми физическими и экономическими понятиями, показать, как некоторые аналитические методы дифференциально-геометрического происхождения, активно используемые в теоретической физике, применимы также и в экономической науке.

Книга состоит из двух частей, разбитых на главы и параграфы. В первой части речь в основном идет о математике, во второй части — о физике и экономике.

Данная книга не претендует на сколько-нибудь последовательное и полное изложение затронутых научных дисциплин: математики — в первой части книги; механики, электродинамики, термодинамики и экономики — во второй. Цель книги — именно демонстрация прикладного использования геометрического аппарата для описания различных систем, как физических, так и экономических, а также аналогий, которые при этом прослеживаются.

В последней главе, посвящённой экономике, имеются параграфы обладающие определённой научной новизной. Нетерпеливый читатель может сразу перейти к её изучению, так как в тексте имеются все необходимые ссылки на используемый математический аппарат, и предыдущие главы могут быть использованы в качестве справочника.

Приводимые в книге выводы формул и доказательства теорем преследуют цель продемонстрировать читателю применение используемого математического аппарата. Таким образом, доказательства служат в качестве примеров. Именно этот критерий, а не важность доказываемых теорем, является причиной использования приводимых доказательств. Этим достигается две цели, с одной стороны книга не перегружена доказательствами и примерами, с другой стороны примеры и доказательства присутствуют, что делает книгу учебником.

В целях удобства ссылок для глав применена сплошная нумерация. Нумерация параграфов и формул производится в каждой главе самостоятельно. При ссылках на формулу или параграф впереди, соответственно, номера формулы или параграфа указывается номер главы.

Например, запись

$$c_v \overset{(9.44)}{=} T\left(\frac{\partial s}{\partial T}\right)_v$$

означает, что левая часть уравнения приравнена правой части на основании формулы номер 44 в главе номер 9.

Искренне благодарю А. Каленюка, внимательно прочитавшего рукопись и сделавшего большое количество ценных замечаний, учтённых при её доработке.

Автор

Часть I

Анализ категорий

Глава 1

Предварительные сведения

В этой главе мы сформулируем первоначальные понятия, без которых невозможно сколь-нибудь глубокое изучение математики, например, понятие множества.

В первую очередь нас будут интересовать множества, используемые в геометрии, т.е. различные фигуры, которые представляют собой непрерывные множества точек. Примерами фигур могут служить отрезок, прямоугольник, круг, сфера, кривая, поверхность и т.д. Кроме фигур, нас будут интересовать пространства, т.е. бесконечные непрерывные множества точек. Самыми простыми пространствами являются прямая линия — одномерное пространство, плоскость — двумерное пространство, а также пространства большей размерности. Фигуры, пространства, а также отдельные точки, из которых они состоят, попадают под определение — объекты.

Мы рассмотрим, как записываются обозначения фигур и пространств, а также как сложные объекты получаются из более простых, например, в результате их прямого произведения.

Еще одним понятием является отображение одних объектов на другие. Мы рассмотрим отображения с определёнными свойствами, которые носят специальные названия, например, преобразование (отображение на себя), операция (отображение произведения множеств), функция (отбражение пространств). В результате окажется, что отображения сами являются объектами, на которые также могут действовать другие отображения. Мы отдельно остановимся на таких специальных отображениях как изоморфизм — взаимно однозначное отображение, позволяющее устанавливать одинаковые с точки зрения математики объекты, а также гомеоморфизм — непрерывный изоморфизм.

Мы получим представление о топологических инвариантах — свойствах, сохраняемых любыми гомеоморфизмами.

Совокупность объектов и их отображений друг на друга даёт понятие категории. В начале первого параграфа мы сформулируем определение математики как науки о различных категориях, т.е. науки о различных классах объектов и заданных на них отображениях.

Отдельно мы рассмотрим элементарные функции — наиболее давно изучаемые в математике отображения чисел.

В дальнейшем мы рассмотрим такую очень важную категорию, как группа — множество, на котором задана специальная операция (отображение). Окажется, что в качестве элементов группы могут выступать отображения. Например, у группы преобразований операция (отображение) задана на множестве преобразований (отображений на себя).

Теорию чисел мы специально рассматривать не будем, но при изучении элементарных функций выяснится, что действительных чисел не всегда хватает для вычисления этих функций. Поэтому мы специально рассмотрим множество комплексных чисел и заданные на нём функции. При этом специально будет подчеркнуто, что комплексные числа являются последним расширением числовой системы, т.е. никаких новых чисел больше не требуется.

Отдельно будет рассмотрен такой объект, как сингулярная цепь, а также её граница. Несмотря на абстрактность этих понятий, они являются очень важными и понадобятся нам в дальнейшем. С ними значительно удобнее работать, чем пытаться переводить на математический язык интуитивные представления о «поверхности, ограничивающей тело» или «линии, ограничивающей поверхность». С помощью этих понятий можно формально записать и использовать при исследованиях и вычислениях фигуры любой формы и их границы.

1.1 Множества

Матема́тика (от греческого $\mu\alpha\theta\eta\mu\alpha$ — наука, знание, изучение, и $\mu\alpha\theta\eta\mu\alpha\tau\iota\kappa o\varsigma$ — любовь к познанию) — наука, изучающая количественные и пространственные соотношения в действительном мире и человеческом воображении. Существуют совершенно иные и весьма разнообразные трактовки предмета математики и её метода, например, как науки о следствиях из непротиворечивых наборов аксиом.

В абстрактном понимании математика представляет собой науку о *категориях* (от греч. $\kappa\alpha\tau\eta\gamma o\rho\iota\alpha$). Любая категория представляет собой совокупность двух неопределяемых понятий: *объектов* (от лат. objectum — предмет) и *морфизмов* (от греч. $\mu o\rho\varphi\eta$ — форма, вид).

Объекты (однотипные математические структуры) могут быть самые разные. Далее будут изучены такие объекты как число, вектор, матрица, множество, группа, топологическое пространство и т.д.

Морфизмы (отображения) также могут быть самые различные, например, алгебраические операции, функции и т.д.

Категория — это совокупность однотипных математических структур (объектов) и отображений (морфизмов) между этими структурами, при выполнении ряда естественных дополнительных условий.

Множество или *совокупность объектов* — это понятие также не определяется, но может быть пояснено с помощью примеров. «Множество есть

многое, мыслимое как единичное» (Г. Кантор).

Запись $\mu \in M$ означает, что объект μ есть элемент множества M (принадлежит множеству M). Знак $\in$ выражает отношение принадлежности элемента некоторому множеству.

Множество можно обозначать фигурными скобками: запись $\{\mu\}$ означает «множество элементов μ». Чтобы определить множество, достаточно указать характеристическое свойство элементов, т.е. такое свойство, которым обладают все элементы этого множества и только они. Запись в фигурных скобках через двоеточие определяет множество через указание свойств его элементов: $\{\,\nu\colon \nu \text{ обладает свойством } S\,\}$, т.е. множество всех элементов ν, которые обладают свойством S. Может случиться, что данным свойством не обладает вообще ни один объект; тогда говорят, что это свойство определяет *пустое множество* $\varnothing$.

Класс или *совокупность множеств* — термин, употребляемый в математике в основном как синоним термина множество. Например, совокупности чисел, векторов, матриц и т.д. образует соответствующие множества, а совокупность различных множеств объектов образует класс этих объектов. Каждая категория состоит из двух классов — класса объектов и класса морфизмов.

Если каждый элемент множества A является в то же время элементом множества B, то множество A называется подмножеством B и записывается так: $A \subset B$ или $B \supset A$.

Объединением множеств $A \cup B \cup \ldots$ называется множество всех тех элементов, каждый из которых есть элемент хотя бы одного из множеств A, B и т.д.

Пересечением множеств $A \cap B \cap \ldots$ называется множество элементов, общих всем данным множествам A, B и т.д.

Дополнением $B \setminus A$ подмножества A в множестве B или *разностью* $B - A$ множеств B и A называется множество всех элементов из B, не принадлежащих A.

Прямым произведением множеств X и Y называется множество $X \times Y$ всевозможных пар объектов (x, y), где $x \in X$, $y \in Y$.

Следует быть осторожным с интуицией, вызванной словом «произведение» в термине «прямое произведение». Большинство привычных из арифметики свойств умножения чисел не верны для прямого произведения множеств.

Прямая степень множества X (n-я декартова степень множества) есть множество

$$X^n = \underbrace{X \times \cdots \times X}_{n} \tag{1.1}$$

всевозможных *последовательностей* $(x^1, \ldots, x^n)$, где $x^1, \ldots, x^n \in X$, т.е. множеств элементов, размещённых в определённом порядке.

Последовательности бывают конечные (состоящие из конечного числа элементов) и бесконечные.

При определённых условиях элементы множеств, полученных прямым произведением, называют также *векторы* или *кортежи* — последовательности конечного числа элементов. Многие математические объекты формально определяются как кортежи.

Число n в X^n называется *показатель степени*, X — *основание*.

В отличие от показателя степени n, числа $1, \ldots, n$, использованные для обозначения элементов последовательности, это не показатели степени, а *индексы* (лат. index — указатель) — числовые или буквенные указатели, которым снабжаются математические выражения, чтобы отличить их друг от друга. Индексы бывают верхние x^n и нижние x_n.

Счётное множество — множество, объекты которого мы можем пересчитать. Например, множество натуральных $\mathbb{N}$, целых $\mathbb{Z}$ и рациональных $\mathbb{Q}$ чисел — счётные множества.

Множество действительных (вещественных) чисел называют *числовой осью* или *числовой прямой* $\mathbb{R}$. Термин ввел И.Барроу (1670). Числовая ось изображает действительные числа в виде линейно упорядоченного множества точек, поэтому понятия «число»и «точка на оси»не различают. Множество действительных чисел пересчитать нельзя — это *несчётное множество*.

Важнейшее свойство числовой оси состоит в ее *непрерывности*. Действительные числа непрерывно заполняют числовую ось, т.е. всякому действительному числу соответствует точка числовой оси, и никаких других точек (чисел) на рассматриваемой прямой нет. Строгая теория действительных чисел была развита лишь во 2-й половине 19 в. трудами К. Вейерштрасса, Р. Дедекинда и Г. Кантора.

Примеры непрерывных множеств:

1. Прямое произведение двух числовых осей образует *плоскость*

$$\mathbb{R}^2 \overset{(1.1)}{=} \mathbb{R} \times \mathbb{R}, \tag{1.2}$$

   каждая точка которой задается двумя числами (x^1, x^2).

2. Прямое произведение трёх числовых осей образует *пространство*

$$\mathbb{R}^3 \overset{(1.1)}{=} \mathbb{R} \times \mathbb{R} \times \mathbb{R}, \tag{1.3}$$

   каждая точка которого задаётся тремя числами (x^1, x^2, x^3).

3. Прямая n-я степень множества действительных чисел $\mathbb{R}$ образует *n-мерное пространство*

$$\mathbb{R}^n \overset{(1.1)}{=} \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n}, \tag{1.4}$$

   каждая *точка* $\nu \in \mathbb{R}^n$ которого, есть n-членная последовательность вещественных чисел $\nu = (x^1, \ldots, x^n)$.

Обратим внимание ещё раз. Точка на числовой оси — это одно действительное число или «1-членная последовательность чисел», а точка в пространстве — это «n-членная последовательность чисел». В отличие от чисел, обозначаемых наклонным шрифтом x, y, z, точки, векторы и прочие «n-членные последовательности чисел» обозначают полужирным шрифтом $\mathbf{x} = (x^1, \ldots, x^n)$ или греческими буквами $\nu = (x^1, \ldots, x^n)$.

Как и для числовой оси интуитивное представление о пространстве связано с его *непрерывностью*, которое является «коренным свойством» или важнейшим свойством пространства, т.е. пространство $\mathbb{R}^n$ есть непрерывное множество n-членных последовательностей $(x^1, \ldots, x^n)$ или точек ν, и никаких других точек, кроме рассматриваемых n-членных последовательностей, в пространстве не существует.

Если $X \subset \mathbb{R}^n$ и $Y \subset \mathbb{R}^m$, то $X \times Y \subset \mathbb{R}^{n+m}$, где $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$ определяется как множество всех пар точек $(\nu, \tau) \in \mathbb{R}^{n+m}$ с $\nu = (x^1, \ldots, x^n) \in X$ и $\tau = (y^1, \ldots, y^m) \in Y$. Это правило распространяется на произведение любого числа множеств.

Показатель степени n называют также *размерностью* — числом измерений пространства $\mathbb{R}^n$ или множества $X \subset \mathbb{R}^n$ и обозначают $n = \dim X$ (от лат. dimension — измерение).

Кроме *размерности множеств* существует понятие *размерности величин* — выражение, показывающее, во сколько раз изменится единица данной величины при изменении *основных единиц измерения*, которые образуют так называемую *систему мер*.

Например, системой мер физических величин является система СГС (сантиметр, грамм, секунда). Остальные единицы измерения различных величин называются *производными единицами измерения* и выражаются через основные единицы при помощи уравнений связи между величинами. СГС система называется также *системой единиц Гаусса* в честь К. Ф. Гаусса, впервые в 1832 предложившего систему с основными единицами: миллиметр, миллиграмм и секунда и применившего эту систему для измерения магнитных величин.

С непрерывностью тесно связано понятие близости. Запись $\nu \, \delta X$ означает «точка ν близка множеству X». *Близость* δ определяется при помощи набора *аксиом* (др.-греч. $\alpha\xi\iota\omega\mu\alpha$ — утверждение, положение), т.е. самоочевидных утверждений, принимаемых без доказательств:

1. никакая точка не близка пустому множеству;
2. $\nu \, \delta X$ для всех $\nu \in X$;
3. если $\nu \, \delta A$ и $A \subset B \subset X$, то $\nu \, \delta B$;
4. если $\nu \, (A \cup B)$, то $\nu \, \delta A$ или $\nu \, \delta B$;
5. если $\tau \, \delta A$ для всех $\tau \in B$ и $\nu \, \delta B$, то $\nu \, \delta A$.

Замкнутое множество содержит все близкие к нему точки.

Открытое множество всегда есть дополнение замкнутого множества, и наоборот, замкнутое множество всегда есть дополнение открытого множества. Например, множество $V \subset \mathbb{R}^n$ замкнуто в $\mathbb{R}^n$, если множество $U = \mathbb{R}^n \setminus V$ открыто. Любая точка множества есть замкнутое множество. Если множество состоит из одной точки, то дополнением замкнутого множества будет пустое множество $\varnothing$, которое открыто.

Пересечение двух (а потому и любого конечного числа) открытых множеств также открыто. Если пересечением множеств будет пустое множество, то оно также открыто. Объединение любого, даже бесконечного числа открытых множеств также открыто.

Пересечение любого числа замкнутых множеств замкнуто. Объединение конечного числа замкнутых множеств замкнуто.

Специальный раздел геометрии — *топология* (от греч. $\tau o \pi o \varsigma$ — место, $\lambda o \gamma o \varsigma$ — слово, учение), — изучает наиболее глубокие геометрические свойства множеств и пространств, основанные на идее непрерывности.

Топологическое пространство есть множество, в котором некоторые подмножества выделены как открытые, причём пересечение конечного числа и объединение любого (даже бесконечного) числа этих открытых множеств открыто.

Окрестностью данной точки в пространстве называется каждое содержащее эту точку открытое множество.

Точка множества называется *внутренней*, если имеет окрестность, целиком принадлежащую множеству. Точка называется *внешней*, если имеет окрестность, целиком не принадлежащую множеству. Остальные точки называются *граничными*. Таким образом, для множества $A \subset \mathbb{R}^n$, точки $\nu \in \mathbb{R}^n$ и её окрестности B всегда осуществляется одна из трех возможностей:

1. Существует такое B, что $\nu \in B \subset A$.
2. Существует такое B, что $\nu \in B \subset \mathbb{R}^n \setminus A$.
3. Для ν всякое B содержит как точки из A, так и точки из $\mathbb{R}^n \setminus A$.

Внутренние точки, обладающие свойством 1, образуют *внутренность* множества A ($\mathrm{Int}A$ от interior). Внешние точки, обладающие свойством 2, образуют *внешность множества* A. Граничные точки, обладающие свойством 3, образуют *границу* ∂A множества A.

Нетрудно видеть, что внутренность всякого множества A всегда есть открытое множество, так как целиком состоит из внутренних точек. То же верно и для внешности A, ибо она есть не что иное, как внутренность $\mathbb{R}^n \setminus A$. Таким образом, их объединение открыто, так что граница ∂A, являясь его дополнением, должна быть замкнутой.

Замыкание $[A]$ *множества* A есть (три равнозначных определения):

1. пересечение всех замкнутых множеств, содержащих рассматриваемое множество;

2. минимальное замкнутое множество, содержащее рассматриваемое множество;

3. совокупность внутренности множества и его границы.

Если множество *замкнуто*, то $[A] = A$. Замыкание множества образовано точками, близкими этому множеству. Их называют *точками прикосновения* данного множества.

Пример. Пусть A — круг на плоскости, тогда $\mathrm{Int}\, A$ — открытый круг, $[A]$ — замкнутый круг, ∂A — окружность (определения см. ниже).

Множество *всюду плотно* в пространстве, если в сколь угодно малой окрестности любой точки пространства есть точка рассматриваемого множества.

Замыкание всюду плотного множества — всё пространство.

Изолированная точка множества — такая точка, что некоторая её окрестность не содержит других точек множества.

Предельная точка множества — в любой окрестности этой точки содержится, по крайней мере, одна точка данного множества, отличная от неё самой. Например, любая внутренняя точка множества есть его предельная точка, а любая изолированная точка множества не является предельной. Но предельная точка множества может и не принадлежать множеству. Например, любая граничная точка открытого множества является его предельной точкой, но не принадлежит множеству.

Производное множество есть совокупность всех его предельных точек. Производное множество получается вычитанием из замыкания множества всех его изолированных точек.

Совершенное множество — множество, совпадающее со своим производным множеством.

Непрерывности противопоставляется *дискретность* (от лат. discretus — разделенный, прерывистый). Например, множества натуральных $\mathbb{N}$, целых $\mathbb{Z}$, рациональных $\mathbb{Q}$ и иррациональных $\mathbb{I}$ чисел — дискретные. Хотя при этом множество рациональных $\mathbb{Q}$ и иррациональных $\mathbb{I}$ чисел всюду плотно в пространстве действительных чисел $\mathbb{R}$.

Связность — свойство топологического пространства, состоящее в том, что его нельзя представить в виде объединения двух отделённых друг от друга частей. Связное множество — множество в топологическом пространстве, состоящее как бы из одного куска, т.е. такое, что при любом его разбиении на два непересекающихся непустых подмножества одно из них содержит предельную точку другого. Например: связными множествами будут отрезок, окружность, круг, шар, интервал, прямая. Объединение связных множеств, имеющих непустое пересечение, связно.

Областью называется открытое связное множество.

Односвязная область — область, в которой любой контур (замкнутая кривая) может быть непрерывно деформирована (стянута) в точку, оставаясь все время в пределах области.

Многосвязная область — область, в которой существуют замкнутые кривые, не стягиваемые в пределах этой области в точку.

1.2 Подмножества (фигуры) в пространстве

В качестве «точек» произвольного пространства может выступать любая непрерывная совокупность объектов, явлений и даже «состояний». Например, невертикальная прямая на плоскости задается как $y = ax + b$. Следовательно, одна прямая определяется парой чисел (a, b) и может рассматриваться как точка плоскости с координатами (a, b). Эта плоскость называется двойственной к исходной плоскости. Её точки — это прямые исходной плоскости.

Современная *геометрия* понимается как наука, изучающая самые разные и причудливые пространства, а также «*фигуры*» или объекты, которые представляют собой различные подмножества (множества) в этих пространствах.

Рассмотрим наиболее распространенные подмножества (фигуры) в пространстве.

Интервал (от лат. intervalum — промежуток, расстояние) обозначаемый $I = (a, b)$ или $a < x < b$, есть простейшее открытое множество на прямой.

Сегмент (от лат. segmentum — отрезок) обозначаемый $a \leqslant x \leqslant b$ или $[a, b]$, есть замкнутое множество на прямой. Круглая скобка обозначает, что соответствующий конец интервала не принадлежит к рассматриваемому множеству, а квадратная — что принадлежит.

Обозначение (a, b) ввёл Г.Ковалевский (1909), $[a, b]$ — Х.Хан (1921).

Бесконечный или несобственный интервал: $(a, +\infty)$ или $(-\infty, a)$ — есть открытое множество (открытая полупрямая). Дополнение открытой полупрямой до прямой, обозначаемое $(-\infty, a]$ и $[a, +\infty)$, есть замкнутое множество (замкнутая полупрямая или луч).

Интервал I также можно определить как связное подмножество $\mathbb{R}$, не сводящееся к одной точке. Он может быть конечным или бесконечным, замкнутым, открытым или полуоткрытым.

Окружность — замкнутая плоская (принадлежащая плоскости) кривая (одномерная поверхность), обозначаемая S^1, все точки которой одинаково удалены от данной точки, которая называется *центром*.

Сфера (от греч. $\sigma\varphi\alpha\ddot{\imath}\rho\alpha$ — шар, мяч) — поверхность в $(n+1)$-мерном пространстве, обозначаемая S^n, все точки которой одинаково удалены от одной точки, *центра сферы*. Отрезок, соединяющий центр сферы (окружности) с какой-либо её точкой, называется *радиусом* r сферы. Сферу также можно получить вращением конца радиуса вокруг центра. Отрезок, соединяющий две точки сферы, называется *хордой*. Наибольшая из хорд, проходящая через центр называется *диаметром*.

Круг D — часть плоскости, ограниченная окружностью и содержащая её центр. *Шар* D^n — часть пространства, ограниченная сферой и содержащая её центр, тогда сфера — граница шара

$$S^n = \partial D^n. \tag{1.5}$$

Внутренность шара (круга) образует открытый шар (круг), который удобно использовать в качестве окрестности точки.

Индукция (от лат. inductio — наведение, побуждение) — метод, когда утверждение $\varphi(n)$, зависящее от натурального числа n, считается доказанным, если доказано $\varphi(1)$ и доказано $\varphi(n+1)$ в предположении, что утверждение $\varphi(n)$ верно.

Тор (от лат. torus — валик, выпуклость, узел) — поверхность

$$T^2 = S^1 \times S^1, \tag{1.6}$$

которая образуется вращением окружности вокруг прямой, лежащей в плоскости этой окружности, но не пересекающей его. Форму тора имеет, например, бублик.

Тор n-мерный есть образуемая по индукции поверхность

$$T^n = T^{n-1} \times S^1 \tag{1.7}$$

в $(n+1)$-мерном пространстве.

Цилиндр есть произведение окружности на сегмент:

$$C = S^1 \times [a, b]. \tag{1.8}$$

Различные степени n-мерного сегмента дают *квадрат* $I^2 = I \times I$, *куб* I^3, *n-мерный куб* I^n. Все кубы связны.

Эллипс (от греч. *ἔλλειψις* — недостаток) — множество точек плоскости, заданное уравнением $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$, где a, b — полуоси. Если $a = b = r$, то эллипс превращается в окружность радиуса r.

Эллипсоид (от греч. *ἔλλειψις* — недостаток и греч. *εἶδος* — вид) — множество точек пространства, заданное уравнением $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 + \left(\frac{z}{c}\right)^2 = 1$, где a, b, c — полуоси *трёхосного эллипсоида*. Если $b = c$, то трёхосный эллипсоид превращается в *эллипсоид вращения*, получаемый вращением эллипса $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$ вокруг полуоси a.

Плоскостью или *m-мерной плоскостью* в пространстве $\mathbb{R}^n$ называется множество $\mathbb{R}^m \times \{0\} = \{\, \nu \in \mathbb{R}^n \colon x^{m+1} = 0, \ldots, x^n = 0 \,\}$. Гиперплоскостью или *$(n-1)$-мерной плоскостью* в пространстве $\mathbb{R}^n$ называется пространство $\mathbb{R}^{n-1}$.

Полупространством в n-мерном пространстве $\mathbb{H}^n \subset \mathbb{R}^n$ называется множество $\{\, \nu \in \mathbb{R}^n \colon x^n \geqslant 0 \,\}$. Например: полупрямая или луч $\mathbb{H}^1 = [0, +\infty)$; полуплоскость $\mathbb{H}^2 = \{\, (x, y) \colon y \geqslant 0 \,\}$ — совокупность точек плоскости, лежащая по одну сторону от оси y (границы полуплоскости); полупространство $\mathbb{H}^3 = \{\, (x, y, z) \colon z \geqslant 0 \,\}$ — совокупность точек пространства, лежащая по одну сторону от плоскости $\{\, (x, y) \,\}$ (границы полупространства) и т.д.

Угол — геометрическая фигура, состоящая из двух различных лучей (замкнутых полупрямых), выходящих из одной точки. Лучи называются сторонами угла, а их общее начало — вершиной. Два угла называют *смежными*, если у них общие вершина и одна сторона, а две другие образуют прямую. Угол, у которого стороны образуют прямую, называется *развёрнутым*. Угол, равный своему смежному углу, называется *прямым*. Угол меньший прямого, называется *острым*, больший прямого и меньший развёрнутого — *тупым*. *Перпендикулярные прямые* (содержащие их плоскости) —

это прямые (плоскости) A и B, взаимное расположение которых составляет прямой угол $A \perp B$. Знак $\perp$ для обозначения перпендикулярности ввёл П. Эригон (1634).

Не следует смешивать понятие угла — геометрической фигуры, и величины этого угла — которая является числом. *Величина* развёрнутого угла равна сумме величин смежных углов и принимается за число π, которое обозначает отношение длины окружности к диаметру. Величины смежных углов будут, соответственно φ и $\pi - \varphi$.

Треугольник (*четырехугольник*, *многоугольник*) — три (четыре, n) отрезка или стороны этой фигуры, соединённые концами в трёх (четырёх, n) вершинах — не лежащих на одной прямой точках плоскости.

Биссектриса (от лат. bis — дважды и seco — рассекаю) — прямая, проходящая через вершину угла и делящая его пополам. Биссектрисы трёх углов треугольника пересекаются в одной точке — центре вписанного в треугольник круга.

Параллелогра́мм (от греч. $\pi\alpha\rho\acute{\alpha}\lambda\lambda o\varsigma$ — параллельный и $\gamma\rho\acute{\alpha}\mu\mu\alpha$ — линия) — это четырёхугольник, у которого противолежащие стороны попарно параллельны, т. е. лежат на параллельных прямых. Частными случаями параллелограмма являются прямоугольник, квадрат и ромб.

Замкнутый параллелограмм это естественный аналог сегмента (отрезка) в $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, определяемый как прямое произведение отрезков

$$P^2 = [a^1, b^1] \times [a^2, b^2], \tag{1.9}$$

т.е. множество всех точек $(x^1, x^2) \in \mathbb{R}^2$ с $x^1 \in [a^1, b^1]$ и $x^2 \in [a^2, b^2]$, а аналог интервала — открытый параллелограмм $(a^1, b^1) \times (a^2, b^2)$.

Параллелепипед в $\mathbb{R}^3$ (от греч. $\pi\alpha\rho\acute{\alpha}\lambda\lambda o\varsigma$ — параллельный и греч. $\varepsilon\pi\iota\pi\varepsilon\delta o\nu$ — плоскость) — шестигранник, противоположные грани (в данном случае четырехугольники) которого попарно параллельны и представляют собой попарно равные параллелограммы.

Замкнутый и открытый параллелепипеды в трёхмерном пространстве определяются как прямое произведение отрезков:

$$P^3 = [a^1, b^1] \times [a^2, b^2] \times [a^3, b^3] \quad \text{и} \quad (a^1, b^1) \times (a^2, b^2) \times (a^3, b^3). \tag{1.10}$$

Замкнутым параллелепипедом в $\mathbb{R}^n$ называется множество

$$P^n = [a^1, b^1] \times \cdots \times [a^n, b^n] \subset \mathbb{R}^n, \tag{1.11}$$

а *открытым параллелепипедом* — множество $(a^1, b^1) \times \cdots \times (a^n, b^n)$.

Открытый параллелепипед, как и открытый шар, удобно рассматривать в качестве окрестности точки, и этим часто пользуются.

Длина как замкнутого $[a, b]$, так и открытого (a, b) интервала определяется как число, равное разнице чисел $b - a$.

Площадь параллелограмма P, как замкнутого, так и открытого, вычисляется как произведение длин интервалов на синус угла между ними:

$$S(P) = (b^1 - a^1)(b^2 - a^2) \sin \varphi. \tag{1.12}$$

За φ принимается угол между образующими параллелограмм интервалами, не превосходящий π.

Объём замкнутого и открытого параллелепипеда P вычисляется как:

$$V(P) = (b^1 - a^1)(b^2 - a^2)(b^3 - a^3)\sin\varphi\sin\theta. \tag{1.13}$$

За φ принимается угол между образующими параллелограмм интервалами, а за θ — угол между параллелограммом и третьим интервалом. Оба угла не превосходят π.

Объём замкнутого и открытого n-мерного параллелепипеда P вычисляется по индукции:

$$V(P) = (b^1 - a^1)\dots(b^n - a^n)\sin\varphi_1\dots\sin\varphi_{n-1}. \tag{1.14}$$

Множества (a, b), $[a, b]$ и $[a, b)$ есть *ограниченные множества* в отличие от множеств $(a, +\infty)$, $(-\infty, a)$, $(-\infty, a]$ и $[a, +\infty)$. Открытые и замкнутые параллелепипеды также есть ограниченные множества. Подмножество ограниченного множества и объединение конечного числа ограниченных множеств также есть ограниченные множества.

Всякое замкнутое и ограниченное множество в $\mathbb{R}^n$ *компактно*. Обратное также верно: всякое компактное множество в $\mathbb{R}^n$ замкнуто и ограничено. Например: Замкнутый интервал $[a, b]$ компактен (этот результат называется теорема Бореля–Лебега). Если множества $A \subset \mathbb{R}^n$ и $B \subset \mathbb{R}^m$ компактны, то $A \times B \subset \mathbb{R}^{n+m}$ компактно. Следовательно, если множества $A_1, \dots, A_n$ компактны, то множество $A_1 \times \dots \times A_n$ компактно. В частности, любой замкнутый параллелепипед компактен. Очевидно, что любое неограниченное подмножество в пространстве некомпактно.

Континуумом называется компактное связное множество.

Определим компактное множество на математическом языке: Семейство Θ открытых множеств называется *открытым покрытием* множества A (или, кратко, покрывает A), если каждая точка $\nu \in A$ принадлежит некоторому множеству семейства Θ, т.е. объединение множеств семейства Θ содержит A. Множество A называется *компактным*, если всякое его открытое покрытие Θ содержит конечное подсемейство (конечное число множеств), также покрывающее A.

Исторически в математике возникло несколько терминов, которые, по сути, являются синонимами: метрика, норма, модуль и абсолютная величина.

Метрика (от греч. *μέτρον* — мера, размер) есть *расстояние* между двумя точками (элементами) множества M: $\nu, \tau \in M \subset \mathbb{R}^n$ или длина соединяющего их отрезка $\xi = \nu - \tau$, т.е. число:

$$\rho(\xi) = \rho(\nu, \tau) = \|\nu - \tau\| \quad \text{или} \quad \rho\colon M \times M \to \mathbb{R}. \tag{1.15}$$

Свойства метрики:

1. $\rho(\nu, \tau) \geqslant 0$, причём $\rho(\nu, \tau) = 0$ тогда и только тогда, когда $\nu = \tau$;

2. $\rho(\nu, \tau) = \rho(\tau, \nu)$;

3. $\rho(\nu, \tau) + \rho(\tau, \upsilon) \geqslant \rho(\nu, \upsilon)$.

Запись вида $x = \rho(\nu, \tau)$ означает отображение ρ элементов ν и τ в элемент x. Расстояние между элементами ν и τ — это число, т.е. в данной записи под метрикой (расстоянием) понимается отображение элементов $\nu, \tau \in M$ в точку (число) на числовой оси.

Абсолютная величина или модуль $|a|$ (от лат. modulus — мера) действительного числа a — неотрицательное число, определяемое по правилу: если $a \geqslant 0$, то $|a| = a$, если $a < 0$, то $|a| = -a$.

Норма или *модуль* (от лат. norma — руководящее начало, правило, образец) — обобщение понятия абсолютной величины числа: $|\nu - \tau| \geqslant 0$, или $\|\nu - \tau\| \geqslant 0$. В дальнейшем мы будем изучать нормы различных классов геометрических объектов.

Пространство с метрикой называется *метрическим пространством*, с нормой — *нормированным пространством*.

Переменная — величина, которая в изучаемой задаче принимает различные значения, причём так, что все допустимые значения переменной полностью определены наперёд заданными условиями.

Предел — одно из основных понятий математического анализа, означающее что некоторая переменная в рассматриваемом процессе её применения неограниченно приближается к какому-то постоянному значению или, как говорят, *сходится*. Понятие предела использовалось ещё Ньютоном во второй половине 17-го века и математиками 18-го века, такими как Эйлер и Лагранж, однако они понимали предел интуитивно. Первые строгие определения предела последовательности (1.16) дали Больцано (1816) и Коши (1821).

Мера множества — обобщённое понятие длины отрезка, площади плоской (лежащей на плоскости) фигуры и объёма тела на множества более общей природы. В качестве примера можно привести определение *меры Лебега* , введённой А. Лебегом в 1902, для ограниченных множеств, лежащих на плоскости. Определение меры напоминает обычный процесс измерения площади.

Основные свойства меры Лебега:

1. Мера любого множества A неотрицательна: $m(A) \geqslant 0$.

2. Мера суммы конечной или счётной системы попарно непересекающихся множеств $A_1, \ldots, A_n, \ldots$: $A = \sum_{n=1}^{\infty} A_n$ равна сумме их мер: $m(A) = \sum_{n=1}^{\infty} m(A_n)$.

3. При перемещении множества как твёрдого тела его мера не меняется.

Для более узких классов множеств мера, совпадающая с мерой Лебега была ранее определена К. Жорданом (1893) и Э. Борелем (1898).

Множество, имеющее меру, называется *измеримым множеством*.

Множество $A \subset \mathbb{R}^n$ имеет (n-мерную) меру нуль $m(A) = 0$, если для всякого $\varepsilon > 0$ существует такое покрытие $\{U_1, U_2, \dots\}$ этого множества замкнутыми параллелепипедами, что $\sum_{i=1}^{\infty} V(U_i) < \varepsilon$. Вместо замкнутых параллелепипедов можно брать открытые.

Свойства меры 0:

1. Множество, содержащее лишь конечное число точек, очевидно, имеет меру 0.

2. Множество, состоящее из бесконечного числа точек, которые могут быть занумерованы в последовательность, также имеет меру 0. Важным и неожиданным примером такого бесконечного множества является множество A всех рациональных чисел между 0 и 1: $m(A) = 0$. При этом множество B всех иррациональных чисел (точек) того же интервала имеет меру 1: $m(B) = 1$. При этом оба множества A и B сходны в том смысле, что всюду плотны на $(0, 1)$, т.е. между любыми двумя точками указанного интервала найдутся как точки множества A, так и точки множества B.

3. Если A имеет меру 0, то $B \subset A$ имеет меру 0.

4. Если $A = A_1 \cup A_2 \cup \dots$ и каждое A_i имеет меру 0, то A имеет меру 0.

5. Если $a < b$, то отрезок $[a, b] \subset \mathbb{R}$ не может иметь меру 0.

Множество $A \subset \mathbb{R}^n$ имеет (n-мерный) *объём* 0, если для всякого $\varepsilon > 0$ существует такое конечное покрытие $\{U_1, \dots, U_n\}$ этого множества замкнутыми параллелепипедами, что $\sum_{i=1}^{n} V(U_i) < \varepsilon$. Очевидно, что множество, имеющее объём 0, имеет также меру 0. Вместо замкнутых параллелепипедов можно, как и раньше, воспользоваться открытыми.

Свойства объёма 0:

1. Компактное множество A, имеющее меру 0, имеет также (n-мерный) объём 0. Это заключение неверно, если A некомпактно (не имеет конечного покрытия). Например, множество всех рациональных чисел между 0 и 1 имеет меру 0, но не может иметь объём 0.

2. Неограниченное множество не может иметь объём 0.

1.3 Отображения

Морфизм или *отображение* ω множества X в множество Y записывается как: $\omega\colon X \to Y$. Отображение элемента $x \in X$ в элемент $y \in Y$ записывается как: $\omega\colon x \mapsto y$ или $y = \omega(x)$. При этом элемент y называется *образ* x, а элемент x называется *прообраз* y.

Область определения $\omega\colon X \to Y$ — есть $U = \{x\colon \omega(x) \in Y\} \subset X$. Запись в фигурных скобках означает «множество элементов x, таких что $\omega(x) \in$

Y». Другими словами, если ω определено не на всём X, а на некотором его подмножестве U, то U — область определения ω.

Область значений отображения ω — есть $V = \{\omega(x)\colon x \in X\} \subset Y$. Другими словами, если множество значений $\omega(X)$, которое называют также *образ* X, есть не всё множество Y, а некоторое его подмножество $V = \omega(X)$, то V — область значений ω.

Отображение $\omega\colon X \to Y$ порождает в произведении множеств $X \times Y$ множество $\Gamma_\omega = \{\,(x, y)\colon y = \omega(x)\,\} \subset X \times Y$, которое называется *графиком отображения* ω и полностью определяет это отображение.

Обратное отображение к $\omega\colon X \to Y$ есть отображение $\omega^{-1}\colon Y \to X$, где $\omega^{-1}(Y) = \{\, y \in Y\colon \omega(y) \in X\,\}$ (если конечно оно существует). Одно из отображений при этом называют правым, а другое — левым.

Инъекция, *инъективное* отображение $\omega\colon X \to Y$ (отображение «в») — когда разные элементы из X имеют разные образы в Y, т.е. для $\nu, \tau \in X$ если $\nu \neq \tau$, тогда $\omega(\nu) \neq \omega(\tau)$, а если $\nu = \tau$, тогда $\omega(\nu) = \omega(\tau)$.

Сюръекция, *сюръективное* отображение $\omega\colon X \to Y$ (отображение «на») — когда любой элемент из Y имеет свой прообраз в X.

Биекция, *биективное* отображение $\omega\colon X \to Y$ (от лат. bi — двойной, двоякий и jacio — бросаю) или *взаимно однозначное* отображение — отображение, инъективное и сюръективное одновременно.

Инъекция является необходимым, а вместе с сюръекцией, достаточным условием биекции.

Изоморфизм (от греч. *ἴσος* — равный и *μορφή* — вид, форма) — это отображение ω, для которого существует взаимно однозначное обратное отображение ω^{-1}. Другими словами, если между двумя множествами X и Y можно установить взаимно однозначные в обе стороны (левую и правую) отображения $\omega\colon X \to Y$ и $\omega^{-1}\colon Y \to X$, то множества X и Y изоморфны (похожи или подобны), а отображение ω есть изоморфизм.

Существование изоморфизмов подметил ещё Р. Декарт, хотя сам термин возник только в середине 19 в. Он называл такие отношения или операции «подобными» и предвидел возможность их «отождествлять». Современная терминология утвердилась после работ Э. Нетер (1918).

С точки зрения современной математики аксиомы любой математической теории определяют систему изучаемую объектов всегда только с точностью до изоморфизма. Другими словами, аксиоматически построенная математическая теория, применимая к какой-либо одной системе объектов, всегда полностью применима к другой, изоморфной ей. Поэтому каждая аксиоматически изложенная математическая теория допускает не одну, а много «интерпретаций», «систем» или «моделей», примеры которых рассматриваются во второй части книги.

Преобразованием называют отображение φ некоторого множества M в себя: $\varphi\colon M \to M$ (биекция множества в себя). *Тождественным преобразованием* называют отображение, переводящее каждый объект множества μ в себя: $\mu \mapsto \mu$ для всех $\mu \in M$.

Инволюция (от лат. involution — изгиб, свёртывание) отображение объекта на себя, квадрат которого (прямое отображение, затем обратное отоб-

ражение) является тождественным отображением. Например, симметрия относительно прямой на плоскости.

Если заданы три множества X, Y, Z и два отображения $\alpha\colon X \to Y$ и $\beta\colon Y \to Z$, то существует отображение $\gamma\colon X \to Z$, определяемое равенством $\gamma(x) = \beta(\alpha(x))$. Это отображение называют *композицией*, *суперпозицией* или *произведением* отображений α и β и обозначают $\beta \circ \alpha\colon X \to Z$ (иногда просто $\beta\alpha$). Композиция $\alpha \circ \beta$, даже если она определена, может не совпадать с $\beta \circ \alpha$.

Алгебраической операцией на множестве A называют отображение ω некоторой прямой степени множества A в множество A: $\omega\colon A^n \to A$, когда каждой последовательности элементов $(a_1, a_2, \ldots, a_n)$ ставится в соответствие один элемент a из A, который обозначается $a = \omega(a_1, a_2, \ldots, a_n)$. Число n — *арность* операции ω.

К алгебраическим операциям относят сложение (вычитание), умножение и взятие противоположного элемента. Однако, деление, например, во множестве всех действительных чисел, алгебраической операцией не является, так как нельзя разделить на нуль. Взятие противоположного элемента также, разумеется, будет алгебраической операцией только на подмножестве положительных чисел, но невозможно на множестве всех действительных чисел.

Исторически первыми были понятия *бинарной* $n = 2$ и *унарной* $n = 1$ операций. Бинарная операция ω над множеством A: $\omega\colon A^2 \to A$ обозначается $a \circ b = c$, где $a, b, c \in A$. *Нульарные операции* — это фиксированные элементы множества, часто называемые просто *нулями*. В дальнейшем мы разберем алгебраические операции сложения и умножения над векторами (векторная алгебра), над матрицами (алгебра матриц) и т.д.

Свойства операций:

Ассоциативность (позднелат. assotiatio — соединение) или *сочетательность*: $a \circ (b \circ c) = (a \circ b) \circ c$; бывает ассоциативность сложения, когда $a + (b + c) = (a + b) + c$ или умножения $a(bc) = (ab)c$. Термин ввел У. Гамильтон (1843).

Дистрибутивность (лат. distributivus — распределительный) или *распределительность* умножения относительно сложения: $a(b + c) = ab + ac$ и $(b + c)a = ba + ca$. Термин ввел Ф. Сервуа (1815).

Коммутативность (позднелат. commutativus — меняющийся), *симметричность* или *переместительность*: $a \circ b = b \circ a$; а также *антикоммутативность*, *антисимметричность* или *кососимметричность*: $a \circ b = -b \circ a$. Бывает коммутативность сложения, когда $a + b = b + a$ или умножения $ab = ba$. Термин ввел Ф. Сервуа (1815).

Алгебраисты склонны объединять хорошо ведущие себя объекты в поле, кольцо или группу.

По определению существует *поле* вещественных чисел в котором возможны все четыре арифметические операции: сложение, вычитание, умножение и деление. Оно задаёт правила игры.

Множество целых чисел не образует поля (отсутствует правило деле-

ния), их скромный удел — *кольцо*, сохраняющее правила сложения, вычитания и умножения.

Рассмотрим термины поле, кольцо и группа в обратном порядке.

Группа, это непустое множество G, на котором задана бинарная алгебраическая операция и выполняются условия:

1. ассоциативность $a \circ (b \circ c) = (a \circ b) \circ c$, где $a, b, c \in G$;
2. существует такой элемент e, называемый *единица*, что для любых $a \in G$ выполняется условие $a \circ e = a$;
3. для любого $a \in G$ существует *обратный элемент* $a^{-1} \in G$, такой что $a \circ a^{-1} = e$.

Пример 1. Множество всех целых чисел с операцией по сложению есть группа. Роль e играет число 0, а роль обратного к числу «z» элемента — число «$-z$».

Пример 2. Для фигуры на плоскости множество всех движений плоскости, самосовмещающих данную фигуру, образует *группу симметрии фигуры*. Операцией служит композиция $\varphi \circ \psi$ двух движений φ и ψ, единицей служит тождественное преобразование плоскости, а обратным к φ элементом служит обратное движение φ^{-1}.

Группа G называется *абелевой* (*коммутативной*), если для любых $a, b \in G$ выполняется условие коммутативности $a \circ b = b \circ a$.

Группой преобразований множества называется такой набор преобразований, если вместе с каждым преобразованием в него входит обратное преобразование и с каждыми двумя преобразованиями — их произведение. Элементом единица в этом случае будет тождественное преобразование множества, переводящее каждый объект в себя.

Понятие группы имеет в алгебре фундаментальное значение. Так понятие изоморфизма возникло в теории групп, где впервые был понят тот факт, что изучение внутренней структуры двух изоморфных систем объектов представляет собой одну и ту же задачу. Понятие группы преобразований — одно из самых фундаментальных для всей математики и одновременно, одно из самых простых: человеческому мозгу свойственно мышление в терминах инвариантов групп преобразований.

Кольцо, это непустое множество R, на котором заданы две бинарные алгебраические операции — сложение и умножение и выполняются условия:

1. коммутативность сложения $a + b = b + a$, где $a, b \in R$;
2. ассоциативность сложения $a + (b + c) = (a + b) + c$, где $a, b, c \in R$;
3. обратимость сложения (возможность вычитания), т.е. уравнение $a + x = d$ имеет решение $x = b - a$, где $a, b, x \in R$;
4. Дистрибутивность умножения относительно сложения $a(b+c) = ab+ac$ и $(b + c)a = ba + ca$, где $a, b, c \in R$.

Из свойств 1–3 ясно, что элементы кольца образуют абелеву группу относительно сложения, называемую аддитивная группа кольца.

Поле, это кольцо R в котором заданы дополнительные условия:

1. коммутативность умножения $ab = ba$, где $a, b \in R$;

2. ассоциативность умножения $a(bc) = (ab)c$, где $a, b, c \in R$.

Поля представляют собой особый подкласс колец.

Если в кольце выполняется только условие 1, то оно называется коммутативным, а если только условие 2 — ассоциативным.

Если в кольце выполняются условия $a^2 = 0$ и $a(bc) + b(ac) + c(ab) = 0$, то оно называется кольцом Ли.

Числовое поле — поле, элементами множества которого являются числа. Другими словами, числовое поле это совокупность чисел, в пределах которой всегда выполнимы и однозначны четыре *арифметических операции* (*действия*) — сложение, вычитание, умножение и деление на число отличное от нуля: $a + b = c$, $a - b = c$, $a \cdot b = c$, $a/b = c$.

Примерами числовых полей и колец могут служить совокупность всех рациональных, действительных и комплексных чисел. Почти все законы элементарной арифметики являются следствиями аксиом, определяющих поле.

Термин «поле» впервые появился в «Теории чисел» П. Дирихле (1871).

Функция (от лат. function — исполнение, осуществление) — отображение между пространствами $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ или их подмножествами $\varphi\colon U \to V$ ($U \subset \mathbb{R}^n, V \subset \mathbb{R}^m$), т.е. отображение $\upsilon = \varphi(\nu)$, относящее каждой точке $\nu = (x^1, \ldots, x^n) \in U$ точку $\upsilon = (y^1, \ldots, y^m) \in V$. Множество U называется *область определения* функции, а множество $V = \varphi(U)$ называется *область значений* функции φ или *образ* U. Координаты точки $x^1, \ldots, x^n$ называются *переменными* функции. Если переменные не зависят от других переменных, то они называются независимыми переменными или *аргументами* (лат. argumentum, здесь — предмет, знак).

Константой (от лат. constantis — постоянный, неизменный) называется величина (часто обозначаемая *const*), противоположная по смыслу переменной и имеющая определённое числовое значение.

Если существует *обратная* функция, то она обозначается $\varphi^{-1}\colon V \to U$, где $\varphi^{-1}(V) = \{\, \nu \in U\colon \varphi(\nu) \in V \,\}$.

Если преобразование $\varphi\colon U \to \mathbb{R}^n$ *взаимно однозначно*, то $\varphi^{-1}\colon \varphi(U) \to \mathbb{R}^n$ определяется требованием, чтобы $\nu = \varphi^{-1}(\upsilon)$ было тем единственным $\nu \in U$, для которого $\varphi(\nu) = \upsilon$.

Для функций $\alpha\colon A \to \mathbb{R}^m$ ($A \subset \mathbb{R}^n$) и $\beta\colon B \to \mathbb{R}^k$ ($B \subset \mathbb{R}^m$) *операция суперпозиции* (*композиции*) есть составление новой *сложной* функции: $(\beta \circ \alpha)(\nu) = \beta(\alpha(\nu))$, где $\nu \in A$. При этом областью определения $\beta \circ \alpha$ служит $A \cap \alpha^{-1}(B)$.

Функция $f\colon \mathbb{R} \to \mathbb{R}$ или $y = f(x)$ называется — *действительная функция действительного переменного*. Переменная функции не обязательно

должна быть действительным числом. Функция $f\colon \mathbb{R}^n \to \mathbb{R}$ или $y = f(x^1, \ldots, x^n)$ есть *действительная функция нескольких действительных переменных* или *координатная функция*.

Последовательность $\nu_1, \nu_2, \ldots \in M$ — функция, определённая на множестве натуральных чисел: $\nu\colon \mathbb{N} \to M$. Чаще всего рассматриваются последовательности, областью определения которых является весь натуральный ряд чисел. Множество значений M последовательности может состоять из объектов любой природы. Значение $\nu(n) \in M$ называют членом (элементом) последовательности ν, имеющим номер n. Последовательность элементов множества действительных чисел $\nu\colon \mathbb{N} \to \mathbb{R}$ называется *числовая последовательность*.

Пределом последовательности называют объект, к которому члены последовательности *сходятся* (приближаются) с ростом номера.

Например, число p называется *пределом числовой последовательности* x_n: $p = \lim\limits_{n\to\infty} x_n$, если последовательность x_n при $n \to \infty$ неограниченно приближается к своему постоянному значению p. Запись $n \to \infty$ означает, что «n стремится к бесконечности». При этом говорят, что последовательность x_n стремится (сходится) к числу p. Другими словами, говорят, что *предел p последовательности x_n существует*, если для любого (сколь угодно малого) положительного ε найдется зависящее от него (достаточно большое) натуральное число N такое, что при всех $n > N$ выполняется условие $|x_n - p| < \varepsilon$.

Последовательность ν_n сходится к точке ν или

$$\nu = \lim_{n\to\infty} \nu_n, \tag{1.16}$$

если для любого $\varepsilon > 0$ существует такое натуральное число N, что $\rho(\nu_n, \nu) < \varepsilon$ для всех $n < N$.

Не у всякой последовательности существует предел. Если у последовательности существует предел, то она называется *сходящейся*, если нет — *расходящейся*.

Определение. Последовательность *абсолютно сходится*, если сходится модуль этой последовательности

$$|\nu| = \lim_{n\to\infty} |\nu_n|. \tag{1.17}$$

Если же последовательность (1.16) сходится, а последовательность (1.17) расходится, то говорят об *условной сходимости*.

Определение. *Фундаментальная последовательность* или *сходящаяся в себе последовательность* или *последовательность Коши* — последовательность ν_n точек метрического пространства такая, что для любого заданного расстояния, существует элемент последовательности, начиная с которого все элементы последовательности находятся друг от друга на расстоянии не более чем заданное. Другими словами для любого $\varepsilon > 0$ существует такое натуральное число N, что $\rho(\nu_i, \nu_j) < \varepsilon$ для всех $i, j > N$.

Всякая сходящаяся последовательность является фундаментальной, но не всякая фундаментальная последовательность сходится.

Полным пространством называется метрическое пространство, в котором сходится любая фундаментальная последовательность.

Если метрическое пространство не полно, его можно пополнить, аналогично тому, как пополняется множество рациональных чисел иррациональными до совокупности всех действительных чисел.

Запись, функция φ имеет в точке τ *предел* ξ:

$$\lim_{\nu \to \tau} \varphi(\nu) = \xi \tag{1.18}$$

означает, что $\varphi(\nu)$ можно сделать сколь угодно близким к ξ, взяв точку ν достаточно близкой к точке τ, но не совпадающей с τ.

Плотность ρ (измеримого) множества $A \subset \mathbb{R}^n$ понимается как предел (если он существует) отношения:

$$\rho = \lim_{m(P) \to 0} \frac{m(A \cap P)}{m(P)}. \tag{1.19}$$

где P — параллелепипед в $\mathbb{R}^n$,
m — мера множества.

Функции могут рассматриваться как объекты для отображений, т.е. над функциями в свою очередь производят отображения. В математике существует ряд отображений разных уровней: Отображение над числами называется *функцией*, отображение над функциями называется *оператором* (позднелат. operator — работник, исполнитель), отображение над операторами — *функтором* (лат. fungo — осуществляю, выполняю) и т.д. В результате функции и операторы, т.е. отображения, становятся объектами для отображений.

Например, над функциями $f, g\colon \mathbb{R}^n \to \mathbb{R}$ можно выполнять *четыре операции* — сложения, вычитания, умножения и деления функций:
$f + g$, $f - g$, $f \cdot g$ и f/g.

1.4 Координаты пространства

Числа в n-членной последовательности (1.4) называют *координатами* (от лат. co — совместно, и ordinatus — упорядоченный, определённый) рассматриваемой точки, а выбранные числовые оси, на которые эта точка проецируется — *координатными осями* пространства.

Системой координат пространства называют совокупность выбранных (совместно определённых) числовых осей. Систему координат удобно выбрать так, чтобы их нулевые точки совпадали. При таком совмещении нулевых точек мы получаем *начало координат*, в котором все координатные оси пересекаются. Координаты пространства должны быть *независимы*. На практике это означает, что мы можем задавать произвольные значения всех

координат. Кроме того, координаты пространства должны быть *существенны*, т.е. должны полностью определять произвольную точку пространства. На практике это означает, что количество независимых координат должно быть равно размерности рассматриваемого пространства. Независимость и существенность являются необходимым и достаточным условиями при выборе системы координат. В остальном выбор системы координат достаточно произволен.

Кроме пространства $\mathbb{R}^n$ вводится понятие *аффинного пространства* A^n, которое отличается от $\mathbb{R}^n$ тем, что в нём «не фиксировано начало координат» и все его точки абсолютно равноправны. Однако иногда аффинное пространство определяют по-другому. Так Андре Арго в своей книге «Математика для электро- и радиоинженеров» [3, с. 238] определяет аффинным пространство, координаты которого имеют разные единицы измерения. Классическим примером служит пространство термодинамической системы, в котором по координатным осям откладываются, например, давление, удельный объём и температура. Другим примером служит Мир — четырёхмерное пространство-время: $\mathbb{R} \times \mathbb{R}^3$ — прямое произведение оси времени t на трёхмерное пространство $\mathbb{R}^3$.

Мысль о многомерном пространстве выражал И. Кант (1746), а о присоединении к пространству в качестве 4-й координаты времени писал Ж. Даламбер (1764). Построение же евклидовой (см. ниже) многомерной геометрии было осуществлено А. Кэли (1843), Г. Грассманом (1844) и Л. Шлефли (1852). Первоначальные сомнения и мистика, связанные со смешением этих обобщений с физическим пространством были преодолены, и n-мерное пространство укрепилось в математике.

Смысл координат состоит в представлении точек пространства последовательностью чисел. Однако это представление может осуществляться в различных системах координат. При рассмотрении различных систем координат нам понадобятся термины из следующих глав, такие как базис (2.17), евклидово пространство (2.68), касательный вектор (с. 110).

Допустим, мы имеем трёхмерное пространство с выбранными координатами x, y, z и хотим задать в этом пространстве новую координатную систему с выбранными координатами u, v, w, которые по каким-то причинам окажутся лучше.

Во-первых, выбранная координатная система может не быть ортогональной. Такой базис называется *косоугольным*. Впрочем такие системы редко встречаются в приложениях, и мы их рассматривать не будем. Если мы выбираем новую координатную систему ортогональной, значит новое векторное пространство — *евклидово*.

Во-вторых, координатные оси не обязательно могут быть прямыми линиями. Например, мы рассматриваем функцию, заданную на какой-либо поверхности, и нам удобно, чтобы координатные оси целиком лежали на этой поверхности, но при этом оставались ортогональными (перпендикулярными друг другу в каждой точке поверхности кривыми линиями). Такие координаты с ортогональными криволинейными осями называют ортогональными *криволинейными координатами*, в отличие от ортогональных прямолиней-

ных координат, которые называют *декартовы координаты* (рис. 1.1).

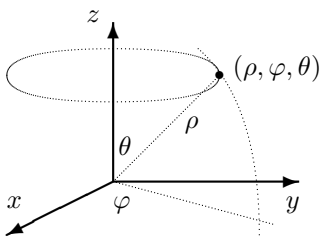


Рис. 1.1: Декартовы (x, y, z) и сферические (ρ, φ, θ) координаты.

В-третьих, выбранная координатная система может не иметь фиксированного начала координат и направления базисных векторов. Действительно, если координатные оси — кривые линии, то базисные векторы в каждой точке этих кривых есть касательные векторы. Очевидно, что в разных точках пространства эти касательные векторы будут иметь разное направление. Тогда в каждой точке пространства (данном начале координат) мы будем иметь три пересекающихся под прямым углом криволинейные координатные оси, соответствующие только этой точке, и три соответствующих координатных направления этих осей. Таким образом, наши криволинейные координат будут функцией рассматриваемой точки пространства $u, v, w = u, v, w(x, y, z)$, а орты криволинейной системы образуют базис в касательном пространстве к координатным кривым, т.е. также являются функциями рассматриваемой точки пространства. Поэтому базис криволинейной системы называют *подвижным репером*.

Таким образом, в ортогональной криволинейной системе координат числовые (координатные) прямые становятся координатными кривыми, а координатные плоскости — координатными поверхностями. Каждая координата образует семейство координатных линий (координатную поверхность). Внутри семейства координатные линии не пересекаются между собой. Кривые из разных семейств пересекаются в каждой точке пространства под прямым углом. Взятые во всей их совокупности координатные линии образуют *координатную сетку*.

Координатные поверхности, образованные семействами координатных линий, также пересекаются под прямым углом в каждой точке.

Наиболее часто используемые ортогональные криволинейные координаты связаны с декартовыми координатами x, y, z следующим образом.

На плоскости:

1. *Полярные* $x = r\cos\varphi$, $y = r\sin\varphi$, где $0 \leqslant r < \infty$, $0 \leqslant \varphi < 2\pi$. Координатные линии — концентрические окружности $r = const$ и лучи $\varphi = const$ (рис. 1.5).

2. *Параболические* $x = u^2 - v^2$, $y = 2uv$, где $-\infty < u < \infty$, $-\infty < v < \infty$. Координатные линии — две системы парабол с противоположно направленными осями, когда $u = cont$, либо $v = const$.

В пространстве:

1. *Цилиндрические* $x = r\cos\varphi$, $y = r\sin\varphi$, $z = z$, где $0 \leqslant r < \infty$, $0 \leqslant \varphi < 2\pi$, $-\infty < z < \infty$ (к полярным координатам добавляется

ось z). Координатные поверхности — круговые цилиндры $r = const$, полуплоскости $\varphi = cont$ и плоскости $z = const$. Название предложил Р. Бальтцер (1882).

2. *Сферические* $x = \rho\cos\varphi\sin\theta$, $y = \rho\sin\varphi\sin\theta$, $z = \rho\cos\theta$, где $0 \leqslant \rho < \infty$, $0 \leqslant \varphi < 2\pi$, $0 \leqslant \theta \leqslant \pi$ (рис. 1.1). Координатные поверхности — концентрические сферы ($\rho = const$), полуплоскости, проходящие через ось z ($\varphi = const$), круговые конусы с вершиной в начале координат и осью z ($\theta = const$). Формулы привёл Ж. Лагранж (1773), название предложил Р. Бальтцер (1882).

3. *Параболические* , $x = u^2 - v^2$, $y = 2uv$, $z = z$ где $-\infty < u < \infty$, $-\infty < v < \infty$, $-\infty < z < \infty$. Координатные поверхности — две системы параболических цилиндров с противоположно направленными осями, когда $u = const$, либо $v = const$ и плоскости $z = const$.

Обратим внимание, что криволинейные координаты могут иметь разные единицы измерения: величины r и ρ имеют меру длины, углы φ и θ — безразмерные величины, а u и v — корень из меры длины.

1.5 Элементарные функции

График (от греч. $\gamma\rho\alpha\varphi\iota\kappa\acute{\upsilon}\varsigma$ — начертанный) функции есть её изображение в пространстве в виде множества точек.

Графиком функции $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ называется множество

$$\{\,(\nu, \upsilon)\colon \upsilon = \varphi(\nu)\,\} \subset \mathbb{R}^n \times \mathbb{R}^m. \tag{1.20}$$

Например, для функции $f\colon A \to \mathbb{R}$, где $A \subset \mathbb{R}^2$, можно получить удобное изображение, построив её график — множество всех точек вида $\Gamma_f = \{\,(x, y, z)\colon z = f(x, y)\,\} \subset \mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$, образующее некоторую фигуру в трёхмерном пространстве.

Функция, обратная к функции $y = f(x)$, обозначается

$$x = f^{-1}(y) \quad \text{или} \quad y = f^{-1}(x). \tag{1.21}$$

Обратная функция однозначна, если исходная функция инъективна.

Если обратная функция однозначна, то $f^{-1}(f(x)) \equiv x$ при всех x из области определения функции $f(x)$ и $f(f^{-1}(x)) \equiv x$ при всех x из области определения функции $f^{-1}(x)$.

Графики функций $y = f(x)$ и $y = f^{-1}(x)$ симметричны относительно биссектрисы первого и третьего координатных углов (рис. 1.2).

Существуют словесный, графический и табличный способ задания функций. Однако наиболее распространенным способом задания функций является *аналитический способ*, при котором функция f задается *формулой* в так называемых *элементарных функциях*.

Выделяют следующие элементарные функции: степенная функция x^a и обратная к ней функция $x^{\frac{1}{a}}$, экспоненциальная функция e^x и обратная

к ней логарифмическая функция $\ln x$, тригонометрические функции $\sin x$, $\cos x$ и обратные к ним функции $\arcsin x$, $\arccos x$.

Функция $x^{\frac{1}{n}} = \sqrt[n]{x}$, где $n \in \mathbb{N}$, называется также *радикал* (от позднелат. radicalis — имеющий корни, лат. radix — корень), знак $\sqrt{}$ является изменённой латинской r, а действие называют *извлечением корня*.

Степенная функция x^2 называется *квадрат* x, а x^3 — *куб* x.

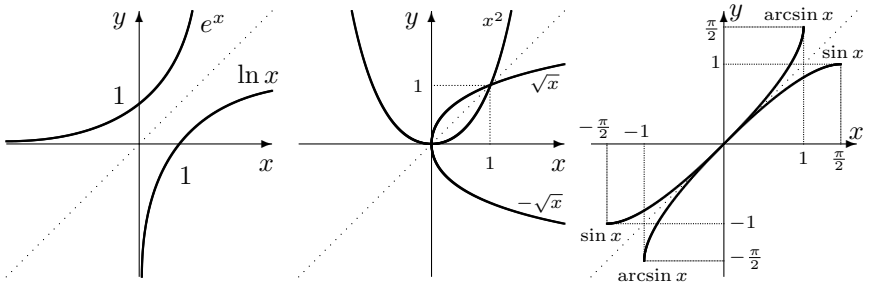


Рис. 1.2: Графики некоторых элементарных функций.

Над элементарными функциями f, g, как и над числами, можно выполнять четыре операции — сложения, вычитания, умножения и деления функций $f+g$, $f-g$, $f \cdot g$ и f/g. Если указанные операции применялись *конечное число раз*, то результат операции также относят к элементарным функциям. Например, показательная функция $a^x = e^{x \ln a}$, тангенс $\operatorname{tg} x = \frac{\sin x}{\cos x}$ и т.д., а также обратные к ним функции $\log_a x$, $\operatorname{arctg} x$ и т.д.

Чётная функция — функция $f(x)$, удовлетворяющая условию: $f(-x) = f(x)$ для всех x из области определения этой функции. Например, x^2, $\cos x$, $\ln |x|$. Сумма, разность, произведение и частное чётных функций являются чётными функциями.

Нечётная функция — функция $f(x)$, удовлетворяющая условию: $f(-x) = -f(x)$ для всех x из области определения этой функции. Например, x^3, $\sin x$, $\operatorname{tg} x$. Сумма и разность нечётных функций являются нечётными функциями, а произведение и частное двух нечётных функций являются чётными функциями. Произведение и частное чётной и нечётной функции являются нечётными функциями.

График чётной функции симметричен относительно оси ординат, а график нечётной функции — относительно начала координат.

Многочлен (полином) одного переменного n-ой степени также является элементарной функцией и имеет вид:

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0, \tag{1.22}$$

где x — действительное переменное,

$a_0, a_1, \ldots, a_n \in \mathbb{R}$ и $n \in \mathbb{N}$ — числа.

Линейная функция одного переменного x — функция вида $ax+b$, где a, b — числа. Линейная функция n переменных, или многочлен первой степени

от n переменных $x^1, \ldots, x^n$ — функция вида $a_0 + a_1 x^1 + \ldots + a_n x^n$, где $a_0, a_1, \ldots, a_n$ — числа.

К числу важнейших свойств многочленов относится *теорема Вейерштрасса*: любую непрерывную (см. ниже) функцию можно с произвольно малой ошибкой (произвольной точностью) заменить многочленом. Точная её формулировка требует, чтобы данная функция была непрерывна на компактной области определения.

Форма (лат. forma) или *однородный многочлен* — многочлен от нескольких переменных, все члены которого имеют одну и ту же степень. Под степенью одночлена $x_1^\alpha x_2^\beta \ldots x_m^\gamma$ понимают число $n = \alpha + \beta + \ldots + \gamma$. В зависимости от степени формы называют линейными $n - 1$, квадратичными $n = 2$, кубическими $n = 3$, и т.д. В зависимости от числа переменных формы называют бинарными $m = 2$ тернарными $m = 3$ и т.д.. Квадратичные формы тесно связаны с линиями и поверхностями второго порядка (см. § 3.6) и являются наиболее важными для приложений. Например: $a_1 x^1 + \ldots + a_n x^n$ — линейная форма, $xy + 2y^2 + z^2$ — тернарная квадратичная форма.

Общее число членов формы равно

$$(x_1 + \ldots + x_m)^N = m^N. \tag{1.23}$$

Очень известной бинарной формой является *биномом Ньютона*

$$(x + y)^N = x^N + C_N^1 x^{N-1} y + \ldots + C_N^n x^{N-n} y^n + \ldots + y^N, \tag{1.24}$$

где C_N^n — коэффициенты,

$n, N \in \mathbb{N}$.

Общее число членов бинома равно сумме коэффициентов:

$$\sum_{n=0}^{N} C_N^n = 2^N \tag{1.25}$$

Для вычисления биноминальных коэффициентов C_N^n нам понадобятся некоторые сведения из *комбинаторики* (от позднелат. combino — соединяю) — раздела математики, посвящённого решению задач выбора и расположения элементов некоторого, обычно конечного, множества в соответствии с определенными правилами (комбинаторику естественно считать просто теорией конечных множеств).

Перестановкой из n элементов называется конечная последовательность (кортеж) длины n, *все элементы которой различны*. Например, два различных элемента x_1, x_2 образуют две перестановки (x_1, x_2), (x_2, x_1). Количество перестановок n элементов равно:

$$n! = n \cdot (n-1) \cdot \ldots \cdot 1 = n \cdot (n-1)! \text{ причём } 0! = 1! = 1. \tag{1.26}$$

Размещением из N элементов по n называется конечная последовательность (кортеж) длины n из этих N, все элементы которой различны. Эле-

менты N принято называть ячейками (или урнами), а элементы n — частицами (или шарами). *Количество размещений* равно:

$$A_N^n = N \cdot (N-1) \cdot \ldots \cdot (N-n+1) = \frac{N!}{(N-n)!}, \text{ так } A_n^n = n!. \qquad (1.27)$$

Сочетанием из N элементов по n называется *произвольный набор* n различных элементов из N (т.е. последовательность элементов не важна). Рассмотрим все размещения из N элементов по n. Разобьем их на различные группы, в каждой из которых размещения различаются только порядком элементов, но не составом, а две различные группы отличаются хотя бы одним элементом. Очевидно, что число различных групп совпадает с числом сочетаний из N элементов по n. Число размещений внутри группы из n элементов равно числу перестановок, т.е. $n!$. *Количество сочетаний* равно:

$$C_N^n = \frac{A_N^n}{n!} = \frac{N!}{n!(N-n)!} = \frac{A_N^n}{A_n^n}. \qquad (1.28)$$

Биноминальные коэффициенты в (1.24) равны числу сочетаний.

Правило для биномиальных коэффициентов:

$$C_N^n = C_{N-1}^{n-1} + C_{N-1}^n. \qquad (1.29)$$

Рациональной дробью называется отношение двух алгебраических многочленов $\frac{P_n(x)}{Q_m(x)}$. Рациональная дробь называется *правильной*, если $n < m$, и *неправильной* в противном случае.

Многочлен $P_n(x)$ есть не только функция переменной x, но и функция переменной n, т.е. последовательность: $(P_1, \ldots, P_2, \ldots)$.

Бесконечный многочлен образует «*ряд степенных функций*» или «*функциональный ряд*». Степенные функции $a_n x^n$ называются *членами ряда*. Многочлен $P_n(x)$ степени n называется *частичной суммой ряда* порядка n. Ряд называется *сходящимся* (*ряд сходится*), если существует предел последовательности его частичных сумм $P_n(x)$.

Например:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots = \sum_{n=0}^{\infty} \frac{x^n}{n!} \text{ или } e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n. \quad (1.30)$$

Таким образом, функцию e^x можно определить не только как ряд, но и как предел последовательности.

Важнейшим средством изучения, изображения и приближенного вычисления функции является представление функции в виде *бесконечного многочлена* или *ряда степенных функций*.

Другие элементарные функции также «разлагаются в ряд»:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots + (-1)^n \frac{x^{2n}}{(2n)!} + \ldots = \sum_{n=1}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}, \qquad (1.31)$$

$$\sin x = \frac{x}{1!} - \frac{x^3}{3!} + \ldots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \ldots = \sum_{n=1}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}, \qquad (1.32)$$

для произвольных действительных переменных x. И

$$\frac{1}{1-x} = 1 + x + x^2 + \ldots + x^n + \ldots, \qquad \text{для } x^2 < 1; \tag{1.33}$$

$$\ln(1+x) = x - \frac{x^2}{2} + \ldots + (-1)^{n+1}\frac{x^n}{n} + \ldots, \qquad \text{для } -1 < x \leqslant 1; \tag{1.34}$$

$$\ln(1-x) = -\left[x + \frac{x^2}{2} + \ldots + \frac{x^n}{n} + \ldots\right], \qquad \text{для } -1 < x \leqslant 1; \tag{1.35}$$

$$\operatorname{arctg} x = x - \frac{x^3}{3} + \ldots + (-1)^n \frac{x^{2n+1}}{2n+1} + \ldots, \qquad \text{для } x^2 < 1. \tag{1.36}$$

Не все функции разлагаются в ряды. Функция, которую можно представить в виде бесконечного ряда, называется *аналитической*.

Элементарные функции, стоящие слева в (1.30–1.36), непрерывные аналитические функции, а стоящие справа члены ряда образуют последовательность, т.е. упорядоченное дискретное множество. Таким образом, непрерывные функции можно изучать, разлагая в дискретные ряды, а дискретные последовательности можно изучать при помощи непрерывных функций. В общем случае говорят, что *математическое описание мира построено на тонкой игре непрерывного и дискретного*. Интуитивно идея непрерывности выражает коренное свойство пространства и времени и имеет, следовательно, фундаментальное значение для познания. Дискретное более заметно.

Способ изучения функций при помощи бесконечных рядов открыл Ньютон. Именно это открытие, а вовсе не дифференциальное и интегральное исчисление или закон всемирного тяготения, Ньютон считал самым важным своим достижением и именно его закодировал в письме к Лейбницу 24 октября 1676 года, в котором описал анализ. Ньютон понимал под *анализом* исследование уравнений при помощи бесконечных рядов. Основное открытие Ньютона, иными словами, заключается в том, что функции раскладываются в бесконечные степенные ряды, ряды подставляются друг в друга, приравниваются коэффициенты при одинаковых степенях, и таким образом один за другим находятся коэффициенты неизвестной функции.

Уравнение — аналитическая запись задачи о разыскании значений неизвестных переменных, при которых значения двух данных функций равны. Неизвестные переменные при этом называют *аргументами*, а значения аргументов, при которых значения функций равны, — *решениями* (*корнями*) уравнения. Пример уравнения: $y = f(x)$, где y — известное значение функции.

Совокупность уравнений, для которых требуется найти значения аргументов, удовлетворяющих одновременно всем этим уравнениям, называется *системой уравнений*. Значения аргументов, удовлетворяющих одновременно всем уравнениям системы — решениями системы.

Например, система n уравнений с n неизвестными имеет вид:

$$\begin{cases} y^1 = f_1(x^1, \ldots, x^n), \\ \quad \ldots\ldots\ldots\ldots\ldots \\ y^n = f_n(x^1, \ldots, x^n). \end{cases} \tag{1.37}$$

Наиболее изучены *алгебраические уравнения*, для которых функции f являются многочленами P_k:

$$P_k(x^1, \ldots, x^n) = 0. \tag{1.38}$$

Функции, удовлетворяющие алгебраическому уравнению, называются *алгебраическими* (например, содержащие только элементарные степенные функции). Аналитические функции, не являющие алгебраическими, называют *трансцендентными* (например, показательные, логарифмические, тригонометрические). Соответственно, числа, не являющиеся корнями никакого многочлена с целыми коэффициентами, называются *трансцендентными* (например, числа π и e).

Одно из самых простых и важных — *квадратное уравнение*:

$$ax^2 + 2bx + c = 0, \text{ где } a \neq 0. \tag{1.39}$$

Решения (корни) квадратного уравнения (алгебраического уравнения 2-й степени) записываются в виде:

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - ac}}{a}. \tag{1.40}$$

Если *дискриминант* $D = ac - b^2 < 0$ отрицателен, уравнение имеет два различных решения. В этом случае уравнение можно разложить на сомножители первой степени $ax^2 + 2bx + c = a(x - x_1)(x - x_2)$, а график функции пересекает ось абсцисс x в двух точках, и функция $y(x) = ax^2 + 2bx + c$ меняет знак при различных значениях x.

При $D = 0$ уравнение имеет два одинаковых решения $x_{1,2} = -\frac{b}{a}$ или как говорят — *корень кратности два*. В этом случае график целиком расположен в верхней или нижней полуплоскости относительно оси абсцисс x, а функция $y(x)$ не меняет знак за исключением одной точки $-\frac{b}{a}$, где $y = 0$ и график касается оси абсцисс x.

При $D > 0$ уравнение не имеет действительных решений (корней). В этом случае график также целиком расположен в одной полуплоскости, и функция $y(x)$ не меняет знак при любых значениях x.

1.6 Функции и непрерывность

Пусть $\varphi\colon A \to \mathbb{R}^m$, где $A \subset \mathbb{R}^n$, и пусть $\nu, \tau \in A$.

Функция $\varphi\colon A \to \mathbb{R}^m$ называется *непрерывной в точке* τ, если

$$\lim_{\nu \to \tau} \varphi(\nu) = \varphi(\tau). \tag{1.41}$$

Другими словами для всякого числа $\varepsilon > 0$ существует такое число $\delta > 0$, что $|\varphi(\nu) - \varphi(\tau)| < \varepsilon$ для всех ν, удовлетворяющих условию $|\nu - \tau| < \delta$.

В терминах окрестности точки определение непрерывности выглядит так: Отображение $\varphi\colon A \to \mathbb{R}^m$ непрерывно в точке ν, если для любой окрестности V точки $\varphi(\nu) \in \mathbb{R}^m$ существует такая окрестность U точки ν, что $\varphi(U) \subset V$.

Понятие непрерывности функции исторически раньше подверглось математической обработке, чем логически предшествующее ему понятие непрерывности множества. Идея непрерывности состоит в том, что близкие друг к другу точки отображаются в близкие точки. Непрерывная функция, это функция, получающая бесконечно малые (по абсолютной величине меньше любого наперёд заданного числа) приращения при бесконечно малых приращениях аргумента.

Функция $\varphi\colon A \to \mathbb{R}^m$ называется (просто) *непрерывной*, если она непрерывна в каждой точке $\tau \in A$.

Если функция φ непрерывна и множество A компактно, то множество $\varphi(A)$ компактно. Множество связно (см. § 1.1) тогда и только тогда, когда всякая непрерывная функция переходит от одного значения к другому, принимая все промежуточные значения.

Точка разрыва — точка, в которой функция не является непрерывной. Например, если функция определена на множестве A и не определена в точке τ, но точка $\tau \notin A$ — предельная точка (см. § 1.1) множества A, то τ — точка разрыва. Другой случай: если $\tau \in A$, но $\lim\limits_{\nu\to\tau} \varphi(\nu) \neq \varphi(\tau)$ (в частности, если этот предел не существует), то τ — точка разрыва.

Если предел $\lim\limits_{\nu\to\tau} \varphi(\nu) = \xi$ существует, то τ — устранимая точка разрыва. Когда $\tau \notin A$ и $\varphi(\tau)$ не определено, мы можем определить $\varphi(\nu) = \xi$, а когда $\tau \in A$, но $\lim\limits_{\nu\to\tau} \varphi(\nu) \neq \varphi(\tau)$, определить $\lim\limits_{\nu\to\tau} \varphi(\nu) = \varphi(\tau)$.

Например, если функция одного переменного $f\colon \mathbb{R} \to \mathbb{R}$ или $f(x)$ имеет в точке a два односторонних предела справа и слева:

$$\lim_{x\to a-0} f(x) = f(a-0), \qquad \lim_{x\to a+0} f(x) = f(a+0), \tag{1.42}$$

тогда точка a — *разрыв первого рода*, а разность $f(a+0) - f(a-0)$ называется *скачок функции в точке* a. Если скачок равен нулю, то a — устранимая точка разрыва или *излом*.

Если функция $f(x)$ определена в окрестности точки a, за исключением, быть может, самой a, и хотя бы один из односторонних пределов не существует, тогда точка a называется *разрыв второго рода*. Например, для функций $\frac{1}{x}$, $\sin\frac{1}{x}$ точка $x = 0$ — разрыв второго рода.

Примеры точек разрыва функции показаны на рис. 1.3.

Особое значение имеют те непрерывные отображения, для которых существуют непрерывные обратные отображения. Пусть U и V — открытые множества в $\mathbb{R}^n$. Непрерывная взаимно однозначная функция $\varphi\colon U \to V$, имеющая обратную непрерывную функцию $\varphi^{-1}\colon V \to U$, есть *гомеоморфизм* (от греч. *όμοιος* — подобный и *μορφη* — форма, образ), т.е. если вза-

имно однозначное отображение (изоморфизм) между двумя множествами непрерывно в обе стороны, то такое отображение есть гомеоморфизм, а эти множества называются *гомеоморфными*, или топологически эквивалентными (топологически неразличимыми).

Главной задачей топологии является выделение и изучение топологических свойств пространств или *топологических инвариантов*, т.е. свойств, сохраняющихся любыми гомеоморфизмами одного топологического пространства на другое.

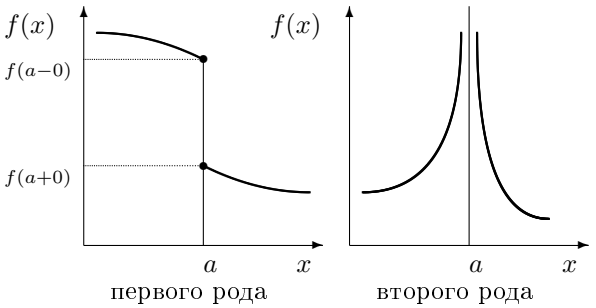


Рис. 1.3: Точки разрыва функции

К числу важнейших, уже известных нам топологических инвариантов относятся связность, компактность и размерность.

Кривая или *линия* — геометрическое понятие, определяемое по разному в разных разделах геометрии, есть непрерывное отображение отрезка в пространство

$$C\colon [0,1] \to \mathbb{R}^n. \tag{1.43}$$

Отображение $I^n\colon [0,1]^n \to \mathbb{R}^n$, где $[0,1]^n = [0,1] \times \ldots \times [0,1]$ называется *стандартный n-мерный куб*, если $I^n(x^1,\ldots,x^n) = (x^1,\ldots,x^n)$ или $I^n(\mathbf{x}) = \mathbf{x}$ для всех $\mathbf{x} \in [0,1]^n$. Например, стандартный нульмерный куб есть число нуль $I^0 = [0,1]^0 = \{0\}$. Обычно $[0,1]^0$ обозначается символом $\{0\}$. Стандартный одномерный куб есть единичный отрезок $I^1 = [0,1]$, стандартный двумерный куб есть единичный квадрат $I^2 = [0,1]^2$ и т.д.

Сингулярный m-мерный куб в A (от лат. singularis — отдельный, особый) есть непрерывная функция $C\colon [0,1]^m \to A$, где A — открытое множество в $\mathbb{R}^n$ с $n \geqslant m$.

Так как функция C непрерывна, а I^m компактен, то образ $C\,(I^m) \subset A$ — компактен. Таким образом, сингулярные кубы представляют собой компактные множества в $\mathbb{R}^n$.

Сингулярный нульмерный куб в A есть функция $C\colon \{0\} \to A$ или $C(0) = (x^1,\ldots,,x^n)$, т.е. точка в A. Сингулярный одномерный куб $C\colon [0,1] \to A$ или $C(s) = \big(x^1(s),\ldots,x^n(s)\big)$ для всех чисел $s \in [0,1]$ есть *компактная кривая* в A. Сингулярный двумерный куб $C\colon [0,1]^2 \to A$ или $C(u,v) = \big(x^1(u,v),\ldots,x^n(u,v)\big)$ для всех пар чисел $u,v \in [0,1]^2$ — *компактная поверхность* в A. Сингулярный m-мерный куб есть *m-мерная компактная поверхность* в $A \subset \mathbb{R}^n$.

Если для всякой точки ν множества $M \subset \mathbb{R}^n$ существует окрестность U, гомеоморфизм $\varphi\colon U \to V$ и выполняется условие:

$$\varphi(U \cap M) = V \cap (\mathbb{R}^m \times \{\,0\,\}) = \{\,\nu \in V\colon x^{m+1} = \ldots = x^n = 0\,\},$$

то M называется *m-мерная поверхность*. Если для некоторых точек ν это условие не выполняется, но выполняется дополнительное условие:

$$\varphi(U \cap M) = V \cap (\mathbb{H}^m \times \{0\}) = \{\nu \in V \colon x^m \geqslant 0 \text{ и } x^{m+1} = \ldots = x^n = 0\},$$

то M называется *m-мерная поверхность с краем*.

Любая точка m-мерной поверхности допускает взаимно однозначное и непрерывное в обе стороны отображение либо на $\mathbb{R}^m$, либо на $\mathbb{H}^m$. Другими словами, окрестность любой точки m-мерной поверхности $U \cap M$ «с точностью до гомеоморфизма» есть либо часть пространства $\mathbb{R}^m \times \{0\}$, либо часть полупространства $\mathbb{H}^m \times \{0\}$.

Важно заметить, что эти условия не могут одновременно выполняться для одной и той же точки ν. Множество всех точек $\nu \in M$, удовлетворяющих второму условию, называется *краем поверхности M* и обозначается ∂M. Как для первого, так и для второго условия $m < n$. Если $m = n$, то M есть не поверхность, а *множество с краем ∂M*.

Гиперповерхностью называется $(n-1)$-мерная поверхность в n-мерном пространстве.

Верхняя грань некоторого множества X действительных чисел обозначается $\sup X$ (от лат. supremum — наивысшее) — наименьшее число b, ограничивающее сверху это множество. На математическом языке: если для каждого $x \in X$ справедливо $x \leqslant b$ и, каково бы ни было $b' < b$, существует такое $x' \in X$, что $x' > b'$.

Нижняя грань некоторого множества X действительных чисел обозначается $\inf X$ (от лат. infimum — наинизшее) — наибольшее число a, ограничивающее его снизу. На математическом языке: если для каждого $x \in X$ справедливо $x \geqslant a$ и, каково бы ни было $a' > a$, существует такое $x' \in X$, что $x' < a'$.

Например: $\inf[a,b) = a$, $\inf(a,b) = a$, $\sup[a,b] = b$, $\sup(a,b) = b$.

Ограниченная функция — функция, множество значений которой ограничено. В данном случае имеется в виду множество значений на числовой оси. Пусть функция $f \colon A \to \mathbb{R}$, где $A \subset \mathbb{R}^n$, ограничена в точке $\tau \in A$, тогда для любого $\delta > 0$ положим:

$$M(\tau, f, \delta) = \sup\{f(\nu) \colon \nu \in A \text{ и } |\nu - \tau| < \delta\} \quad \text{— верхняя грань } f, \tag{1.44}$$

$$m(\tau, f, \delta) = \inf\{f(\nu) \colon \nu \in A \text{ и } |\nu - \tau| < \delta\} \quad \text{— нижняя грань } f. \tag{1.45}$$

Фигурные скобки определяют множество, в данном случае множество значений функции в некоторой окрестности точки τ.

Если функция не непрерывна в точке, но при этом ограничена, то степень её отклонения от непрерывности можно точно измерить.

Колебание $o(f,\tau)$ функции f в точке τ определяется формулой:

$$o(f,\tau) = \lim_{\delta \to 0} \left[M(\tau, f, \delta) - m(\tau, f, \delta) \right]. \tag{1.46}$$

Этот предел всегда существует, так как $M - m$ убывает при убывании величины δ.

С колебанием связаны два важных факта:

1. Ограниченная функция f непрерывна в точке τ тогда и только тогда, когда $o(f,\tau)=0$.

2. Пусть $A\subset\mathbb{R}^n$ — замкнутое множество. Тогда для всякой ограниченной функции $A\colon A\to\mathbb{R}$ и всякого $\varepsilon>0$ множество $\{\,\nu\in A\colon o(f,\nu)\geqslant\varepsilon\,\}$ замкнуто.

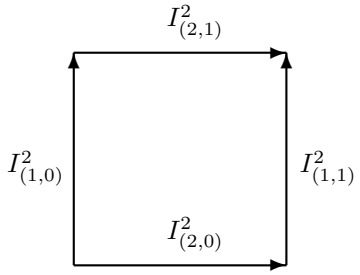
Рассмотрим более подробно понятие *границы множества*.

Границу множества не следует путать с его краем. Например, у открытого множества $A=\{\,\nu\in\mathbb{R}^n\colon |\nu|<1 \text{ или } 1<|\nu|<2\,\}$, граница есть множество $\partial A=\{\,\nu\in\mathbb{R}^n\colon |\nu|=1 \text{ или } |\nu|=2\,\}$. Объединение A с его границей ∂A есть замкнутое множество $M=\{\,\nu\in\mathbb{R}^n\colon |\nu|\leqslant 2\,\}$. Край множества $\partial M=\{\,\nu\in\mathbb{R}^n\colon |\nu|=2\,\}$ не совпадает с границей ∂A.

Определим теперь границу множества через его *грани*.

Грани единичного отрезка $I^1=[0,1]$ есть точки 0 и 1.

Грани единичного квадрата $I^2=[0,1]^2$ показаны на рис. 1.4.



$$I^2_{(1,0)}(x)=(0,x^2),$$
$$I^2_{(1,1)}(x)=(1,x^2),$$
$$I^2_{(2,0)}(x)=(x^1,0),$$
$$I^2_{(2,1)}(x)=(x^1,1).$$

Рис. 1.4: Грани единичного квадрата

Грани стандартного трехмерного куба $I^3=[0,1]^3$ представляют собой шесть единичных ориентированных в пространстве квадратов.

Граница всех этих многомерных кубов $I^n=[0,1]^n$ представляет собой совокупность их граней. Задача заключается в том, чтобы записать эту совокупность неким формальным образом.

Для каждого индекса i от 1 до n стандартного n-мерного куба I^n определим два $(n-1)$-мерных куба $I^n_{(i,0)}$ и $I^n_{(i,1)}$ следующим образом. Для каждого $\nu=(x^1,\ldots,x^{n-1})\in[0,1]^{n-1}$ положим:

$$\begin{aligned} I^n_{(i,0)}(\nu)&=I^n(x^1,\ldots,x^{i-1},0,x^i,\ldots,x^{n-1})=\\ &=(x^1,\ldots,x^{i-1},0,x^i,\ldots,x^{n-1}),\\ I^n_{(i,1)}(\nu)&=I^n(x^1,\ldots,x^{i-1},1,x^i,\ldots,x^{n-1})=\\ &=(x^1,\ldots,x^{i-1},1,x^i,\ldots,x^{n-1}). \end{aligned} \tag{1.47}$$

Назовём $I^n_{(i,0)}$ $(i,0)$-гранью I^n, а $I^n_{(i,1)}$ $(i,1)$-гранью и положим, что гра-

ница ∂I^n имеет вид:

$$\partial I^n = \sum_{i=1}^{n} \sum_{\alpha=0,1} (-1)^{i+\alpha} I^n_{(i,\alpha)}. \tag{1.48}$$

Примеры:

1. Граница единичного отрезка из двух точек 0 и 1 имеет вид:

$$\partial I^1 = -I^1_{(1,0)} + I^1_{(1,1)}. \tag{1.49}$$

2. При обходе единичного квадрата против часовой стрелки, отрезки с противоположным направлением берутся со знаком минус. Тогда граница стандартного квадрата есть объединение четырех граней (единичных отрезков) со следующими коэффициентами:

$$\partial I^2 = -I^2_{(1,0)} + I^2_{(1,1)} + I^2_{(2,0)} - I^2_{(2,1)}. \tag{1.50}$$

3. Граница стандартного трехмерного куба будет состоять из шести граней (единичных ориентированных квадратов), взятых с соответствующими коэффициентами:

$$\partial I^3 = -I^3_{(1,0)} + I^3_{(1,1)} + I^3_{(2,0)} - I^3_{(2,1)} - I^3_{(3,0)} + I^3_{(3,1)}. \tag{1.51}$$

Теперь перейдём от стандартного куба к произвольному компактному множеству. Для произвольного компактного n-мерного множества (сингулярного n-мерного куба) сначала определим (i,α)-грань:

$$C_{(i,\alpha)} = C \circ \left(I^n_{(i,\alpha)}(\nu) \right) \tag{1.52}$$

и затем положим его границей выражение:

$$\partial C = \sum_{i=1}^{n} \sum_{\alpha=0,1} (-1)^{i+\alpha} C_{(i,\alpha)}. \tag{1.53}$$

Например, граница двумерного круга есть одномерная окружность, граница трехмерного шара есть двумерная сфера, граница n-мерной объёмной фигуры есть $(n-1)$-мерная поверхность фигуры и т.д.

Объединение сингулярных m-мерных кубов называется *сингулярная m-мерная цепь* в A и записывается как их формальная сумма:

$$C = \sum a_i C_i, \tag{1.54}$$

где коэффициенты a_i — целые числа.

Цепи появились в связи с тем, что непрерывное множество может состоять из нескольких кусков, которые имеют «самопересечения», «складки» и т.д. В этом случае каждый кусок является сингулярным кубом, а произвольное компактное множество или *фигура* представляет собой сингулярную цепь.

Наконец, определим границу сингулярной n-мерной цепи:

$$\partial C = \partial\left(\sum a_i C_i\right) = \sum a_i \partial C_i. \tag{1.55}$$

Граница сингулярной n-мерной цепи есть $(n-1)$-мерная цепь, т.е. граница n-мерного множества есть его $(n-1)$-мерная поверхность.

Пусть задано преобразование $\psi\colon \mathbb{R}^n \to \mathbb{R}^n$, тогда

$$\partial(\psi C) = \psi(\partial C). \tag{1.56}$$

Теорема: $\partial(\partial C) = 0$ для всякой сингулярной n-мерной цепи в A.

Коротко, $\partial^2 = 0$.

Доказательство. Пусть $i \leqslant j$. Если $\nu = (x^1, \ldots, x^{n-2}) \in [0,1]^{n-2}$, то

$$\begin{aligned}\left(I^n_{(i,\alpha)}\right)_{(j,\beta)}(\nu) &\overset{(1.47)}{=} I^n_{(i,\alpha)}\left(I^{n-1}_{(j,\beta)}(\nu)\right) = \\ &= I^n_{(i,\alpha)}(x^1, \ldots, x^{j-1}, \beta, x^j, \ldots, x^{n-2}) = \\ &= I^n(x^1, \ldots, x^{i-1}, \alpha, x^i, \ldots, x^{j-1}, \beta, x^j, \ldots, x^{n-2}).\end{aligned}$$

Аналогично получаем

$$\begin{aligned}\left(I^n_{(j+1,\beta)}\right)_{(i,\alpha)}(\nu) &= I^n_{(j+1,\beta)}\left(I^{n-1}_{(i,\alpha)}(\nu)\right) = \\ &= I^n_{(j+1,\beta)}(x^1, \ldots, x^{i-1}, \alpha, x^i, \ldots, x^{n-2}) = \\ &= I^n(x^1, \ldots, x^{i-1}, \alpha, x^i, \ldots, x^{j-1}, \beta, x^j, \ldots, x^{n-2}).\end{aligned}$$

Таким образом, $(I^n_{(i,\alpha)})_{(j,\beta)} = (I^n_{(j+1,\beta)})_{(i,\alpha)}$ при $i \leqslant j$.

Следовательно, $(C_{(i,\alpha)})_{(j,\beta)} = (C_{(j+1,\beta)})_{(i,\alpha)}$ при $i \leqslant j$ для любого сингулярного n-мерного куба C. Но

$$\partial(\partial C) = \partial\left(\sum_{i=1}^{n}\sum_{\alpha=0,1} C_{(i,\alpha)}\right) = \sum_{i=1}^{n}\sum_{\alpha=0,1}\sum_{j=1}^{n-1}\sum_{\beta=0,1}(-1)^{i+\alpha+j+\beta}\left(C_{(i,\alpha)}\right)_{(j,\beta)}.$$

В эту сумму $(C_{(i,\alpha)})_{(j,\beta)}$ и $(C_{(j+1,\beta)})_{(i,\alpha)}$ входят с противоположными знаками, потому все члены попарно взаимно уничтожаются, и

$$\partial(\partial C) = 0. \tag{1.57}$$

Например, для границы стандартного квадрата (1.50) получим:

$$\begin{aligned}\partial^2 I^2 = &\left(I^2_{(1,0)}\right)_{(1,0)} - \left(I^2_{(1,0)}\right)_{(1,1)} - \left(I^2_{(1,1)}\right)_{(1,0)} + \left(I^2_{(1,1)}\right)_{(1,1)} - \\ &- \left(I^2_{(2,0)}\right)_{(1,0)} + \left(I^2_{(2,0)}\right)_{(1,1)} + \left(I^2_{(2,1)}\right)_{(1,0)} - \left(I^2_{(2,1)}\right)_{(1,1)} = \\ &= I^2_{(0,0)} - I^2_{(0,1)} - I^2_{(1,0)} + I^2_{(1,1)} - I^2_{(0,0)} + I^2_{(0,1)} + I^2_{(1,0)} - I^2_{(1,1)} = 0.\end{aligned} \tag{1.58}$$

Поскольку теорема верна для всякого сингулярного n-мерного куба, она верна также для любой сингулярной n-мерной цепи.

Например, граница объёмной фигуры, которая есть поверхность фигуры, сама границы не имеет (окружность или контур (замкнутая кривая), сфера и т.д.)

Циклом называется цепь C, которая не имеет границы ($\partial C = 0$).

1.7 Комплексные числа и функции

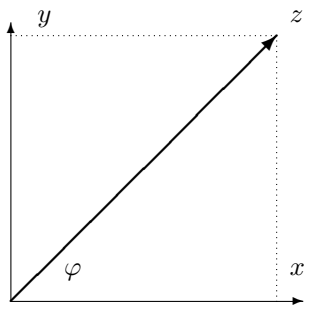


Рис. 1.5: Комплексное число на плоскости

Одной из неразрешимых проблем математики действительных чисел является вычисление корня из отрицательного числа, например, $\sqrt{-\alpha}$, либо вычисление логарифма отрицательного числа, например $\ln(-\alpha)$, где $\alpha > 0$. В этом случае говорят, что таких величин для элементарных функций, заданных на множестве вещественных чисел, не существует. Также оказывается невозможным находить решения тригонометрических уравнений вида $\cos\varphi = \alpha$, где α — произвольное число.

Для решения этих проблем в математике вводится *мнимая единица*, которая обозначается через i и определяется условием

$$i^2 = -1. \tag{1.59}$$

Комплексными числами называют величины вида

$$z = x + iy \quad \text{или} \quad z = (x, y), \tag{1.60}$$

где x и y вещественные числа.

Если действительное число изображают точкой на числовой прямой $\mathbb{R}$, то комплексное число можно изображать точкой $z = (x, y)$ на числовой плоскости $\mathbb{R}^2$. При этом по оси абсцисс откладывают величину x или *вещественную часть* $x = \Re(z)$, а по оси кординат — величину y или *мнимую часть* $y = \Im(z)$ (рис. 1.5).

Мы можем рассматривать вещественную плоскость $\mathbb{R}^2$, которую называют *множество комплексных чисел*, как *комплексную прямую* $\mathbb{C}$.

Вместо того, чтобы говорить о числе или точке z плоскости, можно говорить о направленном отрезке на плоскости, начало которого совпадает с началом координат, конец — с точкой z, а x и y — проекции отрезка на оси координат. Длина отрезка равна расстоянию r от начала координат до рассматриваемой точки, а его направление характеризуется углом φ между отрезком и осью x. В результате получаем *тригонометрическую форму* записи комплексного числа:

$$z = r(\cos\varphi + i\sin\varphi) = (r\cos\varphi, r\sin\varphi), \tag{1.61}$$

где $r = |z|$ — *модуль* комплексного числа, равный длине отрезка,

$\varphi = \arg z$ — *аргумент* комплексного числа.

Эйлер впервые заметил, что из выражений (1.30, 1.31, 1.32, 1.61) получаются формулы, которые теперь так и называют *формулы Эйлера*:

$$e^{ix} = \cos x + i \sin x, \quad e^{-ix} = \cos x - i \sin x. \tag{1.62}$$

Следовательно,

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}, \quad \sin x = \frac{e^{ix} - e^{-ix}}{2i}. \tag{1.63}$$

Наряду с классической и тригонометрической получаем *показательную форму* записи комплексного числа

$$z = r \cdot e^{i\varphi} = r(\cos \varphi + i \sin \varphi) = x + iy. \tag{1.64}$$

Равенство двух комплексных чисел равносильно двум вещественным равенствам. Знаками неравенства комплексные числа соединять нельзя.

Для комплексных чисел можно ввести арифметические операции:

$z_1 \pm z_2 = (x_1 \pm x_2, y_1 \pm y_2)$ — сумма и разность,

$z_1 \cdot z_2 = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1) = r_1r_2\, e^{i(\varphi_1+\varphi_2)} =$
$= (r_1r_2 \cos(\varphi_1 + \varphi_2), r_1r_2 \sin(\varphi_1 + \varphi_2))$ — произведение,

$\dfrac{1}{z} = \left(\dfrac{x}{x^2+y^2}, \dfrac{y}{x^2+y^2}\right) = \dfrac{e^{-i\varphi}}{r} = \left(\dfrac{\cos\varphi}{r}, \dfrac{\sin\varphi}{r}\right)$ — обратное число,

$\dfrac{z_1}{z_2} = \left(\dfrac{x_1x_2 + y_1y_2}{x_2^2+y_2^2}, \dfrac{x_2y_1 - x_1y_2}{x_2^2+y_2^2}\right) = \dfrac{r_1}{r_2}\, e^{i(\varphi_1-\varphi_2)} =$
$= \left(\dfrac{r_1}{r_2}\cos(\varphi_1 - \varphi_2), \dfrac{r_1}{r_2}\sin(\varphi_1 - \varphi_2)\right)$ — частное,

$z^n = r^n e^{in\varphi} = (r^n \cos n\varphi, r^n \sin n\varphi)$ — возведение в степень,

$\sqrt[n]{z} = \sqrt[n]{r} e^{i\frac{\varphi+2k\pi}{n}}$, где $k = 0, 1, \ldots, n-1$ — извлечение корня.

К этим правилам для комплексных чисел надо добавить *принцип аналитического продолжения*, согласно одной из версий которого для функций, определенных с помощью степенных рядов, любое тождество, справедливое для всех вещественных значений независимых переменных, остается справедливым и для их комплексных значений. Например, из тождества $\cos^2 x + \sin^2 x \equiv 1$ следует $\cos^2 ix + \sin^2 ix \equiv 1$ и т.п.

Экспонентой комплексного числа $z = x + iy$ называется число

$$e^z \overset{(1.62)}{=} e^x(\cos x + i \sin x) \overset{(1.30)}{=} \lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n. \tag{1.65}$$

Возведение экспоненты в мнимую степень означает поворот единичного отрезка на соответствующий угол. Следовательно, при мнимых степенях

показательная функция $e^{i\varphi}$ становится периодической. Поворачивая единичный отрезок на различные углы, получаем:

$$1 = e^{i2\pi n}, \quad i = e^{i\left(\frac{\pi}{2}+2\pi n\right)}, \quad -1 = e^{i(\pi+2\pi n)}, \quad -i = e^{i\left(\frac{3}{2}\pi+2\pi n\right)}. \tag{1.66}$$

Логарифм комплексного числа записывается в виде:

$$\ln z = \ln r + i(\varphi + 2\pi n) \tag{1.67}$$

и имеет бесчисленное множество значений.

Из выражений (1.66) получаем:

$$\ln 1 = i2\pi n, \quad \ln i = i(\pi/2 + 2\pi n),$$
$$\ln(-1) = i(\pi + 2\pi n), \quad \ln(-i) = i(3\pi/2 + 2\pi n).$$

Уравнение с тригонометрическими функциями вида $\cos\varphi = \alpha$ решается исходя из уравнений (1.63): $(e^{i\varphi})^2 - 2\alpha e^{i\varphi} + 1 = 0$, тогда $e^{i\varphi} = \alpha \pm \sqrt{\alpha^2 - 1}$. Замечая, что $\alpha - \sqrt{\alpha^2 - 1} = 1/(\alpha + \sqrt{\alpha^2 - 1})$, т.е. $\ln(\alpha - \sqrt{\alpha^2 - 1}) = -\ln(\alpha + \sqrt{\alpha^2 - 1})$, получаем решение

$$\varphi = \pm i \ln(\alpha + \sqrt{\alpha^2 - 1}) + 2\pi n. \tag{1.68}$$

Решая аналогичным способом уравнение $\sin\varphi = \alpha$, либо просто замечая, что $\sin(\varphi + \pi/2) = \cos\varphi$, получим:

$$\varphi = \pm i \ln(\alpha + \sqrt{\alpha^2 - 1}) + \pi(2n + 1). \tag{1.69}$$

Два комплексных числа, отличающихся только знаком мнимой части, называют сопряженными:

$$z = x + iy = re^{i\varphi}; \quad \overline{z} = z^* = x - iy = re^{-i\varphi}. \tag{1.70}$$

Комплексные z и комплексно сопряжённые числа ($\overline{z}$ или z^*) на плоскости всегда располагаются симметрично относительно действительной оси ординат. Любое равенство, связывающее комплексные числа, остается справедливым, если эти числа заменить комплексно сопряжёнными с ними. Квадрат модуля комплексного числа можно вычислить, умножив его на комплексно сопряжённое:

$$z \cdot \overline{z} = r^2 = x^2 + y^2. \tag{1.71}$$

В отличие от многочлена вещественного числа (1.22), *любой* многочлен n-ой степени комплексного числа z можно разложить на сомножители 1-й степени:

$$P(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0 = a_n(z - z_1)\ldots(z - z_n), \tag{1.72}$$

где $a_0, a_1, \ldots, a_n \in \mathbb{R}, \quad a_n \neq 0$.

Числа $z_1, \ldots, z_n$ называются *нулями* или *корнями многочлена* $P(z)$. Этот результат аналогичен утверждению, что, в отличие от вещественных чисел, *любое алгебраическое уравнение комплексного числа степени n имеет n корней.*

Например, $x^3 = 1$ имеет корни $x_1 = 1, x_2 = -\frac{1}{2} + i\frac{\sqrt{3}}{2}, x_3 = -\frac{1}{2} - i\frac{\sqrt{3}}{2}$, в то время как нахождение числа вещественных корней нелегкая задача. Комплексные корни в алгебраическом уравнении всегда встречаются парами — комплексное число и его комплексно сопряженное.

Множество комплексных корней алгебраического уравнения называется множеством *алгебраических чисел.* В отличие от всего множества комплексных чисел, множество алгебраических чисел — счётное.

Таким образом, оказывается, что все нули многочлена (корни алгебраического уравнения) располагаются на плоскости следующим образом. Действительные корни лежат на действительной числовой оси. Комплексные корни располагаются симметричными парами относительно действительной оси. Половина комплексных корней лежит в верхней полуплоскости, половина — в нижней полуплоскости, которая соответствует комплексно сопряженным числам.

Отметим, что из формулы Эйлера (1.62) легко получить множество формул, требующих, при стандартном выводе, большое количество действий. Например:

$$\cos(\varphi + \psi) + i\sin(\varphi + \psi) = e^{i(\varphi+\psi)} =$$
$$= e^{i\varphi}e^{i\psi} = (\cos\varphi + i\sin\varphi) \cdot (\cos\psi + i\sin\psi).$$

Приравнивая действительные и мнимые части, сразу получаем:

$$\cos(\varphi + \psi) = \cos\varphi\cos\psi - \sin\varphi\sin\psi, \tag{1.73}$$
$$\sin(\varphi + \psi) = \sin\varphi\cos\psi + \cos\varphi\sin\psi. \tag{1.74}$$

Впервые, по-видимому, мнимые величины появились в труде Дж. Кардано «Великое искусство, или Об алгебраических правилах» (1545), который счел их бесполезными и непригодными к употреблению. Вплоть до 19 в. комплексные числа оставались для математиков предметом отвлечённых манипуляций, когда норвежец Гаспар Вессель первым ввёл их геометрическое представление. Позднее ирландский математик Уильям Роуэн Гамильтон (1837) развил геометрическую интерпретацию, согласно которой каждое комплексное число задаётся парой обычных чисел.

При введении комплексных чисел элементарные функции удается задать на всем множестве чисел (например, вычислить логарифм и корень отрицательного числа). Кроме того, значения элементарных функций (т.е. переменные обратных функций) задаются на всем множестве чисел, например, значением элементарной тригонометрической функции может быть любое число (а не только на отрезке $[-1, +1]$).

И что самое замечательное, введение комплексных чисел является *последним* расширением числовой системы (после натуральных, целых и действительных), т.е. нам не требуется никаких новых чисел. Более того, в

конце 19 в. было доказано, что всякое расширение понятия числа за пределы комплексных чисел возможно только в случае отказа от каких-либо привычных свойств арифметических действий.

Определение. Функция $f\colon \mathbb{C} \to \mathbb{C}$ или $w = f(z)$ — есть *функция комплексного переменного.*

Запишем комплексную функцию $w = f(z)$ в виде:

$$w = u(x, y) + iv(x, y). \tag{1.75}$$

где $u(x, y)$ и $v(x, y)$ — вещественные функции,

x и y — вещественные числа комплексного числа $z = x + iy$.

Задание комплексной функции равносильно заданию двух вещественных функций от двух вещественных аргументов каждая. Например, равенство $w = e^z$ равносильно системе из двух равенств $u = e^x \cos y$, $v = e^x \sin y$.

Таким образом, отображение вещественной плоскости z в вещественную плоскость w можно записать в виде $u, v\colon \mathbb{R}^2 \to \mathbb{R}^2$.

Поскольку для комплексных чисел определены те же операции, что и для вещественных, то остаются в силе все основные свойства, выведенные для функций вещественного переменного. Например:

$$(f_1 + f_2)(z) = f_1(z) + f_2(z), \quad (f_1 \cdot f_2)(z) = f_1(z) \cdot f_2(z). \tag{1.76}$$

Если мы можем разложить произвольную функцию одного комплексного переменного (аналогично функции действительного переменного, см. (1.30–1.36)) в бесконечный ряд в точке z_0:

$$f(z) = a_0 + a_1(z - z_0) + \ldots + a_n(z - z_0)^n + \ldots, \tag{1.77}$$

то такая функция называется *аналитической* в этой точке.

Функция, аналитическая в точке, непрерывна в этой точке.

Как и для функции действительного переменного для функции комплексного переменного существует проблема неоднозначности.

Функция f определяет отображение плоскости z в плоскость w. Если это значение f единственное, то функция $w = f(z)$ называется *однозначной.* Рассмотрим теперь функцию, обратную к f (обратное отображение) $z = f^{-1}(w)$. Если это значение z единственное, то функция f^{-1} однозначна, а отображение $w = f(z)$ в этом случае называется *однолистным.* Однозначное однолистное отображение иначе называется взаимно однозначным. Однозначная аналитическая функция называется *голоморфной* (от греч. *ὅλος* — весь, целый и *μορφή* — форма, образ). В противном случае отображение $w = f(z)$ называется *многолистным* — оно может быть двулистным, трехлистным и т.д., даже бесконечнолистным. Например, функции $f = \sqrt[n]{z}$, z^n, $\ln z$, $\arcsin z$, $\operatorname{arctg} z$, алгебраические функции и т.д. неоднозначны, одному значению z соответствует n или бесконечное количество значений f.

Область аналитичности *многозначной функции* расширяется и налагается на себя, доставляя новые значения функции в точках плоскости, где она уже была определена. Процесс продолжения функции в большую область,

приводит к понятию аналитической функции, рассматриваемой в целом, т.е. во всей свой естественной области существования. Такой процесс носит название *аналитического продолжения* по Вейерштрассу. Чтобы избавиться от многозначности, функцию f рассматривают не как функцию точек плоскости, а как функцию точек некоторой многолистной поверхности, на которой функция f становится однозначной функцией. Идея перехода к таким поверхностям, которые позволили заменить теорию многозначных аналитических функций теорией однозначных аналитических функций, принадлежит Б. Риману, а сами они носят название *римановых поверхностей* . Листы поверхности носят название *ветвей* аналитической функции. С каждой аналитической функцией комплексного переменного связывается такая риманова поверхность этой функции.

Понятие абстрактной римановой поверхности ввёл Г. Вейль (1913). Термин «голоморфная функция» был введён двумя учениками Коши, Брио (1817–82) и Боке (1819–95).

Класс аналитических функций исключительно важен. Во-первых, он достаточно широк и охватывает большинство функций, встречающихся в математике, естествознании и технике. Во-вторых, он замкнут относительно основных операций арифметики, алгебры и анализа: т.е. при применении этих операций к аналитическим функциям вновь получаются аналитические функции. В-третьих, он обладает свойством единственности, каждая аналитическая функция представляет собой «единую» функцию во всей своей области существования.

В теории аналитических функций принято считать, что кроме обычных, конечных, точек плоскости z имеется еще одна *несобственная*, иначе, бесконечно удаленная точка $z = \infty$. Плоскость с добавленной к ней бесконечно удаленной точкой называется *расширенной плоскостью*. Сфера, касающаяся плоскости, называется *риманова сфера* — одно из возможных геометрических изображений совокупности комплексных чисел, введенное Б. Риманом.

Точка $z = 0$, называется *точкой ветвления* функции $f = \sqrt[n]{z}$, имеющей n ветвей. Другой точкой ветвления этой функции принято считать «бесконечную точку» $z = \infty$.

Точка ветвления $z = 0$ функции $\ln z$ имеет бесконечное число ветвей и называется точкой ветвления бесконечного порядка.

Если функция аналитическая, то соответствующее отображение называется также аналитическим. Представим себе плоскости z и w отображения $w = f(z)$ совмещенными в некоторой точке. При аналитическом отображении вся малая окрестность этой точки испытает, с точностью до малых членов высшего порядка, поступательный перенос, всестороннее растяжение и поворот. Значения модуля и аргумента производной в совмещенной точке служат, соответственно, коэффициентом растяжения и углом поворота. Это означает, что образ каждой малой фигуры будет с точностью до малых членов высшего порядка геометрически подобен прообразу. Ориентация плоскости также сохраняется, т.е. если обходить малый замкнутый контур плоскости в некотором направлении, то и образ будет обходиться в

том же направлении.

Взаимно однозначное (однозначное однолистное) непрерывное отображение некоторой плоской области, при котором сохраняется подобие (форма) бесконечно малых фигур и ориентация плоскости, называется *конформным* (от позднелат. conformis — подобный). Для конформности отображения необходимо и достаточно, чтобы функция и обратная к ней в рассматриваемой области были однозначными и аналитическими.

Если подобие фигур сохраняется, но ориентация меняется на противоположную, то отображение называется *антиконформным*.

Особые точки аналитической функции — точки, в которых нарушается свойство аналитичности. Например, функцию $\frac{1}{z}$ нельзя считать аналитической на *всей* плоскости z. Эта функция аналитическая в области, полученной выбрасыванием из плоскости z точки $z = 0$, в которой непрерывность, а потому и аналитичность нарушаются. Такая точка называется особой точкой аналитической функции или *особенностью*.

Рассматрим *изолированную особую точку* (для неё существует окрестность, свободная от других особых точек) однозначной аналитической функции. Если $f(z)$ — голоморфная функция в кольце $0 < |z - z_0| < r$ с центром в изолированной особой точке z_0, то функция разлагается в этом кольце в *ряд Лорана*:

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n = \sum_{n=1}^{\infty} \frac{a_{-n}}{(z - z_0)^n} + \sum_{n=0}^{\infty} a_n (z - z_0)^n, \tag{1.78}$$

содержащий как положительные, так и отрицательные степени $z - z_0$.

Совокупность членов с неотрицательными степенями — *правильная часть*, с отрицательными степенями — *главная часть* ряда Лорана. Главная часть определяет характер функции f в точке z_0.

Если члены с отрицательными степенями отсутствуют, то z_0 называется *правильной точкой* (*устранимой особой точкой*), так как существует и конечен предел: $\lim\limits_{z \to z_0} f(z) = a_0$. Полагая $f(z_0) = a_0$, получают аналитическую функцию во всем круге $|z - z_0| < r$.

Если число членов с отрицательными степенями конечно:

$$f(z) = \sum_{n=-m}^{+\infty} a_n (z - z_0)^n, \quad m > 0, a_m \neq 0, \tag{1.79}$$

то точка называется *полюсом* функции $f(z)$ порядка или кратности m, в котором $\lim\limits_{z \to\ z_0} f(z) = \infty$. Полюс называется *простым* при $m = 1$.

Если число членов с отрицательными степенями бесконечно, ($a_n \neq 0$ для бесконечного множества отрицательных индексов n), то точка z_0 называется *существенно особой точкой*. Вблизи этой точки $f(z)$ остается однозначной, но в z_0 не существует ни конечного, ни бесконечного предела функции $f(z)$. Например, $\sin(\frac{1}{z})$, $\exp(\frac{1}{z})$.

Целая функция — функция комплексного переменного $f(z)$, голоморфная на всей конечной плоскости, т.е. представимая степенным рядом Тейлора (см. гл. 3), сходящимся при любом z. Бесконечно удалённая точка $z = \infty$ есть изолированная и устранимая особая точка целой функции. Конечных особых точек у целой функции нет. Если $z = \infty$ — полюс для целой функции, то $f(z) = a_0 + a_1 z + \ldots + a_n z^n + \ldots$, сходящимися во всей плоскости. К ним относятся многочлены от z. Если $z = \infty$ — существенно особая точка, то $f(z)$ — трансцендентная целая функция, например, $\sin z$, $\cos z$, e^z.

Глава 2

Линейная алгебра

В этой главе мы рассмотрим такие геометрические объекты как вектор, матрица, тензор. С формальной математической точки зрения эти объекты состоят из наборов (кортежей) чисел, которые называются компонентами (элементами) данного объекта. Однако они не распадаются на отдельные числа, а обладают свойствами самостоятельных объектов.

Множества из этих обектов образуют непрерывные (полные) пространства, имеющие общее название — векторные пространства.

Все эти объекты являются инвариантами — они не зависят от системы координат, хотя их компоненты при этом изменяются.

Оказывается, что с этими объектами до определённой степени можно обращаться как с числами.

Однотипные объекты можно складывать и умножать на числа. Однако перемножение объектов друг на друга имеет свою специфику и определённые ограничения.

Например, для векторов существует внутренне (скалярное) и внешнее произведение. Для матриц операция произведения не коммутативна. При перестановке сомножителей местами результат может отличаться.

Кроме операций, похожих на арифметические, каждый класс геометрических объектов обладает своими специфическими операциями, например, транспонирование матрицы или свёртка тензора.

Отдельный класс объектов — заданные на векторах скалярные функции (линейные формы). Для каждого векторного пространства можно задать сопряжённое или двойственное пространство линейных форм.

2.1 Векторная алгебра

Векторным или *линейным пространством* (над полем K) называется *непрерывное множество* объектов (элементов любой природы) в котором определена операция *линейной суперпозиции*, т.е. заданы операции сложения объектов и умножения объектов на число из поля K.

Вектором (от лат. vector, буквально — несущий) называется элемент этого непрерывного множества.

Примеры:

1. Простейший вектор — это просто точка в пространстве или n-членная последовательность $\nu = (x^1, \ldots, x^n) \in \mathbb{R}^n$. Числа в n-членной последовательности называют *компонентами* вектора.

От начала координат, до заданной точки пространства $\nu = (x^1, \ldots, x^n)$ можно построить направленный отрезок, который имеет специальное название — *радиус-вектор*: $\mathbf{r} = (x^1, \ldots, x^n)$.

Началу координат в $\mathbb{R}^n$ соответствует нулевой вектор $\mathbf{0} = (0, \ldots, 0)$, у которого все компоненты равны нулю, и он не имеет направления.

2. Пусть пространство $\mathbb{R}^n$ непрерывно, и для двух его произвольных точек $\nu = (x^1, \ldots, x^n) \in \mathbb{R}^n$ и $\tau = (y^1, \ldots, y^n) \in \mathbb{R}^n$ можно записать операции линейной суперпозиции над полем $\mathbb{R}$ ($c \in \mathbb{R}$):

$$\begin{aligned} \xi &= \nu + \tau = (x^1 + y^1, \ldots, x^n + y^n) \in \mathbb{R}^n, \\ \zeta &= c\nu = (cx^1, \ldots, cx^n) \in \mathbb{R}^n. \end{aligned} \tag{2.1}$$

Тогда $\mathbb{R}^n$, наделённое операциями линейной суперпозиции над полем вещественных чисел $\mathbb{R}$, есть *векторное пространство* V.

3. Направленный отрезок ξ, соединяющий две точки $\nu = (x^1, \ldots, x^n)$ и $\tau = (y^1, \ldots, y^n)$, есть вектор:

$$\xi = \nu - \tau = (x^1 - y^1, \ldots, x^n - y^n). \tag{2.2}$$

Обычно под вектором понимают не точку, а направленный отрезок, непривязанный к определенной точке пространства (начальная точка может быть выбрано произвольно), т.е. вектор можно произвольно перемещать в пространстве, при условии сохранения его направления.

Компоненты такого вектора — направленного отрезка, это его проекции на координатные оси.

Длина отрезка (1.15) между этими точками ν и τ называется длиной вектора $\rho(\nu, \tau)$, а также нормой $\|\nu - \tau\|$ или модулем $|\nu - \tau|$ вектора.

Понятие линейного пространства имеет в алгебре такое же фундаментальное значение, как и понятие группы. Группа и векторное пространство являются разновидностью *универсальной алгебры*.

В универсальной алгебре *операции рассматриваются как объекты*, которые образуют множество операций (над различными классами объектов), т.е. образуют множество, наделённое алгебраической структурой. Элементами множества операций (универсальной алгебры) служат операции над определенным классом объектов. Например, алгебра чисел, векторная алгебра, алгебра матриц и т.д. Алгебраической структурой называется набор используемых в универсальной алгебре операций. Множество универсальных алгебр с определённым набором операций образует соответствующий класс универсальных алгебр. Так линейная алгебра или векторное пространство есть *алгебра над полем*.

В ходе развития математики первоначально выделились сравнительно немногие типы универсальных алгебр: группы, векторные пространства, ассоциативные кольца и алгебры, модули. В дальнейшем предметом изучения стали также другие классы: неассоциативные кольца и алгебры (в том числе алгебры Ли), решётки (долгое время называвшиеся структурами) и т.д.

Если значение функции $f\colon \mathbb{R}^n \to \mathbb{R}$ или $y = f(x^1, \dots, x^n)$ зависит только от вектора $\mathbf{x} = (x^1, \dots, x^n)$ и не изменяется при преобразовании системы координат, когда компоненты вектора $x^1, \dots, x^n$ изменяются, то такая функция является инвариантом, и её можно записать как числовую функцию от вектора $y = f(\mathbf{x})$. Инвариантная числовая функция от вектора называется *скалярной функцией* n переменных или *скалярным полем* (от лат. scalaris — ступенчатый), а её числовые значения называются *скалярами*, которые не тождественны числам.

Пример. Компоненты вектора — числа, но они не скаляры, так как не являются инвариантами (проекции на оси изменяются при замене осей). А вот длина вектора — одновременно и число, и скаляр, так как от изменения координатных осей не зависит.

Скалярное или *внутреннее произведение* двух векторов $\nu, \tau \in \mathbb{R}^n$ есть скалярная функция, которая связывает нормы векторов и косинус угла φ между ними:

$$\nu \cdot \tau = |\nu||\tau| \cos\varphi, \quad \nu \cdot \nu = |\nu|^2. \tag{2.3}$$

За φ принимается угол между векторами, не превосходящий π.

Величина скалярного произведения — это число, равное произведению норм векторов на косинус угла φ между ними, или *произведение проекции одного вектора на длину второго*.

Для скалярного произведения часто используют обозначение (ν, τ), но во избежание путаницы мы будем использовать это обозначение только для евклидовой структуры (см. ниже).

Свойства скалярного произведения:

1. $c(\nu \cdot \tau) = \nu \cdot (c\tau)$, где $c \in \mathbb{R}$ — *ассоциативность*, данное свойство следует из определения скалярного произведения;
2. $\nu \cdot \tau = \tau \cdot \nu$ *коммутативность* (симметричность);
3. $\nu \cdot \tau = 0$ при ненулевых ν и τ, только если векторы перпендикулярны, что следует из определения с учётом $\cos\frac{\pi}{2}$;
4. $\nu \cdot (\tau + \upsilon) = \nu \cdot \tau + \nu \cdot \upsilon$ — *дистрибутивность* (билинейность), следует из геометрических соображений, так как сумма проекций векторов τ и υ на вектор ν равна проекции вектора $\tau + \upsilon$ на ν;
5. если $\upsilon = \nu + \tau$, то получаем *правило параллелограмма* $\upsilon \cdot \upsilon = \nu \cdot \nu + \tau \cdot \tau + 2(\nu \cdot \tau)$ или $|\upsilon|^2 = |\nu|^2 + |\tau|^2 - 2|\nu||\tau| \cos\varphi$ с учётом угла при построении, так как $\cos(\pi - \varphi) = -\cos\varphi$ (рис. 2.1).

Свойства нормы:

1. $\|\nu\| \geqslant 0$, причем $\|\nu\| = 0$ тогда и только тогда, когда $\nu = (0, \ldots, 0)$, что следует из определения;
2. $\|c\nu\| = |c| \cdot \|\nu\|$, где $c \in \mathbb{R}$, следует из определения;
3. $\|\nu \cdot \tau\| \leqslant \|\nu\| \cdot \|\tau\|$ (причем $\|\nu \cdot \tau\| = \|\nu\| \cdot \|\tau\|$ только если $\tau = c\nu$) — *неравенство Шварца* следует из определения внутреннего произведения с учётом $\cos\varphi \leqslant 1$;
4. $\|\nu + \tau\| \leqslant \|\nu\| + \|\tau\|$ — *аксиома треугольника* следует из правила параллелограмма;
5. если $\nu \cdot \tau = 0$, то из правила параллелограмма получаем *теорему Пифагора* $\|\nu + \tau\|^2 = \|\nu\|^2 + \|\tau\|^2$.

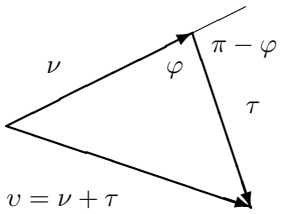


Рис. 2.1: Правило параллелограмма $|\upsilon|^2 = |\nu|^2 + |\tau|^2 - 2|\nu||\tau|\cos\varphi$

Прямая n-я степень множества комплексных чисел $\mathbb{C}$ образует *n-мерное комплексное пространство* $\mathbb{C}^n = \mathbb{C} \times \ldots \times \mathbb{C}$.

Комплексное число задано двумя вещественными числами $z = (x, y)$. Соответственно, точка $\upsilon \in \mathbb{C}^n$ задана двумя вещественными векторами $\upsilon = (\nu, \tau)$ и есть элемент чётномерного вещественного пространства $\mathbb{R}^n$ или $2n$-членная последовательность $\upsilon = (x^1, \ldots, x^n, y^1, \ldots, y^n)$.

Комплексное пространство $\mathbb{C}^n$, наделённое операциями суперпозиции над полем комплексных чисел $\mathbb{C}$, называется *n-мерным линейным комплексным пространством.*

Квадрат модуля (нормы) комплексного числа вычисляется как произведение числа на комплексно сопряжённое $z \cdot \overline{z} \overset{(1.71)}{=} r^2$.

Аналогично комплексному числу, квадрат модуля комплексного вектора в комплексном пространстве $\mathbb{C}^n$ вычисляется как скалярное произведение комплексного вектора на комплексно сопряжённый вектор. Эта аналогия распространяется и на скалярное произведение двух произвольных векторов в комплексном пространстве $\mathbb{C}^n$, которое называют *эрмитово скалярное произведение*:

$$\nu \cdot \overline{\nu} = |\nu|^2, \quad \nu \cdot \overline{\tau} = |\nu||\tau|\cos\varphi. \tag{2.4}$$

Свойства нормы комплексного вектора и эрмитова скалярного произведения двух комплексных векторов:

1. $\nu \cdot \overline{\nu} = |\nu|^2 > 0$ при $\nu \neq 0$, т.е. скалярный квадрат ненулевого вектора есть положительное действительное число;
2. $c(\nu \cdot \overline{\tau}) = (c\nu) \cdot \overline{\tau} = \nu \cdot (\overline{c} \cdot \overline{\tau})$, где $c \in \mathbb{C}$;

3. $\nu \cdot (\overline{\tau} + \overline{v}) = \nu \cdot \overline{\tau} + \nu \cdot \overline{v}$;

4. $\nu \cdot \tau = \overline{\tau \cdot \nu}$ эквивалентно $\nu \cdot \overline{\tau} = \tau \cdot \overline{\nu}$;

5. $\|\nu \cdot \overline{\tau}\|^2 \leqslant (\nu \cdot \overline{\nu})(\tau \cdot \overline{\tau})$ *неравенство Буняковского*, откуда определим:

$$\cos^2 \varphi = \frac{\|\nu \cdot \overline{\tau}\|^2}{(\nu \cdot \overline{\nu})(\tau \cdot \overline{\tau})} \text{ или } \cos \varphi = \pm \frac{\|\nu \cdot \overline{\tau}\|}{\|\nu\| \|\tau\|}. \tag{2.5}$$

Значит между комплексными векторами в $\mathbb{C}^n$ можно определить вещественный угол. Тогда запишем ещё одно свойство:

6. $\nu \cdot \overline{\tau} = 0$ при ненулевых ν и τ, только если векторы ортогональны.

Параллелограмм определён как прямое произведение двух отрезков или интервалов, на которых он построен. Соответственно, если под отрезками понимают вещественные векторы ν и τ, то их прямое произведение есть построенный на них параллелограмм. В линейной алгебре эту алгебраическую операцию называют *внешнее произведение векторов*:

$$P^2(\nu, \tau) \overset{(1.9)}{=} \nu \times \tau. \quad \text{(рис. 2.2)} \tag{2.6}$$

Кососкалярное произведение векторов ν и τ вычисляется как *площадь* — основная числовая характеристика параллелограмма:

$$S = [\nu, \tau] \overset{(1.12)}{=} |\nu||\tau| \sin \varphi. \tag{2.7}$$

За φ принимается угол между векторами, не превосходящий π.

Обозначение $[\nu, \tau]$ часто также называют внешним произведением векторов, но во избежание путаницы мы будем его использовать только для обозначения кососкалярного произведения.

Знак кососкалярного произведения зависит от угла φ (рис. 2.2) и связан с ориентацией пространства. Как длина отрезка может принята положительной или отрицательной, так и площадь параллелограмма может быть принята положительной или отрицательной величиной.

Ориентация пространства (франц. orientation, букв. — направление на восток) — обобщённое понятие направления, задается ориентацией замкнутой поверхности, которая ограничивает кусок пространства.

Направление обхода любой замкнутой кривой, которая ограничивает кусок поверхности, против часовой стрелки считается положительным. При этом говорят, что *поверхность* внутри кривой *ориентирована* правым образом, а площадь внутри кривой считается положительной величиной. Направление обхода замкнутой кривой по часовой стрелке считается отрицательным, при этом поверхность внутри кривой считается ориентированной левым образом, а площадь внутри кривой считается отрицательной величиной.

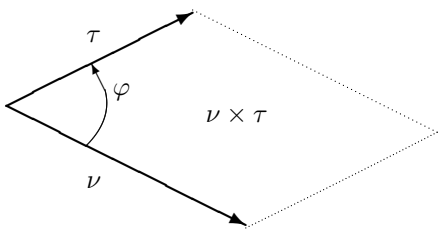


Рис. 2.2: Внешнее произведение

Соответственно, трёхмерный кусок пространства ориентирован правым (левым) образом, если его наблюдаемая наружняя замкнутая поверхность ориентирована правым (левым) образом. От выбора ориентации пространства зависит знак объёмов, ограниченных ориентированными замкнутыми поверхностями, т.е. объём пространства, ориентированного правым образом, считается положительным. Понятие ориентации распространяется также и на многомерные пространства.

Чтобы площадь параллелограмма была положительной, угол φ кососкалярного произведения должен быть положительного вращения, т.е. против часовой стрелки от ν к τ (рис. 2.2).

При изменении направления вращения угол φ, соответственно, $\sin\varphi$ и площадь в кососкалярном произведении меняют знак.

Свойства кососкалярного произведения векторов:

1. $[\nu,\tau] = -[\tau,\nu]$ — кососимметричность;
2. $[\nu,\nu]$ — косоортогональность;
3. $c[\nu,\tau] = [c\nu,\tau] = [\nu,c\tau]$, $c \in \mathbb{R}$ — ассоциативность;
4. $[\nu,\tau+\upsilon] = [\nu,\tau] + [\nu,\upsilon]$ — дистрибутивность.

При преобразованиях системы координат векторы сохраняют длины и образованные ими углы, а значит скалярное и кососкалярное произведения по определению также являются *инвариантами* или геометрическими объектами. Тогда справедливо выражение:

$$|[\nu,\tau]|^2 + |\nu\cdot\tau|^2 \overset{(2.3)}{\underset{(2.7)}{=}} |\nu|^2|\tau|^2(\sin^2\varphi + \cos^2\varphi) = |\nu|^2|\tau|^2. \tag{2.8}$$

Параллелепипед есть *внешнее произведение* трёх векторов ν,τ,υ, на которых он построен

$$P^3(\nu,\tau,\upsilon) \overset{(1.10)}{=} \nu\times\tau\times\upsilon. \tag{2.9}$$

Кососкалярное произведение трёх векторов вычисляется как *объём* — основная числовая характеристика параллелепипеда:

$$V = [\nu,\tau,\upsilon] \overset{(1.13)}{=} |\nu||\tau||\upsilon|\sin\varphi\sin\theta. \tag{2.10}$$

За φ принимается угол между векторами, на которых построен параллелограмм, а за θ — угол между параллелограммом и третьим вектором. Оба угла не превосходят π.

Внешнее произведение k векторов $\xi_1, \ldots, \xi_k$ в n-мерном пространстве даёт построенный на этих векторах k-мерный параллелепипед:

$$P^k(\xi_1, \ldots, \xi_k) \overset{(1.11)}{=} \xi_1 \times \ldots \times \xi_k, \tag{2.11}$$

объём которого равен их *кососкалярному произведению*:

$$V = [\xi, \ldots, \xi_k] \overset{(1.14)}{=} |\xi_1| \cdot \ldots \cdot |\xi_k| \sin\varphi_1 \ldots \sin\varphi_{k-1}. \tag{2.12}$$

Термин вектор ввёл Гамильтон (ок. 1845). Понятия и названия «скалярное произведение» и «векторное произведение» введены У. Гамильтоном (1853); под названием «внутреннее произведение» и «внешнее произведение» их употреблял Г. Грассман (1844). Обозначение $[\nu, \tau]$ введено Г. Грассманом (1844), обозначения $\nu \cdot \tau$ и $\nu \times \tau$ впервые встречаются у Дж. Гиббса (1881), (ν, τ) — у О. Хенричи (1903).

2.2 Базисное представление векторов

Умножение вектора на ненулевое число c означает удлинение вектора на соответствующую величину

$$\xi = c \cdot \zeta. \tag{2.13}$$

Ненулевые векторы ξ и ζ, для которых справедливо это уравнение, имеют одинаковое или противоположное направление и называются *линейно зависимыми* или *коллинеарными*. Одинаковые векторы противоположного направления $\xi = -\zeta$ называют *полярными*.

Вектором направления или *направляющим вектором* прямой называют любой ненулевой вектор $\mathbf{s} = (s^1, \ldots, s^n)$ в n-мерном пространстве, задающий прямую (на которой он лежит).

Пусть вектор $\xi \in \mathbb{R}^n$ соединяет две точки $\nu = (x^1, \ldots, x^n)$ и $\tau = (y^1, \ldots, y^n)$: $\xi = \nu - \tau = (x^1 - y^1, \ldots, x^n - y^n)$ и является коллинеарным вектору направления: $\xi = c \cdot \mathbf{s}$. Ненулевые проекции коллинеарных векторов пропорциональны, поэтому для ненулевых проекций этих двух коллинеарных векторов справедливы следующие соотношения:

$$\frac{x^1 - y^1}{s^1} = \ldots = \frac{x^n - y^n}{s^n} = c. \tag{2.14}$$

В результате мы получили уравнение прямой, которое называется уравнением в отрезках или *каноническим уравнением прямой*.

Если прямая перпендикулярна какой-либо координатной оси, то соответствующая компонента направляющего вектора s^i и коллинеарного ему вектора $x^i - y^i$ равны нулю, и мы записываем каноническое уравнение этой прямой за исключением соотношения для i-ой координаты.

Векторы $\xi_1, \ldots, \xi_n$ называют *линейно независимыми*, если условие:

$$\xi_1 c^1 + \ldots + \xi_n c^n = 0, \tag{2.15}$$

где $c^1, \ldots, c^n$ — вещественные или комплексные числа,

может выполняться только когда $c^1 = \ldots = c^n = 0$.

Если n векторов $\xi_1, \ldots, \xi_n$ в n-мерном вещественном векторном пространстве (или пространстве размерности больше чем n) линейно зависимы, то говорят, что они *компланарны* и лежат в соответствующей $(n-1)$-мерной плоскости или, другими словами, *задают плоскость*.

Допустим, произвольная точка n-мерного пространства с координатами $x_0^1, \ldots, x_0^n$ принадлежит $(n-1)$-мерной плоскости. Тогда $(n-1)$-плоскость есть множество компланарных векторов или точек с координатами $(x^1, \ldots, x^n)$, удовлетворяющее уравнению:

$$\xi_1(x^1 - x_0^1) + \ldots + \xi_n(x^n - x_0^n) = 0 \text{ или } \xi_1 x^1 + \ldots + \xi_n x^n = const. \quad (2.16)$$

где векторы $\xi_1, \ldots, \xi_n$ задают $(n-1)$-мерную плоскость.

Пример: пара коллинеарных векторов лежит на прямой, а тройка векторов, содержащая пару коллинеарных векторов — компланарна и лежит на двумерной плоскости.

Если $n+1$ линейно зависимый вектор задаёт n-мерную плоскость, следовательно, n линейно независимых векторов также задают n-мерное векторное (линейное) пространство, так как мы всегда можем задать с их помощью ещё один линейно зависимый вектор.

Базисом (от греч. $\beta\acute{\alpha}\sigma\iota\varsigma$ — основание) *n-мерного векторного пространства V* называют любые n линейно независимых векторов (вещественных или комплексных) $\xi_1, \ldots, \xi_n$, т.е. таких, что ни один из векторов базиса не может быть записан как линейная суперпозиция оставшихся. Базис дает набор векторов, которых необходимо и достаточно для записи произвольного вектора η n-мерного пространства в виде линейной комбинации базисных векторов, т.е. любых $n-1$ векторов недостаточно, а любые $n+1$ векторов будут линейно зависимы:

$$\eta = \xi_1 x^1 + \ldots + \xi_n x^n = \sum_{j=1}^{n} \xi_j x^j = \xi_j \, x^j, \quad (2.17)$$

где $x^1, \ldots, x^n$ — компоненты вектора в базисе $\xi_1, \ldots, \xi_n$;

j — *индекс суммирования*, который должен иметь ровно два вхождения — одно на верхнем уровне и одно на нижнем (Эйнштейн предложил вообще не записывать символ суммирования $\sum$, имея ввиду неявное суммирование по этому индексу).

Представление вектора в виде линейной комбинации базисных векторов (2.17) называется *разложением вектора по базису*.

Выбор нового базиса означает выбор новой системы координат и задание нового векторного пространства V. Обратно, если мы хотим задать n-мерное векторное пространство, мы должны выбрать в качестве базиса n линейно независимых векторов.

Пространство с базисом из n векторов называется *конечномерным*. *Бесконечномерным* называется пространство, в котором можно найти линейно независимую систему из любого числа векторов.

При изменении базиса вектор не изменяется. Векторы, как геометрические объекты, представляют собой направленные отрезки, и для их существования не требуются базисы. Поэтому векторы остаются *инвариантными* — неизменными по абсолютной величине и направлению при замене осей системы координат.

Однако компоненты векторов зависят от выбранного базиса. При выборе другой системы координат (другого базиса) компоненты вектора изменяются. *Вектор как геометрический объект* обладает свойствами, которые не видны при задании *вектора как последовательности чисел.*

Пусть в новом базисе $\zeta_1, \ldots, \zeta_n$ вектор η имеет вид:

$$\eta = \zeta_1 y^1 + \ldots + \zeta_n y^n = \sum_{i=1}^{n} \zeta_j y^j = \zeta_j \, y^j, \tag{2.18}$$

где $y^1, \ldots, y^n$ — компоненты вектора в базисе $\zeta_1, \ldots, \zeta_n$;

Тогда

$$\eta \overset{(2.17)}{=} \xi_j \, x^j \overset{(2.18)}{=} \zeta_i \, y^i. \tag{2.19}$$

Два вектора равны, только и только тогда, когда равны их компоненты в одном и том же базисе.

Каждой координате пространства $\mathbb{R}^n$ соответствует вектор единичной длины (как говорят, нормированный на единицу) или *орт* (от греч. *ὀρθος* — прямой) — отрезок, направленный вдоль соответствующей оси координат. Орт имеет только одну проекцию, т.е. задается одним числом e_i. Орты пространства $\mathbb{R}^n$ образуют *нормированный* или *стандартный базис* $e_1, \ldots, e_n$, где $e_i = (0, \ldots, 1, \ldots, 0)$, с числом 1 на i-ом месте и 0 на всех остальных. Произвольный вектор $\xi = (x^1, \ldots, x^n) \in \mathbb{R}^n$ может быть записан как линейная суперпозиция ортов:

$$\xi = e_1 x^1 + \ldots + e_n x^n = e_i x^i. \tag{2.20}$$

В комплексном пространстве $\mathbb{C}^n$ нормированным (единичным) вектором или ортом называется вектор e_i, для которого $|e_i| = 1$.

В качестве координат в $\mathbb{R}^3$ обычно выбирают x, y, z. Базис *декартовой* (ортогональной и прямолинейной) системы координат принято задавать ортами i, j, k или ортами e_x, e_y, e_z, соответствующими координатам x, y, z. Тогда, например, радиус-вектор в декартовых координатах имеет вид

$$\mathbf{r} = i \cdot x + j \cdot y + k \cdot z \quad \text{или} \quad \mathbf{r} = e_x x + e_y y + e_z z. \tag{2.21}$$

Дальнейший материал этого параграфа посвящён определению понятия *один-форма* и излагается для вещественного пространства $\mathbb{R}^n$.

Определив в $\mathbb{R}^n$ вектор как линейную суперпозицию базисных векторов (2.20) мы, на самом деле, использовали не одну, а две n-членных последовательности: первая — компоненты вектора $x^1, \ldots, x^n$ с верхним индексом, вторая — базисные векторы $e_1, \ldots, e_n$ с нижним индексом.

Выбор базиса $e_1, \ldots, e_n$ задаёт множество n-членных последовательностей $\xi = (x^1, \ldots, x^n)$ или векторное пространство V. Если же формальным образом произвести выбор координат пространства в данной точке $x^1, \ldots, x^n$ в качестве некого сопряжённого базиса, то тем самым будет задано пространство V^* n-членных последовательностей с нижним индексом $\omega = (a_1, \ldots, a_n)$, *сопряжённое* или *двойственное* векторному пространству V. Это сопряжённое пространство скалярных функций $\omega(\xi)$ задано на элементах $\omega \in V^*$, называемых 1-формы.

Скалярная функция $\omega\colon V \to \mathbb{R}$ называется *линейной функцией* или *линейной формой* (коротко, 1-*формой*) от вектора $\xi \in V$, если:

$$\omega(\xi_1 + \xi_2) = \omega(\xi_1) + \omega(\xi_2), \quad \omega(c\,\xi) = c\omega(\xi), \text{ где } c \in \mathbb{R}; \tag{2.22}$$

т.е. если линейная суперпозиция аргумента даёт линейную суперпозицию функции, то функция называется линейной.

Каждому вектору $\xi = e_1 x^1 + \ldots + e_n x^n \in V$ (каждой точке в пространстве V с выбранными координатами $x^1, \ldots, x^n$) 1-форма $\omega = (a_1, \ldots, a_n)$ относит некоторое число $\omega(\xi)$, равное длине образа вектора ξ в сопряжённом пространстве V^*.

Один-форма может быть записана как линейная суперпозиция:

$$\omega(\xi) = \omega(e_1)\, x^1(\xi) + \ldots + \omega(e_n)\, x^n(\xi) = a_1 x^1 + \ldots + a_n x^n. \tag{2.23}$$

Числа

$$\omega(e_i) = a_i \quad (i = 1, \ldots, n) \tag{2.24}$$

называют *компоненты* 1-формы, n-членную последовательность $a_1, \ldots, a_n$ с нижним индексом называют *ковектор*, число $\omega(\xi)$ называют *свёртка* 1-формы ω с вектором ξ, величины $x^1(\xi), \ldots, x^n(\xi)$, выбранные в качестве сопряжённого базиса, называют *базисные 1-формы от вектора* $\xi \in V$.

Из выражения (2.23) следует, что проекции вектора ξ на координатные оси есть базисные 1-формы

$$x^i(\xi) = x^i \quad (i = 1, \ldots, n). \tag{2.25}$$

Из выражения $\omega(e_i) = a_1 x^1(e_1) + \ldots + a_n x^n(e_n)$ следует что

$$x^i(e_j) = \delta_{ij} \quad (i, j = 1, \ldots, n), \tag{2.26}$$

где δ_{ij} — *символ Кронекера*, который равен нулю $\delta_{ij} = 0$ при $i \neq j$ и единице $\delta_{ij} = 1$ при $i = j$.

Легко убедиться, что величины $x^1, \ldots, x^n$ действительно образуют базис сопряженного пространства V^* или *сопряжённый* (*дуальный*) *базис*, так как

они линейно независимы. Допустим, что это не так, тогда можно записать $c_1x^1 + \ldots + c_nx^n = 0$, где числа c_i не равны нулю одновременно. Тогда $c_1x^1(e_1) + \ldots + c_nx^n(e_1) = c_1 = 0$. Подставляя остальные базисные векторы в выражение, получим, что все эти числа равны нулю $c_1 = \ldots = c_n$, что доказывает линейную независимость $x^1, \ldots, x^n$.

На 1-формах можно задать операцию линейной суперпозиции:

$$(\omega_1 + \omega_2)(\xi) = \omega_1(\xi) + \omega_2(\xi), \quad (c\,\omega)(\xi) = c\,\omega(\xi), \text{ где } c \in \mathbb{R},\ \xi \in V,$$

аналогично соответствующим операциям для векторов:

$$(\xi_1 + \xi_2)(\omega) = \xi_1(\omega) + \xi_2(\omega), \quad (c\,\xi)(\omega) = c\,\xi(\omega), \text{ где } c \in \mathbb{R},\ \omega \in V^*.$$

Поэтому сопряжённое пространство 1-форм V^* является векторным пространством, аналогично векторному пространству V.

Исходное и сопряжённое пространство изоморфны друг другу (ничем не отличаются друг от друга). Просто выбор базисных векторов привёл к делению всех векторов на два вида, т.е. к возникновению двух множеств объектов, одни из которых выражаются через другие.

Обратим внимание ещё раз: Вектор ξ есть точка пространства V с компонентами $x^1, \ldots, x^n$, 1-форма ω есть точка сопряжённого пространства V^* с компонентами $a_1, \ldots, a_n$, называемая также ковектором. Однако 1-форма как функция от вектора $\omega(\xi)$ есть скалярная функция $\omega\colon V^* \to \mathbb{R}$, аналогично этому вектор как функция от 1-формы $\xi(\omega)$ есть скалярная функция $\xi\colon V \to \mathbb{R}$. Термин «двойственное» пространство употребляется потому, что каждую из этих величин (векторов и 1-форм) можно рассматривать как линейную функцию, аргументом которой служит другая величина. Поэтому и говорят, что векторы и 1-формы *двойственны* друг другу. В результате можно записать:

$$\xi(\omega) = \omega(\xi) = \omega \cdot \xi = a_1x^1 + \ldots + a_nx^n. \tag{2.27}$$

Величина $\omega \cdot \xi$ называется скалярным произведением ковектора $\omega = (a_1, \ldots, a_n)$ и вектора $\xi = (x^1, \ldots, x^n)$. В квантовой механике Дираком было введено обозначение вектора скобками $|\xi\rangle$, а ковектора скобками $\langle\omega|$. Скалярное произведение ковектора и вектора изображают скобками $\langle\omega|\xi\rangle$. Можно сказать, что ковектор определён полностью, если задано его скалярное произведение с любым вектором, и наоборот.

Пусть в пространстве задано два различных базиса $e_1, \ldots, e_n$ и $\xi_1, \ldots, \xi_n$. Представим векторы базиса $\xi_1, \ldots, \xi_n$ в базисе $e_1, \ldots, e_n$:

$$\xi_j = e_1\xi_j^1 + \ldots + e_n\xi_j^n \text{ или } \xi_j = e_j\xi_j^i \ (i, j = 1, \ldots, n). \tag{2.28}$$

В этом случае

$$x^i(\xi_j) = \xi_j^i. \tag{2.29}$$

Для произвольного вектора η запишем:

$$\eta = e_ix^i = \xi_jX^j = e_i\xi_j^iX^j \text{ или } x^i = \xi_j^iX^j \ (i, j = 1, \ldots, n). \tag{2.30}$$

Аналогично, для произвольной 1-формы ω запишем:

$$\omega(\xi_i) = A_i = \omega(e_i)\xi^i_j = a_i\xi^i_j \text{ или } A_i = a_i\xi^i_j \ (i,j = 1,\ldots,n). \tag{2.31}$$

Закон преобразования базиса $e_1,\ldots,e_n$ в базис $\xi_1,\ldots,\xi_n$ (от начального к конечному базису) $\xi_j = e_i\xi^i_j$ противоположен закону преобразования компонент вектора $x^i = \xi^i_j X^j$ (от конечных к начальным компонентам) и аналогичен закону преобразования компонент 1-формы $A_i = a_i\xi^i_j$ (от начальных к конечным компонентам). Взаимно противоположные («контра»!) и аналогичные («ко»!) законы преобразования в старой литературе именуются контравариантными и ковариантными, и то, что мы называем просто «векторами», раньше называли «контравариантными векторами», а 1-формы называли «ковариантными векторами» или ковекторами. При современном изложении подчёркивается тот факт, что ни векторы, ни 1-формы не меняются при замене базиса; это геометрические объекты, не зависящие от выбора системы координат.

Векторное исчисление — раздел математики, в котором изучаются свойства операций над векторами. Возникновение векторного исчисления тесно связано с потребностями механики и физики. Понятие вектор возникло как математическая абстракция объектов, характеризующихся величиной и направлением, например: перемещение, скорость, напряжение электрического и магнитного полей. До 19 в. для задания векторов использовался лишь координатный способ и операции над векторами сводились к операциям над их координатами. Лишь в середине 19 в. было создано векторное исчисление, в котором операции проводились непосредственно над векторами, без обращения к координатному способу задания. Основы векторного исчисления были заложены исследованиями У. Гамильтона и Г. Грассмана. Современный вид векторному исчислению придал Дж. Гиббс.

2.3 Матрица и определитель

Рассмотрим систему m линейных уравнений с n-неизвестными:

$$y_i = \alpha_{11}x^1 + \ldots + \alpha_{1n}x^n = \alpha_{ij}x^j \quad (i = 1,\ldots,m). \tag{2.32}$$

Эту систему удобно записать в виде *таблицы чисел* или *двухиндексного массива чисел*, который называется *матрица* (лат. matrix — матка, источник, начало):

$$\alpha = \begin{Vmatrix} \alpha_{11}\ldots\alpha_{1n} \\ \cdots\cdots\cdots \\ \alpha_{m1}\ldots\alpha_{mn} \end{Vmatrix}, \text{ тогда } \begin{Vmatrix} y_1 \\ \cdots \\ y_m \end{Vmatrix} = \begin{Vmatrix} \alpha_{11}\ldots\alpha_{1n} \\ \cdots\cdots\cdots \\ \alpha_{m1}\ldots\alpha_{mn} \end{Vmatrix} \begin{Vmatrix} x^1 \\ \cdots \\ x^n \end{Vmatrix}, \tag{2.33}$$

где коэффициенты α_{ij} $(i = 1,\ldots,m; j = 1,\ldots,n)$ называют *элементами матрицы*.

В составлении двухиндексных массивов используются следующие правила: для любого двухиндексного массива с индексами на одном уровне, например, α_{ij} или α^{ij}, первый индекс — это номер строки, второй индекс — номер столбца; если индексы находятся на разных уровнях α^i_j, тогда верхний — номер строки, а нижний — номер столбца.

Матрицы принято выделять двумя вертикальными черточками $\alpha = \|\alpha_{ij}\|$, либо круглыми скобками $\alpha = (\alpha_{ij})$. Если у матрицы m строк и n столбцов, то $(m \times n)$-матрица называется *прямоугольной*; соответственно, $(n \times n)$-матрица называется *квадратной*; $(n \times 1)$-матрица — *вектор-столбцом*, $(1 \times n)$-матрица — *вектор-строкой*.

Число n квадратной матрицы называют *порядком матрицы*. *Главной диагональю* квадратной матрицы α порядка n называется ряд элементов $\alpha_{11}, \ldots, \alpha_{nn}$. Эти элементы называются *диагональными*. Остальные элементы называются *недиагональными*. Если все диагональные элементы имеют ненулевые значения, а все недиагональные элементы — нулевые, то говорят, что матрица имеет *диагональный вид*. *Треугольная матрица* — квадратная матрица, у которой все элементы, расположенные ниже (или выше) главной диагонали, равны нулю (из $i > j$ или $i < j$ следует $\alpha_{ij} = 0$). У *нулевой матрицы* $\|0\|$ все элементы равны нулю. *Единичная матрица* E — это диагональная матрица, все диагональные элементы которой равны единице $E = \|\delta_{ij}\|$.

Элементы в левой части уравнений (2.32) могут быть записаны не с нижним, а с верхним индексом. Запишем такую систему из n линейных уравнений с n неизвестными и квадратной матрицей:

$$y_i = \alpha^i_1 x^1 + \ldots + \alpha^i_n x^n = \alpha^i_j x^j \quad (i = 1, \ldots, m). \tag{2.34}$$

В этом случае элементы матрицы имеют как верхний, так и нижний индексы, чтобы соответствовать компонентам y^i и x^j.

Заданные компоненты $y^1, \ldots, y^n$ называются *свободными членами* системы уравнений. Матрица α, дополненная столбцом свободных членов, называется *расширенная матрица*:

$$\text{если } \left\|\begin{matrix} y_1 \\ \cdots \\ y_n \end{matrix}\right\| = \left\|\begin{matrix} \alpha^1_1 \ldots \alpha^1_n \\ \cdots\cdots\cdots \\ \alpha^n_1 \ldots \alpha^n_n \end{matrix}\right\| \left\|\begin{matrix} x^1 \\ \cdots \\ x^n \end{matrix}\right\| \text{ то } \left\|\begin{matrix} y^1 & \alpha^1_1 \ldots \alpha^1_n \\ \cdots & \cdots\cdots\cdots \\ y^n & \alpha^n_1 \ldots \alpha^n_n \end{matrix}\right\| \tag{2.35}$$

есть расширенная матрица.

Впервые матрица, как математическое понятие появилось в работах У. Гамильтона, А. Кэли и Дж. Сильвестра в середине 19 в. Основы теории матриц созданы К. Вейерштрассом и Г. Фробениусом во второй половине 19 в. и начале 20 в. Современное обозначение — две вертикальные черточки — ввёл А. Кэли (1841). Символ Кронекера был введён Л. Кронекером (1869).

Запись чисел в виде матрицы оказалась чрезвычайно плодотворной для различных разделов математики, так как оказалось, что эти математические объекты обладают рядом полезных свойств.

Во-первых, *матричная запись системы линейных уравнений позволяет упростить и унифицировать алгоритм их решения.*

Решением системы уравнений (2.35) будут такие значения $x^1, \ldots, x^n$, при подстановке которых уравнения системы превращаются в верные тождества. Система имеет однозначное решение только в том случае, если в ней нет тождественных между собой уравнений.

Мы можем вычислить неизвестные, подставляя их из одних уравнений в другие. При этом мы получим решение для каждого неизвестного в виде *формул Крамера* как частное двух величин:

$$x^i = \frac{\Delta^i}{\Delta}. \tag{2.36}$$

Величина Δ или $\det \alpha$ называется *определитель* квадратной матрицы $\|\alpha^i_j\|$ и представляет собой многочлен, каждый член которого является произведением n элементов, взятых по одному из каждой строки и каждого столбца, и снабжённых определённым знаком:

$$\Delta = \det \alpha = \det \|\alpha^i_j\|^n = \begin{vmatrix} \alpha^1_1 \ldots \alpha^n_1 \\ \cdots\cdots\cdots \\ \alpha^1_n \ldots \alpha^n_n \end{vmatrix} = \sum \pm \alpha^1_{j_1} \cdot \ldots \cdot \alpha^n_{j_n}. \tag{2.37}$$

Суммирование производится по всем перестановкам $j_1, \ldots, j_n$ чисел $1, \ldots, n$, и перед членом берется знак «+», если перестановка $j_1, \ldots, j_n$ чётная, и знак «−», если перестановка $j_1, \ldots, j_n$ нечётная. Таким образом, каждый коэффициент содержит $n!$ членов, из которых половина чётная, а половина нечётная. Каждый член определителя является произведением n элементов, взятых по одному из каждой строки и каждого столбца, и снабжён определённым знаком.

Величина Δ^i представляет собой определитель матрицы α, полученной заменой i-ого столбца коэффициентов матрицы столбцом свободных членов $y^1, \ldots, y^n$.

Из формул Крамера следует, что однозначное решение системы n уравнений с n неизвестными существует, если определитель соответствующей матрицы порядка n отличен от нуля $\det \alpha \neq 0$.

Определитель есть основная числовая характеристика квадратной матрицы. Если $\det \alpha = 0$ то говорят, что квадратная матрица α *вырожденная*, *особая* (*особенная*) или *сингулярная* (от лат. singularis — отдельный, особый), в противном случае если $\det \alpha \neq 0$, квадратная матрица называется *невырожденной* или *неособенной*.

Минор (от лат. minor — меньший) k-го порядка $(m \times n)$-матрицы — определитель квадратной $(k \times k)$-матрицы, составленный из элементов исходной $(m \times n)$-матрицы, стоящих на пересечении её произвольно выделенных k

строк и k столбцов с сохранением их порядка:

$$\det \alpha^{i_1 \dots i_k}_{j_1 \dots j_k} = \begin{vmatrix} \alpha^{i_1}_{j_1} \dots \alpha^{i_k}_{j_1} \\ \dots\dots\dots \\ \alpha^{i_n}_{j_k} \dots \alpha^{i_k}_{j_k} \end{vmatrix} = \sum \pm \alpha^{i_1}_{j_1} \cdot \dots \cdot \alpha^{i_k}_{j_k}. \tag{2.38}$$

Суммирование производится по всем перестановкам $j_1, \dots, j_k$ чисел $1, \dots, k$, и перед членом берется знак «+», если перестановка чётная, и знак «−», если перестановка нечётная. Таким образом, каждый минор содержит $k!$ членов, из которых половина чётная, а половина нечётная.

Каждая матрица имеет $C^k_n C^k_m$ миноров k-го порядка. Минорами 1-го порядка являются элементы матрицы α_{ij}. Если номера строк, в которых расположен минор, совпадают с номерами столбцов, то он называется *главным минором*. Квадратная матрица порядка n имеет C^k_n главных миноров k-го порядка. Главными минорами 1-го порядка являются диагональные элементы матрицы.

Базисный минор матрицы — отличный от нуля минор k-го порядка этой матрицы, такой, что все содержащие его миноры $(k+1)$-го порядка равны нулю или же минор $(k+1)$-го порядка не существует.

Дополнительным минором к минору порядка k квадратной матрицы порядка n называется определитель порядка $n-k$, полученный из исходной матрицы вычёркиванием тех k столбцов (строк), в которых расположен базисный минор.

Дополнительный минор M_{ij} к элементу матрицы α_{ij} получается вычёркиванием i-й строки и j-го столбца исходной матрицы α.

Алгебраическое дополнение элемента α_{ij} матрицы α есть число:

$$A_{ij} = (-1)^{i+j} M_{ij}. \tag{2.39}$$

Название «алгебраическое дополнение» связано с формулами разложения определителя матрицы по строке (по столбцу):

$$\det \alpha = \sum_{j=1}^{n} \alpha_{ij} A_{ij}. \tag{2.40}$$

Ранг матрицы: если среди миноров порядка r вырожденной матрицы α хотя бы один отличен от нуля, и все миноры порядка выше r равны нулю, то число r называется рангом $r = \operatorname{rank} \alpha$ этой матрицы. Порядок любого базисного минора матрицы совпадает с рангом матрицы.

Система уравнений (2.33) с $(m \times n)$-матрицей называется *совместной*, если она имеет хотя бы одно решение, и *несовместной*, если она не имеет ни одного решения. Совместная система называется *определенной*, если она имеет единственное решение, и *неопределенной*, если она имеет более одного решения. В последнем случае каждое ее решение называется *частным решением* системы. Совокупность всех частных решений называется *общим решением*.

Теорема Кронекера-Капели даёт исчерпывающий ответ на вопрос о совместности системы:

1. система совместна, тогда и только тогда, когда ранг расширенной матрицы системы равен рангу основной матрицы;

2. если ранг совместной системы равен числу неизвестных, то система имеет единственной решение;

3. если ранг совместной системы меньше числа неизвестных, то система имеет бесконечное множество решений.

Теорема содержится в лекциях, читавшихся Л. Кронекером в 1883–91. А. Капелли (1892) впервые дал формулировку теоремы с использованием термина «ранг матрицы».

Система линейных уравнений называется однородной, если все свободные члены равны нулю $y^1 = \ldots = y^n = 0$. Однородная система всегда совместна, так как имеет *нулевое* (*тривиальное*) решение $x^1 = \ldots = x^n = 0$. Если ранг однородной системы равен размеру матрицы, то тривиальное решение единственно. Если ранг однородной системы меньше размера матрицы, то она имеет бесчисленное множество решений.

Вычисления по формулам Крамера весьма трудоёмки и применимы только для квадратных матриц. Элементарные преобразования матриц позволяют упростить вычисление решений систем уравнений. *Элементарными преобразованиями матриц* являются:

1. перестановка местами двух параллельных рядов матрицы;

2. умножение всех элементов ряда матрицы на ненулевое число;

3. прибавление ко всем элементам ряда матрицы соответствующих элементов параллельного ряда, умноженного на одно и то же число.

Матрица, полученная из исходной матрицы при помощи элементарных преобразований, эквивалентна исходной матрице. Справедливость указанных преобразований вытекает из естественных аналогичных преобразований соответствующей системы уравнений.

Метод Гаусса — метод последовательного исключения переменных — заключается в том, что с помощью элементарных преобразований матрица приводится к эквивалентной матрице ступенчатого (или треугольного вида).

Матрица, полученная из исходной матрицы при помощи элементарных преобразований, *эквивалентна* исходной матрице. Справедливость указанных преобразований вытекает из естественных аналогичных преобразований соответствующей системы уравнений.

Элементарные преобразования можно производить в матричном виде. Две прямоугольные матрицы α и β эквивалентны, если существуют такие

две квадратные невырожденные матрицы ξ и η, что α и β связаны преобразованием $\beta = \eta\alpha\xi$. Такое преобразование эквивалентно соответствующим элементарным преобразованиям матриц.

Формулы Крамера найдены Г. Крамером (1750). Термин «определитель» в современном его значении ввёл О. Коши (1815), хотя ранее (1801) «детерминантом» К. Гаусс назвал дискриминант квадратичной формы. Идея «определителя» восходит к Г. Лейбницу (1693), который пришёл к определителю при решении систем уравнений. Обозначение — вертикальные линии — ввёл А. Кэли (1841).

Во-вторых, *матричная запись системы линейных уравнений позволяет рассматривать эту систему как линейное отображение.*

При этом оказывается, что *матрица является геометрическим объектом, аналогично таким объектам как точка, вектор, фигура и т.д.*

Система линейных уравнений (2.33) есть *линейное отображение*

$$\alpha\colon \mathbb{R}^n \to \mathbb{R}^m$$

n-членных последовательностей $\xi = (x^1, \ldots, x^n) \in \mathbb{R}^n$ в m-членные последовательности $\eta = (y^1, \ldots, y^m)$ или *линейный оператор*

$$\eta = \alpha\,\xi, \tag{2.41}$$

где $\alpha\,\xi$ — произведение матрицы линейного оператора α на вектор ξ.

Произведение матрицы линейного оператора α на вектор ξ означает, что каждый элемент y^i из η есть сумма произведений элементов i-ой строки матрицы на соответствующие компоненты столбца ξ согласно правилу (2.32) $y_i = \alpha_{ij}x^j$ или (2.34) $y^i = \alpha^i_j x^j$.

Термин «линейный» оператора означает, что линейная суперпозиция аргумента даёт линейную суперпозицию оператора:

$$\alpha\,(\xi + \zeta) = \alpha\,\xi + \alpha\,\zeta \quad \text{и} \quad \alpha\,(c\,\xi) = c\,(\alpha\,\xi) \quad (c \in \mathbb{R}).$$

Разница между терминами линейный оператор и матрица состоит в том, что под оператором понимают вектор-функцию линейного отображения, а под матрицей — таблицу чисел. В остальном их свойства одинаковы, более того каждый линейный оператор задаётся своей матрицей.

Линейное преобразование (2.35) векторного пространства в себя задаётся квадратной невырожденной матрицей

$$\alpha\colon \mathbb{R}^n \to \mathbb{R}^n,$$

которая называется *матрицей преобразования* или *матрицей прямого перехода от одного базиса к другому.*

Тождественное преобразование любого вектора в самого себя задаётся единичной матрицей

$$\xi = E\,\xi.$$

$\mathbb{C}$-*линейный оператор* есть линейное отображение между комплексными пространствами

$$\alpha\colon \mathbb{C}^n \to \mathbb{C}^m.$$

Комплексно сопряжённый $\overline{\alpha}$ к α есть оператор

$$\overline{\alpha}\colon \mathbb{C}^n \to \mathbb{C}^m \quad \text{или} \quad \overline{\alpha\xi} = \overline{\alpha}\,\overline{\xi} \quad (\xi \in \mathbb{C}^n).$$

Комплексно сопряжённый оператор также является $\mathbb{C}$-линейным.

Каждую строку (столбец) матрицы можно рассматривать как n-членную последовательность или вектор, заданный в некотором базисе. Прямоугольная $(m \times n)$-матрица есть *кортеж* (франц. cortege) — конечная последовательность из m элементов (в данном случае — векторов).

Множество всех линейных операторов $\alpha\colon \mathbb{R}^n \to \mathbb{R}^m$ и их матриц α представляет собой прямую степень векторных пространств

$$V^m = V \times \ldots \times V = \mathbb{R}^{mn} \tag{2.42}$$

и само является *векторным пространством* размерности mn над полем $\mathbb{R}$, так как на нём можно задать операции линейной суперпозиции:

$$\alpha\,\xi + \beta\,\xi = (\alpha + \beta)\,\xi, \quad (c\,\alpha)\,\xi = \alpha\,(c\,\xi).$$

Множество всех линейных преобразований $\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$ и их матриц является *векторным пространством* размерности n^2 над полем $\mathbb{R}$.

Линейная суперпозиция матриц имеет вид:

1. Суммой $\alpha + \beta$ двух прямоугольных матриц α и β одного размера называется матрица аналогичного размера с элементами $\alpha_{ij} + \beta_{ij}$.
2. Произведением $c\,\alpha$ матрицы α на число c называют матрицу, элементы которой получаются умножением на это число соответствующих элементов исходной матрицы: $c\,\alpha = \|c\,\alpha_{ij}\|$.

Свойства операции линейной суперпозиции ($a, b \in \mathbb{R}$):

$$\begin{aligned}
&\alpha + \beta = \beta + \alpha,\\
&a\,(\alpha + \beta) = a\,\alpha + a\,\beta,\\
&\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma,\\
&(a + b)\,\alpha = a\,\alpha + b\,\alpha,\\
&(a\,b)\,\alpha = a\,(b\,\alpha).
\end{aligned}$$

Аналогично пространству векторов мы можем ввести *норму линейного оператора* $\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$ или $\eta = \alpha\,\xi$. Нормой α называется число:

$$\|\alpha\| = \sup_{\xi \neq 0} \frac{\|\alpha\,\xi\|}{\|\xi\|}. \tag{2.43}$$

Геометрически $\|\alpha\|$ означает наибольший «коэффициент растяжения» преобразования α.

Легко показать, что норма (2.43) имеет следующие свойства:

1. $\|c\alpha\| = |c|\|\alpha\|$, где $c \in \mathbb{R}$,
2. $\|\alpha + \beta\| \leqslant \|\alpha\| + \|\beta\|$,
3. $\|\alpha\beta\| \leqslant \|\alpha\|\|\beta\|$.

Определим *метрику пространства операторов* или *расстояние* между двумя линейными операторами (метрику) как норму их разности

$$\rho(\alpha, \beta) = \|\alpha - \beta\|. \tag{2.44}$$

Докажем что ρ — метрика.

Из определения (2.43) видно, что: $\rho > 0$, если $\alpha \neq \beta$, $\rho(\alpha, \alpha) = 0$, $\rho(\beta, \alpha) = \rho(\alpha, \beta)$. Неравенство треугольника $\rho(\alpha, \beta) + \rho(\beta, \gamma) \geqslant \rho(\alpha, \gamma)$ вытекает из свойства 2 нормы линейного оператора.

Теорема (без доказательства). Пространство L линейных операторов с метрикой ρ является *полным метрическим пространством*.

Две матрицы равны, если определяемое ими линейное отображение произвольного вектора ξ, дает один и тот же результат:

$$\text{если} \quad \alpha\,\xi = \beta\,\xi, \quad \text{тогда} \quad \alpha = \beta \quad \text{и} \quad \|\alpha_{ij}\| = \|\beta_{ij}\|.$$

Если линейный оператор $\alpha\colon \mathbb{R}^m \to \mathbb{R}^k$ задан $(k \times m)$-матрицей $\|\alpha_{ij}\|$, а оператор $\beta\colon \mathbb{R}^n \to \mathbb{R}^m$ — $(m \times n)$-матрицей $\|\beta_{ij}\|$, то оператор

$$\alpha \circ \beta\colon \mathbb{R}^n \to \mathbb{R}^k$$

задан $(k \times n)$-матрицей $\|\gamma_{ij}\|$ и называется суперпозицией (последовательным применением) двух линейных операторов:

$$\eta = \gamma\,\xi = \alpha \circ \beta\,\xi.$$

Умножением матриц называется операция

$$\gamma = \alpha \circ \beta \quad \text{или} \quad \gamma = \alpha\,\beta,$$

которая даёт матрицу оператора суперпозиции.

Умножение определено только для такой пары матриц, у которой число столбцов первого сомножителя равно числу строк второго. При этом каждый элемент γ_{ij} есть сумма произведений элементов i-ой строки матрицы α на соответствующие элементы j-го столбца матрицы β:

$$\gamma_{ij} = \alpha_{i1}\beta_{1j} + \ldots + \alpha_{im}\beta_{mj} \quad (i = 1, \ldots, k; j = 1, \ldots, n).$$

Умножение матриц обладает следующими свойствами ($a, b, c \in \mathbb{R}$):

$$
\begin{aligned}
&(\alpha\beta)\,\gamma = \alpha\,(\beta\gamma),\\
&(\alpha+\beta)\,\gamma = \alpha\gamma + \beta\gamma,\\
&\alpha\,(\beta+\gamma) = \alpha\beta + \alpha\gamma,\\
&(a+b)\,\alpha = a\,\alpha + b\,\alpha,\\
&c\,(\alpha\beta) = (c\,\alpha)\beta = \alpha\,(c\,\beta).
\end{aligned}
$$

$\mathbb{C}$-линейные операторы обладают аналогичными свойствами над полем комплексных чисел. Комплексно сопряжённые операторы обладают также свойствами ($z \in \mathbb{C}$):

$$
\begin{aligned}
&\overline{\alpha+\beta} = \overline{\alpha} + \overline{\beta},\\
&\overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta},\\
&\overline{z\alpha} = \overline{z}\cdot\overline{\alpha},\\
&\det\overline{\alpha} = \overline{\det\alpha}.
\end{aligned}
$$

Транспонированная матрица (нем. transponieren — перекладывать, от лат. transpono — переставляю) получается из данной (прямоугольной или квадратной) $(m \times n)$-матрицы $\alpha = \|\alpha_{ij}\|$ при замене строк столбцами, т.е. $(n \times m)$-матрица $\alpha^T = \|\alpha_{ji}\|$.

Число строк (столбцов) транспонированной матрицы равно числу столбцов (строк) исходной матрицы. При этом для квадратных матриц справедливо равенство $\det\alpha^T = \det\alpha$.

Для любой невырожденной $\det\alpha \neq 0$ квадратной матрицы α существует единственная *обратная матрица* α^{-1}, такая что:

$$
\alpha^{-1}\alpha = E, \quad \det\alpha^{-1}\cdot\det\alpha = 1. \tag{2.45}
$$

Для существования обратной матрицы необходимо и достаточно, чтобы линейное преобразование $\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$ было *взаимно однозначным* (для любых двух различных векторов $\xi_1 \neq \xi_2$ справедливо $\alpha\,\xi_1 \neq \alpha\,\xi_2$).

Определение обратной матрицы аналогично определению обратной функции, т.е. если она существует, то $\xi = \alpha^{-1}\eta$ при $\eta = \alpha\,\xi$. Тогда систему уравнений (2.35) можно записать при помощи обратной матрицы в виде:

$$
\xi = \alpha^{-1}\eta = \alpha^{-1}\alpha\,\xi.
$$

Таким образом, решение системы уравнений сводится к вычислению обратной матрицы.

Для вычисления обратной матрицы применяется формула:

$$
\alpha^{-1}_{ij} = \frac{A_{ji}}{\Delta}. \tag{2.46}
$$

На практике вычисление обратной матрицы осуществляют так:

1. выписывают транспонированную матрицу;

2. заменяют каждый элемент транспонированной матрицы своим алгебраическим дополнением (2.39);

3. делят полученную матрицу на ее определитель (2.37 или 2.40).

Допустим, что n векторов $\xi_1, \ldots, \xi_n$ ($\xi_i = e_1\xi_i^1 + \cdots + e_n\xi_i^n = e_j\xi_i^j$), заданных в n-мерном пространстве с базисом $e_1, \ldots, e_n$, линейно зависимы, тогда $\xi_1 c^1 + \ldots + \xi_n c^n = \xi_i c^i = e_j \xi_i^j c^i = 0$ $(i, j = 1, \ldots, n)$.

Каждая сумма $\xi_i^j c^j$ представляет собой компоненту соответствующего базисного вектора e_j. Эта компонента равна нулю, так как вектор $(\xi_i^1 c^i, \ldots, \xi_i^n c^i)$ равен нулю, если равны нулю все его компоненты. Совокупность нулей образует нулевой вектор в n-мерном пространстве. Последовательность чисел $c^1, \ldots, c^n$ также образует некий вектор $\mathbf{c}$ в n-мерном пространстве, тогда:

$$\begin{Vmatrix} \xi_1^1 \ldots \xi_n^1 \\ \cdots\cdots\cdots \\ \xi_1^n \ldots \xi_n^n \end{Vmatrix} \begin{Vmatrix} c^1 \\ \cdots \\ c^n \end{Vmatrix} = \begin{Vmatrix} \xi_i^1 c^i \\ \cdots \\ \xi_i^n c^i \end{Vmatrix} = \begin{Vmatrix} 0 \\ \cdots \\ 0 \end{Vmatrix} \quad \text{или} \quad \xi\,\mathbf{c} = 0. \tag{2.47}$$

Множество всех векторов $\mathbf{c} \in V$, для которых $\xi\,\mathbf{c} = 0$, называется *ядром* линейного оператора ξ и обозначается символом $\ker\xi$ (от англ. kernel — ядро). Если существует обратная матрица ξ^{-1}, то из условия $\xi\,\mathbf{c} = 0$ следует, что $\mathbf{c} = \xi^{-1}\xi\,\mathbf{c} = \alpha^{-1}\mathbf{0} = \mathbf{0}$, т.е. $\ker\xi = 0$. Следовательно, если $\ker\xi = 0$ то оператор $\xi\colon \mathbb{R}^n \to \mathbb{R}^n$ действует взаимно однозначно. Таким образом, условие $\ker\xi = 0$ является необходимым и достаточным для существования обратной матрицы.

Если $\ker\xi = 0$, то $\det\alpha \neq 0$.

Образом $\operatorname{im}\alpha$ линейного оператора α называется множество всех векторов η, представляемых в виде $\eta = \alpha\xi$.

Если $\ker\alpha = 0$, то $\operatorname{im}\alpha = V$.

Обратная матрица произведения, равна произведению обратных матриц, взятых в обратном к исходному порядке:

$$(\alpha\beta\ldots\omega)^{-1} = \omega^{-1}\ldots\beta^{-1}\alpha^{-1}. \tag{2.48}$$

Для квадратных матриц γ, α и α' $(\det\gamma \neq 0)$ выражения вида

$$\alpha' = \gamma^{-1}\alpha\gamma \tag{2.49}$$

называются *преобразование подобия.*

Матрицы α и α', связанные преобразованием подобия называются *подобными.* У подобных матриц одинаковый порядок и ранг.

Свойства преобразования подобия: если $\alpha' = \gamma^{-1}\alpha\gamma$ и $\beta' = \gamma^{-1}\beta\gamma$, то

$$\begin{aligned}
&\alpha' + \beta' = \gamma^{-1}(\alpha + \beta)\gamma,\\
&\alpha'\beta' = \gamma^{-1}(\alpha\beta)\gamma,\\
&\alpha'^n = \gamma^{-1}\alpha^n\gamma,\\
&f(\alpha') = \gamma^{-1}f(\alpha)\gamma,\\
&E = \gamma^{-1}E\gamma,\\
&\text{из} \quad [\alpha, \beta] = 0 \quad \text{следует, что} \quad [\alpha', \beta'] = 0,\\
&\det(\gamma^{-1}\alpha\gamma) = \det(\alpha).
\end{aligned}$$

Степень n квадратной матрицы α (порядка k) есть матрица:

$$\alpha^n = \underbrace{\alpha \cdot \ldots \cdot \alpha}_{n} \tag{2.50}$$

Возведение в степень n диагональной матрицы α (порядка k) дает:

$$\begin{Vmatrix} \alpha_{11}^n & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & \alpha_{kk}^n \end{Vmatrix} \tag{2.51}$$

Случай отрицательных степеней n не представляет трудностей, если только матрица α невырожденная.

Степень матрицы дает возможность определить многочлен от квадратной матрицы:

$$P_n(\alpha) = a_n\alpha^n + a_{n-1}\alpha^{n-1} + \ldots + a_1\alpha + a_0E. \tag{2.52}$$

Рассматриваются также аналитические функции от матрицы:

$$f(\alpha) = \sum_{n=0}^{\infty} a_n\alpha^n. \tag{2.53}$$

Можно задать остальные элементарные функции от матрицы:

$$e^{\alpha} \overset{(1.30)}{=} \lim_{n\to\infty}\left(E + \frac{\alpha}{n}\right) = \sum_{n=0}^{\infty}\frac{\alpha^n}{n!}, \quad \text{где } \alpha^0 = E \tag{2.54}$$

$$\cos\alpha \overset{(1.31)}{=} \sum_{n=1}^{\infty}(-1)^n\frac{\alpha^{2n}}{(2n)!}, \tag{2.55}$$

$$\sin\alpha \overset{(1.32)}{=} \sum_{n=1}^{\infty}(-1)^n\frac{\alpha^{2n+1}}{(2n+1)!}. \tag{2.56}$$

Все элементарные функции матрицы также есть матрицы.

Матрицы α комплексно сопряжённого оператора имеет свойство:

$$\overline{e^{\alpha}} = e^{\overline{\alpha}}. \tag{2.57}$$

Пусть матрица оператора $\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$ диагональная, с диагональными элементами $\lambda_1, \ldots, \lambda_n$. Легко видеть, что матрица оператора e^{α} также диагональная, с диагональными элементами $e^{\lambda_1}, \ldots, e^{\lambda_n}$.

Пример. Если рассматривать комплексное число как точку на вещественной плоскости $\mathbb{R}^2$ в базисе $e_1 = 1$, $e_2 = i$, а умножение на комплексное число z как преобразование плоскости $\alpha\colon \mathbb{R}^2 \to \mathbb{R}^2$, то оператор α есть поворот на угол $\arg z$ с растяжением в $|z|$ раз (1.61). Матрица α умножения на комплексное число $z = x + iy$ имеет вид

$$\alpha = \begin{Vmatrix} x & -y \\ y & x \end{Vmatrix}. \tag{2.58}$$

Тогда e^{α} есть оператор умножения на комплексное число e^z.

Умножение матриц — некоммутативная операция, т.е. порядок, в котором производятся операции, небезразличен, и в большинстве случаев матрица $\alpha\beta$ отличается от $\beta\alpha$. Кроме того, возникают проблемы с делением их друг на друга. Поэтому матрицы образуют некоммутативное кольцо. Если отсеять нехорошие матрицы, то можно сконструировать приличное поле. Для этого необходимо выполнение свойства (2.45).

Выражение:

$$[\alpha\beta] = \alpha\beta - \beta\alpha. \tag{2.59}$$

обычно обозначают квадратными скобками и называют *коммутатором* квадратных $(n \times n)$ матриц α и β.

Свойства коммутатора:

1. $[\alpha\beta] = -[\beta\alpha]$ антикоммутативность,
2. $[\alpha, \beta + \gamma] = [\alpha\beta] + [\alpha\gamma]$ дистрибутивность,
3. $c[\alpha\beta] = [c\alpha, \beta] = [\alpha, c\beta]$ (c — число) ассоциативность,
4. $[\alpha[\beta\gamma]] + [\beta[\gamma\alpha]] + [\gamma[\alpha\beta]] = 0$ тождество Якоби,
5. $[\alpha^{-1}, \alpha] = 0$, $[c, \alpha] = 0$ (c — число),
6. если $[\alpha\beta] = 0$, то $\alpha\beta = \beta\alpha$ — матрицы *перестановочны* (косоортогональны),
7. если $[\alpha\beta] = 0$ и α невырожденная матрица, то $\beta = f(\alpha)$,
8. если невырожденная матрица β коммутирует с диагональной матрицей α, то матрица β также диагональная.

Не следует забывать о том, что матрица значительнее числа. У неё есть свои, присущие только ей операции. Например, она имеет горизонтальный и вертикальный размеры, её можно ставить набок и транспонировать (опрокидывать).

Перечислим теперь общие свойства матриц и определителей.

Свойства квадратных матриц:

1. Все квадратные матрицы n-го порядка образуют кольцо.
2. Все квадратные матрицы n-го порядка образуют коммутативную группу относительно операции сложения.
3. Все неособенные квадратные матрицы n-го порядка образуют некоммутативную группу относительно операции умножения.

Свойства треугольных матриц:

1. Диагональная матрица является частным случаем как верхней, так и нижней треугольных матриц.
2. Матрица, обратная диагональной матрицы, также диагональная. Элементы ее, не равные нулю, представляют собой обратные величины диагональных элементов заданной матрицы.
3. Так как определитель треугольной матрицы равен произведению её диагональных элементов, то треугольная (диагональная) матрица является неособенной тогда, и только тогда, когда все её диагональные элементы отличны от нуля.
4. Сумма и произведение двух верхних или нижних треугольных (диагональных) матриц — тоже верхняя или нижняя треугольная (диагональная матрица) и эта операция коммутативная.
5. Все верхние или нижние треугольные и диагональные матрицы n-го порядка образуют три коммутативные группы относительно операции сложения.
6. Все неособенные верхние (нижние) треугольные матрицы образуют некоммутативную группу относительно умножения.
7. Все неособенные диагональные матрицы образуют коммутативную группу относительно умножения.

Каждое строка (столбец) матрицы или определителя представляют собой вектор. Если два уравнения системы тождественны, то два вектора определителя линейно зависимы. Это означает, что одну строку определителя можно сделать нулевой, но определитель при этом не изменится. Определитель с нулевой строкой тождественно равен нулю, так как в каждом члене многочлена будет нулевой сомножитель. Следовательно, определитель составленый из линейно зависимых векторов равен нулю; и наоборот,

если определитель отличен от нуля, то векторы, из которых он составлен, линейно независимы и, следовательно, образуют базис и задают векторное пространство.

Свойства определителя непосредственно следуют из его представления в виде суммы многочленов:

1. определитель не изменится, если поменять местами строки и столбцы (транспонировать);

2. от перемены мест двух строк или двух столбцов, входящих в определитель, его значение меняет знак;

3. общий множитель всех элементов любой строки (столбца) можно вынести за знак определителя;

4. если каждый элемент какой-либо строки (столбца) есть сумма двух слагаемых, то определитель равен сумме двух определителей, причём в одном из них соответствующая строка (столбец) состоит из первых слагаемых, а в другом — из вторых слагаемых, остальные же строки (столбцы) остаются те же, что и в исходном;

5. определитель не изменяется, если к элементам одной строки (столбца) прибавить элементы другой строки (столбца), умноженные на произвольный множитель;

6. определитель равен нулю, если элементы двух строк (или двух столбцов) соответственно пропорциональны;

7. определитель треугольной матрицы равен произведению элементов её главной диагонали;

8. из правила умножения матриц непосредственно видно, что определитель произведения матриц равен произведению определителей: если $\gamma = \alpha\beta$, то $\det\gamma = \det(\alpha\beta) = \det\alpha \cdot \det\beta$, понятно, что это замечание имеет смысл только в случае квадратных матриц.

2.4 Евклидова и эрмитова структуры

Рассмотрим скалярные функции заданные на векторах.

Пусть пространство $V^m = V \times \ldots \times V = \mathbb{R}^{mn}$ — есть m-кратное произведение и $\xi_1, \ldots, \xi_m \in V$ — векторы. Скалярная функция $\omega\colon V \to \mathbb{R}$ или $\omega(\xi_1, \ldots, \xi_m)$ называется *полилинейной функцией* (билинейной, трилинейной и т.д.) или *m-линейной формой* (коротко, *m-формой*), если для всякого $i = 1, \ldots, m$ и каждого $c \in \mathbb{R}$ имеем:

$$\begin{aligned}
&\omega(\xi_1, \ldots, \xi_i + \xi_i', \ldots, \xi_m) = \\
&= \omega(\xi_1, \ldots, \xi_i, \ldots, \xi_m) + \omega(\xi_1, \ldots, \xi_i', \ldots, \xi_m), \\
&\omega(\xi_1, \ldots, c\,\xi_i, \ldots, \xi_m) = c\,\omega(\xi_1, \ldots, \xi_i, \ldots, \xi_m),
\end{aligned} \tag{2.60}$$

т.е. полилинейная форма линейна по каждому аргументу.

Форму $\omega(\xi_1, \ldots, \xi_m)$ называют *симметрической*, если она не меняется при перестановке любых двух различных аргументов:

$$\omega(\xi_1, \ldots, \xi_i, \ldots, \xi_j, \ldots, \xi_m) = \omega(\xi_1, \ldots, \xi_j, \ldots, \xi_i, \ldots, \xi_m). \tag{2.61}$$

Форму $\omega(\xi_1, \ldots, \xi_m)$ называют *кососимметрической* или *внешней*, если она меняет знак при перестановке любых двух аргументов:

$$\omega(\xi_1, \ldots, \xi_i, \ldots, \xi_j, \ldots, \xi_m) = -\omega(\xi_1, \ldots, \xi_j, \ldots, \xi_i, \ldots, \xi_m). \tag{2.62}$$

В этих равенствах векторы ξ_i и ξ_j меняются местами, все же остальные векторы остаются на своём месте.

В этом параграфе мы рассмотрим билинейные симметрические формы, а ниже полилинейные кососимметрические (внешние) формы.

Рассмотрим скалярное произведение двух векторов в n-мерном вещественном пространстве, заданных в базисе $\xi_1, \ldots, \xi_n$:

$$\nu = \xi_1 a^1 + \ldots + \xi_n a^n \text{ и } \tau = \xi_1 b^1 + \ldots + \xi_n b^n. \tag{2.63}$$

Исходя из свойств скалярного произведения (с. 47), получаем:

$$\begin{aligned} \nu \cdot \tau &= (\xi_1 a^1 + \ldots + \xi_n a^n) \cdot (\xi_1 b^1 + \ldots \xi_n b^n) = \\ &= \sum_{i,j=1}^{n} (\xi_i \cdot \xi_j)\, a^i b^j = \sum_{i,j=1}^{n} g_{ij}\, a^i b^j \end{aligned} \tag{2.64}$$

где $g_{ij} = \xi_i \cdot \xi_j$ — скалярное произведение соответствующих пар базисных векторов, причём $g_{ij} = g_{ji}$.

Скалярное произведение вектора ν на себя:

$$\nu \cdot \nu = |\nu|^2 = \sum_{i,j=1}^{n} g_{ij}\, a^i a^j \tag{2.65}$$

задаёт метрику в n-мерном пространстве.

Рассмотрим эрмитово скалярное произведение двух векторов (2.63) в пространстве $\mathbb{C}^n$, заданных в произвольном базисе $\xi_1, \ldots, \xi_n$:

Исходя из свойств эрмитова скалярного произведения (с. 48):

$$\begin{aligned} \nu \cdot \tau &= (\xi_1 a^1 + \ldots + \xi_n a^n) \cdot (\overline{\xi_1 b^1} + \ldots \overline{\xi_n b^n}) = \\ &= \sum_{i,j=1}^{n} (\xi_i \cdot \overline{\xi_j})\, a^i \overline{b^j} = \sum_{i,j=1}^{n} g_{ij}\, a^i \overline{b^j} \end{aligned} \tag{2.66}$$

где $g_{ij} = \xi_i \cdot \overline{\xi_j}$ — эрмитово скалярное произведение соответствующих пар базисных векторов, причём $g_{ij} = \overline{g_{ji}}$.

Эрмитово скалярное произведение вектора ν на себя:

$$\nu \cdot \nu = |\nu|^2 = \sum_{i,j=1}^{n} g_{ij}\, a^i \overline{a^j} \tag{2.67}$$

задаёт метрику в n-мерном комплексном пространстве.

Если стандартный базис состоит из попарно ортогональных (от греч. *ορθογωνιος* — прямоугольный, синоним перпендикулярный) векторов, то говорят, что *базис ортонормированный* (ортогональный и нормированный), в противном случае — базис *косоугольный* или *произвольный*. При ортонормированном базисе: $g_{ij} = e_i e_j = 0$ при $i \neq j$, $g_{ii} = e_i \cdot e_i = 1$.

В n-мерном ортонормированном вещественном пространстве результатом скалярного произведения двух векторов ν и τ будет выражение, называемое *евклидовой структурой* $(\ ,\ )$:

$$\nu \cdot \tau = (\nu, \tau) = a^1 b^1 + \ldots + a^n b^n. \tag{2.68}$$

Конечномерное векторное пространство над полем $\mathbb{R}$, в котором скалярное произведение векторов определено евклидовой структурой, называется *евклидово n-мерное арифметическое пространство* E^n.

Пространство евклидово тогда и только тогда, когда базис ортогональный. Обратное утверждение также справедливо. Евклидова структура аналогична скалярному произведению 1-формы и вектора $(\omega, \xi) = a_1 x^1 + \ldots + a_n x^n$, но выражение для скалярного произведения (ω, ξ) справедливо в любом пространстве, не обязательно евклидовом.

Если векторы перпендикулярны, то, в силу свойства 3 скалярного произведения (см. стр. 47), получим

$$a^1 b^1 + \ldots + a^n b^n = 0. \tag{2.69}$$

Скалярное произведение вектора ν на себя:

$$\nu \cdot \nu = (\nu, \nu) = |\nu|^2 = (a^1)^2 + \ldots + (a^n)^2 \tag{2.70}$$

задаёт *метрику* в евклидовом пространстве.

В n-мерном евклидовом пространстве уравнение $(n-1)$-мерной плоскости (2.16) можно записать как скалярное произведение $\mathbf{n} \cdot \mathbf{r} = const$ векторов, где $\mathbf{n} = (\xi_1, \ldots, \xi_n)$ и $\mathbf{r} = (x^1, \ldots, x^n)$. Вектор $\mathbf{r}$ — это радиус-вектор (см. § 2.1). Допустим теперь, что радиус-вектор $\mathbf{r}$ лежит на плоскости, а $\mathbf{n} \cdot \mathbf{r} = 0$. Тогда в силу свойства 3 скалярного произведения (см. стр. 47) вектор $\mathbf{n}$, очевидно, перпендикулярен радиус-вектору, и поэтому называется *нормалью* плоскости (франц. normal, от лат. normalis — прямой). Если, в свою очередь, $\mathbf{n}$ — нормаль, то уравнение

$$\mathbf{n} \cdot \mathbf{r} = const \tag{2.71}$$

с произвольным радиус-вектором $\mathbf{r}$ задаёт произвольную $(n-1)$-мерную плоскость при помощи её n-мерной нормали.

Если мы хотим выяснить, перпендикулярны или параллельны прямые и плоскости, то мы должны проверить, перпендикулярны или коллинеарны их направляющие векторы и нормали. У параллельных плоскостей коллинеарные (2.14) нормали $\mathbf{n}_1 = c\mathbf{n}_2$. У перпендикулярных плоскостей перпендикулярные нормали $\mathbf{n}_1 \cdot \mathbf{n}_2 = 0$. Прямая параллельна плоскости, если ее

направляющий вектор перпендикулярен нормали плоскости (2.69) $\mathbf{n} \cdot \mathbf{s} = 0$. Прямая перпендикулярна плоскости, если её направляющий вектор $\mathbf{s}$ и нормаль плоскости коллинеарные (2.14) $\mathbf{n} = c\mathbf{s}$.

На свойствах скалярного произведения основаны формулы аналитической геометрии в трёхмерном евклидовом пространстве E^3. В качестве координат пространства E^3 обычно выбирают обозначения x, y, z.

Тогда:

- Каноническое уравнение прямой, проходящей через две точки (x_1, y_1, z_1) и (x_2, y_2, z_2) имеет вид (2.14):

$$\frac{x_2 - x_1}{n} = \frac{y_2 - y_1}{m} = \frac{z_2 - z_1}{k},$$

  где n, m и k — проекции направляющего вектора на оси координат.

- Уравнение прямой в параметрическом виде:

$$x = x_0 + mt, \quad y = y_0 + nt, \quad z = z_0 + kt.$$

- Уравнение (2.16) плоскости в пространстве и прямой на плоскости:

$$ax + by + cz = \mathbf{n} \cdot \mathbf{r} = const, \quad ax + by = \mathbf{n} \cdot \mathbf{r} = const;$$

  где a, b и c — координаты вектора нормали;

  $const$ — величина, характеризующая минимальное расстояние от плоскости (прямой) до начала координат, в частности плоскость, перпендикулярная координатной плоскости (x, y), задается как $z = const$, при этом выражение $ax + by$ также равно константе.

- Расстояние l от точки x_0, y_0 до прямой:

$$l \overset{\substack{(2.16)\\(2.70)}}{=} \frac{\mathbf{n} \cdot \mathbf{r}_0 + c}{|\mathbf{n}|} = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}.$$

- Расстояние l от точки x_0, y_0, z_0 до плоскости:

$$l \overset{\substack{(2.16)\\(2.70)}}{=} \frac{\mathbf{n} \cdot \mathbf{r}_0 + d}{|\mathbf{n}|} = \frac{|ax_0 + by_0 + cy_0 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

- Для перпендикулярных плоскостей (2.69): $a_1a_2 + b_1b_2 + c_1c_2 = 0$.
- Для параллельных плоскостей (2.14): $\frac{a_1}{a_2} = \frac{b_1}{b_2} = \frac{c_1}{c_2}$.
- Для параллельных прямой и плоскости (2.69): $an + bm + ck = 0$.
- Для перпендикулярных прямой и плоскости (2.14): $\frac{a}{n} = \frac{b}{m} = \frac{c}{k}$.

- Если прямая отсекает на осях отрезки a и b, либо плоскость отсекает отрезки a, b и c, то уравнение плоскости (прямой) записывается в виде: $\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1$.

В n-мерном ортонормированном комплексном пространстве результатом эрмитова скалярного произведения двух векторов ν и τ будет выражение, называемое *эрмитовой структурой* $\langle\,,\rangle$:

$$\begin{aligned} \nu \cdot \tau = \langle \nu, \tau \rangle &= a^1\overline{b^1} + \ldots + a^n\overline{b^n}, \\ \nu \cdot \nu = \langle \nu, \nu \rangle &= a^1\overline{a^1} + \ldots + a^n\overline{a^n}. \end{aligned} \tag{2.72}$$

Аналогично n-мерному евклидову пространству, имеющему евклидову структуру, n-мерное комплексное ортонормированное пространство $\mathbb{C}^n$, имеющее эрмитову структуру, называют эрмитово пространство, комплексно евклидово пространство или *унитарное пространство* (франц. unitare, от лат. unitas — единство).

Унитарное пространство не предполагается обязательно конечномерным. В унитарном пространстве точно так же как и в евклидовых пространствах, вводится понятие ортогональности и ортонормированной системы векторов, а в конечномерном случае доказывается существование ортонормированного базиса.

Билинейной формой в евклидовом (унитарном) пространстве называют евклидову (эрмитову) структуру с участием матриц:

$$(\alpha\nu, \tau) = \sum_{i,l=1}^{n} \alpha_{ij}\, a^i b^j \quad \text{или} \quad \langle \alpha\nu, \tau \rangle = \sum_{i,j=1} \alpha_{ij}\, a^i \overline{b^j}, \tag{2.73}$$

где $\alpha\nu$ и τ — векторы.

Сопряжённым линейным преобразованием $\alpha^*\colon N \to M$ к линейному преобразованию $\alpha\colon N \to M$ евклидовых (унитарных) пространств называется такое, что для любых $\nu \in M$ и $\tau \in N$ имеет место равенство: $(\alpha\nu, \tau) = (\nu, \alpha^*\tau)$ или $\langle \alpha\nu, \tau \rangle = \langle \nu, \alpha^*\tau \rangle$, где линейный оператор α^* называется *сопряжённым оператором*. Матрица α^* сопряжённого линейного преобразования называется *сопряжённой матрицей*.

Теорема. Любое линейное преобразование конечномерных евклидовых (или унитарных) пространств обладает сопряжённым, которое определяется единственным образом:

- В евклидовых пространствах сопряжённая матрица является транспонированной матрицей $\alpha^* = \alpha^T$.

- Сопряжёние в унитарном пространстве предполагает не только транспонирование, но и замену полученных элементов на комплексно сопряжённые $\alpha^* = \overline{\alpha^T}$.

Свойства сопряжённых матриц:

1. $(\alpha + \beta)^* = \alpha^* + \beta^*$;

2. $(\lambda\alpha)^* = \overline{\lambda}\alpha^*$, где $\lambda \in \mathbb{C}$;

3. $(\alpha\beta\ldots\omega)^* = \omega^*\ldots\beta^*\alpha^*$;

4. $(\alpha^*)^{-1} = (\alpha^{-1})^*$.

Рассмотрим билинейные формы и квадратные матрицы со специальными свойствами симметрии:

1. Матрица α называется *эрмитовой*, если $\alpha_{ij} = \overline{\alpha_{ji}}$, т.е. элементы, симметричные относительно главной диагонали, являются комплексно сопряжёнными числами. Элементы главной диагонали эрмитовой матрицы обязательно вещественны. Определитель эрмитовой матрицы действительный.

   Суперпозиция эрмитовых матриц дает эрмитову матрицу, возведение в степень эрмитовой матрицы также дает эрмитову матрицу. Обратная к эрмитовой матрице также эрмитова. Произведение двух эрмитовых матриц в общем случае не является эрмитовым, но симметризованное произведение $\gamma = \frac{1}{2}(\alpha\beta + \beta\alpha)$ эрмитово. Любая вещественная функция от эрмитовой матрицы также дает эрмитову матрицу. Если символ f означает действительную функцию двух переменных, и матрицы α и β эрмитовы, то $f(\alpha, \beta)$ также эрмитова матрица.

   Эрмитово сопряженная матрица α^+ получается транспонированием исходной матрицы α и заменой полученных элементов на комплексно сопряжённые, что опять равносильно транспонированию. Поэтому в унитарном пространстве исходная эрмитова матрица и эрмитово сопряжённая матрица равны: $\alpha = \alpha^+ = \alpha^*$, и эрмитову матрицу называют также *самосопряжённой матрицей*.

   Эрмитовой формой в унитарном пространстве называют выражение

$$\langle\alpha\nu, \nu\rangle = \langle\nu, \alpha^+\nu\rangle = \sum_{i,j=1}^{n} \alpha_{ij}\, a^i\overline{a^j}, \tag{2.74}$$

   где α — эрмитова матрица,

   α^+ — эрмитово сопряженная матрица.

2. Частным случаем эрмитовой матрицы является *симметрическая матрица* α, для которой $\alpha_{ij} = \alpha_{ji}$, т.е. элементы, симметричные относительно главной диагонали, равны между собой. Для симметрической матрицы $\alpha = \alpha^*$.

   Квадратичной формой в евклидовом пространстве называют выражение

$$(\alpha\nu, \nu) = (\nu, \alpha\nu) = \sum_{i,j=1}^{n} \alpha_{ij}\, a^i a^j, \tag{2.75}$$

   где α — симметрическая матрица.

Квадратичная (эрмитова) форма будет *положительно-определенной*, если векторы $\alpha\nu$ и ν всегда образуют острый угол, *отрицательно-определенной*, если угол всегда тупой, и *неопределенной*, если при разных значениях ν угол бывает как острым, так и тупым.

Теория квадратичных форм впервые изложена А. Лежандром (1798). Общая теория квадратичных форм создана К. Гауссом (1801); ему же принадлежит термин «квадратичная форма».

3. Матрица α называется *косоэрмитовой* (*альтернирующей*), если $\alpha_{ij} = -\overline{\alpha_{ji}}$. Для косоэрмитовой матрицы $\alpha = -\alpha^+$.

   Элементы главной диагонали косоэрмитовой матрицы чисто мнимые. Определитель косоэрмитовой матрицы порядка n действительный — если n четно, и чисто мнимый — если n нечётно.

4. Частным случаем косоэрмитовой матрицы является *кососимметрическая матрица* α, для которой $\alpha_{ij} = -\alpha_{ji}$. Элементы главной диагонали кососимметрической матрицы равны нулю.

   Для кососимметрической матрицы $\alpha^* = -\alpha$. Тогда

$$\det\alpha = \det\alpha^* = \det(-\alpha) = (-1)^n \det\alpha. \tag{2.76}$$

   Следовательно, определитель кососимметрической матрицы *нечётного порядка* равен нулю $\det\alpha \equiv 0$, так как $\det\alpha = -\det\alpha$.

   Кососимметрической формой в евклидовом пространстве называют выражение

$$(\alpha\nu, \tau) = \sum_{i,j=1}^{n} \alpha_{ij}\, a^i b^j, \tag{2.77}$$

   где α — кососимметрическая матрица.

   Выражение

$$(\alpha\nu, \nu) = \sum_{i,j=1}^{n} \alpha_{ij}\, a^i a^j \equiv 0, \tag{2.78}$$

   тождественно равно нулю в том, и только в том случае, если α — кососимметрическая матрица.

Матрицей Грама называют квадратную матрицу $G = \|g_{ij}\|^n$, составленную из скалярных произведений $g_{ij} = \xi_i \cdot \xi_j$ векторов $\xi_1, \ldots, \xi_n$ евклидова (унитарного) пространства (2.64—2.67).

Матрица Грама является симметрической $g_{ij} = g_{ji}$ (эрмитовой $g_{ij} = \overline{g_{ji}}$) положительно полуопределённой (неотрицательной) матрицей. Справедливо обратное утверждение: любая положительно полуопределённая симметрическая (эрмитова) матрица является матрицей Грама некоторой системы векторов евклидова (унитарного) пространства. Эта система векторов образует базис тогда и только тогда, когда её матрица Грама — положительно определённая.

2.5 Преобразование пространства

Ортогональным (*унитарным*) *преобразованием* называют линейное преобразование σ евклидова (унитарного) пространства V, сохраняющее скалярное (эрмитово скалярное) произведение векторов:

$$(\sigma\nu, \sigma\tau) = (\nu, \tau), \quad \langle\sigma\nu, \sigma\tau\rangle = (\nu, \tau) \quad \text{для любых } \nu, \tau \in V. \tag{2.79}$$

Ортогональное (унитарное) преобразование сохраняет длины векторов и углы между ними. Обратно, всякое линейное преобразование евклидова пространства, сохраняющее длины векторов и углы между ними, есть ортогональное (унитарное) преобразование.

Ортогональные (унитарные) матрицы σ служат матрицами преобразования (перехода) от одного ортонормированного и ортогонального базиса к другому в евклидовом (унитарном) пространстве.

Запишем вектор η в двух ортонормированных базисах (исходном $e_1, \ldots, e_n$ и конечном $\xi_1, \ldots, \xi_n$):

$$\eta = e_1 x^1 + \ldots + e_n x^n = \xi_1 X^1 + \ldots + \xi_n X^n \text{ или } e_i x^i = \xi_j X^j.$$

Представим векторы базиса $\xi_1, \ldots, \xi_n$ в базисе $e_1, \ldots, e_n$

$$\xi_j = e_1 \sigma_j^1 + \ldots + e_n \sigma_j^n \text{ или } \xi_j = e_i \sigma_j^i \ (i, j = 1, \ldots, n). \tag{2.80}$$

Тогда получаем закон преобразования компонент вектора η:

$$\eta = e_i x^i = \xi_j X^j = e_i \sigma_j^i X^j \text{ или } x^i = \sigma_j^i X^j \ (i, j = 1, \ldots, n). \tag{2.81}$$

Запишем этот закон при помощи матрицы:

$$\begin{Vmatrix} x^1 \\ \cdots \\ x^n \end{Vmatrix} = \begin{Vmatrix} \sigma_1^1 \ldots \sigma_n^1 \\ \cdots\cdots\cdots \\ \sigma_1^n \ldots \sigma_n^n \end{Vmatrix} \begin{Vmatrix} X^1 \\ \cdots \\ X^n \end{Vmatrix} \quad \text{или} \quad \mathbf{x} = \sigma\mathbf{X}, \tag{2.82}$$

где $\mathbf{x} = (x^1, \ldots, x^n)$, $\mathbf{X} = (X^1, \ldots, X^n)$.

Каждый j-ый столбец матрицы σ представляет собой базисный вектор ξ_j, заданный в базисе $e_1, \ldots, e_n$.

С учетом свойства 3 сопряжённых матриц (см. с. 74) получаем $\mathbf{x}^* = \mathbf{X}^*\sigma^*$. Если σ является ортогональным преобразованием, то оно должно сохранять скалярное произведение векторов $\mathbf{x}^*\mathbf{x} = \mathbf{X}^*\mathbf{X}$, тогда $\mathbf{x}^*\mathbf{x} = \mathbf{X}^*\sigma^*\sigma\mathbf{X}$ или $\sigma^*\sigma = E$. С учётом $\sigma^{-1}\sigma \overset{(2.45)}{=} E$, получаем:

$$\sigma^* = \sigma^{-1}. \tag{2.83}$$

Запись $\mathbf{x}^*\mathbf{x}$ в данном доказательстве означает, что скалярное произведение осуществляется по правилу умножения матриц, когда каждый элемент произведения есть сумма произведений элементов соответствующей строки и столбца. Определённое таким образом скалярное произведение уже

не симметрично, так как вектор-строку можно отождествить с ковектором (1-формой), а сопряжение транспонирует вектор-столбец в вектор-строку.

Равенство $\sigma^* = \sigma^{-1}$ является необходимым и достаточным условием ортогональности линейного преобразования.

Квадратная матрица σ с действительными (комплексными) элементами есть *ортогональная* (*унитарная*) тогда и только тогда, когда она удовлетворяет любому из следующих условий:

1. столбцы матрицы σ образуют нормированную и ортогональную систему векторов (ортонормированный базис),
2. $\sigma^*\sigma = E$,
3. $\sigma\sigma^* = E$,
4. $\sigma^* = \sigma^{-1}$.

Множество всех ортогональных (унитарных) преобразований евклидова (унитарного) пространства образует группу относительно умножения преобразований (операция умножения матриц преобразований), так называемую *ортогональную* (*унитарную*) *группу*.

Так как величина определителя не изменяется при транспонировании, то можно записать $(\det\sigma)^2 = 1$. Следовательно, определитель ортогональной матрицы может быть равен только $+1$ и -1.

Определитель ортогональной матрицы $\det\sigma$ равен $+1$ (*собственная* ортогональная матрица) при правой ориентации пространства или -1 (*несобственная* ортогональная матрица) при левой ориентации пространства.

Всякое несобственное ортогональное преобразование на декартовой плоскости есть *отражение* относительно некоторой оси, и его матрица, с определителем равным -1, в подходящем ортонормированном базисе имеет вид:

$$\begin{aligned} x^1 &= X^1 \\ x^2 &= -X^2 \end{aligned} \quad \text{или} \quad \sigma = \begin{Vmatrix} 1 & 0 \\ 0 & -1 \end{Vmatrix}. \tag{2.84}$$

Всякое собственное ортогональное преобразование на декартовой плоскости есть *поворот* системы на угол φ (угол считается положительным при повороте против часовой стрелки) с матрицей:

$$\begin{aligned} x^1 &= X^1\cos\varphi - X^2\sin\varphi \\ x^2 &= X^1\sin\varphi + X^2\cos\varphi \end{aligned} \quad \text{или} \quad \sigma = \begin{Vmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{Vmatrix}, \tag{2.85}$$

определитель которой, естественно, равен 1, а для векторов справедливы условия ортогональности и ортонормированности.

Поворот «вправо» или против часовой стрелки на угол $\varphi = 90^\circ$ даёт кососимметрическую матрицу вида

$$\sigma = \begin{Vmatrix} 0 & -1 \\ 1 & 0 \end{Vmatrix}. \tag{2.86}$$

Вообще в пространстве любой размерности поворот вектора на прямой угол осуществляется умножением его на произвольную кососимметрическую матрицу. Однако такой поворот будет ортогональным преобразованием только в том случае, если определитель этой кососимметрической матрицы равен $+1$ и -1. В противном случае преобразование не будет ортогональным, т.е. не даст нового ортонормированного базиса.

Так, например, матрица ортогонального преобразования в нечётномерном пространстве не может быть кососимметрической. Определитель кососимметрической матрицы нечётного порядка всегда равен нулю (2.76) и не является ортогональным преобразованием.

В трехмерной декартовой системе всякое *собственное* ортогональное преобразование есть *поворот* вокруг некоторой оси, а всякое *несобственное* ортогональное преобразование есть *произведение поворота* вокруг оси и *отражения* в перпендикулярной плоскости. Матрицы собственного и несобственного ортогональных преобразований в подходящем ортонормированном базисе имеют вид:

$$\begin{Vmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & \pm 1 \end{Vmatrix} \quad \text{или} \quad \begin{Vmatrix} \pm 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{Vmatrix}, \tag{2.87}$$

где $+1$ соответствует собственному ортогональному преобразованию, а -1 соответствует несобственному ортогональному преобразованию.

Если ось произвольного поворота не совпадает ни с одной из координатных осей, то этот поворот можно представить как три последовательных поворота заданных координатных плоскостей вокруг заданных координатных осей, совершаемых в определенном порядке.

Последовательное выполнение преобразований определяется произведением соответствующих матриц.

Три последовательных угла соответствующих поворотов, например φ, θ и ψ, называются *углы Эйлера*. Тогда матрица произвольного поворота есть произведение трех матриц собственного ортогонального преобразования вида (2.87), т.е. $\chi = \alpha\beta\gamma$.

Аналогично тому, как всякое вращение в трёхмерном пространстве можно разложить на три вращения в координатных плоскостях, вращение в четырёхмерном пространстве можно разложить на шесть вращений в соответствующих координатных плоскостях, вращение в пятимерном пространстве — на десять вращений и т.д.

В произвольном n-мерном евклидовом пространстве любые ортогональные преобразования сводятся к поворотам и отражениям.

Теперь мы готовы обсудить главное качество матриц. В отличие от поворотов и отражений, которые сохраняют длину векторов при ортогональном преобразовании, преобразования линейного оператора над *собственным* вектором приводят к удлинению и сжатию этого вектора.

Запишем систему уравнений так называемой задачи на собственные зна-

чения и собственные векторы:

$$\alpha\,\xi = \lambda\,\xi, \tag{2.88}$$

где λ — число,

ξ — неизвестный вектор.

Мы видим, что по условию задачи в результате преобразований матрицы α над вектором ξ должно произойти удлинение или сжатие вектора на скалярную величину λ с сохранением его направления, т.е. векторы ξ и $\alpha\,\xi$ коллинеарные. Если вектор ξ таков, что преобразования, которые выполняет над ним линейный оператор α, сводятся к удлинению или сжатию, то он называется *собственным вектором*, его направление — *собственным направлением* и коэффициент его удлинения или сжатия — *собственным значением* матрицы α.

Чтобы вычислить собственные значения λ составим определитель следующего вида:

$$\begin{vmatrix} \alpha_{11}-\lambda & \dots & \alpha_{1n} \\ \cdots & \cdots & \cdots \\ \alpha_{n1} & \dots & \alpha_{nn}-\lambda \end{vmatrix} = 0 \text{ или } |\alpha - \lambda E| = 0. \tag{2.89}$$

Это определитель равен следующему выражению:

$$\Delta(\lambda) = \lambda^n + \sum_{i=1}^{n-1}(-1)^{n-i}a_i\lambda^{n-i} + a_n = 0 \tag{2.90}$$

которое называется *характеристическим уравнением* матрицы α.

Характеристическое уравнение является многочленом порядка n и имеет n корней $\lambda_1, \dots, \lambda_n$, вообще говоря, комплексных. Теория комплексных чисел появилась, отчасти, для отыскания этих корней. В случае системы из двух линейных уравнений с двумя неизвестными мы получаем квадратное характеристическое уравнение (1.39), корни которого (1.40) являются собственными значениями системы.

Кратность корня λ характеристического уравнения (т.е. количество одинаковых корней) называется алгебраической *кратностью* (или просто кратностью). Если кратность равна 1, то собственное значение называется простым. Совокупность всех собственных значений $\lambda_1, \dots, \lambda_n$ называется *спектром* матрицы.

Коэффициент a_1 в (2.90) не что иное, как сумма элементов главной диагонали, которую называют *следом* ($\mathrm{tr}\alpha$ или $\mathrm{Sp}\alpha$) (нем. Spur, англ. trase — след) матрицы α:

$$a_1 = \sum_{i=1}^{n}\alpha_{ii} = \mathrm{Sp}\alpha. \tag{2.91}$$

Коэффициент a_n равен определителю матрицы α:

$$a_n = \det\|\alpha_{ij}\|. \tag{2.92}$$

По теореме Виета след матрицы равен сумме, а определитель — произведению корней характеристического уравнения:

$$a_1 = \alpha_{11} + \ldots + \alpha_{nn} = \operatorname{Sp} \|\alpha_{ij}\| = \lambda_1 + \ldots + \lambda_n, \tag{2.93}$$

$$a_n = \det \|\alpha_{ij}\| = \lambda_1 \cdot \ldots \cdot \lambda_n. \tag{2.94}$$

Свойства следа матрицы:

1. $\operatorname{Sp} \alpha' = \operatorname{Sp}(\beta^{-1}\alpha\beta) = \operatorname{Sp} \alpha$ — важнейшим свойством следа матрицы, как и определителя матрицы, является его независимость от выбора системы координат, базиса и собственных векторов;
2. $\operatorname{Sp}(\alpha\beta) = \operatorname{Sp}(\beta\alpha)$, $\operatorname{Sp}(\alpha\beta\gamma) = \operatorname{Sp}(\gamma\alpha\beta) = \operatorname{Sp}(\beta\gamma\alpha)$, $\operatorname{Sp}(\alpha\beta - \beta\alpha) = 0$;
3. $\operatorname{Sp}(\alpha + \beta) = \operatorname{Sp} \alpha + \operatorname{Sp} \beta$, $\operatorname{Sp}(c\alpha) = c\operatorname{Sp} \alpha$ — линейность.

Для комплексно сопряжённого оператора: $\operatorname{Sp} \overline{\alpha} = \overline{\operatorname{Sp} \alpha}$.

Если у характеристического уравнения имеются комплексные корни, то коэффициенты a_1 и a_n все равно будут действительными, так как и произведение и сумма комплексно сопряженных чисел всегда дают действительное число.

Теорема о связи определителя и следа. Пусть $\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$ — линейный оператор, $\varepsilon \in \mathbb{R}$, тогда при $\varepsilon \to 0$:

$$\det(E + \varepsilon\alpha) = 1 + \varepsilon \operatorname{Sp} \alpha + O(\varepsilon^2), \tag{2.95}$$

Доказательство. Определитель оператора $E + \varepsilon\alpha$ равен произведению собственных чисел. Собственные числа с учётом кратностей равны $1 + \varepsilon\lambda_i$, тогда

$$\det(E + \varepsilon\alpha) = (1 + \varepsilon\lambda_1) \cdot \ldots \cdot (1 + \varepsilon\lambda_n) = 1 + \varepsilon \sum_{i=1}^{n} \lambda_i + O(\varepsilon^2),$$

что и требовалось доказать.

Теорема. Для любого линейного оператора $\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$ и $\mathbb{C}$-линейного оператора $\alpha\colon \mathbb{C}^n \to \mathbb{C}^n$ справедливо выражение

$$\det e^{\alpha} = e^{\operatorname{Sp} \alpha} \tag{2.96}$$

Доказательство для линейного оператора $\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$:

$$\det e^{\alpha} \overset{(2.54)}{=} \det \left(\lim_{m\to\infty} \left(E + \frac{\alpha}{m} \right)^m \right) = \lim_{m\to\infty} \left(\det \left(E + \frac{\alpha}{m} \right)^m \right),$$

ибо определитель матрицы — многочлен и, следовательно, непрерывная функция от элементов. Далее

$$\det \left(E + \frac{\alpha}{m} \right)^m \overset{(2.95)}{=} \left(\det \left(E + \frac{\alpha}{m} \right) \right)^m = \left(1 + \frac{\operatorname{Sp} \alpha}{m} + O\left(\frac{1}{m^2} \right) \right)^m.$$

Остаётся заметить, что $\lim\limits_{m\to\infty}\left(1+\frac{\operatorname{Sp}\alpha}{m}+O\left(\frac{1}{m^2}\right)\right)^m = e^a$ для любого $a\in\mathbb{R}$, в частности для $a=\operatorname{Sp}\alpha$.

Следствие 1. Оператор e^α невырожден.

Следствие 2. Оператор e^α сохраняет ориентацию $\mathbb{R}^n$ (т.е. $\det e^\alpha > 0$).

Вернёмся, однако к собственным значениям и векторам.

Свойства собственных значений и собственных векторов:

1. Если две матрицы коммутируют, то у них одинаковые собственные векторы, и наоборот.

2. Подобные матрицы (2.49) имеют одинаковые собственные значения, а их собственные векторы ξ и ξ' связаны соотношениями:
$$\xi' = \gamma^{-1}\xi \quad \text{или} \quad \xi = \gamma\xi'. \tag{2.97}$$

3. Собственные значения сопряжённых вещественных матриц совпадают. Собственные значения сопряжённых матриц над полем комплексных чисел комплексно сопряжены.

4. Матрица невырождена тогда и только тогда, когда все её собственные значения отличны от нуля.

5. Все собственные значения эрмитовой и вещественной симметрической матриц вещественны.

6. Все собственные значения эрмитовой и вещественной симметрической матриц различны.

7. Для эрмитовых и вещественных симметрических матриц всегда можно вычислить собственные векторы.

8. Все собственные векторы эрмитовой и вещественной симметрической матриц ортогональны, и из них можно составить базис.

9. Все собственные значения ортогональных и унитарных матриц по модулю равны единице; собственные подпространства, отвечающие собственным значениям, ортогональны.

10. Определитель кососимметрической матрицы α нечётного $2n+1$ порядка равен нулю (2.76) и поэтому имеет, по крайней мере, один собственный вектор $\xi\neq 0$ с собственным значением 0.

    Например, в трёхмерном пространстве матрица вида
$$\omega = \begin{Vmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{Vmatrix}. \tag{2.98}$$
    имеет ненулевой собственный вектор $\xi=(\omega_x,\omega_y,\omega_z)$ с нулевым собственным значением и даёт поворот на угол 90°.

Докажем утверждение 5 — все собственные значения эрмитовой матрицы вещественны. Эрмитова матрица характеризуется равенством: $\alpha^+ = \alpha$. Пусть ξ собственный вектор, а λ — соответствующее собственное значение. Тогда: $\alpha\xi = \lambda\xi$. Умножим обе части уравнения на ξ^+ (эрмитово-сопряжённый собственный вектор): $\xi^+\alpha\xi = \lambda\xi^+\xi$. С учётом свойств сопряжённых матриц (с. 74), получаем:

$$\xi^+\alpha^+\xi^{++} = \lambda^*\xi^+\xi^{++} \text{ или } \xi^+\alpha\xi = \lambda^*\xi^+\xi, \text{ так как } \xi^{++} = \xi.$$

Тогда получаем: $\lambda\xi^+\xi = \lambda^*\xi^+\xi$ или $\lambda = \lambda^*$.

Если число собственных векторов квадратной матрицы равно её порядку, и эти векторы линейно независимы, то из них можно составить базис. Однако базис собственных векторов не обязательно существует даже над полем комплексных чисел. Любое линейное преобразование α на полем комплексных чисел имеет хотя бы один собственный вектор, однако линейное преобразование действительного линейного пространства может не иметь ни одного собственного вектора. Например, таков поворот евклидовой плоскости на любой угол, не кратный π.

Так как согласно 5 и 6 все собственные значения *эрмитовых и вещественных симметрических матриц* вещественны и различны, а согласно 7 и 8 для них всегда можно вычислить собственные векторы и составить из них базис (все векторы линейно независимы, т.е. ортогональны), то поэтому симметрические матрицы имеют особое прикладное значение.

Теорема. Матрица линейного преобразования имеет диагональный вид в данном базисе тогда и только тогда, когда базис линейного пространства состоит из собственных векторов (по определению линейно независимых); при этом по диагонали матрицы стоят собственные значения (которые должны быть различны).

Доказательство. Преобразуем систему уравнений $\theta = \alpha\xi$ при помощи σ. Запишем $\theta = \sigma\Theta$, $\xi = \sigma\Xi$. Тогда $\sigma\Theta = \alpha\sigma\Xi$, откуда получаем:

$$\Theta = \sigma^{-1}\alpha\sigma\Xi = \sigma^*\alpha\sigma\Xi. \tag{2.99}$$

Примем n собственных векторов $\xi_1, \ldots, \xi_n$ эрмитовой или вещественной симметрической матрицы α за базис и составим из них матрицу $\sigma = (\xi_1, \ldots, \xi_n)$, а из собственных значений $\lambda_1, \ldots \lambda_n$ — диагональную матрицу λ. Эти матрицы связаны между собой как $\alpha\sigma = \sigma\lambda$. Тогда

$$\sigma^{-1}\alpha\sigma = \begin{Vmatrix} \lambda_1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & \lambda_n \end{Vmatrix}. \tag{2.100}$$

Таким образом, матрица ортогонального (унитарного) преобразования, составленная из собственных векторов, задает *собственное ортогональное* (*унитарное*) *преобразование* и преобразует исходную матрицу линейного преобразования α к диагональному виду.

Множество собственных ортогональных преобразований евклидова пространства образует нормальную подгруппу ортогональной группы преобразований, называемую *специальной ортогональной группой*.

Преобразуем систему координат с помощью матрицы σ. Запишем

$$\sigma^{-1}\alpha\sigma \overset{(2.100)}{=} \lambda E \quad \text{или} \quad |\sigma^{-1}\alpha\sigma - \lambda E| = 0. \tag{2.101}$$

Мы получили, что характеристическое уравнение остается инвариантным при всех преобразованиях системы координат.

Слово «характеристический» в математике всегда означает «связанный инвариантно». В данном случае характеристическое уравнение — уравнение, не зависимое от выбора базиса.

Все собственные значения эрмитовой и симметрической матриц вещественны. Это означает, что любая квадратичная (эрмитова) форма с помощью невырожденного ортогонального (унитарного) преобразования координат или может быть приведена к *диагональному виду*:

$$(\nu, \alpha\nu) = \sum_{i=1}^{n} \lambda_i \xi_i^2, \quad \langle \nu, \alpha\nu \rangle = \sum_{i=1}^{n} \lambda_i |\xi_i|^2. \tag{2.102}$$

где λ_i — диагональные компоненты матрицы, есть множество действительных чисел (собственных значений матрицы α).

Таким образом, нет такой квадратичной формы, которую нельзя привести к главным осям, т.е. к простому выражению указанного выше вида.

Значение величины квадратичной формы при ортогональном преобразовании координат не изменяется, т.е. квадратичная форма является инвариантом или геометрическим объектом в евклидовом пространстве.

Ранг матрицы квадратичной формы, называемый рангом квадратичной формы, равен числу отличных от нуля коэффициентов канонической (принятой за образец) формы (2.102) и не меняется при линейных преобразованиях.

Для квадратичных форм имеет место закон инерции квадратичных форм: число положительных и число отрицательных квадратов в нормальном виде квадратичной формы не зависит от способа приведения квадратичной формы к нормальному виду с помощью невырожденных ортогональных преобразований.

Квадратичная форма является *положительно определенной*, *отрицательно определенной*, *неотрицательной*, *неположительной*, *неопределённой* и *тождественно равной нулю* в том и только в том случае, если действительные собственные значения λ_i матрицы α соответственно все положительные, все отрицательные, все неотрицательные, все неположительные, имеют разные знаки или все равны нулю.

Знакоопределенными называют положительно и отрицательно определенные квадратичные формы. *Неопределенными* называют квадратичные формы, которые принимают как положительные, так и отрицательные значения.

2.6 Многомерные параллелепипеды

Как мы знаем, пространство $\mathbb{R}^n$, наделённое операциями линейной суперпозиции, есть векторное пространство V, произвольный вектор (направленный отрезок) которого есть линейная комбинация его проекций на координатные оси: $\nu = e_1 a^1 + \ldots + e_n a^n$. Внешнее произведение двух векторов: $\nu = e_1 a^1 + \ldots + e_n a^n$ и $\tau = e_1 b^1 + \ldots + e_n b^n$ в $\mathbb{R}^n$, даёт соответствующий параллелограмм $\nu \times \tau$. В отличие от направленного отрезка параллелограмм невозможно проектировать на координатные оси, и поэтому необходимо рассматривать его проекции на координатные плоскости.

Примеры: В двумерном пространстве мы имеем одну проекцию, совпадающую с исходным параллелограммом, на единственную координатную плоскость; в трёхмерном пространстве существуют три проекции на три координатных плоскости; в пространстве большей размерности ситуация сложнее и, в прямом смысле, многограннее. Так в четырёхмерном пространстве координатных плоскостей шесть, в пятимерном — десять, а в общем случае пространства $\mathbb{R}^n$ таких плоскостей C_n^2.

Кососкалярное произведение двух векторов в n-мерном пространстве (не обязательно евклидовом) даёт нам площадь построенного на них параллелограмма, которую, исходя из свойств кокоскалярного произведения, получаем в следующем виде:

$$\begin{aligned}[\nu, \tau] &= \left[(e_1 a^1 + \ldots + e_n a^n), (e_1 b^1 + \ldots + e_n b^n)\right] = \\ &= \sum_{1 \leqslant i < j \leqslant n} [e_i, e_j] \cdot (a^i b^j - a^j b^i).\end{aligned} \tag{2.103}$$

Величины $a^i b^j - a^j b^i$ есть ориентированные площади проекций параллелограмма $\nu \times \tau$ на координатные плоскости i, j параллельно остальным координатным направлениям, а величины $[e_i, e_j]$ есть ориентированные площади параллелограммов, построенных на базисных векторах e_i, e_j.

Обратим внимание ещё раз. В трёхмерном пространстве вектор имеет три проекции на координатные оси, и ориентированный параллелограмм также имеет три проекции на координатные плоскости. В n-мерном пространстве вектор имеет n проекций на координатные оси, однако параллелограмм имеет не n, как интуитивно могло показаться, а C_n^2 проекций на координатные плоскости, так как этих плоскостей C_n^2.

Кососкалярное произведение двух векторов на плоскости есть площадь построенного на этих векторах параллелограмма:

$$[\nu, \tau] = [e_x, e_y] \cdot (a^x b^y - a^y b^x) \overset{(2.37)}{=} [e_x, e_y] \cdot \begin{vmatrix} a^x & a^y \\ b^x & b^y \end{vmatrix}. \tag{2.104}$$

Если ориентированная плоскость евклидова, то $[e_x, e_y] = 1$. Следовательно *определитель*, составленный из двух векторов в E^2, *равен площади построенного на этих векторах параллелограмма.*

В трёхмерном пространстве мы получим три компоненты:

$$[\nu,\tau] = [e_x,e_y]\cdot(a^xb^y - a^yb^x) + [e_y,e_z]\cdot(a^yb^z - a^zb^y) + \\ + [e_z,e_x]\cdot(a^zb^x - a^xb^z). \tag{2.105}$$

В евклидовом пространстве E^3 полученный результат можно представить как трёхмерный вектор υ в исходном базисе e_x, e_y, e_z, перпендикулярный параллелограмму $\nu\times\tau$:

$$\upsilon = \nu\times\tau = e_x(a^yb^z - a^zb^y) + e_y(a^zb^x - a^xb^z) + e_z(a^xb^y - a^yb^x). \tag{2.106}$$

Полученное выражение (2.106) $\upsilon = \nu\times\tau$ называется *векторным произведением* в трёхмерном евклидовом пространстве и обозначает не параллелограмм, а перпендикулярный ему вектор.

Векторное произведение $\upsilon = \nu\times\tau$ можно также понимать как преобразование $\upsilon = \omega\tau$ вектора τ в вектор υ с поворотом на 90° и растяжением, которое задаётся матрицей вида (2.98):

$$\omega = \begin{Vmatrix} 0 & -a^z & a^y \\ a^z & 0 & -a^x \\ -a^y & a^x & 0 \end{Vmatrix}. \tag{2.107}$$

Вектор $\nu = (a^x, a^y, a^z)$ есть ненулевой собственный вектор матрицы ω с нулевым собственным значением.

Обратим внимание ещё раз. В отличие от внешнего произведения, векторное произведение, это частный случай, справедливый только для пространства E^3. В пространствах большей размерности векторное произведение двух векторов невозможно, так как количество проекций параллелограмма будет равно $C_n^2 \neq n$. Однако для E^3 этот частный случай очень важен, так как широко применяется в приложениях.

Векторное произведение базисных векторов имеет вид

$$e_x = e_y\times e_z, \quad e_y = e_z\times e_x, \quad e_z = e_x\times e_y. \tag{2.108}$$

Направления векторов e_x, e_y, e_z, как и направления векторов ν, τ, υ, должны правильно ориентировать пространство (рис. 2.3). Тройка векторов называется *правой*, если при построении трех указанных векторов из одной точки их концы образуют треугольник, обход которого в направлении $e_x \to e_y \to e_z$ или $\nu \to \tau \to \upsilon$ должен происходить против часовой стрелки (в положительном направлении). Если обход происходит по часовой стрелке, тогда тройка векторов называется *левой*.

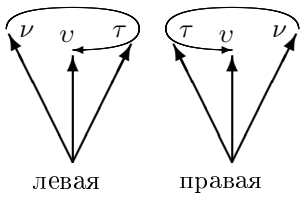


Рис. 2.3: Ориентация пространства

Правая (левая) тройка векторов располагается так, как могут быть расположены соответственно большой, указательный и средний пальцы правой

(левой) руки. Тройки одного и того же класса могут быть приведены в совпадение *вращением*. Тройки разных классов переходят друг в друга путем *зеркального отражения*. Таким образом, необходимость выбора одной из троек связана с проблемой зеркальной ориентации пространства и не может быть устранена вращением.

Не существует инвариантного геометрического определения этих двух классов троек: чтобы определить, например, правые тройки, необходимо конкретно указать какую-либо тройку (ссылкой на пальцы человеческой руки, на буравчик определенной нарезки и т.п.). Очевидно, что все имеющие геометрический и физический смысл соотношения не могут зависеть от того, какой из классов троек мы условимся называть правым. Это утверждение принято формулировать так: все соотношения должны быть *зеркально инвариантны*. Именно ориентация векторов при векторном произведении привела к понятию правой и левой ориентации пространства. Современное название «правая» и «левая» системы предложены Дж. Максвеллом (1873).

Положительно ориентированная правая тройка векторов выбирается в качестве стандартной ориентации пространства.

Так как направление вектора $\upsilon = \nu \times \tau$ зависит от выбранной ориентации пространства, то их называют *аксиальными* или *псевдовекторами* в отличие от прочих векторов (*истинных* или *полярных*), направление которых не зависит от выбора троек или координат.

При зеркальном отражении системы координат все компоненты полярного вектора изменят знак, а компоненты аксиального вектора останутся неизменными. Во всём остальном аксиальные векторы ведут себя как истинные (полярные) векторы.

Обратим внимание. Любая кососимметрическая матрица третьего порядка (2.107) состоит из компонентов аксиального вектора.

Скалярное произведение псевдовектора на вектор даёт *псевдоскаляр*. В то время, как истинный скаляр вполне инвариантен относительно любых преобразований координат, псевдоскаляр изменяет знак при зеркальном отражении пространства.

Свойства векторного произведения:

1. $c(\nu \times \tau) = c\nu \times \tau = \nu \times c\tau, \quad c \in \mathbb{R}$;
2. $\nu \times \tau = -\tau \times \nu, \qquad \nu \times \nu = 0$;
3. $\nu \times (\tau + \upsilon) = \nu \times \tau + \nu \times \upsilon$;
4. $\nu \cdot (\tau \times \upsilon) = \tau \cdot (\upsilon \times \nu) = \upsilon \cdot (\nu \times \tau)$, тогда $\nu \cdot (\tau \times \nu) = \tau \cdot (\nu \times \tau) = 0$;
5. $\nu \times (\tau \times \upsilon) = \tau(\nu \cdot \upsilon) - \upsilon(\nu \cdot \tau)$ или $(\nu \times \tau) \times \upsilon = \tau(\nu \cdot \upsilon) - \nu(\tau \cdot \upsilon)$;
6. $(\nu \times \tau) \cdot (\upsilon \times \varsigma) = (\nu \cdot \upsilon)(\tau \cdot \varsigma) - (\nu \cdot \varsigma)(\tau \cdot \upsilon)$;
7. $(\nu \times \tau) \times (\upsilon \times \varsigma) = (\nu \times \upsilon \times \varsigma)\tau - (\tau \times \upsilon \times \varsigma)\nu = (\nu \times \tau \times \varsigma)\upsilon - (\nu \times \tau \times \upsilon)\varsigma$;
8. $|\nu \times \tau|^2 = (\nu \cdot \nu)(\tau \cdot \tau) - (\nu \cdot \tau)^2$.

В ориентированном евклидовом трёхмерном пространстве кососкалярное произведение трёх векторов $[\nu, \tau, \upsilon]$ называется скалярно-векторным $\nu \cdot (\tau \times \upsilon)$ или *смешанным произведением* (ν, τ, υ) и равно объёму параллелепипеда, построенного на этих векторах:

$$[\nu, \tau, \upsilon] = \nu \cdot (\tau \times \upsilon) = (\nu, \tau, \upsilon) = \begin{vmatrix} a^x & a^y & a^z \\ b^x & b^y & b^z \\ c^x & c^y & c^z \end{vmatrix} \quad (\nu, \tau, \upsilon \in E^3). \tag{2.109}$$

Смешанное произведение можно также понимать как обычную билинейную форму (2.73) $(\nu, \tau, \upsilon) = (\alpha\nu, \upsilon)$, где

$$\alpha = \begin{Vmatrix} 0 & b^z & -b^y \\ -b^z & 0 & b^x \\ b^y & -b^x & 0 \end{Vmatrix}. \tag{2.110}$$

Вектор $\tau = (b^x, b^y, b^z)$ есть ненулевой собственный вектор кососимметрической матрицы α с нулевым собственным значением.

В пространстве большей размерности смешанное произведение для вектора задать нельзя, так как размерность параллелограмма $\tau \times \upsilon$ больше размерности вектора ν.

Если определитель векторов смешанного произведения равен нулю, то эти *векторы компланарны* — они лежат в одной плоскости или в параллельных плоскостях. В этом случае мы можем составить на этих трёх векторах (трёх точках) уравнение данной плоскости:

$$\begin{vmatrix} x & y & z & 1 \\ a^x & a^y & a^z & 1 \\ b^x & b^y & b^z & 1 \\ c^x & c^y & c^z & 1 \end{vmatrix} = 0. \tag{2.111}$$

Рассмотрим k векторов $\xi_1, \ldots, \xi_k$, заданных в n-мерном пространстве (не обязательно евклидовом): $\xi_i = e_1\xi_i^1 + \ldots + e_n\xi_i^n$ $(i = 1, \ldots, k)$.

Внешнее произведение k векторов $\xi_1, \ldots, \xi_k$ в n-мерном пространстве даёт построенный на этих векторах k-мерный параллелепипед $\xi_1 \times \cdots \times \xi_k$. Ориентированный объём k-мерного параллелепипеда, построенного на векторах $\xi_1, \ldots, \xi_k$ в n-мерном пространстве, можно записать как сумму произведений двух видов величин:

$$[\xi_1, \ldots, \xi_k] = \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} [e_{i_1}, \ldots, e_{i_k}] \cdot b^{i_1 \ldots i_k}. \tag{2.112}$$

Величина $b^{i_1 \ldots i_k} = \det \xi^{i_1 \ldots i_k}_{j_1 \ldots j_k}$ есть определитель, в котором суммирование производится по всем перестановкам $j_1, \ldots, j_k$ чисел $1, \ldots, k$, и перед членом берется знак «+», если перестановка чётная, и знак «−», если перестановка нечётная. Каждая такая величина равна площади проекции k-мерного

параллелепипеда на соответствующую k-мерную координатную плоскость $j_1, \ldots, j_k$ параллельно остальным координатным направлениям. Величина $[e_{i_1}, \ldots, e_{i_k}]$ есть ориентированный объём параллелепипеда, построенного на этих базисных векторах. Всего в n-мерном пространстве имеется C_n^k различных k-мерных координатных плоскостей и проекций.

Если число векторов, на которых задан параллелепипед, совпадает с размерностью пространства $k = n$: $\xi_1, \ldots, \xi_n$, то получаем:

$$[\xi_1, \ldots, \xi_n] = [e_1, \ldots, e_n] \cdot \sum \pm \xi^1_{j_1} \cdot \ldots \cdot \xi^n_{j_n}. \tag{2.113}$$

Полученная в выражении сумма $\sum \pm \xi^1_{j_1} \cdot \ldots \cdot \xi^n_{j_n}$ есть *определитель* $\det \xi$ квадратной $(n \times n)$-матрицы ξ коэффициентов векторов. Если мы рассмотрим параллелепипед, построенный на любых k векторах из этих n векторов, то его проекции $b^{i_1 \ldots i_k}$ на k-мерные координатные плоскости будут минорами этого определителя.

Ориентация объёма n-мерного параллелепипеда $[e_1, \ldots, e_n]$ означает, что это кососкалярное произведение должно быть положительным. Таким образом, базисы n-мерного пространства $\xi_1, \ldots, \xi_n$ можно разбить на две группы. Каждая из групп называется *ориентацией* μ векторного пространства V. Первая ориентация обозначается символом $\mu = [\xi_1, \ldots, \xi_n]$, а вторая ориентация — символом $\mu = -[\xi_1, \ldots, \xi_n]$.

Стандартная ориентация евклидова пространства E^n есть $\mu = [e_1, \ldots, e_n] = 1$. Тогда в ориентированном n-мерном евклидовом пространстве, определитель $\det \xi$, составленный из n-векторов $\xi_1, \ldots, \xi_n$, будет равен ориентированному объёму n-мерного параллелепипеда, построенного на этих векторах: $[\xi_1, \ldots, \xi_n] = \det \xi$, если $\xi_1, \ldots, \xi_n \in E^n$. Рёбра параллелепипеда задаются столбцами матрицы.

Определение. *Определитель Грама* $\det G = \Gamma(\xi_1, \ldots, \xi_n)$ евклидова или унитарного пространства равен квадрату объёма n-мерного параллелепипеда, построенного на векторах $\xi_1, \ldots, \xi_n$: $\Gamma(\xi_i) = (\xi_i, \xi_i)$.

Если векторы $\xi_1, \ldots, \xi_n$ линейно независимы, то *определитель Грама* $\det G = \Gamma(\xi_1, \ldots, \xi_n)$ матрицы $G = \|g_{ij}\|^n$, где $g_{ij} = \xi_i \cdot \xi_j$, всегда больше нуля в силу свойства 1 эрмитова скалярного произведения (см. с. 48):

$$\sum_{i,j=1}^{n} g_{ij} a^i \overline{a^j} \overset{(2.67)}{=} |\nu|^2 > 0 \quad \text{при} \quad \nu \neq 0. \tag{2.114}$$

В случае линейно зависимых векторов $\xi_1, \ldots, \xi_n$ найдутся такие не равные одновременно нуля числа $a^1, \ldots, a^n$, что $0 = \xi_1 a^1 + \ldots + \xi_n a^n$, тогда $\sum_{i,j=1}^{n} g_{ij} a^i \overline{a^j} = 0$, и $\det \|g_{ij}\|^n = 0$ равен нулю.

Следовательно,

$$\det \|g_{ij}\|^n \geqslant 0. \tag{2.115}$$

Квадрат объёма параллелепипеда не превышает произведения квадратов длины его сторон, что выражается неравенством:

$$\Gamma(\xi_1, \ldots, \xi_n) \leqslant \Gamma(\xi_1) \cdot \ldots \cdot \Gamma(\xi_n), \text{ где } \Gamma(\xi_i) = (\xi_i, \xi_i). \tag{2.116}$$

Определитель Грама впервые рассматривался И. Грамом (1879).

2.7 Внешние алгебраические формы

Внешние формы есть скалярные функции, заданные не просто на векторах (т.е. в различных точках пространства), а на векторных объектах. Появление внешних форм связано с необходимостью вычислять скалярные кососимметричные функции, заданные на таких векторных объектах, как параллелограммы и многомерные параллелепипеды. Как оказалось, любой определитель матрицы порядка n также есть пример внешней n-формы (полилинейной кососимметрической функции).

Аналогично тому, как проекции вектора ξ на координатные оси есть базисные 1-формы $x^1(\xi), \ldots, x^n(\xi)$ (2.25) — проекции k-мерного параллелепипеда (внешнего произведения k векторов $\xi_1, \ldots, \xi_k$) в n-мерном пространстве на k-мерные координатные плоскости $j_1, \ldots, j_k$ параллельно остальным координатным направлениям можно рассматривать как *базисные k-формы* или как внешнее произведение k базисных 1-форм:

$$x^{j_1} \wedge \ldots \wedge x^{j_k}(\xi_1, \ldots, \xi_k) = b^{i_1 \ldots i_k} \quad (1 \leqslant i_1 < \ldots < i_k \leqslant n). \tag{2.117}$$

Значком $\wedge$ принято обозначать внешнее произведение скалярных функций на векторах, в отличие от внешнего произведения векторов, обозначаемых значком $\times$.

Выражение $x^{j_1} \wedge \ldots \wedge x^{j_k}(e_{j_1}, \ldots, e_{j_k}) = \delta_{i_1 j_1} \cdot \ldots \cdot \delta_{i_k j_k}$, аналогично выражению (2.26) $x^i(e_j) = \delta_{ij}$ для базисных 1-форм. Легко получить, что выражения $x^{j_1} \wedge \ldots \wedge x^{j_k}$ линейно независимы и, следовательно, образуют базис размерностью C_n^k.

Если число векторов $\xi_1, \ldots, \xi_n$, на которых задан параллелепипед, совпадает с размерностью пространства, то $x^1 \wedge \ldots \wedge x^n = \det \xi$. Тогда в ориентированном n-мерном евклидовом пространстве базисная n-форма $x^1 \wedge \ldots \wedge x^n$ равна ориентированному объёму n-мерного параллелепипеда, построенного на векторах $\xi_1, \ldots, \xi_n$:

$$x^1 \wedge \ldots \wedge x^n = [\xi_1, \ldots, \xi_n] = \det \xi, \text{ если } \xi_1, \ldots, \xi_n \in E^n. \tag{2.118}$$

Внешней формой степени k, или k-формой, называется скалярная функция $\omega^k \colon V^k \to \mathbb{R}$ от k векторов $\xi_1, \ldots, \xi_k \in V$, которая k-линейна и кососимметрична ($c_1, c_2 \in \mathbb{R}$):

$$\begin{aligned} &\omega^k(c_1\xi_1' + c_2\xi_1'', \xi_2, \ldots, \xi_k) = c_1\omega^k(\xi_1', \xi_2, \ldots, \xi_k) + \\ &+ c_2\omega^k(\xi_1'', \xi_2, \ldots, \xi_k), \\ &\omega(\xi_{i_1}, \ldots, \xi_{i_k}) = \pm\omega(\xi_1, \ldots, \xi_k). \end{aligned} \tag{2.119}$$

Знак «+» берется, если перестановка $i_1, \ldots, i_k$ чётная, и знак «−», если перестановка $i_1, \ldots, i_k$ нечётная.

Множество всех k-форм становится вещественным линейным пространством $\Lambda^k(V)$, если ввести операцию сложения

$$(\omega_1+\omega_2)^k(\xi_1,\ldots,\xi_k)=\omega_1^k(\xi_1,\ldots,\xi_k)+\omega_2^k(\xi_1,\ldots,\xi_k) \tag{2.120}$$

и умножения на число $c\in\mathbb{R}$

$$(c\omega)^k(\xi_1,\ldots,\xi_k)=c\omega^k(\xi_1,\ldots,\xi_k). \tag{2.121}$$

Любая внешняя k-форма однозначно представляется в виде:

$$\omega^k(\xi_1,\ldots,\xi_k)=\sum_{1\leqslant i_1<\ldots i_k\leqslant n}\omega^k(e_{i_1},\ldots,e_{i_k})\cdot x^{i_1}\wedge\ldots\wedge x^{i_k}(\xi_1,\ldots,\xi_k). \tag{2.122}$$

Линейное (векторное) пространство $\Lambda^k(V)$ задано на базисе из базисных k-форм и имеет размерность C_n^k.

В частности, при $k=n$ пространства $\Lambda^k(V)$ равна $C_n^n=1$.

Тогда всякая n-форма в $\mathbb{R}^n$ есть либо ориентированный объём параллелепипеда, при некотором выборе единицы объёма, либо нуль:

$$\omega^n(\xi_1,\ldots,\xi_n)=\omega^n(e_1,\ldots,e_n)\cdot x^1\wedge\ldots\wedge x^n. \tag{2.123}$$

Если мы не знаем разложения векторов $\xi_1,\ldots,\xi_k$ по базису $e_1,\ldots,e_n$, т.е. не знаем $x^{i_1}\wedge\ldots\wedge x^{i_k}$, но знаем разложение $\omega^k(\xi_1,\ldots,\xi_k)$ по векторам $\xi_1,\ldots,\xi_k$, то получаем:

$$\omega^k(\xi_1,\ldots,\xi_k)=\begin{vmatrix}\omega_1(\xi_1) & \ldots & \omega_k(\xi_1)\\ \cdots & \cdots & \cdots\\ \omega_1(\xi_k) & \ldots & \omega_k(\xi_k)\end{vmatrix}=\omega_1\wedge\ldots\wedge\omega_k(\xi_1,\ldots,\xi_k). \tag{2.124}$$

Величины

$$\omega_j(\xi_i)=\omega_j(e_1)x^1+\ldots+\omega_j(e_1)x^n=a_{1j}x^1+\ldots+a_{nj}x^n\ (j=1,\ldots,k)$$

есть 1-формы, заданные на соответствующем векторе, а величина

$$\omega_1\wedge\ldots\wedge\omega_k(\xi_1,\ldots,\xi_k)$$

называется *внешнее произведение внешних 1-форм.*

Значение внешнего произведения 1-форм на параллелепипеде $\xi_1\times\ldots\times\xi_k$ равно ориентированному объёму образа параллелепипеда в ориентированном евклидовом координатном пространстве $\mathbb{R}^k$ с координатами $\omega_1,\ldots,\omega_k$ при отображении $\xi\mapsto(\omega_1(\xi),\ldots,\omega_k(\xi))$.

Числа $\omega_j(e_i)=a_{ij}$ $(i=1,\ldots,n;j=1,\ldots,k)$ есть компоненты 1-форм, которые они относят базисным векторам $e_1,\ldots,e_n$.

Аналогично, величина

$$\omega^k(e_{i_1},\ldots,e_{i_k})=a_{i_1,\ldots,i_k}=\sum\pm a_{i_1j_1}\cdot\ldots\cdot a_{i_kj_k}=\det\|a_{ij}\|^k \tag{2.125}$$

есть соответствующая компонента внешней k-формы.

Суммирование производится по всем перестановкам $j_1, \ldots, j_k$ чисел $1, \ldots, k$, и перед членом берется знак «+», если перестановка чётная, и знак «−», если перестановка нечётная.

В результате для внешней k-формы можно записать

$$\omega^k(\xi_1, \ldots, \xi_k) = \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} a_{i_1 \ldots i_k} \cdot x^{i_1} \wedge \ldots \wedge x^{i_k}. \tag{2.126}$$

В частности, при $k = n$:

$$\omega^n(\xi_1, \ldots, \xi_n) = \det \omega \cdot x^1 \wedge \ldots \wedge x^n. \tag{2.127}$$

Если мы знаем разложение векторов $\xi_1, \ldots, \xi_k$ по базису $e_1, \ldots, e_n$: $\xi_i = e_1\xi_i^1 + \ldots + e_n\xi_i^n$, но не знаем разложение $\omega^k(\xi_1, \ldots, \xi_k)$ по векторам $\xi_1, \ldots, \xi_k$, то получаем:

$$\omega^k(\xi_1, \ldots, \xi_k) = \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \omega^k(e_{i_1}, \ldots, e_{i_k}) \cdot \det \xi_{j_1 \ldots j_k}^{i_1 \ldots i_k}. \tag{2.128}$$

В частности, для векторов $\xi = (a^1, \ldots, a^n)$ и $\eta = (b^1, \ldots, b^n)$ имеем:

$$\omega^2(\xi, \eta) = \sum_{i<j} \omega^2(e_i, e_j) \cdot (a^i b^j - a^j b^i). \tag{2.129}$$

В ориентированном евклидовом пространстве 2-форма билинейна, кососимметрична и, как всякая билинейная форма, может быть записана в виде евклидовой структуры

$$\omega^2(\xi, \eta) \overset{\substack{(2.73)\\(2.77)}}{=} (\alpha\xi, \eta), \tag{2.130}$$

где α — кососимметрическая матрица.

Соответственно, для векторов $\xi_1, \ldots, \xi_n$ имеем:

$$\omega^n(\xi_1, \ldots, \xi_n) = \omega^n(e_1, \ldots, e_n) \cdot \det \xi. \tag{2.131}$$

В ориентированном евклидовом пространстве E^n получаем

$$\omega^k = \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \det \|a_{ij}\|^k \det \xi_{j_1 \ldots j_k}^{i_1 \ldots i_k} \quad \text{и} \quad \omega^n = \det \omega \cdot \det \xi. \tag{2.132}$$

Внешним произведением $\omega^k \wedge \omega^l$ *k-формы* ω^k *в* $\mathbb{R}^n$ *на l-форму* ω^l *в* $\mathbb{R}^n$ называется $(k+l)$-форма в $\mathbb{R}^n$, значение которой на $k+l$ векторах $\xi_1, \ldots, \xi_k, \xi_{k+1}, \ldots, \xi_{k+l} \in \mathbb{R}^n$ равно:

$$\omega^{k+l}(\xi_1, \ldots, \xi_{k+l}) = \omega^k \wedge \omega^l = \sum \pm \omega^k(\xi_{i_1}, \ldots, \xi_{i_k}) \omega^l(\xi_{j_1}, \ldots, \xi_{j_l}). \tag{2.133}$$

Знак «+» берется, если перестановка $i_1 < \ldots < i_k, j_1 < \ldots < j_l$ номеров $1, 2, \ldots, k+l$ чётная, и знак «−», если перестановка нечётная.

Иными словами, каждое разбиение $k+l$ векторов $\xi_1, \ldots, \xi_{k+l}$ на две группы (из k и из l векторов) дает одно слагаемое в сумме. Это слагаемое равно произведению значения k-формы ω^k на k векторах первой группы на значение l-формы ω^l на l векторах второй группы со знаком «+» или «−» в зависимости от того, как упорядочены векторы в группах.

Например, если $k+l=2$, то разбиений всего два: ξ_1, ξ_2 и ξ_2, ξ_1. Поэтому $(\omega_1 \wedge \omega_2)(\xi_1, \xi_2) = \omega_1(\xi_1)\omega_2(\xi_2) - \omega_2(\xi_1)\omega_1(\xi_2)$.

Внешнее произведение внешних форм:

1. косокоммутативно: $\omega^k \wedge \omega^l = (-1)^{kl}\omega^l \wedge \omega^k$,
2. дистрибутивно: $(\lambda_1\omega_1^k + \lambda_2\omega_2^k) \wedge \omega^l = \lambda_1\omega_1^k \wedge \omega^l + \lambda_2\omega_2^k \wedge \omega^l$,
3. ассоциативно: $(\omega^k \wedge \omega^l) \wedge \omega^m = \omega^k \wedge (\omega^l \wedge \omega^m)$.

Пример. В трёхмерном евклидовом пространстве каждому вектору $\nu = (a^x, a^y, a^z) \in E^3$ можно сопоставить две формы ($\xi, \eta \in E^3$):

1. $\omega_\nu^1(\xi) = (\nu, \xi) = a^x x + a^y y + a^z z$ — скалярное произведение,
2. $\omega_\nu^2(\xi, \eta) = (\nu, \xi, \eta) = a^x y \wedge z + a^y z \wedge x + a^z x \wedge y$ — смешанное произведение.

Тогда внешнее произведение двух 1-форм в E^3:

$$\omega_\nu^1(\xi) = a^x x + a^y y + a^z z, \quad \omega_\tau^1(\eta) = b^x x + b^y y + b^z z,$$

превращается в произведение векторных произведений:

$$\begin{aligned} \omega_\nu^1 \wedge \omega_\tau^1 &= (a^x b^y - a^y b^x)\, x \wedge y + \\ &+ (a^y b^z - a^z b^y)\, y \wedge z + (a^z b^x - a^x b^z)\, z \wedge x = \\ &= (\nu \times \tau) \cdot (\xi \times \eta) = \omega_{\nu\times\tau}^2. \end{aligned} \tag{2.134}$$

Также внешнее произведение 1-формы и 2-формы в E^3:

$$\omega_\nu^1(\xi) = a^x x + a^y y + a^z z, \quad \omega_\tau^2(\eta, \zeta) = b^x y \wedge z + b^y z \wedge x + b^z x \wedge y,$$

есть произведение скалярного и смешанного произведения векторов:

$$\omega_\nu^1 \wedge \omega_\tau^2 = (a^x b^x + a^y b^y + a^z b^z)\, x \wedge y \wedge z = (\nu \cdot \tau)(\xi, \eta, \zeta). \tag{2.135}$$

Рассмотрим теперь внешний квадрат внешней формы:

$$\omega^k \wedge \omega^k = (-1)^{k^2}\omega^k \wedge \omega^k. \tag{2.136}$$

Очевидно, что если k — нечётно, то её внешний квадрат тождественно равен нулю $\omega^k \wedge \omega^k = -\omega^k \wedge \omega^k = 0$. Однако, если k — четно, то её внешний квадрат не равен тождественно нулю. Отсюда легко понять, что 2-форма ω^2 тождественно равна нулю, если пространство нечётное $\mathbb{R}^{n+1}$, и не равна тождественно нулю, если пространство чётное $\mathbb{R}^{2n}$.

Кососимметрическая 2-форма называется *вырожденной*, если $\omega^2(\xi,\eta)\equiv 0$ при любом η и $\xi\neq 0$. Форма, которая тождественно равна нулю $\omega^2(\xi,\eta)\equiv 0$ при любом η, только если $\xi=0$, называется *невырожденной*. В нечётномерном пространстве кососимметрическая 2-форма всегда вырожденная, а в чётномерном пространстве всегда можно выбрать невырожденную кососимметрическую 2-форму.

Обратим внимание ещё раз. Незначительное, казалось бы, различие в свойствах внешней чётной и нечётной форм — неравенство или тождественное равенство нулю внешнего квадрата, привело к принципиальной разнице в свойствах чётномерного и нечётномерного пространств.

Как мы уже знаем, определитель кососимметрической матрицы нечётного порядка равен нулю $\det\alpha\equiv 0$ (2.76), и имеет, по крайней мере, один собственный вектор $\xi\neq 0$ с собственным значением 0 (свойство 10 собственных векторов, стр. 81), который называется *нулевым вектором* вырожденной формы ω^2 в пространстве $\mathbb{R}^{2n+1}$. Обязательное наличие такого ненулевого вектора отражает существование оси у всякого вращения нечётномерного пространства $\mathbb{R}^{2n+1}$ всех векторов.

Все различные нулевые векторы ω^2 образуют линейное подпространство. Если все нулевые векторы коллинеарны, то размерность этого подпространства минимальная возможная (т.е. 1 — нулевые векторы образуют линию), а форма ω^2 называется *неособой*. В случае чётномерного пространства $\mathbb{R}^{2n}$ форма ω^2 неособая, если нулевые векторы отсутствуют (размерность подпространства нулевых векторов равна 0).

Невырожденная билинейная кососимметрическая 2-форма ω^2, заданная в $\mathbb{R}^{2n}$, называется *симплектическая линейная структура* (от лат. simplex — простой). Соответственно, любое чётномерное пространство $\mathbb{R}^{2n}$ вместе с симплектической линейной структурой называется *симплектическим линейным пространством*.

Рассмотрим теперь примеры симплектических пространств.

Например, комплексное пространство имеет естественную симплектическую структуру. Вектор линейного комплексного пространства $\mathbb{C}^n$ есть n-членная последовательность комплексных чисел $\xi=(z^1,\ldots,z^n)$ или $2n$-членная последовательность $\xi=(x^1,\ldots,x^n,y^1,\ldots,y^n)$ вещественных чисел чётномерного вещественного пространства $\mathbb{R}^{2n}$.

Эрмитово скалярное произведение в унитарном пространстве (эрмитова структура) $\langle\xi,\eta\rangle=a^1\overline{b^1}+\ldots+a^n\overline{b^n}$ состоит из n произведений пар комплексных чисел: $a^j=x^j_\xi+iy^j_\xi$ и $b^j=x^j_\eta+iy^j_\eta$. Тогда запишем:

$$\begin{aligned}\langle\xi,\eta\rangle &= a^1\overline{b^1}+\ldots+a^n\overline{b^n}=\sum_{j=1}^n(x^j_\xi+iy^j_\xi)(x^j_\eta-iy^j_\eta)=\\ &=\sum_{j=1}^n(x^j_\xi x^j_\eta+y^j_\xi y^j_\eta)+i\sum_{j=1}^n(y^j_\xi x^j_\eta-x^j_\xi y^j_\eta)=\\ &=\sum_{j=1}^n(x^j_\xi x^j_\eta+y^j_\xi y^j_\eta)+i\sum_{j=1}^n x^j\wedge y^j(\xi,\eta).\end{aligned}\tag{2.137}$$

Вещественная часть эрмитовой структуры билинейна, симметрична, невырожденна и есть евклидова структура:

$$\Re\langle\xi,\eta\rangle = x^1_\xi x^1_\eta + y^1_\xi y^1_\eta + \ldots + x^n_\xi x^n_\eta + y^n_\xi y^n_\eta = (\xi,\eta). \tag{2.138}$$

Мнимая часть эрмитовой структуры есть кососкалярное произведение или 2-форма, которая билинейна, кососимметрична, невырождена и есть симплектическая структура:

$$\begin{aligned}\Im\langle\xi,\eta\rangle &= (y^1_\xi x^1_\eta - x^1_\xi y^1_\eta) + \ldots + (y^n_\xi x^n_\eta - x^n_\xi y^n_\eta) = \\ &= x^1\wedge y^1 + \ldots + x^n \wedge y^n = [\xi,\eta] = \omega^2(\xi,\eta).\end{aligned} \tag{2.139}$$

Невырожденность кососкалярного произведения означает, что если $[\xi,\eta] = 0$ при любом η, то тогда $\xi = 0$ (или, вектор косоортогональный всему пространству — нулевой). Аналогично, невырожденность скалярного произведения означает, что если $(\xi,\xi) = 0$, то $\xi = 0$.

Таким образом, эрмитова структура в $\mathbb{C}^n$ состоит из вещественной и мнимой частей, которые являются вещественными билинейными формами. Первая из них симметрична, вторая кососимметрична и обе невырождены, первая называется евклидовой структурой, а вторая называется симплектической структурой:

$$\langle\xi,\eta\rangle = (\xi,\eta) + i[\xi,\eta], \quad (\xi,\eta) = (\eta,\xi), \quad [\xi,\eta] = -[\eta,\xi]. \tag{2.140}$$

Мнимая часть комплексного пространства является лишь одним из примеров симплектического пространства.

Симплектическая структура возникает в любом чётномерном пространстве, каждая точка которого есть два n-мерных вектора с координатами $\mathbf{p} = (p_1,\ldots,p_n)$ и $\mathbf{q} = (q_1,\ldots,q_n)$ соответственно, при надлежащем выборе базиса (в подобных обозначениях традиционно используют нижние индексы). Рассмотрим это подробнее.

Если $(e_i,e_j) = 0$, то векторы e_i, e_j *ортогональны*.

Если $[e_i,e_j] = 0$, то векторы e_i, e_j *косоортогональны*.

Евклидова структура в $\mathbb{R}^n$ возникает при выборе базиса $e_1,\ldots,e_n$, для которого справедливы соотношения:

$$(e_i,e_j) = \delta_{ij} \quad (i,j = 1,\ldots,n). \tag{2.141}$$

Симплектическая структура в $\mathbb{R}^{2n}$ возникает при выборе базиса $e_{q_1},\ldots,e_{q_n},e_{p_1},\ldots,e_{p_n}$, для которого справедливы соотношения:

$$[e_{q_i},e_{p_j}] = \delta_{ij}, \quad [e_{q_i},e_{q_j}] = [e_{p_i},e_{p_j}] = 0 \quad (i,j = 1,\ldots,n). \tag{2.142}$$

Иными словами в симплектическом пространстве каждый базисный вектор косоортогонален всем базисным векторам, кроме одного, с ним сопряжённого. Если принять векторы симплектического базиса за координатные

орты, то мы получим симплектическую систему координат $p_1, \ldots, p_n, q_1, \ldots, q_n$, в которой симплектическая структура принимает *стандартный вид*:

$$\begin{aligned}\omega^2(\xi,\eta) = [\xi,\eta] = (q_1^\xi p_1^\eta - p_1^\xi q_1^\eta) + \ldots + (q_n^\xi p_n^\eta - p_n^\xi q_n^\eta) = \\ = p_1 \wedge q_1 + \ldots + p_n \wedge q_n = \mathbf{q}^\xi \mathbf{p}^\eta - \mathbf{p}^\xi \mathbf{q}^\eta = \mathbf{p} \wedge \mathbf{q}.\end{aligned} \tag{2.143}$$

Мы можем задать в $\mathbb{R}^{2n}$ евклидову структуру, если выберем базис $e_{q_1}, \ldots, e_{q_n}, e_{p_1}, \ldots, e_{p_n}$, для которого справедливы соотношения:

$$(e_{q_i}, e_{q_j}) = \delta_{ij}, \quad (e_{q_i}, e_{p_j}) = 0 \quad (i,j = 1, \ldots, n). \tag{2.144}$$

Тогда получаем

$$\begin{aligned}(\xi,\eta) &= p_1^\xi p_1^\eta + q_1^\xi q_1^\eta + \ldots + p_n^\xi p_n^\eta + q_n^\xi q_n^\eta, \\ (\xi,\xi) &= p_1^2 + q_1^2 + \ldots + p_n^2 + q_n^2.\end{aligned} \tag{2.145}$$

Геометрически симплектическая структура ω^2 означает сумму площадей ориентированных проекций параллелограмма на n координатных двумерных плоскостей $(p_1, q_1); \ldots; (p_n, q_n)$.

Рассмотрим внешний квадрат симплектической структуры:

$$\omega^2 \wedge \omega^2 = -2 \sum_{i<j} p_i \wedge p_j \wedge q_i \wedge q_j. \tag{2.146}$$

Внешняя k-я степень ω^2 равна:

$$\underbrace{\omega^2 \wedge \ldots \wedge \omega^2}_{k} = \pm k! \sum_{i<j} p_{i_1} \wedge \ldots \wedge p_{i_k} \wedge q_{i_1} \wedge \ldots \wedge q_{i_k}. \tag{2.147}$$

В частности

$$\underbrace{\omega^2 \wedge \ldots \wedge \omega^2}_{n} = \pm n!\, p_1 \wedge \ldots \wedge p_n \wedge q_1 \wedge \ldots \wedge q_n. \tag{2.148}$$

есть, с точностью до множителя, объём $2n$-мерного параллелепипеда в симплектическом пространстве $\mathbb{R}^{2n}$.

Симплектическая структура, как всякая билинейная форма, может быть записана в виде евклидовой структуры:

$$[\xi,\eta] = \omega^2(\xi,\eta) \overset{\substack{(2.73)\\(2.77)}}{=} (I\xi,\eta), \tag{2.149}$$

где I — кососимметрическая матрица вида

$$I = \begin{Vmatrix} \mathbf{0} & -E \\ E & \mathbf{0} \end{Vmatrix}, \tag{2.150}$$

E — единичная матрица порядка n,

$\mathbf{0}$ — нулевая матрица порядка n.

При $n = 1$ (на плоскости p, q) I есть просто поворот на угол $90°$ «вправо» или против часовой стрелки (2.86), а в общем случае I есть поворот «вправо» на угол $90°$ в каждой из n плоскостей p_i, q_i. Таким образом I — оператор преобразования.

Два последовательных поворота на $90°$ дают полярные векторы, т.е. $I^2 = -E$, где E — единичная матрица порядка $2n$, что также легко проверяется при перемножении матриц I.

Вектор n-мерного линейного комплексного пространства $\upsilon \in \mathbb{C}^n$ задан двумя вещественными векторами $\upsilon = (\nu, \tau)$ и есть элемент чётномерного вещественного пространства $\mathbb{R}^{2n}$ или $2n$-членная последовательность $\upsilon = (x^1, \ldots, x^n, y^1, \ldots, y^n)$. Также комплексный вектор может быть записан как $\upsilon = \nu + i\tau$. Назовём подпространство $\mathbb{R}^n$ векторов вида $\nu + i\mathbf{0}$, $\nu \in \mathbb{R}^n$ *вещественной плоскостью*, а подпространство $\mathbb{R}^n$ векторов вида $\mathbf{0} + i\tau$, $\tau \in \mathbb{R}^n$ *мнимой плоскостью*. Тогда оператор (2.150) $I\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ изоморфно отображает вещественную плоскость в мнимую, а мнимую — в вещественную, при этом $I^2 = -E$.

Обозначим через $\varsigma\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ *оператор комплексного сопряжения*: $\varsigma(\nu + i\tau) = \nu - i\tau$. Действие ς обозначается часто просто чертой сверху. В отличие от оператора I, оператор ς представляет собой не поворот, а отображение мнимой плоскости относительно вещественной плоскости. Оператор ς совпадает с единичным E на вещественной плоскости и с минус единичным $-E$ — на мнимой. Он инволютивен: $\varsigma^2 = E$.

Как все евклидовы пространства одинаковой размерности изоморфны друг другу, так и все симплектические пространства одинаковой размерности изоморфны. Линейное преобразование $S\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ называется *симплектическим*, если оно сохраняет кососкалярное произведение $[S\xi, S\eta] = [\xi, \eta]$, аналогично тому, как линейное преобразование $\sigma\colon V \to V$ называется ортогональным (унитарным) если оно сохраняет скалярное (эрмитово-скалярное) произведение $(\sigma\xi, \sigma\eta) = (\xi\eta)$.

Теорема. Чтобы преобразование было симплектическим, необходимо и достаточно, чтобы его матрица S в симплектической системе координат удовлетворяла соотношению

$$S^* I S = I \tag{2.151}$$

где S^* — матрица, сопряжённая к S.

Доказательство. Условие симплектичности ($[S\xi, S\eta] = [\xi, \eta]$ для всех ξ, η) с помощью оператора I записывается через скалярное произведение в виде $(IS\xi, S\eta) = (I\xi, \eta)$ или еще $(S^* I S\xi, \eta) = (I\xi, \eta)$, что и требовалось доказать. Оператор симплектического преобразования S называется *симплектическим оператором*. Оператор преобразования I также симплектический.

Множество всех симплектических преобразований в $\mathbb{R}^{2n}$ образует *симплектическую группу*, аналогично тому, как множество всех ортогональных (унитарных) преобразований евклидова (унитарного) пространства образует ортогональную (унитарную) группу.

2.8 Тензоры

Все рассмотренные в данной главе геометрические объекты можно обобщить под общим названием тензоры.

Тензор (от лат. tendo, буквально — напрягаю, растягиваю) или поливектор (от греч. $\pi o\lambda\acute{\upsilon}\varsigma$ — многочисленный, обширный) — геометрический объект, который описывается n-мерным массивом (таблицей) чисел, где n — *валентность* или *ранг* тензора.

Так скаляр или инвариант, который в любом базисе имеет одну и ту же компоненту (число, которое можно рассматривать как нульмерный массив с единственным элементом), есть тензор нулевой валентности или нулевого ранга. Вектор и ковектор (1-форма), которые задаются одномерным массивом (строкой или столбцом) есть тензоры первой валентности. Матрица α линейного преобразования $y^i = \alpha^i_j x^j$ есть 1 раз ковариантный и 1 раз контравариантный (смешанный) тензор второй валентности, который отождествляют с линейным преобразованием и иногда называют *аффинором*. Также существует два раза ковариантный тензор, например, матрица Грама, общая компонента которого может быть записана в виде g_{ij} и при свертывании даёт квадратичную $\alpha_{ij}x^i x^j$ или билинейную $\alpha_{ij}x^i y^j$ формы, и два раза контравариантный тензор, общая компонента которого может быть записана в виде: t^{ij}. Существует четыре типа тензоров третьей валентности: t^{ijk}, t^{ij}_k, t^i_{jk}, t_{ijk}. Массивы большей размерности представляют собой тензоры более высокого ранга. В частности, r раз ковариантный и s раз контравариантный тензор t есть зависящая от выбора базиса совокупность чисел $t^{i_1\ldots i_r}_{j_1\ldots j_s}$. При смене базиса компоненты тензора преобразуются по следующей формуле:

$$t^{k_1\ldots k_r}_{m_1\ldots m_s} = A^{j_1}_{m_1}\ldots A^{j_s}_{m_s} B^{k_1}_{i_1}\ldots B^{k_s}_{i_r} t^{i_1\ldots i_r}_{j_1\ldots j_s}. \tag{2.152}$$

При этом сам тензор как «геометрическая сущность» от выбора базиса не зависит. Например вектор является частным случаем тензора: компоненты вектора меняются при смене координатных осей, но сам вектор остаётся неизменным.

Представление тензора в компонентах — классический подход к определению тензора, более распространённый в физической литературе. Термин первоначально относился к тензору деформации, связанному с возникающими упругими деформациями при малых растяжениях.

Тензор в более современном определении — полилинейная функция (2.60) называется m-тензором на V, а множество всех m-тензоров, обозначаемое $\mathfrak{I}^m(V)$, становится вещественным *векторным пространством*, если для каждой пары $S, T \in \mathfrak{I}^m(V)$ и каждого $c \in \mathbb{R}$ имеем:

$$\begin{aligned}(S+T)(\xi_1,\ldots,\xi_m) &= S(\xi_1,\ldots,\xi_m) + T(\xi_1,\ldots,\xi_m),\\ (cS)(\xi_1,\ldots,\xi_m) &= cS(\xi_1,\ldots,\xi_m).\end{aligned} \tag{2.153}$$

В отличие от полилинейной функции, для которой операция линейной суперпозиции определена на множестве V её аргументов ξ, для тензора

операция линейной суперпозиции определена на множестве $\mathfrak{I}^m(V)$ самих тензоров. Поэтому множество тензоров образует векторное пространство. При этом сами тензоры есть скалярные функции $T\colon V^m \to \mathbb{R}$.

Тензорные операции можно считать прямым обобщением матричных операций (умножение матриц между собой и с векторами), а также векторных операций, таких, как скалярное произведение. Эти операции, если исходить из современного (аксиоматического) определения, прямо вытекают из полилинейности тензоров, точно так же, как и матричные операции вытекают из линейности линейных операторов.

Пусть тензор $S \in \mathfrak{I}^m(V)$ и тензор $T \in \mathfrak{I}^k(V)$. Операция $\otimes$ *тензорное произведение* связывает различные пространства $\mathfrak{I}^m(V)$ формулой:

$$S \otimes T(\xi_1, \ldots, \xi_m, \xi_{m+1}, \ldots, \xi_{m+k}) = S(\xi_1, \ldots, \xi_m)\, T(\xi_{m+1}, \ldots, \xi_{m+k}) \quad (2.154)$$

Заметим, что порядок сомножителей здесь существенен, поскольку произведения $S \otimes T$ и $T \otimes S$ отнюдь не равны.

Например, совокупность чисел $t^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ (r раз ковариантный и s раз контравариантный тензор t) образует полилинейную форму от s векторов пространства V и r векторов сопряжённого пространства V^* и является элементом тензорного произведения пространств:

$$\underbrace{V \otimes \ldots \otimes V}_{s} \otimes \underbrace{V^* \otimes \ldots \otimes V^*}_{r} \quad (2.155)$$

где число $n = r + s$ называется валентностью тензора t.

Свойства тензорного произведения:

1. $(S_1 + S_2) \otimes T = S_1 T + S_2 T$,
2. $S \otimes (T_1 + T_2) = S T_1 + S T_2$,
3. $(cS) \otimes T = S \otimes (cT) = c(S \otimes T)$,
4. $(S \otimes T) \otimes U = S \otimes (T \otimes U) = S \otimes T \otimes U$.

Теорема: Пусть $\nu_1, \ldots, \nu_n$ — базис пространства V и $\varphi_1, \ldots \varphi_n$ — дуальный базис сопряжённого пространства V^*, $\varphi_i(\nu_j) = \delta_{ij}$. Тогда множество всех тензорных произведений m-го порядка

$$\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_m}, \quad i \leqslant i_1 < \ldots < i_m \leqslant n \quad (2.156)$$

является базисом тензорного пространства $\mathfrak{I}^m(V)$ (множества тензоров $T(\xi_1, \ldots, \xi_m)$), которое в силу этого имеет размерность n^m.

Доказательство: Заметим, что

$$\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_m}(\nu_{j_1}, \ldots, \nu_{j_m}) = \varphi_{i_1}(\nu_{j_1}), \ldots, \varphi_{i_m}(\nu_{j_m}) =$$
$$= \delta_{i_1 j_1} \ldots \delta_{i_m j_m} = \begin{cases} 1, & \text{если } j_1 = i_1, \ldots, j_m = i_m, \\ 0 & \text{во всех остальных случаях.} \end{cases}$$

Если дано m векторов $\xi_1, \ldots, \xi_m$, где $\xi_i = x_i^1\nu_1 + \ldots + x_i^n\nu_n$, то для любого $T \in \mathfrak{I}^m(V)$ в силу свойств (2.60) полилинейной формы имеем:

$$T(\xi_1, \ldots, \xi_m) = \sum_{i_1,\ldots,i_m=1}^{n} x_1^{i_1} \ldots x_m^{i_m} T(\nu_{i_1}, \ldots, \nu_{i_m}).$$

В силу выражения $\varphi_i(\xi) = x^i$ и свойства тензорного произведения

$$\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_m}(\xi_1, \ldots, \xi_m) = x_1^{i_1} \cdot \ldots \cdot x_m^{i_m}. \tag{2.157}$$

Тогда:

$$T(\xi_1, \ldots, \xi_m) = \sum_{i_1,\ldots,i_m=1}^{n} T(\nu_{i_1}, \ldots, \nu_{i_m})\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_m}(\xi_1, \ldots, \xi_m). \tag{2.158}$$

Так что, $\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_m}$ порождают $\mathfrak{I}^m(V)$.

Предположим теперь, что $c_{j_1}, \ldots, c_{j_m}$ — числа, для которых

$$\sum_{i_1,\ldots,i_m=1}^{n} c_{j_1} \ldots c_{j_m} \varphi_{i_1} \otimes \ldots \otimes \varphi_{i_m}(\xi_1, \ldots, \xi_m) = 0. \tag{2.159}$$

Применение обеих частей этого равенства к $\nu_{i_1}, \ldots, \nu_{i_m}$ дает $c_{j_1} = \ldots = c_{j_m} = 0$. Следовательно, все выражения $\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_m}$ линейно независимы в силу определения линейной независимости векторов. Значит множество произведений $\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_m}(\xi_1, \ldots, \xi_m)$ равно n^m (аналогично общему числу членов многочлена (1.23)) и образует базис $\mathfrak{I}^m(V)$.

На тензоры можно распространить также важную конструкцию, хорошо известную в случае сопряжённых пространств. Именно, всяким линейным отображением $f\colon V \to W$ определяется линейное отображение $f^*\colon \mathfrak{I}^m(W) \to \mathfrak{I}^m(V)$, действующее по формуле

$$f^*T(\xi_1, \ldots, \xi_m) = T(f^*(\xi_1), \ldots, f^*(\xi_m)) \tag{2.160}$$

для любых $T \in \mathfrak{I}^m(V)$ и $\xi_1, \ldots, \xi_m \in V$. Звездочка сверху указывает, что f^* сопряжённое отображение, т.е. действует в сторону, противоположную f. Иными словами, значение функции f^*T на векторах равно значению тензора T на образах этих векторов.

При этом

$$f^*(S \otimes T) = f^*S \otimes f^*T. \tag{2.161}$$

Пусть $f\colon \mathbb{R}^k \to \mathbb{R}^n$ — линейное отображение, и ω^m внешняя m-форма на $\mathbb{R}^n$. Тогда существует *сопряжённое отображение* $f^*\colon \mathfrak{I}^m(\mathbb{R}^n) \to \mathfrak{I}^m(\mathbb{R}^k)$, и на $\mathbb{R}^k$ возникает m-форма $f^*\omega^m$, значение которой на m-векторах $\xi_1, \ldots, \xi_m \in \mathbb{R}^n$ равно значению ω^m на их образах:

$$(f^*\omega^m)(\xi_1, \ldots, \xi_m) = \omega^m(f^*(\xi_1), \ldots, f^*(\xi_m)). \tag{2.162}$$

Например:

$$f^*(\lambda_1\omega_1 + \lambda_2\omega_2) = \lambda_1 f^*(\omega_1) + \lambda_2 f^*(\omega_2),$$
$$f^*(\lambda_1\omega_1 \wedge \lambda_2\omega_2) = \lambda_1 f^*(\omega_1) \wedge \lambda_2 f^*(\omega_2).$$

Форма ω^m является *инвариантом преобразования* $f\colon \mathbb{R}^n \to \mathbb{R}^n$ тогда и только тогда, когда $f^*\omega^m = \omega^m$. Если формы ω^m и ω^n — инварианты преобразования f, то $\omega^m \wedge \omega^n$ — также инвариант f.

Внутреннее произведение в пространстве тензоров определяется как $T = (\,,\,) \in \mathfrak{I}^2(V)$ — 2-тензор на V, который *симметричен*, т.е. $T(\xi,\zeta) = T(\zeta,\xi)$ для всех $\xi,\zeta \in V$, и *положительно определен*, т.е. $T(\xi,\xi) > 0$ для всех $\xi \neq 0$.

Чтобы выделить внутреннее произведение векторов $(\,,\,)$ мы называем его *стандартным внутренним произведением* на $\mathbb{R}^n$. Следующая теорема показывает, что внутреннее произведение тензоров не слишком далее идущее обобщение стандартного.

Теорема: Если T — внутреннее произведение на V, то V обладает базисом $e_1,\ldots,e_n$, для которого $T(e_i,e_j) = \delta_{ij}$. (Такой базис называется ортонормированным относительно T, т.е. евклидовым). Следовательно, существует такой изоморфизм $f\colon \mathbb{R}^n \to V$, что $T(f(\xi),f(\zeta)) = (\xi,\zeta)$ для всех $\xi,\zeta \in \mathbb{R}^n$. Другими словами, $f^*(T) = (\,,\,)$.

Внешние формы есть кососимметрические тензоры. Пространство $\Lambda^k(v)$ всех кососимметрических k-тензоров является, очевидно, подпространством в $\mathfrak{I}^k(V)$.

Тензоры — естественное обобщение тех понятий, с которыми мы уже имели дело. Их алгебра устроена просто, и они имеют множество разнообразных применений. Главная проблема для тех, кто впервые сталкивается с тензорами, — это невозможность их «визуализации»: их трудно нарисовать. По-видимому, в большинстве случаев лучше отказаться от попыток рисовать тензоры и работать с ними в рамках данного определения, т.е. думать о тензорах как о полилинейных функциях на векторах. Возникновение тензорного исчисления было подготовлено в 19 в. развитием теории алгебраических форм, с одной стороны, и теории квадратичных дифференциальных форм — с другой. Исследования в области теории дифференциальных квадратичных форм были непосредственно связаны с дифференциальной геометрией: с геометрией поверхностей (К. Гаусс) и с геометрией многомерных метрических пространств (Б. Риман). Современную форму тензорному исчислению придал Г. Риччи-Курбастро, поэтому тензорное исчисление иногда называют исчислением Риччи. Идеи Г. Риччи-Курбастро первоначально не получили широкого распространения. Внимание к ним возросло после появления общей теории относительности, математическая часть которой целиком основана на тензорном исчислении.

Глава 3

Дифференцирование

Дифференцирование (от лат. differentia — разность) математических объектов можно понимать аналитически — как метод анализа этих объектов или геометрически — как метод аппроксимации.

Анализ (от др.-греч. $\alpha\nu\alpha\lambda\upsilon\sigma\iota\varsigma$ — разложение, расчленение) — операция мысленного или реального расчленения целого объекта исследования на составные части.

Аппроксимация (от лат. approximo — приближаюсь) — замена одних объектов другими, более простыми или удобными, но в том или ином смысле близкими к исходным объектам.

Во многих математических книгах сделан упор на применяемый при дифференцировании аналитический аппарат. Данная глава также в основном посвящена тому, как проводить вычисления при дифференцировании различных математических объектов. Но, прежде чем перейти к аналитическим построениям, полезно уяснить хотя бы на нескольких примерах, что означает дифференцирование с точки зрения геометрии.

Рассмотрим плоскую (лежащую на плоскости) кривую. С геометрической точки зрения дифференцирование кривой в данной точке означает построение касательной прямой — прямой, проходящей через рассматриваемую точку кривой и совпадающей с ней в этой точке. В результате кривая заменяется (аппроксимируется) касательной прямой, которую значительно проще исследовать.

Дифференцирование кривой в каждой точке означает замену кривой множеством касательных — бесконечно малых открытых интервалов.

Очевидно, что касательную можно построить не всегда. Во-первых, кривая должна быть непрерывна в точке дифференцирования. Следовательно, *необходимым условием* дифференцируемости как кривой, так и любого другого дифференцируемого математического объекта, является его непрерывность, т.е. *всякий дифференцируемый объект непрерывен.*

Обратное утверждение не верно. Допустим, что непрерывная кривая имеет *излом*, т.е. рассматриваемая точка является *угловой точкой.* В этой точке мы можем аппроксимировать кривую двумя различными касательны-

ми, слева и справа от излома, т.е. непрерывная кривая не дифференцируема в точке излома. Можно построить непрерывную кривую, целиком состоящую из изломов, т.е. не дифференцируемую ни в одной точке (впервые на такие функции было указано Б. Больцано (1830)). Поэтому, во-вторых, дифференцируемые кривые должны не иметь изломов. Такие кривые называют гладкими.

Если теперь перейти к рассмотрению непрерывной поверхности, то её дифференцируемость или гладкость в данной точке означает возможность построить касательную плоскость — плоскость, содержащую касательные ко всем гладким кривым, проходящим через эту точку.

Дифференцирование поверхности (пространства) в каждой точке означает замену поверхности множеством касательных кружков (шариков) — бесконечно малых открытых окрестностей точки.

Гладкость — свойство функции или геометрической фигуры (кривой, поверхности и т.д.), состоящее в том, что у каждой точки данного объекта имеется окрестность, допускающая дифференцируемую параметризацию (касательную). В этом смысле говорят о гладкой функции, гладкой кривой, гладкой поверхности и т.д.

В то время как условие непрерывности является лишь необходимым, то условие гладкости является достаточным для дифференцируемости математического объекта.

Особое значение придаётся дифференцированию таких объектов как скалярные и векторные поля.

Скалярное поле (скалярная функция) — область в пространстве, в каждой точке $\mathbf{x}$ которой задан скаляр y.

Векторное поле — область в пространстве, из каждой точки $\mathbf{x}$ которой торчит вектор $\mathbf{a} = (a_1, \ldots, a_n)$ (на математическом языке — каждой точке отнесено значение некоторого вектора).

Введение понятия скалярных и векторных полей наряду со скалярами и векторами привело к разделению векторного исчисления на *векторную алгебру* и *векторный анализ*. В векторной алгебре изучают линейные операции и различные произведения векторов. В векторном анализе вместо скаляров и векторов изучают средствами математического анализа векторные и скалярные функции (векторные и скалярные поля).

Понятие «векторное поле» было введено Фарадеем (ок. 1830), т.е. раньше, чем алгебраическое понятие «поле», введённое в «Теории чисел» П. Дирихле (1871).

Если график скалярной функции (скалярного поля) есть поверхность в расширенном пространстве с координатами $x^1, \ldots, x^n, y$, а дифференцирование скалярного поля в каждой точке представляет собой построение касательной n-мерной плоскости к графику, то представить график векторного поля и, соответственно, дать геометрическую интерпретацию дифференцирования векторного поля значительно сложнее.

Исторически важнейшую роль в понимании физиками и математиками сущности векторных полей и методов работы с ними (векторного анализа) сыграла гидродинамика — наука о движении жидкости. При гидродинами-

ческой интерпретации векторное поле — это поле скоростей движущейся жидкости.

Рассмотрим векторное поле на плоскости. Его можно представить как поле скоростей при движении слоя жидкости очень малой глубины (достаточно малой, чтобы не учитывать движения перпендикулярно плоскости). Гидродинамические аналогии позволяют выделить три типичных случая векторного поля, представленных на рис. 3.1.

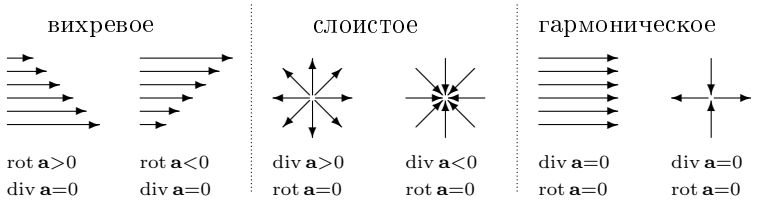


Рис. 3.1: Типичные случаи векторного поля

Вихревое поле — в окрестности рассматриваемой точки поля соседние векторы скорости различаются по длине. Если установить в данной точке поля маленькое колесо с лопастями, то разница в величине скоростей жидкости на лопастях заставит колесо вращаться либо против часовой стрелки (вправо), либо по часовой стрелке (влево), образуя *вихрь*.

Пример 1. Рассмотрим движении вязкой жидкости в трубке. У края трубки слой жидкости «прилипает» к стенке и остаётся неподвижным, при удалении от края скорость жидкости нарастает и становится максимальной по центру трубки. Поэтому соседние слои жидкости движутся с разной скоростью, образуя вихрь в каждой точке поля.

Пример 2. Рассмотрим вращающийся диск. Скорость точек диска увеличивается при удалении от его центра, поэтому соседние слои (окружности) на диске движутся с разной скоростью, образуя вихрь в каждой точке. Аналогичное движение происходит во вращающемся смерче.

Слоистое поле — в данной точке поля имеется отверстие или *источник*. Поэтому жидкость будет либо вытекать из этого отверстия, образуя *исток*, либо утекать в это отверстие, образуя *сток*.

Пример 3. Если в качестве векторного поля взять совокупность направлений наискорейшего спуска (стекания жидкости) на земной поверхности, то вершины и впадины будут источниками, причём вершины (расходящиеся направления спуска) будут истоками, а впадины (сходящиеся направления спуска) будут стоками.

Комбинация первого и второго случая не даёт нам ничего нового и поэтому не выделяется отдельно. Однако отдельно выделяется случай *гармонического поля* — когда в данной точке отсутствуют как вихри, так и источники.

Пример 4. Горные перевалы одновременно являются и вершинами и впадинами. Если количество сходящихся и расходящихся направлений спуска на перевале одинаково, то перевал не является источником.

Аналитически дифференцирование можно понимать как категорию, объектами которой являются дифференцируемые функции (функции, которые можно дифференцировать), а действующими на функции отображениями являются *операторы дифференцирования*, называемые также *дифференциальными операторами*.

Дифференцирование векторного поля $\mathbf{a}$ позволяет, в частности, вычислить точки, где имеются вихри и (или) источники. Для этих вычислений применяются два специальных дифференциальных оператора, которые подробно описаны в настоящей главе.

Оператор, обозначаемый $\operatorname{div}\mathbf{a}$, называется *дивергенцией* (от лат. divergentia — расходимость) и вычисляет скаляр в данной точке векторного поля. Точка является истоком при $\operatorname{div}\mathbf{a} > 0$ и стоком при $\operatorname{div}\mathbf{a} < 0$.

Оператор, обозначаемый $\operatorname{rot}\mathbf{a}$, называется *ротор* (от лат. roto — вращаю(сь)) или *вихрь* и вычисляет в данной точке поля аксиальный вектор (псевдовектор), направленный по оси вращения воображаемого маленького колеса с лопастями. При рассмотрении векторного поля на плоскости псевдовектор не принадлежит рассматриваемой плоскости векторного поля и состоит всего из одной компоненты, величина которой является числом. Закрученный против часовой стрелки (вправо) ротор торчит из плоскости вверх и считается положительным ($\operatorname{rot}\mathbf{a} > 0$), закрученный по часовой стрелке (влево) ротор торчит из плоскости вниз и считается отрицательным ($\operatorname{rot}\mathbf{a} < 0$).

Выбор положительного направления ротора естественно связан с выбором правой ориентации пространства и правилом правой руки. Для ротора это правило можно сформулировать следующим образом: Если обхватить воображаемое маленькое колесо с лопастями правой рукой так, чтобы четыре пальца указывали направление вращения колеса, то отставленный большой палец покажет направление ротора.

Если мы рассмотрим векторное поле не на плоскости, а в трёхмерном евклидовом пространстве, то проекции векторного поля на координатные плоскости позволят вычислить при дифференцировании три компоненты ротора. Поэтому ротор трёхмерного пространства уже не является числом и не может быть положительным или отрицательным. Однако направление ротора по-прежнему определяется по правилу правой руки, но уже принадлежит пространству векторного поля.

Если мы рассмотрим векторное поле в пространстве большей размерности, то количество компонент ротора, равное количеству координатных плоскостей C_n^2, оказывается больше размерности пространства n, и в виде вектора вычислить ротор невозможно. Однако оказывается, что для любого нечётномерного пространства (когда $n = 2k + 1$) возможно вычислить при дифференцировании *вектор направления* ротора, принадлежащий пространству векторного поля, что отражает существование оси у всякого вращения нечётномерного пространства.

В отличие от ротора дивергенция векторного поля является скаляром и без проблем вычисляется при любой размерности пространства.

Векторное поле, в котором отсутствуют вихри ($\operatorname{rot}\mathbf{a} = 0$), называется *по-*

тенциальным, *ньютоновым*, *слоистым* или *безвихревым*. Векторное поле, в котором отсутствуют источники ($\operatorname{div}\mathbf{a} = 0$), называется *соленоидальным* (от греч. *σωλήν* — трубка и *εἶδος* — вид), *лапласовым*, *трубчатым* или *вихревым*. Векторное поле, в котором отсутствуют как источники, так и вихри, называется *гармоническим*.

Термин «дивергенция» ввел У. Клиффорд (1878); ему же принадлежит обозначение div. Величину, отличную от дивергенции знаком, рассматривал Дж. Максвелл (1873), называя ее конвергенцией. Термин «ротор» ввёл У. Клиффорд (1878); ему же принадлежит обозначение rot, впрочем, ранее (1855, 1856) это обозначение (как и curt) применял Дж. Максвелл. Слово «соленоидальный» ввёл А. Ампер, термин «соленоидальное поле» принадлежит У. Томсону (лорду Кельвину).

3.1 Дифференцирование скалярной функции

График скалярной функции $f\colon \mathbb{R}^n \to \mathbb{R}$ или $z = f(x^1, \ldots, x^n)$ представляет собой изображение функции в виде n-мерной поверхности, состоящей из множества точек $(x^1, \ldots, x^n, f)$ в $(n+1)$-мерном пространстве с координатами $x^1, \ldots, x^n, z$. *Дифференцирование* скалярной функции в точке $\tau = (x_0^1, \ldots, x_0^n)$ означает аппроксимацию этой поверхности n-мерной плоскостью, которая касается графика функции в этой точке.

Касательная n-мерная плоскость к графику функции $(x^1, \ldots, x^n, f)$ в точке τ задаётся уравнением (2.16):

$$f(x^1, \ldots, x^n) - f(x_0^1, \ldots, x_0^n) = a_1(x^1 - x_0^1) + \ldots + a_n(x^n - x_0^n). \qquad (3.1)$$

Существенным требованием к касательной n-мерной плоскости является то, что она не должна быть параллельна оси z (не должна содержать вектора, параллельного оси z).

Если считать окрестность точки дифференцирования очень небольшой (бесконечно малой), то соответствующие *бесконечно малые разности* аргументов $x^i - x_0^i$ и функции $f - f(\tau)$ относительно точки дифференцирования τ обозначаются символами $dx^1, \ldots, dx^n, df$ и называются *дифференциалами* (от лат. differentia — разность). Тогда

$$df = a_1(x^1, \ldots, x^n)\, dx^1 + \ldots + a_n(x^1, \ldots, x^n)\, dx^n. \qquad (3.2)$$

Как мы видим, дифференциал задает уравнение касательной плоскости в рассматриваемой точке дифференцирования скалярной функции. В этом и заключается его геометрический смысл.

Функция $f\colon \mathbb{R}^n \to \mathbb{R}$ дифференцируема на множестве U, если f дифференцируема в каждой точке $\tau \in U$. Функция f вообще *дифференцируема*, если её можно продолжить до дифференцируемой функции на некотором открытом множестве, содержащем U.

Например, дифференцирование функции $z = f(x)$ одного переменного в точке x_0 (рис. 3.2) означает аппроксимацию кривой $\{x, f\}$ в бесконечно

малой окрестности этой точки касательной с уравнением

$$df = a(x_0)\,dx.$$

Производной функции f в точке x_0 называется величина $a(x_0) = f'(x_0)$ в уравнении касательной, равная тангенсу угла наклона касательной, и вычисляемая как предел (если, конечно, он существует):

$$f'(x_0) = \frac{df}{dx} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}. \tag{3.3}$$

Определению (3.3) можно придать вид, более соответствующий представлению об аппроксимации функции. Введём обозначение $\Delta x = x - x_0$. Говорят, что функция $f\colon \mathbb{R} \to \mathbb{R}$ *дифференцируема* в точке $x_0 \in \mathbb{R}$, если существует такая функция $f'(x_0)\colon \mathbb{R} \to \mathbb{R}$, что

$$\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0) - f'(x_0)\,\Delta x}{\Delta x} = 0. \tag{3.4}$$

Аппроксимация кривой $f(x)$ в точке x_0 касательной $\Delta f = f'(x_0)\,\Delta x$ означает, что хотя величины Δx и $f(x_0 + \Delta x) - f(x_0)$ малы, но величина $f(x_0 + \Delta x) - f(x_0) - \Delta f$, которая есть расстояние между кривой и касательной, имеет *больший порядок малости*.

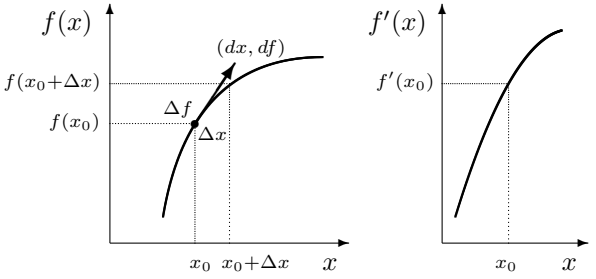


Рис. 3.2: Дифференцирование $z = f(x)$.

Если функция f дифференцируема на некотором открытом интервале, то её производная на этом интервале обозначается f'.

Само собой разумеется, что эта производная также представляет собой функцию, которую можно изобразить на графике (рис. 3.2).

Хотя в курсах анализа при введении производной учат, что $f' = \frac{df}{dx}$ не дробь, а единый символ $\frac{d}{dx}$, обозначающий *оператор дифференцирования*, который действует на функцию f, мы обращаемся с этим символом как с дробью, состоящей из двух *операторов* d — дифференциалов.

Лейбниц не стал бы вводить это сложное обозначение, если бы не имел в виду самой настоящей дроби. Дело в том, что df и dx — вовсе не таинственные «бесконечно малые величины», а вполне конечные числа, точнее — функции вектора (dx, df) — касательного к кривой $z = f(x)$ в каждой точке дифференцирования. Этот вектор задаёт направление касательной $df = f'dx$ в каждой дифференцируемой точке кривой $f(x)$.

Вычислим производные элементарных функций. При вычислениях необходимо учитывать, что все члены разложения, которые после деления на Δx всё ещё содержат в числителе множитель Δx, имеют больший порядок малости и в пределе равны нулю:

1. $\frac{dc}{dx} \overset{(3.3)}{=} \lim\limits_{\Delta x\to 0} \frac{c-c}{\Delta x} = 0$ — производная числа c равна нулю,

2. $\frac{dx^n}{dx} = \lim\limits_{\Delta x\to 0} \frac{(x+\Delta x)^n - x^n}{\Delta x} =$
$\overset{(1.24)}{=} \lim\limits_{\Delta x\to 0} (nx^{n-1} + \frac{n(n-1)}{2} x^{n-2}\Delta x + \ldots) = nx^{n-1},$

3. $\frac{de^x}{dx} \overset{(1.30)}{=} \frac{d}{dx}(1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots) =$
$= 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots = e^x,$

4. $\frac{d\ln x}{dx} \overset{(3.3)}{=} \lim\limits_{\Delta x\to 0} \frac{\ln(x+\Delta x) - \ln x}{\Delta x} = \lim\limits_{\Delta x\to 0} \frac{\ln(1+\frac{\Delta x}{x})}{\Delta x} =$
$\overset{(1.34)}{=} \lim\limits_{\Delta x\to 0} \left(\frac{1}{x} - \frac{\Delta x}{2x^2} + \ldots\right) = \frac{1}{x},$

5. $\frac{d\sin x}{dx} \overset{(3.3)}{=} \lim\limits_{\Delta x\to 0} \frac{\sin(x+\Delta x) - \sin x}{\Delta x} \overset{(1.74)}{=} \lim\limits_{\Delta x\to 0} \frac{\sin\Delta x\cdot\cos x + (\cos\Delta x - 1)\sin x}{\Delta x} =$
$\overset{\substack{(1.32)\\(1.31)}}{=} \cos x \cdot \lim\limits_{\Delta x\to 0} \left(1 - \frac{\Delta x^2}{3!} + \ldots\right) + \lim\limits_{\Delta x\to 0} \left(-\frac{\sin x\cdot\Delta x}{2!} + \ldots\right) = \cos x,$

6. $\frac{d\cos x}{dx} \overset{(3.3)}{=} \lim\limits_{\Delta x\to 0} \frac{\cos(x+\Delta x) - \cos x}{\Delta x} =$
$\overset{(1.73)}{=} \lim\limits_{\Delta x\to 0} \frac{(\cos\Delta x - 1)\cdot\cos x - \sin\Delta x\cdot\sin x}{\Delta x} = -\sin x.$

Существуют правила дифференцирования и соответствующие таблицы для всех элементарных функций. При этом оказалось, что производная элементарной функции всегда дает элементарную функцию.

Тоерема о дифференцировании сложной функции. Если функция $g\colon \mathbb{R} \to \mathbb{R}$ дифференцируема в точке $x_0 \in \mathbb{R}$, а функция $f\colon \mathbb{R} \to \mathbb{R}$ дифференцируема в точке $g(x_0) \in \mathbb{R}$, то композиция $f \circ g\colon \mathbb{R} \to \mathbb{R}$ дифференцируема в x_0 и тогда:

$$\frac{df}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx} = f'_g(g(x_0)) \cdot g'_x(x_0). \tag{3.5}$$

Правила вычисления производных легко получают из определения производной и теоремы о дифференцировании сложной функции:

$$(f+g)' = f' + g', \tag{3.6}$$
$$(cf)' = cf' \quad (c \in \mathbb{R}), \tag{3.7}$$
$$(f\cdot g)' = f'g + g'f, \tag{3.8}$$
$$\left(\frac{f}{g}\right)' = \frac{f'g - g'f}{g^2}. \tag{3.9}$$

Теорема об обратной функции. Если функция $f\colon \mathbb{R} \to \mathbb{R}$ или $y = f(x)$ непрерывно дифференцируема на открытом интервале $U \subset \mathbb{R}$, содержащем точку a, и $f'(a) \neq 0$ для всех $a \in U$, тогда эта функция взаимно однозначна и имеет обратную функцию $x = f^{-1}(y)$, также дифференцируемую на открытом интервале $V \subset \mathbb{R}$, содержащем точку $b = f(a)$.

При этом:

$$\left(f^{-1}(b)\right)'_y = \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{f'(f^{-1}(b))}. \tag{3.10}$$

В случае функции $f\colon \mathbb{R}^n \to \mathbb{R}$ или $y = f(x^1, \ldots, x^n)$ величины $a_1, \ldots, a_n$ в (3.1) вычисляют аналогично производной функции одного переменного в предположении, что кроме аргумента x^i все аргументы $x^j \neq x^i$ сохраняют постоянные значения, и функция нескольких переменных превращается в функцию одного переменного x^i. Эти величины обозначаются как $a_i = f'_{x^i}$ и называются *частные производные*:

$$f'_{x^i} = \left(\frac{\partial f}{\partial x^i}\right)_{x^j \neq x^i} = \frac{\partial f}{\partial x^i} \quad (i = 1, \ldots, n). \tag{3.11}$$

Удобство обозначения оператора $\frac{\partial}{\partial x}$ состоит в том, что при растяжении оси x, $\frac{\partial}{\partial x}$ ведёт себя как $\frac{1}{x}$.

Эластичность функции или относительная производная

$$E_i(f) = \frac{x^i}{f}\frac{\partial f}{\partial x^i} \tag{3.12}$$

есть предел отношения относительного приращения $\frac{\Delta f}{f}$ функции f к относительному приращению $\frac{\Delta x^i}{x^i}$ независимой переменной x^i.

Для сложной функций $f \circ g\colon \mathbb{R}^n \to \mathbb{R}$ при вычислении частной производной имеет место *цепное правило*

$$f'_{x^i} \overset{(3.5)}{=} f'_g \cdot g'_{x^i} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial x^i} \quad (i = 1, \ldots, n). \tag{3.13}$$

Набор компонент $f'_{x^i} = \frac{\partial f}{\partial x^i}$ $(i = 1, \ldots, n)$ во всех дифференцируемых точках функции $f(x^1, \ldots, x^n)$ имеет специальное название — *градиент* функции f и обозначается $\operatorname{grad} f$ (от лат. gradiens — шагающий):

$$\operatorname{grad} f = \left(\frac{\partial f}{\partial x^1}, \ldots, \frac{\partial f}{\partial x^n}\right). \tag{3.14}$$

Градиент скалярной функции, как n-членная последовательность, представляет собой вектор и является *инвариантом*, т.е. не меняется при изменении системы координат, хотя его компоненты при этом изменяются и поэтому являются не скалярными, а координатными функциями.

С другой стороны, градиент есть действующий на скалярное поле дифференциальный оператор, который производит векторное поле.

Пример. Дифференцирование функции двух переменных $z = f(x, y)$ в точке $\tau = (x_0, y_0)$ означает аппроксимацию поверхности $\{x, y, f\}$ в бесконечно малой окрестности этой точки касательной плоскостью

$$df = f'_x\, dx + f'_y\, dy, \tag{3.15}$$

которая может быть представлена в пространстве $\{x, y, z\}$ как скалярное произведение вектора $\operatorname{grad} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$ и вектора $\xi = (dx, dy)$:

$$df = \operatorname{grad} f \cdot \xi. \tag{3.16}$$

Если же принять точку τ за начало координат, то касательная плоскость представляет собой двумерное пространство с собственными координатами dx, dy (рис. 3.3).

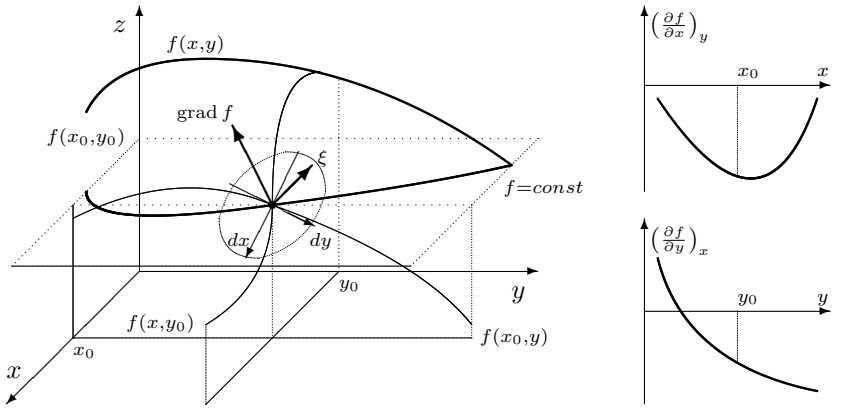


Рис. 3.3: Дифференцирование функции $z = f(x, y)$.

Изменение функции f при движении по поверхности её графика зависит от направления вектора ξ: $df(\xi)$. Если выбрано направление, совпадающее с направлением $\operatorname{grad} f$, то скалярное произведение этих векторов даст максимальную величину $df(\xi)$. Таким образом, градиент указывает в сторону наибыстрейшего возрастания функции f. Если выбрано направление, перпендикулярное направлению $\operatorname{grad} f$, то получаем $df(\xi) = 0$ или $f(x, y) = const$, т.е. кривую на графике с одним и тем же постоянным значением функции (*линию уровня*), образованную сечением поверхности функции горизонтальной плоскостью $z = f(x_0, y_0)$.

Если график $\{x, y, f\}$ образует выпуклую вверх двумерную поверхность (холм), то градиент указывает на вершину холма и перпендикулярен одномерным *линиям уровня*, в данном случае соединяющим точки одинаковой высоты на поверхности холма.

Мы получили, что $\operatorname{grad} f$ в каждой дифференцируемой точке графика функции указывает направление наибольшего возрастания функции и перпендикулярен в каждой точке кривой $f = const$, т.е. является *нормалью* к кривым (линиям) уровня функции.

Линию уровня называют также *изолиния* — линия на графике функции, в каждой точке которой значение либо самой функции, либо одной из её координат сохраняет одинаковое значение.

Например, сечения графика $\{x, y, f\}$ плоскостями $x = x_0$ и $y = y_0$ представляют собой изолинии $f(x_0, y)$ и $f(x, y_0)$. При условии дифференцируемости функции $f(x, y)$ в каждой точке эти линии имеют производные $\left(\frac{\partial f}{\partial x}\right)_y$ и $\left(\frac{\partial f}{\partial y}\right)_x$, графики которых также изображены на рис. 3.3.

Поверхность уровня $f(x, y, z) = const$ в случае функции трёх переменных представляет собой двумерную поверхность в трёхмерном пространстве графика функции $\{x, y, z, f\}$ с нормалью $\operatorname{grad} f$, а касательная плоскость при дифференцировании графика функции представляет собой трёхмерное пространство с координатами dx, dy, dz (рис. 3.4).

Градиент функции трёх переменных $\operatorname{grad} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$ в каждой точке (x_0, y_0, z_0) перпендикулярен (является нормалью) поверхности уровня $f(x, y, z) = const$, которой принадлежит данная точка.

Пример. Рассмотрим сферу радиуса $r = const$. Тогда

$$\operatorname{grad} r = \frac{\mathbf{r}}{r} \tag{3.17}$$

есть единичный вектор направленный по радиус-вектору $\mathbf{r}$ или нормаль к поверхности сферы в каждой её точке.

Доказательство (3.17) для $r = \sqrt{x^2 + y^2 + z^2}$

$$\operatorname{grad}_x r = \frac{\partial r}{\partial x} = \frac{\partial \sqrt{x^2 + y^2 + z^2}}{\partial x} \overset{(3.13)}{=} \frac{x}{r}. \tag{3.18}$$

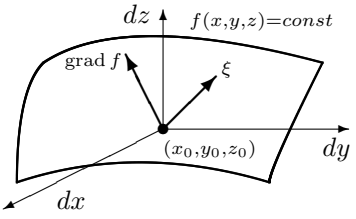


Рис. 3.4: Касательное пространство

В общем случае функции $z = f(x^1, \ldots, x^n)$ касательная плоскость образована множеством *касательных векторов*

$$\xi = (dx^1, \ldots, dx^n),$$

называется *n-мерным касательным пространством* в точке τ с координатами $dx^1, \ldots, dx^n$ и обозначается $T\mathbb{R}^n_\tau$, а точка дифференцирования τ принимается за начало координат этого нового пространства.

В этом случае $(n-1)$-мерное *множество уровня* — множество точек пространства, на котором функция $f(x^1, \ldots, x^n) = const$ имеет постоянное значение, называется также *геометрическая связь* на n-мерной поверхности графика функции.

Градиент называется также *производной функции* f в точке τ. Другими словами, говорят, что функция $z = f(x^1, \ldots, x^n)$ *дифференцируема* в точке $\tau \in \mathbb{R}^n$, если существует такое отображение $f'(\tau)\colon \mathbb{R}^n \to \mathbb{R}$, что выполняется условие:

$$\lim_{\xi \to 0} \frac{f(\tau + \xi) - f(\tau) - f'(\tau) \cdot \xi}{\|\xi\|} = 0. \tag{3.19}$$

где $\xi = (dx^1, \ldots, dx^n)$ — касательный вектор.

Заметим, что ξ есть отрезок из n-мерного пространства $T\mathbb{R}^n_\tau$, а $f(\tau+\xi) - f(\tau) - f'(\tau)\cdot\xi$ — число, так что обозначение расстояния $\|\xi\|$ в определении (3.19) существенно.

Тогда мы можем записать *дифференциал* (3.2) в следующем виде

$$df(\xi) = f'(\tau)\cdot\xi = \frac{\partial f}{\partial x^1}\,dx^1 + \ldots + \frac{\partial f}{\partial x^n}\,dx^n \overset{(2.17)}{=} \frac{\partial f}{\partial x^i}\,dx^i, \tag{3.20}$$

где *оператор дифференцирования* d имеет вид

$$d = dx^1\frac{\partial}{\partial x^1} + \ldots + dx^n\frac{\partial}{\partial x^n}. \tag{3.21}$$

Производную f' назвают также *векторной пространственной производной скалярной функции*.

Дифференцируемость функции одного переменного равносильна условию существованию производной. Чтобы функция $z = f(x^1, \ldots, x^n)$ была дифференцируемой (гладкой) функцией, к условию существования всех частных производных достаточно добавить условие их непрерывности.

Касательное пространство $T\mathbb{R}^n_\tau$ к графику функции f в точке τ состоит из множества касательных векторов ξ, т.е. из множества всех пар (τ, ξ), где точка τ задана, а вектор ξ пробегает $\mathbb{R}^n$. Пару векторов (τ, ξ) или вектор ξ, торчащий из точки τ, можно записать как ξ_τ.

Отображение $\varphi\colon T\mathbb{R}^n \to \mathbb{R}^n$, сопоставляющее каждому вектору ξ точку τ в которой вектор касается $\mathbb{R}^n$, называют *естественной проекцией*. Прообраз точки τ при естественной проекции $\varphi^{-1}(\tau)$, есть касательное пространство $T\mathbb{R}^n_\tau$.

Если функция f дифференцируема на всём пространстве $\mathbb{R}^n$, то объединение всех касательных пространств $\bigcup\limits_{\tau\in\mathbb{R}^n} T\mathbb{R}^n_\tau$ во всех разных точках τ пространства $\mathbb{R}^n$ даёт *касательное расслоение* пространства $\mathbb{R}^n$ и обозначается $T\mathbb{R}^n$, а $T\mathbb{R}^n_\tau$ есть *слой расслоения* над точкой τ.

Например, для дифференцируемой кривой объединение касательных в каждой точке кривой в некую двумерную поверхность представляет собой «касательное расслоение», а каждая касательная является «слоем расслоения» над соответствующей точкой.

Размерность касательного расслоения $T\mathbb{R}^n$ вдвое больше размерности $\mathbb{R}^n$, а его точка это пара (τ, ξ) или $2n$-членная последовательность

$$(\tau, \xi) = (x^1, \ldots, x^n, dx^1, \ldots, dx^n).$$

3.2 Различные виды производных

Производная вектор–функции. *Вектор–функция* n переменных $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ или $\mathbf{y} = \varphi(\mathbf{x})$ представляет собой набор из m скалярных функций $y^1 = f_1(\mathbf{x}), \ldots, y^m = f_m(\mathbf{x})$.

Вектор–функция $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ дифференцируема в точке $\tau \in \mathbb{R}^n$, если существует такой линейный оператор $\varphi'(\tau)\colon \mathbb{R}^n \to \mathbb{R}^m$, что:

$$\lim_{\xi \to 0} \frac{\|\varphi(\tau+\xi) - \varphi(\tau) - \varphi'(\tau)\cdot\xi\|}{\|\xi\|} = 0. \tag{3.22}$$

Величина $\varphi'(\tau)\cdot\xi$ есть произведение матрицы $\varphi'(\tau)$ на вектор ξ. Заметим, что ξ есть отрезок из n-мерного пространства $T\mathbb{R}^n_\tau$, а $\varphi(\tau+\xi)-\varphi(\tau)-\varphi'(\tau)\cdot\xi$ — отрезок из $\mathbb{R}^m$, так что обозначения расстояний между точками $\|\,\|$ существенны и в этом определении.

Матрица $\varphi'(\tau)$ линейного оператора $\varphi'(\tau)\colon \mathbb{R}^n \to \mathbb{R}^m$ называется производной функции $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ в точке $\tau \in \mathbb{R}^n$ и есть $(m \times n)$-матрица, которая называется ещё *матрицей Якоби*, и также может быть названа *тензорной пространственной производной вектор–функции*:

$$\varphi'(\tau) = \frac{\partial\varphi}{\partial\mathbf{x}} = \left\| \begin{array}{ccc} \dfrac{\partial y^1}{\partial x^1} & \cdots & \dfrac{\partial y^1}{\partial x^n} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial y^m}{\partial x^1} & \cdots & \dfrac{\partial y^m}{\partial x^n} \end{array} \right\| = \left\| \begin{array}{c} \operatorname{grad} f_1(\tau) \\ \cdots \\ \operatorname{grad} f_m(\tau) \end{array} \right\|. \tag{3.23}$$

Если функция $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ дифференцируема в точке τ, то существует *единственное* линейное отображение $\varphi'(\tau)\colon \mathbb{R}^n \to \mathbb{R}^m$. Непрерывными и дифференцируемыми при этом являются все функции $f_1, \ldots, f_m$.

Функция $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ дифференцируема на множестве $U \subset \mathbb{R}^n$ (вообще *дифференцируема*), если она дифференцируема в каждой точке $\tau \in U$ (на некотором открытом множестве, содержащем U).

Теорема о дифференцировании сложной функции. Если $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ дифференцируема в точке $\tau \in \mathbb{R}^n$, а функция $\psi\colon \mathbb{R}^m \to \mathbb{R}^p$ дифференцируема в точке $\varphi(\tau) \in \mathbb{R}^m$, то композиция $\psi\circ\varphi\colon \mathbb{R}^n \to \mathbb{R}^p$ дифференцируема в τ и тогда:

$$(\psi\circ\varphi)' = \frac{\partial\psi}{\partial\varphi}\cdot\frac{\partial\varphi}{\partial\mathbf{x}} = \psi'_\varphi(\varphi(\tau))\cdot\varphi'_\mathbf{x}(\tau). \tag{3.24}$$

Пусть $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ — функция преобразования.

Определитель квадратной $(n \times n)$-матрицы Якоби называют *якобианом* и обозначают:

$$\det\varphi'_\mathbf{x}(\tau) = J = \frac{\partial(f_1,\ldots,f_n)}{\partial(x^1,\ldots,x^n)} = \left| \begin{array}{ccc} \dfrac{\partial y^1}{\partial x^1} & \cdots & \dfrac{\partial y^1}{\partial x^n} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial y^n}{\partial x^1} & \cdots & \dfrac{\partial y^n}{\partial x^n} \end{array} \right| = \left| \begin{array}{c} \operatorname{grad} f_1 \\ \cdots \\ \operatorname{grad} f_n \end{array} \right|. \tag{3.25}$$

Если градиенты компонентов вектор–функции линейно независимы, то якобиан отличен от нуля. В этом случае мы можем вычислить производную вектор-функции, обратной к исходной вектор–функции.

Теорема об обратной функции. Предположим, что $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ или $\mathbf{y} = \varphi(\mathbf{x})$ непрерывно дифференцируема в некотором открытом множестве U, содержащем точку τ, и $\det \varphi'_{\mathbf{x}}(\tau) \neq 0$. Тогда существует открытое множество V содержащее точку $\upsilon = \varphi(\tau)$, такое, что отображение $\varphi\colon U \to V$ имеет непрерывное обратное отображение $\varphi^{-1}\colon V \to U$ или $\mathbf{x} = \varphi^{-1}(\mathbf{y})$, дифференцируемое для всех точек $\upsilon \in V$ и удовлетворяющее соотношению:

$$\left(\varphi^{-1}(\upsilon)\right)'_{\mathbf{y}} = \left[\varphi'_{\mathbf{x}}(\varphi^{-1}(\upsilon))\right]' = \frac{\partial \varphi^{-1}}{\partial \mathbf{y}} = \begin{Vmatrix} \dfrac{\partial x^1}{\partial y^1} & \cdots & \dfrac{\partial x^1}{\partial y^n} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial x^n}{\partial y^1} & \cdots & \dfrac{\partial x^n}{\partial y^n} \end{Vmatrix}. \tag{3.26}$$

Таким образом, матрица Якоби обратной функции равна матрице, обратной к матрице Якоби исходной функции.

Обратим внимание ещё раз: Если дифференцируемая функция преобразования пространства в себя имеет взаимно однозначную дифференцируемую обратную функцию, то якобиан преобразования отличен от нуля, а матрица Якоби обратной функции равна матрице, обратной к матрице Якоби исходной функции. Этот вывод показывает связь понятий алгебры и дифференциального исчисления, а также имеет наглядное геометрическое истолкование: производная вектор–функции является линейным оператором, взаимно однозначно преобразующим любой ненулевой вектор касательного пространства в ненулевой.

Пусть U и V — открытые множества в $\mathbb{R}^n$, тогда дифференцируемая функция $\varphi\colon U \to V$, имеющая дифференцируемую обратную функцию, называется *диффеоморфизм* — гомеоморфизм, являющийся дифференцируемым отображением, или заменой переменных (некой функции) из U на переменные из V, гладкой вместе с обратной заменой.

Пример. Связь между координатами разных систем, например декартовой и криволинейной, в каждой точке пространства осуществляется матрицей Якоби:

$$dx = \frac{\partial x}{\partial u}\,du + \frac{\partial x}{\partial v}\,dv + \frac{\partial x}{\partial w}\,dw \quad \text{и т. д.} \tag{3.27}$$

Выберем базис $e_1, \ldots, e_m$ пространства $\mathbb{R}^m$. Разумеется, эти базисные векторы отличаются от векторов $e_1, \ldots, e_n$ координатного пространства $\mathbb{R}^n$, так как принадлежат разным пространствам. Тогда вектор-функция может быть записана в виде вектора $\mathbf{y} = e_1 y^1 + \ldots + e_m y^m$, как точка пространства $\mathbb{R}^m$, аналогично записи для точки $\mathbf{x} = e_1 x^1 + \ldots + e_n x^n$ пространства $\mathbb{R}^n$.

Для любой точки $\varphi(\tau) \in \mathbb{R}^m$, где $\tau \in \mathbb{R}^n$ — точка дифференцирования функции, мы можем задать касательное пространство $T\mathbb{R}^m_{\varphi(\tau)}$ с координатами $dy^1, \ldots, dy^m$ и базисом $e_1^{\varphi(\tau)}, \ldots, e_m^{\varphi(\tau)}$, состоящее из множества касательных векторов $\eta = \varphi'_{\mathbf{x}}(\tau) \cdot \xi = e_1^{\varphi(\tau)} dy^1 + \ldots + e_m^{\varphi(\tau)} dy^m$.

Мы можем ввести обозначения для касательных векторов пространств $\mathbb{R}^n$ и $\mathbb{R}^m$ как для дифференциалов:

$$d\mathbf{x} = \xi, \quad d\mathbf{y}(\xi) = \eta = \varphi'_{\mathbf{x}}\, d\mathbf{x} = \varphi'_{\mathbf{x}}(\tau) \cdot \xi. \tag{3.28}$$

Производная функции комплексного переменного. Функцию комплексного переменного $f\colon \mathbb{C} \to \mathbb{C}$ называют *дифференцируемой* в точке $z_0 \in \mathbb{C}$, если существует предел:

$$f'(z_0) = \frac{df}{dz} = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}, \tag{3.29}$$

который называется производной функции комплексного переменного.

Если f дифференцируема в каждой точке z открытого множества $A \subset \mathbb{C}$ и f' непрерывна на A, то функцию f также называют *аналитической* на A. Это определение аналитической функции отличается от определения на стр. 40 по форме, но аналогично ему по сути.

Поскольку для комплексных чисел определены те же операции, что и для вещественных, то остаются в силе все основные свойства производной и формулы дифференцирования, выведенные для функций вещественного переменного. Например:

$$\frac{d(f_1 + f_2)}{dz} = \frac{df_1}{dz} + \frac{df_2}{dz}, \quad \frac{d(f_1 f_2)}{dz} = \frac{df_1}{dz} f_2 + f_1 \frac{df_2}{dz}. \tag{3.30}$$

Запишем функцию комплексного переменного в виде $f(z) = u + iv$, где $u(x, y)$ и $v(x, y)$ — функции двух действительных переменных.

Так как при вычислении значений функции $f(z)$ над ее комплексным аргументом z алгебраические действия могут производиться без расчленения z на вещественную и мнимую части, то оказывается, что для дифференцируемости (аналитичности) комплексной функции $f = u + iv$, недостаточно одной дифференцируемости функций вещественных переменных $u(x, y)$ и $v(x, y)$ по вещественным аргументам x и y.

Действительно, если производная $\frac{df}{dz}$ существует, то при различных способах изменения z соответствующие изменения f таковы, что предел (3.29) не зависит от того, каким образом Δz стремится к нулю.

Например, пусть при дифференцировании функции f изменяется только x, тогда $dz = dx$ и

$$\frac{df}{dz} = \frac{\frac{\partial f}{\partial x}\, dx}{dx} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}. \tag{3.31}$$

Если же изменяется только y, то $dz = i\,dy$, поэтому

$$\frac{df}{dz} = \frac{\frac{\partial f}{\partial y}\, dy}{i\,dy} = -i \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}. \tag{3.32}$$

Приравнивая два выражения для производной, полученных различными способами, находим:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}; \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \tag{3.33}$$

Формулы (3.33) называются *условия Коши–Римана*.

С учётом условий (3.33) производную функции $f(z)$ можно представить в следующих равносильных формах:

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y} = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y} = \frac{\partial v}{\partial y} + i\frac{\partial v}{\partial x}. \tag{3.34}$$

Существование производной функции комплексного переменного — это *структурное свойство* функции, тесно связанное с геометрическими представлениями. Такой подход был предложен О. Коши и развит Б. Риманом в 19 в. Рассмотренный на стр. 40 аналитический подход, основанный на возможности представления функции степенным рядом, систематически развивался К. Вейерштрассом.

Решающее значение в построении теории комплексных функций сыграл «выход в комплексную область» — переход от действительного переменного к комплексному переменному.

Надо сказать, что функция комплексного переменного имеет громадное прикладное значение. Такие дисциплины, как электродинамика и гидродинамика, вообще не мыслимы без использования этой функции. Дело в том, что переход к комплексным числам обычно не усложняет, а упрощает задачу. Например, в первой главе мы упоминали, что всякое уравнение степени n имеет ровно n комплексных корней. В то время как нахождение вещественных корней — нелегкая задача. В конце 19 в. и начале 20 в., например, трактаты по гидродинамике в основном состояли из длинных выкладок с использованием элементарных и специальных функций. По образному выражению одного из известных американских гидродинамиков С. Голдстейна, за этими выкладками никак нельзя было увидеть саму воду, нельзя представить, что она мокрая. С использованием комплексных переменных решение часто удаётся существенно упростить.

Производные высших порядков. Производная f' дифференцируемой функции одного переменного $y = f(x)$ в свою очередь может иметь производную, которую называют *второй производной* (*производной второго порядка*) функции f и обозначают f'' или $\frac{d^2f}{dx^2}$.

Тогда f' или $\frac{df}{dx}$ называют *первой производной*.

Частные производные второго порядка скалярной функции $y = f(x^1, \ldots, x^n)$ вычисляются как частные производные от частных производных первого порядка и записываются в виде:

$$f''_{x^ix^j} = \frac{\partial^2 f}{\partial x^i \partial x^j}, \quad f''_{x^ix^j} = \frac{\partial^2 f}{\partial (x^i)^2} \tag{3.35}$$

Производная $f''_{x^ix^j}$ называется *смешанной*, а $f''_{x^ix^j}$ называется *чистой* частной производной второго порядка.

Матрица, составленная из частных производных второго порядка скалярной функции, называется её *второй производной*:

$$\frac{\partial^2 f}{\partial \mathbf{x}^2} = f''(\tau) = \begin{Vmatrix} f''_{x^1x^1} & \cdots & f''_{x^1x^n} \\ \cdots & \cdots & \cdots \\ f''_{x^nx^1} & \cdots & f''_{x^nx^n} \end{Vmatrix}. \tag{3.36}$$

Определитель $f''(\tau)$ называется *гессиан* или определитель Гессе:

$$\det f'' = \begin{vmatrix} f''_{x^1x^1} & \cdots & f''_{x^1x^n} \\ \cdots & \cdots & \cdots \\ f''_{x^nx^1} & \cdots & f''_{x^nx^n} \end{vmatrix} \tag{3.37}$$

Определитель Гессе есть инвариант, т.е. он не изменяется при линейных преобразованиях координат.

Определение. Дифференцируемая функция $f\colon \mathbb{R}^n \to \mathbb{R}$ дважды дифференцируема в точке $\tau \in \mathbb{R}^n$, если существует такой линейный оператор $f''\colon \mathbb{R}^n \to \mathbb{R}^n$, что выполняется условие:

$$\lim_{\xi \to 0} \frac{f(\tau+\xi) - f(\tau) - f'(\tau)\cdot\xi - \xi^* \cdot f'' \cdot \xi}{\xi \cdot \xi} = 0, \tag{3.38}$$

где $\xi^* \cdot f'' \cdot \xi$ — квадратичная форма,

$\xi \cdot \xi$ — скалярное произведение.

Скалярное произведение $\xi \cdot \xi$, где ξ — касательный вектор, называется *первая квадратичная форма* или *метрика* касательного пространства:

$$\xi \cdot \xi = \sum_{i,j=1}^{n} (e_i^\tau \cdot e_j^\tau)\, dx^i dx^j = \sum_{i,j=1}^{n} g_{ij}(\tau)\, dx^i dx^j = ds^2 \geqslant 0, \tag{3.39}$$

где ds — *линейный элемент* пространства или отрезок, соединяющий две близкие точки пространства,

g_{ij} — дважды ковариантный метрический тензор или положительно полуопределённая симметрическая ($g_{ij} = g_{ji}$) матрица Грама.

Квадратичная форма $\xi^* \cdot f'' \cdot \xi$ называется *вторая квадратичная форма* или *дифференциал второго порядка* скалярной функции:

$$d^2 f = f'' d\mathbf{x}^2 = \xi^* \cdot f'' \cdot \xi = \sum_{i,j=1}^{n} f''_{x^ix^j} dx^i dx^j. \tag{3.40}$$

Геометрический смысл дифференциала первого порядка df (с. 105) заключается в задании уравнения касательной плоскости. При этом величина $f(\tau+\xi) - f(\tau) - f'(\tau)\cdot\xi$ в определении производной функции $f\colon \mathbb{R}^n \to \mathbb{R}$ (3.19) есть *расстояние* между n-мерной поверхностью графика функции и n-мерной касательной плоскостью и имеет больший порядок малости, чем расстояние $\|\xi\|$, поэтому предел их отношения равен нулю, и оценку этого расстояния с помощью df произвести нельзя.

Величина дифференциала второго порядка $d^2 f$ вычисляет *расстояние* между n-мерной поверхностью графика функции и n-мерной касательной плоскостью с точностью до *второго порядка малости* по величине $\|\xi\|$. Причем знак этой величины позволяет определить каким образом в данной точке относительно касательной плоскости располагается поверхность (выше или ниже плоскости). Имеется ввиду произвольная точка в окрестности точки дифференцирования, так как в самой точке дифференцирования все $dx^i = 0$, и $d^2 f = 0$.

Геометрический смысл дифференциала второго порядка состоит в том, что поведение его знака в окрестности точки дифференцирования (знакоопределённый, знакопеременный и т.д.) позволяет определить геометрию этой точки (см. с. 147).

Первая (3.39) и вторая (3.40) квадратичные формы являются примерами *дифференциальных форм*, т.е. многочленов от дифференциалов переменных, каждый член которых имеет относительно дифференциалов одну и ту же степень.

Если функция $y = f(x^1, \ldots, x^n)$ дважды дифференцируема в точке $\tau \in \mathbb{R}^n$, то существует единственное линейное отображение $f'' \colon \mathbb{R}^n \to \mathbb{R}^n$. При этом все функции f'_{x^i} являются непрерывными и дифференцируемыми на открытом множестве, содержащем точку τ.

Для того чтобы функция $y = f(x^1, \ldots, x^n)$ была дважды дифференцируемой функцией, к условию существования всех частных производных $f''_{x^i x^j}$ достаточно добавить условие их непрерывности. Доказано, что в этом случае $f''_{x^i x^j} = f''_{x^j x^i}$, и f'' — симметрическая матрица.

Значки d в df и d^2 в $d^2 f$ называют *операторами дифференцирования*:

$$d = \sum_{i=1}^{n} dx^i \frac{\partial}{\partial x^i}, \quad d^2 = \sum_{i,j=1}^{n} dx^i dx^j \frac{\partial^2}{\partial x^i \partial x^j} \tag{3.41}$$

Если функция $y = f(x^1, \ldots, x^n)$, определенная на некотором множестве U в $\mathbb{R}^n$, k раз дифференцируема, и все ее частные производные до порядка k включительно непрерывны на U, то говорят о функции *дифференцируемой класса* C^k или о *гладкости порядка* k. Тогда

$$d^k f = \left(dx^1 \frac{\partial}{\partial x^1} + \ldots + dx^n \frac{\partial}{\partial x^n} \right)^k f. \tag{3.42}$$

Функциями класса C^0 называются непрерывные, но недифференцируемые функции.

Мы будем считать, что в случаях «общего положения», т.е. во всех случаях кроме некоторых исключительных, встречающиеся нам функции *гладкие*, т.е. непрерывно дифференцируемые нужное число раз. Изломы и разрывы функций относятся к исключительным случаям, которые мы будем рассматривать отдельно.

Производные и дифференциалы выше второго порядка редко используются в приложениях. Значительно больше распространение имеют *бесконечно дифференцируемые функции класса* C^∞, которые можно разложить

в точке дифференцирования τ в бесконечный ряд:

$$f(\tau+\xi)=\sum_{k=0}^{\infty}\left(dx^1\frac{\partial}{\partial x^1}+\ldots+dx^n\frac{\partial}{\partial x^n}\right)^k f(\tau)= \\ =f(\tau)+\sum_{i=1}^{n}dx^i\frac{\partial f}{\partial x^i}+\sum_{i,j=1}^{n}dx^i\frac{\partial^2 f}{\partial x^i\partial x^j}+\ldots. \tag{3.43}$$

Функцию, которую можно разложить в бесконечный ряд, называют аналитической, поэтому бесконечно дифференцируемые функции также называют аналитическими.

Для функции одного переменного $y=f(x)$, бесконечно дифференцируемой в точке x_0, получаем *ряд Тейлора*:

$$f(x)=\sum_{k=0}^{\infty}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k=f(x_0)+f'(x_0)(x-x_0)+\ldots, \tag{3.44}$$

который может быть записан в форме Пеано:

$$f(x)=P_n(x)+r_n(x), \tag{3.45}$$

где $P_n(x)=\sum\limits_{k=0}^{\infty}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k$ — многочлен Тейлора,

$r_n(x)=o((x-x_0)^n)$ — остаточный член $\left(\lim\limits_{x\to x_0}r_n(x)=0\right)$.

Рассмотрим важный частный случай. Разложим функции кривой $\nu(s)=(x^1(s),\ldots,x^n(s))$ (см. стр. 31) в ряд Тейлора относительно малого, но конечного смещения ε:

$$x^i(s+\varepsilon)=x^i(s)+\varepsilon\frac{dx^i}{ds}+\frac{\varepsilon^2}{2!}\frac{d^2x^i}{ds^2}+\ldots\overset{(1.30)}{=}e^{\varepsilon\frac{dx^i}{ds}},\quad \nu(s+\varepsilon)=e^{\varepsilon\frac{d\nu}{ds}}. \tag{3.46}$$

Оператор $e^{\varepsilon\frac{d\nu}{ds}}$ называется *экспонентой от оператора* $\varepsilon\frac{d}{ds}$ или *оператором сдвига* (*конечного смещения*). Оператор $\varepsilon\frac{d}{ds}$ представляет бесконечно малое движение вдоль кривой, а его экспонента даёт *конечное смещение* ε.

Термин «дифференциал» и обозначение df для «бесконечно малой разницы» введены Г. Лейбницем (1675). Термин «производная», «вторая производная» и др., ввёл Ж. Лагранж (1797), обозначения $f'(x), f''(x)$ — он же (1770, 1779), а $\frac{df}{dx}$ — Г. Лейбниц (1675). Частные производные появились в трудах И. Ньютона, Г. Лейбница, Я. и И. Бернулли, обозначения $\frac{\partial f}{\partial x}$ ввёл А. Лежандр (1786), f'_x — Ж. Лагранж (1797, 1801), $\frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x\partial y}$ — К. Якоби (1837). Термин «градиент» ввел Дж. Максвелл (1873); ему же принадлежит обозначение grad. Определитель Гессе введён О. Гессе (1844). В отдельных случаях интегрирующим множителем пользовался И. Бернулли (1691–92). Систематическую разработку метода предпринял Л. Эйлер (1732–33). Ряд Тейлора был изобретён И. Ньютоном и опубликован его учеником Б. Тейлором (1712); ряд, сводящийся к ряду Тейлора простым преобразованием, опубликован И. Бернулли (1694). Свойства якобиана впервые изучил К. Якоби (1833, 1841).

3.3 Дифференциальные формы

Мы называли градиент вектором, но если внимательно посмотреть на уравнение (3.20), то видно, что дифференциал df скалярной функции есть скалярное произведение ковектора $f'(\tau) = (a_1, \ldots, a_n)$ и касательного вектора $\xi = (dx^1, \ldots, dx^n)$, т.е. дифференциал есть 1-форма от касательного вектора ξ, а градиент есть ковектор.

Величины $\frac{\partial f}{\partial x^i}$ $(i = 1, \ldots, n)$ градиента есть компоненты *дифференциальной* 1-*формы* $df(\xi)$ от касательного вектора ξ в точке τ.

Касательное пространство $T\mathbb{R}^n_\tau$, есть сопряжённое исходному пространству $\mathbb{R}^n$ в точке τ. Координаты $dx^1, \ldots, dx^n$ касательного пространства линейно независимы, образуют дуальный базис, сопряжённый базису $e_1, \ldots, e_n$ исходного пространства, и имеют название *базисные дифференциальные* 1-*формы* векторного пространства 1-форм.

По аналогии с выражениями (2.23–2.26) для алгебраических 1-форм запишем выражения для дифференциальных 1-форм:

$$dx^i(e_j) = \delta^i_j,\ dx^i(\xi) = dx^i,\ df(e_i) = \frac{\partial f}{\partial x^i},\ df(\xi) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\, dx^i. \tag{3.47}$$

Выражения для алгебраических и дифференциальных 1-форм похожи с учётом того, что базис, координаты и компоненты алгебраических 1-форм заданы на всём пространстве, а для дифференциальных 1-форм являются функциями рассматриваемой точки τ.

Дифференциал скалярной функции df это не единственная дифференциальная 1-форма, которая может быть задана на базисе $dx^1, \ldots, dx^n$.

Всякая дифференциальная 1-форма ω в пространстве $\mathbb{R}^n$ (скалярная функция от касательного вектора ξ) с выбранной системой координат $x^1, \ldots, x^n$ однозначно представляется в виде:

$$\omega(\xi) = a_1(x^1, \ldots, x^n)\, dx^1 + \ldots + a_n(x^1, \ldots, x^n)\, dx^n \overset{(2.17)}{=} a_i\, dx^i. \tag{3.48}$$

где $a_i(x^1, \ldots, x^n)$ — гладкие функции.

Обратно этому, любая гладкая вектор-функция в $\mathbb{R}^n$ с компонентами $\mathbf{a} = (a_1, \ldots, a_n)$ и касательный вектор $\xi = (dx^1, \ldots, dx^n)$ аналогично выражению (2.27) дают дифференциальную 1-форму:

$$\omega^1_{\mathbf{a}}(\xi) = (\mathbf{a}, \xi). \tag{3.49}$$

1-форма (3.48, 3.49) в отличие от выражения (3.2) не обязательно является дифференциалом какой-либо функции и называется *дифференциальным выражением Пфаффа* (*пфаффовой формой*).

Дифференциальная 1-*форма в точке дифференцирования* τ есть гладкое отображение касательного пространства $T\mathbb{R}^n_\tau$ в прямую: $\omega_\tau \colon T\mathbb{R}^n_\tau \to \mathbb{R}$. *Дифференциальная* 1-*форма на пространстве* $\mathbb{R}^n$ есть гладкое отображение касательного расслоения $T\mathbb{R}^n$ в прямую: $\omega \colon T\mathbb{R}^n \to \mathbb{R}$, линейное на

каждом касательном пространстве $T\mathbb{R}^n_\tau$. Множество всех дифференциальных 1-форм образует линейное пространство $\Lambda^1(\mathbb{R}^n)$.

В то время как векторы (касательный вектор ξ и векторное поле $\mathbf{a}$), ковекторы (grad), скалярные функции (метрика (3.39) $ds^2 = \xi \cdot \xi$ и 1-форма (3.49) $\omega^1_{\mathbf{a}}$) есть *инварианты*, их компоненты — базисные формы dx^i и компоненты форм a_i *не являются инвариантами* и имеют в каждой системе координат свой специальный вид.

Пример. Рассмотрим трёхмерное евклидово пространство декартовых (x, y, z) и криволинейных координат (u, v, w).

1. Инварианты не зависят от системы координат:
   $\xi = d\mathbf{r}$ — в классической записи векторного анализа касательный вектор ξ трёхмерного евклидова (декартова или криволинейного) касательного пространства называют *элементом длины*;
   $ds^2 = d\mathbf{r} \cdot d\mathbf{r}$;
   $df = \operatorname{grad} f \cdot d\mathbf{r}$;
   $\mathbf{a} = e_x a_x + e_y a_y + e_z a_z = e_u a_u + e_v a_v + e_w a_w$;
   $\omega^1_{\mathbf{a}} = \mathbf{a} \cdot d\mathbf{r}$.

2. Для декартовой системы координат:
   $dx(e_x) = dy(e_y) = dz(e_z) = 1, \quad dx(e_y) = dy(e_z) = 0$ и т. д.;
   $d\mathbf{r} = e_x\, dx + e_y\, dy + e_z\, dz$;
   $ds^2 = dx^2 + dy^2 + dz^2$;
   $df = \frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy + \frac{\partial f}{\partial z}\, dz$;
   $\omega^1_{\mathbf{a}} = a_x\, dx + a_y\, dy + a_z\, dz$.

3. Для криволинейной системы координат нам понадобятся коэффициенты Ламе L_u, L_v, L_w, введенные Г. Ламе (1859):
   $L_u^2 = \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2$ и т. д.
   Тогда:
   $du(e_u) = \frac{1}{L_u}$, $dv(e_v) = \frac{1}{L_v}$, $dw(e_w) = \frac{1}{L_w}$;
   $d\mathbf{r} = e_u L_u du + e_v L_v dv + e_w L_w dw$;
   $ds^2 = L_u^2 du^2 + L_v^2 dv^2 + L_w^2 dw^2$;
   $df = \frac{\partial f}{\partial u}\, du + \frac{\partial f}{\partial v}\, dv + \frac{\partial f}{\partial w}\, dw$;
   $\omega^1_{\mathbf{a}} = a_u L_u\, du + a_v L_v\, dv + a_w L_w\, dw$;
   $\operatorname{grad} f = \frac{1}{L_u}\frac{\partial f}{\partial u} e_u + \frac{1}{L_v}\frac{\partial f}{\partial v} e_v + \frac{1}{L_w}\frac{\partial f}{\partial w} e_w$.

4. Для цилиндрических координат:
   $dr(e_r) = dz(e_z) = 1$, $d\varphi(e_\varphi) = \frac{1}{r}$;
   $L_r = L_z = 1$, $L_\varphi = r$;
   $ds^2 = dr^2 + r^2 d\varphi^2 + dz^2$;
   $\omega^1_{\mathbf{a}} = a_r dr + r a_\varphi d\varphi + a_z dz$;
   $\operatorname{grad} f = \frac{\partial f}{\partial r} e_r + \frac{1}{r}\frac{\partial f}{\partial \varphi} e_\varphi + \frac{\partial f}{\partial z} e_z$.

5. Для сферических координат:
   $d\rho(e_\rho) = 1$, $d\varphi(e_\varphi) = \frac{1}{\rho \sin\theta}$, $d\theta(e_\theta) = \frac{1}{\rho}$;
   $L_\rho = 1$, $L_\varphi = \rho \sin\theta$, $L_\theta = \rho$;

$ds^2 = d\rho^2 + \rho^2 \sin^2\theta d\varphi^2 + \rho^2 d\theta^2$;
$\omega^1_{\mathbf{a}} = a_\rho d\rho + \rho\cos\theta a_\varphi d\varphi + \rho a_\theta d\theta$;
$\operatorname{grad} f = \frac{\partial f}{\partial \rho} e_\rho + \frac{1}{\rho\sin\theta}\frac{\partial f}{\partial \varphi} e_\varphi + \frac{1}{\rho}\frac{\partial f}{\partial \theta} e_\theta$.

6. Для параболических координат:
$du(e_u) = dv(e_v) = \frac{1}{2\sqrt{u^2+v^2}}$, $dz(e_z) = 1$;
$L_u = L_v = 2\sqrt{u^2+v^2}$, $L_z = 1$;
$\operatorname{grad} f = \frac{1}{2\sqrt{u^2+v^2}} \left(\frac{\partial f}{\partial u} e_u + \frac{\partial f}{\partial v} e_v\right) + \frac{\partial f}{\partial z} e_z$.

Аналогично алгебраическим формам, в касательном пространстве можно задать функции не только на касательных векторах, но и на бесконечно малых параллелограммах, а также на бесконечно малых многомерных параллелепипедах.

Бесконечно малые параллелограммы и параллелепипеды образуются при внешнем произведении бесконечно малых векторов касательного пространства.

Пример. Бесконечно малый параллелограмм, построенный на двух касательных векторах ξ, η трёхмерного евклидова пространства, записывают как векторное произведение и называют *элементом площади*:

$$\begin{aligned} d\sigma = \xi \times \eta &\overset{(2.106)}{=} e_x\, dy \wedge dz + e_y\, dz \wedge dx + e_z\, dx \wedge dy \\ &= e_u L_v L_w dv \wedge dw + e_v L_u L_w dw \wedge du + e_w L_u L_v du \wedge dv, \end{aligned} \tag{3.50}$$

а бесконечно малый параллелепипед, построенный на трёх касательных векторах ξ, η, ζ, записывают как кососкалярное или смешанное произведение и называют *элементом объёма*:

$$dV = [\xi, \eta, \zeta] \overset{(2.109)}{=} dx \wedge dy \wedge dz = L_u L_v L_w du \wedge dv \wedge dw. \tag{3.51}$$

Элементы площади $d\sigma$ и объёма dV также есть инварианты, т.е. их величина не зависит от системы координат, однако вид их записи (компоненты) зависит от выбранной системы координат.

Для сферических координат:

$$d\sigma = e_\rho \rho^2 \sin\theta\, d\varphi \wedge d\theta + e_\varphi \rho\, d\theta \wedge d\rho + e_\theta \rho \sin\theta\, d\rho \wedge d\varphi; \tag{3.52}$$

$$dV = \rho^2 \sin\theta\, d\rho \wedge d\varphi \wedge d\theta. \tag{3.53}$$

Дифференциальная k-форма в точке дифференцирования τ есть внешняя k-форма ω^k_τ на n-мерном касательном пространстве $T\mathbb{R}^n_\tau$, т.е. k-линейная кососимметричная скалярная функция от k касательных векторов $\xi_1, \ldots, \xi_k$ (от *k-мерного бесконечно малого параллелепипеда*, построенного на этих векторах в точке дифференцирования τ). Если такая форма задана в каждой точке пространства $\mathbb{R}^n$ и если она дифференцируема, то говорят, что задана дифференциальная k-форма ω^k. Всякая дифференциальная k-форма в пространстве $\mathbb{R}^n$ с выбранной системой координат $x^1, \ldots, x^n$ однозначно представляется в виде:

$$\omega^k = \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} a_{i_1 \ldots i_k}(x^1, \ldots, x^n)\, dx^{i_1} \wedge \ldots \wedge dx^{i_k}, \tag{3.54}$$

где $a_{i_1 \ldots i_k}(x^1, \ldots, x^n)$ — гладкие функции.

Базис $dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ $(1 \leqslant i_1 < \ldots < i_k \leqslant n)$ линейного пространства $\Lambda^k(\mathbb{R}^n)$ дифференциальных k-форм имеет размерность C_n^k. В частности, при $k = n$ размерность пространства $C_n^n = 1$. Тогда

$$\omega^n(\xi_1, \ldots, \xi_n) = a(x^1, \ldots, x^n)\, dx^1 \wedge \ldots \wedge dx^n. \tag{3.55}$$

Пример 1. Каждая дифференциальная 2-форма в трёхмерном евклидовом пространстве имеет три компоненты $a_{i_1 i_2}$, поэтому ей можно сопоставить векторное поле $\mathbf{a} = (a_x, a_y, a_z)$, которое в этом случае используется как нижний индекс в обозначении 2-формы:

$$\omega^2_{\mathbf{a}}(\xi, \eta) = (\mathbf{a}, \xi, \eta) = \mathbf{a} \cdot d\sigma. \tag{3.56}$$

Соответствие (3.56) между векторным полем и формой инвариантно, однако формулы перехода от полей к формам и обратно имеют в каждой системе координат свой специальный вид:

1. Для декартовой системы координат
   $\omega^2_{\mathbf{a}} = a_x\, dy \wedge dz + a_y\, dz \wedge dx + a_z\, dx \wedge dy.$

2. В общем, для криволинейной системы координат
   $\omega^2_{\mathbf{a}} = a_u L_v L_w\, dv \wedge dw + a_v L_u L_w\, dw \wedge du + a_w L_u L_v\, du \wedge dv.$

3. Для цилиндрических координат
   $\omega^2_{\mathbf{a}} = a_r r\, d\varphi \wedge dz + a_\varphi\, dz \wedge dr + a_z r\, dr \wedge d\varphi.$

4. Для сферических координат
   $\omega^2_{\mathbf{a}} = a_\rho \rho^2 \cos\theta\, d\varphi \wedge \theta + a_\varphi \rho\, d\theta \wedge d\rho + a_\theta \rho \cos\theta\, d\rho \wedge d\varphi.$

Пример 2. Каждая дифференциальная 3-форма в трёхмерном евклидовом пространстве имеет одну компоненту a:

$$\omega^3(\xi, \eta, \zeta) = a\, dV. \tag{3.57}$$

Как и в предыдущем примере, формулы перехода от полей к формам и обратно зависят от системы координат:

1. Для декартовой системы координат
   в трёхмерном пространстве
   $\omega^3 = a(x, y, z)\, dx \wedge dy \wedge dz.$

2. Для трёхмерной криволинейной системы координат
   $\omega^3 = a(u, v, w) L_u L_v L_w\, du \wedge dv \wedge dw.$

Дифференциальная внешняя k-форма ω^k как всякая гладкая функция имеет дифференциал

$$d\omega^k \overset{(3.20)}{=} \frac{\partial \omega^k}{\partial x^i}\, dx^i = \omega^{k+1}. \tag{3.58}$$

При записи дифференциала k-формы в развёрнутом виде мы получаем *операцию внешнего дифференцирования*:

$$\begin{aligned} d\omega^k &\overset{(3.54)}{=} d\left(\sum_{1\leqslant i_1<\ldots<i_k\leqslant n} a_{i_1\ldots i_k}\, dx^{i_1}\wedge\ldots\wedge dx^{i_k}\right) = \\ &= \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} da_{i_1\ldots i_k}\wedge dx^{i_1}\wedge\ldots\wedge dx^{i_k} = \\ &\overset{(3.20)}{=} \sum_{1\leqslant i_1<\ldots<i_k\leqslant n}\ \sum_{j=1}^{n} \frac{\partial a_{i_1\ldots i_k}}{\partial x^j}\, dx^j\wedge dx^{i_1}\wedge\ldots\wedge dx^{i_k}. \end{aligned} \tag{3.59}$$

Свойства внешнего дифференцирования:

$$d(\omega_1+\omega_2) = d\omega_1 + d\omega_2, \tag{3.60}$$

$$d(\omega_1^k\wedge\omega_2^l) = d\omega_1^k\wedge\omega_2^l + (-1)^k\omega_1^k\wedge d\omega_2^l, \tag{3.61}$$

$$d(d\omega) = 0 \quad \text{или кратко } d^{\,2} = 0. \tag{3.62}$$

Докажем свойство (3.62):

$$d\omega = \sum_{1\leqslant i_1<\ldots<i_k\leqslant n}\ \sum_{j=1}^{n} \frac{\partial a_{i_1\ldots i_k}}{\partial x^j}\, dx^j\wedge dx^{i_1}\wedge\ldots\wedge dx^{i_k},$$

тогда

$$d(d\omega) = \sum_{1\leqslant i_1<\ldots<i_k\leqslant n}\ \sum_{j=1}^{n}\sum_{l=1}^{n} \frac{\partial a_{i_1\ldots i_k}}{\partial x^j\partial x^l}\, dx^l\wedge dx^j\wedge dx^{i_1}\wedge\ldots\wedge dx^{i_k}.$$

Но в этой сумме члены каждой пары

$$\frac{\partial a_{i_1\ldots i_k}}{\partial x^j\partial x^l}\, dx^l\wedge dx^j\wedge dx^{i_1}\wedge\ldots\wedge dx^{i_k} \text{ и } \frac{\partial a_{i_1\ldots i_k}}{\partial x^l\partial x^j}\, dx^j\wedge dx^l\wedge dx^{i_1}\wedge\ldots\wedge dx^{i_k}$$

взаимно уничтожаются, и в результате получаем $d(d\omega) = 0$.

Пример 1. Скалярная функция f рассматривается как *форма нулевой степени*, дифференцирование которой даёт 1-форму df:

$$df(\xi) \overset{(3.16)}{=} \operatorname{grad} f\cdot\xi = (\operatorname{grad} f, \xi) = \omega^1_{\operatorname{grad} f}(\xi).$$

Пример 2. Внешнее дифференцирование 1-формы даёт 2-форму:

$$d\omega^1(\xi) = \omega^2(\xi,\eta) = \sum_{i,j=1}^{n}\left(\frac{\partial a_j}{\partial x^i} - \frac{\partial a_i}{\partial x^j}\right) dx^i\wedge dx^j(\xi,\eta), \tag{3.63}$$

компоненты которой $\omega_{ij} = \frac{\partial a_j}{\partial x^i} - \frac{\partial a_i}{\partial x^j} = -\omega_{ji}$ можно записать в виде кососимметрической матрицы

$$\left\| \begin{array}{ccc} 0 & \dots & \omega_{1n} \\ \cdots & 0 & \cdots \\ -\omega_{1n} & \dots & 0 \end{array} \right\|. \tag{3.64}$$

Пример 3. Внешнее дифференцирование 1-формы в трёхмерном евклидовом пространстве даёт 2-форму, состоящую из трёх компонент, которой можно сопоставить векторное поле ротора $\operatorname{rot}\mathbf{a}$ — дифференциального оператора, качественно рассмотренного на рис. 3.1:

$$d\omega^1_{\mathbf{a}}(\xi) = \omega^2_{\operatorname{rot}\mathbf{a}}(\xi,\eta) = (\operatorname{rot}\mathbf{a},\xi,\eta) = \operatorname{rot}\mathbf{a}\cdot d\sigma. \tag{3.65}$$

1. Для декартовой системы координат
$\operatorname{rot}\mathbf{a} = e_x\left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}\right) + e_y\left(\frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x}\right) + e_z\left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right)$.

2. Для криволинейной системы координат
$\operatorname{rot}\mathbf{a} = \frac{e_u}{L_vL_w}\left[\frac{\partial}{\partial v}(a_wL_w) - \frac{\partial}{\partial w}(a_vL_v)\right] + \frac{e_v}{L_uL_w}\left[\frac{\partial}{\partial w}(a_uL_u) - \frac{\partial}{\partial u}(a_wL_w)\right] + \frac{e_w}{L_uL_v}\left[\frac{\partial}{\partial u}(a_vL_v) - \frac{\partial}{\partial v}(a_uL_u)\right]$.

3. Для цилиндрических координат
$\operatorname{rot}\mathbf{a} = \frac{e_r}{r}\left[\frac{\partial a_z}{\partial\varphi} - \frac{\partial(ra_\varphi)}{\partial z}\right] + e_\varphi\left[\frac{\partial a_r}{\partial z} - \frac{\partial a_z}{\partial r}\right] + \frac{e_z}{r}\left[\frac{\partial(ra_\varphi)}{\partial r} - \frac{\partial a_r}{\partial\varphi}\right]$.

4. Для сферических координат
$\operatorname{rot}\mathbf{a} = \frac{e_\rho}{\rho\sin\theta}\left[\frac{\partial a_\theta}{\partial\varphi} - \frac{\partial(a_\varphi\sin\theta)}{\partial\theta}\right] + \frac{e_\varphi}{\rho}\left[\frac{\partial a_\rho}{\partial\theta} - \frac{\partial(\rho a_\theta)}{\partial\rho}\right] + \frac{e_\theta}{\rho}\left[\frac{\partial(\rho a_\varphi)}{\partial\rho} - \frac{1}{\sin\theta}\frac{\partial a_\rho}{\partial\varphi}\right]$.

Пример 4. Как упоминалось на стр. 93, для каждой кососимметричной 2-формы в нечётномерном пространстве существует по крайней мере один неравный нулю вектор ζ, для которого $\omega^2(\zeta,\eta) = 0$ при любых η. Такая 2-форма называется вырожденной, а вектор $\zeta \neq 0$ называется нулевым вектором формы ω^2. Если ω^2 — неособая форма в нечётномерном пространстве, то нулевые векторы $\zeta \neq 0$ формы ω^2 лежат на одной прямой. Эта прямая инвариантно связана с формой ω^2.

Пусть теперь M^{2k+1} — нечётномерное многообразие, $\mathbf{a}$ — векторное поле, $\omega^1_{\mathbf{a}}(\xi) \overset{(3.49)}{=} (\mathbf{a},\xi)$ — 1-форма на M. Тогда запишем

$$d\omega^1_{\mathbf{a}}(\xi) \overset{(3.63)}{=} \omega^2_\zeta(\xi,\eta) = [\zeta,\xi,\eta]. \tag{3.66}$$

Из формулы (3.66) видно, что в каждой точке $\tau \in M$ существует направление $\zeta = (b^1,\dots,b^n)$ для которого

$$d\omega^1_{\mathbf{a}}(\zeta) = [\zeta,\zeta,\eta] = 0. \tag{3.67}$$

Обязательное наличие такого ненулевого вектора ζ отражает существование оси у всякого вращения нечётномерного пространства. Пусть далее

2-форма неособа, тогда направление ζ определено однозначно, называется «*направлением ротора*» формы ω^1 в точке τ и инвариантно связано с формой ω^2.

Таким образом, хотя для произвольного нечётномерного пространства невозможно вычислить ротор в явном виде, но возможно вычислить «направление ротора».

Векторные линии поля ζ называются *линиями ротора* или *характеристиками* формы ω^1.

Переход от 2-формы $\omega^2_{\operatorname{rot}\mathbf{a}}$ к полю ротора $\zeta = \operatorname{rot}\mathbf{a}$ не инвариантная операция, она зависит от евклидовой структуры E^{2k+1} (выбранной системы координат) касательного пространства TM_τ. Однако направление ротора ζ инвариантно связано с формой ω^2, а значит и с формой ω^1.

Пример 5. Внешнее дифференцирование 2-формы в трёхмерном евклидовом пространстве даёт 3-форму, единственная компонента (3.57) которой представляет собой дивергенцию векторного поля $\operatorname{div}\mathbf{a}$ — дифференциальный оператор, качественно рассмотренный на рис. 3.1:

$$d\omega^2_{\mathbf{a}}(\xi,\eta) = \omega^3(\xi,\eta,\zeta) = \operatorname{div}\mathbf{a}\, dV. \tag{3.68}$$

1. Для декартовой системы координат
   $\operatorname{div}\mathbf{a} = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}$.
2. Для криволинейной системы координат
   $\operatorname{div}\mathbf{a} = \frac{1}{L_u L_v L_w}\left[\frac{\partial}{\partial u}(a_u L_v L_w) + \frac{\partial}{\partial v}(a_v L_u L_w) + \frac{\partial}{\partial w}(a_w L_u L_v)\right]$.
3. Для цилиндрических координат
   $\operatorname{div}\mathbf{a} = \frac{1}{r}\left[\frac{\partial(r a_r)}{\partial r} + \frac{\partial a_\varphi}{\partial \varphi}\right] + \frac{\partial a_z}{\partial z}$.
4. Для сферических координат
   $\operatorname{div}\mathbf{a} = \frac{1}{\rho^2 \sin\theta}\left[\frac{\partial(a_\rho \rho^2 \sin\theta)}{\partial \rho} + \frac{\partial(\rho a_\varphi)}{\partial \varphi} + \frac{\partial(a_\theta \rho \sin\theta)}{\partial \theta}\right]$.

Дивергенция вычисляет источники, а ротор — вихри векторного поля.

Ненулевую дивергенцию можно понимать как количество векторных линий, которые возникают или исчезают в рассматриваемой точке пространства. Если же $\operatorname{div}\mathbf{a} = 0$, то в данной области не исчезает и не возникает никаких новых линий векторного поля. Как мы знаем (см. рис. 3.1), такое векторное поле называется соленоидальным (вихревым).

Отличие ротора (вихря) от нуля говорит о завихрённости потока. Если же $\operatorname{rot}\mathbf{a} = 0$, то вихри отсутствуют. Как мы знаем (см. рис. 3.1), такое векторное поле называется потенциальным (слоистым).

Векторные линии потенциального поля возникают из источников и не имеют вихрей, а векторные линии соленоидального поля представляют собой замкнутые вокруг вихревых точек кривые и не имеют источников.

В безвихревом гравитационном поле сорванный с ветки листок падает прямо на землю, не кружась в вихревом поле ветра.

Дивергенция является суммой диагональных элементов (следом), а ротор — комбинацией недиагональных элементов матрицы Якоби.

Если градиент называют векторной пространственной производной скалярного поля, матрицу Якоби — тензорной производной векторного поля, то дивергенцию и ротор называют, соответственно, *скалярной и векторной пространственными производными векторного поля.*

Если же рассматривать дивергенцию и ротор как преобразование векторного поля, то они имеют собственные матрицы.

Потенциальным полям с нулевым ротором отвечают диагональные матрицы с действительными собственными значениями.

Вихревым полям с нулевой дивергенцией отвечают кососимметрические матрицы с нулевой диагональю и чисто мнимыми собственными значениями. В нечётномерном пространстве одно собственное значение обязательно равно нулю, что указывает на наличие оси вращения пространства, задаваемой соответствующим собственным вектором.

Теорема Гельмгольца. Если дивергенция и ротор векторного поля $\mathbf{c}$ определены в каждой точке конечной открытой области пространства $U \subset \mathbb{R}^3$, то всюду в U поле $\mathbf{c}$ может быть представлено в виде суммы потенциального $\mathbf{d}$ ($\operatorname{rot}\mathbf{d} = 0$) и соленоидального $\mathbf{b}$ ($\operatorname{div}\mathbf{b} = 0$) полей:

$$\mathbf{c} = \mathbf{d} + \mathbf{b}. \tag{3.69}$$

Другими словами, произвольное дифференцируемое векторное поле всегда есть сумма потенциального и соленоидального полей.

Примеры вычисления ротора и дивергенции.

Пример 1.

$$\operatorname{rot}\dot{\mathbf{r}} \overset{(3.113)}{=} \operatorname{rot}(\omega \times \mathbf{r}) \overset{(3.168)}{=} 2\,\omega. \tag{3.70}$$

Направление вектора угловой скорости вращения ω точки пространства совпадает с направлением вихря (векторного поля скоростей) в этой точке.

Пример 2.

$$\operatorname{rot}\mathbf{r} = 0; \tag{3.71}$$

$$\operatorname{div}\mathbf{r} = 3. \tag{3.72}$$

Поле радиус-вектора имеет исток в начале координат и не имеет вихрей.

При втором внешнем дифференцировании форм в трёхмерном евклидовом пространстве получаем два правила для дифференциальных операторов второго порядка:

$$d(df) = d(\omega^1_{\operatorname{grad} f}) = \operatorname{rot}(\operatorname{grad} f) \cdot d\sigma = 0 \quad \text{или} \quad \operatorname{rot}(\operatorname{grad} f) = 0; \tag{3.73}$$

$$d(d\omega^1_{\mathbf{a}}) = d(\omega^2_{\operatorname{rot}\mathbf{a}}) = \operatorname{div}(\operatorname{rot}\mathbf{a})\, dV = 0 \quad \text{или} \quad \operatorname{div}(\operatorname{rot}\mathbf{a}) = 0. \tag{3.74}$$

Эти два правила можно интерпретировать следующим образом:

Если векторное поле $\mathbf{d}$ есть градиент скалярного поля f

$$\mathbf{d} = \operatorname{grad} f, \tag{3.75}$$

то, так как «вихрь градиента равен нулю» $\operatorname{rot}(\operatorname{grad} f) = 0$, получаем:

$$\operatorname{rot} \mathbf{d} = 0. \tag{3.76}$$

Следовательно, $\mathbf{d}$ есть потенциальное поле, а скалярное поле f называется *скалярный потенциал* (от лат. potentia — сила) поля $\mathbf{d}$.

Если векторное поле $\mathbf{b}$ есть ротор векторного поля $\mathbf{a}$

$$\mathbf{b} = \operatorname{rot} \mathbf{a}, \tag{3.77}$$

то, так как «дивергенция вихря равна нулю» $\operatorname{div}(\operatorname{rot} \mathbf{a}) = 0$, получаем:

$$\operatorname{div} \mathbf{b} = 0. \tag{3.78}$$

Следовательно, $\mathbf{b}$ есть соленоидальное поле, а векторное поле $\mathbf{a}$ называется *векторный потенциал* поля $\mathbf{b}$.

Можно записать ещё один дифференциальный оператор второго порядка, который представляет собой скалярное произведение оператора набла на градиент и называется *оператор Лапласа* или *лапласиан*:

$$\Delta f = \operatorname{div}(\operatorname{grad} f) = \frac{\partial^2 f}{\partial {x^1}^2} + \ldots + \frac{\partial^2 f}{\partial {x^n}^2}. \tag{3.79}$$

В отличие от операторов $\operatorname{rot}(\operatorname{grad} f) = 0$ и $\operatorname{div}(\operatorname{rot} \mathbf{a}) = 0$ оператор Лапласа не обязательно равен нулю.

Выражение $\Delta f = \rho$ называется *уравнение Пуассона*.

Функция $f(x^1, \ldots, x^n)$, удовлетворяющая уравнению Пуассона есть *скалярный потенциал*, а функция $\rho(x^1, \ldots, x^n)$ называется *плотность распределения потенциала*.

Если в какой-либо части пространства $\rho = 0$, то выражение $\Delta f = 0$, называется *уравнение Лапласа*, а функция $f(x^1, \ldots, x^n)$, удовлетворяющая уравнению Лапласа, называется *гармоническая функция*.

При известном вихревом поле векторный потенциал определён только с точностью до градиента некого скалярного поля f, удовлетворяющего уравнению Лапласа $\operatorname{div}(\operatorname{grad} f) = \Delta f$, так как $\operatorname{rot}(\operatorname{grad} f) = 0$.

При известном потенциальном поле скалярный потенциал определён только с точностью до постоянной величины c, так как $\operatorname{grad} c = 0$.

Примеры вычисления лапласианов.

Пример 1. Рассмотрим трёхмерное евклидово пространство:

1. Для декартовой системы координат
   $\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$.
2. В общем, для криволинейной системы координат
   $\Delta f = \frac{1}{L_u L_v L_w} \left[\frac{\partial}{\partial u} \left(\frac{L_v L_w}{L_u} \frac{\partial f}{\partial u} \right) + \frac{\partial}{\partial v} \left(\frac{L_u L_w}{L_v} \frac{\partial f}{\partial v} \right) + \frac{\partial}{\partial w} \left(\frac{L_u L_v}{L_w} \frac{\partial f}{\partial w} \right) \right]$.
3. Для цилиндрических координат
   $\Delta f = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r} \frac{\partial f}{\partial r} + \frac{1}{r^2} \frac{\partial^2 f}{\partial \varphi^2} + \frac{\partial^2 f}{\partial z^2}$.

4. Для сферических координат
$\Delta f = \frac{\partial^2 f}{\partial \rho^2} + \frac{2}{\rho}\frac{\partial f}{\partial \rho} + \left[\frac{1}{\rho^2}\frac{\partial^2}{\partial \theta^2} + \frac{\operatorname{ctg}\theta}{\rho^2}\frac{\partial}{\partial \theta} + \frac{1}{\rho^2 \sin^2\theta}\frac{\partial^2}{\partial \varphi^2}\right] f.$

5. Для параболических координат
$\Delta f = \frac{1}{4(u^2+v^2)}\left[\frac{\partial^2 f}{\partial u^2} + \frac{\partial^2 f}{\partial v^2}\right] + \frac{\partial^2 f}{\partial z^2}.$

Пример 2. Пусть $f = u + iv$ — комплексная функция, аналитическая в области $A \subset \mathbb{C}$. Тогда $u(x,y)$ и $v(x,y)$ связаны условиями Коши-Римана (3.33). Возьмем производные по x от первого равенства (3.33) и по y от второго равенства (3.33), получим:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x \partial y}; \quad \frac{\partial^2 v}{\partial x \partial y} = -\frac{\partial u}{\partial y^2}, \quad \text{откуда}$$

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \text{ аналогично получаем } \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0. \tag{3.80}$$

Выражения (3.80) есть уравнения Лапласа, а удовлетворяющие этим уравнениям (решения уравнений) — функции $u(x,y)$ и $v(x,y)$ есть гармонические функции.

Итак, в качестве вещественной и мнимой частей аналитической функции комплексного переменного, могут быть взяты не какие угодно, а только гармонические функции $u(x,y)$ и $v(x,y)$. Кроме того, эти функции должны быть связаны друг с другом условиями Коши–Римана и называются поэтому *сопряженными гармоническими функциями*.

Пример 3. Согласно (3.193) $\frac{1}{r}$ — гармоническая функция.

Пример 4. Электростатический потенциал в пространстве, свободном от зарядов, удовлетворяет уравнению Лапласа, а при наличии заряда — уравнению Пуассона, при этом плотность распределения потенциала равна плотности электрических зарядов.

Если набор операторов $\frac{\partial}{\partial x^i}$ $(i = 1, \ldots, n)$ представить как векторно-дифференциальный оператор ∇, который называется *набла* (от греческого $\upsilon\alpha\beta\lambda\alpha$ — арфа, форма которой напоминает значок ∇) и имеет, при своем применении, как черты вектора, так и черты оператора дифференцирования, то дифференциальные операторы $\operatorname{grad} f$, Δf, $\operatorname{div} \mathbf{a}$ и $\operatorname{rot} \mathbf{a}$ можно записать следующим образом:

Под градиентом подразумевается произведение ∇ на скаляр f:

$$\operatorname{grad} f = \nabla f. \tag{3.81}$$

Под лапласианом понимается квадрат ∇:

$$\Delta f = \nabla \cdot (\nabla f) = \nabla^2 f. \tag{3.82}$$

Под дивергенцией в $\mathbb{R}^n$ понимается скалярное произведение ∇ на векторное поле $\mathbf{a} = (a_1, \ldots, a_n)$:

$$\operatorname{div} \mathbf{a} = \nabla \cdot \mathbf{a} = \frac{\partial a_1}{\partial x^1} + \ldots + \frac{\partial a_n}{\partial x^n}. \tag{3.83}$$

В трёхмерном евклидовом пространстве под ротором понимается векторное произведение ∇ на векторное поле $\mathbf{a} = (a_x, a_y, a_z)$:

$$\operatorname{rot} \mathbf{a} = \nabla \times \mathbf{a}. \tag{3.84}$$

Форма ω называется *замкнутой* если имеет нулевой дифференциал (нулевой градиент) некоторой формы $d\omega = 0$ и *точной* если сама является дифференциалом (градиентом) $\omega = d\eta$ для некоторой формы η.

Точная форма от формы нулевой степени или от заданной на многообразии функции f есть *полный дифференциал* df этой функции.

Всякая точная форма $\omega = d\eta$ *замкнута*, так как $d\omega = d(d\eta) = 0$. Но если форма ω замкнута $d\omega = 0$, то не обязательно точна, так как она не обязательно есть дифференциал какой-либо формы, т.е. форма η может и не существовать.

Теорема. Если для дифференциала

$$df = a_1 dx^1 + \ldots + a_n dx^n \tag{3.85}$$

выполняются условия

$$\frac{\partial a_i}{\partial x^j} = \frac{\partial a_j}{\partial x^i} \quad \text{для всех } i, j = 1, \ldots, n, \tag{3.86}$$

то df есть *полный дифференциал* функции f. Тогда

$$a_i = \frac{\partial f}{\partial x^i}, \qquad \frac{\partial a_i}{\partial x^j} = \frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial^2 f}{\partial x^j \partial x^i} = \frac{\partial a_j}{\partial x^i}. \tag{3.87}$$

Доказательство

$$d^2 f \overset{(3.63)}{=} \sum_{i,j=1}^{n} \left(\frac{\partial a_j}{\partial x^i} - \frac{\partial a_i}{\partial x^j} \right) dx^i \wedge dx^j \overset{(3.86)}{=} 0,$$

что и требовалось доказать.

Примеры полных дифференциалов.

Пример 1. Для функции двух переменных условие (3.86) имеет вид $df = a_x\, dx + a_y\, dy$ — полный дифференциал, если

$$\left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y} \right) = 0. \tag{3.88}$$

Пример 2. Для функции трёх переменных условие (3.86) имеет вид $df = a_x\, dx + a_y\, dy + a_z\, dz$ — полный дифференциал, если

$$\operatorname{rot} \mathbf{a} = 0, \quad \text{где} \quad \mathbf{a} = (a_x, a_y, a_z). \tag{3.89}$$

Если условие (3.86) не выполняется, то иногда удаётся найти функцию $R(x^1, \ldots, x^n)$, которая превратит дифференциал df в полный дифференциал dU некоторой функции U:

$$R(a_1 dx^1 + \ldots + a_n dx^n) = dU. \tag{3.90}$$

Такая функция R называется *интегрирующий множитель*.

Функция, дифференциал которой не является полным, но имеет интегрирующий множитель, называется *голономной*. Функция, для дифференциала которой невозможно найти интегрирующий множитель, называется *неголономной*. Голономную функцию f можно восстановить по её дифференциалу df (проинтегрировать имеющуюся производную функции), а неголономную — нельзя.

Доказано, что для дифференциала функции двух переменных интегрирующий множитель всегда можно найти, поэтому функция двух переменных всегда голономна.

Функция трёх переменных голономна в том случае, если

$$\mathbf{a} \cdot \operatorname{rot} \mathbf{a} = 0. \tag{3.91}$$

Для большего количества переменных интегрирующий множитель можно найти не всегда.

Дифференциальные формы введены Э. Картаном в начале 20-го века.

Уравнение Пуассона впервые было изучено С. Пуассоном (1812). Уравнение Лапласа изучено раньше, чем уравнение Пуассона, и встречается у Л. Эйлера (1761) и Ж. Д'Аламбера (1761) в связи с задачами гидромеханики. Однако широкую известность это уравнение получило после появления работ П. Лапласа (1782, 1799) по теории гравитационного потенциала и небесной механике. Обозначение Δ ввёл Р. Мёрфи (1833). Идея потенциала принадлежит Ж. Лагранжу (1775) и П. Лапласу (1782). Теория потенциала создана независимо Дж. Грином (1828) и К. Гауссом (1840). Термин «потенциальная функция» предложил Дж. Грин, термин «потенциал» — К. Гаусс.

3.4 Многообразие

Теперь у нас есть все основные определения, чтобы сформулировать обобщение для дифференцируемых подмножеств пространства $\mathbb{R}^n$. Эти подмножества или поверхности называют *многообразия*.

Понятие дифференцируемого или гладкого многообразия играет в геометрии и в анализе столь же фундаментальную роль, как в алгебре понятие группы и линейного пространства.

Приведём примеры возникновения многообразий в задачах:

1. Две точки на плоскости $\mathbb{R}^2$ образуют четырехмерное *конфигурационное пространство* $\mathbb{R}^4 = \mathbb{R}^2 \times \mathbb{R}^2$. *Конфигурация* (от позднелат.

configuratio — расположение) — конечное множество точек, прямых, плоскостей, связанных между собой отношениями принадлежности (инцидентности). Размерность конфигурационного пространства называется *числом степеней свободы*. Если точки жестко связаны между собой и образуют отрезок (в четырехмерном конфигурационном пространстве задана геометрическая связь), то независимых координат у отрезка останется только три: две координаты местоположения на плоскости и угол поворота отрезка, если местоположение выбрано осью вращения. В результате отрезок на плоскости имеет конфигурационным пространством многообразие $\mathbb{R}^2 \times S^1$, где S^1 — окружность.

2. В трехмерном пространстве $\mathbb{R}^3$ две точки образуют шестимерное конфигурационное пространство $\mathbb{R}^6 = \mathbb{R}^3 \times \mathbb{R}^3$. Если точки жестко связаны, то они образуют отрезок или «задают связь». Если закрепить один из концов отрезка (ещё три связи по числу координат), то полученная конфигурация с четырьмя связями называется «сферический маятник», конфигурационным пространством которого будет многообразие двумерная сфера: $S^2 = \{\, (x, y, z) \colon x^2 + y^2 + z^2 = r^2 \,\}$.

3. Если закрепить ось сферического маятника в одной плоскости, то мы получим «плоский маятник», конфигурационное пространство которого есть многообразие окружность S^1.

4. Если прикрепить к концу плоского маятника второй плоский маятник, то мы получим «плоский двойной маятник», конфигурационное пространство которого есть прямое произведение двух окружностей, т.е. многообразие двумерный тор $T^2 = S^1 \times S^1$.

5. Конфигурационное пространство сферического двойного маятника есть прямое произведение двух сфер $S^2 \times S^2$.

Если рассмотреть многообразие «локально» — в окрестности произвольной его точки, то любая точка m-мерного многообразия допускает взаимно однозначное и дифференцируемое в обе стороны отображение либо на пространство $\mathbb{R}^m$, либо на полупространство $\mathbb{H}^n$. Другими словами, окрестность любой точки m-мерного многообразия $U \cap M$ «с точностью до диффеоморфизма» есть либо часть пространства $\mathbb{R}^m \times \{\, 0 \,\} \subset \mathbb{R}^n$, либо часть полупространства $\mathbb{H}^m \times \{\, 0 \,\} \subset \mathbb{R}^n$. Однако эти условия не могут выполняться одновременно для одной и той же точки ν. Множество всех точек $\nu \in M$, удовлетворяющих второму условию, называется *краем* многообразия M и обозначается ∂M.

Отметим два крайних случая: точка в $\mathbb{R}^n$ есть нульмерное многообразие, а открытое подмножество в $\mathbb{R}^n$ есть n-мерное многообразие.

Рассмотрим сферу, которая касается плоскости в начале координат. Все точки плоскости взаимно однозначно и непрерывно проецируются на сферу с помощью лучей, проведенных через точку сферы, противоположную точке касания. Такое проецирование сферы на плоскость называется стереографической проекцией, а плоскость называется *картой многообразия*

«сфера». Однако точка сферы, из которой проводятся лучи, проецируется на бесконечность, и вблизи неё необходима другая карта с другой системой координат, т.е. не существует такого координатного отображения открытой плоскости, которое годилось бы для всей замкнутой сферы. Таким образом, сфера и сфера с единственной «выколотой» точкой — это два существенно разных объекта. Сфера — замкнутое множество, а сфера с выколотой точкой — открытое множество. В общем случае не существует координатной системы эквивалентной координатной системе пространства такой же размерности, которая годилась бы для всего замкнутого многообразия. Поэтому, на многообразии система координат задаётся локально, только в пределах карты, и необходимо задавать несколько карт с различными системами координат. Например, каждую полусферу можно рассмотреть как кусок плоскости, а сферу представлять склеенной из двух кусков.

Обратим внимание ещё раз: Карта многообразия с системой координат есть область пространства, которая взаимно однозначно проецируется на соответствующую область многообразия. Поэтому говорят, что многообразие локально диффеоморфно $\mathbb{R}^n$ или $\mathbb{H}^n$ (неразличимо с ними), но не обязательно диффеоморфно им в целом (глобально). Поскольку так устроено каждое m-мерное многообразие, то говорят любые два многообразия одинаковой размерности и одного класса гладкости локально диффеоморфны. Таким образом, n-мерное многообразие есть топологическое пространство локально диффеоморфное пространству $\mathbb{R}^n$, которое может быть описано набором из n вещественных чисел. *Можно представить себе многообразие как составленное из кусков пространства* $\mathbb{R}^n$*, склеенных друг с другом посредством диффеоморфизмов.*

При переходе к рассмотрению глобальных структур (многообразия в целом) ситуация в корне меняется. Рассмотрим ограниченные многообразия. Во-первых, многообразия могут быть открытыми и замкнутыми (замкнутые ограниченные многообразия компактны). Во-вторых, многообразия могут состоять из отдельных кусков (частей), либо из одного куска. Многообразие *связно* (состоит из одного куска), если его нельзя разбить на два непересекающихся открытых подмногообразия. В-третьих, многообразия могут быть ориентируемыми и неориентируемыми. Ориентируемые поверхности называются еще двусторонними, а неориентируемые поверхности — односторонними.

Всякое связное одномерное многообразие диффеоморфно в целом (глобально) либо окружности S^1, либо интервалу прямой $\mathbb{R}^1$.

Примерами компактных двумерных многообразий являются сфера S^2, тор T^2 (диффеоморфный сфере с одной ручкой, похожей на гирю для атлетов), крендель (диффеоморфный сфере с двумя ручками) и «сфера с n ручками». Сферы с ручками составляют важнейший класс двумерных многообразий — *замкнутые* (в том смысле, что не имеют края) *ориентируемые двумерные многообразия* или *замкнутые поверхности*. В курсах топологии доказывается, что всякое двумерное компактное связное ориентируемое многообразие диффеоморфно сфере с ручками. Сферы с разным количеством ручек не диффеоморфны.

Другой класс двумерных многообразий составляют *компактные незамкнутые* (имеющие край) *ориентируемые двумерные многообразия*, или *поверхности с краем*, которые получаются из замкнутой поверхности удалением конечного числа непересекающихся дисков.

Ещё один класс двумерных многообразий образуют *компактные неориентируемые двумерные многообразия*. Они также могут быть замкнутыми или с краем. Примеры: лист Мёбиуса (с краем), который получается склеиванием концов закрученной на пол-оборота полоски; проективная плоскость (замкнута), которая получается склеиванием листа Мёбиуса и диска; поверхность Клейна (замкнута), которая получается при склеивании двух листов Мёбиуса краями.

Этими тремя классами исчерпываются все компактные связные двумерные многообразия. Имеется классификация и открытых двумерных многообразий. Про трехмерные многообразия мало что известно. Полная классификация многообразий трех измерений не найдена даже для случая замкнутых многообразий. Только недавно Г. Перельман доказал гипотезу Пуанкаре, что всякое компактное односвязное трёхмерное многообразие диффеоморфно трёхмерной сфере.

Запишем теперь свойства многообразий на математическом языке. Определение многообразия. Если для всякой точки ν множества $M \subset \mathbb{R}^n$ существует окрестность U и диффеоморфизм $\varphi\colon U \to V$, то при выполнении условия, приведённого на стр. 31 для гомеоморфизма:

$$\varphi(U \cap M) = V \cap (\mathbb{R}^m \times \{\,0\,\}) = \{\,\nu \in V\colon x^{m+1} = \ldots = x^n = 0\,\},$$

M называется m-мерное многообразие. Если для некоторых точек ν это условие не выполняется, но выполняется дополнительное условие:

$$\varphi(U \cap M) = V \cap (\mathbb{H}^m \times \{\,0\,\}) = \{\,\nu \in V\colon x^m \geqslant 0 \text{ и } x^{m+1} = \ldots = x^n = 0\,\},$$

то M называется*m-мерное многообразие с краем*.

Диффеоморфизм $\varphi\colon U \to V$ или $\xi = \varphi(\nu) = (f_1(\nu), \ldots, f_n(\nu))$, где $f_i\colon U \to \mathbb{R}$ предполагает, что $\det \varphi'(\nu)$ для всех $\nu \in U$. Это значит, что все векторы $\operatorname{grad} f_1, \ldots, \operatorname{grad} f_n$ линейно независимы.

Говорят, что M есть *вложенное* в пространство $\mathbb{R}^n$ многообразие размерности $m = n - k$, если пересечение окрестности U каждой точки $\nu \in M$ с M задаётся уравнениями $f_{m+1} = const, \ldots, f_n = const$, т.е. заданы уравнения k геометрических связей, и векторы $\operatorname{grad} f_{m+1}, \ldots, \operatorname{grad} f_n$ в ν линейно независимы. Геометрические связи делают часть координат пространства $\mathbb{R}^n$ зависимыми, оставшиеся независимые координаты в окрестности точки ν: $y^1 = f_1(x^1, \ldots, x^n), \ldots, y^m = f_m(x^1, \ldots, x^n)$ определяют размерность многообразия и могут образовать локальную систему координат вложенного многообразия:

$$x^1 = f_1^{-1}(y^1, \ldots, y^m), \ldots, x^m = f_m^{-1}(y^1, \ldots, y^m). \tag{3.92}$$

Таким образом, вложенное в пространство многообразие есть гладкая поверхность, заданная при помощи поверхностей уровня или связей.

В действительности оказывается, что всякое многообразие размерности $\dim M = n$ вложено в пространство $\mathbb{R}^N$ достаточно большой размерности. Например, достаточно $N > 2n$. Таким образом, абстрактное понятие многообразия охватывает не больший круг объектов, чем «n-мерные поверхности в N-мерном пространстве». Этот фундаментальный факт приводится в виде теоремы без доказательства и в дальнейшем не используется. Теорема. Всякое многообразие диффеоморфно подмногообразию пространства достаточно большой размерности.

Теорема. Для всякой точки ν n-мерного многообразия $M \subset \mathbb{R}^N$ выполнено следующее «координатное условие»: Существуют окрестность $U \subset \mathbb{R}^N$ точки ν, открытое множество $V \subset \mathbb{R}^n$ и взаимно однозначная дифференцируемая функция $\psi\colon V \to M \subset \mathbb{R}^N$ или $\psi = (f_1^{-1}, \ldots, f_n^{-1})$, такая что $\psi(V) = U \cup M$, и $\psi'(\upsilon)$ имеет ранг n для всякой точки $\upsilon = (y^1, \ldots, y^n) \in V$. Такая функция ψ называется *системой координат* многообразия в окрестности точки ν.

Картой n-мерного многообразия M называется открытое множество $V \subset \mathbb{R}^n$ вместе с системой координат $\psi\colon V \to M$.

Для всяких двух карт $\psi_1\colon V_1 \to M$ и $\psi_2\colon V_2 \to M$ отображение $\psi_2^{-1} \circ \psi_1\colon \psi_1^{-1}(\psi_2(V_2)) \to \mathbb{R}^n$ дифференцируемо и имеет невырожденный якобиан. Действительно, $\psi_2^{-1}(\nu)$ состоит из $f_1, \ldots, f_n$ компонент диффеоморфизма $\varphi(\nu) = (f_1(\nu), \ldots, f_N(\nu))$. Композиция $\psi_2^{-1} \circ \psi_1$ есть преобразование координат между картами.

Говорят, что на многообразии M задана *структура дифференцируемого многообразия*, если M снабжено конечным или счетным набором карт, так что каждая точка многообразия изображена хотя бы на одной карте. Карты V_1, V_2 называются *совместными*, если функции ψ_1, ψ_2 дифференцируемы. Если преобразования координат есть гладкие функции класса C^k, то многообразие называется *дифференцируемым или гладким многообразием*, или *структурой класса* k. Если $k = \infty$, то многообразие называется *аналитическим многообразием*.

Обратим внимание еще раз. Пусть многообразие M покрыто счетным числом открытых множеств $U_1, U_2, \ldots$. Для каждого U_i имеется диффеоморфизм $\varphi_i\colon U_i \to V_i$, где V_i — *открытая клетка* (диффеоморфная открытому шару D^n область в $\mathbb{R}^N$). Число $\dim M = n$ есть размерность многообразия M. Счетное число открытых клеток $V_1, V_2, \ldots$, для которых имеются диффеоморфизмы $\psi_i\colon V_i \to M$ (системы локальных координат многообразия M), есть набор карт для M. Если $U_i \cap U_j \neq \varnothing$, то существует диффеоморфизм $\varphi_{ji} = \varphi_j \varphi_i^{-1}$ множества $\varphi_i(U_i \cap U_j)$ на $\varphi_j(U_i \cap U_j)$. При этом диффеоморфизм $\psi_{ji} = \psi_j^{-1}\psi_i$, есть преобразование координат между совместными картами (есть условие совместности карт атласа или условие перехода между ними).

Атласом называется совокупность совместных друг с другом карт. Два атласа *эквивалентны*, если их объединение есть снова атлас (если любая карта первого атласа согласована с любой картой второго).

Примеры атласов:

1. Евклидово пространство $\mathbb{R}^n$ есть многообразие, атлас которого состоит из единственной карты.

2. Сфера S^2 есть многообразие, с атласом из двух карт $\mathbb{R}^2$, в стереографической проекции. Аналогичная конструкция годится и для n-мерной сферы $S^n = \left\{ (x_1, \ldots, x_{n+1})\colon x_1^2 + \ldots + x_{n+1}^2 = r^2 \right\}$ в виде двух $\mathbb{R}^n$.

3. Конфигурационное пространство плоского маятника — окружность S^1 — есть многообразие с атласом из двух карт, прямых $\mathbb{R}^1$, которым соответствует изменение угловой координаты $\varphi\colon \mathbb{R}^1 \to S^1$ в интервалах $(-\pi, \pi)$ и $(0, 2\pi)$.

4. Отрезок на плоскости имеет конфигурационным пространством многообразие $\mathbb{R}^2 \times S^1$, которое покрывается двумя картами.

Аналогично касательному расслоению пространства $\mathbb{R}^n$ с каждым гладким многообразием связано другое многообразие вдвое большей размерности, называемое касательным расслоением.

В каждой точке τ n-мерного многообразия M мы можем построить *n-мерное касательное пространство* TM_τ, которое имеет структуру линейного пространства $\mathbb{R}^n$ и, очевидно, не зависит от выбора системы координат (от карты). Объединение касательных пространств к многообразию M в разных точках $\bigcup\limits_{\tau \in M} TM_\tau$, где τ пробегает M, имеет естественную структуру дифференцируемого многообразия, размерность которого вдвое больше размерности M. Это многообразие называется *касательным расслоением* многообразия M и обозначается TM. В этом случае касательное пространство $TM_\tau \subset TM$ называется *слоем расслоения над точкой* τ, а многообразие M — *базой расслоения*. Расслоение TM является, по крайней мере, локально, прямым произведением вида $U \times F$, где U — открытое множество базы M, а F представляет собой типичный слой. Все пространство расслоения в целом (не локально) обычно не может быть представлено в виде прямого произведения базы и слоя.

Первая квадратичная форма (3.39) гладкой многомерной поверхности (многообразия M) всегда положительно определена и задаёт метрику касательного к поверхности пространства или длину касательного вектора в каждой точке τ: $(\,,)\colon TM_\tau \to \mathbb{R}$ или $\xi \mapsto (\xi, \xi)$, которая позволяет измерять данной точке поверхности длины кривых, углы между векторами, объёмы и т.д. Эта фиксированная положительно определённая квадратичная форма (ξ, ξ) на многообразии называется также *римановой метрикой* (*структурой*). Дифференцируемое многообразие M с римановой метрикой в каждом касательном пространстве TM_τ называется *римановым многообразием*. Пусть W — карта атласа M с координатами $x^1, \ldots, x^n$. Тогда риманова метрика задается формулой (3.39).

Выражение:

$$|\xi| = ds = \sqrt{\sum_{i,j=1}^{n} g_{ij} dx^i dx^j}, \tag{3.93}$$

аналогичное определению линейного элемента пространства (см. стр. 116) называют линейным элементом ds касательного пространства на римановом многообразии.

Если координаты вещественного пространства имеют разные единицы измерения, например, пространство термодинамики, в котором по координатным осям откладываются давление, объем и температура P, T, V, то метрика в таком пространстве не имеет смысла, так как члены метрики невозможно сложить. Риманова метрика на поверхности (вложенном многообразии), заданной в аффинном пространстве, координаты которого имеют разные единицы измерения, имеет смысл, даже в этом случае, так как за счёт величин g_{ij} единицы измерения членов в правой части формы одинаковые, и их можно складывать.

Мы пришли к общей римановой геометрии, в основе которой лежат три идеи. Первая идея — признание того, что вообще возможна геометрия, отличная от евклидовой, была впервые развита Н. И. Лобачевским. Вторая — это идущее от К. Гаусса понятие внутренней геометрии поверхностей и её аналитический аппарат в виде квадратичной формы, определяющей линейный элемент поверхности. Третья идея — это понятие об n-мерном пространстве, выдвинутое и разработанное в простейших случаях в 1-й половине 19 в. рядом геометров. Понятие многомерного многообразия было впервые сформулировано Б. Риманом в его лекции «О гипотезах, лежащих в основании геометрии» (1854).

Для системы координат $\psi\colon V \to \mathbb{R}^n$ в точке $\upsilon \in V$ можно выбрать базис $e_1^\upsilon, \ldots, e_m^\upsilon$. Часто необходимо выбрать ориентацию μ_τ в каждом касательном пространстве TM_τ многообразия M. Такие ориентации называют *согласованными* , если для каждой системы координат $\psi\colon V \to \mathbb{R}^n$ и каждой пары $\upsilon, \varsigma \in W$ равенство $[e_1^\upsilon, \ldots, e_m^\upsilon] = \mu_{\psi(\upsilon)}$ выполнено тогда и только тогда, когда $[e_1^\varsigma, \ldots, e_m^\varsigma] = \mu_{\psi(\varsigma)}$. Предположим, что выбраны согласованные ориентации μ_τ. Если система координат $\psi\colon V \to \mathbb{R}^n$ такова, что $[e_1^\upsilon, \ldots, e_m^\upsilon] = \mu_{\psi(\upsilon)}$ для некоторого, а потому и для каждого $\upsilon \in V$ то говорят, что ψ *сохраняет ориентацию*.

Многообразие M, допускающее выбор согласованных ориентаций μ_τ, называется *ориентируемым многообразием*, а всякий такой выбор μ_τ — *ориентацией* μ этого многообразия. Классическим примером неориентируемого многообразия является лист Мёбиуса — закрученная на пол-оборота и склеенная концами полоска.

Если M есть m-мерное многообразие с краем и $\tau \in \partial M$, то $T\partial M_\tau$ есть $(m-1)$-мерное подпространство m-мерного векторного пространства TM_τ. Существуют в точности два единичных вектора в TM_τ, перпендикулярных к $T\partial M_\tau$. Только один из этих единичных векторов равен некоторому базисному вектору системы координат $\psi\colon V \to \mathbb{R}^n$ с $V \subset \mathbb{H}^n$ и $\psi(0) = \tau$. Этот

единичный вектор $n(\tau)$ называется *ортом внешней нормали* и не зависит от системы координат ψ.

Пусть μ — ориентация на m-мерном многообразии с краем M.

Для всякой точки $\tau \in \partial M$ выберем $[n(\tau), e_1^\upsilon, \ldots, e_{m-1}^\upsilon] = \mu_\tau$. Тогда $[e_1^\upsilon, \ldots, e_{m-1}^\upsilon] = \partial\mu_\tau$ есть согласованная ориентация на ∂M. Таким образом, если M ориентируемо, то ∂M также ориентируемо, и ориентация на M определяет ориентацию $\partial\mu$ на ∂M, называемую *индуцированной ориентацией*. Соответственно, в полупространстве $\mathbb{H}^m$ с его стандартной ориентацией, индуцированной ориентацией на $\mathbb{R}^{m-1} = \{\, \tau \in \mathbb{H}^m \colon x^m = 0 \,\}$ служит стандартная ориентация, умноженная на $(-1)^m$.

Назовём вектор-функцию, которая относит каждой точке многообразия $\tau \in M$ некоторый вектор $\xi(\tau) \in TM_\tau$, *векторным полем* ξ на многообразии M.

Дифференциальная 1*-форма* в точке дифференцирования многообразия τ есть гладкое отображение $\omega_\tau \colon TM_\tau \to \mathbb{R}$ касательного пространства TM_τ в прямую (аналогично (3.48)):

$$\omega = a_1(x^1, \ldots, x^n) \cdot dx^1 + \ldots + a_m(x^1, \ldots, x^n) \cdot dx^n, \tag{3.94}$$

где $a_i(x^1, \ldots, x^n)$ — гладкие функции.

Дифференциальная 1*-форма* на многообразии M есть гладкое отображение касательного расслоения TM в прямую $\omega \colon TM \to \mathbb{R}$, линейное на каждом касательном пространстве TM_τ.

Функция ω, которая каждой точке многообразия $\tau \in M$ относит $\omega(\tau) \in \Lambda^k(TM_\tau)$, называется формой k-ой степени на M и может быть записана в аналогичном (3.54) виде:

$$\omega^k = \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} a_{i_1 \ldots i_k}(x^1, \ldots, x^n) \cdot dx^{i_1} \wedge \ldots \wedge dx^{i_k}, \tag{3.95}$$

где $a_{i_1 \ldots i_k}(x^1, \ldots, x^n)$ — гладкие функции.

Сложение, умножение на число и внешнее умножение форм на M определяются поточечно: в каждой точке $\tau \in M$ надо сложить, умножить на число или внешне перемножить соответствующие алгебраические внешние формы на касательном пространстве TM_τ.

Координаты касательного пространства $dy^1, \ldots, dy^m$ есть базис линейного пространства $T\mathbb{R}^n_{\varphi(\tau)}$ дифференциальных 1-форм, а $dy^{i_1} \wedge \ldots \wedge dy^{i_k}$ есть базис линейного пространства $\Lambda^k(\mathbb{R}^m)$ дифференциальных k-форм. Причём

$$dy^i = \frac{\partial y^i}{\partial x^1}\, dx^1 + \ldots + \frac{\partial y^i}{\partial x^n}\, dx^n \quad (i = 1, \ldots, m). \tag{3.96}$$

Тогда

$$dy^{i_1} \wedge \ldots \wedge dy^{i_k} = \sum_{1 \leqslant j_1 < \ldots, < j_k \leqslant n} \frac{\partial(y^{i_1}, \ldots, y^{i_k})}{\partial(x^{j_1}, \ldots, x^{j_k})}\, dx^{j_1} \wedge \ldots \wedge dx^{j_k} \tag{3.97}$$

Рассмотрим преобразование $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ переменных (координат) $x^1, \ldots, x^n$ в координаты $y^1 = f_1(x^1, \ldots, x^n), \ldots, y^n = f_1(x^1, \ldots, x^n)$.

Величина

$$dV = dx^1 \wedge \ldots \wedge dx^n \tag{3.98}$$

есть объём n-мерного бесконечно малого параллелепипеда и носит название *элемент объёма* dV в ориентированном n-мерном евклидовом пространстве, хотя форма dV, вообще говоря, не является дифференциалом какой-либо формы $(n-1)$-й степени.

Тогда якобиан

$$dy^1 \wedge \ldots \wedge dy^n = \frac{\partial(y^1, \ldots, y^n)}{\partial(x^1, \ldots, x^n)}\, dx^1 \wedge \cdots \wedge dx^n. \tag{3.99}$$

характеризует растяжение (сжатие) элемента объёма dV при преобразовании координат от переменных $x^1, \ldots, x^n$ к переменным $y^1, \ldots, y^n$.

Формулы (3.96–3.99) преобразования координат

$$\begin{aligned}
dy^i &= \frac{\partial y^i}{\partial x^1}\, dx^1 + \ldots + \frac{\partial y^i}{\partial x^n}\, dx^n \quad (i = 1, \ldots, n),\\
dy^{i_1} \wedge \ldots \wedge dy^{i_k} &= \sum_{1 \leqslant j_1 < \ldots, < j_k \leqslant n} \frac{\partial(y^{i_1}, \ldots, y^{i_k})}{\partial(x^{j_1}, \ldots, x^{j_k})}\, dx^{j_1} \wedge \ldots \wedge dx^{j_k},\\
dy^1 \wedge \ldots \wedge dy^n &= \frac{\partial(y^1, \ldots, y^n)}{\partial(x^1, \ldots, x^n)}\, dx^1 \wedge \cdots \wedge dx^n.
\end{aligned}$$

справедливы и в том случае, если мы рассматриваем локальные координаты на многообразии.

3.5 Производная по направлению

Кривая в пространстве (см. стр. 31) представляет собой отображение $\nu\colon I \to \mathbb{R}^n$ или $\nu(s) = (x^1(s), \ldots, x^n(s))$, где число s называют *параметр*. Производная кривой $\nu(s)$ по параметру s имеет вид:

$$\nu'_s = \left(\frac{dx^1}{ds}, \ldots, \frac{dx^n}{ds}\right). \tag{3.100}$$

Производные функций $x^1(s), \ldots, x^n(s)$ есть компоненты *вектора направления* ν'_s касательной к кривой в каждой точке дифференцирования, совпадающего по направлению с вектором $\xi = (dx^1, \ldots, dx^n)$.

Если мы дифференцируем функцию $f\colon \mathbb{R}^n \to \mathbb{R}$ не на всём пространстве $\mathbb{R}^n$, а только на заданной в пространстве кривой $\nu\colon I \to \mathbb{R}^n$, то мы получаем производную функции одного переменного, так как гладкая кривая имеет размерность $\mathbb{R}^1$. Такая производная называется *производная по направлению* s и вычисляется по правилу сложной функции

$$f'_s = \frac{df}{ds} = \frac{\partial f}{\partial x^1}\frac{dx^1}{ds} + \ldots + \frac{\partial f}{\partial x^n}\frac{dx^n}{ds} = \operatorname{grad} f \cdot \nu'_s. \tag{3.101}$$

Величина $df(\xi)$ есть *образ* в координатном пространстве касательного вектора ξ. Функция $f(s)$ вдоль выбранной кривой с параметром s есть образ этой кривой на n-мерной поверхности графика функции. В результате дифференцирования мы аппроксимируем (заменяем) в окрестности точки τ бесконечно малую дугу кривой, выбранной на n-мерной поверхности функции f, бесконечно малым отрезком df, т.е. линейной функцией, которую значительно проще исследовать. Дифференцирование всей кривой на поверхности означает её замену ломаной, состоящей из бесконечно малых звеньев df.

Если мы дифференцируем по параметру s функцию $f\colon \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}$ или $z = f(x^1,\dots,x^n,s)$, которая явно зависит от параметра s, то мы должны записать производную функции в следующем виде

$$f'_s = \frac{df}{ds} = \frac{\partial f}{\partial x^1}\frac{dx^1}{ds} + \dots + \frac{\partial f}{\partial x^n}\frac{dx^n}{ds} + \frac{\partial f}{\partial s} = \operatorname{grad} f\cdot\nu'_s + \frac{\partial f}{\partial s}. \tag{3.102}$$

Выражение (3.102) называют также *полной производной* функции $z = f(x^1,\dots,x^n,s)$. Соответственно, производная по направлению (3.101) есть полная производная функции $z = f(x^1,\dots,x^n)$.

Диференциал

$$df = \frac{\partial f}{\partial x^1}\,dx^1 + \dots + \frac{\partial f}{\partial x^n}\,dx^n + \frac{\partial f}{\partial s}\,ds. \tag{3.103}$$

называют *полным дифференциалом* функции $z = f(x^1,\dots,x^n,s)$, а дифференциал (3.20) — полным диференциалом функции $z = f(x^1,\dots,x^n)$.

Другими словами, полная производная и полный дифференциал должны включать при вычислении *все* частные производные функции.

Если дифференциал функции не является полным, то на практике это обычно означает, что мы не смогли определить все координаты, на которых задана рассматриваемая функция.

Пример 1. Если в качестве параметра выбрать время t, то отображение $\nu\colon I\to\mathbb{R}^n$ называют *движением*, кривую $\nu(t)$ обозначают $\mathbf{x}(t)$ и называют *траекторией движения*, производную вектора ν'_s по времени обозначают $v = \dot{\mathbf{x}} = (\dot{x}^1,\dots,\dot{x}^n)$ и называют *скоростью*. В свою очередь производную скорости по времени $\dot{v} = \ddot{\mathbf{x}} = (\ddot{x}^1,\dots,\ddot{x}^n)$ называют *ускорением*.

Пример 2. Скорость скалярной функции $z = f(x^1,\dots,x^n,t)$:

$$\dot{f} = \operatorname{grad} f\cdot\dot{\mathbf{x}} + \frac{\partial f}{\partial t}. \tag{3.104}$$

Величина $\operatorname{grad} f\cdot\dot{\mathbf{x}}$ получена только за счет перехода точки ν вдоль своей траектории от одних значений функции к другим и называется *переносной* (*конвективной*) скоростью. Величина $\frac{\partial f}{\partial t}$ дает изменение функции в неподвижной точке, полученное из-за *нестационарности* функции, и называется *местной* (*локальной*) скоростью.

Пример 3. Полная производная функции $z = f(x^1,\dots,x^n)$ по x^1:

$$\frac{df}{dx^1} = \frac{\partial f}{\partial x^1} + \frac{\partial f}{\partial x^2}\frac{dx^2}{dx^1} + \dots + \frac{\partial f}{\partial x^n}\frac{dx^n}{dx^1}. \tag{3.105}$$

Пример 4. Скорость вектор-функции $\mathbf{y} = \varphi(t, \nu)$ имеет вид:

$$\dot{\varphi} = \frac{\partial \varphi}{\partial \mathbf{x}}\,\dot{\mathbf{x}} + \frac{\partial \varphi}{\partial t}. \tag{3.106}$$

Пример 5. *Комплексной кривой* называют функцию φ: $\varphi\colon I \to \mathbb{C}^n$, $I \subset \mathbb{R}$ — отображение интервала I вещественной оси t в линейное комплексное пространство $\mathbb{C}^n$. Производная кривой φ в точке $t_0 \in I$ определяется обычным образом, это вектор пространства $\mathbb{C}^n$:

$$\dot{\varphi}(t_0) = \frac{d\varphi}{dt} = \lim_{\Delta t \to 0} \frac{\varphi(t_0 + \Delta t) - \varphi(t_0)}{\Delta t}. \tag{3.107}$$

Если $n > 1$, то, в отличие от $\mathbb{C}$, две кривые в $\mathbb{C}^n$ перемножать нельзя. Однако можно умножить $\varphi\colon I \to \mathbb{C}^n$ на $f\colon I \to \mathbb{C}$:

$$(f \cdot \varphi)(t) = f(t) \cdot \varphi(t). \tag{3.108}$$

Свойства производной комплексной кривой:

$$\frac{d\overline{\varphi}}{dt} = \overline{\frac{d\varphi}{dt}}, \quad \frac{d(\varphi_1 + \varphi_2)}{dt} = \frac{d\varphi_1}{dt} + \frac{d\varphi_2}{dt}, \quad \frac{d(f\varphi)}{dt} = \frac{df}{dt}\varphi + f\frac{\varphi}{dt}. \tag{3.109}$$

Разумеется, мы предполагаем, что производные существуют.

Пример 6. Рассмотрим *бесконечно малый поворот* радиус-вектора $\mathbf{r}$ в пространстве с координатами x, y, z вокруг оси z, т.е. такое ортогональное преобразование, при котором составляющие вектора изменяются бесконечно мало. Матрица бесконечно малого поворота σ задаёт почти тождественное преобразование E, отличающееся от него лишь бесконечно малым оператором ε: $\sigma = E + \varepsilon$. Если матрица конечного вращения вокруг оси z на угол φ имеет вид (2.87):

$$\begin{Vmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{Vmatrix}$$

то матрица σ бесконечно малого поворота на угол $d\varphi$ (так как $\cos d\varphi \approx 1$, $\sin d\varphi \approx d\varphi$) имеет вид:

$$\sigma = E + \varepsilon = \begin{Vmatrix} 1 & -d\varphi & 0 \\ d\varphi & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix} \text{ или } \varepsilon = \begin{Vmatrix} 0 & -d\varphi & 0 \\ d\varphi & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix} = d\varphi \begin{Vmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix}.$$

Таким образом, ε есть кососимметрическая матрица. Обратная матрица равна $\sigma^{-1} = E - \varepsilon$. Тогда получаем $\sigma\sigma^{-1} = (E + \varepsilon)(E - \varepsilon) \approx E$.

Произвольный (не ортогональный осям) бесконечно малый поворот задаётся кососимметрической матрицей следующего вида:

$$\varepsilon = \begin{Vmatrix} 0 & -d\Omega_z & d\Omega_y \\ d\Omega_z & 0 & -d\Omega_x \\ -d\Omega_y & d\Omega_x & 0 \end{Vmatrix} \quad \text{или} \quad \begin{aligned} dx &= z\,d\Omega_y - y\,d\Omega_z, \\ dy &= x\,d\Omega_z - z\,d\Omega_x, \\ dz &= y\,d\Omega_x - x\,d\Omega_y. \end{aligned} \tag{3.110}$$

Компоненты $d\Omega_i$ связаны с элементами матрицы ε_{jk} соотношением:

$$d\Omega_i = \frac{1}{2}\sum_{j,k=1}^{3} \delta_{ijk}\varepsilon_{jk}, \tag{3.111}$$

где δ_{ijk} — называется *символ Леви–Чивиты*, равный нулю, если среди индексов i, j, k имеются одинаковые, и равный ± 1 в зависимости от чётности или нечётности перестановки индексов i, j, k.

В результате мы можем записать векторное уравнение:

$$d\mathbf{r} = d\Omega \times \mathbf{r}, \tag{3.112}$$

где $d\Omega$ — вектор, компоненты которого равны $d\Omega_x, d\Omega_y, d\Omega_z$.

Однако представления ортогонального преобразования в векторной форме ещё не достаточно для доказательства того, что $d\Omega$ есть вектор. Истинные векторы, в частности должны выдерживать любые ортогональные преобразования, оставаясь при этом инвариантами. Оказывается, что величина $d\Omega$ в основном «выдерживает» эти «испытания на вектор» за исключением операции отражения, т.е. как результат векторного произведения является псевдовектором или аксиальным вектором.

Вектор $\omega = \frac{d\Omega}{dt}$ называется *угловая скорость вращения* и направлен вдоль оси бесконечно малого поворота, совершающегося в момент времени t. Эта ось называется *мгновенной осью вращения*.

В результате мы можем записать *уравнение вращения*

$$\dot{\mathbf{r}} = \omega \times \mathbf{r}. \tag{3.113}$$

На вектор $\mathbf{r}$, который мы здесь дифференцируем, не было наложено никаких условий. Произвольность этого вектора можно подчеркнуть, записав общее уравнение в операторной форме:

$$\left(\frac{d}{dt}\right) = \left(\frac{d}{dt}\right)_{\text{перемещение}} + \omega \times_{\text{вращение}} . \tag{3.114}$$

Мы разделили изменение произвольного вектора, связанное с его перемещением относительно начала координат и с его вращением.

Пример 7. Пусть $\mathbf{a}$ — векторное поле в области U. Кривая $\nu\colon I \to \mathbb{R}^n$ в каждой точке которой касательный вектор $d\mathbf{x}(s)$ имеет направление векторного поля $\mathbf{a}$, называется *векторная линия* или *линия тока* . Тогда с учётом (2.14) можно записать *уравнение векторной линии*:

$$\frac{dx^1}{a_1} = \ldots = \frac{dx^n}{a_n} = \frac{d\mathbf{x}}{\mathbf{a}}. \tag{3.115}$$

Пусть $\mathbf{a}$ — поле скоростей частиц стационарного потока жидкости. Тогда векторные линии

$$\frac{dx}{a_x} = \frac{dy}{a_y} = \frac{dz}{a_z}, \tag{3.116}$$

есть траектории частиц жидкости.

Пример 8. Пусть $f = u + iv$ — комплексная функция, аналитическая в области $A \subset \mathbb{C}$, где $u(x, y)$ и $v(x, y)$ связаны условиями Коши-Римана (3.33). Зададим в области A векторное поле $\mathbf{a} = (a_x, a_y)$:

$$a_x = \frac{\partial u}{\partial x} \overset{(3.33)}{=} \frac{\partial v}{\partial y}; \quad a_y = \frac{\partial u}{\partial y} \overset{(3.33)}{=} -\frac{\partial v}{\partial x}. \tag{3.117}$$

Тогда

$$du = a_x\,dx + a_y\,dy = (\mathbf{a}, \xi), \tag{3.118}$$

есть скалярное произведение вектора $\mathbf{a}$ и вектора $\xi = (dx, dy)$, а

$$dv = a_x\,dy - a_y\,dx = \begin{vmatrix} a_x & a_y \\ dx & dy \end{vmatrix} = [\mathbf{a}, \xi], \tag{3.119}$$

есть кососкалярное произведение вектора $\mathbf{a}$ и вектора ξ.

Если $du = 0$, то $u(x, y) = const$ — линия уровня, в каждой точке перпендикулярная вектору $\mathbf{a} = \operatorname{grad} u$.

Если $dv = 0$, то $v(x, y) = const$ — векторная линия, имеющая в каждой точке направление вектора $\mathbf{a}$:

$$a_x\,dy - a_y\,dx = 0 \quad \text{или} \quad \frac{dx}{a_x} = \frac{dy}{a_y}. \tag{3.120}$$

Таким образом, каждая из функций $u(x, y)$ и $v(x, y)$ задает семейство кривых на плоскости. Кривые из разных семейств пересекаются друг с другом под прямым углом в каждой точке плоскости. Внутри каждого из семейств, кривые не пересекаются между собой.

Пример 9. Пусть $\mathbf{a}$ — векторное поле в области U, $f\colon U \to \mathbb{R}$ — дифференцируемая функция. *Производной* скалярного поля f *по направлению* s векторного поля $\mathbf{a} = (a_1, \ldots, a_n)$ называется величина:

$$f'_s = \frac{\partial f}{\partial x^1}\,a_1 + \ldots + \frac{\partial f}{\partial x^n}\,a_n = \operatorname{grad} f \cdot \mathbf{a}. \tag{3.121}$$

Производную по направлению векторного поля можно записать в виде

$$L_{\mathbf{a}} f = f'_s \text{ или } L_{\mathbf{a}} = \mathbf{a}\nabla = a_1 \frac{\partial}{\partial x^1} + \ldots + a_n \frac{\partial}{\partial x^n}. \tag{3.122}$$

Скалярный оператор $L_{\mathbf{a}} = \mathbf{a}\nabla$ получен скалярным умножением произвольного вектора поля $\mathbf{a}$ на оператор Гамильтона ∇, стоящий *справа* от $\mathbf{a}$ (в отличие от $\nabla\,\mathbf{a} \overset{(3.83)}{=} \operatorname{div}\mathbf{a}$). Этот дифференциальный оператор первого порядка называют также *производной Ли* и обозначают $L_{\mathbf{a}}$ (L — в честь Софуса Ли).

Производная функции по направлению векторного поля это геометрический объект, по самому своему определению не зависящий ни от какой

системы координат. Поскольку производная Ли взаимно однозначно задается полем $\mathbf{a}$, то их часто отождествляют.

Как мы знаем совокупность объектов и отображений называется категорией. Совокупность всех линейных пространств (совокупность объектов) и их линейных отображений друг в друга также образует соответствующую категорию.

Множество F всех бесконечно дифференцируемых функций $f\colon U \to \mathbb{R}$ имеет естественную структуру вещественного линейного пространства (так как сложение функций сохраняет дифференцируемость), и даже кольца (так как произведение бесконечно дифференцируемых функций дифференцируемо), или, лучше сказать, $\mathbb{R}$-алгебры (кольца, для элементов которого определено умножение на числа, удовлетворяющее обычным требованиям). Совокупность всех векторных полей $\mathbf{a}$ по разным направлениям (разным параметрам s) образует линейное пространство векторных полей. Производная функции из F по направлению поля $\mathbf{a}$ снова принадлежит F (здесь существенна бесконечная дифференцируемость). Итак, дифференцирование по направлению гладкого поля $\mathbf{a}$ есть отображение $L_\xi\colon F \to F$ алгебры бесконечно дифференцируемых функций в себя. Свойства этого отображения:

1. $L_{\mathbf{a}}(c_1 f_1 + c_2 f_2) = c_1 L_{\mathbf{a}} f_1 + c_2 L_{\mathbf{a}} f_2$, $c_1, c_2 \in \mathbb{R}$, линейность;
2. $L_{\mathbf{a}}(f_1 f_2) = f_1 L_{\mathbf{a}} f_2 + f_2 L_{\mathbf{a}} f_1$ формула Лейбница;
3. $L_{\mathbf{a}+\mathbf{b}} = L_{\mathbf{a}} + L_{\mathbf{b}}$;
4. $L_{f\mathbf{a}} = f L_{\mathbf{a}}$.

Алгебраисты называют отображение $L_{\mathbf{a}}$ (коммутативного) кольца в себя *дифференцированием*, если оно обладает свойствами 1 и 2 отображения. Все дифференцирования кольца благодаря свойствам 3 и 4 образуют *модуль* над этим кольцом (модуль над кольцом — обобщение линейного пространства над $\mathbb{R}$: элементы модуля можно складывать между собой и умножать на элементы кольца).

Аналитики называют отображение $L_{\mathbf{a}}$ *линейным однородным дифференциальным оператором первого порядка*. Это название связано с тем, что согласно 1 и 2, оператор $L_{\mathbf{a}}F \to F$ $\mathbb{R}$-линеен. Аналогичный $L_{\mathbf{a}}$ оператор производной Ли вдоль векторного поля $\mathbf{a}$ можно определить не только для функций, но и для произвольных дифференциально-геометрических объектов, переносимых диффеоморфизмами (векторных полей, форм, тензоров) — производная каждого из объектов будет объектом той же природы.

Французы называют оператор $L_{\mathbf{a}}$ *производной рыбака*: рыбак сидит на месте и дифференцирует объекты, проносимые мимо него фазовым потоком (см. § 5.1).

Пусть заданы два векторных поля:

$$\nu'_t = \mathbf{a}(\nu), \quad \nu'_s = \mathbf{b}(\nu). \tag{3.123}$$

Рассмотрим смешанную производную функции f по двум разным направлениям, которая есть результат последовательного воздействия (композиции) двух дифференциальных операторов:

$$L_{\mathbf{b}}L_{\mathbf{a}}f = f''_{ts} = \sum_{i=1}^{n} b_i \frac{\partial}{\partial x^i} \sum_{j=1}^{n} a_j \frac{\partial f}{\partial x^j} = \sum_{i,j=1}^{n} b_i \frac{\partial a_j}{\partial x^i} \frac{\partial f}{\partial x^j} + \sum_{i,j=1}^{n} b_i a_j \frac{\partial^2 f}{\partial x^i \partial x^j}.$$

Операция композиции двух дифференциальных операторов, вообще говоря, не коммутативна:

$$(L_{\mathbf{b}}L_{\mathbf{a}} - L_{\mathbf{a}}L_{\mathbf{b}})f = f''_{ts} - f''_{st} = \sum_{i,j=1}^{n} \left(b_i \frac{\partial a_j}{\partial x^i} - a_i \frac{\partial b_j}{\partial x^i} \right) \frac{\partial f}{\partial x^j}. \tag{3.124}$$

На первый взгляд коммутатор двух операторов дифференцирования — это дифференциальный оператор второго порядка, но так как слагаемое со вторыми производными $f''_{x^i x^j}$ при вычитании пропадает, то коммутатор $L_{\mathbf{b}}L_{\mathbf{a}} - L_{\mathbf{a}}L_{\mathbf{b}}$ есть линейный дифференциальный оператор первого порядка.

Скобкой Пуассона (коммутатором) векторных полей $\mathbf{a}$ и $\mathbf{b}$:

$$\mathbf{c} = \{\mathbf{a}, \mathbf{b}\} \tag{3.125}$$

называется векторное поле $\mathbf{c} = (c_1, \ldots, c_n)$, для которого оператор дифференцирования есть выражение:

$$L_{\mathbf{c}} = L_{\mathbf{b}}L_{\mathbf{a}} - L_{\mathbf{a}}L_{\mathbf{b}} \text{ или } L_{\mathbf{c}} = c_1 \frac{\partial}{\partial x^1} + \ldots + c_n \frac{\partial}{\partial x^n} \tag{3.126}$$

где $c_j = \{\mathbf{a}, \mathbf{b}\}_j = \sum\limits_{i=1}^{n} \left(b_i \frac{\partial a_j}{\partial x^i} - a_i \frac{\partial b_j}{\partial x^i} \right)$.

Свойства скобки Пуассона векторных полей $\mathbf{a}, \mathbf{b}, \mathbf{c}$ ($c \in \mathbb{R}$):

1. $\{\mathbf{a}, \mathbf{b} + \mathbf{c}\} = \{\mathbf{a}, \mathbf{b}\} + \{\mathbf{a}, \mathbf{c}\}$, $c\{\mathbf{a}, \mathbf{b}\} = \{c\mathbf{a}, \mathbf{b}\} = \{\mathbf{a}, c\mathbf{b}\}$ линейность;
2. $\{\mathbf{a}, \mathbf{b}\} = -\{\mathbf{b}, \mathbf{a}\}$ кососимметричность;
3. $\{\mathbf{a}, \{\mathbf{b}, \mathbf{c}\}\} + \{\mathbf{b}, \{\mathbf{c}, \mathbf{a}\}\} + \{\mathbf{c}, \{\mathbf{a}, \mathbf{b}\}\} = 0$ тождество Якоби.

Алгеброй Ли называется линейное пространство с бинарной операцией, обладающей свойствами 1, 2 и 3. Следовательно, скобка Пуассона превращает линейное пространство векторных полей в алгебру Ли.

3.6 Дифференциальная геометрия

Рассмотрим график функции одного переменного $y = f(x)$, который представляет собой плоскую (лежащую на плоскости) кривую. Введем для кривой понятие *порядок касания*. Плоская кривая имеет два типа точек касания. Если кривая лежит по одну сторону от касательной, то прямая называется *касательной первого порядка*. Если кривая пересекает касательную

в точке касания, то прямая называется *касательной второго порядка*, а точка касания, называется *точкой перегиба* кривой.

Названия касательных обусловлены тем, что *в точках перегиба и только в них вторая производная кривой всегда равна нулю*

$$f'' = 0. \tag{3.127}$$

Причём это условие не зависит от выбора системы координат, т.е. *точки перегиба есть инварианты*. Плоская кривая не имеет касательных выше второго порядка. Например, порядок касания оси x с графиком $y = x^{2n}, n \in \mathbb{N}$ первый, а с графиком $y = x^{2n+1}, n \in \mathbb{N}$ — второй.

Обозначим через α угол между касательными в двух соседних точках 1 и 2, и через Δs длину дуги между этими точками. Величина:

$$k = \lim_{\Delta s \to 0} \frac{\alpha}{\Delta s} = \frac{1}{R} \tag{3.128}$$

называется *кривизной* кривой в данной точке с радиусом R (рис. 3.5).

Другими словами, плоская гладкая кривая имеет в каждой точке кроме касательной еще и *соприкасающуюся окружность*, характеризующую искривленность кривой в данной точке. Центр и радиус R этой окружности называют *центром кривизны* и *радиусом кривизны*. Множество центров кривизны плоской кривой называется её *эволютой* (от лат. evolutus — развёрнутый). Сама кривая по отношению к своей эволюте называется её *эвольвентой* (от лат. evolvents — разворачивающий).

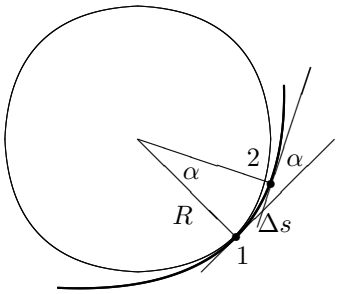


Рис. 3.5: Кривизна

Если за исходную кривую взять эволюту γ, то её эвольвента η — это траектория, которую описывает конец натянутой нити, сматываемой с исходной кривой (рис. 3.6). В точке касания эвольвенты и эволюты у эвольвенты имеется острие, обозначенное далее пунктиром. Эвольвента в этом месте имеет бесконечную кривизну и не гладкая. Плоскость, в которой лежит соприкасающаяся окружность, называется *соприкасающейся плоскостью*. Прямая линия, соединяющая центр кривизны с касательной точкой, перпендикулярна касательной и называется нормалью. Эволюта γ является огибающей нормалей эвольвенты η.

Понятия эволюты и эвольвенты введены Х. Гюйгенсом (1673) и присутствовали во многих старых учебниках анализа, начиная от первого учебника Лопиталя и приблизительно до Гурса, но в современных курсах анализа имеется тенденция о них умалчивать, хотя в курсах дифференциальной геометрии им уделено должное место.

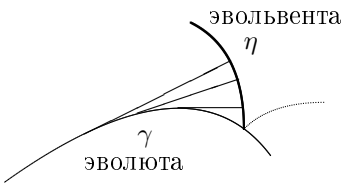


Рис. 3.6: Образование эвольвенты при помощи нити

Пусть плоская кривая задана функцией $y = f(x)$. Дифференциал второго порядка функции равен её отклонению от касательной прямой на бесконечно малом расстоянии от точки дифференцирования, т.е.

$$d^2 f = \alpha\, ds, \text{ или } k = \frac{d^2 f}{ds^2} \qquad (3.129)$$

Таким образом, кривизна кривой k есть отношение её второй и первой дифференциальных форм (см. стр. 117) причём знак кривизны задаётся только второй дифференциальной формой, так как первая дифференциальная форма всегда положительно определена. В соответствии со значением $d^2 f$ *кривизна будет положительной при искривлении линии против часовой стрелки* и *отрицательной при искривлении против часовой стрелки. В точке перегиба кривизна равна нулю* так как $d^2 f = 0$.

Когда линия — прямая, угол α всегда равен нулю, т.е. $k = 0$ во всех точках прямой. Прямая линия имеет бесконечный радиус кривизны. Если кривая — окружность радиуса R, то $\Delta s = \alpha R$ и $k = \frac{1}{R}$ во всех точках окружности. Произвольная линия $y = f(x)$ в разных точках имеет различный радиус кривизны. Запишем $\operatorname{tg}\alpha = \frac{\sqrt{ds^2 - dx^2}}{dx} = f'(x)$ или $ds^2 = (f'^2 + 1)dx^2$. При этом $d^2 f = f''(x)dx^2$, тогда $k \overset{(3.129)}{=} \frac{f''}{f'^2+1}$.

Если кривая не является плоской, то у неё тоже можно построить касательную, соприкасающуюся окружность и соприкасающуюся плоскость. Однако её соприкасающаяся плоскость при переходе от точки к точке меняется. Поэтому в трёхмерном пространстве, кроме кривизны кривой вводится ещё понятие кручения.

Если обозначить через β угол между соприкасающимися плоскостями в двух соседних точках и через Δs длину дуги между этими точками, то величина:

$$\sigma = \lim_{\Delta s \to 0} \frac{\beta}{\Delta s} \qquad (3.130)$$

называется *кручением* данной кривой линии в соответствующей точке.

Касательная к кривой в каждой точке имеет вектор направления ν_s''. Прямая линия, соединяющая центр кривизны пространственной кривой с касательной точкой называется *главной нормалью*. Вектор ν_s'' перпендикулярен вектору ν_s' и направлен по главной нормали. Нормаль, перпендикулярная соприкасающейся плоскости, имеет направление $\nu_s' \times \nu_s''$ и называется *бинормалью*.

Кривизна и кручение кривой в трёхмерном евклидовом пространстве вычисляются по формулам:

$$k = \sqrt{\frac{[\nu_s', \nu_s'']^2}{(\nu_s', \nu_s')^3}}, \quad \sigma = \frac{(\nu_s', \nu_s'', \nu_s''')}{[\nu_s', \nu_s'']^2}; \qquad (3.131)$$

где (ν'_s, ν'_s) — скалярное произведение,
$[\nu'_s, \nu''_s]$ — кососкалярное произведение,
$(\nu'_s, \nu''_s, \nu'''_s)$ — смешанное произведение.

Для того, чтобы вторая из этих формул давала правильно знак кручения, необходимо выбрать систему координат с «правой» ориентацией. Кривизна и кручение суть дифференциальные инварианты относительно изменения параметра s. Это значит, что их величина не изменяется при переходе к новой системе координат. Точки, в которых $\nu'_s \neq 0$ кривая имеет касательную и простирается вдоль касательной в обе стороны. Вблизи точки, где $\nu'_s = 0$, кривая может иметь иное строение; в таких случаях точка называется *нерегулярной* или *особой*.

График гладкой функции двух переменных $z = f(x, y)$ представляет собой двумерную поверхность в трёхмерном пространстве.

Пусть вторая квадратичная форма (3.40) имеет вид:

$$d^2 f = Ldy^2 + 2Mdxdy + Ndx^2, \tag{3.132}$$

где $L = f''_{yy}$, $M = f''_{xy}$, $N = f''_{xx}$,

Если дифференцирование скалярной функции позволяет аппроксимировать в точке дифференцирования её поверхность (график) касательной плоскостью, то повторное дифференцирование позволяет определить особенности этой поверхности в рассматриваемой точке.

Если $d^2 f = 0$, то в этих точках поверхность и касательная плоскость пересекаются, и мы можем записать квадратное уравнение:

$$Lu^2 + 2Mu + N = 0 \tag{3.133}$$

где $u = \frac{dy}{dx}$.

Корни квадратного уравнения (1.40):

$$u_{1,2} = -\frac{M \pm \sqrt{M^2 - LN}}{L} \tag{3.134}$$

задают пересечение поверхности касательной плоскостью в окрестности рассматриваемой точки. Дискриминант уравнения есть гессиан:

$$LN - M^2 = \Delta = \begin{vmatrix} f''_{xx} & f''_{xy} \\ f''_{yx} & f''_{yy} \end{vmatrix}, \text{ где } f''_{xy} = f''_{yx} \tag{3.135}$$

Если $\Delta > 0$, то вторая квадратичная форма (3.132) знакоопределённая при произвольных dx и dy (дифференциал $d^2 f$ знакоопределён), а поверхность в некоторой окрестности рассматриваемой точки касания выпукла и располагается по одну сторону от касательной плоскости. В этом случае точка поверхности называется *эллиптической*.

Если $\Delta < 0$, то форма (3.132) — знакопеременная. Это означает, что поверхность в окрестности точки касания располагается по разные стороны от касательной плоскости. В этом случае пересечение поверхности и касательной плоскости дает две «касательные прямые» $dy = u_1 dx$ и $dy = u_2 dx$.

Такая точка поверхности называется *гиперболической*. В окрестности гиперболической точки поверхность имеет седлообразный характер и называется *минимакс*. В географии минимакс – это *седловина*, которая наблюдается, например, в точке перевала через горный хребет; другой пример — это просто седло. Величины u_1 и u_2 есть функции точки касания, поэтому пересечение поверхности и касательной плоскости образует две пересекающиеся кривые линии.

Если $\Delta = 0$, то поверхность целиком расположена по одну сторону от касательной плоскости, обе касательные совпадают (имеет место вырождение) и поверхность в окрестности точки касания содержит касательную прямую. Такая касательная точка поверхности называется *параболической*. Величина u есть функция точки касания, поэтому пересечение поверхности и плоскости дает кривую, целиком состоящую из параболических точек, которая называется *параболической линией*.

Бывают поверхности, например *сфера*, *эллипсоид* (поверхность, полученная деформацией сферы вдоль взаимно перпендикулярных осей, например $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$), *параболоид вращения* с уравнением $z = x^2 + y^2$ и т.д., у которых все точки эллиптические; такие поверхности выпуклы *в целом*. Бывают поверхности, например *гиперболический параболоид* с уравнением $z = x^2 - y^2$, у которых все точки гиперболические. Однако бывают поверхности, обладающие точками обоих типов; такой является, например, поверхность *тора*, т.е. бублика идеальной формы. В этом случае куски поверхности, заполненные эллиптическими точками, от кусков, заполненных гиперболическими точками, отделяются параболическими линиями. На торе это точки, в которых касательная плоскость перпендикулярна оси тора, заполняют две окружности. Впрочем, бывают поверхности, например *цилиндр* $x^2 + y^2 = const$ или *конус* $x^2 + y^2 = z^2$, сплошь заполненные параболическими точками.

Запишем метрику касательного пространства ds^2 (3.39) в виде:

$$ds^2 = Edy^2 + 2Fdxdy + Gdx^2, \tag{3.136}$$

где $E = g_{yy}$, $F = g_{xy}$, $G = g_{xx}$.

Величина K называется *внешней* или *гауссовой кривизной поверхности* в рассматриваемой точке и выражается соотношением:

$$K = \frac{LN - M^2}{EG - F^2}. \tag{3.137}$$

Как и в случае плоской кривой, частное величин (3.132) и (3.136):

$$k = \frac{1}{R} = \frac{d^2 f}{ds^2}. \tag{3.138}$$

называется *нормальной кривизной поверхности* в направлении s (или кривизной кривой на поверхности функции, задаваемой параметром s).

Величина нормальной кривизны поверхности, как и в случае кривой, равна отношению второй и первой квадратичных форм. Знак кривизны также задаётся второй квадратичной формой.

Как мы видим, кривизны кривой и поверхности очень похожи, однако они различаются тем, что первая измеряется в единственном направлении вдоль собственной кривой, а вторая имеет бесконечное множество значений в зависимости от выбранного направления вдоль поверхности. Поэтому нормальная кривизна поверхности в разных направлениях имеет разные названия. Так максимальные значения нормальной кривизны в данной точке называются *главными кривизнами*, а соответствующие направления s, называются *главными направлениями* поверхности в рассматриваемой точке. Направления s, в которых нормальная кривизна обращается в нуль ($d^2 f = 0$), называются *асимптотическими* (от греч. $\alpha\sigma\upsilon\mu\pi\tau\omega\tau o\varsigma$ — несовпадающий, не касающийся).

Так, например, в гиперболической точке ($\Delta < 0$) касательная плоскость пересекает поверхность по двум асимптотическим направлениям, в эллиптической точке ($\Delta > 0$) асимптотических направлений нет, а в параболической точке ($\Delta = 0$) плоскость касается участка поверхности с одним асимптотическим направлением.

Запишем классификацию точек поверхности ([29, О. А. Платонова] и [22, Е. Е. Ландис], даётся по [8, В. И. Арнольд «Теория катастроф»]):

1. В *эллиптической области* поверхность выпукла, все касательные прямые на касательной плоскости имеют первый порядок в каждой эллиптической точке, вторая квадратичная форма знакоопределённая, а дискриминант и внешняя кривизна положительные.

   Точки, в окрестности которых поверхность не выпукла, имеют касательные выше первого порядка. Они образованы точками пересечения поверхности и касательной плоскости (имеющей нулевую кривизну) и называются *асимптотическими касательными*. Естественно, что в эллиптической области асимптотических линий нет.

2. В *гиперболической области* каждая точка имеет две касательных второго порядка. Внутри области гиперболичности касательные второго порядка в каждой точке задают два направления, двигаясь по которым мы получаем сеть *асимптотических линий* на поверхности в трехмерном пространстве. Через каждую точку области гиперболичности проходят, таким образом, две асимптотические линии и множество неасимптотических. При этом вторая квадратичная форма знакопеременная, а дискриминант и внешняя кривизна отрицательные.

   Две эти области эллиптическую и гиперболическую разделяет общая граница — параболическая линия:

3. *Параболическая линия* имеет в каждой своей точке одну касательную второго порядка (оба асимптотических направления совпадают). Вторая квадратичная форма, дискриминант и внешняя кривизна равны нулю.

4. Рассматривая асимптотические линии внутри области гиперболичности, мы можем найти их точки перегиба. Так как асимптотические

линии проходят через каждую точку области гиперболичности, то их точки перегиба могут образовывать непрерывную особую линию:

Кривую перегиба асимптотических линий. Точки этой линии имеют касательную третьего порядка. Наконец на этой кривой выделены еще особые точки трех типов:

5. *Точка двойного перегиба* (касательная четвертого порядка), т.е. точка, где кривая перегиба в свою очередь имеет точку перегиба.

6. *Перегиб обеих асимптотических линий* (две касательные третьего порядка), т.е. точка самопересечения кривой перегиба, где обе асимптотические линии имеют перегиб.

7. *Общие точки линий* 3 и 4, т.е. точки перегиба параболической линии.

Для поверхностей общего положения в точках 6 происходит пересечение двух ветвей линий перегиба под ненулевым углом, а в точках 7 — касание (первого порядка) линий 3 и 4.

Ясно, что тип точки поверхности не меняется при любой замене координат, т.е. указанная классификация точек поверхности является *геометрически инвариантной* и зависит только от знака гессиана второй квадратичной формы (3.40) в рассматриваемой точке. Это правило справедливо не только для двумерной поверхности в трёхмерном пространстве, а распространяется на пространство любой размерности. Так график гладкой функции n-переменных $y = f(x^1, \ldots, x^n)$ представляет собой n-мерную поверхность в $(n+1)$-мерном пространстве. Если величина гессиана в точке дифференцирования положительна, отрицательна или равна нулю (матрица вырождена), то рассматриваемая точка называется соответственно эллиптической, гиперболической или параболической.

Обратим внимание ещё раз. Параболические поверхности, например, конуса или цилиндра, имеют нулевую кривизну как у плоскости, в отличие от эллиптических поверхностей, например, сферы радиуса r, которая имеет кривизну $\frac{1}{r^2}$. Этот результат отражает тот факт, что из листа бумаги можно свернуть конус или цилиндр, но нельзя получить никакого куска сферы без складок, растяжений или разрывов. Причина заключается в различии *внутренней геометрии* этих поверхностей: никакую часть сферы нельзя *изометрически* (греч. $\iota\sigma o\mu\varepsilon\tau\rho\iota\alpha$ — равномерность) отобразить на плоскость. *Изометрическое отображение* — взаимно однозначное отображение одного метрического пространства в другое сохраняющее расстояние между точками. Если пространства изометричны друг другу, то они оказываются гомеоморфными. *Внутренняя геометрия* — это совокупность фактов, которые могут быть получены при помощи измерений, производимых на самой поверхности, без обращения к объемлющему пространству. Деформация поверхности, при которой её внутренняя геометрия всё время остаётся неизменной, называется *изгибанием* поверхности и возможна, если поверхность физически реализована из гибкого, но нерастяжимого ма-

териала, и деформация происходит так, чтобы не возникало изломов или разрывов.

В достаточно малой окрестности точки дифференцирования гладкая поверхность совпадает, с точностью до величин высшего порядка малости, с касательной плоскостью в этой точке. Поэтому в такой окрестности соотношения длин на поверхности будут такими же, как на плоскости, конечно с точностью до малых величин высшего порядка. Таким образом, в малых областях поверхности имеет место, с точностью до малых величин высшего порядка, евклидова геометрия. Будучи евклидовой в бесконечно малом, в целом рассматриваемая поверхность евклидовой не является и имеет свою внутреннюю геометрию.

Одной из основных особенностей графика непрерывной скалярной функции являются точки *экстремума* (лат. extremum — крайнее) — *локального максимума* или *локального минимума* функции, т.е. точки, в некоторой окрестности которых непрерывная функция имеет верхнюю (нижнюю) грань. Экстремальная точка будет соответственно самой высокой или самой низкой точкой поверхности в рассматриваемой окрестности точки. Экстремум не является геометрическим инвариантом, так как при изменении системы координат (например, при повороте) экстремальная точка может перестать быть таковой.

Необходимым условием существования экстремума в рассматриваемой точке является либо обращение в нуль первой производной (градиента) функции, т.е. $\operatorname{grad} f = 0$ ($\frac{\partial f}{\partial x^i} = 0$ для всех $i = 1, \ldots, n$), либо отсутствие производной (в случае излома графика функции). Например, вершина конуса $x^2 + y^2 = z^2$ есть экстремум $z = 0$, но производная в этой точке не существует.

Достаточное условие существования экстремума сформулировано только для гладких функций и заключается в том, что рассматриваемая точка должна быть эллиптической ($\Delta > 0$), т.е. точка экстремума гладкой функции есть обязательно эллиптическая точка, в которой градиент функции обращается в нуль.

Если функция выпукла вверх и везде, кроме точки касания τ, находится под касательной плоскостью, то в некоторой окрестности τ $d^2 f < 0$ везде, кроме точки касания, и τ — максимум. Если функция выпукла вниз и везде, кроме точки касания, находится над касательной плоскостью, то $d^2 f > 0$ везде, кроме точки касания, и τ — минимум.

Для гладкой функции одного переменного, когда $d^2 f = f''(x_0) dx^2$, экстремум означает перемену знака первой производной. Функция $f(x)$ имеет максимум в некоторой точке x_0, если $f''(x_0) < 0$ (слева от этой точки $f'(x_0) > 0$, а справа — $f'(x_0) < 0$). Функция $f(x)$ имеет минимум в точке x_0 при $f''(x_0) > 0$ (слева $f'(x_0) < 0$, справа $f'(x_0) > 0$).

Кроме понятия экстремума, существует понятие *условного экстремума*. Допустим, мы рассматриваем гладкую кривую в пространстве и касательную к ней в некоторой точке. Если это касательная первого порядка, то говорят, что кривая имеет условный экстремум относительно касательной, так как кривая выпукла относительно этой касательной. В случае гладкой

функции двух переменных любая эллиптическая точка поверхности функции может стать точкой условного экстремума, который можно представить как точку касания рассматриваемой поверхности и касательной плоскости. Таким образом условный экстремум существует только у гладких функций и определяется кривизной поверхности в рассматриваемой точке. Поэтому в отличие от безусловного локального экстремума условный локальный экстремум уже не зависит от системы координат, т.е. *геометрически инвариантен*.

Задачи на условный экстремум возникают во многих вопросах механики, геометрии (например, отыскание прямоугольника наименьшего периметра, имеющего заданную площадь) и т.д.

Задача на условный экстремум есть задача отыскания нужной точки на вложенном многообразии (см. стр. 133) или на гладкой поверхности в пространстве, заданной при помощи геометрических связей, которые в случае условного экстремума называются *условиями*.

Например, точка касания кривой и касательной первого порядка является точкой условного экстремума, а касательная в этом случае называется условием условного экстремума.

Пусть M есть гладкая поверхность в пространстве $\mathbb{R}^n$, размерности $m = n - k$, и заданы уравнения k геометрических связей (условий) $f_1 = const, \ldots, f_k = const$. При этом векторы $\operatorname{grad} f_1, \ldots, \operatorname{grad} f_k$ линейно независимы. Тогда метод решения задачи на условный экстремум координатной функции $y = f(x^1, \ldots, x^n)$ при заданных условиях, заключается в сведении этой задачи к задаче на безусловный экстремум вспомогательной функции — так называемой *функции Лагранжа*, которая имеет вид: $L = f + \lambda_1 f_1 + \ldots + \lambda_k f_k$, где $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ — множители Лагранжа.

Необходимое условие экстремума функции Лагранжа в рассматриваемой точке записывается в виде $\frac{\partial L}{\partial x^i} = 0$ для всех $i = 1, \ldots, n$ или $\operatorname{grad} L = (0, \ldots, 0)$ при линейно независимых $\operatorname{grad} f_1, \ldots, \operatorname{grad} f_k$.

Пример. Для гладкой функции f двух переменных в трёхмерном пространстве при наличии только одной связи $g = const$ рассматриваемые поверхности касаются в точке условного экстремума. Их градиенты в этой точке противоположны (или совпадают) по направлению и отличаются по длине на множитель λ.

Достаточное условие экстремума функции Лагранжа выписывается в виде квадратичной формы $d^2L = \sum_{i,j=1}^{n} L''_{x^i x^j} dx^i dx^j$. Точка условного экстремума есть эллиптическая точка. Это означает, что нам достаточно проверить знакоопределённость квадратичной формы (3.40) для исходной поверхности $d^2 f = \sum_{i,j=1}^{n} f''_{x^i x^j} dx^i dx^j$ и не обязательно вычислять квадратичную форму для функции Лагранжа.

Рассмотрим график x, y, f функции двух переменных, представляющий собой двумерную поверхность в трехмерном пространстве. Допустим, на поверхности есть *излом*, проходящий вдоль некоторой линии. Линия излома $\alpha\beta$, разделяющая две гладких области, называется *пограничной кривой*, на которой частные производные функции имеют *разрыв первого рода*. (рис. 3.7).

Линия $y = const$ образующаяся на поверхности графика при сечении этой поверхности плоскостью $y = const$, имеет излом в точке τ. Следовательно, производная $\left(\frac{\partial f}{\partial x}\right)_y$ меняется в этой точке скачком. Очевидно (сделайте сечения поверхности плоскостями $x = const$ и $f = const$), что скачком в этой точке изменяются также производные $\left(\frac{\partial f}{\partial y}\right)_x$ и $\left(\frac{\partial y}{\partial x}\right)_f$.

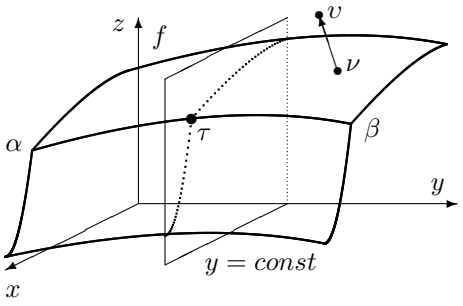


Рис. 3.7: Пограничная кривая

Таким образом, по одну сторону пограничной кривой частные производные функции имеют одно значение, а по другую сторону — значение производных скачком меняется. Следовательно, ввиду того, что частные производные на изломе функции $f(x, y)$ имеют разрыв, касательную плоскость ни в одной точке на пограничной кривой $\alpha\beta$ мы построить не сможем, т.е. функция на этой кривой не дифференцируема.

Получим теперь дифференциальные соотношения, определяющие величину скачка функции при переходе через пограничную кривую.

Если пограничная кривая гладкая, то она задаётся гладкой функцией $f(x(s), y(s))$, где s — параметр кривой. Тогда у этой не дифференцируемой на изломе функции $f(x, y)$ при её изменении *вдоль гладкой пограничной кривой* может быть вычислен дифференциал df (прямая линия, касательная к кривой), причём двумя равноправными путями. Если частные производные, взятые с одной стороны пограничной кривой, записать с индексом (1), а с другой стороны с индексом (2) (на рис. 3.7 либо сверху от кривой, либо снизу), то соотношение для дифференциала функции вдоль гладкой пограничной кривой примет вид:

$$df = f'_s\, ds = \left(\frac{\partial f}{\partial x}\right)_y^{(1)} dx + \left(\frac{\partial f}{\partial y}\right)_x^{(1)} dy = \left(\frac{\partial f}{\partial x}\right)_y^{(2)} dx + \left(\frac{\partial f}{\partial y}\right)_x^{(2)} dy.$$

Отнимая одно уравнение от другого, получим:

$$\left(\frac{\partial f}{\partial x}\right)_y^{(1)} - \left(\frac{\partial f}{\partial x}\right)_y^{(2)} = \left[\left(\frac{\partial f}{\partial y}\right)_x^{(2)} - \left(\frac{\partial f}{\partial y}\right)_x^{(1)}\right] \frac{dy}{dx}. \tag{3.139}$$

Соотношение (3.139) определяет величину скачка производной $\left(\frac{\partial f}{\partial x}\right)_y$ при переходе через гладкую пограничную кривую.

Обратим внимание ещё раз. Здесь $\left(\frac{\partial f}{\partial x}\right)_y^{(1)}$ и $\left(\frac{\partial f}{\partial x}\right)_y^{(2)}$ — частные производные, взятые в точке τ при пересечении линии $y = const$ с пограничной кривой, причём первая из них взята с одной стороны кривой, вторая — с

другой стороны; $\left(\frac{\partial f}{\partial y}\right)_x^{(1)}$ и $\left(\frac{\partial f}{\partial y}\right)_x^{(2)}$ — частные производные, взятые в точке τ при пересечении линии $x = const$ с пограничной кривой, по одну и по другую сторону от этой кривой; $\frac{dy}{dx}$ — *полная производная* функции $y(x)$ вдоль пограничной кривой, df — *полный дифференциал* функции f вдоль пограничной кривой.

Допустим теперь, что в пространстве x, y, z задано скалярное поле $u = u(x, y, z)$, которое имеет (или не имеет) излом или разрыв на двумерной поверхности (x, y, f), т.е. на графике скалярной функции двух переменных в трехмерном пространстве. Для выяснения этого обстоятельства мы должны рассмотреть *производную по нормали* или *нормальную производную* скалярного поля u к поверхности (x, y, f). Рассмотрим нормаль к поверхности (x, y, f) в точке ν (рис. 3.7), которая есть вектор $\mathbf{n} = \upsilon - \nu$, т.е. υ, как и ν — точка на векторе нормали к поверхности. Длина нормали $n = |\upsilon - \nu|$ есть растояние между точками υ и ν.

Производная по нормали от u в точке ν имеет вид:

$$\frac{\partial u}{\partial n} = \lim_{\upsilon \to \nu} \frac{u(\upsilon) - u(\nu)}{|\upsilon - \nu|}. \tag{3.140}$$

Таким образом, производная по нормали есть производная по направлению (3.101) нормали. Если $\mathbf{n}$ — единичный вектор нормали, то

$$\frac{\partial u}{\partial n} = \mathbf{n} \cdot \operatorname{grad} u = \frac{\partial u}{\partial x}\cos(nx) + \frac{\partial u}{\partial y}\cos(ny) + \frac{\partial u}{\partial z}\cos(nz), \tag{3.141}$$

где (nx), (ny), (nz) — углы между вектором нормали и соответствующими осями координат.

Само собой, что можно рассмотреть нормаль с каждой стороны поверхности (для выпуклой поверхности говорят — внешнюю и внутреннюю нормаль). Если производные по нормалям с двух сторон поверхности не равны

$$\frac{\partial u}{\partial n}^{(2)} \neq \frac{\partial u}{\partial n}^{(1)} \tag{3.142}$$

то функция u имеет излом на поверхности (x, y, f), а её частные производные имеют разрыв на этой поверхности.

3.7 Преобразования производных

Рассмотрим важный частный случай вычисления неопределённых выражений типа $\lim\limits_{x \to x_0} \frac{f(x)}{g(x)}$, где $\lim\limits_{x \to x_0} f(x) = \lim\limits_{x \to x_0} g(x) = 0$. Основным инструментом раскрытия таких неопределённостей $\frac{0}{0}$ служит ряд Тейлора, с помощью которого выделяется главная часть функции.

Функции f и g представляют в виде ряда окрестности точки x_0 (если это возможно) до первого не равного нулю члена

$$f(x) = a(x - x_0)^n + o\left((x - x_0)^n\right), \qquad a \neq 0;$$
$$g(x) = b(x - x_0)^m + o\left((x - x_0)^m\right), \qquad b \neq 0.$$

В результате получаем

$$\lim_{x\to x_0}\frac{f(x)}{g(x)}=\frac{a}{b}\lim_{x\to x_0}(x-x_0)^{n-m}=\begin{cases}0, & \text{если } n>m,\\ \dfrac{a}{b}, & \text{если } n=m,\\ \infty, & \text{если } n<m.\end{cases}\tag{3.143}$$

Другим общим методом раскрытия таких неопределённостей и неопределённостей, сводимых к ним, является правило Лопиталя.

Теорема Коши (*обобщённая формула конечных приращений*). Пусть в промежутке $[a,b]$ заданы две функции $f(x)$ и $g(x)$, имеющие всюду на $[a,b]$ производные $f'(x)$ и $g'(x)$, причём $g'(x)$ не обращается в нуль ни в одной точке открытого промежутка (a,b). В таком случае между a и b существует такая точка c, $a<c<b$, что

$$\frac{f(b)-f(a)}{g(b)-g(a)}=\frac{f'(c)}{g'(c)}.\tag{3.144}$$

Правило Лопиталя. Пусть функции $f(x)$ и $g(x)$ удовлетворяют всем условиям теоремы Коши, и, кроме того, $f(a)=g(a)=0$. Если существует (конечный или бесконечный) предел A отношения их производных, то он равен пределу отношения функций:

$$A=\lim_{x\to a}\frac{f'(x)}{g'(x)}=\lim_{x\to a}\frac{f(x)}{g(x)}.\tag{3.145}$$

Пример. Пусть $\varphi(\nu)$ — гладкая функция, а $\nu(t)$ — гладкая кривая. Тогда $\partial\varphi=\partial\nu=0$ в точке дифференцирования. По правилу Лопиталя получаем

$$\frac{\partial\dot\varphi}{\partial\dot\nu}=\frac{\partial\varphi}{\partial\nu}\quad\text{или}\quad\frac{\partial\dot\varphi}{\partial\varphi}=\frac{\partial\dot\nu}{\partial\nu}.\tag{3.146}$$

Однородная функция, это функция $f\colon\mathbb{R}^n\to\mathbb{R}$ такая, что для любого $\nu\in\mathbb{R}^n$ и $\lambda\in\mathbb{R}$ выполняется равенство

$$f(\lambda\nu)=\lambda^r f(\nu),\tag{3.147}$$

где r называют *порядком однородности* или *степенью однородности*.

Различают *положительно однородные функции*, для которых равенство (3.147) выполняется только при $\lambda>0$, и а*бсолютно однородные функции*, для которых выполняется равенство $f(\lambda\nu)=|\lambda|^r f(\nu)$.

Свойства однородной функции:

1. Если функция f является многочленом от n переменных, то она будет однородной функцией степени r в том и только в том случае, когда f — однородный многочлен степени r (см. стр. 26), в частности в этом случае r должно быть целым.

   Пример: Функция $f(x,y)=ax^2+bxy+cy^2$ однородная степени $r=2$, так как $f(\lambda x,\lambda y)=\lambda^2ax^2+\lambda^2bxy+\lambda^2cy^2=\lambda^2f(x,y)$.

2. Однородная функция в нуле равна нулю, если она там определена $f(\mathbf{0}) = 0$.

3. *Лемма Эйлера.* Однородные функции пропорциональны скалярному произведению своего градиента на вектор своих переменных с коэффициентом равным порядку однородности r:

$$\nabla f \cdot \nu = rf. \tag{3.148}$$

Лемма доказывается дифференцированием (3.147) по λ при $\lambda = 1$:

$$\frac{df(\lambda\nu)}{d\lambda} = \nabla f \cdot \nu = r\lambda^{r-1} f.$$

Пример: Пусть $\nu = (x^1, \ldots, x^n)$, $f(\nu) = \sum_{i,j=1}^{n} \alpha_{ij} x^i x^j = (\nu, \alpha\nu)$ — однородная функция степени $r = 2$, тогда

$$\nabla f \cdot \nu = 2f. \tag{3.149}$$

Очень важное преобразование производных и дифференциалов имеет специальное название *преобразование Лежандра* — вспомогательный математический приём, состоящий в переходе от функции на линейном пространстве к функции на сопряжённом пространстве.

Пусть $\xi \in V^*$ — 1-форма (ковектор), $\eta \in V$ — вектор, $(\xi \cdot \eta)$ — скалярное произведение вектора и 1-формы. Тогда запишем:

$$d(\xi \cdot \eta) = \eta\, d\xi + \xi\, d\eta, \quad (\xi \cdot \eta)' = \eta \cdot \xi'_s + \xi \cdot \eta'_s. \tag{3.150}$$

В одномерном случае получаем:

$$d(xy) = y\, dx + x\, dy, \quad (xy)' = yx' + xy'. \tag{3.151}$$

Записанные выражения носят название *правило Лейбница.*

Пусть теперь $z = f(\xi)$, где $\xi \in V$, — функция на линейном пространстве, и $w = g(\eta)$, где $\eta \in V^*$, — функция на сопряжённом пространстве. Найдем условия, при которых справедливо *преобразование Лежандра*:

$$(\xi \cdot \eta) = f(\xi) + g(\eta) \quad \text{или} \quad d(\xi \cdot \eta) = f'_\xi\, d\xi + g'_\eta\, d\eta. \tag{3.152}$$

Исходя из выражения $d(\xi \cdot \eta) = \eta\, d\xi + \xi\, d\eta$, получаем:

$$\xi = \frac{\partial g}{\partial \eta} = g'_\eta, \quad \eta = \frac{\partial f}{\partial \xi} = f'_\xi. \tag{3.153}$$

Таким образом,

$$\text{если} \quad \xi = g'_\eta \quad \text{и} \quad \eta = f'_\xi, \quad \text{то} \quad (\xi \cdot \eta) = f(\xi) + g(\eta).$$

Геометрический смысл преобразования Лежандра состоит в переходе от поверхности (графика) функции $z = f(\xi)$ в $(n+1)$-мерном пространстве к описанию её как огибающей n-параметрического семейства касательных плоскостей $z = (\xi \cdot \eta) - g(\eta)$, где компоненты 1-формы η — параметры семейства.

Уравнения $\eta = f'_\xi$ разрешимы, т.е. преобразование Лежандра существует, если $\det f''_{\xi\xi} \neq 0$. Если $\det f''_{\xi\xi} > 0$, то оно переводит выпуклую функцию f в выпуклую функцию g. В этом случае преобразование Лежандра инволютивно: применённое повторно к g, оно даёт f.

Неопределённости вида $\frac{0}{0}$ (без обозначений) исследовал Г. Лопиталь (1696), символ $\frac{0}{0}$ ввёл И. Бернулли (1730). Правило Лопиталя найдено И. Бернулли и сообщено им Г. Лопиталю, опубликовавшему это правило в 1696. Преобразование Лежандра введено А. Лежандром (1789).

Преобразование второй производной:

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) \overset{(3.10)}{=} \frac{d}{dx}\left(\frac{dx}{dy}\right)^{-1} \overset{(3.5)}{=} \frac{d}{dy}\left(\frac{dx}{dy}\right)^{-1}\frac{dy}{dx} = \overset{(3.9)}{=} -\frac{d^2x}{dy^2}\left(\frac{dy}{dx}\right)^3. \tag{3.154}$$

Преобразования частных производных и якобианов.

Частные производные имеют особенности по сравнению с производными функции одного переменного. Если производная функции определена однозначно, то частная производная вычисляется при условии постоянства величины, которая стоит в виде индекса. При разных условиях (разных индексах) частная производная будет иметь, разумеется, разное значение:

$$\left(\frac{\partial f}{\partial x}\right)_y \neq \left(\frac{\partial f}{\partial x}\right)_z. \tag{3.155}$$

Можно возразить, что при вычислении частной производной по одному переменному все остальные переменные закреплены, и поэтому вычисление однозначно. Это справедливо, но только при условии, что мы правильно определили и закрепили все переменные, от которых зависит наша функция. Если же это не так, то частная производная вычисляется при условии закрепления только тех переменных, которые указаны в индексах. Причем это не обязательно могут быть независимые переменные, это могут быть и функции от переменных.

Аналогично (3.154) можно показать, что

$$\left(\frac{\partial^2 y}{\partial x^2}\right)_z = -\left(\frac{\partial^2 x}{\partial y^2}\right)_z \left(\frac{\partial y}{\partial x}\right)^3_z. \tag{3.156}$$

С учётом теоремы об обратных величинах и правила дифференцирования сложной функции запишем

$$\left(\frac{\partial y}{\partial x}\right)_z \overset{(3.10)}{=} \frac{1}{\left(\frac{\partial x}{\partial y}\right)_z}, \quad \left(\frac{\partial f}{\partial x}\right)_z \overset{(3.5)}{=} \left(\frac{\partial f}{\partial y}\right)_z \left(\frac{\partial y}{\partial x}\right)_z \tag{3.157}$$

Тогда получаем:

$$\left(\frac{\partial f}{\partial x}\right)_z \left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial f}\right)_z = 1. \tag{3.158}$$

С учётом записи для дифференциала

$$df \overset{(3.15)}{=} \left(\frac{\partial f}{\partial x}\right)_y dx + \left(\frac{\partial f}{\partial y}\right)_x dy$$

функции двух переменных получаем:

$$\left(\frac{\partial f}{\partial z}\right)_u = \left(\frac{\partial f}{\partial x}\right)_y \left(\frac{\partial x}{\partial z}\right)_u + \left(\frac{\partial f}{\partial y}\right)_x \left(\frac{\partial y}{\partial z}\right)_u. \tag{3.159}$$

Если принять $z = x$, то получаем:

$$\left(\frac{\partial f}{\partial x}\right)_u = \left(\frac{\partial f}{\partial x}\right)_y + \left(\frac{\partial f}{\partial y}\right)_x \left(\frac{\partial y}{\partial x}\right)_u \tag{3.160}$$

Если теперь принять $u = f$, то получаем:

$$\left(\frac{\partial f}{\partial x}\right)_y = -\left(\frac{\partial f}{\partial y}\right)_x \left(\frac{\partial y}{\partial x}\right)_f \text{ или } \left(\frac{\partial f}{\partial x}\right)_y \left(\frac{\partial x}{\partial y}\right)_f \left(\frac{\partial y}{\partial f}\right)_x = -1. \tag{3.161}$$

Заметим, что выражения (3.158) и (3.161) очень похожи, но вычисляются при разных индексах и дают разный результат.

Запишем якобиан двух функций $v = f(x, y)$ и $w = g(x, y)$:

$$\frac{\partial(f, g)}{\partial(x, y)} = \left(\frac{\partial f}{\partial x}\right)_y \left(\frac{\partial g}{\partial y}\right)_x - \left(\frac{\partial f}{\partial y}\right)_x \left(\frac{\partial g}{\partial x}\right)_y. \tag{3.162}$$

Якобианы обладают следующими основными свойствами:

$$\frac{\partial(f, g)}{\partial(x, y)} = -\frac{\partial(g, f)}{\partial(x, y)} = -\frac{\partial(f, g)}{\partial(y, x)}; \tag{3.163}$$

$$\left(\frac{\partial f}{\partial x}\right)_y = \frac{\partial(f, y)}{\partial(x, y)} = \frac{\partial(y, f)}{\partial(y, x)}; \tag{3.164}$$

$$\frac{\partial(f, g)}{\partial(x, y)} = \frac{\partial(f, g)}{\partial(z, u)} \frac{\partial(z, u)}{\partial(x, y)}; \tag{3.165}$$

$$\frac{\partial(x, y)}{\partial(x, y)} = 1; \quad \frac{\partial(f, f)}{\partial(x, y)} = 0; \quad \frac{\partial(f, c)}{\partial(x, y)} = 0, \text{ если } c = const. \tag{3.166}$$

Использование соотношений (3.163–3.166) позволяет достаточно просто производить преобразования частных производных.

Докажем, например, соотношение (3.161):

$$\left(\frac{\partial x}{\partial y}\right)_f \overset{(3.164)}{=} \frac{\partial(x,f)}{\partial(y,f)} \overset{(3.163)}{=} -\frac{\partial(f,x)}{\partial(y,f)} \overset{(3.165)}{=} -\frac{\partial(f,x)}{\partial(y,x)}\frac{\partial(y,x)}{\partial(y,f)} = \overset{(3.164)}{=} -\left(\frac{\partial f}{\partial y}\right)_x \left(\frac{\partial x}{\partial f}\right)_y.$$

Правила вычисления дифференциальных операторов.

Как мы уже видели (3.17, 3.71, 3.72) можно производить непосредственное вычисление дифференциальных операторов. Например, прямым вычислением можно получить следующие выражения для радиус-вектора $\mathbf{r}$ и постоянного векторного поля $\mathbf{a} = const$:

$$\operatorname{grad}(\mathbf{a}\cdot\mathbf{r}) = \mathbf{a}, \tag{3.167}$$

$$\operatorname{rot}(\mathbf{a}\times\mathbf{r}) = 2\mathbf{a}, \tag{3.168}$$

$$\operatorname{rot}\frac{\mathbf{a}\times\mathbf{r}}{r^3} = \frac{\mathbf{a}}{r^3} - \frac{3(\mathbf{a}\cdot\mathbf{r})\mathbf{r}}{r^5} = \operatorname{grad}\left(\frac{\mathbf{a}\cdot\mathbf{r}}{r^3}\right). \tag{3.169}$$

Доказательство (3.168) для $\mathbf{a} = (a_x, a_y, a_x)$

$$\mathbf{a}\times\mathbf{r} \overset{(2.106)}{=} e_x(a_y z - a_z y) + e_y(a_z x - a_x z) + e_z(a_x y - a_y x),$$

$$\operatorname{rot}_x(\mathbf{a}\times\mathbf{r}) \overset{(3.63)}{=} \frac{\partial}{\partial y}(a_x y - a_y x) - \frac{\partial}{\partial z}(a_z x - a_x z) = 2a_x.$$

Если скалярная функция g является сложной функцией нескольких скаляров $f_1, \ldots, f_n$, которые сами представляют собой функции координат, то имеет место формула:

$$\operatorname{grad} g(f) = \nabla g(f) = g'_f \nabla f = g'_f(f)\operatorname{grad} f, \tag{3.170}$$

$$\operatorname{grad} g(f_1, \ldots, f_n) = \frac{\partial g}{\partial f_1}\operatorname{grad} f_1 + \ldots + \frac{\partial g}{\partial f_n}\operatorname{grad} f_n. \tag{3.171}$$

В частности, например, для радиус-вектора $\mathbf{r}$:

$$\operatorname{grad}\left(\frac{1}{r}\right) \overset{(3.170)}{=} \frac{\partial}{\partial r}\left(\frac{1}{r}\right)\operatorname{grad} r \overset{\text{стр. 106}}{=} -\frac{\mathbf{r}}{r^3}, \tag{3.172}$$

$$\operatorname{grad}\left(\frac{1}{r^3}\right) \overset{(3.170)}{=} \frac{\partial}{\partial r}\left(\frac{1}{r^3}\right)\operatorname{grad} r \overset{\text{стр. 106}}{=} -\frac{3\mathbf{r}}{r^5}. \tag{3.173}$$

Дифференцирование суммы скалярных или векторных функций происходит по обычному правилу дифференцирования:

$$\operatorname{grad}(f+g) = \nabla(f+g) = \nabla f + \nabla g = \operatorname{grad} f + \operatorname{grad} g, \tag{3.174}$$

$$\operatorname{div}(\mathbf{a}+\mathbf{b}) = \nabla(\mathbf{a}+\mathbf{b}) = \nabla\mathbf{a} + \nabla\mathbf{b} = \operatorname{div}\mathbf{a} + \operatorname{div}\mathbf{b}, \tag{3.175}$$

$$\operatorname{rot}(\mathbf{a}+\mathbf{b}) = \nabla\times\mathbf{a} + \nabla\times\mathbf{b} = \operatorname{rot}\mathbf{a} + \operatorname{rot}\mathbf{b}. \tag{3.176}$$

Дифференцирование произведения функций происходит по правилу дифференцирования сложной функции:

$$\operatorname{grad}(fg) = \nabla(fg) = f\nabla g + g\nabla f = f\operatorname{grad} g + g\operatorname{grad} f, \tag{3.177}$$
$$\operatorname{div}(f\mathbf{a}) = \nabla(f\mathbf{a}) = f\nabla\mathbf{a} + \mathbf{a}\nabla f = f\operatorname{div}\mathbf{a} + \mathbf{a}\cdot\operatorname{grad} f, \tag{3.178}$$
$$\operatorname{rot}(f\mathbf{a}) = \nabla\times(f\mathbf{a}) = f\nabla\times\mathbf{a} + (\nabla f)\times\mathbf{a} = f\operatorname{rot}\mathbf{a} + \operatorname{grad} f\times\mathbf{a}, \tag{3.179}$$
$$\operatorname{grad}(\mathbf{a}\cdot\mathbf{b}) = \nabla(\mathbf{a}\cdot\mathbf{b}) = (\mathbf{b}\nabla)\,\mathbf{a} + (\mathbf{a}\nabla)\,\mathbf{b} + \mathbf{b}\times\operatorname{rot}\mathbf{a} + \mathbf{a}\times\operatorname{rot}\mathbf{b}, \tag{3.180}$$
$$\operatorname{div}(\mathbf{a}\times\mathbf{b}) = \nabla(\mathbf{a}\times\mathbf{b}) = \mathbf{b}(\nabla\times\mathbf{a}) - \mathbf{a}(\nabla\times\mathbf{b}) = \mathbf{b}\operatorname{rot}\mathbf{a} - \mathbf{a}\operatorname{rot}\mathbf{b}, \tag{3.181}$$
$$\operatorname{rot}(\mathbf{a}\times\mathbf{b}) = \nabla\times(\mathbf{a}\times\mathbf{b}) = (\mathbf{b}\nabla)\,\mathbf{a} - (\mathbf{a}\nabla)\,\mathbf{b} + \mathbf{a}\operatorname{div}\mathbf{b} - \mathbf{b}\operatorname{div}\mathbf{a}. \tag{3.182}$$

Например:

$$\operatorname{rot}\left(\frac{\mathbf{a}}{r}\right) = \frac{\mathbf{a}\times\mathbf{r}}{r^3}, \quad (\mathbf{a} = const) \tag{3.183}$$
$$\operatorname{grad}\left(\frac{f}{r}\right) = \frac{\operatorname{grad} f}{r} - f\frac{\mathbf{r}}{r^3}. \tag{3.184}$$

Операция скалярного оператора (3.122) $\mathbf{a}\nabla$ над произвольным скаляром f даёт скаляр:

$$(\mathbf{a}\nabla)f = a_x\frac{\partial f}{\partial x} + a_y\frac{\partial f}{\partial y} + a_z\frac{\partial f}{\partial z}. \tag{3.185}$$

Операция $\mathbf{a}\nabla$ над произвольным вектором $\mathbf{b} = (b_x, b_y, b_z)$ даёт вектор:

$$(\mathbf{a}\nabla)\cdot\mathbf{b} = a_x\frac{\partial\mathbf{b}}{\partial x} + a_y\frac{\partial\mathbf{b}}{\partial y} + a_z\frac{\partial\mathbf{b}}{\partial z}. \tag{3.186}$$

компонента которого, например, по оси x равна:

$$((\mathbf{a}\nabla)\cdot\mathbf{b})_x = a_x\frac{\partial b_x}{\partial x} + a_y\frac{\partial b_y}{\partial y} + a_z\frac{\partial b_z}{\partial z} = (\mathbf{a}\nabla)b_x. \tag{3.187}$$

Вторые производные вычисляются по следующим правилам:

$$\operatorname{div}(\operatorname{grad} f) = \nabla^2 f \overset{(3.79)}{=} \Delta f, \tag{3.188}$$
$$\operatorname{rot}(\operatorname{grad} f) = \nabla\times(\nabla f) \overset{(3.73)}{=} 0, \tag{3.189}$$
$$\operatorname{div}(\operatorname{rot}\mathbf{a}) = \nabla(\nabla\times\mathbf{a}) \overset{(3.74)}{=} 0, \tag{3.190}$$
$$\operatorname{grad}\operatorname{div}\mathbf{a} = \nabla(\nabla\mathbf{a}) = \nabla\times\nabla\times\mathbf{a} + \nabla^2\mathbf{a} = \operatorname{rot}\operatorname{rot}\mathbf{a} + \nabla^2\mathbf{a}, \tag{3.191}$$
$$\operatorname{rot}\operatorname{rot}\mathbf{a} = \nabla\times\nabla\times\mathbf{a} = \nabla(\nabla\mathbf{a}) - \nabla^2\mathbf{a} = \operatorname{grad}\operatorname{div}\mathbf{a} - \nabla^2\mathbf{a}. \tag{3.192}$$

Например:

$$\Delta\left(\frac{1}{r}\right) = 0, \tag{3.193}$$
$$\Delta\left(\frac{f}{r}\right) = \frac{\Delta f}{r} - 2\frac{\mathbf{r}\cdot\operatorname{grad} f}{r^3}. \tag{3.194}$$

Доказательство (3.193)

$$\operatorname{div}(\operatorname{grad})\left(\frac{1}{r}\right) \overset{(3.172)}{=} -\operatorname{div}\left(\frac{\mathbf{r}}{r^3}\right) \overset{\substack{(3.72)\\(3.178)}}{=} -3\frac{1}{r^3} + \mathbf{r}\frac{3\mathbf{r}}{r^5} = 0.$$

Выражение $\operatorname{grad}\operatorname{div}\mathbf{a} = \nabla(\nabla\mathbf{a})$ есть вектор:

$$\nabla(\nabla\mathbf{a}) = \left(e_x\frac{\partial}{\partial x} + e_y\frac{\partial}{\partial y} + e_z\frac{\partial}{\partial z}\right)\left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}\right). \tag{3.195}$$

Выражение $\nabla^2\mathbf{a}$ также вектор, но отличный от предыдущего и похожий на лапласиан (3.188):

$$\nabla^2\mathbf{a} = (\nabla\nabla)\mathbf{a} = \frac{\partial^2\mathbf{a}}{\partial x^2} + \frac{\partial^2\mathbf{a}}{\partial y^2} + \frac{\partial^2\mathbf{a}}{\partial z^2}, \tag{3.196}$$

слагающая которого, например, по оси x равна

$$\left(\nabla^2\mathbf{a}\right)_x = \nabla^2 a_x = \frac{\partial^2 a_x}{\partial x^2} + \frac{\partial^2 a_x}{\partial y^2} + \frac{\partial^2 a_x}{\partial z^2}. \tag{3.197}$$

Глава 4

Интегрирование

Интегрирование — действие, обратное дифференцированию, т.е. восстановления искомой функции по её производной.

Пример. Если мы восстанавливаем функцию одного действительного переменного $F(x)$ по имеющейся *непрерывной* производной $f(x)$:

$$f(x) = \frac{dF(x)}{dx}, \quad F(x) = \int f(x)\,dx + c, \tag{4.1}$$

то говорят, что мы производим *операцию интегрирования*, обозначаемую значком $\int$. Значение функции $F(x)$ можно вычислить только с точностью до произвольного постоянного числа c, так как производная от любого постоянного числа всегда равна нулю (см. стр. 106).

Функция $F(x)$ называется *первообразной* функции $f(x)$, а выражение $\int f(x)\,dx + c$ — *неопределённым интегралом*.

Из взаимно обратного характера операций интегрирования и дифференцирования следует возможность получения из формул дифференцирования (см. стр. 106) соответствующих формул интегрирования:

1. $\int x^n\,dx = \frac{x^{n+1}}{n+1} + c$,
2. $\int e^x\,dx = e^x + c$,
3. $\int \frac{dx}{x} = \ln x + c$,
4. $\int \cos x\,dx = \sin x + c$,
5. $\int \sin x\,dx = -\cos x + c$.

Одна из основных теорем интегрального исчисления устанавливает, что каждая непрерывная функция действительного переменного имеет «неопределенный интеграл» или «первообразную». Но в отличие от производной элементарной функции первообразная элементарной функции не всегда является элементарной функцией. Поэтому если такие приёмы как разложение функций в ряд, правило дифференцирования сложной функции и т.д.,

позволяют вычислить производную практически любой функции, то для вычисления интеграла в аналитическом виде таких универсальных методов не существует. Вычисление интеграла известной функции называется вычислением «квадратуры». Но даже в тех случаях, когда интеграл выражается через элементарные функции (т.е., как говорят, берётся в конечном виде или в квадратурах), нет единых рецептов, которые позволяли бы найти такое выражение, т.е. записать результат через элементарные функции (*в квадратурах*).

В отличии от теории дифференциального исчисления теория интегрального исчисления представляет собой, во-первых, набор доказательств об интегрируемости (существовании интеграла) различных классов функций, а, во-вторых, набор приемов по отысканию интеграла различных классов функций. Часть этих доказательств и приёмов мы и рассмотрим в настоящей главе.

Пример. Приём вычисления неопределенных интегралов при помощи функции комплексного переменного. Запишем очевидное уравнение

$$\int e^{(a+ib)x} dx = \frac{e^{(a+ib)x}}{a+ib} + C_1 + iC_2, \quad (a, b \in \mathbb{R}).$$

Замечаем, что:

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

Приравнивая действительные и мнимые части, получаем

$$\int e^{ax} \cos(bx)\, dx = \frac{e^{ax}}{a^2+b^2}(a\cos(bx) + b\sin(bx)) + C_1, \tag{4.2}$$

$$\int e^{ax} \sin(bx)\, dx = \frac{e^{ax}}{a^2+b^2}(a\sin(bx) - b\cos(bx)) + C_2. \tag{4.3}$$

Интеграл — одно из важнейших понятий математики, возникшее в связи с потребностью, с одной стороны, отыскивать функции по их производным, а с другой — вычислять площади, объемы, длины дуг, работу сил и т.п. Слово «интеграл» впервые в печати употребил Я. Бернулли (1690). Возможно, термин образован от латинского integer — целый. По другому предположению, Я. Бернулли произвел термин от латинского integro — приводить в прежнее состояние, восстанавливать.

4.1 Интеграл

Разбиением R замкнутого интервала $[a, b]$ называется последовательность $t_0, t_1, \ldots, t_N$, где $a = t_0 < t_1 < \ldots < t_N = b$. Разбиение R делит интервал $[a, b]$ на N интервалов.

Разбиение параллелепипеда $P = [a_1, b_1] \times \ldots \times [a_n, b_n]$ определяется как семейство $R = (R_1, \ldots, R_n)$, где каждое R_i есть разбиение отрезка $[a_i, b_i]$.

Если R_i делит $[a_i, b_i]$ на N_i интервалов, то $R = (R_1, \ldots, R_n)$ делит $[a_1, b_1] \times \ldots \times [a_n, b_n]$ на $N = N_1 \cdot \ldots \cdot N_n$ параллелепипедов. Мы будем называть их *параллелепипедами S разбиения R*.

Рассмотрим не обязательно непрерывную, но ограниченную функцию $y = f(x^1, \ldots, x^n)$ или $f\colon P \to \mathbb{R}$, где P — замкнутый параллелепипед в $\mathbb{R}^n$, а R — его разбиение. Для каждого параллелепипеда S разбиения R определим верхнюю и нижнюю грани:

$$M_S(f) \overset{(1.44)}{=} \sup\{\, f(\nu)\colon \nu \in S\,\}, \quad m_S(f) \overset{(1.45)}{=} \inf\{\, f(\nu)\colon \nu \in S\,\}. \tag{4.4}$$

Определение. Пусть $V(S)$ — объём S, тогда *нижняя* L (от англ. lower — нижний) и *верхняя* U (от англ. upper — верхний) *суммы* f для R определяются формулами:

$$L(f, R) = \sum_S m_S(f) \cdot V(S), \quad U(f, R) = \sum_S M_S(f) \cdot V(S). \tag{4.5}$$

Очевидно, что $L(f, R) \leqslant U(f, R)$. Если R' продолжение разбиения R (т.е. каждый параллелепипед из разбиения R' содержится в некотором параллелепипеде из разбиения R), то

$$L(f, R) \leqslant L(f, R') \leqslant U(f, R') \leqslant U(f, R),$$

откуда следует, что верхняя грань всех нижних сумм $\sup\{L(f, R)\}$ не превосходит нижней грани всех верхних сумм $\inf\{U(f, R)\}$.

Верхняя грань всех нижних сумм $\sup\{L(f, R)\}$ называется *верхним*, а нижняя грань всех верхних сумм $\inf\{U(f, R)\}$ — *нижним интегралом* f по P. Эти интегралы всегда существуют и достигаются при измельчении разбиения до бесконечно малых параллелепипедов.

Функция f называется *интегрируемой* на параллелепипеде P, если она ограничена и $\sup\{L(f, R)\} = \inf\{U(f, R)\}$. При таком бесконечно мелком разбиении мы заменяем знак суммы $\sum$ на знак интеграла $\int$, а объём параллелепипеда $V(S)$ на элемент объёма dV. Эта величина обозначается $\int_P f$ или $\int_P f dV$ и называется *интегралом* f по P

$$\sup\left\{\sum_S m_S(f) V(S)\right\} = \inf\left\{\sum_S M_S(f) V(S)\right\} = \int\limits_P f dV = \int\limits_P f. \tag{4.6}$$

Пример. Если $f\colon [a, b] \to \mathbb{R}$ или $y = f(x)$, где $a \leqslant b$, то выражение

$$\int\limits_{[a,b]} f = \int\limits_a^b f = \int\limits_a^b f(x)\, dx = F(x)\big|_a^b = F(b) - F(a) \tag{4.7}$$

называют *определённым интегралом* непрерывной функции $f(x)$ с нижним пределом a и верхним пределом b.

Геометрически определенный интеграл вычисляется как площадь, заключённая между кривой $f(x)$ и осью абсцисс x, ограниченная слева и справа вертикальными прямыми $x = a$ и $x = b$.

Часто используют обозначение

$$\int\limits_P f dV = \int\limits_P f\, dx^1 \wedge \ldots \wedge dx^n = \underbrace{\int \ldots \int}_{n} f\, dx^1 \wedge \ldots \wedge dx^n, \tag{4.8}$$

где $dV = dx^1 \wedge \ldots \wedge dx^n$ — элемент объёма,

$\underbrace{\int \ldots \int}_{n}$ — *кратный интеграл.*

Пример. Интеграл $V(A) = \int_A dV$ называется n-мерным объёмом множества A. Разумеется, одномерный объём часто называют *длиной*, а двумерный — *площадью*. Если A есть n-мерное многообразие, а $dV = dx^1 \wedge \ldots \wedge dx^n$ есть элемент объёма на многообразии A, то под *объёмом многообразия* A понимают $V(A) = \int_A dV$ в предположении, что этот интеграл существует (объём многообразия ограничен).

Теорема о среднем. Пусть функция f непрерывна на P, тогда найдётся такая точка $\tau \in P$, что

$$\int\limits_P f\, dV = f(\tau) V(P). \tag{4.9}$$

Тогда $f(\tau)$ называют *среднее значение* на P и обозначают $\overline{f}$.

Доказательство

$$m_S(f) \leqslant f \leqslant M_S(f).$$

Теорема. Пусть P — замкнутый параллелепипед, $f\colon P \to \mathbb{R}$ ограниченная функция и B — множество её точек разрыва. Функция f интегрируема на P тогда и только тогда, когда B — множество меры 0.

Доказательство этой теоремы мы не приводим, но оно сводится к тому, что раз множество B имеет объём 0, а колебание функции (1.46) имеет предел, то всегда можно выбрать такое разбиение параллелепипеда, что разница между верхней и нижней суммами f стремится к нулю.

Интегралы по другим множествам легко сводятся к интегралам от параллелепипеда. Если $A \subset P$ для некоторого замкнутого параллелепипеда P и $f\colon P \to \mathbb{R}$ — интегрируемая функция, то под $\int_A f$ понимается $\int_P f\chi_A$, где χ_A — *характеристическая функция*:

$$\chi_A(\nu) = \begin{cases} 0 & \text{при } \nu \notin A, \\ 1 & \text{при } \nu \in A. \end{cases} \tag{4.10}$$

Функция χ_A также должна быть интегрируема. Функция $\chi_A\colon P \to \mathbb{R}$ интегрируема тогда и только тогда, когда граница множества A имеет меру 0 (и, следовательно, объем 0). Доказательство сводится к тому, что функция

χ_A разрывна только на границе множества A. Таким образом, множество точек разрыва функции χ_A совпадает с границей A и требуемый результат следует из предыдущей теоремы.

Ограниченное множество A, граница которого имеет меру 0, называется *измеримым по Жордану*.

Определение. Функция f называется *интегрируемой по Риману* на множестве A: $\int_A f$, если множество A — измеримо по Жордану.

Свойства интеграла. Пусть функции $f, g\colon A \to \mathbb{R}$ интегрируемы, тогда справедливы выражения:

1. функция $f+g$ интегрируема и $\int_A f+g = \int_A f + \int_A g$,
2. $\int_A cf = c\int_A f$ для всякой постоянной c,
3. если $f \leqslant g$, то $\int_A f \leqslant \int_A g$,
4. $\left|\int_A f\right| \leqslant \int_A |f|$.

Теорема Фубини. Пусть $A \subset \mathbb{R}^n$ и $B \subset \mathbb{R}^m$ — замкнутые параллелепипеды и $f\colon A \times B \to \mathbb{R}$ — интегрируемая функция, тогда:

$$\int\limits_{A\times B} f = \int\limits_A \left(\int\limits_B f\right). \tag{4.11}$$

Интегралы в правой части (предполагается, что они существуют) называются *повторными интегралами* для f.

Если $A = [a_1, b_1] \times \ldots \times [a_n, b_n]$ и $f\colon A \to \mathbb{R}$ — достаточно хорошая функция, то повторное применение теоремы Фубини даёт

$$\int\limits_A f = \int\limits_{a_n}^{b_n} dx^n \ldots \int\limits_{a_1}^{b_1} f(x^1, \ldots, x^n)\, dx^1. \tag{4.12}$$

Таким образом, *повторные интегралы* вычисляются при помощи последовательного вычисления обычных интегралов.

Пример 1. Вычислим площадь S поверхности сферы радиуса r.

С учётом $r = const$, $dr = 0$ получаем

$$d\sigma \overset{(3.52)}{=} r^2 \sin\theta\, d\varphi \wedge d\theta \quad (0 \leqslant \varphi < 2\pi,\ 0 \leqslant \theta \leqslant \pi).$$

Тогда

$$\begin{aligned} S = \int\limits_S d\sigma &\overset{(4.12)}{=} r^2 \int\limits_0^{2\pi} d\varphi \int\limits_0^{\pi} \sin\theta\, d\theta = \\ &\overset{\text{(с. 163)}}{\overset{(4.7)}{=}} -r^2\varphi\big|_0^{2\pi} \cos\theta\big|_0^{\pi} = 4\pi r^2. \end{aligned} \tag{4.13}$$

Пример 2. Вычислим объём V шара радиуса r:

$$dV \overset{(3.53)}{=} \rho^2 \sin\theta \, d\rho \wedge d\varphi \wedge d\theta \quad (0 \leqslant \rho \leqslant r,\ 0 \leqslant \varphi < 2\pi,\ 0 \leqslant \theta \leqslant \pi).$$

Тогда

$$\begin{gathered} V = \int\limits_V dV \overset{(4.12)}{=} \int\limits_0^r \rho^2 \, d\rho \int\limits_0^{2\pi} d\varphi \int\limits_0^{\pi} \sin\theta \, d\theta = \\ \overset{\substack{(4.7) \\ (\text{с. } 163)}}{=} -\frac{\rho^3}{3}\Big|_0^r \varphi|_0^{2\pi} \cos\theta|_0^{\pi} = \frac{4}{3}\pi r^3. \end{gathered} \tag{4.14}$$

Пример 3. Если переменные подынтегральной функции разделяются

$$f(x, y) = u(x)\, v(y)$$

то мы можем записать:

$$\iint\limits_S f \, d\sigma = \int\limits_{y_1}^{y_2} dy \int\limits_{x_1}^{x_2} f(x, y) \, dx = \int\limits_{y_1}^{y_2} v(y) \, dy \cdot \int\limits_{x_1}^{x_2} u(x) \, dx. \tag{4.15}$$

Пример 4. Если область интегрирования $A \times B$ ограничена прямыми $y = y_1$, $y = y_2$ и кривыми $x = \varphi_1(y)$, $x = \varphi_2(y)$, при этом $A = [y_1, y_2]$ и $B = [\varphi_1(y), \varphi_2(y)]$ — одномерные отрезки, а f непрерывная функция, то геометрически вычисление повторного интеграла

$$\begin{gathered} \int\limits_{A\times B} f = \int\limits_{\varphi_1(y)}^{\varphi_2(y)} dx \int\limits_{y_1}^{y_2} f(x, y) \, dy = \int\limits_{y_1}^{y_2} dy \int\limits_{\varphi_1(y)}^{\varphi_2(y)} f(x, y) \, dx = \\ \overset{(4.7)}{=} \int\limits_{y_1}^{y_2} (F(\varphi_2) - F(\varphi_1)) \, dy = \int\limits_{y_1}^{y_2} F(y) \, dy \end{gathered} \tag{4.16}$$

означает возможность вычисления объёма между поверхностью функции $z = f(x, y)$ и плоскостью (x, y) как путём его разбиения на элементарные столбики, так и путём его разбиения на элементарные слои.

Причём для слоёв мы можем записать:

$$\frac{\partial F(y)}{\partial y} = \int\limits_{\varphi_1(y)}^{\varphi_2(y)} \frac{\partial f(x, y)}{\partial y} \, dx. \tag{4.17}$$

Теорема о замене переменных. Пусть $A \subset \mathbb{R}^n$ — открытое множество и $\varphi\colon A \to \mathbb{R}^n$ — такая взаимно однозначная непрерывно дифференцируемая функция, что $\det \varphi'_\nu(\tau) \neq 0$, для всех $\tau \in A$. Тогда

$$\int\limits_{\varphi(A)} f = \int\limits_A (f \circ \varphi) \det \varphi' \tag{4.18}$$

для любой интегрируемой функции $f\colon \varphi(A) \to \mathbb{R}$ или

$$\int\limits_{\varphi(A)} f \cdot dy^1 \wedge \ldots \wedge dy^n = \int\limits_A f(\varphi(\tau)) \frac{\partial(y^1, \ldots, y^n)}{\partial(x^1, \ldots, x^n)} \cdot dx^1 \wedge \ldots \wedge dx^n. \qquad (4.19)$$

Запишем некоторые *правила интегрирования*:

1. Формулы *интегрирования по частям* следуют из правила Лейбница (3.151):

$$\int f\, dg = f \cdot g - \int g\, df, \qquad (4.20)$$

$$\int f \frac{dg}{dt}\, dV = \int \frac{d(fg)}{dt}\, dV - \int g \frac{df}{dt}\, dV. \qquad (4.21)$$

2. *Правило замены переменной* следует из теоремы о замене переменных (4.19):

$$\int f(y)\, dy = \int f[\varphi(x)] \varphi'_x(x)\, dx, \qquad (4.22)$$

где $y = \varphi(x)$, $dy = \varphi'_x(x)\, dx$.

Рассмотрим *интеграл от функции комплексного переменного*:

$$\int\limits_{z_1}^{z_2} f(z)\, dz. \qquad (4.23)$$

Ясно, что в отличие от действительных чисел, которые лежат на одной числовой оси, точки z_1 и z_2 на плоскости можно соединить разными кривыми. Оказывается, что если $f(z)$ аналитическая функция и если в области, ограниченной различными путями $f(z)$ однозначна (либо область лежит на одной ветви) и нигде не обращается в бесконечность, то интеграл не зависит от выбора пути. Этот факт представляется вполне естественным, если мы вспомним определение полного дифференциала. Два действительных числа полностью определяют комплексное число, а две функции двух действительных переменных полностью определяют одну функцию одной комплексного переменного. Следовательно, дифференциал в подынтегральном выражении является полным и полностью определен, тогда интеграл зависит только от начальных и конечных значений первообразной. Указанный вывод равносилен следующему утверждению, известному под именем *теоремы Коши*: интеграл, взятый по замкнутому контуру от функции, аналитической всюду внутри этого контура и на нем, равен нулю. Направление обхода контура против часовой стрелки по общему правилу считается положительным.

Рассмотрим *изолированную особую точку* голоморфной функции. При обходе особой точки интеграл по замкнутому контуру или циклу (такой интеграл имеет специальное обозначение $\oint$) может быть равен нулю, но может и отличаться от нуля. Интеграл будет зависеть от того, сколько раз

и в каком направлении мы сделали обход особой точки, но не зависит от того, по какому пути производился этот обход:

$$\oint \frac{dz}{z} = \oint \frac{ire^{i\varphi}d\varphi}{re^{i\varphi}} = i\oint d\varphi = 2\pi i, \tag{4.24}$$

$$\oint \frac{dz}{z^n} = 0, \quad \text{где} \quad n = 2, 3, 4, \ldots. \tag{4.25}$$

Интеграл (4.24) дает при интегрировании функцию $\ln z$ с бесконечным числом ветвей, приращение которой по замкнутому контуру равно $2\pi i$. Второй интеграл дает при интегрировании однозначную функцию $\frac{z^{-n+1}}{1-n}$, приращение которой по замкнутому контуру равно нулю.

Для того чтобы вычислить интеграл по замкнутому контуру произвольной аналитической функции мы должны разложить нашу функцию в ряд и почленно проинтегрировать.

Интеграл по замкнутому контуру от положительных членов ряда Лорана всегда равен нулю. Также равен нулю интеграл от конечного числа членов ряда с отрицательными степенями, кроме первого члена ряда a_{-1}. Для a_{-1} интеграл, исходя из формулы (4.24), равен:

$$\oint \frac{a_{-1}}{z - z_0}\, dz = 2\pi i a_{-1}. \tag{4.26}$$

Вычетом функции $f(z)$ в изолированной особой точке z_0 называется коэффициент a_{-1} ряда Лорана (1.78), с центром в точке z_0:

$$a_{-1} = \mathrm{Res}_{z_0} f(z). \tag{4.27}$$

Например:

$$\mathrm{Res}_{\,0} \frac{1}{z} = 1. \tag{4.28}$$

Функции, представимые в виде отношения двух голоморфных функций, называются *мероморфными функциями*. Таковыми являются рациональные функции $\mathrm{tg}\, z$ и $\mathrm{ctg}\, z$, эллиптические функции и др. Мероморфная в области D функция голоморфна в этой области за исключением, быть может, конечного или счетного множества полюсов; в полюсах значения мероморфной функции считаются равными бесконечности. Вычет мероморфной функции $f(z) = \frac{g(z)}{h(z)}$ в точке $z = z_0$, где числитель отличен от нуля, а знаменатель имеет нуль первого порядка, равен:

$$\mathrm{Res}_{z_0} \frac{g(z)}{h(z)} = \frac{g(z)}{h'(z)}. \tag{4.29}$$

Интеграл от произвольной аналитической функции по замкнутому контуру, содержащему особую точку $z = z_0$, равен:

$$\oint f(z)\, dz \overset{(4.26)}{\underset{(4.27)}{=}} 2\pi i \mathrm{Res}_{z_0} f(z) \quad \text{и} \quad \oint \frac{g(z)}{h(z)}\, dz \overset{(4.29)}{=} 2\pi i \frac{g(z)}{h'(z)}. \tag{4.30}$$

Интеграл по замкнутому контуру, содержащему несколько особых точек, равен произведению $2\pi i$ на сумму вычетов подынтегральной функции во всех особых точках, расположенных внутри контура.

Понятие определенного интеграла дал О. Коши (1823). О. Коши применял своё определение только к случаям «общего положения», т.е. к непрерывным функциям и конечным пределам интегрирования a и b. Определения интеграла для разрывов функции или бесконечных пределов интегрирования в различных вариациях последовательно давали Б. Риман (1853), Т. Стилтьес (1894), А. Лебег (1902), А. Данжуа (1912) и А. Я. Хинчин (1915).

4.2 Теорема Стокса

Если в качестве подынтегральной функции взять дифференциальную форму ω, заданную на k-мерном бесконечно малом параллелепипеде $\xi_1 \times \ldots \times \xi_k$, то $\int_{[0,1]^k} \omega$ есть интеграл на стандартном k-мерном кубе в $\mathbb{R}^n$.

Если теперь $C\colon [0,1]^k \to A$ сингулярный k-мерный куб в A, то

$$\int\limits_C \omega = \int\limits_{[0,1]^k} C^*(\omega) \tag{4.31}$$

где C^* — функция, *сопряжённая* к C, которая действует по формуле (2.162)

$C^*(\omega(\xi_1, \ldots, \xi_k)) = \omega(C(\xi_1), \ldots, C(\xi_k))$.

Звёздочка сверху указывает, что C^* действует в сторону, противоположную C. Значение формы $C^*(\omega)$ на k-векторах равно значению формы ω на образах этих векторов. При этом форма $C^*(\omega)$ задана в каждой точке $\tau \in [0,1]^k$, а форма ω задана в точке $C(\tau)$. Особое определение требуется для случая $k = 0$. Форма нулевой степени есть функция $\omega = f$, а нульмерный куб есть точка $C(0) = (x^1, \ldots, x^n)$ в A, тогда

$$\int\limits_{C(0)} f = f(C(0)) = f(x^1, \ldots, x^n). \tag{4.32}$$

Пусть $\omega = f\, dx^1 \wedge \ldots \wedge dx^k$ — форма k-й степени на $[0,1]^k$. Тогда

$$\int\limits_{I^k} f\, dx^1 \wedge \ldots \wedge dx^k = \int\limits_{[0,1]^k} (I^k)^*(f\, dx^1 \wedge \ldots \wedge dx^k). \tag{4.33}$$

Наконец интеграл от дифференциальной k-формы по сингулярной k-мерной цепи $C = \sum a_i C_i$ определим формулой

$$\int\limits_C \omega = \sum a_i \int\limits_{C_i} \omega. \tag{4.34}$$

Тот факт, что $d^2 = 0$ для дифференциальных форм (см. (3.62)) и $\partial^{\,2} = 0$ для цепей (см. стр. 35) наводит на мысль, что между формами и цепями d

и ∂ существует какая-то связь. Связь между формами и цепями наиболее четко выражается теоремой Стокса.

Теорема Стокса. Пусть ω — форма $(k-1)$-й степени на открытом множестве A и C — сингулярная k-мерная цепь в A. Тогда

$$\int\limits_C d\omega = \int\limits_{\partial C} \omega. \tag{4.35}$$

Доказательство проведём сначала для $C = I^k$. Тогда ω есть сумма форм $(k-1)$-й степени вида $f\,dx^1 \wedge \ldots \wedge \widehat{dx^i} \wedge \ldots \wedge dx^k$, где $\widehat{dx^i}$ отсутствующий член, и достаточно доказать теорему для каждой из них. А это достигается непосредственным вычислением. Заметим, что

$$\int\limits_{[0,1]^{k-1}} {I^k_{(j,\alpha)}}^* (f\,dx^1 \wedge \ldots \wedge \widehat{dx^i} \wedge \ldots \wedge dx^k) =$$

$$= \begin{cases} 0 & \text{при } j \neq i, \\ \displaystyle\int\limits_{I^k} f(x^1, \ldots, \alpha, \ldots, x^n)\, dx^1 \wedge \ldots \wedge dx^k & \text{при } j = i. \end{cases}$$

Тогда с учётом определения ∂I^k (1.48)

$$\int\limits_{\partial I^k} f\,dx^1 \wedge \ldots \wedge \widehat{dx^i} \wedge \ldots \wedge dx^k =$$

$$\overset{(4.33)}{=} \sum_{j=1}^{k} \sum_{\alpha=0,1} (-1)^{j+\alpha} \int\limits_{[0,1]^{k-1}} {I^k_{(j,\alpha)}}^* (f\,dx^1 \wedge \ldots \wedge \widehat{dx^i} \wedge \ldots \wedge dx^k) =$$

$$= (-1)^{i+1} \int\limits_{I^k} f(x^1, \ldots, 1, \ldots, x^n)\, dx \wedge \ldots \wedge dx^k +$$

$$+ (-1)^i \int\limits_{I^k} f(x^1, \ldots, 0, \ldots, x^n)\, dx \wedge \ldots \wedge dx^k.$$

С учётом свойств внешнего дифференцирования (3.59) получаем

$$\int\limits_{I^k} d(f\,dx^1 \wedge \ldots \wedge \widehat{dx^i} \wedge \ldots \wedge dx^k) =$$

$$= \int\limits_{[0,1]^k} \frac{\partial f}{\partial x^i}\, dx_i \wedge dx^1 \wedge \ldots \wedge \widehat{dx^i} \wedge \ldots \wedge dx^k =$$

$$= (-1)^{i-1} \int\limits_{[0,1]^k} \frac{\partial f}{\partial x^i}\, dx^1 \wedge \ldots \wedge dx^k.$$

С учётом выражения для определённого интеграла (4.7) и теоремы Фубини (4.12) получаем

$$\int\limits_{[0,1]^k} \frac{\partial f}{\partial x^i}\, dx^1 \wedge \ldots \wedge dx^k = \int\limits_{[0,1]^{k-1}} [f(x^1, \ldots, 1, \ldots, x^n) -$$

$$- f(x^1, \ldots, 0, \ldots, x^n)]\, dx^1 \wedge \ldots \wedge \widehat{dx^i} \wedge \ldots \wedge dx^k.$$

Тогда

$$\int\limits_{I^k} d(f\,dx^1 \wedge \ldots \wedge \widehat{dx^i} \wedge \ldots \wedge dx^k) =$$
$$= (-1)^{i+1} \int\limits_{I^k} f(x^1, \ldots, 1, \ldots, x^n)\, dx \wedge \ldots \wedge dx^k +$$
$$+ (-1)^i \int\limits_{I^k} f(x^1, \ldots, 0, \ldots, x^n)\, dx \wedge \ldots \wedge dx^k.$$

Таким образом,

$$\int\limits_{I^k} d\omega = \int\limits_{\partial I^k} \omega. \tag{4.36}$$

Если теперь C — произвольный сингулярный k-мерный куб, то

$$\int\limits_{\partial C} \omega \overset{(4.31)}{=} \int\limits_{\partial I^k} C^*(\omega). \tag{4.37}$$

Поэтому

$$\int\limits_C d\omega \overset{(4.31)}{=} \int\limits_{I^k} C^*(d\omega) = \int\limits_{I^k} d(C^*\omega) \overset{(4.36)}{=} \int\limits_{\partial I^k} C^*\omega \overset{(4.37)}{=} \int\limits_{\partial C} \omega. \tag{4.38}$$

Наконец, для произвольной сингулярной k-мерной цепи $C = \sum a_i C_i$

$$\int\limits_C d\omega \overset{(4.34)}{=} \sum a_i \int\limits_{C_i} d\omega \overset{(4.38)}{=} \sum a_i \int\limits_{\partial C_i} \omega \overset{(4.34)}{=} \int\limits_{\partial C} \omega,$$

что и требовалось доказать.

Впервые формулировка теоремы Стокса появилась в виде приписки к письму сэра Уильяма Томсона (лорда Кельвина) к Стоксу, датированному 2 июля 1850 г. Опубликована она была в качестве восьмого вопроса к экзаменам на смитовскую премию 1854 г. Этот конкурсный экзамен, которому подвергались лучшие студенты-математики Кембриджского университета, с 1849 по 1882 г. проводился проф. Стоксом.

Современниками Стокса были даны, по крайней мере, три доказательства: одно опубликовал Томсон, другое было изложено в «Трактате о натуральной философии» Томсона и Тейта и третье предложил Максвелл в «электричестве и магнетизме».

Теорема Стокса обладает тремя важными характерными признаками многих больших теорем:

1. она тривиальна;

2. тривиальна она потому, что все входящие в неё выражения определены надлежащим образом;

3. она имеет важные следствия (изложены ниже).

Пусть на m-мерном многообразии с краем M заданы форма k-й степени ω и сингулярный k-мерный куб C. Интеграл от ω по C определяется точно так же, как и раньше в (4.31).

Интегралы по сингулярным цепям определяются аналогично (4.34). В случае $m = k$ может оказаться, что существуют такие открытое множество $M \supset [0,1]^k$ и система координат $\psi\colon W \to \mathbb{R}^n$, что $C(\tau) = \psi(\tau)$ для всех $\tau \in [0,1]^k$. Сингулярные k-мерные кубы в M всегда будут считаться принадлежащими к этому типу. Если M ориентировано, то будем говорить, что C в M сориентирован, если ψ сохраняет ориентацию.

Теорема. Пусть M — ориентированное k-мерное многообразие, $C_1, C_2\colon [0,1]^k \to M$ — два сориентированных сингулярных k-мерных куба в M и ω — форма k-й степени на M. Тогда

$$\int\limits_{C_1} \omega = \int\limits_{C_2} \omega, \quad \text{если} \quad \omega = 0 \quad \text{вне} \quad C_1([0,1]^k) \cap C_2([0,1]^k). \tag{4.39}$$

Доказательство проведём при помощи функции $\varphi\colon C_1 \to C_2$, которая есть преобразование координат.

Пусть $\omega = f\,dx^1 \wedge \ldots \wedge dx^k$, тогда

$$\begin{aligned}
\int\limits_{C_2} \omega &= \int\limits_{[0,1]^k} C_2^*(\omega) = \int\limits_{[0,1]^k} \omega(C_2) = \\
&= \int\limits_{C_2([0,1]^k)} f\,dy^1 \wedge \ldots \wedge dy^k = \\
&= \int\limits_{C_1([0,1]^k)} f(\varphi(\tau)) \frac{\partial(y^1, \ldots, y^k)}{\partial(x^1, \ldots, x^k)}\,dx^1 \wedge \ldots \wedge dx^k = \\
&= \int\limits_{[0,1]^k} \omega(C_1) = \int\limits_{[0,1]^k} C_1^*(\omega) = \int\limits_{C_1} \omega.
\end{aligned}$$

Если в M найдётся такой сориентированный сингулярный k-мерный куб C, что $\omega = 0$ вне $C([0,1]^k)$, то по определению положим

$$\int\limits_{C} \omega = \int\limits_{M} \omega \tag{4.40}$$

Предыдущая теорема показывает, что $\int_M \omega$ не будет зависеть от выбора C. Если M, ориентированное k-мерное многообразие с краем, то ∂M имеет индуцированную ориентацию $\partial\mu$. Тогда грань $C_{k,0}$ сориентирована, если k чётно, и несориентирована, если k нечётно, т.е. $\int_{C_{k,0}} = (-1)^k \int_{\partial M} \omega$. С другой стороны, $C_{k,0}$ входит с коэффициентом $(-1)^k$ в ∂C. Поэтому

$$\int\limits_{\partial C} = \int\limits_{(-1)^k C_{(k,0)}} \omega = (-1)^k \int\limits_{C_{k,0}} \omega = \int\limits_{\partial M} \omega. \tag{4.41}$$

Теорема Стокса для многообразий. Пусть M — компактное ориентированное m-мерное многообразие с краем, где ∂M наделено индуцированной ориентацией, и ω — форма $(m-1)$-й степени на M. Тогда

$$\int\limits_M d\omega = \int\limits_{\partial M} \omega. \tag{4.42}$$

Цепь C, которая не имеет границы $\partial C = 0$, называется *цикл*. Интеграл по циклу принято обозначать значком $\oint$. Край многообразия всегда есть цикл, так как $\partial^2 = 0$. Тогда теорема Стокса будет иметь вид:

$$\int\limits_C d\omega = \oint\limits_{\partial C} \omega, \quad \int\limits_M d\omega = \oint\limits_{\partial M} \omega. \tag{4.43}$$

Теорема. Пусть M_1, M_2 — компактные m-мерные многообразия с краем, где $\partial M_1, \partial M_2$ наделены ориентациями, индуцированными стандартными ориентациями многообразий M_1, M_2, причём $M_2 \subset M_1 \setminus \partial M_1$ и ω — форма $(m-1)$-степени на M_1. Тогда

$$\oint\limits_{\partial M_1} \omega = \oint\limits_{\partial M_2} \omega. \tag{4.44}$$

Интегральным инвариантом преобразования $\psi\colon M \to M$ называется дифференциальная k-форма ω, если интегралы ω по любой k-мерной цепи C и по ее образу ψC при отображении одинаковы:

$$\int\limits_C \omega = \int\limits_{\psi C} \omega. \tag{4.45}$$

Относительным интегральным инвариантом дифференцируемого преобразования $\psi\colon M \to M$ называется дифференциальная k-форма ω, если для всякого цикла C

$$\oint\limits_C \omega = \oint\limits_{\psi C} \omega. \tag{4.46}$$

Теорема. Пусть ω — относительный интегральный инвариант преобразования ψ, тогда $d\omega$ — абсолютный интегральный инвариант ψ

$$\int\limits_C d\omega = \int\limits_{\psi C} d\omega. \tag{4.47}$$

Доказательство:
$\int_C d\omega \overset{(4.43)}{=} \oint_{\partial C} \omega \overset{(4.46)}{=} \oint_{\psi(\partial C)} \overset{(1.56)}{=} \oint_{\partial(\psi C)} \overset{(4.43)}{=} \int_{\psi C} d\omega.$

Обратное не верно, если $d\omega^k$ абсолютный интегральный инвариант, то ω^k может не быть относительным интегральным инвариантом, если в M есть k-мерная замкнутая цепь, не являющаяся границей.

Теорема. Интеграл дифференциала по циклу равен нулю:

$$\oint\limits_C d\omega = 0, \quad \text{если} \quad \partial C = 0. \tag{4.48}$$

Доказательство $\oint_C d\omega \overset{(4.43)}{=} \int_{\partial C} \omega = 0$, так как $\partial C = 0$.

Теорема. Интеграл замкнутой формы по границе цепи равен нулю:

$$\oint\limits_{\partial C} \omega = 0, \quad \text{если} \quad d\omega = 0. \tag{4.49}$$

Доказательство $\oint_{\partial C} \omega \overset{(4.43)}{=} \int_C d\omega = 0$, так как $d\omega = 0$.

Теорема. Интеграл от точной формы по циклу равен нулю:

$$\oint\limits_{C} \omega = 0, \quad \text{если} \quad \omega = d\eta, \partial C = 0. \tag{4.50}$$

Доказательство $\oint_C \omega = \oint_C d\eta \overset{(4.43)}{=} \int_{\partial C} \eta = 0$, так как $\partial C = 0$.

Полным дифференциалом называется точная форма от заданной на многообразии функции f, для которого справедливо

$$\oint\limits_{C} df = 0. \tag{4.51}$$

Если интеграл полного дифференциала функции по циклу равен нулю, то изменение функции f при перемещении из точки 1 в точку 2 не зависит от пути интегрирования (от цепи) и равно разности значений функции в этих точках:

$$\int\limits_{1}^{2} df = f_2 - f_1. \tag{4.52}$$

В этом случае мы всегда можем восстановить функцию f по её полному дифференциалу df, а функция f называется *функция состояния*. Если 1-форма (пфаффова форма df) не является полным дифференциалом f, тогда функцию f называют *функцией процесса*. Изменение функции процесса f при перемещении из точки 1 в точку 2 также вычисляется как интеграл, но значение интеграла будет зависеть от пути интегрирования (цепи):

$$\Delta f_{1-2} = \int\limits_{(1-2)} df. \tag{4.53}$$

Как уже упоминалось на стр. 139, если дифференциал функции не является полным, то обычно это означает, что мы не смогли определить все координаты, на которых задана рассматриваемая функция.

Рассмотрим особенности применения теоремы Стокса на нечётномерных многообразиях.

Пусть C — замкнутая кривая (цикл) на M^{2n+1}. Совокупность всех линий ротора (с. 125), проходящих через C, образуют «*трубку ротора*». Назовём интеграл 1-формы $\omega^1_\xi(\eta) = (\xi, \eta)$ по C — *циркуляцией ротора*, а интеграл 2-формы $\omega^2_{\operatorname{rot}\xi}(\eta, \mu) = \omega^1_\xi(\eta)$ по поверхности S, ограниченной замкнутой кривой C, — *потоком ротора*.

Многомерная лемма Стокса (частный случай (4.44)). Интеграл 1-формы ω по любому из двух одномерных циклов C_1, C_2, охватывающих одну и ту же трубку ротора, одинаков, если $C_1 - C_2 = \partial S$, где S — кусок трубки ротора:

$$\oint\limits_{C_1} \omega = \oint\limits_{C_2} \omega. \tag{4.54}$$

Доказательство: Значение $d\omega$ на любой линии ротора равно нулю, тогда $\oint_{C_1} \omega - \oint_{C_2} \omega = \oint_{\partial S} \omega \overset{(4.43)}{=} \int_S d\omega = 0$.

Теорема. При интегрировании замкнутой формы ω^k по k-мерному циклу можно заменять цикл на другой при условии, что их разность $C_1^k - C_2^k = \partial C^{k+1}$ есть граница $(k+1)$-мерной цепи («плёнки»):

$$\oint\limits_{C_1^k} \omega^k = \oint\limits_{C_2^k} \omega^k, \quad \text{если} \quad d\omega^k = 0, \quad C_1^k - C_2^k = \partial C^{k+1}. \tag{4.55}$$

Такие два цикла Пуанкаре назвал *гомологичными* (греч. *ομολογία* — согласие, соответствие; от *ομος* — равный, одинаковый и *λογος* — слово, учение).

4.3 Интегрирование в E^3

Так как предмет *векторного анализа* состоит в изучении дифференцирования и интегрирования скалярных и векторных полей, то в качестве противоядия после большого количества абстрактных математических определений рассмотрим интегрирование векторных и скалярных полей в трёхмерном евклидовом пространстве E^3.

Криволинейным интегралом в классическом векторном анализе называют интеграл от формы первой степени по сингулярной одномерной цепи (по какой-либо кривой).

Поверхностным интегралом в классическом векторном анализе называют интеграл от формы второй степени по сингулярной двумерной цепи (по какой-либо поверхности).

Более сложные интегралы от дифференциальной k-формы по сингулярной k-мерной цепи в классическом векторном анализе не определены. Обобщение определения криволинейных и поверхностных интегралов на их многомерный аналог (многообразие) в принятых классическим векторным анализом обозначениях невозможно.

В отличие от пределов интегрирования функции одного переменного, которые лежат на одной оси, две точки в пространстве можно соединить разными кривыми. Поэтому в обозначениях классического анализа определение криволинейного интеграла не является однозначным, а определение поверхностного интеграла ещё более сложно.

При этих определениях используют понятия:

Интегралов 1-го рода — при интегрировании скалярного поля f:

1. $\int_C f(\nu)\,ds$, где C — заданная кривая, ds — дифференциал её дуги, ν — точка на кривой;

2. $\int_S f(\nu)\,d\sigma$, где S — заданная поверхность, $d\sigma$ — дифференциал её площади, ν — точка на поверхности.

Интегралов 2-го рода — при интегрировании векторного поля $\mathbf{a}$:

1. $\int_C \mathbf{a}\,d\mathbf{r} = \int_C P dx + Q dy + R dz$, где C — заданная кривая, $d\mathbf{r}$ — элемент длины;

2. $\iint_S \mathbf{a}\,d\sigma = \iint_S P dy dz + Q dz dx + R dx dy$, где S — заданная поверхность, $d\sigma$ — элемент площади.

Вычисление интеграла 1-го рода на плоской кривой сводится к определённому, а на поверхности — к повторному интегралу:

$$\int\limits_C f(\nu)\,ds = \int\limits_a^b f[x, y(x)]\sqrt{1 + y'^2}\,dx,$$

$$\int\limits_S f(\nu)\,d\sigma = \int\limits_a^b dx \int\limits_{\varphi_1(x)}^{\varphi_2(x)} f(x, y)\,dy.$$

При вычислении интегралов 2-го рода кривые и поверхности предполагаются ориентированными, а площади проекций не равны величинам $dydz, dzdx, dxdy$, что вызывает сложности в определении интегралов. Кроме того с подобными выражениями почти невозможно работать, и их быстро преобразуют в интегралы 1-го рода.

Когда в векторном анализе рассматриваются интегралы на циклах $\oint \mathbf{a}\,d\mathbf{r}$ и $\oiint_S \mathbf{a}\,d\sigma$, или на замкнутых кривых и поверхностях, их доказательства и сами формулировки существенно основываются на интуитивных представлениях о «поверхности, ограничивающей тело» или «линии, ограничивающей поверхность».

Поэтому мы будем дублировать обозначения классического векторного анализа записями в терминах дифференциальных форм.

Впервые криволинейный интеграл встречается у А. Клеро (1743), а в общем виде их ввёл О. Коши (1825).

В ориентированном трехмерном пространстве E^3 интеграл

$$\int\limits_C \mathbf{a}\,d\mathbf{r} = \int\limits_C \omega^1_{\mathbf{a}}(\xi) = \int\limits_C (\mathbf{a}, \xi) \tag{4.56}$$

называют *циркуляцией вектора* $\mathbf{a}$ по контуру C, а интеграл

$$\iint\limits_S \mathbf{a}\,d\sigma = \int\limits_S \omega^2_{\mathbf{a}}(\xi, \eta) = \int\limits_S (\mathbf{a}, \xi, \eta) \tag{4.57}$$

называют *потоком вектора* $\mathbf{a}$ через поверхность S.

Следующие три классические теоремы векторного анализа — Стокса, Гаусса-Остроградского и Грина — являются непосредственными следствиями теоремы Стокса (4.43).

Теорема Стокса. Пусть $S \subset \mathbb{R}^3$ — компактное двумерное ориентированное многообразие (поверхность) с краем $\partial S = C$ (контуром, ограничивающим поверхность, и наделённым индуцированной ориентацией). Если функции $P, Q, R\colon V \to \mathbb{R}$ — дифференцируемы, то

$$\oint\limits_C P dx + Q dy + R dz = \iint\limits_S \left[\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dxdy + \right. \\ \left. + \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) dydz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) dzdx \right]. \tag{4.58}$$

Направление обхода контура C должно быть согласовано с ориентацией поверхности S.

В обозначениях векторного анализа теорема имеет вид:

$$\oint\limits_C \mathbf{a}\, d\mathbf{r} \overset{\substack{(4.43)\\(3.65)}}{=} \iint\limits_S \operatorname{rot} \mathbf{a}\, d\sigma \tag{4.59}$$

где $d\sigma$ — вектор элемента площади с направлением внешней нормали к поверхности S.

Формула (4.59) предложена Дж. Стоксом (1854). Смысл формулы Стокса состоит в том, что циркуляция векторного поля по замкнутому контуру C равна потоку вихря поля через поверхность S. Рассмотрим, например, поле скоростей частиц в потоке.

Векторной трубкой поля называют совокупность всех векторных линий поля (3.116), проходящих через некоторую замкнутую кривую C. Площадь S, ограниченную кривой C, называют *сечением* трубки. Поток вектора $\iint_S \mathbf{a}\, d\sigma$ через это сечение называют *интенсивностью* векторной трубки. Рассмотрим, например, движение жидкости. Каждой частице жидкости соответствует своя векторная линия движения. Контур векторной трубки состоит из одних и тех же частиц жидкости.

С каждым векторным полем $\mathbf{a}$ связано поле его вихря $\mathbf{b} = \operatorname{rot} \mathbf{a}$ (3.77). Интегральные кривые $\mathbf{b}$ принято называть *линиями ротора* или *вихревыми линиями*. Векторная трубка поля вихря (совокупность всех вихревых линий, проходящих через некоторую замкнутую кривую C) называется *вихревой трубкой* поля или *трубкой ротора*. *Поток вихря*

$$\iint\limits_S \operatorname{rot} \mathbf{a}\, d\sigma = \int\limits_S \omega^2_{\operatorname{rot} \mathbf{a}}(\xi, \zeta) = \int\limits_S (\operatorname{rot} \mathbf{a}, \xi, \zeta) = \int\limits_S d\omega^1_{\mathbf{a}}(\xi) \tag{4.60}$$

через любые сечения вихревой трубки постоянен и называется *напряжением* вихревой трубки. В частности, если вихревая трубка сужается (воронка утекающей в раковине воды), то интенсивность вихря растет, а напряжение сохраняется. Если рассмотреть движение жидкости, то контур

вихревой трубки состоит из одних и тех же частиц жидкости. Циркуляция скорости жидкости по контуру, состоящему из одних и тех же частиц, также сохраняется. Другими словами, циркуляция скорости жидкости по контуру сохраняется, когда контур переносится жидкостью на новое место. Это утверждение является содержанием, так называемой, *гидродинамической леммы* или леммы Стокса — трёхмерного случая многомерной леммы Стокса.

Лемма Стокса (частный случай (4.54)): Циркуляция поля $\mathbf{a}$ по любым двум замкнутым кривым C_1, C_2, охватывающим одну и ту же трубку ротора, одинакова:

$$\oint\limits_{C_1} \mathbf{a}\, d\mathbf{r} = \oint\limits_{C_2} \mathbf{a}\, d\mathbf{r}. \tag{4.61}$$

Доказательство непосредственно следует из теоремы Стокса:

$$\oint\limits_{C_1} \mathbf{a}\, d\mathbf{r} - \oint\limits_{C_2} \mathbf{a}\, d\mathbf{r} = \iint\limits_{S} \operatorname{rot} \mathbf{a}\, d\sigma = 0,$$

где $C_1 - C_2 = \partial S$, а S — кусок трубки ротора. Вектор $d\sigma$ есть вектор элемента площади с направлением внешней нормали к трубке, а вектор $\operatorname{rot}\mathbf{a}$ касается трубки, значит их скалярное произведение равно нулю.

Как уже упоминалось (3.76), если векторное поле является градиентом скалярного поля: $\mathbf{a} = -\operatorname{grad} f$, то его вихрь равен нулю $\operatorname{rot}\mathbf{a} = 0$. Циркуляция такого векторного поля по замкнутому контуру также равна нулю: $\oint_C \mathbf{a}\, d\mathbf{r} = \iint_S \operatorname{rot}(\operatorname{grad} f)\, d\sigma = 0$.

Это правило справедливо для односвязных многообразий, когда мы предполагаем, что всегда сможем путем непрерывной деформации стянуть наш замкнутый контур в точку, т.е. область определения нашего вектора обладает свойством *односвязности*. Таким образом, когда мы говорим о потенциальном поле, мы имеем ввиду только однозначный потенциал. Допустим теперь, что мы задали циркуляцию вектора на двумерном не односвязном многообразии, которое представляет собой тор.

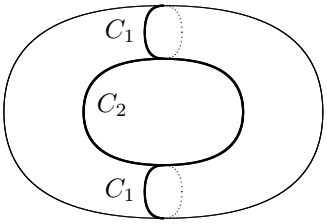


Рис. 4.1: Неприводимые контуры тора C_1 и C_2

Если мы выберем на поверхности тора так называемый неприводимый контур, который нельзя путем непрерывной деформации стянуть в точку, а два различных неприводимых контура нельзя перевести путем непрерывной деформации друг в друга, то ситуация будет иной. Потенциал, вычисленный для произвольной точки тора, не будет однозначным, а циркуляция поля по неприводимому замкнутому контуру не будет равна нулю. На двумерном торе T^2 можно выбрать два таких неприводимых контура (рис. 4.1).

Если обойти замкнутый контур, то из потенциала вычтется циркуляция поля по этому контуру.

На n-мерном инвариантном торе T^n можно выбрать ровно n неприводимых контуров $C_1, \ldots, C_n$, которые не могут быть стянуты в точку или переведены непрерывным образом друг в друга.

Неприводимый контур можно стянуть в точку только одним способом, сначала обойти контур против часовой стрелки, а затем по часовой стрелке. В этом случае получится два интеграла по контуру равных между собой, но противоположных по знаку. Такой двойной контур можно стянуть в точку и интеграл по нему будет равен нулю.

Таким образом, безвихревое поле $\operatorname{rot} \mathbf{a}$ в неодносвязной области не обязано быть потенциальным, и может быть циркуляционным, когда $\oint_C \mathbf{a}\, d\mathbf{r} \neq 0$. Такими будут, например, магнитное поле внутри торовидной катушки, по которой проходит постоянный ток; поле скоростей при безвихревом течении жидкости, обтекающей замкнутый канал, и т.д.

Кроме формулы для вихря запишем еще одну формулу, заменяющую двойной интеграл интегралом по замкнутому контуру:

$$\iint\limits_S d\sigma \times \operatorname{grad} f = \oint\limits_C f\, d\mathbf{r}. \tag{4.62}$$

Теорема Гаусса–Остроградского. Пусть $V \subset \mathbb{R}^3$ — компактное трёхмерное многообразие (область трёхмерного пространства) с краем $S = \partial V$ (поверхность, ограничивающая область), функции $P, Q, R\colon \to \mathbb{R}$ — дифференцируемы. Тогда

$$\begin{aligned} &\oiint\limits_S P\cos(nx)\, dy\, dz + Q\cos(ny)\, dz\, dx + R\cos(nz)\, dx\, dy = \\ &= \iiint\limits_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx\, dy\, dz. \end{aligned} \tag{4.63}$$

где n внешняя нормаль к поверхности S.

Доказательство непосредственно следует из теоремы Стокса, при записи теоремы в обозначениях векторного анализа:

$$\oiint\limits_S \mathbf{a}\, d\sigma \overset{\substack{(4.43)\\(\text{с. }126)}}{=} \iiint\limits_V \operatorname{div} \mathbf{a}\, dV. \tag{4.64}$$

Формула опубликована М. В. Остроградским (1838). Смысл формулы Остроградского состоит в том, что интеграл от дивергенции вектора $\mathbf{a}$, распространенный на объем V, равен потоку вектора, направленному по внешней нормали через замкнутую поверхность S. Формула Остроградского также верна только для односвязных областей.

Допустим, мы имеем положительную дивергенцию поля $\operatorname{div} \mathbf{a} > 0$. Это означает, что количество векторных линий входящих в объем V через замкнутую поверхность S меньше количества векторных линий, выходящих из этого объема. Следовательно, внутри поверхности S существуют точки поля, которые являются «истоками» векторных линий $\mathbf{a}$. Истоки поля,

где $\operatorname{div}\mathbf{a} < 0$ иногда называют стоками поля. Интегрирование происходит по объёму, следовательно, $\operatorname{div}\mathbf{a}$ является «плотностью истоков» векторных линий. Если число входящих и выходящих векторных линий одинаково, то дивергенция равна нулю $\operatorname{div}\mathbf{a} = 0$. Это означает, что $\omega = \mathbf{a}\,d\sigma$ есть замкнутая форма $d\omega = \operatorname{div}\mathbf{a} = 0$, или что поток вектора $\mathbf{a}$ через край области равен нулю $\oiint_S \mathbf{a}\,d\sigma = 0$.

Пример. *Уравнение неразрывности*:

$$\operatorname{div}\mathbf{j} + \frac{\partial\rho}{\partial t} = 0 \quad \text{или} \quad \oiint\limits_S \mathbf{j}\,d\sigma = -\frac{\partial}{\partial t}\iiint\limits_V \rho\,dV, \tag{4.65}$$

где ρ — плотность, $\dot{\mathbf{r}}$ — скорость, а $\mathbf{j} = \rho\dot{\mathbf{r}}$ — поток вещества.

Доказательство. Пусть $\mathbf{a} = \dot{\mathbf{r}}$ означает скорость течения жидкости, имеющей объёмную плотность $\rho(t, x, y, z)$. Вычислим количество жидкости, протекающее через замкнутую поверхность S за единицу времени:

$$\oiint\limits_S \rho\dot{\mathbf{r}}\,d\sigma.$$

Количество жидкости, занимающее объём V, ограниченный S, выражается интегралом $\iiint_V \rho\,dV$. Скорость изменения этой величины вычисляется как

$$\frac{d}{dt}\iiint\limits_V \rho\,dV = \iiint\limits_V \frac{\partial\rho}{\partial t}\,dV$$

и равно с обратным знаком количеству жидкости, протекающей через поверхность S:

$$\oiint\limits_S \rho\dot{\mathbf{r}}\,d\sigma \overset{(4.64)}{=} \iiint\limits_V \operatorname{div}(\rho\dot{\mathbf{r}}) = -\iiint\limits_V \frac{\partial\rho}{\partial t}\,dV,$$

что и требовалось доказать.

При этом предполагается, что в объёме нет источников жидкости.

Уравнение (4.65) легко распространяется на n-мерный случай.

Пример. Запишем полную производную функции ρ:

$$\dot{\rho} \overset{(3.104)}{=} \operatorname{grad}\rho\cdot\dot{\mathbf{r}} + \frac{\partial\rho}{\partial t} \overset{(3.178)}{=} \operatorname{div}(\rho\dot{\mathbf{r}}) - \rho\operatorname{div}\dot{\mathbf{r}} + \frac{\partial\rho}{\partial t} \overset{(4.65)}{=} -\rho\operatorname{div}\dot{\mathbf{r}},$$

откуда

$$\operatorname{div}\dot{\mathbf{r}} = -\frac{1}{\rho}\frac{d\rho}{dt}. \tag{4.66}$$

Этот пример наглядно демонстрирует смысл дивергенции. Дивергенция (расходимость) поля скоростей $\dot{\mathbf{r}}$ даёт относительное изменение плотности элемента жидкости, находящегося в данном месте.

Если жидкость несжимаема ($\rho = const$ или $\dot{\rho} = 0$), то это изменение должно равняться нулю, и мы получим *условие несжимаемости*:

$$\operatorname{div}\dot{\mathbf{r}} \overset{(4.66)}{=} 0. \tag{4.67}$$

Уравнение неразрывности широко используется, например, в гидро- и аэромеханике, где выражает закон сохранения вещества для движущейся жидкости (газа).

Кроме теоремы Остроградского или формулы для дивергенции запишем еще две формулы, еще два частных случая, заменяющие тройной интеграл двойным:

$$\iiint\limits_V \operatorname{grad} f\, dV = \oiint\limits_S f\, d\sigma, \tag{4.68}$$

$$\iiint\limits_V \operatorname{rot} \mathbf{a}\, dV = \oiint\limits_S d\sigma \times \mathbf{a}. \tag{4.69}$$

Используя символический оператор ∇, можно указать удобное для запоминания операторное соотношение, из которого формула замены тройных интегралов двойными получаются в виде частных случаев. Это соотношение имеет вид:

$$\iiint\limits_V dV \cdot \nabla = \oiint\limits_S d\sigma. \tag{4.70}$$

Теорема Грина. Пусть $S \subset \mathbb{R}^2$ — компактное двумерное ориентированное многообразие с краем $\partial S = C$, наделённым индуцированной ориентацией, функции $P, Q\colon \to \mathbb{R}$ — дифференцируемы. Тогда

$$\oint\limits_C P dx + Q dy = \iint\limits_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dxdy. \tag{4.71}$$

Эта формула является частным случаем теоремы Стокса и была известна ещё Л. Эйлеру (1771). Две другие впервые опубликованы Дж. Грином (1828) в связи с исследованием по теории потенциала.

Пусть $V \subset \mathbb{R}^3$ — компактный объём, ограниченный поверхностью $S = \partial V$; $u, v\colon V \to \mathbb{R}$ — дифференцируемы.

Тогда первая формула имеет вид:

$$\oiint\limits_S v\frac{\partial u}{\partial n}\, d\sigma = \iiint\limits_V \left(\frac{\partial v}{\partial x}\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial z}\frac{\partial u}{\partial z}\right) dV + \iiint\limits_V v\Delta u\, dV, \tag{4.72}$$

где $d\sigma$ — вектор элемента площади с направлением внешней нормали к поверхности S,

dV — элемент объёма,

$\frac{\partial u}{\partial n}$ — производная по направлению внешней нормали к S,

Δ — оператор Лапласа.

Доказательство. Подставим функции $P = v\frac{\partial u}{\partial x}, Q = v\frac{\partial u}{\partial y}, R = v\frac{\partial u}{\partial z}$ в уравнение (4.63) и с учётом (3.142) сразу получим уравнение (4.72).

Из нее сразу получаем вторую формулу Грина:

$$\oiint\limits_S \left(v\frac{\partial u}{\partial n} - u\frac{\partial v}{\partial n}\right) d\sigma = \iiint\limits_V (v\Delta u - u\Delta v)\, dV. \tag{4.73}$$

В обозначениях векторного анализа формулы имеют вид:

$$\oiint\limits_S v\mathrm{grad}\, u\, d\sigma = \iiint\limits_V (\mathrm{grad}\, v\, \mathrm{grad}\, u + v\Delta u)\, dV, \tag{4.74}$$

$$\oiint\limits_S (v\mathrm{grad}\, u - u\mathrm{grad}\, v)\, d\sigma = \iiint\limits_V (v\Delta u - u\Delta v)\, dV. \tag{4.75}$$

4.4 Несобственные интегралы

Интегралы, у которых область интегрирования конечна и подынтегральная функция в этой области не обращается в бесконечность, называются интегралами в собственном смысле слова или, коротко, *собственными*. Если хотя бы одно из этих условий не выполнено, то интеграл называется *несобственный* (*сингулярный*). Так если один из пределов интегрирования бесконечен, то говорят, что интеграл имеет *особенность*.

Например, интеграл вида:

$$\int\limits_a^\infty f(x)\, dx \tag{4.76}$$

имеет особенность на верхнем пределе.

Если переменная интегрирования x имеет непосредственный физический смысл, то реально она изменяется не до бесконечности, а до какого-то очень большого, но конечного предела Ω:

$$\int\limits_a^\infty f(x)\, dx = \int\limits_a^\Omega f(x)\, dx + \int\limits_\Omega^\infty f(x)\, dx.$$

Может оказаться, что интеграл, хотя и зависит от этого предела, но при достаточно больших его значениях практически не изменяется:

$$\lim_{\Omega\to\infty} \int\limits_\Omega^\infty f(x)\, dx = 0, \quad \int\limits_a^\infty f(x)\, dx = \lim_{\Omega\to\infty} \int\limits_a^\Omega f(x)\, dx. \tag{4.77}$$

Такой интеграл называется *сходящимся*.

Например ($a \neq 0$)

$$\int\limits_0^\infty e^{-ax} dx = \frac{1}{a} \int\limits_0^\infty e^{-ax} d(ax) = -\frac{1}{a}\, e^{-ax}\Big|_0^\infty = -\frac{1}{a}(0-1) = \frac{1}{a}. \tag{4.78}$$

Основной вклад в сходящийся интеграл дает его «конечная» («собственная») часть, тогда как вклад особенности при достаточно большом пределе интегрирования как угодно мал. Если значение интеграла с ростом предела интегрирования не «устанавливается», а стремится к бесконечности или колеблется, то интеграл называется *расходящимся*.

Интегралы, имеющие особенности, как сходящиеся, так и расходящиеся, могут, тем не менее, мажорироваться какой-нибудь функцией. Наиболее часто этот факт распознается при помощи сравнения подынтегральной функции $f(x)$ со степенной функцией $\frac{const}{x^n}$, от которой интеграл легко берется:

$$\int\limits_{x_0}^{\infty} \frac{const}{x^n}\,dx = \frac{const}{(n-1)x_0^{n-1}} - \frac{const}{(n-1)\infty^{n-1}}.$$

Здесь надо различать два случая. Если $n > 1$, то последний член в правой части равен нулю, и в этом случае интеграл сходится. Если $n \leqslant 1$, то при $n = 1$ последний член в правой части равен $const \cdot \ln\infty$, а при $n < 1$ равен $\frac{const\cdot\infty^{1-n}}{n-1}$, и интеграл в любом случае расходится.

Интегралы (4.2) и (4.3) при интегрировании на интервале $[0, \infty)$ и при отрицательных значениях степени показательной функции, очевидно, сходящиеся, так как показательная функция стремится к нулю быстрее любой степени x. Получаем:

$$\int\limits_0^{\infty} e^{-ax}\cos(bx)\,dx = \frac{a}{a^2+b^2}, \qquad \text{где } a > 0; \tag{4.79}$$

$$\int\limits_0^{\infty} e^{-ax}\sin(bx)\,dx = \frac{a}{a^2+b^2}, \qquad \text{где } a > 0. \tag{4.80}$$

В сходящемся интеграле (4.80) примем $b = 1$, а величину a примем в качестве переменной $(0 < a_1 \leqslant a \leqslant a_2 < \infty)$:

$$\int\limits_{a_1}^{a_2} da \int\limits_0^{\infty} e^{-ax}\sin x\,dx \overset{(4.16)}{=} \int\limits_0^{\infty} \sin x\,dx \int\limits_{a_1}^{a_2} e^{-ax}da = \int\limits_0^{\infty} \left(e^{-a_1 x} - e^{-a_2 x}\right)\frac{\sin x}{x}\,dx;$$

$$\int\limits_{a_1}^{a_2} da \int\limits_0^{\infty} e^{-ax}\sin x\,dx \overset{(4.80)}{=} \int\limits_{a_1}^{a_2} \frac{da}{a^2+1} = \operatorname{arctg} a_2 - \operatorname{arctg} a_1.$$

В результате получаем:

$$\int\limits_0^{\infty} \left(e^{-a_1 x} - e^{-a_2 x}\right)\frac{\sin x}{x}\,dx = \operatorname{arctg} a_2 - \operatorname{arctg} a_1.$$

Приняв $a_1 = 0$ и перейдя к пределу при $a_2 \to \infty$, получим:

$$\int\limits_0^{\infty} \frac{\sin x}{x}\,dx = \frac{\pi}{2}, \quad \int\limits_{-\infty}^{+\infty} \frac{\sin x}{x}\,dx = \pi. \tag{4.81}$$

Для получения второй формулы (4.81) нужно заменить x на $-x$.

Рассмотрим интеграл:

$$\iint\limits_S e^{-x^2-y^2}\,d\sigma \overset{(4.15)}{=} \int\limits_{-\infty}^{+\infty} e^{-x^2}dx \cdot \int\limits_{-\infty}^{+\infty} e^{-y^2}dy = I^2. \tag{4.82}$$

Заменим переменные $x = \rho\cos\varphi$, $y = \rho\sin\varphi$, $x^2 + y^2 = \rho^2$

$$\iint\limits_S e^{-\rho^2}\rho\,d\varphi\,d\rho = \int\limits_0^{2\pi} d\varphi \cdot \int\limits_0^{\infty} e^{-\rho^2}\rho\,d\rho = \pi. \tag{4.83}$$

Приравнивая (4.82) и (4.83) ($I^2 = \pi$), получаем

$$\int\limits_{-\infty}^{+\infty} e^{-x^2}dx = \sqrt{\pi}. \tag{4.84}$$

Применим формулу (4.17) к сходящемуся интегралу:

$$\int\limits_0^{\infty} e^{-\alpha x}dx = \frac{1}{\alpha}, \quad \alpha > 0.$$

Дифференцируя по параметру α n раз, получим:

$$\int\limits_0^{\infty} (-x)^n e^{-\alpha x}dx = (-1)\ldots(-n)\alpha^{-(n+1)}.$$

Полагая $\alpha = 1$ и сокращая на $(-1)^n$, получим:

$$\int\limits_0^{\infty} x^n e^{-x}dx = n!. \tag{4.85}$$

Для вычисления несобственных интегралов также можно применять формулы для вычетов функции комплексного переменного.

Теорема. Допустим функция f определена не только для вещественной переменной x: $f(x)$, но и для комплексной переменной z: $f(z)$ и при этом удовлетворяет следующим условиям ($m, k, n \in \mathbb{N}$):

1. $f(z)$ голоморфна во всей верхней комплексной полуплоскости, за исключением m полюсов $z_1, \ldots, z_m$, расположенных над вещественной остью x, и k полюсов $z_{m+1}, \ldots, z_{m+k}$, расположенных на самой вещественной оси x;

2. если $|z|$ стремится к бесконечности, то $zf(z)$ стремится к нулю равномерно относительно аргумента z ($0 \leqslant z \leqslant \pi$);

3. $\int\limits_0^{+\infty} f(x)\,dx$ и $\int\limits_{-\infty}^{0} f(x)\,dx$ сходятся.

В этом случае несобственный интеграл

$$\int\limits_{-\infty}^{+\infty} f(x)\,dx = 2\pi i \sum_{j=1}^{m} \operatorname{Res}_{z_j} f(z) + \pi i \sum_{j=m+1}^{m+k} \operatorname{Res}_{z_j} f(z). \tag{4.86}$$

вычисляется как умноженная на $2\pi i$ сумма вычетов, расположенных в верхней комплексной полуплоскости, и умноженная на πi сумма вычетов на самой вещественной оси.

Несобственный интеграл

$$\int\limits_{-\infty}^{+\infty} f(x)e^{inx}\,dx = 2\pi i \sum_{j=1}^{m} \operatorname{Res}_{z_j} f(z) + \pi i \sum_{j=m+1}^{m+k} \operatorname{Res}_{z_j} f(z). \tag{4.87}$$

вычисляется аналогично, причём условие 2 становится менее обременительным: если $|z|$ стремится к бесконечности, то $f(z)$ стремится к нулю равномерно относительно аргумента z $(0 \leqslant z \leqslant \pi)$.

Пример. Применим формулу Эйлера (1.62) к (4.87) и разделим вещественную и мнимую части, тогда:

$$\int\limits_{0}^{\infty} f(x)\cos(nx)\,dx = \pi i \sum_{j=1}^{m} \operatorname{Res}_{z_j} f(z) + \frac{\pi i}{2} \sum_{j=m+1}^{m+k} \operatorname{Res}_{z_j} f(z), \tag{4.88}$$

$$\int\limits_{0}^{\infty} f(x)\sin(nx)\,dx = \pi \sum_{j=1}^{m} \operatorname{Res}_{z_j} f(z) + \frac{\pi}{2} \sum_{j=m+1}^{m+k} \operatorname{Res}_{z_j} f(z). \tag{4.89}$$

С учётом (4.28) $\operatorname{Res}_0 \frac{1}{z} = 1$ и (4.89) мы сразу получаем формулу (4.81):

$$\int\limits_{0}^{\infty} \frac{\sin x}{x}\,dx = \frac{\pi}{2} \operatorname{Res}_0 \frac{1}{z} = \frac{\pi}{2}.$$

4.5 Скалярный и векторный потенциалы

Несобственные интегралы широко применяются для вычисления скалярных и векторных *потенциалов* (3.75–3.79), для которых оказываются справедливы выражения следующего вида:

$$\varphi_V(\xi) = \iiint\limits_V \frac{\rho(\nu)}{r}\,dV, \tag{4.90}$$

$$\varphi_S(\xi) = \iint\limits_S \frac{\varrho(\nu)}{r}\,d\sigma, \tag{4.91}$$

$$\varphi_S(\xi) = \iint\limits_S \varrho(\nu) \frac{\partial}{\partial n}\left(\frac{1}{r}\right) d\sigma, \tag{4.92}$$

где $\varphi_V(\xi)$ — *потенциал объёма*,
$\varphi_S(\xi)$ — *потенциал поверхности*,
$\rho(\nu)$ — *объёмная плотность*,
$\varrho(\nu)$ — *поверхностная плотность*,
$r = |\xi - \nu|$ — расстояние от точки вычисления потенциала ξ до точки интегрирования ν.

Поверхность S считается незамкнутой и достаточно хорошей (гладкой) или дифференцируемой в каждой точке. Это означает, что можно построить касательную плоскость в каждой точке поверхности.

При вычислении потенциала под интегралом в точке ξ возникает неопределённость, так как $\nu = \xi$, следовательно, $r = 0$ и подынтегральная функция в точке $r = 0$ обращается в бесконечность. Значит, (4.90–4.92) — несобственные интегралы. Поэтому нужно использовать специальные методы интегрирования, чтобы раскрыть эту неопределённость.

Обратим внимание:

1. Сразу видно, что потенциал не растёт до бесконечности при значении $r = 0$, так как

$$\lim_{r\to 0} \frac{dV}{r} = 0, \quad \lim_{r\to 0} \frac{d\sigma}{r} = 0, \quad \lim_{r\to 0} \frac{\mathbf{r}\, dV}{r^3} = 0; \tag{4.93}$$

в то время как, например,

$$\lim_{r\to 0} \frac{\mathbf{r}\, d\sigma}{r^3} \neq 0. \tag{4.94}$$

Доказательство.

Чтобы раскрыть неопределённость $\frac{0}{0}$ при вычислении потенциала в точке ξ $(r = 0)$ поместим начало координат в точку ξ, рассмотрим окрестность точки ξ в виде шара D радиусом r и запишем элементы площади и объёма в сферических координатах

$$d\sigma \overset{(3.52)}{=} e_r r^2 \sin\theta\, d\varphi \wedge d\theta + e_\varphi r\, d\theta \wedge dr + e_\theta r \sin\theta\, dr \wedge d\varphi,$$
$$dV \overset{(3.53)}{=} r^2 \sin\theta\, dr \wedge d\varphi \wedge d\theta.$$

Так как интегрирование ведётся по поверхности шара ∂D то $dr = 0$ и соответствующие члены будут обращаться в нуль. Тогда с учётом $\mathbf{r} \parallel e_r$ получим

$$\lim_{r\to 0} \frac{d\sigma}{r} = \lim_{r\to 0} [\, e_r r \sin\theta\, d\varphi \wedge d\theta + e_\varphi d\theta \wedge dr + e_\theta \sin\theta\, dr \wedge d\varphi\,] = 0,$$
$$\lim_{r\to 0} \frac{\mathbf{r} d\sigma}{r^3} = \lim_{r\to 0} [\sin\theta\, d\varphi \wedge d\theta\,] \neq 0, \quad \text{так как не содержит } r,$$
$$\lim_{r\to 0} \frac{dV}{r} = \lim_{r\to 0} [\, r \sin\theta\, dr \wedge d\varphi \wedge d\theta\,] = 0,$$
$$\lim_{r\to 0} \frac{\mathbf{r}\, dV}{r^3} = \lim_{r\to 0} [\sin\theta\, dr \wedge d\varphi \wedge d\theta\,] = 0.$$

2. Несобственные интегралы (4.90) сходятся при $\xi \to \infty$, если ρ и μ интегрируемы во всей области, так как если $\xi \to \infty$, то $r \to \infty$. Тогда

$$\lim_{r\to\infty} \iiint\limits_V \frac{\rho(\nu)}{r}\, dV \leqslant \lim_{r\to\infty} \iiint\limits_V \frac{|\rho(\nu)|}{r}\, dV \leqslant \lim_{r\to\infty} \frac{const}{r} = 0, \tag{4.95}$$

$$\lim_{r\to\infty} \iint\limits_S \frac{\varrho(\nu)}{r}\, d\sigma \leqslant \lim_{r\to\infty} \iint\limits_S \frac{|\varrho(\nu)|}{r}\, d\sigma \leqslant \lim_{r\to\infty} \frac{const}{r} = 0. \tag{4.96}$$

Следствие. Любой потенциал на бесконечности равен нулю.

Теория потенциала возникла из потребностей физики как чисто прикладной метод решения уравнений. В первоначальном понимании — это учение о свойствах сил, действующих по закону всемирного тяготения. В формулировке этого закона, данной И. Ньютоном, речь идет только о силах взаимного притяжения, действующих на две материальные частицы малых размеров (материальные точки), прямо пропорциональных произведению масс этих частиц и обратно пропорциональных квадрату расстояния между частицами. Поэтому первой и важнейшей с точки зрения небесной механики и геодезии задачей было изучение сил притяжения материальной точки ограниченным гладким материальным телом – сфероидом и, в частности, эллипсоидом (ибо многие небесные тела имеют именно эту форму).

После первых частных достижений И. Ньютона и других ученых основное значение здесь имели работы Ж. Лагранжа, А. Лежандра и П. Лапласа. Ж. Лагранж установил, что поле сил тяготения является, как говорят теперь, — потенциальным, и ввел функцию, которую позже Дж. Грин назвал потенциальной, а К. Гаусс — просто потенциалом. Ныне достижения этого первоначального периода обычно входят в курсы классической небесной механики.

Далее К. Гаусс и его современники обнаружили, что метод потенциалов применим не только для решения задач теории тяготения, но и вообще для решения широкого круга задач математической физики, в частности электростатики и магнетизма. В связи с этим стали рассматриваться потенциалы не только масс, но и, например, зарядов и токов.

Покажем, что для скалярных и векторных потенциалов справедливы уравнения (4.90, 4.91).

Пусть скалярная функция $\varphi(\xi)$ конечна и непрерывна во всём объёме V, имеет внутри этого объёма конечные производные как первого, так и второго второго порядка и удовлетворяет
внутри V — уравнению Пуассона

$$\Delta\varphi_V = -4\pi\rho, \tag{4.97}$$

снаружи V — уравнению Лапласа

$$\Delta\varphi_V = 0, \tag{4.98}$$

а сплошность (непрерывность) её градиента может нарушаться на отдельных «поверхностях разрыва» S (изменяться скачком на конечную величину по различным сторонам поверхности), при этом

$$\nabla\varphi_S^{(2)} - \nabla\varphi_S^{(1)} = 4\pi\varrho, \tag{4.99}$$

где индекс $^{(2)}$ соответствует наружней, индекс $^{(1)}$ — внутренней сторонам поверхности S, а поверхностная плотность $\varrho(\nu)$ непрерывна на поверхности S и представляет собой вектор нормали в данной точке поверхности. Тогда для вычисления этой функции справедливо уравнение

$$\varphi(\xi) = \iiint\limits_V \frac{\rho(\nu)}{r}\,dV + \iint\limits_S \frac{\varrho(\nu)}{r}\,d\sigma. \tag{4.100}$$

где $r = |\xi - \nu|$ — расстояние от точки вычисления функции до точки интегрирования.

Доказательство.

Воспользуемся теоремой Грина (4.75), где $v = \frac{1}{r}$ и $u = \varphi$

$$\oiint\limits_S \left(\varphi\nabla\left(\frac{1}{r}\right) - \frac{\nabla\varphi}{r}\right) d\sigma \overset{(4.75)}{=} \iiint\limits_V \left(\varphi\,\Delta\left(\frac{1}{r}\right) - \frac{\Delta\varphi}{r}\right) dV \overset{\substack{(4.98)\\(3.193)}}{=} 4\pi \iiint\limits_V \frac{\rho(\nu)}{r}\,dV.$$

Так как теорема Грина применима только к таким участкам пространства, в которых оба скаляра v и u, и их производные непрерывны, то точку ξ необходимо исключить из области интегрирования V. С этой целью исключим из области V окрестность точки ξ в виде маленького шара D и добавим к поверхности S сферу ∂D. Тогда

$$4\pi \iiint\limits_{[V-D]} \frac{\rho(\nu)}{r}\,dV = \oiint\limits_S \left(\varphi\nabla\left(\frac{1}{r}\right) - \frac{\nabla\varphi}{r}\right) d\sigma + \oiint\limits_{\partial D} \left(\varphi\nabla\left(\frac{1}{r}\right) - \frac{\nabla\varphi}{r}\right) d\sigma,$$

где второй член в правой части равен

$$\oiint\limits_{\partial D} \left(\varphi\nabla\left(\frac{1}{r}\right) - \frac{\nabla\varphi}{r}\right) d\sigma \overset{\substack{(4.9)\\(3.172)}}{=} \overline{\varphi} \oiint\limits_{\partial D} \frac{\mathbf{r}\,d\sigma}{r^3} - \overline{\nabla\varphi} \oiint\limits_{\partial D} \frac{d\sigma}{r} =$$

$$\overset{\substack{(4.93)\\(4.94)\\(3.52)}}{=} \overline{\varphi} \oiint\limits_{\partial D} \sin\theta\,d\varphi\wedge d\theta \overset{(4.13)}{=} 4\pi\overline{\varphi}.$$

Перед $\overline{\varphi}$ стоит знак $+$, а не $-$, так как уравнение (3.172) записано для радиус-вектора, направленного от центра, а в нашем случае внешняя к объёму интегрирования $V - D$ нормаль к поверхности сферы ∂D направлена к её центру — точке ξ и прямо противоположна радиус-вектору $\mathbf{r}$.

Среднее значение $\overline{\varphi}$ на поверхности маленькой сферы ∂D с центром в точке ξ, очевидно, имеет пределом значение $\varphi(\xi)$ в этой точке.

Тогда получаем

$$\iiint\limits_V \frac{\rho(\nu)}{r}\,dV = \lim_{D\to 0}\iiint\limits_{[V-D]} \frac{\rho(\nu)}{r}\,dV = \frac{1}{4\pi}\oiint\limits_S \left(\varphi\,\nabla\left(\frac{1}{r}\right) - \frac{\nabla\varphi}{r}\right)d\sigma + \lim_{\partial D\to\xi}\overline{\varphi} =$$
$$= \frac{1}{4\pi}\oiint\limits_S \left(\varphi\,\nabla\left(\frac{1}{r}\right) - \frac{\nabla\varphi}{r}\right)d\sigma + \varphi(\xi).$$

Рассмотрим теперь поверхностный интеграл — первый член в правой части. Поверхность интегрирования может состоять, во-первых, из пограничной поверхности объёма V, однако результат интегрирования по этой поверхности всегда можно принять равным нулю, так как при удалении границы объёма на бесконечность справедливо уравнение (4.96), во-вторых, из незамкнутой поверхности S, на которой по условию скалярная функция φ остаётся непрерывной $\varphi^{(2)} = \varphi^{(1)}$, а её градиент имеет разрыв (4.99).

Так как S не замкнута, то интегрирование по замкнутой поверхности сводится к *двукратному* интегрированию: один раз по внутренней, а другой раз по наружней стороне этой поверхности. Причём на внутренней стороне поверхности величина $d\sigma$ меняет знак. Тогда получаем

$$\oiint\limits_S \left(\varphi\,\nabla\left(\frac{1}{r}\right) - \frac{\nabla\varphi}{r}\right)d\sigma = \iint\limits_S \left(\left(\varphi^{(2)} - \varphi^{(1)}\right)\nabla\left(\frac{1}{r}\right) - \frac{\nabla\varphi^{(2)} - \nabla\varphi^{(1)}}{r}\right)d\sigma =$$
$$\overset{(4.99)}{=} -4\pi\iint\limits_S \frac{\varrho}{r}\,d\sigma.$$

Вместе с предыдущей формулой это даёт искомое уравнение (4.100).

Докажем теперь обратную теорему. Скалярный потенциал объёма $\varphi_V(\xi)$ (4.90) удовлетворяет внутри V — уравнению Пуассона (4.97), снаружи V — уравнению Лапласа (4.98).

Доказательство. Чтобы раскрыть неопределённость $\frac{0}{0}$ при вычислении потенциала в точке ξ ($r = 0$) исключим из области интегрирования V окрестность точки ξ в виде маленького шара D

Запишем

$$\varphi_V(\xi) = \iiint\limits_{[V-D]} \frac{\rho(\nu)}{r}\,dV + \iiint\limits_D \frac{\rho(\nu)}{r}\,dV = \varphi_1(\xi) + \varphi_2(\xi)$$

Член $\varphi_1(\xi)$ удовлетворяет уравнению Лапласа $\Delta\varphi_1(\xi) = 0$, т.е. является гармонической функцией

$$\Delta\varphi_1(\xi) = \iiint\limits_{[V-D]} \Delta\left(\frac{\rho(\nu)}{r}\right)dV =$$
$$\overset{(3.194)}{=} \iiint\limits_{[V-D]} \frac{\Delta\rho(\nu)}{r}\,dV - \iiint\limits_{[V-D]} 2\frac{\mathbf{r}\nabla\rho(\nu)}{r^3}\,dV \overset{(4.93)}{=} 0.$$

Для члена $\varphi_2(\xi)$ получаем

$$\Delta\varphi_2(\xi) = \iiint\limits_D \nabla^2 \left(\frac{\rho(\nu)}{r}\right) dV \overset{(4.70)}{=} \oiint\limits_{\partial D} \nabla \left(\frac{\rho(\nu)}{r}\right) d\sigma =$$
$$\overset{(3.184)}{=} \oiint\limits_{\partial D} \frac{\nabla\rho(\nu)}{r}\, d\sigma - \oiint\limits_{\partial D} \frac{\mathbf{r}\rho(\nu)}{r^3}\, d\sigma =$$
$$\overset{\substack{(4.93)\\(4.94)}}{=} - \oiint\limits_{\partial D} \rho(\nu)\sin\theta\, d\varphi \wedge d\theta \overset{\substack{(4.12)\\(4.13)}}{=} -4\pi\rho(\xi).$$

Если теперь ξ находится вне области V, то в шаре D плотность тождественна нулю $\rho(\xi) \equiv 0$. Тогда $\Delta\varphi_V(\xi) = 0$, что и требовалось доказать для (4.98). Если же ξ находится внутри области V, то $\Delta\varphi_V(\xi) = -4\pi\rho(\xi)$, что и требовалось доказать для (4.97).

Мы показали, что выражение (4.90) $\varphi_V(\xi) = \iiint_V \frac{\rho(\nu)}{r}\, dV$ является решением (интегралом) дифференциального уравнения Пуассона(4.97) $\Delta\varphi_V = -4\pi\rho$. Следовательно, для векторного потенциала объёма $\varphi_V = (\varphi_V^x, \varphi_V^y, \varphi_V^z)$ выражения

$$\varphi_V^x(\xi) = \iiint\limits_V \frac{\rho^x(\nu)}{r}\, dV, \quad \varphi_V^y(\xi) = \iiint\limits_V \frac{\rho^y(\nu)}{r}\, dV, \quad \varphi_V^z(\xi) = \iiint\limits_V \frac{\rho^z(\nu)}{r}\, dV.$$

есть решения следующих дифференциальных уравнений Пуассона

$$\nabla^2\varphi_V^x = -4\pi\rho^x, \quad \nabla^2\varphi_V^y = -4\pi\rho^y, \quad \nabla^2\varphi_V^z = -4\pi\rho^z.$$

Эти три уравнения для слагающих вектора φ_V, согласно (3.196, 3.197), равнозначны одному векторному уравнению

$$\nabla^2\varphi_V(\xi) = -4\pi\rho(\xi), \tag{4.101}$$

которое и является искомым дифференциальным уравнением для векторного потенциала.

До сих пор мы считали потенциал φ непрерывным, и только для его производных непрерывность могла нарушаться. Покажем, что выражение (4.92) $\varphi_S(\xi) = \iint_S \varrho \frac{\partial}{\partial n}\left(\frac{1}{r}\right) d\sigma$ применимо в случае, когда существует поверхность, на которой нарушена непрерывность самого потенциала.

Очевидно, что на «поверхности разрыва» может нарушаться непрерывность потенциала поверхности, а не потенциала объёма. Это означает, что либо на двух сторонах одной поверхности, либо на двух весьма близких и параллельных друг другу поверхностях наблюдаются различные поверхностные плотности.

Наиболее распространённым считается примеры из электродинамики, когда образуется *двойной электрический слой*, либо двойной магнитный слой, или, как его принято называть, *магнитный листок*. В этом случае

на противолежащих элементах обеих поверхностей (двух сторон одной поверхности) поверхностные плотности равны по величине и противоположны по знаку $\mu = \mu^{(2)} = -\mu^{(1)}$ (рис. 4.2).

Потенциал этого слоя в некоторой точке ξ поля будет равен

$$\varphi_S(\xi) \overset{(4.91)}{=} \iint\limits_S \frac{\mu^{(2)}}{r^{(2)}}\, d\sigma + \iint\limits_S \frac{\mu^{(1)}}{r^{(1)}}\, d\sigma = \iint\limits_S \mu \left(\frac{1}{r^{(2)}} - \frac{1}{r^{(1)}} \right) d\sigma.$$

Величина $\frac{1}{r^{(2)}} - \frac{1}{r^{(1)}}$ есть приращение численной величины обратного значения радиус-вектора $\mathbf{r}$, проведённого от точки ξ к рассматриваемому элементу двойного слоя, при перемещении начальной точки вектора от внутренней поверхности к внешней. Пусть $\mathbf{n}$ есть направление нормали к двойному слою, направленному от внутренней его стороны к внешней, а δ — толщина слоя. Если расстояние между слоями мало по сравнению с расстоянием от точки ξ, то можно приближённо положить:

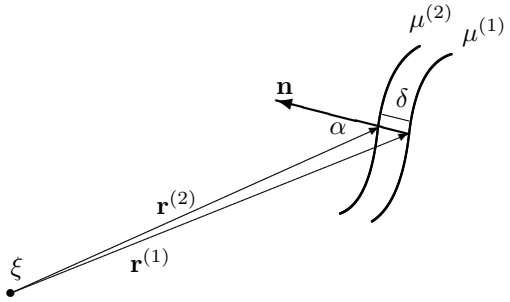


Рис. 4.2: Поверхность разрыва потенциала φ_S

$$r = r^{(2)} \approx r^{(1)}, \quad r^{(2)} - r^{(1)} \approx \delta \cos\alpha,$$

где α — угол между вектором нормали $\mathbf{n}$ и радиус-вектором $\mathbf{r}$, проведённым от точки наблюдения к рассматриваемой точки поверхности. Ввиду малости расстояния δ безразлично, к какому именно слою проведён этот радиус-вектор $\mathbf{r}$.

Таким образом, потенциал двойного слоя принимает вид (4.92), что и требовалось доказать:

$$\varphi_S(\xi) = -\iint\limits_S \frac{\mu\delta\cos\alpha}{r^2}\, d\sigma \overset{(2.3)}{=} -\iint\limits_S \frac{\mu\delta\,\mathbf{n}\cdot\mathbf{r}}{r^3}\, d\sigma \overset{\substack{(3.172)\\(3.141)}}{=} \iint\limits_S \varrho \frac{\partial}{\partial n}\left(\frac{1}{r}\right) d\sigma,$$

где $\varrho = \mu\delta$ — условная «поверхностная плотность», например, величина дипольного момента в электродинамике.

4.6 Дельта-функция Дирака (δ-функция)

Исследуем несобственный интеграл вида:

$$\frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} e^{ixy} dy. \tag{4.102}$$

Представим этот несобственный интеграл как предел интеграла по конечному промежутку $[-y, +y]$:

$$\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} e^{ixy}dy = \frac{1}{2\pi}\lim_{y\to\infty}\int\limits_{-y}^{+y} e^{ixy}dy = \lim_{y\to\infty}\frac{e^{ixy}-e^{-ixy}}{2\pi ix} =$$

$$= \lim_{y\to\infty}\frac{\sin(xy)}{\pi x} = \lim_{y\to\infty} F_y,$$

где $F_y = \frac{\sin(xy)}{\pi x}$.

Функция F_y является функцией двух переменных x, y. Рассматривая различные значения y, мы получим семейство кривых F_y. Граница $F(x)$ области, заполненной семейством кривых F_y, касающихся этой границы, носит название *огибающей* семейства кривых.

Если мы рассмотрим поверхность функции двух переменных F_y в трёхмерном пространстве (F_y, x, y); то при фиксированном значении x_0 мы получим кривую $F_y(x_0, y)$, максимум (минимум) которой даст точку огибающей. Таким образом, мы получаем условие для уравнения огибающей семейства кривых:

$$\left(\frac{\partial F_y(x,y)}{\partial y}\right)_x = 0. \tag{4.103}$$

Условие (4.103) дает нам уравнение $y = y(x)$, связывающее между собой x и y. Подставив это уравнение в F_y, мы получим функцию огибающей семейства кривых $F(x) = F_y(x, y(x))$:

$$\left(\frac{\partial F_y(x,y)}{\partial y}\right)_x = \frac{\cos(xy)}{\pi} = 0, \quad \text{т.е. } \sin(xy) = \pm 1,\ F(x) = \pm\frac{1}{\pi x}. \tag{4.104}$$

График функции $F_y(x, y)$ и ее огибающей $F(x)$ изображен на рис. 4.3:

Интеграл от $F_y(x, y)$ по x при произвольном значении y равен:

$$\int\limits_{-\infty}^{+\infty} F_y(x,y)\,dx = \int\limits_{-\infty}^{+\infty}\frac{\sin(xy)}{\pi x}\,dx \quad = \frac{1}{\pi}\int\limits_{-\infty}^{+\infty}\frac{\sin(xy)}{xy}\,d(xy) = \frac{1}{\pi}\int\limits_{-\infty}^{+\infty}\frac{\sin z}{z}\,dz = 1.$$

При $x = 0$ значение $F_y(0, y) = y$.

Функция F_y симметрична относительно переменной x: $F_y(x, y) = F_y(-x, y)$.

При увеличении y главный максимум F_y увеличивает свою высоту и, соответственно, уменьшает ширину, что вызывает соответствующее увеличение числа колебаний функции на фиксированном интервале.

Неограниченно увеличивая y, получим функцию со следующими свойствами:

1. функция равна нулю при всех $x \neq 0$;
2. функция бесконечна при $x = 0$;
3. интеграл от функции по переменной x, взятый в пределах от $-\infty$, до $+\infty$, равен 1.

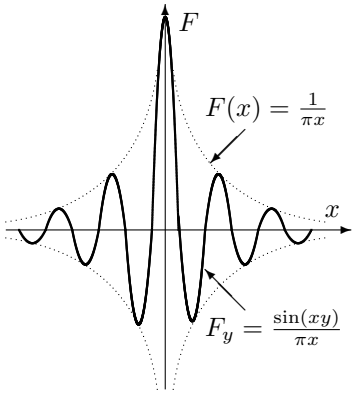


Рис. 4.3: Графики функций $F(x)$ и $F_y(x, y)$

Функция с этими свойствами называется *дельта-функция Дирака* и обозначается $\delta(x)$. Функция $\delta(x)$ чрезвычайно удобна и широко применяется в физике.

Мы получили соотношения:

$$\delta(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} e^{ixy} dx = \lim_{y\to\infty} \frac{\sin(xy)}{\pi x}, \tag{4.105}$$

$$\int\limits_{-\infty}^{+\infty} \delta(x)\, dx = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} e^{ixy} dx\, dy = 1. \tag{4.106}$$

Можно получить еще два выражения для функции $\delta(x)$:

$$\delta(x) = \lim_{y\to\infty} \frac{y}{\pi} \frac{1}{1 + (xy)^2}, \tag{4.107}$$

$$\delta(x) = \lim_{y\to\infty} \sqrt{\frac{y}{\pi}} e^{-x^2 y}. \tag{4.108}$$

Из свойств $\delta(x)$ следуют основные соотношения для нее. Так как $\delta(x) = 0$ при всех $x \neq 0$, поэтому промежуток интегрирования можно сузить до малой окрестности точки нуль $[-\varepsilon, +\varepsilon]$, где ε — малая величина. На этом промежутке $f(x) \approx f(0)$, следовательно:

$$\begin{aligned} \int\limits_{-\infty}^{+\infty} \delta(x) f(x)\, dx = \int\limits_{-\varepsilon}^{+\varepsilon} \delta(x) f(x)\, dx = \int\limits_{-\varepsilon}^{+\varepsilon} \delta(x) f(0)\, dx = \\ = f(0) \int\limits_{-\varepsilon}^{+\varepsilon} \delta(x)\, dx = f(0) \int\limits_{-\infty}^{+\infty} \delta(x)\, dx = f(0). \end{aligned} \tag{4.109}$$

Делая аналогичные выкладки, можно получить:

$$\int\limits_{-\infty}^{+\infty} \delta(x - a) f(x)\, dx = f(a). \tag{4.110}$$

Отметим еще некоторые интересные формулы для дельта-функции. Прежде всего:

$$\int\limits_{-\infty}^{+\infty} |\lambda|\delta(\lambda x)\,dx = \int\limits_{-\infty}^{+\infty} \delta(\lambda x)\,d(\lambda x) = 1, \quad \text{при } \lambda \neq 0.$$

Из этого вычисления следует:

$$\delta(\lambda x) = \frac{\delta(x)}{|\lambda|}, \quad \text{при } \lambda \neq 0. \tag{4.111}$$

Если функция $f(x)$ обращается в нуль лишь при $x = 0$, то вблизи нуля с точностью до малых высшего порядка можно написать

$$f(x) = f(0) + x\,f'(0) = x\,f'(0).$$

Тогда

$$\delta(f(x)) = \frac{\delta(x)}{|f'(0)|}. \tag{4.112}$$

Наконец, из равенства функции $\delta(x-a)$ нулю при $x \neq a$ вытекает:

$$f(x)\,\delta(x-a) = f(a)\,\delta(x-a) \tag{4.113}$$

При помощи δ-функции легко записать некоторые, имеющие большое значение специальные функции, например — *единичную функцию*:

$$e(x) = \int\limits_{-\infty}^{x} \delta(x)\,dx, \quad \text{при} \begin{cases} x < 0, & e(x) = 0, \\ x > 0, & e(x) = 1. \end{cases} \tag{4.114}$$

При интегрировании единичной функции получают:

$$f(x) = \int\limits_{-\infty}^{x} e(x)\,dx, \quad \text{при} \begin{cases} x < 0, & f(x) = 0, \\ x > 0, & f(x) = x. \end{cases} \tag{4.115}$$

С $e(x)$ тесно связана функция $\operatorname{sign} x$ (от лат. signum — знак):

$$\operatorname{sign} x = \frac{x}{|x|} = 2\,e(x) - 1. \tag{4.116}$$

В пространстве $\mathbb{R}^n$ дельта-функция имеет вид:

$$\delta(x^1, \ldots, x^n) = \delta(x^1) \cdot \ldots \cdot \delta(x^n). \tag{4.117}$$

С дельта-функцией непосредственно связано еще одно полезное расширение интеграла — *интеграл Стильтьеса*. Допустим, мы рассматриваем интеграл от некой функции (называемой в механике силой) $f(x)$, действующей на стержень длины l с распределенной плотностью $\rho(x)$:

$$\int\limits_{(l)} f(x)\rho(x)\,dx. \tag{4.118}$$

Если на отрезке стержня есть распределенная плотность $\rho_p(x)$ и точечная масса m_a в точке a, то общая масса стержня записывается как:

$$m = m_a + \int_0^l \rho_p(x)\,dx. \tag{4.119}$$

В этом случае интеграл (4.118) записать невозможно, так как точечная масса не учитывается функцией $\rho_p(x)$, но можно понимать его как *интеграл по мере* μ, (он также называется интегралом Стилтьеса), который имеет вид:

$$\int_{(l)} f(x)\,d\mu(x). \tag{4.120}$$

где $\mu(x)$ играет роль функции m.

Если функция $\mu(x)$ дифференцируема: $d\mu = \rho(x)\,dx$, то интеграл (4.120) превращается в интеграл (4.118). В противном случае мы можем при помощи дельта-функции определить общую плотность $\rho(x)$ как:

$$\rho(x) = \rho_p(x) + m_a\delta(x-a).$$

Тогда полная масса m равна:

$$m = \int_0^l \rho(x)\,dx. \tag{4.121}$$

Допуская дельтообразные слагаемые, интеграл вида (4.121) можно переписать в интеграл вида (4.118), даже если функция μ не дифференцируема в обычном смысле слова.

Глава 5

Фазовое пространство

Уравнения, в которых содержатся производные искомых функций, называются *дифференциальными уравнениями*. Если уравнения не содержат производных выше первого порядка, то они называются *обыкновенными дифференциальными уравнениями*.

На самом деле в данной главе изложены элементы теории обыкновенных дифференциальных уравнений. Однако, оказалось, что систему обыкновенных дифференциальных уравнений удобно изучать в так называемом фазовом пространстве, поэтому на протяжении всей главы мы будем изучать свойства этого пространства.

Изложение теории обыкновенных дифференциальных уравнений в настоящей главе сильно отличается от традиционного. Как всегда мы в первую очередь будем стремиться всюду выявить геометрическую, качественную сторону изучаемых понятий. В соответствии с этим принципом при изложении материала не используется ни одной сколько-нибудь сложной формулы. Зато появляется целый ряд понятий, которые при традиционном изложении остаются в тени и не включены ни в курсы анализа, ни в курсы геометрии (фазовое пространство и фазовые потоки, векторные поля и однопараметрические группы диффеоморфизмов).

Теория дифференциальных уравнений — одно из основных орудий математического естествознания. Эта теория позволяет изучать всевозможные эволюционные процессы, обладающие свойствами *детерминированности*, *конечномерности* и *дифференцируемости*. Формулировка состояния какой-либо системы (физической, химической, биологической, экономической и т.д.) в виде точки в соответствующем фазовом пространстве и какого-либо закона природы в виде дифференциального уравнения сводит любую задачу об эволюции процесса к *геометрической задаче* о поведении фазовых кривых данного векторного поля в этом фазовом пространстве.

5.1 Однопараметрический диффеоморфизм

Действие диффеоморфизма (гладкого преобразования) можно рассматривать двояко: как замену координат пространства (многообразия) и как перенос векторного поля на новое место. Выше мы рассматривали первый случай, теперь рассмотрим второй.

Движение точки вдоль кривой $\nu(s) = (x^1(s), \ldots, x^n(s))$, заданной параметром $s \in \mathbb{R}$, можно рассматривать как перемещающее точку преобразование пространства $\varrho\colon U \to U$, где $U \in \mathbb{R}^n$, которое зависит от параметра s. Если движение вдоль кривой можно рассматривать в обоих направлениях, то функция $\varrho(x^1, \ldots, x^n)$ имеет обратную взаимно однозначную и взаимно дифференцируемую функцию ϱ^{-1}, т.е. является диффеоморфизмом. Кроме того в процессе перемещения от одного значения параметра s к другому функция ϱ также может изменяться, т.е. функция зависит также и от параметра s. Поэтому рассматриваемая функция есть диффеоморфизм с параметром s или *однопараметрический диффеоморфизм* $\varrho^s\colon \mathbb{R} \times U \to U$ или $\tau = \varrho^s(\nu)$, где $\mathbb{R}$ — ось s.

Множество (совокупность всех возможных) диффеоморфизмов ϱ^s обладает свойствами:

1. $\varrho^{s+t} = \varrho^s \circ \varrho^t$,
2. $\varrho^\circ \varrho^0 = \varrho^s$,
3. $\varrho^{-s} \circ \varrho^s = \varrho^0$, где $(\varrho^s)^{-1} = \varrho^{-s}$

и поэтому образует группу преобразований (см. § 1.3) или *однопараметрическую группу диффеоморфизмов пространства* $\mathbb{R}^n$: $\{\varrho^t\}$.

Из свойства 1 следует, что ϱ^0 — тождественное преобразование, оставляющее каждую точку на месте, поэтому свойства 2 и 3 вытекают из свойства 1. Поэтому свойство 1 называют *групповым свойством*.

Часто использую запись $\varrho^s\nu \equiv \varrho^s(\nu)$. В качестве параметра s обычно используют время t, а преобразование ϱ^t, переводящее состояние в момент 0 в состояние в момент t, называют *преобразованием за время* t. Однопараметрическая группа диффеоморфизмов пространства — это математический эквивалент физического понятия «двусторонне детерминированный процесс». Процесс называется *детерминированным* (от лат. determinans, род. падеж determinantis — определяющий), если весь его будущий ход $\varrho^t\nu$ и всё его прошлое $\varrho^{-t}\nu$ однозначно определяются состоянием в настоящее время $\varrho^0\nu$.

Каждое положение точки в пространстве (состояние системы) раньше называли *фазой*. Поэтому область пространства $U \subset \mathbb{R}^n$, называют *фазовым пространством*, каждую точку $\nu \in U$ этого пространства — *фазовой точкой*, а траекторию её движения $\nu(t)$ — *фазовой траекторией* или *фазовой кривой*. Совокупность всех фазовых траекторий в фазовом пространстве называют *фазовым портретом*, а группу диффеоморфизмов ϱ^t — *фазовым*

потоком. Произведение $\mathbb{R} \times U$ называют $(n+1)$-мерным *расширенным фазовым пространством* с координатами $t, x^1, \ldots, x^n$. График $\Gamma_\nu(t) = (t, \nu(t))$ фазовой кривой $\nu(t)$ в расширенном фазовом пространстве называют *интегральной кривой*, которую не следует путать с фазовой кривой — проекцией интегральной кривой на фазовое пространство вдоль оси t.

Скорость $\dot{\nu}$ фазовой точки называется также скоростью выхода точки из ν или *вектором фазовой скорости потока в точке* $\nu \in U$:

$$\dot{\nu}(t) = \frac{d}{dt}\left(\varrho^t \nu(0)\right) = \left.\frac{d}{dt}\right|_{t=0} \varrho^t(x^1, \ldots, x^n). \tag{5.1}$$

Вектор скорости потока $\dot{\nu}$ каждой фазовой точки задаёт касательное пространство в этой точке. Во всех точках области U векторы скорости образуют *векторное поле* $\xi = \dot{\nu} = (a_1, \ldots, a_n)$ фазовой скорости потока.

Уравнение векторного поля $\dot{\nu} = \xi$ есть система обыкновенных дифференциальных уравнений, разрешённых относительно производных:

$$\dot{\nu} = \xi(t, \nu) \text{ или } \dot{x}^i = a_i(t, x^1, \ldots, x^n), \quad i = 1, \ldots, n. \tag{5.2}$$

Система дифференциальных уравнений (5.2) называется также *динамическая система* (от греч. *δυναμικός* — сильный и *συστημα* — целое, составленное из частей) — в первоначальном значении механическая система с конечным числом *степеней свободы* (числом переменных n). Фазовую точку ν называют также *изображающей точкой* динамической системы, а каждое её положение в пространстве или фаза является текущим *состоянием* динамической системы.

Таким образом, *с каждой однопараметрической группой диффеоморфизмов* $\varrho^t(\nu)$ *связана динамическая система, заданная векторным полем фазовой скорости* $\dot{\nu} = \xi$.

Если правая часть уравнения (5.2) не зависит от t:

$$\dot{\nu} = \xi(\nu) \tag{5.3}$$

то уравнение называется *автономной системой*.

Если в качестве параметра выбрано не время t, а параметр s, то автономная и неавтономная системы имеют вид: $\nu'_s = \xi(\nu)$ и $\nu'_s = \xi(s, \nu)$.

Система (5.2) из n обыкновенных дифференциальных уравнений называется *нормальной формой* системы дифференциальных уравнений и эквивалентна *канонической форме* уравнения (одному дифференциальному уравнению порядка n), которая имеет следующий вид:

$$\frac{d^n x}{dt^n} = f\left(t, x, \frac{dx}{dt}, \ldots, \frac{d^{n-1}x}{dt^{n-1}}\right). \tag{5.4}$$

Очевидно, что уравнение (5.4) может быть преобразовано в систему (5.2) введением дополнительно $n-1$ неизвестных функций:
$x^1 = \frac{dx}{dt}, x^2 = \frac{d^2x}{dt^2}, \ldots, x^{n-1} = \frac{d^{n-1}x}{dt^{n-1}}$.

Следовательно, порядок неавтономной системы n обыкновенных дифференциальных уравнений также равен n. Если система автономная, то функции $a_1, \ldots, a_n$ не зависят от t, и порядок системы может быть уменьшен на единицу. Для этого достаточно исключить время и записать (5.2) в виде *уравнения векторной линии* (3.115):

$$\frac{dx^1}{a_1} = \ldots = \frac{dx^n}{a_n} = \frac{d\mathbf{x}}{\mathbf{a}} \tag{5.5}$$

Векторная линия совпадает с соответствующей фазовой кривой.

Отдельные точки τ, в которых вектор фазовой скорости обращается в нуль $\nu'_s = \xi(\tau) = (0, \ldots, 0)$, называются *неподвижными точками* потока, *стационарными точками*, *особыми точками* потока или *положениями равновесия*. Фазовая кривая в особой точке состоит из одной единственной изолированной точки.

Рассмотрим примеры фазовых потоков и динамических систем. Пусть фазовое пространство двумерно: $U \subset \mathbb{R}^2$. (рис. 5.1).

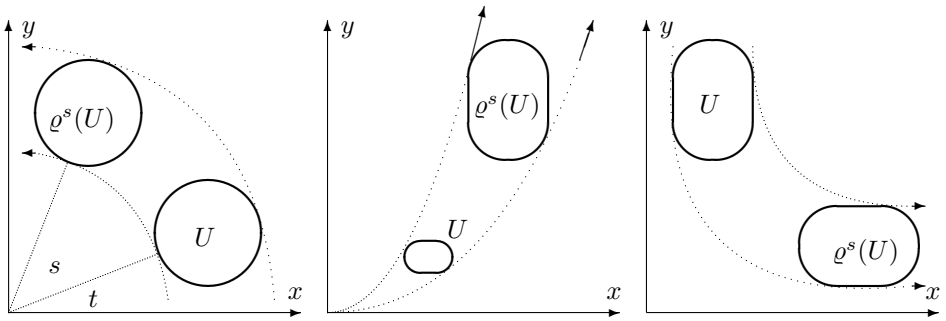


Рис. 5.1: Действие фазового потока на область U

1. Фазовый поток ϱ^s — *поворот* плоскости на угол s вокруг начала координат (точки 0). Очевидно, групповое свойство выполнено. Поворот удобно рассматривать на комплексной плоскости. Тогда $z = x + iy = re^{it}$, фазовый поток $\varrho^s(z) = ze^{is} = re^{i(t+s)}$. Фазовые кривые — неподвижная точка ($z = 0$) и окружности с центром в начале координат $x^2 + y^2 = r^2$ или $x = r\cos(t+s), y = r\sin(t+s)$, которые образуют фазовый портрет. Площади областей при поворотах не изменяются. Система имеет вид $x'_s = -y, y'_s = x$.

2. Фазовый поток $\varrho^s(x, y) = (e^s x, e^{2s} y)$ — *растяжение*. Групповое свойство вытекает из свойства экспоненты $e^{s+t} = e^s e^t$. Фазовые кривые — неподвижная точка 0, половины координатных осей и половины парабол. Площади областей увеличиваются при действии ϱ^s в e^{3s} раз. Система имеет вид $x'_s = x, y'_s = 2y$.

3. Фазовый поток $\varrho^s(x,y) = (e^s x, e^{-s} y)$ — *гиперболический поворот* (сжатие в e^s раз в направлении оси y и растяжение в e^s раз в направлении оси x). Фазовые кривые — неподвижная точка 0, половины координатных осей и половины гипербол $xy = const$. Гиперболические повороты сохраняют площади областей, хотя и сильно искажают их форму. Система имеет вид $x'_s = y, y'_s = x$.

Функция $\varphi\colon \mathbb{R} \to U$, которая удовлетворяет дифференциальным уравнениям динамической системы: $\varphi'_s = \xi(s,\nu)$ (автономной или неавтономной) называется её *решением*, а процедура отыскания решений — *интегрированием системы*. Каждая фазовая кривая $\nu(s)$ удовлетворяет этим уравнениям и является решением динамической системы, поэтому основная задача теории дифференциальных уравнений состоит в отыскании фазовых кривых по заданному векторному полю.

Отыскание решения в аналитическом виде, т.е. запись результата через элементарные функции исторически называется вычислением «в квадратурах». В общем виде эта задача не поддаётся средствам современной математики и, по-видимому, в некотором смысле неразрешима.

Параметр s называют также *направлением векторного поля* ξ. *Полем направлений* в фазовом пространстве U называют соответствие, которое каждой точке $\nu \in U$ сопоставляет проходящую через ν прямую, содержащую вектор ξ. Эта прямая есть касательная (дифференциал) фазовой кривой в данной точке: $d\nu(s)$. Поле направлений в фазовом пространстве есть геометрический эквивалент динамической системы, а фазовые кривые, которые в каждой точке имеют соответствующее касательное направление, есть её геометрические решения.

Полем направлений в расширенном фазовом пространстве $\mathbb{R} \times U$ называют соответствие, которое каждой точке $(s,\nu) \in \mathbb{R} \times U$ сопоставляет проходящую через эту точку прямую, содержащую вектор $(1, \xi(s,\nu))$. Указанная прямая есть касательная (дифференциал) $d\Gamma = (ds, d\nu(s))$ интегральной кривой $\Gamma_\nu(s) = (s, \nu(s))$ в данной точке расширенного фазового пространства. Нначальное состояние динамической системы (s_0, ν_0), задаёт интегральную кривую, которая проходит через эту точку.

Геометрическая интерпретация фазовых траекторий приводит к мысли, что через каждую точку фазового пространства автономной системы с заданным непрерывным полем направлений пройдет только одна фазовая траектория. Особая точка также есть единственная фазовая траектория в данном положении равновесия, состоящая из одной точки.

Так как в любой точке фазового пространства фазовая траектория одна, то фазовые кривые нигде не пересекутся. В противном случае, если бы фазовые кривые пересекались, то из одной точки фазового пространства выходило бы больше одной траектории, которые имели бы разные направления.

В отношении расширенного фазового пространства это предположение звучит так: через каждую точку расширенного фазового пространства проходит одна и только одна интегральная кривая соответствующей системы

дифференциальных уравнений. Эквивалентная формулировка звучит так: через каждую точку области, где задано поле направлений, проходит одна и только одна интегральная кривая этого поля.

Замечание. Фазовые кривые неавтономной системы (образы решений в фазовом пространстве) могут пересекаться, не совпадая. Поэтому за решениями неавтономных систем лучше следить по интегральным кривым в расширенном фазовом пространстве.

В отношении *существования* фазовой (интегральной) кривой сформулированная гипотеза оказывается правильной. Доказательство этого предложения принадлежит Дж. Пеано (1890). В отношении же единственности фазовой (интегральной) кривой гипотеза оказывается, вообще говоря, ошибочной, если векторное поле ξ не является гладким (дифференцируемым). Однако условие гладкости векторного поля ξ может быть ослаблено и оказывается, что достаточным условием единственности решения системы уравнений $\nu'_s = \xi(s, \nu)$ является условие Липшица: Вектор-функция $\xi\colon U \to \mathbb{R}^n$ удовлетворяет условию Липшица, если существует такая положительная константа $\mathcal{L}$, что $|\xi(\nu) - \xi(\tau)| \leqslant \mathcal{L}|\nu - \tau|$ для всех $\nu, \tau \in U$. Это означает, что вектор-функция не должна быстро меняться при малом изменении аргументов.

Теорема существования и единственности. Пусть правая часть системы уравнений $\nu'_s = \xi(s, \nu)$ в некоторой области пространства $\mathbb{R} \times U$ непрерывна и удовлетворяет условию Липшица по ν, т.е. в каждой плоскости $s = const$, с одной и той же константой $\mathcal{L}$. Тогда через каждую точку области U проходит одна и только одна фазовая кривая.

Возникает вопрос, всякое ли удовлетворяющее условию Липшица векторное поле является полем фазовой скорости потока? Ответ на этот вопрос отрицательный. *Фазовый поток возникает только на компактном фазовом пространстве.* В свою очередь гладкое или хотя бы удовлетворяющее условию Липшица векторное поле на компактном многообразии всегда определяет фазовый поток.

Основной метод решения и исследования дифференциальных уравнений — это подбор подходящей замены переменных, т.е. в геометрических терминах — подходящего диффеоморфизма, упрощающего данное поле направлений, векторное поле или фазовый поток.

Пусть M и N области, а $\varphi\colon M \to N$ — диффеоморфизм области M линейного пространства в область N линейного пространства.

Если $\tau = \varphi(\nu) \in N$, где $\nu \in M$, то векторное поле ξ в M переходит при диффеоморфизме φ в векторное поле η в N по правилу (3.96):

$$\eta = \varphi'_\nu(\tau) \cdot \xi \text{ или } dy^i = \frac{\partial y^i}{\partial x^1}\, dx^1 + \ldots + \frac{\partial y^i}{\partial x^n}\, dx^n \quad (i = 1, \ldots, n).$$

Соответственно, интегральные кривые исходного поля направлений в M переходят при диффеоморфизме φ в интегральные кривые поля направлений в N, полученного действием φ на исходное поле.

Если $\{\varrho^t\colon M \to M\}$ — однопараметрическая группа диффеоморфизмов (фигурные скобки обозначают множество, т.е. группу диффеоморфизмов

ϱ^t или поток), то образом потока $\{\varrho^t\}$ под действием диффеоморфизма φ называется поток $\{\varsigma^t\colon N \to N\}$, где $\varsigma^t = \varphi\varrho^t\varphi^{-1}$.

Если рассматривать диффеоморфизм φ как «замену переменных», то преобразование ς^t — это просто преобразование ϱ^t, «записанное в новых координатах». Потоки $\{\varrho^t\}$ и $\{\varsigma^t\}$ называют иногда эквивалентными, а диффеоморфизм φ — эквивалентностью.

Выпрямлением поля направлений называется диффеоморфизм, переводящий его в поле параллельных направлений. Интегральные кривые исходного поля переводятся при этом в интегральные прямые. Все гладкие поля направлений в областях одинакового числа измерений локально диффеоморфны (переводятся друг в друга диффеоморфизмом).

Теорема о выпрямлении. Всякое гладкое поле направлений выпрямляемо в окрестности каждой точки. Если поле k раз непрерывно дифференцируемо (класса C^k, $q \leqslant k \leqslant \infty$), то и выпрямляющий диффеоморфизм можно выбрать класса C^k.

Выпрямлением векторного поля называется диффеоморфизм, превращающий его в поле параллельных векторов одинаковой длины в евклидовом пространстве. Векторы векторного поля ξ определяют поле направлений в этой области фазового пространства, следовательно, всякое гладкое векторное поле локально выпрямляемо в окрестности каждой неособой точки (точки, где вектор поля отличен от нуля $\xi \neq 0$).

Выпрямление поля было открыто Ньютоном, который зашифровал метод доказательства теоремы о выпрямлении в знаменитом «втором письме» Ньютона к секретарю Королевского общества Ольденбургу (от 24 октября 1676 года) в виде второй (длинной) анаграммы.

В современных терминах метод Ньютона состоит в следующем.

Пусть дано уравнение $\dot\nu = \xi(t,\nu)$. Будем искать выпрямляющий диффеоморфизм $\upsilon = \varphi(t,\nu)$, для которого $\upsilon = \nu$ при $t = 0$ (время не преобразовываем). Диффеоморфизм приводит уравнение $\dot\nu = \xi(t,\nu)$ к уравнению $\dot\upsilon = 0$, фазовые траектории которого — прямые. Из условия $\dot\upsilon = 0$ получаем уравнение $\frac{\partial\varphi}{\partial t} + \frac{\partial\varphi}{\partial\nu}\cdot\dot\nu \overset{(3.106)}{\equiv} 0$. Разложим ξ и φ в ряды Тейлора по степеням t: $\varphi = \varphi_0 + t\varphi_1 + t^2\varphi_2 + \ldots$, $\dot\nu = \xi = \xi_0 + t\xi_1 + t^2\xi_2 + \ldots$. Тогда $\varphi_0(\nu) \equiv \nu$, поэтому $\frac{\partial\varphi}{\partial\nu} = E + t\varphi_1' + \ldots$. В результате получаем

$$(\varphi_1 + 2t\varphi_2 + \ldots) + (E + t\varphi_1' + \ldots)\cdot(\xi_0 + t\xi_1 + t^2\xi_2 + \ldots) = 0.$$

На основании единственности коэффициентов ряда Тейлора приравняем нулю коэффициенты при одинаковых степенях t и последовательно получим: $\varphi_1+\xi_0 = 0, 2\varphi_2+\varphi_1'\xi_0+\xi_1 = 0,\ldots$. Мы можем последовательно («рекуррентно») найти сначала φ_1, потом φ_2 и так все члены искомого ряда. В этом состоит *метод Ньютона интегрирования дифференциальных уравнений с помощью рядов*. Все дальнейшее развитие анализа линейных пространств и их отображений даже и сегодня следует по намеченному Ньютоном пути.

Пример. Решить методом Ньютона уравнение $\dot x = x$ с начальным условием $x_0 = 1$ при $t = 0$. Решение. $x = x_0 + tx_1 + t^2x_2 + t^3x_3 + \ldots$. Тогда получаем $1 + tx_1 + t^2x_2 + \ldots$, следовательно, $x_1 = 1$, $x_2 = \frac{x_1}{2} = \frac{1}{2!}$,

$x_3 = \frac{x_2}{3} = \frac{1}{3!}, \ldots x_n = \frac{1}{n!}$. Так и был впервые выведен ряд для экспоненты. В результате решение уравнения: $x = x_0 e^t$.

Эйлер заметил, что ряды иногда расходятся, и обратил внимание на необходимость доказательства сходимости, чем много занимались в 19-м веке. Сходимость рядов в аналитическом случае была доказана Коши, Пикар перенес теорему Коши на случаи конечной гладкости.

Вместо t можно использовать любой малый параметр s, от которого гладко зависит решение уравнения: $\dot{\nu} = \xi(t, \nu, s)$. Для выпрямления этого векторного поля применим маленькую хитрость и будем рассматривать «расширенное уравнение», дополнив его уравнением $\dot{s} = 0$ в расширенном фазовом пространстве (s, ν).

Описанный метод постоянно используется во всех приложениях теории дифференциальных уравнений под названием *теории возмущений* или *метод малого параметра*. При этом не следует забывать, что хорошее приближение мы получим лишь в малой окрестности точки.

Мы уже знаем, что в окрестности неособой точки ($\xi \neq 0$) поле устроено просто: оно выпрямляется диффеоморфизмом. Чтобы рассмотреть устройство поля в малой окрестности особой точки τ, т.е. точки, где $\xi = 0$, применяет приём так называемой линеаризации.

Разложим функцию ξ в ряд Тейлора:

$$\xi(\nu) - \xi(\tau) = \xi'(\tau) \cdot (\nu - \tau) + o(\nu - \tau), \tag{5.6}$$

где производная векторного поля $\xi'(\tau)$ есть $(n \times n)$-матрица Якоби

$$\xi'(\tau) = \alpha = \left\| \frac{\partial a_i}{\partial x^j} \right\|^n. \tag{5.7}$$

и рассмотрим уравнение $\dot{\nu} = \xi$ с точностью до первого члена.

Пусть $\tau = (0, \ldots, 0)$, кроме того $\xi(\tau) = (0, \ldots, 0)$. Получаем:

$$\dot{\nu} = \alpha\nu. \tag{5.8}$$

Уравнение (5.8) называется *линейным*, отбрасывание остальных членов называется *линеаризацией уравнения* $\dot{\nu} = \xi$ в особой точке τ, поле $\alpha\nu$ называется линейной частью поля ξ в точке τ. Если система неавтономная, то линейный оператор зависит от времени: $\alpha = \alpha(t)$.

В результате приёма линеаризации мы всегда можем получить решение уравнения, так как линейные автономные уравнения — едва ли не единственный большой класс дифференциальных уравнений, для которых имеется полная теория, позволяющая полностью их решить.

Особая точка поля называется *невырожденной*, если оператор α линейной части поля в этой точке невырожденный. Собственные значения оператора линеаризации векторного поля в особой точке называют иногда собственными значениями особой точки.

При решении уравнения $\dot{x} = ax$, где a — число, (с начальным условием x_0 при $t = 0$) методом Ньютона получим: $x = x_0 e^{at}$. Так как $\frac{d}{dt} e^{\alpha t} = \alpha e^{\alpha t}$,

где α — матрица линейного оператора, то оказывается, что и в общем случае решение уравнения $\dot{\nu} = \alpha\nu$, с начальным условием $\nu(0) = \nu_0$ имеет вид:

$$\nu(t) = \nu_0 e^{\alpha t}. \tag{5.9}$$

Поэтому запишем следующую теорему.

Теорема о производящем операторе. Всякая однопараметрическая группа $\{\varrho^t : U \to U\}$ линейных преобразований имеет вид

$$\varrho^t = e^{\alpha t}, \tag{5.10}$$

где $\alpha : U \to U$ — некоторый линейный оператор.

Оператор α называют *производящим оператором* группы $\{\varrho^t\}$. Имеется взаимно однозначное соответствие между линейными дифференциальными уравнениями $\dot{\nu} = \alpha\nu$ и их фазовыми потоками $\{\varrho^t\}$; при этом фазовый поток состоит из линейных диффеоморфизмов.

Следствие 1 (формула Лиувилля). Фазовый поток $\{\varrho^t\}$ линейного уравнения $\dot{\nu} = \alpha\nu$ за время t меняет объем любой фигуры в e^{at} раз (см. рис. 5.1), где $a = \operatorname{Sp}\alpha$. Действительно,

$$\det \varrho^t = \det e^{\alpha t} \overset{(2.96)}{=} e^{\operatorname{Sp}\alpha t} = e^{t\operatorname{Sp}\alpha} = e^{at}.$$

Следствие 2. Если $\operatorname{Sp}\alpha = 0$ (след матрицы оператора α равен 0), то фазовый поток уравнения $\dot{\nu} = \alpha\nu$ сохраняет объёмы, т.е. ϱ^t сохраняет объём любой фигуры. Действительно, $\det \varrho^t = e^0 = 1$.

В частности, дивергенция линейного векторного поля $\xi(\nu) = \alpha\nu$, — это след оператора $\alpha \overset{(5.7)}{=} \left\| \frac{\partial a_i}{\partial x^j} \right\|^n$:

$$\operatorname{div}\xi \overset{(2.91)}{=} \operatorname{Sp}\alpha. \tag{5.11}$$

Дивергенция векторного поля определяет скорость растяжения объёмов соответствующим фазовым потоком. Если $\operatorname{div}\xi = 0$, то фазовый поток сохраняет объёмы любой области фазового пространства.

5.2 Общее решение линейного уравнения

При практическом решении дифференциальных уравнений оператор α задан своей матрицей в некотором базисе, и требуется явно вычислить матрицу оператора $e^{\alpha t}$ в том же базисе. Если все n собственных чисел оператора α вещественны и различны, то матрица α будет диагональная в собственном базисе. К этому базису и надо перейти, если матрица оператора α дана в другом базисе. Поэтому нужно

1. составить вековое или характеристическое уравнение
   $\det|\alpha - \lambda E| = 0$,

2. найти его корни $\lambda_1, \ldots, \lambda_n$ (мы предполагаем, что они вещественны и различны),

3. найти собственные векторы $\xi_1, \ldots, \xi_n$ из линейных уравнений $\alpha\xi_k = \lambda_k\xi_k$, $\xi_k \neq 0$,

4. разложить начальное условие по собственным векторам $\nu_0 = \sum_{k=1}^{n} a_k\xi_k$,

5. написать ответ

$$\nu(t) = a_1 e^{\lambda_1 t}\xi_1 + \ldots + a_n e^{\lambda_n t}\xi_n, \tag{5.12}$$

где $\xi_1, \ldots, \xi_n$ — не зависящие от начальных условий постоянные векторы, $a_1, \ldots, a_n$ — зависящие от начальных условий числа.

Векторы $\xi_1, \ldots, \xi_n$ называют *фундаментальной системой решений*. Если считать, что решения линейного уравнения образуют *пространство решений*, то векторы $\xi_1, \ldots, \xi_n$ образуют базис пространства решений, а решение $\nu(t)$ есть вектор этого пространства. Пространство решений уравнения линейно и изоморфно U.

Определение. Пусть $\alpha\colon \mathbb{C}^n \to \mathbb{C}^n$ — $\mathbb{C}$-линейный оператор. Системой n линейных однородных дифференциальных уравнений первого порядка с комплексными постоянными коэффициентами (линейным уравнением с фазовым пространством $\mathbb{C}^n$) называют уравнение

$$\dot{v} = \alpha v, \quad v \in \mathbb{C}^n. \tag{5.13}$$

Решением уравнения $\dot{v} = \alpha v$ с начальным условием $v(t_0) \in \mathbb{C}^n$, где $t_0 \in \mathbb{R}$, называется комплексная кривая $v\colon I \to \mathbb{C}^n$, если для всякого во-первых, $t \in I$, $\dot{v} = \alpha v(t)$ и во-вторых, $t_0 \in I$, где $I \subset \mathbb{R}$ — интервал вещественной оси.

Теорема. Решение уравнения $\dot{v} = \alpha v$ с начальным условием $v(0) = v_0$ даётся формулой

$$v(t) = v_0 e^{\alpha t}, \tag{5.14}$$

так как $\frac{d}{dt}e^{\alpha t} = \alpha e^{\alpha t}$.

Теорема. Всякая однопараметрическая группа $\{\varrho^t(t \in \mathbb{R})\}$ $\mathbb{C}$-линейных преобразований пространства $\mathbb{C}^n$ имеет вид

$$\varrho^t = e^{\alpha t}, \tag{5.15}$$

где $\alpha\colon \mathbb{C}^n \to \mathbb{C}^n$ — некоторый $\mathbb{C}$-линейный оператор.

Теорема. Если n корней $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ характеристического уравнения $\det|\alpha - \lambda E| = 0$ попарно различны, то $\mathbb{C}^n$ разлагается в прямое произведение $\mathbb{C}^n = \mathbb{C}^1_1 \times \ldots \times \mathbb{C}^1_n$ инвариантных относительно α и $e^{\alpha t}$ одномерных подпространств, причём в каждом одномерном подпространстве $\mathbb{C}^1_k$ величина $e^{\alpha t}$ сводится к умножению на комплексное число $e^{\lambda_k t}$. Тогда всякое решение уравнения (5.13) $\dot{v} = \alpha v$ имеет вид:

$$v(t) = c_1 e^{\lambda_1 t}\xi_1 + \ldots + c_n e^{\lambda_n t}\xi_n \tag{5.16}$$

где $\xi_1, \ldots, \xi_n$ — не зависящие от начальных условий постоянные векторы,
$c_1, \ldots, c_n$ — зависящие от начальных условий комплексные числа.

Каждый $\xi_k \in \mathbb{C}^n$ $(k = 1, \ldots, n)$ есть собственный вектор оператора α с собственным значением $\lambda_k \in \mathbb{C}$. Выберем собственные векторы с вещественными собственными значениями вещественными, а с комплексно сопряжёнными — комплексно сопряжёнными. Чтобы вектор $v(0) = c_1\xi_1 + \ldots + c_n\xi_n$ был вещественным, необходимо и достаточно, чтобы $c_1\xi_1 + \ldots + c_n\xi_n = \overline{c}_1\overline{\xi}_1 + \ldots + \overline{c}_n\overline{\xi}_n$. Для этого коэффициенты при комплексно сопряжённых векторах должны быть комплексно сопряжёнными, а при вещественных — вещественными.

Если $z_1, \ldots, z_n$ — линейная система координат в $\mathbb{C}^n$, а $\lambda_k = -\gamma_k + i\varpi_k$ $(k = 1, \ldots, n)$, то вещественная или мнимая часть каждой координаты $z_l = x_l + iy_l$ будет меняться со временем как

$$\begin{aligned} x_l &= \sum_{k=1}^{n} r_{kl} e^{-\gamma_k t} \cos(\varpi_k t + \vartheta_{kl}) = \\ &= \sum_{k=1}^{n} \left(a_{kl} e^{-\gamma_k t} \cos \varpi_k t + b_{kl} e^{-\gamma_k t} \sin \varpi_k t \right) \end{aligned} \tag{5.17}$$

где r, ϑ, a, b — зависящие от начальных условий вещественные числа.

Воспользуемся результатами исследования комплексного уравнения для изучения вещественного случая.

Пусть $\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$ — вещественный линейный оператор, задающий вещественное линейное уравнение

$$\dot{\nu} = \alpha\nu, \quad \nu \in \mathbb{R}^n. \tag{5.18}$$

Для его решения проведём комплексификацию вещественного пространства, увидим, что решения комплексифицированного уравнения являются решениями исходного вещественного уравнения, и запишем общее решение вещественного линейного уравнения.

Комплексификацией вещественного n-мерного линейного пространства $\mathbb{R}^n$ называется n-мерное комплексное пространство ${}^{\mathbb{C}}\mathbb{R}^n = \mathbb{C}^n$, которое строится следующим образом. Точки пространства ${}^{\mathbb{C}}\mathbb{R}^n$ — это пары векторов $v = (\nu, \tau); \nu, \tau \in \mathbb{R}^n$, обозначаемые $v = \nu + i\tau$. Операции сложения и умножения на комплексные числа определяются обычным способом. Поэтому два вещественных уравнения $\dot{\nu} = \alpha\nu$, $\dot{\tau} = \alpha\tau$ образуют одно уравнение $\dot{v} = \dot{\nu} + i\dot{\tau} = \alpha(\nu + i\tau) = \alpha v$. В результате комплексификация линейного оператора $\alpha\colon \mathbb{R}^n \to \mathbb{R}^m$ это оператор ${}^{\mathbb{C}}\alpha\colon {}^{\mathbb{C}}\mathbb{R}^n \to {}^{\mathbb{C}}\mathbb{R}^m$, действующий по следующему правилу: ${}^{\mathbb{C}}\alpha(\nu + i\tau) = \alpha\nu + i\alpha\tau$.

Время при комплексификации остаётся вещественным.

Комплексное сопряжение. Пусть $\alpha\colon {}^{\mathbb{C}}\mathbb{R}^n \to {}^{\mathbb{C}}\mathbb{R}^m$ — $\mathbb{C}$-линейный оператор. Комплексно сопряжённым к α оператором $\overline{\alpha}$ называется оператор $\alpha\colon {}^{\mathbb{C}}\mathbb{R}^n \to {}^{\mathbb{C}}\mathbb{R}^m$, определённый соотношением

$$\overline{\alpha v} = \overline{\alpha}\,\overline{v} \quad \text{для всякого } v \in {}^{\mathbb{C}}\mathbb{R}^n. \tag{5.19}$$

Оператор $\overline{\alpha}$ $\mathbb{C}$-линейный.

Комплексный линейный оператор $\alpha\colon {}^{\mathbb{C}}\mathbb{R}^n \to {}^{\mathbb{C}}\mathbb{R}^m$ является комплексификацией вещественного тогда и только тогда, когда

$$\overline{\alpha} = \alpha. \tag{5.20}$$

Комплексификация уравнения (5.18) — это уравнение (5.13) с комплексным фазовым пространством

$$\dot{v} = {}^{\mathbb{C}}\alpha v, \quad v \in \mathbb{C}^n = {}^{\mathbb{C}}\mathbb{R}^n. \tag{5.21}$$

Лемма 1. Решения уравнения (5.21) с комплексно сопряжёнными начальными условиями комплексно сопряжены.

Доказательство. Пусть $v(t)$ — решение с начальным условием $v(t_0) = v_0$, тогда $\overline{v}(t_0) = \overline{v}_0$. Покажем, что $\overline{v}$ — решение. Тогда лемма будет доказана (ввиду единственности).

При любом значении t имеем

$$\frac{d\overline{v}}{dt} \overset{(3.109)}{=} \overline{\frac{dv}{dt}} = \overline{{}^{\mathbb{C}}\alpha v} \overset{(5.19)}{=} \overline{{}^{\mathbb{C}}\alpha}\,\overline{v} \overset{(5.20)}{=} {}^{\mathbb{C}}\alpha\overline{v},$$

что и требовалось доказать.

Следствие. Решение уравнения (5.21) с вещественным начальным условием вещественно и удовлетворяет уравнению (5.18). Ибо если бы $\overline{v} \neq v$, то нарушалась бы теорема о единственности.

Лемма 2. Функция $v(t)$ тогда и только тогда является решением комплексифицированного уравнения (5.21), когда ее вещественная и мнимая части удовлетворяют исходному уравнению (5.18).

Действительно, ${}^{\mathbb{C}}\alpha(\nu + i\tau) = \alpha\nu + i\alpha\tau$, поэтому уравнение (5.21) распадается в прямое произведение

$$\dot{\nu} = \alpha\nu, \quad \dot{\tau} = \alpha\tau, \quad \nu, \tau \in \mathbb{R}^n.$$

Лемма 3. Если $\xi \in \mathbb{C}^n$ — собственный вектор оператора ${}^{\mathbb{C}}\alpha$ с собственным значением λ, то $\overline{\xi}$ — собственный вектор с собственным значением $\overline{\lambda}$. Кратности собственных чисел λ и $\overline{\lambda}$ совпадают.

Действительно, поскольку (5.20) $\overline{\alpha} = \alpha$, уравнение ${}^{\mathbb{C}}\alpha\xi = \lambda\xi$ эквивалентно ${}^{\mathbb{C}}\alpha\overline{\xi} = \overline{\lambda\xi}$ и характеристическое уравнение имеет вещественные коэффициенты.

Теорема. Пространство $\mathbb{R}^n$ распадается на r инвариантных относительно оператора α одномерных подпространств (где r — число вещественных собственных чисел λ) и w инвариантных относительно оператора α двумерных подпространств (где $w = \frac{n-r}{2}$ — число комплексно сопряжённых пар собственных чисел $\lambda, \overline{\lambda}$).

Действительно, вещественному собственному числу отвечает вещественный собственный вектор и, значит, одномерное инвариантное α подпространство в $\mathbb{R}^n$, а комплексно сопряжённой паре вещественных чисел отвечает по лемме 3 комплексно сопряжённая пара собственных векторов и, значит, двумерное инвариантное α подпространство в $\mathbb{R}^n$.

Каждое общее решение уравнения (5.18) записывается аналогично (5.12) и (5.16) единственным образом (при фиксированном выборе собственных векторов) в виде

$$\nu(t) = \sum_{k=1}^{r} a_k e^{\lambda_k t}\xi_k + \sum_{k=r+1}^{r+w} c_k e^{\lambda_k t}\xi_k + \overline{c}_k e^{\overline{\lambda}_k t}\overline{\xi}_k. \tag{5.22}$$

где a_k — вещественные числа,

c_k — комплексные числа.

Формулу (5.22) можно переписать в виде

$$\nu(t) = \sum_{k=1}^{r} a_k e^{\lambda_k t}\xi_k + 2\Re \sum_{k=r+1}^{r+w} c_k e^{\lambda_k t}\xi_k. \tag{5.23}$$

Рассмотрим случаи одномерного и двумерного подпространства.

Запишем линейное уравнение

$$\dot{z} = \lambda z \quad (z, \lambda \in \mathbb{C}, t \in \mathbb{R}), \tag{5.24}$$

фазовое пространство которого — комплексная прямая $\mathbb{C}$.

Мы уже знаем его решение

$$z(t) = z_0 e^{\lambda t}. \tag{5.25}$$

Во всех случаях кроме редко встречающегося и неинтересного случая $\lambda = 0$, *точка $z = 0$ является единственной неподвижной точкой фазового потока и единственной особой точкой соответствующего уравнению $\dot{z} = \lambda z$ векторного поля.*

Изучим поведение фазовых кривых вблизи этого единственного положения равновесия.

Система из двух обыкновенных дифференциальных уравнений эквивалентна канонической форме уравнения второго порядка (5.4), которое мы получаем, подставляя одно уравнение в другое:

$$\ddot{x} + 2\gamma\dot{x} + \omega^2 x = 0. \tag{5.26}$$

Мы уже знаем решение уравнения $x = x_0 e^{\lambda t}$. При подстановке этого решения мы получаем характеристическое (2.90) квадратное уравнение:

$$\lambda^2 + 2\gamma\lambda + \omega^2 = 0, \tag{5.27}$$

корни (собственные числа) которого вычисляются по формуле (1.40):

$$\lambda_{1,2} = -\gamma \pm \sqrt{\gamma^2 - \omega^2} = -\gamma \pm i\varpi, \tag{5.28}$$

а общее решение (5.12) записывается в виде линейной суперпозиции:

$$x = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}. \tag{5.29}$$

Если подставить $z = x + iy$ и $\lambda = -\gamma + i\varpi$ в уравнение (5.24), и приравнять вещественные и мнимые части, то мы получим систему из двух уравнений

$$\begin{array}{l} \dot{x} = -\gamma x - \varpi y, \\ \dot{y} = -\gamma y + \varpi x; \end{array} \quad \text{или} \quad \begin{Vmatrix} \dot{x} \\ \dot{y} \end{Vmatrix} = \begin{Vmatrix} -\gamma & -\varpi \\ \varpi & -\gamma \end{Vmatrix} \begin{Vmatrix} x \\ y \end{Vmatrix}; \tag{5.30}$$

эквивалентную исходному уравнению (5.24)

$$\dot{z} = \lambda z, \quad \text{где} \quad \lambda = \begin{Vmatrix} -\gamma & -\varpi \\ \varpi & -\gamma \end{Vmatrix}. \tag{5.31}$$

Условием линейной независимости уравнений является неравенство нулю определителя матрицы $\det \lambda = \omega^2 = \gamma^2 + \varpi^2 \neq 0$, что выполняется во всех случаях, кроме случая $\lambda = 0$.

Таким образом, преобразование фазового потока $\varrho^t = e^{\lambda t}$ есть преобразование плоскости (x, y) оператором $\lambda\colon \mathbb{R}^2 \to \mathbb{R}^2$ с поворотом на угол $\arg \lambda$ и растяжением в $|\lambda|$ раз (см. рис. 5.1).

Вид кривой (5.25) $z(t) = z_0 e^{\lambda t}$ зависит от значения $\lambda_{1,2} = -\gamma \pm i\varpi$:

1. Если $|\gamma| > |\omega|$, то $\lambda_{1,2}$ — вещественные числа.

   Если λ вещественное число, то функция $e^{\lambda t}$ вещественна. Тогда при $z = x + iy$ мы можем записать уравнение $\dot{z} = \lambda z$, как два одномерных вещественных уравнения $\dot{x} = \lambda x$, $\dot{y} = \lambda y$ с одномерными подпространствами решений. В этом случае фазовый поток уравнения $\dot{z} = \lambda z$ состоит из растяжения в $e^{\lambda t}$ раз (см. рис. 5.1).

   При $\gamma > 0$ фазовые траектории сходятся в начало координат (особая точка называется *устойчивый узел*), а при $\gamma < 0$ расходятся из начала координат (особая точка — *неустойчивый узел*) (см. рис. 5.2).

2. Если $\gamma = 0$, $\omega \neq 0$, то $\lambda_{1,2} = \pm i\omega$ — чисто мнимые.

   По формуле Эйлера (1.62) $e^{\lambda t} = e^{i\omega t} = \cos \omega t + i \sin \omega t$. В этом случае фазовый поток уравнения $\dot{z} = \lambda z$ — это семейство поворотов относительно $z = 0$ на угол ωt (см. рис. 5.1).

   В полярных координатах $z = re^{i\varphi}$ уравнение $\dot{z} = \dot{r}\frac{z}{r} + i\dot{\varphi}z = \lambda z = i\omega z$ эквивалентно системе двух вещественных уравнений:

   $$\dot{r} = 0, \quad \dot{\varphi} = \omega.$$

   Фазовые кривые — окружности, а их центр точка $z = 0$ называется особая точка типа *центр* (см. рис. 5.2):

   $$x^2 + y^2 = r^2 = const; \quad \text{или} \quad r = const, \quad \varphi = \omega t + \varphi_0.$$

   Изменение знака при ω меняет направление обхода окружности. При $+\omega$ окружность обходится против часовой стрелки, а при $-\omega$ — по часовой стрелке.

3. Если $|\gamma| < |\omega|$, то $\lambda_{1,2} = -\gamma \pm i\varpi$ ($\gamma, \varpi \in \mathbb{R}$) — комплексно сопряжённые числа.

   Действие фазового потока — поворот с растяжением (см. рис. 5.1).

   В полярных координатах $z = re^{i\varphi}$ уравнение $\dot{z} = \dot{r}\frac{z}{r} + i\dot{\varphi}z = \lambda z = (-\gamma + i\varpi)z$ эквивалентно системе двух вещественных уравнений:

$$\dot{r} = -\gamma r, \quad \dot{\varphi} = \varpi.$$

   Решение системы выглядит так:

$$r = r_0 e^{-\gamma t}, \quad \varphi = \varpi t + \varphi_0.$$

   При соответствующем выборе начала отсчёта углов фазовая кривая задаётся уравнением:

   $r = e^{-\frac{\gamma}{\varpi}\varphi}$ или $\varphi = -\frac{\varpi}{\gamma}\ln r$ при условии $\varphi_0 = -\frac{\varpi}{\gamma}\ln r_0$.

   Такая кривая называется *логарифмической спиралью*. Логарифмические спирали накручиваются на начало координат при $\gamma > 0$ (такая особая точка называется *устойчивый фокус*), либо раскручиваются от начала координат при $\gamma < 0$ (такая особая точки называется *неустойчивый фокус*) (см. рис. 5.2). Изменение знака при ϖ меняет направление спирали. При $+\varpi$ спираль накручивается или раскручивается против часовой стрелки, а при $-\varpi$ — по часовой стрелке.

4. Мы считали γ и ω действительными числами. Если теперь $\omega^2 < 0$, т.е. ω — мнимое число, то $\lambda_{1,2}$ — вещественные числа, причём $\lambda_2 < 0 < \lambda_1$. При $\gamma = 0$ фазовый поток представляет собой гиперболический поворот (см. рис. 5.1). Фазовые кривые — неподвижная точка $z = 0$, половины координатных осей и половины гипербол:

$$x^2 - y^2 = const.$$

   Особая точка $z = 0$ будет точкой типа *седло* (рис. 5.2).

   При $\gamma \neq 0$ особая точка по прежнему седло, а фазовый портрет получит дополнительное растяжение.

Таким образом, значения корней $\lambda_{1,2}$ характеристического уравнения задают классификацию особых точек на плоскости:

Условие на собственные числа	**Тип особой точки**
Действительные, $0 < \lambda_2 < \lambda_1$	Неустойчивый узел
Действительные, $\lambda_2 < 0 < \lambda_1$	Седло
Действительные, $\lambda_2 < \lambda_1 < 0$	Устойчивый узел
Чисто мнимые, $\lambda_1 = i\omega, \lambda_2 = -i\omega$	Центр
Комплексно сопряженные, $\Re(\lambda) > 0$	Неустойчивый фокус
Комплексно сопряженные, $\Re(\lambda) < 0$	Устойчивый фокус

Как показал еще А.Пуанкаре, поведение фазовых кривых (фазовый портрет) в окрестности положения равновесия на фазовой плоскости в системе общего положения (т.е. в большинстве систем) ограничивается перечисленными случаями.

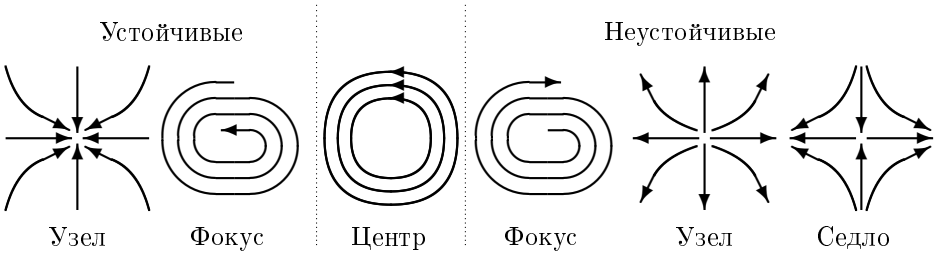


Рис. 5.2: Типичные фазовые портреты в окрестности точки равновесия

Как мы видим на рис. 5.2, два типичных фазовых портрета являются устойчивыми, а три — неустойчивыми, эти случаи разделены пограничным состоянием типа центр. Из представленных выше в таблице условий на собственные числа легко заметить, что положение равновесия асимптотически устойчиво, если собственные числа λ лежат в левой комплексной полуплоскости $\Re\lambda < 0$.

Из рисунка 5.2 видно, что фазовые кривые линейных уравнений без чисто мнимых собственных чисел ведут себя достаточно просто.

Линейные уравнения с чисто мнимыми собственными числами доставят нам примеры более сложного поведения фазовых кривых.

Рассмотрим как определяется устойчивость систем вблизи положения равновесия в общем случае. Запишем уравнение

$$\dot{\nu} = \xi(\nu), \quad \nu \in U \subset \mathbb{R}^n, \tag{5.32}$$

где ξ гладкое векторное поле. Выберем систему координат так, чтобы положение равновесия $\xi = 0$ было началом координат: $\nu = 0$.

Определение. Положение равновесия $\nu = 0$ уравнения (5.32) называется *устойчивым* (или *устойчивым по Ляпунову*), если система будет оставаться недалеко от положения равновесия. На математическом языке: для любого $\varepsilon > 0$ существует $\delta > 0$ такое, что если $\nu_0 < \delta$ для всякого начального положения ν_0, то решение $\nu(t)$ удовлетворяет условию $\nu(t) < \varepsilon$ в течение неограниченного времени $t \in [0, \infty)$.

Определение. Положение равновесия $\nu = 0$ уравнения (5.32) называется *асимптотически устойчивым*, если оно устойчиво и

$$\lim_{t\to\infty} \nu(t) = 0$$

для всякого решения $\nu(t)$ с начальным условием ν_0, лежащим в достаточно малой окрестности нуля.

Мы рассмотрели поведение системы (5.24) на фазовой плоскости (x, y). Однако поведение системы зависит не только от указанных переменных, но и от параметров γ и ω^2. Поэтому мы можем формальным образом ввести пространство параметров, и рассмотреть поведение системы в этом пространстве. Если мы формальным образом рассмотрим пространство параметров (ω^2, γ), где на оси абсцисс откладывается значение ω^2, а на оси ординат — γ (рис. 5.3), то устойчивые состояния лежат в области верхнего правого угла плоскости ($\omega^2 > 0, \gamma > 0$), а все остальные — неустойчивые. Их разделяет *граница устойчивости*, образованная положительными координатными полуосями. При переходе параметра через определенное значение происходит изменение поведения фазовых кривых в зависимости от того, какая особая точка находится в рассматриваемой области. Устойчивые состояния возможны при неотрицательных значениях этих параметров.

Область устойчивости делится линией $\omega^2 = \gamma^2$ на две части: устойчивых фокусов и устойчивых узлов. В нижнем правом углу лежат разделенные линией $\omega^2 = \gamma^2$ области неустойчивых фокусов и неустойчивых узлов. Точки типа фокуса лежат внутри параболы, а точки типа узла в правой полуплоскости вне параболы. Положительная полуось ω^2 образована точками, соответствующими состояниям равновесия типа центр. Область седел лежит левее оси ординат γ. Предельный случай (когда $\omega^2 = \gamma^2$) не представляет большого интереса, ибо этот случай, как и всякий, когда соотношение между параметрами системы точно фиксировано, не может быть точно реализован на практике и выполняет лишь роль границы между фокусами и узлами. Состояния равновесия типа центр, также не представляют большого интереса. Если условие параметра $\gamma = 0$ хоть немного нарушается, то система уходит в другое состояние равновесия. Поэтому состояние равновесия типа центр физически имеет значение лишь как граница между устойчивым и неустойчивым фокусом. Ось абсцисс при $\omega^2 = 0$ является границей между седлами и узлами и образует точки типа «седло-узел». Другие же пять типов состояний равновесия (фокусы, узлы и седла) сохраняют свой тип, т.е. обладают устойчивостью по отношению к малым изменениям параметров системы. Поэтому говорят, что эти пять типов состояний равновесия являются «грубыми».

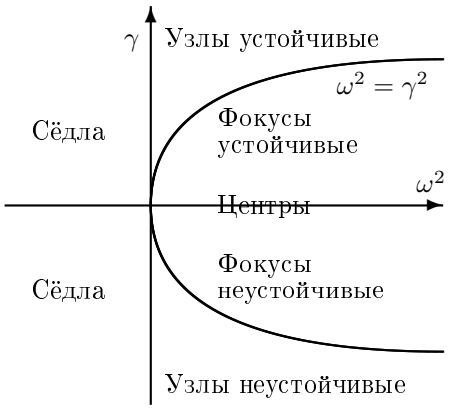


Рис. 5.3: Плоскость параметров

В соответствии с общей стратегией Пуанкаре подобный вид или, как говорят, подобную *особенность* имеют все системы общего положения, за-

висящие от двух параметров при подходящем подборе координат. Все более сложные случаи превращаются в указанные при общем малом изменении («шевелении») системы, т.е. не должны встречаться в природе, и их на первый взгляд можно не рассматривать. Эта точка зрения обесценивает большую часть теории дифференциальных уравнений и вообще математического анализа, где традиционно основное внимание уделяется малоценным, но трудным для исследования случаям необщего положения.

При замене координат $y = \omega^2 + \gamma; x = \omega^2 - \gamma$ мы можем записать типичную особенность границы области устойчивости положений равновесия в общих двухпараметрических семействах эволюционных систем (с фазовым пространством любой размерности) в виде $y = |x|$. Формула $y > |x|$ описывает область устойчивости при подходящем выборе координат на плоскости или в пространстве параметров, вообще говоря, криволинейных (рис. 5.4).

Область устойчивости при большем числе параметров (например, для трехпараметрических систем) также располагается углами наружу, вклиниваясь зияющими вершинами в область неустойчивости.

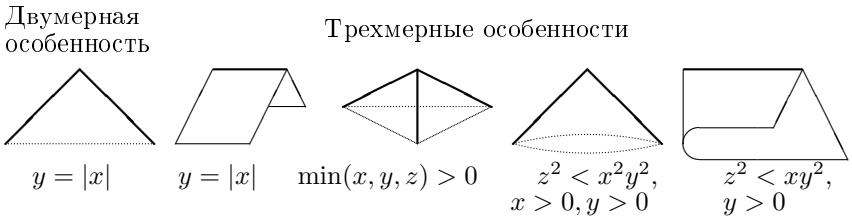


Рис. 5.4: Особенности и области устойчивости

Таким образом, для системы, принадлежащей особой части границы устойчивости, при малом изменении параметров более вероятно попадание в область неустойчивости, чем в область устойчивости. Это проявление общего принципа, согласно которому все хорошее (например, устойчивость) более хрупко, чем плохое.

По-видимому, все хорошие объекты удовлетворяют нескольким требованиям *одновременно*, плохим же считается объект, обладающий хотя бы одним из ряда недостатков.

При увеличении числа параметров число типов особенностей быстро растет, однако, как доказал Л.В. Левантовский [23] , оно остается конечным (с точностью до гладких замен параметров) при любом конечном числе параметров, сохраняется и принцип хрупкости.

5.3 Первые интегралы

Многие геометрические понятия можно описывать двумя способами: на языке точек пространства или же с помощью функций, заданных на нём.

Такая дуализация часто оказывается полезной в самых разных отделах математики. В частности, векторные поля можно описывать не только с помощью скоростей движений, но и как дифференцирование функций.

Определение. Функция f называется *первым интегралом* системы $\nu'_s = \xi(\nu)$, если её производная по направлению поля ξ равна нулю:

$$L_\xi f \equiv 0. \tag{5.33}$$

Наименование «первый интеграл» осталось с тех времён, когда пытались решить все дифференциальные уравнения путём интегрирования и называли интегралом то, что сейчас называют решением. Следующие два свойства первого интеграла очевидно эквивалентны соотношению (5.33) и могли бы быть приняты за его определение.

1. Функция постоянна $f = const$ вдоль каждой фазовой кривой.

2. Каждая фазовая кривая поля ξ принадлежит одной и только одной поверхности (множеству уровня) $f(x^1, \ldots, x^n) = const$.

Таким образом, при наличии первых интегралов фазовые кривые целиком лежат на их поверхности.

Пример. Рассмотрим систему, фазовым пространством которой является вся плоскость: $\dot\nu = \nu$ или $\dot x = x$, $\dot y = y$. Фазовые кривые представляют собой расходящиеся из начала координат лучи. Первый интеграл — непрерывная на всей плоскости функция, постоянная на каждом луче, т.е. просто константа. Никаких других, отличных от постоянной, первых интегралов система не имеет.

Непостоянные первые интегралы встречаются редко. Зато в тех случаях, когда они есть и когда их удаётся найти, награда бывает весьма значительной. Отсутствие непостоянных первых интегралов связано с топологическим устройством фазовых кривых. В общем случае фазовые кривые не укладываются в целом на поверхности (множестве уровня) никакой функции, поэтому непостоянного первого интеграла и нет. Однако локально, в окрестности неособой точки, фазовые кривые устроены просто и непостоянные первые интегралы существуют.

Например, мы выбрали в n-мерном фазовом пространстве U с заданным векторным полем ξ неособую точку ν_0 (и имеем единственную интегральную кривую, которая проходит через эту точку). Тогда существует такая окрестность V точки ν_0, что уравнение $\dot\nu = \xi(\nu)$ в V имеет $n-1$ функционально независимый первый интеграл $f_1, \ldots, f_{n-1}$, которые образуют полную систему первых интегралов.

Определение. Набор k функций *функционально независим* в окрестности точки ν_0, если ранг производной в точке ν_0 отображения $\varphi\colon U \to \mathbb{R}^k$, где $\varphi = (f_1, \ldots, f_k)$, заданного этими функциями, равен k. В свою очередь, это означает, что $\operatorname{grad} f_1, \ldots, \operatorname{grad} f_k$ линейно независимы.

Определение. *Полной системой* зависящих или не зависящих от времени *первых интегралов* системы дифференциальных уравнений, называется

такой набор первых интегралов, что любой другой первый интеграл является функцией от первых интегралов набора.

Пусть теперь $f(t,\nu)$ — гладкая функция на расширенном фазовом пространстве неавтономного системы уравнений $\dot\nu = \xi(t,\nu)$. Составим автономную систему, фазовые кривые которой будут интегральными кривыми исходного уравнения. Для этого расширим систему, добавив к исходным тривиальное уравнение $\dot t = 1$: $\chi = (t,\nu)$, $\dot\chi = \eta(\chi)$, $\eta(t,\nu) = (1,\nu)$. Если функция f является первым интегралом этого расширенной автономной системы уравнений, то она называется зависящим от времени первым интегралом системы $\dot\nu = \xi(t,\nu)$. При этом каждая интегральная кривая этой системы целиком лежит на одной из поверхностей $f(t,\nu) = const$. В некоторой окрестности точки (t_0,ν_0) будет полная система n первых интегралов.

Определение. Гладкая функция, постоянная на фазовых кривых автономной системы дифференциальных уравнений, называется (не зависящим от времени) *первым интегралом* этой системы уравнения. Зависящий от времени первый интеграл (вообще говоря, неавтономной системы уравнений) — это функция в расширенном фазовом пространстве, постоянная на интегральных кривых.

Уравнения с частными производными изучены гораздо хуже, чем с обыкновенными. Однако линейное однородное уравнение с частными производными (5.33) $L_\xi f = 0$, где ξ — известное векторное поле в области U, а f — неизвестная функция, удаётся свести к обыкновенному дифференциальному уравнению (5.3) $\dot\nu = \xi(\nu)$, называемому также *уравнением характеристик*. При этом фазовые кривые векторного поля ξ называются *характеристиками*, а первые интегралы — *решениями* уравнения $L_\xi f = 0$. Сущность связи между уравнением с частными производными и уравнением характеристик состоит в том, что движение сплошной среды можно описывать как с помощью обыкновенных дифференциальных уравнений её частиц, так и с помощью уравнений с частными производными для поля. Термин «характеристика» в данном случае означает «инвариантна относительно выбора системы координат».

Задачей Коши для уравнения $L_\xi f = 0$ называется задача об определении функции f по её значениям (начальной функции) на заданной гиперповерхности (начальной поверхности) в фазовом пространстве. Совокупность начальной функции заданной на начальной гиперповерхности называется начальным условием. Задача Коши не всегда имеет решение. Действительно, вдоль каждой характеристики f постоянно. Но характеристика может пересекать начальную поверхность несколько раз. Если значения начальной функции в этих точках различны, то соответствующая задача Коши не имеет решения ни в какой области, содержащей указанную характеристику.

Если характеристика пересекает начальную гиперповерхность *трансверсально* (под ненулевым углом, т.е. не является касательной), то точка пересечения называется *нехарактеристической*. У такой нехарактеристической точки всегда существует окрестность, в которой задача Коши имеет решение, и притом только одно.

5.4 Симплектическое многообразие

Как мы уже знаем с каждым пространством или гладким многообразием связано другое пространство или многообразие вдвое большей размерности, называемое касательным расслоением. Но если расслоению пространства мы уделили несколько слов, то расслоение многообразия следует рассмотреть подробнее.

Пусть V — n-мерное многообразие с выбранными локальными координатами $q_1, \ldots, q_n$(в подобных обозначениях традиционно используют нижние индексы). Дифференцирование многообразия в точке $\tau = (q_1, \ldots, q_n)$ даёт касательное пространство TV_τ, состоящее из касательных векторов $dq_1, \ldots, dq_n$. Тогда касательное расслоение TV есть $2n$-мерное дифференцируемое многообразие с локальными координатами $q_1, \ldots, q_n, dq_1, \ldots, dq_n$. Обозначим через $p_1, \ldots, p_n$ компоненты 1-формы на касательном пространстве TV_τ в точке τ. Тогда:

$$\omega^1 = p_1\, dq_1 \ldots + p_n\, dq_n = \mathbf{p}\, d\mathbf{q}. \tag{5.34}$$

Эта дифференциальная 1-форма называется *кокасательным вектором* к V в точке τ, аналогично тому, как алгебраическая 1-форма называется ковектором к векторному пространству.

Множество всех кокасательных к V в точке τ векторов образует n-мерное линейное пространство T^*V_τ с координатами $p_1, \ldots, p_n$, сопряжённое к касательному пространству TV_τ, которое называют *кокасательным пространством* к многообразию V в точке τ. Объединение кокасательных пространств к многообразию во всех его точках T^*V называется *кокасательным расслоением* V и имеет естественную структуру чётномерного дифференцируемого многообразия $M = T^*V$ с локальными координатами $p_1, \ldots, p_n, q_1, \ldots, q_n$ или $\mathbf{p}, \mathbf{q}$.

Вектор-функция, заданная в каждой точке многообразия, образует *векторное поле на многообразии* M, которое, как и касательное пространство на многообразии, не зависит от карты. Пусть область $U \subset M$ есть чётномерное фазовое пространство с координатами $\nu = (\mathbf{p}, \mathbf{q})$, на котором задано два векторных поля $\dot{\nu} = \xi(\nu)$ и $\nu'_s = \eta(\nu)$.

Тогда в точке $\tau \in M$: $\xi = (\dot{\mathbf{p}}, \dot{\mathbf{q}})$, $\eta = (\mathbf{p}'_s, \mathbf{q}'_s)$:

$$\omega^2(\xi, \eta)\, dt\, ds \overset{(2.149)}{=} (I\xi, \eta)\, dtds \overset{(2.143)}{=} d\mathbf{q}^\xi d\mathbf{p}^\eta - d\mathbf{p}^\xi d\mathbf{q}^\eta = d\mathbf{p} \wedge d\mathbf{q}, \tag{5.35}$$

где I — кососимметрическая матрица вида (2.150).

Таким образом, чётномерное дифференцируемое многообразие M имеет *точную дифференциальную* 2-*форму*, которая возникает при внешнем дифференцировании 1-формы (5.34):

$$\omega^2 = d\omega^1 = dp_1 \wedge dq_1 + \ldots + dp_n \wedge dq_n = d\mathbf{p} \wedge d\mathbf{q}. \tag{5.36}$$

Всякая точная форма замкнута, поэтому

$$d\omega^2 \overset{(5.36)}{=} d(d\omega^1) \overset{(3.62)}{=} 0. \tag{5.37}$$

При этом форма $\omega^2 = d\mathbf{p} \wedge d\mathbf{q}$ невырожденная, кососимметрическая и по определению является симплектической линейной структурой, аналогичной (2.143). Но в данном случае симплектическая структура задана не на линейном пространстве, а на многообразии.

Чётномерное многообразие M^{2n} с заданной на нём дифференциальной формой ω^2 называется *симплектическим многообразием*.

Пусть теперь $H(p_1, \ldots, p_n, q_1, \ldots, q_n)$ или $H(\mathbf{p}, \mathbf{q})$ — произвольная функция на симплектическом многообразии M. Тогда dH есть дифференциальная 1-форма на M: $dH(\eta) = \omega^1_\xi(\eta) = \omega^2(\xi, \eta)\, ds$, которой соответствует в каждой точке некоторый касательный к M вектор η:

$$dH(\eta) = \omega^2(\xi, \eta)\, ds \overset{(5.35)}{=} -\dot{\mathbf{p}}\, d\mathbf{q} + \dot{\mathbf{q}}\, d\mathbf{p} = \sum_{i=1}^{n} \dot{p}_i dq_i + \sum_{i=1}^{n} \dot{q}_i dp_i.$$

То же самое в другой записи: (5.38)

$$dH(\eta) = \omega^1_\xi(\eta) = \omega^2(\xi, \eta)\, ds \overset{(5.35)}{=} (I\xi, \eta) ds.$$

Откуда получаем векторное поле фазовых скоростей системы:

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}, \quad \dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}; \qquad \text{или} \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad \dot{q}_i = \frac{\partial H}{\partial p_i}.$$

То же самое в другой записи: (5.39)

$$\xi = (\dot{\mathbf{p}}, \dot{\mathbf{q}}) = I \operatorname{grad} H \qquad \text{или} \qquad \xi = (\dot{\mathbf{p}}, \dot{\mathbf{q}}) = (-H'_{\mathbf{q}}, H'_{\mathbf{p}}).$$

Обычно в математической литературе принято обозначение $\xi = I dH$, но dH скалярная функция, и произведение матрицы I на скаляр не совсем понятно, а $\operatorname{grad} H = (H'_{\mathbf{p}}, H'_{\mathbf{q}})$ — вектор:

$$\begin{Vmatrix} \dot{\mathbf{p}} \\ \dot{\mathbf{q}} \end{Vmatrix} = \begin{Vmatrix} \mathbf{0} & -E \\ E & \mathbf{0} \end{Vmatrix} \begin{Vmatrix} H'_{\mathbf{p}} \\ H'_{\mathbf{q}} \end{Vmatrix}.$$

Теорема Лиувилля. Фазовый поток $\varrho^t_H \colon \mathbb{R} \times U \to U \subset M$ векторного поля $\xi = (\dot{\mathbf{p}}, \dot{\mathbf{q}})$ сохраняет симплектическую структуру ω^2.

Доказательство. По теореме (5.10) группа однопараметрических преобразований, заданная векторным полем фазового пространства имеет вид $\varrho^t = e^{\alpha t}$. Согласно следствию 2 (5.11) этой теоремы, если $\operatorname{Sp} \alpha = 0$, то фазовый поток векторного поля сохраняет объём любой фигуры, так как $\det \varrho^t = e^{\operatorname{Sp} \alpha t} = e^0 = 1$. При этом $\operatorname{div} \xi = \operatorname{Sp} \alpha = 0$.

Так как:

$$\operatorname{Sp} I = 0 \quad \text{(все диагональные элементы матрицы } I \text{ — нулевые)};$$

$$\operatorname{div} \xi \overset{(3.83)}{=} \frac{\partial \dot{\mathbf{p}}}{\partial \mathbf{p}} + \frac{\partial \dot{\mathbf{q}}}{\partial \mathbf{q}} \overset{(5.39)}{=} \frac{\partial}{\partial \mathbf{p}} \left(-\frac{\partial H}{\partial \mathbf{q}} \right) + \frac{\partial}{\partial \mathbf{q}} \left(\frac{\partial H}{\partial \mathbf{p}} \right) \equiv 0,$$

то фазовый поток ϱ^t_H векторного поля ξ сохраняет объём любой фигуры.

Геометрически форма $\omega^2 = d\mathbf{p} \wedge d\mathbf{q}$ означает сумму площадей ориентированных проекций элементарного (бесконечно малого) параллелограмма на n координатных плоскостей $(p_1, q_1), \ldots, (p_n, q_n)$. Так как фазовый поток сохраняет объём любой фигуры, то он сохраняет сумму площадей элементарного параллелограмма, что и требовалось доказать.

Гомотопия (от греч. *ομος* — равный, одинаковый и *τοπος* — место) — формализация интуитивного представления о непрервной деформируемости одного отображения в другое. Фазовый поток ϱ^t представляет собой семейство диффеоморфизмов, непрерывно зависящих от параметра t. Это семейство называют *гомотопией, связывающей* при $0 \leqslant t \leqslant \tau$ диффеоморфизм ϱ^0 с диффеоморфизмом ϱ^τ *«непрерывным путём»*.

Пусть M — произвольное многообразие, C – k-мерная цепь в M. Тогда гомотопия ϱ^t при $0 \leqslant t \leqslant \tau$ даёт $(k+1)$-мерную цепь в M, которую называют *следом* цепи C и обозначают JC. Можно сказать, что JC — это цепь, которую цепь C *заметает* при гомотопии ϱ^t, $0 \leqslant t \leqslant \tau$: «торцы» цепи JC образованы начальным C и конечным $\varrho^\tau C$ положениями, а «боковая поверхность» $J(\partial C)$ заметена границей ∂C (рис. 5.5).

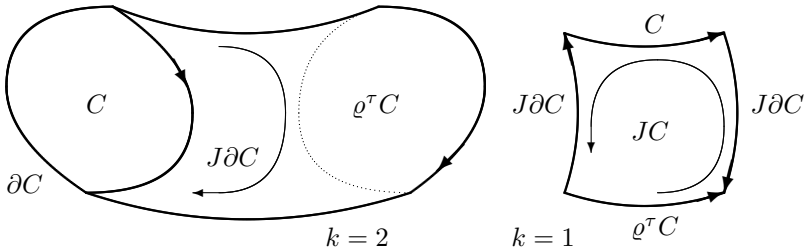


Рис. 5.5: След цепи при гомотопии

При $k = 1$ линия C заметает криволинейный четырёхугольник JC, границы линии C — две точки, которые заметают боковые границы четырёхугольника $J\partial C$.

При $k = 2$ круг C заметает «колбаску» JC, граница круга — окружность ∂C, которая заметает «боковую поверхность колбаски» $J\partial C$.

При правильном выборе ориентации граней (см. стр. 49) получаем для границы JC следующее выражение:

$$\partial(JC) = \varrho^\tau C - J(\partial C) - C. \tag{5.40}$$

Теорема Лиувилля (в интегральном виде). Задающая симплектическую структуру форма ω^2 *является интегральным инвариантом* (4.45) гамильтонова фазового потока

$$\int\limits_{\varrho_H^\tau C} \omega^2 = \int\limits_C \omega^2. \tag{5.41}$$

Один-форма dH также есть интегральный инвариант ϱ_H^t.

Доказательство. Рассмотрим 1-цепь из одного куска (см. стр. 34) $\gamma\colon [0,1] \to M$ или $\gamma(s)$ на симплектическом многообразии (M^{2n}, ω^2).

По теореме Фубини (4.11):

$$\int\limits_{J\gamma}\omega^2 \overset{(4.12)}{=} \int\limits_0^1\int\limits_0^\tau \omega^2(\xi,\eta)\,dt\,ds \overset{(5.38)}{=} \int\limits_0^\tau dt \int\limits_{\varrho_H^t\gamma} dH \quad \text{или} \quad \frac{\partial}{\partial t}\int\limits_{J\gamma}\omega^2 = \int\limits_{\varrho_H^t\gamma} dH.$$

Если цепь γ замкнута, то $\partial\gamma = 0$, тогда

$$\int\limits_{\varrho_H^t\gamma} dH \overset{(4.48)}{=} \int\limits_{\partial(\varrho_H^t\gamma)} H = 0 \quad \text{и, значит,} \quad \int\limits_{J\gamma}\omega^2 = 0. \tag{5.42}$$

Рассмотрим 2-цепь C, такую, что $\gamma = \partial C$, тогда:

$$0 \overset{(5.37)}{=} \int\limits_{JC} d\omega^2 \overset{(4.43)}{=} \oint\limits_{\partial(JC)} \omega^2 \overset{(5.40)}{=} \int\limits_{\varrho_H^\tau C}\omega^2 - \int\limits_C \omega^2 - \int\limits_{J(\partial C)}\omega^2 \overset{\gamma=\partial C}{\overset{(5.42)}{=}} \int\limits_{\varrho_H^\tau C}\omega^2 - \int\limits_C \omega^2,$$

что и требовалось доказать.

Каноническим преобразованием называется любое преобразование $\psi\colon M \to M$, например $(\mathbf{p},\mathbf{q}) \mapsto (\mathbf{P},\mathbf{Q})$, фазового пространства системы если оно сохраняет его симплектическую структуру ω^2 или, как говорят, имеет ω^2 *интегральным инвариантом.*

Тогда следствие теоремы Лиувилля можно сформулировать так:

Преобразование фазового пространства, осуществляемое фазовым потоком ϱ_H^t, каноническое. Это означает, что различные симплектические многообразия в каждой своей точке *изоморфны и отображаются друг в друга с сохранением «площадей».*

Таким образом, локально каждое симплектическое многообразие изоморфно стандартному симплектическому пространству.

В отличие от симплектических — римановы структуры $(\ ,\ )\colon TV_\tau \to \mathbb{R}$ или $\xi \mapsto (\xi,\xi)$, как правило, неизоморфны друг другу и не отображаются друг на друга с сохранением «площадей».

Для их различения Риман и ввел свою кривизну.

Следствие. $\omega^1 = \mathbf{p}\,d\mathbf{q}$ — относительный интегральный инвариант (4.46) преобразования ϱ_H^t, так как у него есть абсолютный интегральный инвариант $\omega^2 = d\omega^1 = d\mathbf{p}\wedge d\mathbf{q}$ (4.47).

Следствие. Как мы знаем (2.148) $\omega^{2n} = \underbrace{\omega^2\wedge\ldots\wedge\omega^2}_{n}$ и есть, с точностью до множителя, объём $2n$-мерного параллелепипеда в симплектическом пространстве $\mathbb{R}^{2n}$.

Тогда на симплектическом многообразии M с локальными координатами $p_1,\ldots,p_n,q_1,\ldots,q_n$ и дифференциальной симплектической структурой (5.36) $\omega^2 = dp_1\wedge dq_1 + \ldots + dp_n\wedge dq_n$ при различных внешних степенях дифференциальной формы ω^2 каждая из форм

$$\omega^{2k} = \sum_{i_1<\ldots<i_k} dp_{i_1}\wedge\ldots\wedge dp_{i_k}\wedge dq_{i_1}\wedge\ldots\wedge dq_{i_k}, \tag{5.43}$$

например, $\left(\omega^2\right)^2 = \omega^2 \wedge \omega^2$, $\left(\omega^2\right)^3 = \omega^2 \wedge \omega^2 \wedge \omega^2 \ldots \omega^{2n}$, является *абсолютным интегральным инвариантом* фазового потока.

Следовательно, при каноническом преобразовании сохраняется элемент фазового объёма на M, который равен

$$dV = D = \varrho_H^t D = dp_1 \wedge \ldots \wedge dp_n \wedge dq_1 \wedge \ldots \wedge dq_n, \tag{5.44}$$

т.е. *фазовый поток сохраняет объём фазового пространства.*

Следствие. Определитель любого симплектического преобразования $S\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ равен единице (например, $\det I = 1$).

Доказательство. Как мы уже знаем, канонические преобразования сохраняют внешние степени формы ω^2. Так как внешняя n-я степень есть (с точностью до постоянного множителя) элемент объёма в $\mathbb{R}^{2n}$. Значит, симплектические преобразования S стандартного $\mathbb{R}^{2n} = \{(\mathbf{p}, \mathbf{q})\}$ сохраняют элемент объёма, так что $\det S = 1$.

Например, в случае $n = 1$, $T^*M = \mathbb{R}^2$, $\omega^2 = dp \wedge dq$ — это элемент площади, который является интегральным инвариантом всякого преобразования ψ с якобианом 1.

Теорема Пуанкаре о возвращении (следствие теоремы Лиувилля). Пусть ϱ — сохраняющее объём преобразование, переводящее *ограниченную область* D евклидова пространства в себя: $\varrho D = D$. Тогда точка τ области D возвращается в свою окрестность U, т.е. $\varrho^n \tau \in U$ при некотором $n > 0$. Другими словами, почти всякая движущаяся точка многократно возвращается к своему исходному положению.

Доказательство. Рассмотрим образы окрестности U:

$$U, \varrho U, \varrho^2 U, \ldots, \varrho^n U, \ldots.$$

Все они имеют одинаковый положительный объём. Если бы они не пересекались, объём D был бы бесконечен. Значит при некоторых $k > l \geqslant 0$: $\varrho^k U \cap \varrho^l U \neq 0$. Следовательно, $\varrho^{k-l} U \cap U \neq 0$. Тогда $\varrho^{k-l} \tau \in U$, что и требовалось доказать.

Несколько парадоксальным выводом из теорем Пуанкаре и Лиувилля является следующее предсказание: если открыть перегородку, отделяющую камеру с газом от камеры с вакуумом, то через некоторое время молекулы газа почти наверное снова соберутся в первой камере. Разгадка парадокса в том, что «некоторое время» больше времени существования Солнечной системы.

С другой стороны теорема о возвращении не требует, чтобы молекулы газа возвращались одновременно, а периодическое возвращение каждой молекулы парадоксом не является.

Определение. Векторное поле $\xi = I \operatorname{grad} H$ называется *гамильтоновым векторным полем*, функция H — *функцией Гамильтона.* Динамическая система, заданная на симплектическом многообразии, называется *гамильтоновой системой.* Уравнения (5.39) называются *каноническими уравнениями Гамильтона*, а пары переменных p_i, q_i — *канонически сопряжёнными парами координат.* Число степеней свободы гамильтоновой системы из $2n$

уравнений по-прежнему считается n, как у исходного n-мерного многообразия V с выбранными локальными координатами $q_1, \ldots, q_n$. Фазовый поток $\varrho_H^t \colon \mathbb{R} \times U \to U \subset M$ называется *гамильтоновым фазовым потоком с функцией Гамильтона* H.

Обратим внимание ещё раз. Координаты $\mathbf{p}, \mathbf{q}$ фазового пространства $U \subset M$ есть локальные координаты на M, и при попытке продолжить функцию H на всё многообразие M, мы можем получить «многозначную функцию Гамильтона». Ибо замкнутая 1-форма $\omega^1_\xi(\eta) = \omega^2(\eta, \xi)$ на неодносвязном многообразии (например, на торе T^2) может не быть точной, т.е. дифференциалом $dH = \omega^1$, но локально она точна. Поэтому, всякая однопараметрическая группа диффеоморфизмов, сохраняющих симплектическую структуру ω^2, задаёт локально гамильтоново векторное поле и является локально гамильтоновым фазовым потоком, но необязательно гамильтоново в целом. Например, в $\mathbb{R}^{2n}$ всякая замкнутая 1-форма является дифференциалом функции и поэтому всякая однопараметрическая группа канонических (сохраняющих $d\mathbf{p} \wedge d\mathbf{q}$) диффеоморфизмов всегда является гамильтоновым фазовым потоком. Но, например, на компактном многообразии без края симплектическая форма не может быть точной.

Теорема. Функция H является первым интегралом ($H = const$) фазового потока ϱ_H^t (сформулирован общий аналог закона сохранения энергии, так как в механике H оказывается полной энергией механической системы, поэтому первые интегралы гамильтоновой системы исторически называют также *интегралы движения*).

Доказательство. Производная H по направлению вектора поля $\xi = I\operatorname{grad} H$ (вдоль фазовых кривых) равна значению dH на векторе ξ, тогда получаем $dH(\xi) = \omega^2(\xi, \xi)\, dt$ или

$$\dot H = \omega^2(\xi,\xi) \overset{(5.38)}{=} \sum_{i=1}^n \left(\dot q_i \frac{\partial H}{\partial q_i} + \dot p_i \frac{\partial H}{\partial p_i} \right) \overset{(5.39)}{=} \sum_{i=1}^n \left(\frac{\partial H}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial H}{\partial q_i} \right) = 0,$$

т.е. $\dot H = 0$ или $H = const$ на многообразии M.

Обратная теорема. Первый интеграл на симплектическом многообразии является функцией Гамильтона.

Доказательство

$$-\left(\frac{\partial H}{\partial p_i}\right)_{q_i}\left(\frac{\partial q_i}{\partial H}\right)_{p_i} \overset{(3.161)}{=} \left(\frac{\partial q_i}{\partial p_i}\right)_H \overset{(3.164)}{=} \frac{\partial(q_i, H)}{\partial(p_i, H)} =$$
$$\overset{(3.165)}{=} \frac{\partial(q_i, H)}{\partial(t, H)}\frac{\partial(t, H)}{\partial(p_i, H)} \overset{(3.164)}{=} \left(\frac{\partial q_i}{\partial t}\right)_H \left(\frac{\partial t}{\partial p_i}\right)_H \overset{(H=const)}{=} \frac{\dot q_i}{\dot p_i}.$$

Таким образом, если $H = const$, то для H справедливы уравнения (5.39), что и требовалось доказать.

Введём новую *функцию Лагранжа* или *лагранжиан* $L(\mathbf{q}, \dot{\mathbf{q}})$:

$$L(\mathbf{q}, \dot{\mathbf{q}}) \overset{(3.152)}{=} \mathbf{p} \cdot \dot{\mathbf{q}} - H(\mathbf{p}, \mathbf{q}). \tag{5.45}$$

Преобразование Лежандра $L(\mathbf{q},\dot{\mathbf{q}}) = \mathbf{p}\cdot\dot{\mathbf{q}} - H(\mathbf{p},\mathbf{q})$ не зависит от системы координат: оно сопоставляет функции $L\colon TV \to \mathbb{R}$ на касательном расслоении функцию $H\colon T^*V \to \mathbb{R}$ на кокасательном.

Тогда

$$dL = \frac{\partial L}{\partial \mathbf{q}}\,d\mathbf{q} + \frac{\partial L}{\partial \dot{\mathbf{q}}}\,d\dot{\mathbf{q}} = \mathbf{p}\,d\dot{\mathbf{q}} + \dot{\mathbf{q}}\,d\mathbf{p} - dH \overset{(5.38)}{=} \mathbf{p}\,d\dot{\mathbf{q}} + \dot{\mathbf{p}}\,d\mathbf{q}. \tag{5.46}$$

В результате можно записать

$$\begin{aligned} &\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}}, \quad \dot{\mathbf{p}} = \frac{\partial L}{\partial \mathbf{q}} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right), \quad \text{или} \\ &p_i = \frac{\partial L}{\partial \dot{q}_i}, \quad \frac{\partial L}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = 0 \qquad (i = 1,\ldots,n). \end{aligned} \tag{5.47}$$

Уравнения (5.47) называются *уравнениями Эйлера–Лагранжа.*

Теорема Эммы Нётер. Всякой однопараметрической группе диффеоморфизмов ϱ^s многообразия лагранжевой системы, сохраняющих функцию Лагранжа ($\frac{dL}{ds} = 0$), соответствует первый интеграл уравнений движения. Пусть V — гладкое многообразие, $\mathbf{q} = \mathbf{q}(t,s)$ — локальные координаты, $L\colon TV \to \mathbb{R}$ — гладкая числовая функция на касательном расслоении, $\varrho^s\colon V \to V$ — диффеоморфизм. Тогда соответствующая L система уравнений Лагранжа имеет первый интеграл $\mathcal{I}\colon TV \to \mathbb{R}$ вида:

$$\mathcal{I}(\mathbf{q},\dot{\mathbf{q}}) = \frac{\partial L}{\partial \dot{\mathbf{q}}}\frac{d\varrho^s}{ds}\bigg|_{s=0} = \frac{\partial L}{\partial \dot{\mathbf{q}}}\,\mathbf{q}'_s \overset{(5.47)}{=} \mathbf{p}\cdot\mathbf{q}'_s. \tag{5.48}$$

Доказательство. По условию $\frac{dL}{ds} = 0$. С учётом $\frac{d\dot{\mathbf{q}}}{ds} \overset{(3.87)}{=} \frac{d\mathbf{q}'_s}{dt}$ получаем:

$$\begin{aligned} 0 = \frac{dL(\mathbf{q},\dot{\mathbf{q}})}{ds} &\overset{(3.101)}{=} \frac{\partial L}{\partial \mathbf{q}}\frac{d\mathbf{q}}{ds} + \frac{\partial L}{\partial \dot{\mathbf{q}}}\frac{d\dot{\mathbf{q}}}{ds} = \\ &\overset{(5.47)}{=} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right)\mathbf{q}'_s + \frac{\partial L}{\partial \dot{\mathbf{q}}}\frac{d\mathbf{q}'_s}{dt} \overset{(3.101)}{=} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\mathbf{q}'_s\right) = \frac{d\mathcal{I}}{dt}. \end{aligned}$$

Значит $\frac{d\mathcal{I}}{dt} = 0$ или $\mathcal{I} = const$, что и требовалось доказать.

Теорема Нётер установлена в работах учёных гёттингенской школы Д. Гильберта, Ф. Клейна и Э. Нётер. В наиболее распространенной формулировке теорема была доказана Эмми Нётер в 1918 году.

Вычислим полную производную по времени некоторой функции $F(t,\mathbf{p},\mathbf{q})$ на симплектическом многообразии:

$$\begin{aligned} \dot{F} &\overset{(3.104)}{=} \sum_{i=1}^{n}\left(\dot{q}_i\frac{\partial F}{\partial q_i} + \dot{p}_i\frac{\partial F}{\partial p_i}\right) + \frac{\partial F}{\partial t} \overset{(3.122)}{=} L_\xi F + \frac{\partial F}{\partial t} = \\ &\overset{(5.39)}{=} \sum_{i=1}^{n}\left(\frac{\partial H}{\partial p_i}\frac{\partial F}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial F}{\partial p_i}\right) + \frac{\partial F}{\partial t} = \{H,F\} + \frac{\partial F}{\partial t}, \end{aligned} \tag{5.49}$$

где выражение

$$\{H, F\} = \omega^2(\operatorname{grad} F, \operatorname{grad} H) \tag{5.50}$$

называется скобка Пуассона функций F и H.

Скобкой Пуассона функций F и H, заданных на симплектическом многообразии называется производная функции F по направлению гамильтонова фазового потока с функцией Гамильтона.

Кроме того, скобка Пуассона функций F и H равна «кососкалярному произведению» векторов скоростей фазовых потоков с функциями Гамильтона F и H. Также, скобка Пуассона функций F и H является кососимметричной билинейной функцией от F и H:

Скобка Пуассона введена С. Пуассоном (1809).

Если $F = const$, т.е. является первым интегралом гамильтоновой системы (фазового потока с функцией Гамильтона H), тогда необходимо и достаточно чтобы её скобка Пуассона с H была равна нулю:

$$\{H, F\} = 0.$$

Теорема Пуассона. Если функции F и G — первые интегралы, то их скобка Пуассона также есть первый интеграл:

$$\text{Если} \quad \{F, G\} = 0 \quad \text{тогда, очевидно} \quad \frac{d}{dt}\{F, G\} = 0. \tag{5.51}$$

Если имеет место выражение $\{F, G\} = 0$, то говорят, что функции F и G находятся в *инволюции*.

Первые интегралы гамильтоновой системы по традиции обозначают прописными буквами H, F и т.д.

Свойства скобок Пуассона:

1. $\{F, G\} = -\{G, F\}$ — антикоммутативность;
2. $\{F, c\} = 0$, где $c \in \mathbb{R}$;
3. $\{H, aF + bG\} = a\{H, F\} + b\{H, G\}$, где $a, b \in \mathbb{R}$ — билинейность;
4. $\{H, F \cdot G\} = \{H, F\} \cdot G + \{H, G\} \cdot F$ — тождество Лейбница;
5. $\{H, \{F, G\}\} + \{F, \{G, H\}\} + \{G, \{H, F\}\} = 0$ — тождество Якоби;
6. $\frac{\partial}{\partial t}\{F, G\} = \left\{\frac{\partial F}{\partial t}, G\right\} + \left\{F, \frac{\partial G}{\partial t}\right\}$;
7. $\{F, q_i\} = \frac{\partial F}{\partial p_i}$, $\{p_i, F\} = \frac{\partial F}{\partial q_i}$, где $i = 1, \ldots, n$;
8. $\{q_i, q_j\} = 0$, $\{p_i, p_j\} = 0$, $\{p_i, q_j\} = -\{q_j, p_i\} = \delta_{ij}$ полученные выражения называют также *фундаментальными скобками Лагранжа*;
9. Пусть ξ и η — гамильтоновы поля с функциями Гамильтона F и G, тогда $\mu = \{\xi, \eta\}$ — также гамильтоново поле с функцией Гамильтона $H = \{F, G\}$.

10. Гамильтоновы векторные поля на симплектическом многообразии образуют подалгебру алгебры Ли всех полей, а первые интегралы гамильтонова фазового потока образуют подалгебру алгебры Ли всех функций.

11. Преобразование $(\mathbf{p},\mathbf{q}) \mapsto (\mathbf{P},\mathbf{Q})$ каноническое тогда и только тогда, когда для любых F и G справедливо $\{F,G\}_{\mathbf{p},\mathbf{q}} = \{F,G\}_{\mathbf{P},\mathbf{Q}}$. При этом симплектические структуры совпадают $d\mathbf{p}\wedge d\mathbf{q} = d\mathbf{P}\wedge d\mathbf{Q}$.

5.5 Производящая функция

Если в качестве нечётномерного многообразия (расширенного фазового пространства неавтономной гамильтоновой системы) мы рассмотрим расширенное фазовое пространство с координатами $t,\mathbf{p},\mathbf{q}$, то первым интегралом будет функция $H(t,\mathbf{p},\mathbf{q})$, а преобразование Лежандра

$$L(t,\mathbf{q},\dot{\mathbf{q}}) \overset{(3.152)}{=} \mathbf{p}\cdot\dot{\mathbf{q}} - H(t,\mathbf{p},\mathbf{q}) \tag{5.52}$$

задаёт функцию $L(t,\mathbf{q},\dot{\mathbf{q}})$, причём, очевидно, что

$$\frac{\partial L}{\partial t} = -\frac{\partial H}{\partial t}. \tag{5.53}$$

Теорема. Линии ротора (см. стр. 125) формы $\omega^1 = \mathbf{p}d\mathbf{q} - Hdt$ в $(2n+1)$-мерном расширенном фазовом пространстве $(t,\mathbf{p},\mathbf{q})$ однозначно проектируются на ось t, т.е. задаются функциями $\mathbf{p}=\mathbf{p}(t)$, $\mathbf{q}=\mathbf{q}(t)$. Эти функции удовлетворяют системе канонических дифференциальных уравнений (5.39) с функцией Гамильтона $H(t,\mathbf{p},\mathbf{q})$: $\dot{\mathbf{p}} = -H'_{\mathbf{q}}$, $\dot{\mathbf{q}} = H'_{\mathbf{p}}$. Иными словами, линии ротора формы $\omega^1 = \mathbf{p}d\mathbf{q} - Hdt$ суть траектории фазового потока в расширенном фазовом пространстве, т.е. интегральные кривые канонических уравнений Гамильтона.

Доказательство. Дифференциал формы $\omega^1 = \mathbf{p}d\mathbf{q} - Hdt$ равен:

$$d\omega^1 \overset{(3.59)}{=} d\mathbf{p}\wedge d\mathbf{q} - H'_{\mathbf{p}}\,d\mathbf{p}\wedge dt - H'_{\mathbf{q}}\,d\mathbf{q}\wedge dt.$$

или

$$d\omega^1 = \sum_{i=1}^{n}\left(dp_i\wedge dq_i - \frac{\partial H}{\partial p_i}\,dp_i\wedge dt - \frac{\partial H}{\partial q_i}\,dq_i\wedge dt\right).$$

Из этого выражения видно, что матрица α порядка $2n+1$ 2-формы $\omega^2 = d\omega^1 = (\alpha\xi,\eta)$ в координатах $t,\mathbf{p},\mathbf{q}$ имеет вид:

$$\alpha = \begin{Vmatrix} \mathbf{0} & -E & H'_{\mathbf{p}} \\ E & \mathbf{0} & H'_{\mathbf{q}} \\ -H'_{\mathbf{p}} & -H'_{\mathbf{q}} & 0 \end{Vmatrix}. \tag{5.54}$$

Ранг этой матрицы равен $2n$ (левый верхний $2n$-угол невырожден). Поэтому 2-форма неособая. Непосредственно проверяется, что вектор $(-H'_{\mathbf{q}}, H'_{\mathbf{p}}, 1)$ — собственный вектор матрицы α с собственным значением нуль (умножение матрицы на вектор дает 0). Значит, он задаёт направление линий ротора формы ω^1, что и требовалось доказать.

Теорема. Пусть две замкнутые кривые C_1, C_2 охватывают одну и ту же трубку фазовых траекторий поля $\dot{\mathbf{p}} = -H'_{\mathbf{q}}$, $\dot{\mathbf{q}} = H'_{\mathbf{p}}$. Тогда интегралы формы $\omega^1 = \mathbf{p}\,d\mathbf{q} - H dt$ по ним одинаковы (рис. 5.6):

$$\oint\limits_{C_1} \{\mathbf{p}\,d\mathbf{q} - H\,dt\} = \oint\limits_{C_2} \{\mathbf{p}\,d\mathbf{q} - H\,dt\}. \tag{5.55}$$

Доказательство теоремы непосредственно следует из многомерной леммы Стокса (4.54).

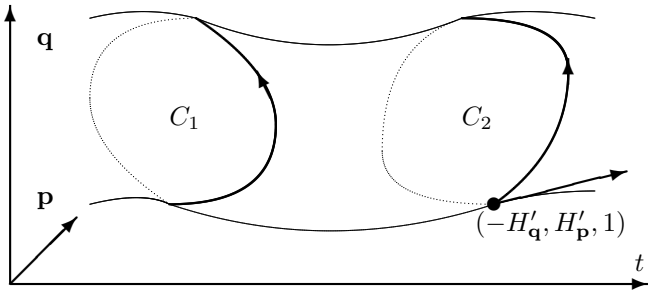


Рис. 5.6: Гамильтоново поле и линии ротора формы $\mathbf{p}\,d\mathbf{q} - H dt$

Следствие 1. Форма $\omega^1 = \mathbf{p}\,d\mathbf{q} - H dt$ — есть относительный интегральный инвариант (4.46):

$$\oint\limits_{C} \{\mathbf{p}\,d\mathbf{q} - H\,dt\} = \oint\limits_{\varrho^\tau_H C} \{\mathbf{p}\,d\mathbf{q} - H\,dt\}. \tag{5.56}$$

Рассмотрим, в частности, кривые, составленные из одновременных состояний, т.е. лежащие в плоскостях $t = const$. Вдоль таких кривых $dt = 0$. Тогда получаются три важных следствия.

Следствие 2. Фазовый поток сохраняет при гомотопии интеграл формы $\omega = \mathbf{p}\,d\mathbf{q}$ по замкнутым кривым в пространстве $\mathbf{p}, \mathbf{q}$:

$$\oint\limits_{C} \mathbf{p}\,d\mathbf{q} = \oint\limits_{\varrho^\tau_H C} \mathbf{p}\,d\mathbf{q}. \tag{5.57}$$

Форма $\omega = \mathbf{p}\,d\mathbf{q}$ называется *относительным интегральным инвариантом Пуанкаре*.

Следствие 3. Из теоремы Стокса непосредственно следует:

$$\oint\limits_{C} \mathbf{p}\,d\mathbf{q} \overset{(4.43)}{=} \int\limits_{S} d\mathbf{p} \wedge d\mathbf{q}, \tag{5.58}$$

где $C = \partial S$,

S — двумерная ориентированная цепь,

$d\omega = \omega^2 = d\mathbf{p} \wedge d\mathbf{q}$ — абсолютный интегральный инвариант.

Следствие 4. Фазовый поток сохраняет при гомотопии сумму площадей ориентированных проекций поверхности на n координатных плоскостей p_i, q_i (симплектическую структуру):

$$\int\limits_S d\mathbf{p} \wedge d\mathbf{q} = \int\limits_{\varrho_H^\tau S} d\mathbf{p} \wedge d\mathbf{q}, \tag{5.59}$$

Мы ещё раз доказали теорему Лиувилля (5.41).

Таким образом, определение канонического преобразования $\varrho_H^\tau \colon M \to M$ может быть записано тремя эквивалентными способами:

1. Если оно имеет ω^2 интегральным инвариантом, т.е. ϱ_H^τ сохраняет симплектическую структуру.
2. Если $\int_{\varrho_H^\tau C} \omega^2 = \int_C \omega^2$, т.е. ϱ_H^τ сохраняет сумму площадей проекций любой поверхности.
3. Если $\oint_C \mathbf{p}\, d\mathbf{q} = \oint_{\varrho_H^\tau C} \mathbf{p}\, d\mathbf{q}$, т.е. форма $\mathbf{p}\, d\mathbf{q}$ относительный интегральный инвариант ϱ_H^τ.

Замечание. Пусть $\varrho \colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ — каноническое преобразование фазового пространства, переводящее точку с координатами $(\mathbf{p}, \mathbf{q})$ в точку с координатами $(\mathbf{P}, \mathbf{Q})$. Для любой замкнутой кривой C имеем

$$\oint\limits_C (\mathbf{p}\, d\mathbf{q} - \mathbf{P}\, d\mathbf{Q}) = \oint\limits_C \mathbf{p}\, d\mathbf{q} - \oint\limits_C \mathbf{P}\, d\mathbf{Q} = 0 \tag{5.60}$$

ввиду каноничности ϱ_H^t. Тогда 1-форма $\mathbf{p}\, d\mathbf{q} - \mathbf{P}\, d\mathbf{Q}$ есть полный дифференциал функции S, и интеграл $\int_{\mathbf{p}_0\mathbf{q}_0}^{\mathbf{p}_1,\mathbf{q}_1} (\mathbf{p}\, d\mathbf{q} - \mathbf{P}\, d\mathbf{Q})$ не зависит от пути интегрирования, а лишь от начальной и конечной точек.

Замкнутая форма $\omega^1 = \mathbf{p}d\mathbf{q} - H dt$ называется *интегральным инвариантом Пуанкаре-Картана*. С этой формой инвариантным (не зависящим от карт) образом связано семейство линий на нечётномерном многообразии — линий ротора. На карте $t, \mathbf{p}, \mathbf{q}$ эти линии изображаются траекториями фазового потока $\dot{\mathbf{p}} = -H'_{\mathbf{q}}$, $\dot{\mathbf{q}} = H'_{\mathbf{p}}$ с функцией Гамильтона $H(t, \mathbf{p}, \mathbf{q})$. Умножив преобразование Лежандра (5.45) на dt получим $L dt = \mathbf{p} d\mathbf{q} - H dt$. Тогда можно выбрать такую функцию $S(t, \mathbf{q})$, что

$$dS = L\, dt = \mathbf{p}\, d\mathbf{q} - H\, dt \tag{5.61}$$

Точная дифференциальная форма dS является полным дифференциалом функции S.

Доказательство

$$\frac{\partial S}{\partial \mathbf{q} \partial t} = \frac{\partial \mathbf{p}}{\partial t} \overset{(5.39)}{=} -\frac{\partial H}{\partial \mathbf{q}} = \frac{\partial S}{\partial t \partial \mathbf{q}} \quad \text{значит} \quad \frac{\partial S}{\partial \mathbf{q} \partial t} = \frac{\partial S}{\partial t \partial \mathbf{q}},$$

откуда согласно (3.87) следует, что dS — полный дифференциал функции S, что и требовалось доказать.

В результате запишем

$$L \overset{(5.61)}{=} \dot{S}, \tag{5.62}$$

$$\mathbf{p} \overset{(3.102)}{=} \frac{\partial S}{\partial \mathbf{q}}, \tag{5.63}$$

$$H \overset{(3.102)}{=} -\frac{\partial S}{\partial t}. \tag{5.64}$$

Так как функция H есть первый интеграл ($H = const, \dot{H} = 0$), то для функции S справедливо так называемое *волновое уравнение*

$$\nabla^2 S = \frac{1}{\dot{\mathbf{q}}^2} \frac{\partial^2 S}{\partial t^2}. \tag{5.65}$$

Доказательство

$$0 = \dot{H} \overset{(3.102)}{=} \frac{\partial H}{\partial \mathbf{q}} \dot{\mathbf{q}} + \frac{\partial H}{\partial \mathbf{p}} \dot{\mathbf{p}} + \frac{\partial H}{\partial t} \overset{\substack{(5.39)\\(5.47)}}{=} \frac{\partial H}{\partial \mathbf{q}} \dot{\mathbf{q}} + \frac{\partial L}{\partial \mathbf{q}} \dot{\mathbf{q}} + \frac{\partial H}{\partial t} =$$

$$= \frac{\partial (H + L)}{\partial \mathbf{q}} \dot{\mathbf{q}} + \frac{\partial H}{\partial t} \overset{(5.52)}{=} \frac{\partial \mathbf{p}}{\partial \mathbf{q}} \dot{\mathbf{q}}^2 + \frac{\partial H}{\partial t} \overset{\substack{(5.63)\\(5.64)}}{=} \frac{\partial^2 S}{\partial \mathbf{q}^2} \dot{\mathbf{q}}^2 - \frac{\partial^2 S}{\partial t^2}.$$

Пусть теперь $T, \mathbf{P}, \mathbf{Q}$ — локальная система координат в некоторой карте расширенного фазового пространства, и $S(T, \mathbf{P}, \mathbf{Q})$, $G(T, \mathbf{P}, \mathbf{Q})$ — такие функции, что

$$\mathbf{p}\, d\mathbf{q} - H\, dt = \mathbf{P}\, d\mathbf{Q} - G\, dT + dS. \tag{5.66}$$

Левая и правая части суть формы на расширенном фазовом пространстве. Так как $ddS = 0$ и не влияет на линии ротора, то траектории фазового потока $\dot{\mathbf{p}} = -H'_{\mathbf{q}}$, $\dot{\mathbf{q}} = H'_{\mathbf{p}}$ изображаются на карте $T, \mathbf{P}, \mathbf{Q}$ интегральными кривыми канонических уравнений вида:

$$\frac{d\mathbf{P}}{dT} = -\frac{\partial G}{\partial \mathbf{Q}}, \quad \frac{d\mathbf{Q}}{dT} = \frac{\partial G}{\partial \mathbf{P}}. \tag{5.67}$$

Теорема. Пусть $\varrho\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ — каноническое преобразование фазового пространства, переводящее точку с координатами $(\mathbf{p}, \mathbf{q})$ в точку с координатами $(\mathbf{P}, \mathbf{Q})$. Функции $\mathbf{P}(\mathbf{p}, \mathbf{q})$, $\mathbf{Q}(\mathbf{p}, \mathbf{q})$ можно рассматривать как новые координаты в фазовом пространстве, в которых канонические уравнения имеют канонический вид со старой функцией Гамильтона $G(t, \mathbf{P}, \mathbf{Q}) = H(t, \mathbf{p}, \mathbf{q})$:

$$\frac{d\mathbf{P}}{dt} = -\frac{\partial H}{\partial \mathbf{Q}}, \quad \frac{d\mathbf{Q}}{dt} = \frac{\partial H}{\partial \mathbf{P}}. \tag{5.68}$$

Доказательство. 1-форма $\mathbf{p}d\mathbf{q} - \mathbf{P}d\mathbf{Q} = dS(\mathbf{p}, \mathbf{q})$ есть полный дифференциал функции S. Следовательно, в расширенном фазовом пространстве $\mathbf{p}d\mathbf{q} - Hdt = \mathbf{P}d\mathbf{Q} - Hdt + dS$, и применим предыдущий вывод.

Следовательно, *при любом каноническом преобразовании канонические уравнения сохраняют свой вид, а также величину функции Гамильтона.* Так как координаты независимы, то якобиан не равен нулю

$$J = \frac{\partial(P_1, \ldots, P_n, Q_1, \ldots, Q_n)}{\partial(p_1, \ldots, p_n, q_1, \ldots, q_n)} = \frac{\partial(\mathbf{P}, \mathbf{Q})}{\partial(\mathbf{p}, \mathbf{q})} \neq 0. \tag{5.69}$$

Каноническое преобразование называют *свободным*, если в окрестности некоторой точки $(\mathbf{p}_0, \mathbf{q}_0)$ за независимые координаты можно принять $\mathbf{Q}, \mathbf{q}$. Тогда в этой точке отличен от нуля якобиан:

$$\frac{\partial(\mathbf{Q}, \mathbf{q})}{\partial(\mathbf{p}, \mathbf{q})} = \det \frac{\partial \mathbf{Q}}{\partial \mathbf{q}} \neq 0. \tag{5.70}$$

В этом случае функция $S(\mathbf{p}, \mathbf{q}) = S_1(\mathbf{Q}, \mathbf{q})$ называется *производящей функцией свободного канонического преобразования.* Подчеркнём, что S_1 не есть функция на фазовом пространстве $\mathbb{R}^{2n}$: эта функция задана в области прямого произведения $\mathbb{R}^n_{\mathbf{q}} \times \mathbb{R}^n_{\mathbf{Q}}$ некоторых двух n-мерных координатных пространств. Тогда из $\mathbf{p}d\mathbf{q} - \mathbf{P}d\mathbf{p}Q = dS(\mathbf{p}, \mathbf{q})$ следует, что

$$\frac{\partial S_1(\mathbf{Q}, \mathbf{q})}{\partial \mathbf{q}} = \mathbf{p}, \quad \frac{\partial S_1(\mathbf{Q}, \mathbf{q})}{\partial \mathbf{Q}} = -\mathbf{P}.$$

Теорема. Пусть $S_1(\mathbf{Q}, \mathbf{q})$ — функция, заданная в окрестности некоторой точки $(\mathbf{Q}_0, \mathbf{q}_0)$ прямого произведения двух координатных евклидовых пространств. Если

$$\det \frac{\partial^2 S_1}{\partial \mathbf{Q} \partial \mathbf{q}}\bigg|_{\mathbf{Q}_0 \mathbf{q}_0} \neq 0, \tag{5.71}$$

то функция S_1 является производящей функцией некоторого свободного канонического преобразования $\varrho(\mathbf{p}, \mathbf{q}) = (\mathbf{Q}, \mathbf{q})$.

Из этой теоремы следует, что каноническое преобразование задаётся всего одной функцией $2n$ переменных — своей производящей функцией, что даёт выгоду во всех вычислениях, связанных с каноническими преобразованиями.

Пусть теперь функция $S(t, \mathbf{Q}, \mathbf{q})$ произвольная невырожденная функция старых координат, новых координат и времени. Пусть

$$\det \frac{\partial^2 S_1}{\partial \mathbf{Q} \partial \mathbf{q}} \neq 0.$$

Тогда она задаёт каноническое преобразование по правилу:

$$\mathbf{p} = \frac{\partial S}{\partial \mathbf{q}}, \quad \mathbf{P} = -\frac{\partial S}{\partial \mathbf{Q}}, \quad G = H + \frac{\partial S}{\partial t}. \tag{5.72}$$

На свойствах канонического преобразования построен *метод интегрирования канонических уравнений Гамильтона–Якоби. Уравнением*

Гамильтона–Якоби называется уравнение с частными производными первого порядка, в которое явно не входит значение неизвестной функции S, т.е. уравнение вида: $H(t, \frac{\partial S}{\partial x}\mathbf{q}, \mathbf{q}) = 0$. Идея метода состоит в следующем, если нам удастся найти каноническое преобразование, приводящее функцию Гамильтона к такому виду, что канонические уравнения удастся проинтегрировать, то тем самым мы сумеем проинтегрировать и исходные канонические уравнения. Например, канонические уравнения, в которых функция Гамильтона зависит от одних лишь переменных $\mathbf{Q}$: $H = G(\mathbf{Q})$, легко интегрируются. Действительно, $\dot{\mathbf{Q}} = G'_{\mathbf{p}}$, $\dot{\mathbf{P}} = -G'_{\mathbf{Q}}$, откуда $\mathbf{P}_t = \mathbf{P}_0 - tG'_{\mathbf{Q}}$. Тогда с учётом $\frac{\partial S}{\partial \mathbf{q}} = \mathbf{p}$ мы можем записать $H(t, \frac{\partial S(\mathbf{Q},\mathbf{q})}{\partial \mathbf{q}}, \mathbf{q}) = G(\mathbf{Q})$, где *после* дифференцирования вместо $\mathbf{q}$ следует поставить $\mathbf{q}(\mathbf{Q}, \mathbf{P})$.

Теорема Якоби. Если найдено решение $S(\mathbf{Q}, \mathbf{q})$ уравнения Гамильтона–Якоби $H(t, \frac{\partial S(\mathbf{Q},\mathbf{q})}{\partial \mathbf{q}}, \mathbf{q}) = G(\mathbf{Q})$, зависящее от n параметров $Q_1, \ldots, Q_n$ и такое, что $\det S''_{\mathbf{Q},\mathbf{q}} \neq 0$, то канонические уравнения $\dot{\mathbf{p}} = -H'_{\mathbf{q}}$, $\dot{\mathbf{q}} = H'_{\mathbf{p}}$ решаются явно в квадратурах. При этом функции $\mathbf{Q}(\mathbf{p}, \mathbf{q})$, определённые уравнениями $S'_{\mathbf{q}}(\mathbf{Q}, \mathbf{q}) = \mathbf{p}$ являются n первыми интегралами уравнений $\dot{\mathbf{p}} = -H'_{\mathbf{q}}$, $\dot{\mathbf{q}} = H'_{\mathbf{p}}$.

Далеко не все канонические преобразования свободные. Поэтому в качестве новых независимых координат не обязательно принимают координаты $\mathbf{Q}, \mathbf{q}$. Пусть, например, за независимые локальные координаты можно принять $\mathbf{P}, \mathbf{q}$. Тогда запишем:

$$\mathbf{p}\, d\mathbf{q} + \mathbf{Q}\, d\mathbf{P} = d(\mathbf{P} \cdot \mathbf{Q} + S(\mathbf{p}, \mathbf{q})) = dR, \tag{5.73}$$

где $R = \mathbf{P} \cdot \mathbf{Q} + S(\mathbf{p}, \mathbf{q})$ также называется производящей функцией.

5.6 Функциональный анализ

Рассмотрим интеграл от линейного элемента ds касательного пространства на римановом многообразии (см. (3.93)). Определение длины дуги как интеграла $\int ds$ (суммы всех бесконечно малых величин ds) соответствует как бы измерению длин «бесконечно малым масштабом» (как это отметил еще Б. Риман). Наикратчайший путь $\int ds$ между двумя точками называется *геодезической линией* (от греч. $\gamma\varepsilon\omega\delta\alpha\iota\sigma\acute{\iota}\alpha$, букв. — деление земли, $\gamma\eta$ — Земля и $\delta\alpha\acute{\iota}\omega$ — разделяю) и играет роль *прямой на римановом многообразии*. Через каждую точку риманова многообразия в любом направлении проходит геодезическая линия, и притом единственная. На плоскости геодезические линии суть прямые, на круговом цилиндре — винтовые линии, на сфере — окружности, центры которых совпадают с центром сферы (большие круги). Геодезическая линия обладает тем свойством, что её главная нормаль является нормалью к поверхности (вложенному многообразию). Так как определение геодезических линий связано только с измерениями на поверхности, они относятся к объектам внутренней геометрии поверхности.

Задача поиска наикратчайшей дуги (геодезической) относится к задачам *вариационного исчисления*. Для определения этого термина нам пона-

добится несколько новых понятий. Во-первых, расширим понятие пространства и введём термин — *пространство функций* или *функциональное пространство*, т.е. элементы пространства не являются числами, а представляют собой функции. Во-вторых, введём понятие *функционал* — числовая функция, заданная на пространстве, элементы которого не являются числами. Функционал имеет аргументом элементы линейного пространства (например, векторы) или функционального пространства и возвращает в качестве результата скаляр. В этом смысле алгебраические формы являются функционалами. Однако, наиболее часто понятие функционала используется именно в смысле функции, заданной на функциональном пространстве. В-третьих, расширим понятие дифференциала функции и введем *вариацию функционала*, понимая под этим термином его дифференциал. В этом смысле дифференциальные формы и дифференциальные операторы являются вариациями функционалов.

Однако, наиболее часто понятие вариации функционала используется именно в смысле дифференциала функции заданной на функциональном пространстве. Раздел математики, изучающей соответствующую категорию (функциональные пространства и их отображения — функционалы) называется *функциональный анализ*, а вариационное исчисление — это раздел функционального анализа, в котором изучаются вариации функционалов.

Например, множество всех возможных путей или кривых C между двумя точками на многообразии (гладкой поверхности) образует непрерывное множество скалярных функций (длина пути есть скаляр), т.е. образует функциональное пространство. Функция $S(C)$, заданная на этом функциональном пространстве есть функционал. Вариация функционала в заданной точке есть его дифференциал по соответствующему направлению. Функционал по наикратчайшему пути или геодезической линии называется *экстремаль* функционала, которая является расширением понятия экстремума функции. Кандидатами в экстремали будут направления, по которым вариация функционала равна нулю $\delta S = 0$.

Обращение вариации в нуль (или первой вариации, аналогично первому дифференциалу) достаточно для отыскания так называемых «стационарных точек» в функциональном пространстве. Действительно ли они являются экстремалями, т.е. достигается ли на них локальный экстремум, надо исследовать отдельно. Однако оказывается, что во многих прикладных задачах важнее найти не экстремали, а именно стационарные точки.

Лагранж рассматривал по преимуществу функционалы, связанные с названной в его честь функцией Лагранжа или, что то же самое, с интегральным инвариантом Пуанкаре-Картана, т.е. функционалы вида:

$$S \overset{(5.61)}{=} \int_{t_0}^{t_1} L(t, \mathbf{q}, \dot{\mathbf{q}})\, dt = \int \{\mathbf{p} d\mathbf{q} - H dt\}\,. \tag{5.74}$$

Обозначим $t_0 = 0$, а $t_1 = \tau$. Тогда функционал Φ при перемещении от

точки $\mathbf{q}_0$ к точке $\mathbf{q}_\tau$ по некоторой кривой C есть интеграл:

$$\Phi(C) = S(\tau, \mathbf{q}_\tau) - S(0, \mathbf{q}_0) = S_{\mathbf{q}_0,0}(\tau, \mathbf{q}) \overset{(4.7)}{=} \int_0^\tau L(t, \mathbf{q}, \dot{\mathbf{q}})\, dt. \tag{5.75}$$

Функционал $\Phi(C)$ на геодезической линии C называют *действие*.

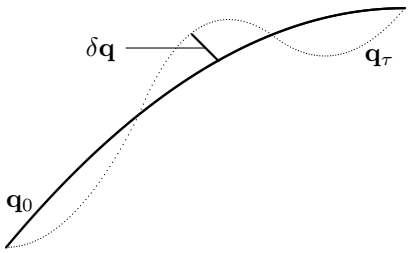


Рис. 5.7: Вариация экстремали

Функционал на геодезической линии есть экстремаль. Необходимым условием экстремали является обращение в нуль вариации функционала $\delta S = 0$. Чтобы определение экстремали было корректным, нужно принять некоторые предосторожности: нужно потребовать, чтобы экстремали, выходящие из точки $(t_0, \mathbf{q}_0)$ к разным точкам $(t, \mathbf{q})$, более не пересекались (при достаточно малом $|t - t_0|$), а образовывали так называемой «центральное поле экстремалей». Тогда экстремаль дифференцируема во всех точках в окрестности точки t_0.

Выясним, как изменяется значение функционала между точками $\mathbf{q}_0$ и $\mathbf{q}_\tau$ при замене экстремали другой достаточно близкой к ней траекторией, которая может быть получена бесконечно малой «деформацией» δ экстремали, т.е. при помощи *варьирования* (рис. 5.7).

При изменении координат на величины δq_i и $\delta \dot{q}_i = \frac{d}{dt}(\delta q_i)$, где δq_i — произвольные гладкие функции, принимающие малые значения, возникает вариация функционала. При этом время не варьируется, т.е. $\delta t = 0$. Начальная и конечная точки также остаются неподвижными, т.е. все

$$\delta q_i(0) = \delta q_i(\tau) = 0. \tag{5.76}$$

Также очевидно, что

$$\delta \int_0^\tau L\, dt = \int_0^\tau (\delta L)\, dt \quad \text{и} \quad \int \sum = \sum \int.$$

Тогда получаем:

$$0 = \delta S = \delta \int_0^\tau L(t, \mathbf{q}, \dot{\mathbf{q}})\, dt = \sum_{i=1}^n \int_0^\tau \left(\frac{\partial L}{\partial q_i} \delta q_i + \frac{\partial L}{\partial \dot{q}_i} \frac{d}{dt}(\delta q_i) \right) dt.$$

При интегрировании по частям:

$$\int_0^\tau \frac{\partial L}{\partial \dot{q}_i} \frac{d}{dt}(\delta q_i)\, dt \overset{(4.20)}{=} \frac{\partial L}{\partial \dot{q}_i}\, \delta q_i - \int_0^\tau \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) \delta q_i\, dt.$$

Тогда

$$0 = \delta S = \sum_{i=1}^{n} \frac{\partial L}{\partial \dot{q}_i}\, \delta q_i \Big|_0^\tau + \sum_{i=1}^{n} \int_0^\tau \left[\frac{\partial L}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i} \right) \right] \delta q_i \, dt. \tag{5.77}$$

Первый член в (5.77) равен нулю в силу задания граничных условий (5.76). Остается интеграл, который должен быть равен нулю при произвольных и независимых значениях δq_i. Это возможно только в том случае если подынтегральное выражение тождественно обращается в нуль.

Тогда:

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i} \right) = 0 \quad (i = 1, \ldots, n).$$

В результате мы получили уравнения Эйлера–Лагранжа (5.47).

Если мы рассмотрим интегральный инвариант Пуанкаре–Картана, то его интеграл $\int \{\mathbf{p} d\mathbf{q} - H dt\}$ имеет экстремалью линию ротора, так как этот интеграл по «бесконечно малому параллелограмму, проходящему через направление ротора», равен нулю. Если это рассуждение покажется недостаточно строгим, его можно заменить выкладкой:

$$\begin{aligned} 0 &= \delta \int_0^\tau \left(\sum_{i=1}^{n} p_i \dot{q}_i - H \right) dt = \\ &= \sum_{i=1}^{n} \int_0^\tau \left(\dot{q}_i\, \delta p_i + p_i\, \delta \dot{q}_i - \frac{\partial H}{\partial p_i}\, \delta p_i - \frac{\partial H}{\partial q_i}\, \delta q_i \right) dt = \\ &= \sum_{i=1}^{n} p_i\, \delta q_i |_0^\tau + \sum_{i=1}^{n} \int_0^\tau \left[\left(\dot{q}_i - \frac{\partial H}{\partial p_i} \right) \delta p_i - \left(\dot{p}_i + \frac{\partial H}{\partial q_i} \right) \delta q_i \right] dt. \end{aligned}$$

Откуда сразу следуют уравнения Гамильтона, которые, как мы знаем, задают векторное поле фазовых скоростей и линии ротора — траектории фазового потока в расширенном фазовом пространстве.

При варьировании интегрального инварианта Пуанкаре–Картана допускается значительно более широкий класс кривых, чем при варьировании функции Лагранжа, так как при этом никаких ограничений не накладывается. Может показаться удивительным, но оба рассмотрения при этом эквивалентны: из экстремальности в более узком классе вариаций ($\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}}$) следует экстремальность при любых вариациях. Объяснение состоит в том, что при фиксированном $\dot{\mathbf{q}}$ величина $\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}}$ экстремизирует $\mathbf{p}\dot{\mathbf{q}} - H$ при преобразовании Лежандра.

Если путь C от $\mathbf{q}_0$ к $\mathbf{q}_\tau$ параметризован так, что функция Гамильтона имеет фиксированное значение $H = H(\mathbf{q}, \frac{\partial L}{\partial \dot{\mathbf{q}}}) = const$, то запишем

$$\Phi(C) = \int_{\mathbf{q}_0}^{\mathbf{q}_\tau} \mathbf{p}\, d\mathbf{q} = \int_0^\tau (\mathbf{p} \cdot \dot{\mathbf{q}})\, dt = \int_0^\tau \frac{\partial L}{\partial \dot{\mathbf{q}}}\, \dot{\mathbf{q}}\, dt. \tag{5.78}$$

Экстремаль функционала $\Phi(C) = \int \mathbf{p}d\mathbf{q}$ называют «*укороченное действие*». Важно отметить, что отрезок $0 \leqslant t \leqslant \tau$, параметризующий кривую C, не фиксирован, и может быть разным у сравниваемых кривых. Зато одинаковым должно быть значение функции Гамильтона H.

Геодезические линии впервые появились в работах Я. Бернулли (1697–98), И. Бернулли и Л. Эйлера (1728–32). Термин «геодезическая» введён П.Лапласом (1798–99) применительно к «кратчайшим линиям» на земной поверхности. Геодезические линии на произвольной поверхности изучал Ж. Лиувилль (1844).

Понятие вариации функционала используется в теории экстремальных задач для получения необходимых и достаточных условий экстремума. Именно такой смысл вкладывается в этот термин, начиная с работы 1762 года Ж. Лагранжа. Впервые вариант решения поставленной задачи был предложен Л. Эйлером (1744).

Функциональный анализ развивался параллельно с развитием теоретической физики, в первую очередь квантовой и статистической механики, как самостоятельный раздел математики сложился на рубеже 19 и 20 вв., и окончательно оформился в 20–30-е годы 20 в.

5.7 Теорема Лиувилля–Арнольда

Настоящий параграф посвящён важному частному случаю, когда динамическую систему удаётся проинтегрировать в квадратурах.

Как мы уже знаем, разные фазовые кривые автономной системы не пересекаются (см. стр. 204). Для одной фазовой кривой существует три возможности: фазовая кривая либо не самопересекается, либо сводится к одной точке, либо является замкнутой фазовой кривой (диффеоморфной окружности). Незамкнутые фазовые кривые, хотя и не могут самопересекаться, могут сложным образом навиваться сами на себя.

В зависимости от того, размещается фазовая кривая в неограниченной или в конечной (например, замкнутые кривые) областях фазового пространства для $t = (-\infty, \infty)$, движение называется соответственно *инфинитным* или *финитным*.

При наличии непостоянных первых интегралов фазовые кривые целиком лежат на их поверхности, и система часто совершает финитное движение. При этом замкнутые фазовые кривые задают периодическое (циклическое) движение (процесс) или *колебание*. Системы, совершающие колебания, называют *осцилляторами* (от лат. oscillo — качаюсь). *Частота колебания* ν — характеристика периодического процесса, равная числу полных циклов, совершённых за единицу времени. Время полного цикла T, есть обратная частоте величина, называется *периодом*:

$$\nu = \frac{\omega}{2\pi} = \frac{1}{T}. \tag{5.79}$$

Величина ω называется *круговая частота*.

Если частота непостоянна, но меняется достаточно медленно, то фазовые кривые при этом незамкнутые и сложным образом навивают сами на себя, а движение называется условно-периодическим или квазипериодическим.

Теорема Лиувилля–Арнольда (условия интегрируемости системы $2n$ уравнений с периодическим или условно-периодическим движением). Если в системе с n степенями свободы (т.е. в $2n$-мерном фазовом пространстве) известны n независимых первых интегралов в инволюции, то система интегрируема в квадратурах.

Дадим точную формулировку теоремы Лиувилля–Арнольда. Предположим, что на симплектическом $2n$-мерном многообразии даны n функций $F_1(\nu), \ldots, F_n(\nu)$ в инволюции $\{F_i, F_j\} \equiv 0$. Рассмотрим множество уровня функций этих функций: $M_\psi = \{\nu\colon F_i(\nu) = f_i; i = 1, \ldots, n\}$; (если $\psi = (f_1, \ldots, f_n)$, то $\Psi(\nu) = \psi$). Предположим теперь, что на M_ψ n функций F_i независимы (см. стр. 217) (т.е. $\operatorname{grad} F_1, \ldots, \operatorname{grad} F_n$ линейно независимы в каждой точке M_ψ). Примем независимые функции F_i за n-независимых координат $\Psi = (F_1, \ldots, F_n)$. Тогда

1. M_ψ — гладкое многообразие, инвариантное относительно фазового потока с функцией Гамильтона $H = F_1$.

2. Если многообразие M_ψ компактно и связно, то оно диффеоморфно n-мерному тору T^n с угловыми координатами $\varphi = (\varphi, \ldots, \varphi)$.

3. Фазовый поток с функцией Гамильтона H определяет на M_ψ условно-периодическое движение (однопараметрическую группу диффеоморфизмов $T^n \to T^n$), т.е. в угловых координатах: $\dot\varphi = \omega$, $\omega(\psi) = (\omega_1, \ldots, \omega_n)$. Величины $\omega_1, \ldots, \omega_n$ называют *частотами* условно-периодического движения. Так как $\Psi(\nu) = \psi$, то во введённых координатах Ψ, φ фазовый поток с функцией Гамильтона $H = F_1$ записывается в виде особенно простой системы $2n$ обыкновенных дифференциальных уравнений: $\dot\Psi = 0$, $\dot\varphi = \omega(\Psi)$.

4. Канонические уравнения с функцией Гамильтона H интегрируются в квадратурах: $\Psi(t) = \Psi(0)$, $\varphi(t) = \varphi(0) + \omega(\Psi(0))\, t$.

По условию $\operatorname{grad} F_1, \ldots, \operatorname{grad} F_n$ линейно независимы в каждой точке M_ψ, т.е. определяют вложенное в $2n$-мерное фазовое пространство подмногообразие M_ψ (см. стр. 133). Оказалось, что это подмнообразие представляет собой n-мерный тор (диффеоморфно тору).

Симплектическая структура фазового пространства определяет оператор I, переводящий $\operatorname{grad} F_i$ в векторное поле $I\operatorname{grad} F_i$ фазовых скоростей системы с функцией Гамильтона F_i. Из независимости различных $\operatorname{grad} F_i$ и невырожденности изоморфизма I следует независимость $I\operatorname{grad} F_i$ в каждой точке M_ψ. Поля $I\operatorname{grad} F_i$ *попарно коммутируют*, так как F_i находятся в инволюции $\{F_i, F_j\} \equiv 0$.

По этой же причине для любых пар $i, j = 1, \ldots, n$ производная функции F_i по направлению поля $I\mathrm{grad}\, F_j$ равна нулю, т.е. все поля $I\mathrm{grad}\, F_i$ *касаются* M_ψ. Следовательно, многообразие M_ψ *нулевое*, т.е. 2-форма $\omega^2 = d\mathbf{p} \wedge \mathbf{q}$ обращается в нуль на M_ψ. Таким образом, касательные к M_ψ n векторов $I\mathrm{grad}\, F_i$ попарно косоортогональны и образуют базис в касательной плоскости к многообразию M_ψ в точке ν.

Переменные Ψ, φ не являются, вообще говоря, симплектическими координатами, но, оказывается, существуют функции от Ψ: $\mathbf{I} = \mathbf{I}(\Psi) = (I_1, \ldots, I_n)$, такие, что переменные $\mathbf{I}, \varphi$ уже есть симплектические координаты.

Для построения координат $I_1, \ldots, I_n$ используем 1-форму $\omega^1 = \mathbf{p}\, d\mathbf{q}$. Пусть $C_1, \ldots, C_n$ — базисные одномерные циклы тора M_ψ, т.е. приращение координаты φ_i на цикле C_j равно 2π, если $i = j$ и нулю, если $i \neq j$.

Тогда положим:

$$I_i(\psi) = \frac{1}{2\pi} \oint\limits_{C_i} \mathbf{p}\, d\mathbf{q}. \tag{5.80}$$

Так как 2-форма $\omega^2 = d\mathbf{p} \wedge d\mathbf{q} = d\omega^1 = 0$, то 1-форма $\omega^1 = \mathbf{p}\, d\mathbf{q}$ — замкнута. Поэтому интеграл (5.80) не зависит от выбора кривой C_i, представляющей базисный цикл. Действительно, по теореме Стокса:

$$\oint\limits_{C'_i} \mathbf{p}\, d\mathbf{q} - \oint\limits_{C_i} \mathbf{p}\, d\mathbf{q} = \iint\limits_{\sigma} d\mathbf{p} \wedge d\mathbf{q} = 0, \quad \text{где} \quad \partial\sigma = C'_i - C_i.$$

Чтобы построить каноническое преобразование $\mathbf{p}, \mathbf{q} \mapsto \mathbf{I}, \varphi$ будем искать его производящую функцию $S(\mathbf{I}, \mathbf{q})$ по правилу (5.72):

$$\mathbf{p} = \frac{\partial S(\mathbf{I}, \mathbf{q})}{\partial \mathbf{q}}, \quad \varphi = \frac{\partial S(\mathbf{I}, \mathbf{q})}{\partial \mathbf{I}} = (\varphi_1, \ldots, \varphi_n). \tag{5.81}$$

Функция

$$S(\mathbf{I}, \mathbf{q}) = \int\limits_{\mathbf{q}_0}^{\mathbf{q}_1} \mathbf{p}(\mathbf{I}, \mathbf{q})\, d\mathbf{q},$$

не меняется при деформации пути интегрирования и мы можем принять её за производящую функцию канонического преобразования $\mathbf{p}, \mathbf{q} \mapsto \mathbf{I}, \varphi$.

Если при заданных значениях $f_i = F_i(\nu)$ величины I_i независимы: $\det \frac{\partial \mathbf{I}}{\partial \psi} \neq 0$, то в окрестности тора M_ψ можно принять переменные $\mathbf{I}, \varphi$ за координаты.

При обходе замкнутой кривой C_i интеграл $\int \mathbf{p}\, d\mathbf{q}$ получает приращение, равное площади Π_i, ограниченной кривой C_i

$$\Delta_i S = \oint\limits_{C_i} \mathbf{p}\, d\mathbf{q} = \Pi_i = 2\pi I_i.$$

Поэтому S — «многозначная функция» на кривой C_i: она определена с точностью до прибавления кратного Π_i. На производную $\mathbf{p} = \frac{\partial S(\mathbf{I},\mathbf{q})}{\partial \mathbf{q}}$ это слагаемое не влияет, но оно приводит к неоднозначности $\varphi = \frac{\partial S(\mathbf{I},\mathbf{q})}{\partial \mathbf{I}}$:

$$\Delta_i \varphi_j = \Delta_i \frac{\partial S}{\partial I_j} = \frac{\partial}{\partial I_j} \Delta_i S = \frac{\partial}{\partial I_j} 2\pi I_i = 2\pi \delta_{ij}.$$

Величины $I_1, \ldots, I_n$ являются первыми интегралами системы с функцией Гамильтона $H = F_1$. Нетрудно сообразить, что переменные $I_1, \ldots, I_n$ имеют размерность действия (5.75), называются *переменными действия* и вместе с *угловыми переменными* $\varphi_1, \ldots, \varphi_n$ образуют в окрестности тора M_ψ систему канонических координат «*действие–угол*». В переменных действие-угол система $2n$ дифференциальных уравнений фазового потока аналогична (5.39) и имеет вид:

$$\dot{\mathbf{I}} = -\frac{\partial H}{\partial \varphi} = 0, \quad \dot{\varphi} = \frac{\partial H}{\partial \mathbf{I}} \equiv \omega(\mathbf{I}) = (\omega_1, \ldots, \omega_n). \tag{5.82}$$

Уравнения (5.82) немедленно интегрируются:

$$\mathbf{I}(t) = \mathbf{I}(0), \quad \varphi(t) = \varphi(0) + \omega t.$$

Преобразование $\mathbf{p}, \mathbf{q} \mapsto \mathbf{I}, \varphi$ каноническое, т.е. исходная симплектическая структура ω^2 выражается через $\mathbf{I}, \varphi$ по формуле:

$$\omega^2 = dp_1 \wedge dq_1 + \ldots + dp_n \wedge dq_n = dI_1 \wedge d\varphi_1 + \ldots + dI_n \wedge d\varphi_n. \tag{5.83}$$

Обратим внимание ещё раз. Чтобы проинтегрировать неавтономную систему $2n$ обыкновенных дифференциальных уравнений, нужно знать $2n$ первых интегралов. На практике оказывается, что если дана каноническая система дифференциальных уравнений, то во многих случаях движение системы оказывается условно-периодическим, и тогда достаточно знать лишь n первых интегралов ($F_1 = H, F_2, \ldots, F_n$) — каждый из них позволяет понизить порядок системы не на одну, а на две единицы.

Замечание. Если каноническая система интегрируется методом Гамильтона-Якоби, то она имеет n первых интегралов в инволюции.

Переменные действия $I_1, \ldots, I_n$ или относительные интегральные инварианты Пуанкаре называют также *адиабатическими инвариантами* (от др.-греч. $\alpha\delta\iota\alpha\beta\alpha\tau o\varsigma$ — непроходимый) — величинами, остающимися практически неизменными при медленном (адиабатическом) изменении параметров системы (например, ω).

Циклическое движение даёт n частот по n неприводимым контурам тора $C_1, \ldots, C_n$. Частоты $\omega_1, \ldots, \omega_n$ называются независимыми, когда они линейно независимы над полем рациональных чисел: если $k_1\omega_1 + \ldots + k_n\omega_n = 0$, где $k_1, \ldots, k_n \in \mathbb{Z}$, может выполняться только когда $k_1 = \ldots = k_n = 0$. Когда частоты зависимы, то целочисленный вектор $\mathbf{k} = (k_1, \ldots, k_n) \in \mathbb{Z}^n$ называется *соотношением* между частотами при $k_1\omega_1 + \ldots + k_n\omega_n = 0$, а система называется *вырожденной*.

В процессе движения траектория остается все время на торе. Поэтому говорят также о существовании инвариантных торов в интегрируемом случае. Фазовая траектория на торе называется *обмоткой тора*. Если частоты независимы, то система невырождена, траектория незамкнута и всюду плотна (обмотка плотно заполняет тор). При этом

$$\det \frac{\partial \omega}{\partial \mathbf{I}} = \det \frac{\partial^2 H}{\partial \mathbf{I}^2} \neq 0 \quad \text{или} \quad \det \frac{\partial \omega_i}{\partial I_j} = \det \frac{\partial^2 H}{\partial I_i \partial I_j} \neq 0. \tag{5.84}$$

где члены матрицы $\frac{\partial \omega}{\partial \mathbf{I}}$ удовлетворяют условию симметрии: $\frac{\partial \omega_i}{\partial I_j} = \frac{\partial \omega_j}{\partial I_i}$.

Если частоты зависимы, то траектория замкнута и не заполняет тор плотно. *Замыкания траекторий* вырожденной системы — это торы размерности меньше n. Если между частотами имеется r соотношений $\mathbf{k}_1, \ldots, \mathbf{k}_r$, где $1 \leqslant r \leqslant n$, то говорят об r-кратном вырождении, а замыкание траектории на T^n представляет собой тор T^{n-r}, движение на котором — условно-периодическое с $n-r$ независимыми частотами, а траектории всюду плотно заполняют тор $(n-r)$-ого порядка.

Если система невырожденна, а частоты независимы, то инвариантные торы $\mathbf{I} = const$ определены однозначно, несмотря на неоднозначность в выборе переменных действие-угол, в построении которых всегда имеется некоторый произвол. Напротив, в вырожденной системе можно построить такие системы переменных действие-угол, что торы в одной и в другой системе будут разными. Это объясняется тем, что замыкания траекторий вырожденной системы — это торы размерности меньше n, и их можно по-разному соединять в n-мерные торы.

При $n = 2$ двумерные торы T^2, соответствующие различным значениям интегралов $I_1, I_2 = const$, *делят пространство* (вложены друг в друга и не пересекаются). При $n > 2$ торы T^n не делят фазовое пространство и пересекаются. Это легко понять из следующих соображений. В $2n$-мерном фазовом пространстве поверхность имеет размерность $2n-1$. Границы поверхности, которые должны делить пространство на различные области, имеют размерность $2n-2$. Если торы делят пространство, то их размерность n должна удовлетворять условию $n \geqslant 2n-2$ или $n \leqslant 2$.

Например, движение на двумерном торе характеризуется двумя частотами: ω_1 и ω_2. В общем случае эти частоты несоизмеримы, и тогда фазовая траектория всюду плотно покрывает поверхность тора и является незамкнутой. Если отношение этих частот $\frac{\omega_1}{\omega_2}$ есть рациональное число, то возникает так называемый *внутренний резонанс*, при котором траектория замыкается через конечное число оборотов на торе и, следовательно, не покрывают поверхность тора. Очевидно, что условие резонанса означает вырождение, при котором частоты являются зависимыми, в противном случае частоты независимы друг от друга.

Рассмотрим систему с двумя степенями свободы, в которой частоты ω_1 и ω_2 сильно различаются $\omega_1 \gg \omega_2$. Пусть для простоты $\omega_1 = \omega_1(I_1), \omega_2 = \omega_2(I_2)$, $H = const$ и $k_1\omega_1 = k_2\omega_2$, где $k_1, k_2 \in \mathbb{Z}$ — условие резонанса. Тогда можно выразить I_2 через H и I_1, а I_1 через ω_1, и функция $\omega_2 = \omega_2(H, \omega_1)$

имеет на плоскости (ω_1, ω_2) вид некоторой кривой (рис. 5.8).

Условию резонанса соответствует на плоскости (ω_1, ω_2) семейство прямых $\omega_2 = \frac{k_1}{k_2}\omega_1$ с различными наклонами $\frac{k_1}{k_2}$. Точки их пересечения с кривой $\omega_2 = \omega_2(H, \omega_1)$ дают резонансные точки. Как известно, в точках резонанса система становится неустойчивой, поэтому области в окрестности этих точек считаются «опасными».

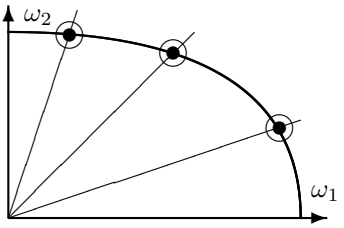


Рис. 5.8: Резонансные точки при $n = 2$ и «опасные области» в их окрестности

При $n > 2$ инвариантные торы пересекаются между собой. Поэтому опасная область представляет собой не просто окрестности точек, а систему каналов, покрывающую всю гиперповерхность постоянной энергии. Отмеченный топологический эффект сразу приводит к необычайно важному следствию для анализа устойчивости системы: попав даже в сколь угодно узкий резонансный канал, система может уйти сколь угодно далеко на поверхности постоянной энергии. Эти соображения сформулированы В. И. Арнольдом и легли в основу представлений о диффузии Арнольда [6].

Глава 6

Функция вероятности

Функция вероятности — одна из разновидностей функций, имеющая очень большое значение для всей науки. Поэтому функции вероятности следует посвятить отдельную главу. Хотя функция вероятности определяется по тем же правилам, что и любая другая функция, и не имеет никаких специфических свойств.

6.1 Предварительные определения

Теория вероятностей — математическая наука, позволяющая по вероятностям одних случайных событий находить вероятности других случайных событий, связанных каким-либо образом с первыми.

Случайное (*элементарное*) *событие* (исход) — событие, наступление или ненаступление (возможно мысленное) которого в некотором испытании (эксперименте) зависит от ряда случайных факторов и для которого постулируется определенная вероятность его наступления при определённых, неограниченно воспроизводимых условиях.

Пространство элементарных событий — множество Ω всех различных исходов случайного эксперимента. Пространство элементарных событий называется дискретным, если число его элементов конечно или счётно. Любое пространство элементарных событий не являющееся дискретным, называется недискретным, и при этом, если наблюдаемыми результатами (нельзя произносить случайными событиями) являются точки того или иного числового арифметического или координатного пространства, то пространство называется непрерывным (*континуум*).

Функция вероятности P имеет свою область определения в Ω.

Кроме того, задаются определенные алгебраические операции над множеством элементарных событий, т.е. некая алгебра A.

Совокупность пространства элементарных событий, алгебры и функции вероятности определяет *вероятностное пространство*:

$$(\Omega, A, P) \tag{6.1}$$

где Ω — пространство элементарных событий,

A — σ-алгебра подмножеств Ω (непустая совокупность подмножеств множества Ω, замкнутая относительно теоретико-множественных операций (объединения, пересечения, дополнения), производимых в конечном числе),

P — вероятность (функция вероятности), определенная для элементов класса A как неотрицательная, нормированная, аддитивная функция множества.

Случайное событие ω — подмножество Ω ($\omega \subset \Omega$), являющееся элементом класса A, обладает одной особенностью, оно обязательно наступает и поэтому называется *достоверным*, а событие $\varnothing$ (пустое множество) никогда не наступает и поэтому называется *невозможным*.

Каждое испытание заканчивается одним и только одним из случайных событий (исходов) $\omega_1, \omega_2, \ldots$.

В классическом определении *вероятность* P события ω вычисляется как отношение числа «благоприятствующих» данному событию исходов n, к общему числу «равновозможных» исходов N при тех или иных неограниченно повторяющихся условиях. Эта вероятность выражается числом p:

$$P(\omega) = \frac{n}{N} = p. \tag{6.2}$$

Следовательно, с каждым исходом $\omega_1, \omega_2, \ldots$ связывается неотрицательное число — вероятность этого исхода (события) $p_1, p_2, \ldots$.

Сумма вероятностей всех событий является *достоверным событием*, т.е. наступает с вероятностью 1: $p_1 + p_2 + \ldots = 1$.

Термин вероятность не имеет исчерпывающего определения. Гипотеза вероятности предполагает, что существует некое нормальное число наступлений некого данного события при большом числе повторений.

Статистическое определение вероятности позволяет вычислить новые вероятности по уже вычисленной ранее вероятности. В современной теории вероятностей свойства вероятности формулируются в виде аксиом. Однако ни эти аксиомы, ни классический, ни статистический подходы не дают исчерпывающего определения вероятности.

Понятие вероятность имеет смысл не для всех случайных событий, а лишь для тех из них, которые обладают статистической однородностью, или устойчивостью. Универсального, математически строгого определения статистической устойчивости также не существует.

Частота события в серии из N испытаний имеет вид:

$$\frac{m_N}{N}, \tag{6.3}$$

где μ_N — число тех испытаний, в которых событие ω наступило,

N — общее число испытаний (повторных осуществлений условий).

Наличие у события ω при определенных условиях вероятности, равной p, проявляется в том, что почти в каждой достаточно длинной серии

испытаний частота события $\frac{m_N}{N}$ приблизительно равна p:

$$\frac{m_N}{N} \approx p.$$

Мы можем теперь сформулировать одну из важнейших теорем теории вероятностей, найденную Я. Бернулли и опубликованную уже после его кончины в его книге «Искусство предположений» (1713).

Теорема Бернулли — закон больших чисел в простейшей форме:

Если вероятность наступления некоторого случайного события ω в последовательности N независимых испытаний постоянна и равна p, то частота события мало отличается от вероятности p или для любых $\varepsilon > 0$ и $\eta > 0$ при достаточно большом числе испытаний N $N \to \infty$ имеет место неравенство:

$$P\left\{\left|\frac{m_N}{N} - p\right| < \varepsilon\right\} \geqslant 1 - \eta. \tag{6.4}$$

Создателем теории вероятностей считается П. Лаплас, хотя до него первые шаги в этой области были сделаны Б. Паскалем, П. Ферма, Я. Бернулли и другими. Классический труд Лапласа «Аналитическая теория вероятностей» издавался трижды при его жизни — в 1812, 1814 и 1820; в качестве введения к последним изданиям была помещена работа «Опыт философии теории вероятностей», в которой в популярной форме разъясняются основные положения и значение теории вероятностей.

6.2 Операции с вероятностями

Событие ζ называется *объединением* (или суммой) событий, если оно имеет вид: «наступает или ω_1, или ω_2, или ...»

$$\zeta = \omega_1 \cup \omega_2 \cup \ldots. \tag{6.5}$$

Событие ξ называется *совмещением* (или *произведением*) событий, если оно имеет вид: «наступает и ω_1, и ω_2, и ...»

$$\xi = \omega_1 \cap \omega_2 \cap \ldots. \tag{6.6}$$

События ω_1 и ω_2 называются *несовместными*, если их одновременное осуществление (совмещение) невозможно.

Теорема сложения вероятностей: если события несовместны, то вероятность их объединения равна сумме их вероятностей:

$$P(\omega_1 \cup \omega_2 \cup \ldots) = P(\omega_1) + P(\omega_2) + \ldots. \tag{6.7}$$

Теорема умножения вероятностей: вероятность совмещения независимых событий равна произведению вероятностей этих событий

$$P(\omega_1 \cap \omega_2 \cap \ldots) = P(\omega_1) \cdot P(\omega_2) \cdot \ldots. \tag{6.8}$$

Для зависимых событий теорема звучит так:

Вероятность совмещения событий $\omega_1, \omega_2, \ldots$ равна вероятности события ω_1 умноженной на вероятность события ω_2, взятую при условии, что ω_1 наступило, умноженной на вероятность события ω_3, взятую при условии, что ω_1 и ω_2 наступили и т.д.

При решении вероятностных задач часто бывает важно определить вероятность события, когда о нем имеются некоторые дополнительные сведения (условия). Обычная ситуация: нужно найти вероятность события после того, как стало известно, что произошел некоторый исход, благоприятствующий событию.

Условная вероятность события ζ при условии осуществления события ξ (с $P(\xi) > 0$) определяется формулой:

$$P(\zeta|\xi) = \frac{P(\zeta \cap \xi)}{P(\xi)} \quad \text{или} \quad P(\zeta \cap \xi) = P(\xi)P(\zeta|\xi) = P(\zeta)P(\xi|\zeta). \tag{6.9}$$

Здесь $P(\zeta|\xi)$ условная вероятность, а $P(\xi)$ и $P(\zeta \cap \xi)$ безусловные вероятности.

Пусть теперь $\zeta = \omega_1 \cup \omega_2 \cup \ldots$ и $P(\zeta) = P(\omega_1) + P(\omega_2) + \ldots$, тогда получаем одну из основных формул теории вероятностей формулу *полной вероятности*:

$$P(\xi) = P(\omega_1)P(\xi|\omega_1) + P(\omega_2)P(\xi|\omega_2) + \ldots. \tag{6.10}$$

Из (6.9) следует:

$$P(\zeta|\xi) = \frac{P(\zeta)P(\xi|\zeta)}{P(\xi)}. \tag{6.11}$$

Заменим в (6.11) ζ на ω_n и, с учетом (6.10), получим *теорему Байеса*:

$$P(\omega_n|\xi) = \frac{P(\omega_n)P(\xi|\omega_n)}{P(\omega_1)P(\xi|\omega_1) + P(\omega_2)P(\xi|\omega_2) + \ldots} \tag{6.12}$$

Рассмотрим в качестве примера задачу, которая представляет значительный интерес для ряда основных вопросов физики, химии, биологии, инженерного дела и пр. Имеется N ячеек и n частиц. Частицы наудачу размещаются по ячейкам. Нужно найти вероятность каждого из возможных размещений.

Существует три различных подхода, выработанных в физике и получивших соответственно наименование статистик Максвелла–Больцмана, Бозе–Эйнштейна и Ферми–Дирака:

- *Статистика Максвелла–Больцмана.*

  Каждая из всех n различных частиц с вероятностью $\frac{1}{N}$ может попасть в каждую из ячеек, независимо от положения других частиц. Число всех возможных различных расположений частиц по ячейкам, как легко понять, равно N^n. Найдем теперь вероятность того, что в первой ячейке окажутся n_1 частиц, во второй — n_2, в N-й — n_N.

Понятно, что некоторые из чисел могут оказаться нулями. Повторив почти дословно рассуждения, проведенные нами при выводе формулы для числа сочетаний (1.28), мы получаем, что число всех возможных способов размещения n частиц по N ячейкам в количестве $n_1, \ldots, n_N$ соответственно равно:

$$\frac{n!}{n_1! \ldots n_N!}. \tag{6.13}$$

Тогда искомая вероятность указанного размещения равна:

$$P(n_1, \ldots, n_N) = \frac{n!}{n_1! \ldots n_N! N^n}. \tag{6.14}$$

В качестве частного случая рассмотрим эту задачу при $n \leqslant N$. Чему равна вероятность того, что в определенных ячейках окажется по одной частице, а в остальных по 0 частиц? Искомая вероятность, как это вытекает из формулы, равна:

$$p = \frac{n!}{N^n}. \tag{6.15}$$

Если ячейки неразличимы, то вероятность того, что по одной частице окажется в каких-то ячейках, будет больше на число сочетаний, т.е. в C_N^n раз:

$$P = C_N^n p = \frac{N!}{N^n (N-n)!} \tag{6.16}$$

- *Статистика Бозе–Эйнштейна.*

  Допустим, наши частицы и ячейки неразличимы и $n \leqslant N$. В этом случае число всех возможных способов размещения частиц по ячейкам окажется намного меньше чем в (1.27). Расположим на прямой $N+1$ вертикальную черточку. Каждую ячейку будем рассматривать как промежуток между двумя соседними черточками. Две крайние черточки оставим неподвижным и между ними поместим n частиц. Станем теперь переставлять всеми возможными способами $N-1$ внутреннюю черточку и n частиц. Число перестановок черточек и частиц равно $N+n-1$. Среди перестановок, однако, имеются тождественные. Действительно, за различные перестановки мы, во-первых, считали те, в которых поменялись местами черточки. Таким образом, каждое распределение мы считали $(N-n)!$ раз. Во-вторых, мы считали различными частицы, и тем самым каждое распределение мы снова считали $n!$ раз. Отсюда число различных в смысле Бозе-Эйнштейна распределений частиц по ячейкам равно:

$$C_{N+n-1}^n = \frac{(N+n-1)!}{n!(N-1)!}. \tag{6.17}$$

Вероятность попадания по одной частице в определенные ячейки равна:

$$p = \frac{1}{C_{N+n-1}^n} = \frac{n!(N-1)!}{(N+n-1)!}. \tag{6.18}$$

Вероятность попадания в какие-то n ячеек по одной частице равна:

$$P = \frac{C_N^n}{C_{N+n-1}^n} = \frac{N!(N-1)!}{(N+n-1)!(N-n)!}. \tag{6.19}$$

- *Статистика Ферми–Дирака.*

  В этой статистике не только уничтожена индивидуальность частиц, но и предполагается, что в каждой ячейке может находиться либо 0 частиц, либо 1 частица. Общее число распределений n частиц по N ячейкам ($n \leqslant N$) в статистике Ферми–Дирака равно C_N^n.

  Интересовавшие нас вероятности p и P в статистике Ферми–Дирака равны:

$$p = \frac{1}{C_N^n} = \frac{n!(N-n)!}{N!}, \quad P = 1. \tag{6.20}$$

6.3 Биноминальное распределение

Рассмотрим последовательность N испытаний, в каждом из которых вероятность наступления определенного события ω одна и та же и равна p, а вероятность ненаступления события равна $q = 1 - p$. Испытания предполагаются независимыми, т.е. вероятность появления события ω в каждом из испытаний не зависит от того, появлялось ли это событие в других испытаниях. Последовательность независимых испытаний с двумя исходами носит название *последовательности испытаний Бернулли.*

Вероятность того, что событие ω наступит при определенных n испытаниях, а при остальных $N - n$ не наступит, в силу теоремы умножения вероятностей равна $p^n q^{N-n}$. Но событие ω может произойти при любых наборах n из N возможных испытаний. Число всех наборов n элементов из N равно C_N^n. Поэтому, в силу теоремы сложения вероятностей, искомая вероятность равна:

$$P_N(n) = C_N^n p^n q^{N-n}. \tag{6.21}$$

Эта вероятность носит название *формула Бернулли* или *биноминальное распределение.*

Формула (6.21) дает в частности, вероятность того, что событие ω произойдет во всех N испытаниях $P_N(N) = p^N$, а также вероятность того, что оно не произойдет ни разу $P_N(0) = q^N$.

Сумма вероятностей всех событий является достоверным событием, т.е. наступает с вероятностью 1:

$$\sum_{n=0}^{N} P_N(n) = 1 = \sum_{n=0}^{N} C_N^n p^n q^{N-n} = (p+q)^N. \tag{6.22}$$

Мы получили частный случай формулы (1.24).

Практическое значение имеют два различных предела биноминального распределения при больших N, которые мы сейчас получим: локальная предельная теорема Муавра-Лапласа и распределение Пуассона.

Рассмотрим, как функция вероятности $P_N(n)$ зависит от n:

- Функция $P_N(n)$ возрастает, если $\frac{P_N(n+1)}{P_N(n)} = \frac{N-n}{n+1}\frac{p}{q} > 1$, т.е. если $n < Np - q$.
- Функция $P_N(n)$ достигает максимума или убывает если: $n = Np - q$ или $n > Np - q$.
- $q \ll Np$ при достаточно большом значении N , следовательно, функция $P_N(n)$ достигает максимума когда $n \approx Np$.

Воспользуемся формулой Дж. Стирлинга (1730) для разложения натурального логарифма $\ln n!$ в бесконечный ряд:

$$\ln n! = n\ln n - n + \ln\sqrt{2\pi n} + \frac{s_1}{n} - \frac{s_2}{n^3} + \ldots + (-1)^{k+1}\frac{s_k}{n^{2k-1}} + \ldots. \quad (6.23)$$

где числа s_k могут быть выписаны в явном виде, например,

$$s_1 = \frac{1}{12}, \quad s_2 = \frac{1}{360}.$$

Ряд этот расходится, однако при любом натуральном k верно равенство:

$$\ln n! = n\ln n - n + \ln\sqrt{2\pi n} + \frac{s_1}{n} - \frac{s_2}{n^3} + \ldots + (-1)^{k+1}\frac{s_k\theta}{n^{2k-1}} \quad (6.24)$$

где $0 < \theta < 1$.

Для нас представляют интерес такие следствия:

$$\ln n! \approx n\ln n \quad \text{или} \quad n! \approx n^n, \quad (6.25)$$

$$\ln n! \approx n\ln n - n \quad \text{или} \quad n! \approx e^{-n}n^n, \quad (6.26)$$

$$\ln n! \approx n\ln n - n + \ln\sqrt{2\pi n} \quad \text{или} \quad n! \approx \sqrt{2\pi n}e^{-n}n^n. \quad (6.27)$$

Заменяя в формуле (6.21) факториалы по формуле (6.27), получим:

$$P_N(n) = \sqrt{\frac{N}{2\pi n(N-n)}}\frac{e^n e^{N-n}}{e^N}\left(\frac{Np}{n}\right)^n\left(\frac{Nq}{N-n}\right)^{N-n}. \quad (6.28)$$

Введем новую переменную ε:

$$n = Np + \varepsilon \approx Np, \quad N - n = Nq - \varepsilon \approx Nq.$$

Тогда:

$$P_N(n) = \sqrt{\frac{N}{2\pi n(N-n)}}\, e^{-(Np+\varepsilon)\ln\left(1+\frac{\varepsilon}{Np}\right)-(Nq-\varepsilon)\ln\left(1-\frac{\varepsilon}{Nq}\right)}.$$

Для достаточно больших N, когда $\frac{\varepsilon}{N} \to 0$ воспользуемся разложениями показателя экспоненты в ряд (1.34, 1.35):

$$(Np+\varepsilon)\ln\left(1+\frac{\varepsilon}{Np}\right)+(Nq-\varepsilon)\ln\left(1-\frac{\varepsilon}{Nq}\right)=$$
$$=\frac{\varepsilon^2}{2N}\left(\frac{1}{p}+\frac{1}{q}\right)+\ldots=\frac{\varepsilon^2}{2Npq}+\ldots.$$

В результате, получим:

$$P_N(n)=\frac{1}{\sqrt{2\pi Npq}}\,e^{-\frac{\varepsilon^2}{2Npq}}=\frac{1}{\sqrt{2\pi}\sigma}\,e^{-\frac{\varepsilon^2}{2\sigma^2}}, \tag{6.29}$$

где $\sigma=\sqrt{Npq}$.

Полученное соотношение известно как *локальная предельная теорема Муавра–Лапласа*, доказанная П. Лапласом в книге «Аналитическая теория вероятностей» (1812). Один частный случай локальной теоремы был известен А. Муавру (1730).

Встречаются испытания Бернулли, в которых N относительно велико $N\to\infty$, p относительно мало, а произведение $\lambda=Np\approx n$ не мало, но и не велико. Тогда заменяя входящие в формулу (6.21) факториалы по формуле (6.26), получим:

$$P_N(n)=\frac{e^{N-n}}{n!e^N}(Np)^n\left(\frac{Nq}{N-n}\right)^{N-n}=\frac{\lambda^n e^{-n}}{n!}\left(\frac{N-\lambda}{N-n}\right)^{N-n}=$$
$$=\frac{\lambda^n e^{-n}}{n!}\left(\frac{1-\frac{\lambda}{N}}{1-\frac{n}{N}}\right)^{N-n}=\frac{\lambda^n}{n!}e^{(N-n)\ln\left(1-\frac{\lambda}{N}\right)-(N-n)\ln\left(1-\frac{n}{N}\right)-n}.$$

Воспользуемся разложением показателя экспоненты в ряд (1.35):

$$(N-n)\ln\left(1-\frac{\lambda}{N}\right)-(N-n)\ln\left(1-\frac{n}{N}\right)-n=-\lambda+\frac{\lambda n}{N}-\frac{n^2}{N}-\ldots.$$

В результате получим знаменитое *распределение Пуассона* или пуассоновское приближение для биноминального распределения по имени С. Пуассона (1837):

$$P(n)=\frac{\lambda^n}{n!}e^{-\lambda}. \tag{6.30}$$

При этом

$$\sum_{n=0}^{\infty}P(n)=\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}=e^{-\lambda+\lambda}=1. \tag{6.31}$$

6.4 Случайная величина

Случайная величина — числовая функция, определённая на множестве исходов. Совокупность значений, которые может принимать случайная величина ξ, и вероятностей, с которыми она их принимает $P(\xi)$, называют *распределением случайной величины*. Если различные значения случайной величины ξ образуют конечную или бесконечную последовательность $x_1, x_2, \ldots$ с соответствующими вероятностями $p_1, p_2, \ldots$, то случайная величина называется *дискретной*.

Сумма:

$$M\xi = \overline{x} = \sum_{n=1}^{N} x_n p_n \tag{6.32}$$

значений случайной величины, умноженных на соответствующие вероятности, называется *математическим ожиданием* случайной величины ξ или *средним значением* $\overline{x}$ случайной величины ξ. Термин «математическое ожидание» ввел П. Лаплас (1795).

Допустим, в системе существует N возможных исходов, при которых случайная величина x принимает значения x_i, с вероятностью каждого исхода p_i, $i = 1, \ldots, N$. Математическое ожидание или среднее значение величины $\overline{x}$ равно

$$\overline{x} = p_1 x_1 + \ldots + p_N x_N, \quad \text{при этом} \quad p_1 + \ldots + p_N = 1. \tag{6.33}$$

Количество *информации* I (от лат. information — разъяснение, изложение), необходимое, чтобы точно указать, какой из вариантов исхода в системе осуществится, вычисляют как математическое ожидание от логарифма вероятности

$$I = -\sum_{i=1}^{N} p_i \ln p_i. \tag{6.34}$$

Знак минус стоит для того, чтобы значение величины информации I было положительным. Описанный способ вычисления информации был предложен К. Шенноном (1948).

Первые отчетливые предложения об общих способах измерения количества информации принадлежат, по видимому Р. Фишеру (в связи с вопросами математической статистики) и Р. Хартли (в связи с вопросами хранения информации в запоминающих устройствах и передачей информации по каналам связи). Свое окончательное выражение эти предложения нашли в теории информации, созданной американским ученым К. Шенноном (1948), и теперь информация является основным понятием кибернетики. Законов информации известно немного, причем связаны они в основном с проблемой ее оптимальной передачи.

Если исходы равновероятны, то

$$p_i = \frac{1}{N}, \quad \overline{x} = \frac{x_1 + \ldots + x_N}{N}, \quad I = -N \frac{1}{N} \ln\left(\frac{1}{N}\right) = \ln N. \tag{6.35}$$

Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий слагаемых:

$$M(\xi + \zeta) = M\xi + M\zeta. \tag{6.36}$$

Математическое ожидание произведения двух независимых (дополнительное условие) случайных величин равно произведению математических ожиданий сомножителей:

$$M(\xi\zeta) = M\xi \cdot M\zeta. \tag{6.37}$$

Например, при бросании правильной кости с шестью гранями, случайная величина ξ принимает значения $1, 2, 3, 4, 5, 6$ с вероятностями $p_i = \frac{1}{6}$, $i = 1, 2, 3, 4, 5, 6$. При бросании монеты случайное событие ξ может принимать два значения «орел» и «решка» (например, 0 и 1) с вероятностью $\frac{1}{2}$ для каждого значения.

В случае биноминального распределения при каждом испытании событие ω может принимать два значения 1 или 0. Вероятность первого значения равна p, вероятность второго — q.

Математическое ожидание биноминального распределения $P(\mu) = C_N^n p^n q^{N-n}$ (6.21) равно

$$M\mu = Np. \tag{6.38}$$

Доказательство.

Математическое ожидание в любом испытании с номером n равно p:

$$M\mu_n = 1 \cdot p + 0 \cdot q = p. \tag{6.39}$$

где μ_n означает число появлений события в испытании с номером n.

Число появлений события в N испытаниях равно

$$\mu = \mu_1 + \ldots + \mu_N. \tag{6.40}$$

Тогда математическое ожидание появления события в N испытаниях равно $M\mu \overset{(6.36)}{=} Np$.

В распределении Пуассона математическое ожидание $M\mu = \lambda$.

Если это доказательство выглядит не слищком строгим, то можно предложить более длинный путь. Заметим, что

$$M\mu = \sum_{n=0}^{N} nP_N(n) = 0 \cdot^0 q^N + \sum_{n=1}^{N} nP_N(n) = \sum_{n=1}^{N} nP_N(n).$$

Тогда

$$\sum_{n=1}^{N} nP_N(n) = \sum_{n=1}^{N} \frac{N!n}{n!(N-n)!} p^n q^{N-n} = Np \sum_{n=1}^{N} \frac{(N-1)!}{(n-1)!(N-n)!} p^{n-1} q^{N-n},$$

Произведем замену $n-1=k$ и получим:

$$\sum_{n=1}^{N} nP_N(n) = Np \sum_{k=0}^{N-1} \frac{(N-1)!}{k!(N-1-k)!} p^k q^{N-1-k} = Np \sum_{k=0}^{N-1} C_{N-1}^k p^k q^{N-1-k}.$$

С учетом (6.22) получаем математическое ожидание (6.38).

Случайная величина может отклоняться от среднего значения как в большую, так и в меньшую сторону, что вызывает ошибку при её измерении. Если при измерении не допускаются систематические ошибки, связанные с особенностями наблюдателя и измерительного прибора, то среднее значение случайной ошибки равно 0:

$$M(\xi - M\xi) = 0 \tag{6.41}$$

Поэтому для измерения ошибки используют *дисперсию* (от лат. dispersion — рассеяние) — меру отклонения случайной величины от её среднего значения, вычисляемую как квадрат отклонения:

$$D\xi = M(\xi - M\xi)^2. \tag{6.42}$$

Дисперсии можно придать другую форму:

$$D\xi = M\left[\xi^2 - 2\xi M\xi + (M\xi)^2\right] = M\xi^2 - 2(M\xi)^2 + (M\xi)^2.$$

В результате:

$$D\xi = M\xi^2 - (M\xi)^2 = \sum_{n=1}^{N} x_n^2 p_n - \left(\sum_{n=1}^{\infty} x_n p_n\right)^2. \tag{6.43}$$

Дисперсия суммы независимых случайных величин равна сумме их дисперсий:

$$D(\xi_1 + \xi_2 + \ldots) = D\xi_1 + D\xi_2 + \ldots. \tag{6.44}$$

Вычислим дисперсию биноминального распределения.

Мы можем вычислить величину

$$M\mu^2 = \sum n = 0^N n^2 P_N(n) = N^2 p^2 + Npq.$$

Тогда дисперсия биноминального распределения равна:

$$D\mu = M\mu^2 - (M\mu)^2 = Npq.$$

Но мы можем сразу получить этот результат, не вычисляя предварительно величину $M\mu^2$, а вычислив лишь $M\mu_n^2$:

$$M\mu_n^2 = 1^2 p + 0^2 q = p.$$

Тогда дисперсия любого испытания с номером n равна:

$$D\mu_n = M\mu_n^2 - (M\mu_n)^2 = p - p^2 = pq.$$

Значит

$$D\mu = D(\mu_1 + \ldots + \mu_N) = Npq. \tag{6.45}$$

С учетом (6.29) получаем

$$D\mu = \sigma^2, \tag{6.46}$$

где σ — *среднее квадратичное отклонение.*

Вычислим дисперсию в распределении Пуассона

$$\begin{aligned} D\mu = M\mu^2 - (M\mu)^2 &= \sum_n n^2 \frac{\lambda^n}{n!} e^{-\lambda} - \lambda^2 = \\ &= \lambda \sum_n n \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda} - \lambda^2 = \lambda \sum_k (k+1) \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 = \\ &= \lambda(\lambda + 1) - \lambda^2 = \lambda. \end{aligned} \tag{6.47}$$

Закон больших чисел, сформулированный в виде теоремы Бернулли (6.4) представляет собой общий принцип, в силу которого совместное действие случайных факторов приводит при весьма общих условиях к результату, почти не зависящему от случая. Теперь мы можем сформулировать его в виде теоремы Чебышева, доказанной в работе «О средних величинах» (1867) в следующем виде:

Для попарно независимых случайных величин $\xi_1, \ldots, \xi_N$ для любого $\varepsilon > 0$ и при достаточно большом числе испытаний N ($N \to \infty$) имеет место неравенство:

$$P\left\{\left|\frac{\xi_1 + \ldots + \xi_N}{N} - \frac{M\xi_1 + \ldots + M\xi_N}{N}\right| > \varepsilon\right\} \to 0. \tag{6.48}$$

6.5 Непрерывно распределенные величины

До сих пор мы рассматривали случайные величины с дискретной последовательностью (спектром) значений. Если перейти к случайным величинам с непрерывным спектром значений, то полученные закономерности (6.22), (6.32) и (6.43) нужно переписать в следующем виде:

$$\sum_{n=1}^{N} P(\xi) = 1, \qquad \int_{-\infty}^{+\infty} \rho(x)\,dx = 1, \tag{6.49}$$

$$M\xi = \overline{x} = \sum_{n=1}^{N} x_n p_n, \qquad Mx = \overline{x} = \int_{-\infty}^{+\infty} x\rho(x)\,dx, \tag{6.50}$$

$$\overline{x^2} = \int_{-\infty}^{+\infty} x^2 \rho(x)\,dx, \tag{6.51}$$

$$D\xi = M\xi^2 - (M\xi)^2, \qquad Dx = \overline{x^2} - \overline{x}^2. \tag{6.52}$$

где $\rho(x)$ — *плотность вероятности* (плотность распределения вероятностей) случайной величины.

Локальная предельная теорема Муавра–Лапласа (6.29):

$$P_N(\xi) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\xi - M\xi)^2}{2\sigma^2}}$$

при переходе непрерывно распределённым случайным величинам превращается в *нормальное распределение* (плотность вероятности случайной величины) следующего вида:

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\overline{x})^2}{2\sigma^2}} \tag{6.53}$$

График функции $\rho(x)$ имеет вид симметричного относительно $\overline{x}$ купола, с уходящими на бесконечность «хвостами», тем более острого, чем меньше значение σ. Причем в силу (4.84):

$$\int\limits_{-\infty}^{+\infty} \rho(x)\,dx = \int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\overline{x})^2}{2\sigma^2}}\,dx = 1. \tag{6.54}$$

Нормальное распределение служит хорошим приближением, когда рассматриваемая случайная величина представляет собой сумму большого числа независимых случайных величин, максимальная из которых мала по сравнению со всей суммой.

Допустим теперь, что наша случайная величина x является аргументом некоторой функции $f(x)$. В этом случае мы можем вычислить среднее значение нашей функции:

$$\overline{f} = \int\limits_{-\infty}^{+\infty} f(x)\rho(x)\,dx. \tag{6.55}$$

Плотности вероятностей впервые появились у К. Гаусса в «Теории движения небесных тел» (1809).

Часть II

Анализ систем

В первой части мы не определяли понятия объект и система, понимая под ними некую математическую абстракцию. Но при рассмотрении материального мира требуется дополнительно определить как эти, так и несколько новых понятий.

Объект (лат. objectum — предмет) — предмет (явление, процесс) наблюдаемый субъектом и существующий независимо от сознания субъекта (в реальной действительности).

Субъект (лат. subjectus — лежащий внизу, находящийся в основе) — носитель действия или наблюдатель (существо, наделённое волей), который познает, мыслит и действует, в отличие от объекта, на которого направлены мысль и действие.

Система (от греч. *σύστημα* — составленный) — множество объектов, находящихся в отношениях и связях друг с другом, образующих определенную целостность, единство.

Модель (франц. modele, от лат. modulus — мера, образец) — некоторый материальный или мысленно представляемый объект или явление, замещающий оригинальный объект или явление (прототип), с сохранением некоторых важных его свойств и опусканием несущественных свойств, в которых модель может отличаться от прототипа.

Математическая модель — приближённое описание какого-либо класса явлений внешнего мира, выраженное при помощи математической символики. Анализ математической модели позволяет проникнуть в сущность изучаемых явлений.

Таким образом, при наблюдении материального мира субъект выделяет в нём некую систему, составленную из взаимодействующих объектов, и, так как природа бесконечно сложна, при помощи математической символики составляет приближённое описание этой системы.

В первую очередь субъект стремится учесть основные решающие факторы, определяющие те именно черты процесса, которые его интересуют. Он заменяет бесконечно сложное природное явление некой моделью, которая, по его мнению, позволит адекватно описать и исследовать интересующий его процесс. Ответ на вопрос о том, как далеко можно идти в этом направлении и все же получить удовлетворительные результаты, может дать, в конечном счете, только опыт. Как подчеркивали Л. И. Мандельштам и А. А. Андронов, выбор такой идеализации очень часто далеко не тривиален и требует глубокого понимания сущности происходящих явлений. Общеизвестно, что когда удается подобрать удачную математическую модель для описания какого-либо реального процесса, то полученное описание в свою очередь позволяет по-новому взглянуть на изучаемое явление, обратить внимание на те стороны процесса, которые до этого находились в тени.

Во второй части книги описаны три физических системы: механическая, электродинамическая и термодинамическая; а также экономическая система. Однако описание этих систем не являлось для автора самоцелью. Цель заключалась в демонстрации заинтересованному читателю, что «экспериментально открытые физические законы» из казалось бы совершенно

разных разделов физики имеют общую природу. Все эти законы могут быть выведены из геометрических свойств пространства и только из них. Более того, между этими разделами существуют прямые аналогии: законы различных физических дисциплин имеют одинаковый аналитический вид, а входящие в них функции имеют аналогии с функциями из другого раздела.

Однако самое удивительное заключается в том, что все сказанное справедливо и для экономической системы. Автор исходит из гипотезы, что поведение различных систем, как физических так и, например, экономических, основано на общих принципах, в противоположность существующему мнению, что экономическая система слишком сложна и обладает специфическими чертами, чтобы ее описывать с применением математического аппарата естественных наук.

Для подтверждения этой гипотезы автор попытался применить используемые в теоретической физике интегро-дифференциальные уравнения для описания соответствующих экономико-математических моделей и интерпретировать полученные решения таким образом, чтобы они не противоречили известным экономическим законам. По мнению автора при таком подходе не только подтверждается гипотеза общности законов, но и значительно расширяются наши представления о функционировании экономической системы, так как математический аппарат и способы решения различных задач в естественных науках значительно лучше разработаны, чем в науках об обществе.

Насколько эта попытка удалась — судить читателю.

Глава 7

Механическая система

7.1 Механика Гамильтона и Лагранжа

Механика (греч. $\mu\eta\chi\alpha\nu\iota\kappa\grave{\eta}$ — искусство построения машин) — раздел физики, изучающий взаимодействие объектов, имеющих массу, в силовом поле. В этом случае соответствующую динамическую систему называют *механической системой*.

Таким образом, основной величиной, характеризующей свойства вещества в изучаемой системе, является масса, а в качестве векторного поля изучаемой системы выступает поле механических сил.

Рассмотрим n-мерное риманово многообразие V с выбранными локальными координатами $q_1, \dots, q_n$ и метрикой (3.93)

$$T = \sum_{i,j=1}^{n} \alpha_{ij}(q_1, \dots, q_n)\, \dot{q}_i \dot{q}_j \overset{(2.75)}{=} (\alpha\dot{\mathbf{q}}, \dot{\mathbf{q}}) = (\dot{\mathbf{q}}, \alpha\dot{\mathbf{q}}), \tag{7.1}$$

где $\alpha(\mathbf{q})$ — симметрическая матрица ($\alpha_{ij} = \alpha_{ji}$),

на котором задана дифференцируемая функция координат $U(q_1, \dots, q_n)$ или $U\colon V \to \mathbb{R}$.

Тогда касательное расслоение TV есть $2n$-мерное симплектическое многообразие с локальными координатами $q_1, \dots, q_n, dq_1, \dots, dq_n$ или $\mathbf{q}, d\mathbf{q}$, а кокасательное расслоение T^*V есть $2n$-мерное симплектическое многообразие с локальными координатами $p_1, \dots, p_n, q_1, \dots, q_n$ или $\mathbf{p}, \mathbf{q}$.

На кокасательном расслоении можно задать произвольную дифференцируемую *функцию Гамильтона* $E = H(\mathbf{p}, \mathbf{q})$ или $H\colon T^*V \to \mathbb{R}$, которая обязательно будет первым интегралом $E = const$ фазового потока с векторным полем (5.39)

$$\dot{\mathbf{p}} = -\frac{\partial E}{\partial \mathbf{q}}, \quad \dot{\mathbf{q}} = \frac{\partial E}{\partial \mathbf{p}} \tag{7.2}$$

и полным дифференциалом $dE = \dot{\mathbf{q}}\, d\mathbf{p} - \dot{\mathbf{p}}\, d\mathbf{q}$.

При помощи преобразования Лежандра (5.45) $L(\mathbf{q},\dot{\mathbf{q}}) = \mathbf{p}\cdot\dot{\mathbf{q}} - H(\mathbf{p},\mathbf{q})$ мы можем ввести *функцию Лагранжа* $L\colon TV \to \mathbb{R}$ на касательном расслоении, для которой справедливы уравнения (5.46), (5.47)

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}}, \quad \dot{\mathbf{p}} = \frac{\partial L}{\partial \mathbf{q}} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right), \quad dL = \mathbf{p}\, d\dot{\mathbf{q}} + \dot{\mathbf{p}}\, d\mathbf{q}. \tag{7.3}$$

Фазовый поток с векторным полем (7.2) называют называют *механикой Гамильтона*, а поток с полем (7.3) — *механикой Лагранжа*. Соответственно, дифференциальные уравнения динамической системы называются в гамильтоновой механике уравнениями Гамильтона, в лагранжевой механике — уравнениями Лагранжа.

Если теперь определить функцию Лагранжа как

$$L(\mathbf{q},\dot{\mathbf{q}}) = T(\mathbf{q},\dot{\mathbf{q}}) - U(\mathbf{q}), \tag{7.4}$$

тогда получаем

$$\mathbf{p} = \frac{\partial T}{\partial \dot{\mathbf{q}}}, \quad \dot{\mathbf{p}} \overset{(7.3)}{=} -\operatorname{grad} U; \tag{7.5}$$

$$E \overset{(5.45)}{=} \frac{\partial T}{\partial \dot{\mathbf{q}}}\,\dot{\mathbf{q}} - L \overset{(3.149)}{=} 2T - L = T + U. \tag{7.6}$$

При этом:

- V называют *конфигурационным пространством* (примеры конфигурационных пространств рассматриваются в § 3.4);
- $\mathbf{q} = (q_1,\dots,q_n)$ называют *обобщёнными координатами*;
- $\mathbf{p} = \frac{\partial T}{\partial \dot{\mathbf{q}}} = (p_1,\dots,p_n)$ называют *обобщёнными импульсами*;
- $\mathbf{q}(t)$ называют *траекторией движения*;
- $\dot{\mathbf{q}} = \frac{\partial L}{\partial \mathbf{p}} = (\dot{q}_1,\dots,\dot{q}_n)$ называют *обобщёнными скоростями*;
- $\mathbf{F} = \dot{\mathbf{p}} = \frac{\partial L}{\partial \mathbf{q}} = (\dot{p}_1,\dots,\dot{p}_n)$ называют *обобщёнными силами*;
- $T(\mathbf{q},\dot{\mathbf{q}})$ называют *кинетической энергией*;
- $U(\mathbf{q})$ называют *потенциальной энергией*;
- $E(\mathbf{p},\mathbf{q}) = T + U = const$ называют *полной энергией*.

Механическая система, где обобщённые силы $\dot{\mathbf{p}} = \mathbf{F}$ вычисляются как градиент потенциальной энергии со знаком минус (7.5), назвается *потенциальной* или *консервативной системой*, обобщённые силы называются *потенциальными* или *консервативными силами*.

Интеграл силы $\mathbf{F}$ по кривой C называется *работа силы* A

$$A = \int\limits_C \mathbf{F}\, d\mathbf{q}. \tag{7.7}$$

Следует отметить, что обобщённые координаты $\mathbf{q}$ и обобщённые силы $\mathbf{F}$ представляют собой достаточно произвольные переменные и могут не иметь размерности длины и силы. Однако произведение $\mathbf{F}\, d\mathbf{q}$ всегда имеет размерность работы. Такие пары называют *сопряжёнными переменными* или *взаимными величинами*.

Если $\mathbf{F}$ — потенциальная сила, то 1-форма $dU = -\mathbf{F}\, d\mathbf{q}$ есть полный дифференциал, и работа силы не зависит от формы кривой C, а определяется начальной и конечной точками этой кривой:

$$\int\limits_1^2 \mathbf{F}\, d\mathbf{q} \overset{(4.52)}{=} U_1 - U_2 \quad \text{или} \quad \oint\limits_C \mathbf{F}\, d\mathbf{q} \overset{(4.51)}{=} 0. \tag{7.8}$$

Часть пространства (ограниченная или неограниченная), в каждой точке которой действует определённая по численной величине и направлению сила называется *силовое поле*. Если силовое поле зависит только от координат точки $U(\mathbf{q})$, то оно называется *стационарным*, а если ещё и от времени $U(t, \mathbf{q})$ — нестационарным. Если силовое поле не зависит ни от координат, ни от времени $U = const$, то оно называется *однородным*. Если элементарная работа сил поля равна полному дифференциалу поля: $dU = -\mathbf{F}\, d\mathbf{q}$, то силовое поле называется *потенциальным полем* или *потенциалом*.

Внешняя среда — окружение механической системы, оказывающее на неё воздействие. Если внешняя среда не оказывает воздействия, то механическая система называется *изолированной* (*замкнутой*). Воздействие внешней среды на механическую систему определяется числом внешних степеней свободы.

Сама механическая система также представляет собой среду.

Дискретная среда — механическая система, обладающая счётным множеством внутренних степеней свободы. Её движение в трёхмерном евклидовом пространстве описывается координатами и импульсами отдельных *материальных точек* — объектов, размерами которых можно в данных условиях пренебречь.

Допустим, механическая система состоит из k материальных точек.

Тогда:

- пространство $V = \underbrace{E^3 \times \ldots \times E^3}_{k}$ соответствует $n = 3k$-мерному конфигурационному пространству;
- координаты $x^1, \ldots, x^k, y^1, \ldots, y^k, z^1, \ldots, z^k$ соответствуют $n = 3k$ обобщённым координатам $\mathbf{q}$;
- импульсы $p_x^1, \ldots, p_x^k, p_y^1, \ldots, p_y^k, p_z^1, \ldots, p_z^k$ соответствуют $n = 3k$ обобщённым импульсам $\mathbf{p}$;

- скорости $\dot{x}^1, \ldots, \dot{x}^k, \dot{y}^1, \ldots, \dot{y}^k, \dot{z}^1, \ldots, \dot{z}^k$ соответствуют $n = 3k$ обобщённым скоростям $\dot{\mathbf{q}}$;
- силы $\dot{p}_x^1, \ldots, \dot{p}_x^k, \dot{p}_y^1, \ldots, \dot{p}_y^k, \dot{p}_z^1, \ldots, \dot{p}_z^k$ соответствуют $n = 3k$ обобщённым силам $\dot{\mathbf{p}}$.

Сплошная среда — механическая система, обладающая несчётным множеством внутренних степеней свободы. Её движение в трёхмерном евклидовом пространстве, описывается координатами и скоростями поля плотности. *Объёмная плотность* ρ в трёхмерном евклидовом пространстве является *скаляром* (представляет собой скалярное поле). *Поверхностная плотность* ϱ на двумерной евклидовой поверхности является *вектором* (представляет собой векторное поле), направленным по нормали к рассматриваемой точке поверхности. В зависимости от задач, к этим полям могут добавляться поля других физических величин.

Переход от сплошной к дискретной среде осуществляют при помощи δ-функции Дирака:

$$\rho(\mathbf{r}) \overset{(4.110)}{=} \sum_{i=1}^{k} x_i \, \delta(\mathbf{r} - \mathbf{r}_i), \tag{7.9}$$

где $x_1, \ldots, x_k$ — величины, характеризующие свойства, связанные с изучаемыми объектами.

Если плотность сплошной среды не зависит явно от времени, то такая сплошная среда называется *несжимаемой*.

Механика сплошных сред делится на механику твёрдого тела, гидродинамику, газодинамику. Каждая из этих дисциплин также делится на более узкие разделы. Так, механика твёрдого тела делится на теорию упругости, теорию пластичности, теорию трещин и так далее.

Дискретная и сплошная среда — часто и успешно используемые в физике модели для систем, соответственно, с небольшим числом объектов (т.е. степеней свободы) или с очень большим множеством объектов. Однако с точки зрения математической строгости следует помнить об одной неточности: все реальные системы обладают пусть большим, но конечным множеством степеней свободы. Дискретная же среда может обладать бесконечным, хотя и счётным, множеством степеней свободы, а сплошная среда обладает не просто бесконечным, а несчётным множеством степеней свободы.

Как известно (см. стр. 204) через каждую точку фазового пространства проходит только одна фазовая траектория. В случае механической системы этот общий формальный математический результат решения системы обыкновенных дифференциальных уравнений (динамической системы) установлен экспериментально, т.е. является законом природы, и называется *принцип детерминированности Ньютона*. В случае механики Гамильтона это означает, что если знать начальное состояние механической системы (начальные значения $\mathbf{p}$ и $\mathbf{q}$), то можно однозначно определить всё её предыдущее и дальнейшее движение. Никаких дополнительных сведений, например, знания начальных значений $\dot{\mathbf{p}}$, для этого не нужно. В случае механики

Лагранжа это означает, что если знать начальное состояние механической системы (начальные значения $\dot{\mathbf{q}}$ и $\mathbf{q}$), то можно однозначно определить всё её предыдущее и дальнейшее движение. Никаких дополнительных сведений, например, знания начальных значений $\mathbf{p}$, для этого не нужно.

Мы не удивляемся этому факту, так как узнаём его очень рано.

Обратим внимание ещё раз. Из геометрического свойства фазового пространства — через каждую точку фазового пространства проходит только одна фазовая траектория — не следует принцип детерминированности Ньютона, но из экспериментального установления этого принципа, как закона природы, следует, что механическая система есть динамическая система, заданная на фазовом пространстве $T^*V = (\mathbf{p}, \mathbf{q})$.

Если мы рассмотрим расширенное фазовое пространство с координатами $t, \mathbf{p}, \mathbf{q}$, то линии ротора формы (5.61) $\omega^1 = dS = L\,dt = \mathbf{p}\,d\mathbf{q} - H\,dt$, которую называют интегральным инвариантом Пуанкаре–Картана задают траектории фазового потока в расширенном пространстве, т.е. интегральные кривые канонических уравнений Гамильтона (7.2). Интеграл (5.75) $\Phi(C) = \int_0^\tau L\,dt$ вычисляет функционал $\Phi(C)$ по наикратчайшей дуге C (геодезической), пройденной динамической системой за время на отрезке $0 \leqslant t \leqslant \tau$, который называется действие.

Экспериментально установлено, что переход механической системы из одного состояния в другое (между двумя точками расширенного фазового пространства) всегда происходит по геодезической линии. При этом проекция геодезической на конфигурационное пространство с обобщёнными координатами $\mathbf{q}$ называется *траекторией движения* системы. Обязательное движение механической системы по геодезической линии называется в механике *принципом наименьшего действия Гамильтона*.

С точки зрения принципа наименьшего действия природа достигает своей цели прямейшим путем, следовательно, с наименьшей затратой средств. Поэтому более удачен был бы термин «принцип наименьшей затраты средств, при наибольшем действии». Но после того как термин «действие» был санкционирован Гельмгольцем и Планком, всякая замена его другим термином была бы бесперспективной.

Простейший пример наикратчайшей траектории — распространение луча света по прямой, но сразу следует отметить, что наикратчайшая траектория, это не обязательно прямая линия. Рассмотрим *брахистохрону* (от греч. $\beta\rho\acute{\alpha}\chi\iota\sigma\tau\sigma\varsigma$ — кратчайший и $\chi\rho\acute{o}\nu\sigma\varsigma$ — время) — кривую скорейшего спуска, например, при стекании воды с холма. Задача о её нахождении была поставлена в 1696 году Иоганном Бернулли. Одно из решений дано в знаменитом учебнике Г. Лопиталя. В однородном поле тяжести брахистохрона представляет собой перевёрнутую циклоиду (от греч. $\kappa\upsilon\kappa\lambda o\varepsilon\iota\delta\acute{\eta}\varsigma$ — круглый) — кривую, которую описывает относительно земли точка на ободе катящегося колеса. Верхняя точка траектории есть точка возврата циклоиды (точка касания земли).

Экстремаль функционала (5.78) $\Phi(C) = \int_{\mathbf{q}_0}^{\mathbf{q}_\tau} \mathbf{p}d\mathbf{q} = \int_0^\tau 2T\,dt$ или «укороченного действия» называют в механике *принципом наименьшего действия Мопертюи* (Эйлера–Лагранжа–Якоби).

Впервые в несколько туманной формулировке этот принцип был сформулирован Мопертюи в 1746 г. Мопертюи пришёл к этому принципу из ощущения, что совершенство вселенной требует определенной экономии в природе и противоречит любым бесполезным расходам энергии. Естественное движение должно быть таким, чтобы сделать некоторую величину минимальной. Нужно было только найти эту величину, что он и продолжал делать. Она являлась произведением продолжительности (времени) движения в пределах системы на удвоенную величину, которую мы теперь называем кинетической энергией системы.

Хотя Мопертюи указывали ещё на письмо Лейбница, относящееся к 1707 г. (оригинал этого письма не сохранился), он ревностно защищал свой приоритет, не останавливаясь даже перед использованием своей власти в качестве президента Берлинской академии.

Исторически принцип Мопертюи был сформулирован раньше принципа наименьшего действия, и в отличие от последнего ограничивается более узким вопросом об определении лишь самой траектории. Более определенную математическую форму принципу придали только Эйлер и, в особенности, Лагранж в 1760 г. Эйлер (в «Reflexions sur quelques loix generales de la nature», 1748) принимает принцип наименьшего количества действия, называя его «усилием». Его выражение соответствует тому, что мы теперь назвали бы потенциальной энергией, так что его утверждение наименьшего количества действия в статике эквивалентно принципу, что система тел в покое примет конфигурацию, которая минимизирует полную потенциальную энергию.

7.2 Механика Ньютона

Ньтоновой механикой называется частный случай лагранжевой механики, когда вместо конфигурационного пространства V рассматривают привычное нам трёхмерное евклидово пространство E^3.

Например, самой простой механической системой является *материальная точка* — объект очень маленьких размеров (точнее, размерами которого можно пренебречь) с массой m, в евклидовом трёхмерном пространстве E^3 с координатами x, y, z, на котором задана потенциальная энергия — функция кординат $U(x, y, z)$.

Движение материальной точки в трехмерном пространстве понимается как перемещение точки с координатами $\mathbf{r} = (x, y, z)$ вдоль какой-то линии $\mathbf{r}(t)$, которая называется *траекторией* движения системы, со скоростью $\dot{\mathbf{r}} = (\dot{x}, \dot{y}, \dot{z}) = \mathbf{v} = (v_x, v_y, v_z)$. *Количество движения* или *импульс* понимается как произведение массы на скорость

$$\mathbf{p} = (p_x, p_y, p_z) = m\dot{\mathbf{r}} = m\mathbf{v}. \tag{7.10}$$

Трёхмерное евклидово пространство механической системы, состоящей из одной материальной точки соответствует конфигурационному пространству лагранжевой системы, координаты материальной точки x, y, z соответ-

ствуют обобщённым координатам, импульсы p_x, p_y, p_z соответствуют обобщённым импульсам.

Согласно принципу детерминированности Ньютона зададим начальное состояние механической системы в момент времени t_0:

$$\mathbf{p}_0 = m\mathbf{v}_0 = (0,0,0), \quad \mathbf{r}_0 = (x_0, y_0, z_0).$$

Тогда мы можем записать

$$T \overset{(7.5)}{=} \int\limits_{\dot{\mathbf{q}}_0}^{\dot{\mathbf{q}}} \mathbf{p}\, d\dot{\mathbf{q}} \overset{(7.10)}{=} \int\limits_{\mathbf{0}}^{\mathbf{v}} m\mathbf{v}\, d\mathbf{v} = \frac{m\mathbf{v}^2}{2} = \frac{mv^2}{2}, \quad \text{где } v = |\mathbf{v}|. \tag{7.11}$$

Исторически уравнения движения в ньютоновой механике также выписываются при помощи уравнений Ньютона или так называемых трёх законов Ньютона. Помимо этих законов Исаак Ньютон («Математические начала натуральной философии», 1686) также дал строгие определения таких физических понятий, как количество движения или импульс (не вполне ясно использованное у Декарта) и сила. Он ввёл в физику (которую до этого называли «натуральной философией») понятие массы как меры инерции и, одновременно, гравитационных свойств (ранее пользовались понятием *вес*). При этом, надо сказать, что законы Ньютона были впервые сформулированы не Ньютоном, хотя и называются его именем.

В своих построениях механической системы Ньютон исходил не из интегральных (H и L), а из дифференциальных ($\dot{\mathbf{p}}$ и $\dot{\mathbf{q}}$) функций, т.е. задавал векторное поле фазового пространства.

Выражение $\mathbf{F} = \dot{\mathbf{p}}$, которое мы ввели в лагранжевой механике как определение обобщённых сил, в ньютоновой механике носит название второго закона Ньютона

$$\mathbf{F} = m\dot{\mathbf{v}}, \tag{7.12}$$

где $\dot{\mathbf{v}} = \ddot{\mathbf{r}}$ — *ускорение* материальной точки.

Второй закон Ньютона: в инерциальной системе отсчета ускорение $\ddot{\mathbf{r}}$, которое получает материальная точка, прямо пропорционально равнодействующей всех приложенных к ней сил $\mathbf{F}$ и обратно пропорционально её массе m.

Второй закон Ньютона может выполняться только при выполнении первого закона Ньютона.

Первый закон Ньютона или *закон инерции*: существуют такие, называемые *инерциальными* (от лат. inertis — бездеятельный), системы отсчёта, относительно которых материальная точка при отсутствии внешних воздействий сохраняет состояние покоя или равномерного прямолинейного движения.

Инерциальные системы обладают двумя свойствами:

1. Все системы координат, движущиеся относительно инерциальной системы равномерно и прямолинейно, также инерциальные, т.е. все инерциальные системы отсчёта равноправны. Время одинаково во всех системах отсчёта — «абсолютное время». Пусть $\mathbf{r}'$ и $\mathbf{r}$ радиус-векторы одной точки в двух инерциальных системах, тогда скорости $\dot{\mathbf{r}}'$ и $\dot{\mathbf{r}}$ этой точки в двух инерциальных системах, движущихся относительно друг друга со скоростью $\mathbf{v}$, связаны *преобразованием Галилея* (термин предложен Ф. Франком (1909)):

$$\dot{\mathbf{r}} = \dot{\mathbf{r}}' + \mathbf{v}. \tag{7.13}$$

2. *Постулат эквивалентности* — все законы природы во все моменты времени одинаковы (инвариантны) во всех инерциальных системах, т.е. мы не можем посредством физических экспериментов отличить «неподвижную систему» от «равномерно движущейся», а можем лишь констатировать, что все системы эквивалентны и движутся относительно друг друга. Это значит, что законы физики в них проявляются одинаково, и записи этих законов имеют одинаковую форму в разных инерциальных системах отсчёта.

Действительно, второй закон Ньютона может выполняться только при выполнении первого закона Ньютона. Допустим, материальная точка в произвольной системе с координатами $\mathbf{r}'$ движется равномерно и прямолинейно относительно инерциальной системы $\mathbf{r}$ со скоростью $\mathbf{v} = const$. Тогда $\dot{\mathbf{r}} = \dot{\mathbf{r}}' + \mathbf{v}$ и $\dot{\mathbf{v}} = 0$. Следовательно, второй закон Ньютона вполняется одновременно в двух этих системах

$$\mathbf{F} = m\ddot{\mathbf{r}} = m(\ddot{\mathbf{r}}' + \dot{\mathbf{v}}) = m\ddot{\mathbf{r}}'.$$

Обратно, система, в которой справедлив второй закон Ньютона, является инерциальной. Например, система, связанная с телом, вращающимся с угловой скоростью ω относительно инерциальной системы, не является инерциальной и, очевидно, не удовлетворяет второму закону Ньютона, так как к нему нужно добавить член, учитывающий влияние вращения (3.113): $\dot{\mathbf{r}} = \omega \times \mathbf{r}$.

Факт (или гипотеза) существования инерциальных систем называется *принципом относительности Галилея* (1632), который после множества опытов заключил, что для движения свободного тела с постоянной скоростью не нужно какой-либо внешней причины. Этот факт на первый взгляд казался удивительным, хотя основания у него точно такие же, как и для очевидного всем утверждения, что для сохранения неподвижным объектом состояния покоя, какой-либо внешней причины также не нужно. Надо отметить, что Галилей допускал свободное движение не только по прямой, но и по окружности (видимо, из астрономических соображений), хотя такая система не является инерциальной.

Так как скорость относительного движения инерциальных систем $\mathbf{v}$ выбирается произвольно, то предположение о существовании хотя бы одной

инерциальной системы отсчёта приводит к выводу о существовании *бесконечного множества* таких систем, движущихся друг относительно друга со всевозможными постоянными скоростями.

Следует отметить, что понятие инерциальной системы отсчёта — абстрактная модель, то есть некий идеальный объект, рассматриваемый вместо реального объекта (примерами абстрактной модели служат абсолютно твердое тело или нерастяжимая невесомая нить). Реальные системы отсчёта всегда связаны с каким-либо объектом или объектами и соответствие реально наблюдаемого движения тел в таких системах с результатами расчётов будет неполным.

Третий закон Ньютона объясняет взаимодействие двух объектов с индексами i и j: тела действуют друг на друга с силами, имеющими одинаковую природу, направленными вдоль одной и той же прямой, равными по модулю и противоположными по направлению

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji} \quad \text{или} \quad \mathbf{F}_{ij} + \mathbf{F}_{ji} = 0, \quad \mathbf{F}_{ii} = 0. \tag{7.14}$$

Рассмотрим случай дискретной среды. Допустим, что механическая система состоит из k материальных точек с массами $m_1, \ldots, m_k$ и радиус-векторами $\mathbf{r}_1, \ldots, \mathbf{r}_k$. Также в пространстве задана потенциальная энергия $U(\mathbf{r}_1, \ldots, \mathbf{r}_k)$.

Тогда для вычисления величин применяется свойство аддитивности.

Аддитивность (лат. additivus от additio — прибавляю) или *экстенсивность* — свойство величин по отношению к сложению, состоящее в том, что значение величины, соответствующее целому объекту, равно сумме значений величин, соответствующих его частям, в некотором классе возможных разбиений объекта на части.

Пример 1. Аддитивность площади (объёма) в математике означает, что площадь (объём) равна сумме площадей её частей, если этих частей конечное число.

Пример 2. *Аддитивность* (*экстенсивность*) *величины* в физике означает, что величина чего-то равна сумме величин составных частей. Например, к таким физическим величинам относятся: энергия (полная, кинетическая и потенциальная), количество вещества его плотность и масса, импульс, сила, в отличие от *интенсивных величин*, например, таких как температура.

Средний радиус-вектор $\mathbf{r}$ рассматриваемой системы определяет точку, называемую *центр масс системы* (средне взвешенное значение радиус-векторов материальных точек), и вычисляется с учётом общей массы m всех точек системы:

$$\mathbf{r} = \frac{1}{m} \sum_{i=1}^{k} m_i \mathbf{r}_i, \quad m = \sum_{i=1}^{k} m_i. \tag{7.15}$$

Пусть теперь $\mathbf{F}_i^{\text{внеш}}$ — внешняя сила, действующая на i-ю точку, тогда

$$\mathbf{p}_i \overset{(7.10)}{=} m_i \mathbf{v}_i = m_i \dot{\mathbf{r}}_i, \tag{7.16}$$

$$\dot{\mathbf{p}}_i = m_i \ddot{\mathbf{r}}_i \overset{(7.12)}{=} \mathbf{F}_i = \mathbf{F}_i^{\text{внеш}} + \sum_{j=1}^{k} \mathbf{F}_{ji}. \tag{7.17}$$

Последовательно запишем вычисления

$$\mathbf{v} = \dot{\mathbf{r}} \overset{(7.15)}{=} \frac{1}{m} \sum_{i=1}^{k} m_i \dot{\mathbf{r}}_i, \tag{7.18}$$

$$\dot{\mathbf{v}} = \ddot{\mathbf{r}} = \frac{1}{m} \sum_{i=1}^{k} m_i \ddot{\mathbf{r}}_i, \tag{7.19}$$

$$\mathbf{p} = \sum_{i=1}^{k} \mathbf{p}_i = \sum_{i=1}^{k} m_i \dot{\mathbf{r}}_i = m\mathbf{v}, \tag{7.20}$$

$$\dot{\mathbf{p}} = \sum_{i=1}^{k} \dot{\mathbf{p}}_i = \sum_{i=1}^{k} m_i \ddot{\mathbf{r}}_i = m\ddot{\mathbf{r}} = \sum_{i=1}^{k} \mathbf{F}_i^{\text{внеш}} + \sum_{i,j=1}^{k} \mathbf{F}_{ji}. \tag{7.21}$$

С учётом (7.14) $\sum_{i,j=1}^{k} \mathbf{F}_{ji} = 0$, получаем

$$\dot{\mathbf{p}} \overset{(7.21)}{=} m\ddot{\mathbf{r}} = \sum_{i=1}^{k} \mathbf{F}_i^{\text{внеш}} = \mathbf{F}^{\text{внеш}}. \tag{7.22}$$

Согласно принципу детерминированности Ньютона зададим начальное состояние механической системы в момент времени t_0

$$\mathbf{v}_0^i = (0,0,0), \quad \mathbf{r}_0^i = (x_0^i, y_0^i, z_0^i), \quad i = 1, \ldots, k.$$

Запишем выражение для *полной кинетической энергии* системы:

$$T \overset{(7.5)}{=} \sum_{i=1}^{k} \int_{\mathbf{0}}^{\mathbf{v}} m_i \mathbf{v}_i \, d\mathbf{v}_i = \sum_{i=1}^{k} \frac{m_i \mathbf{v}_i^2}{2}. \tag{7.23}$$

Получим

$$L \overset{(7.4)}{=} \sum_{i=1}^{k} \frac{m_i \mathbf{v}_i^2}{2} - U(\mathbf{r}_1, \ldots, \mathbf{r}_k), \quad \text{откуда} \quad \dot{\mathbf{p}}_i \overset{(7.3)}{=} \frac{\partial L}{\partial \mathbf{r}_i} = -\frac{\partial U}{\partial \mathbf{r}_i}. \tag{7.24}$$

Запишем выражение для *полной потенциальной энергии* системы:

$$U = -\sum_{i=1}^{n} \int_{\mathbf{r}_0}^{\mathbf{r}} \dot{\mathbf{p}}_i \, d\mathbf{r}_i \overset{(7.17)}{=} -\sum_{i=1}^{n} \int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{F}_i^{\text{внеш}} \, d\mathbf{r}_i - \frac{1}{2} \sum_{i \neq j} \int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{F}_{ji} \, d\mathbf{r}_i =$$

$$= \sum_{i=1}^{n} U_i + \frac{1}{2} \sum_{i \neq j} U_{ij}.$$

Внутренние силы взаимодействия $\mathbf{F}_{ji}$ мы считаем консервативными.

Член $\frac{1}{2}\sum_{i\neq j} U_{ij}$ называют внутренней потенциальной энергией системы. Она, вообще говоря, отлична от нуля и, что весьма важно, может изменяться вместе с изменением самой системы с течением времени.

Для частного класса систем — твёрдых тел — внутренний потенциал есть величина постоянная. Формально, твёрдое тело можно определить как систему материальных точек, расстояния между которыми постоянны и не могут изменяться со временем. Так как полный потенциал во всех случаях есть величина, определённая лишь с точностью до некой константы интегрирования, то постоянный внутренний потенциал можно при исследовании движения системы совершенно не рассматривать.

Уравнения, полученные для механической системы из k точек, могут быть распространены на сплошную среду, например, на газ, жидкость и т.д. В этом случае материальная точка с массой m_i и радиус-вектором $\mathbf{r}_i$ заменяется на объёмную *плотность вещества* $\rho(\mathbf{r})$ в рассматриваемой точке пространства, с радиус-вектором $\mathbf{r}$. Вместо импульса частицы $\mathbf{p}_i = m_i\mathbf{v}_i$ в заданной точке пространства вычисляется *плотность потока*:

$$\mathbf{j} = \rho\,\dot{\mathbf{r}} = \rho\,\mathbf{v}, \tag{7.25}$$

где $\mathbf{v}$ — *скорость вещества* в данной точке пространства.

Действующая в рассматриваемой точке пространства сила понимается как *плотность силы* $\mathbf{f}$, а второй закон Ньютона (7.12) имеет вид

$$\mathbf{f} = \rho\,\dot{\mathbf{v}}. \tag{7.26}$$

Суммирование частиц заменяется интегрированием по объёму тела V так, будто тело является совокупностью дискретных частиц и мы получаем уравнения для непрерывной среды, аналогичные уравнениям для системы из k частиц

$$m = \iiint\limits_V \rho(\mathbf{r})\,dV \qquad \text{аналогично (7.15)} \qquad m = \sum_{i=1}^{k} m_i, \tag{7.27}$$

$$\mathbf{r} = \frac{1}{m}\iiint\limits_V \rho(\mathbf{r})\,\mathbf{r}\,dV \qquad \text{аналогично (7.15)} \qquad \mathbf{r} = \frac{1}{m}\sum_{i=1}^{k} m_i\mathbf{r}_i, \tag{7.28}$$

$$\mathbf{p} = \iiint\limits_V \mathbf{j}\,dV \qquad \text{аналогично (7.20)} \qquad \mathbf{p} = \sum_{i=1}^{k} m_i\dot{\mathbf{r}}_i. \tag{7.29}$$

7.3 Вращение механической системы

Момент количества движения, кинетический момент или *момент импульса* материальной точки относительно начала координат понимается как векторное произведение радиус-вектора на импульс, т.е. аксиальный или псевдовектор $\mathbf{L} = (L_x, L_y, L_z)$ (см. стр. 85):

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}. \tag{7.30}$$

Момент силы или *вращающий момент силы*, действующий на материальную точку относительно начала координат понимается как векторное произведение радиус-вектора на приложенную к точке силу:

$$\mathbf{N} = \mathbf{r} \times \mathbf{F}. \tag{7.31}$$

Таким образом, моменты импульса и силы — аксиальные векторы.

Если механическая система состоит из k материальных точек, то

$$\mathbf{L}_i = \mathbf{r}_i \times \mathbf{p}_i, \quad \mathbf{L} = \sum_{i=1}^{k} \mathbf{r}_i \times \mathbf{p}_i = \sum_{i=1}^{k} \mathbf{L}_i; \tag{7.32}$$

$$\mathbf{N}_i = \mathbf{r}_i \times \mathbf{F}_i, \quad \mathbf{N} = \sum_{i=1}^{k} \mathbf{r}_i \times \mathbf{F}_i = \sum_{i=1}^{k} \mathbf{N}_i. \tag{7.33}$$

Таким образом, моменты импульса и силы — экстенсивные величины, *зависящие от выбора начала координат.*

Запишем *уравнение моментов*:

$$\dot{\mathbf{L}}_i = \mathbf{N}_i, \quad \dot{\mathbf{L}} = \sum_{i=1}^{k} \mathbf{r}_i \times \mathbf{F}_i^{\text{внеш}} = \mathbf{N}^{\text{внеш}}. \tag{7.34}$$

Доказательство

$$\dot{\mathbf{L}}_i = \mathbf{r}_i \times m_i\ddot{\mathbf{r}}_i + \dot{\mathbf{r}}_i \times m_i\dot{\mathbf{r}}_i \overset{\text{стр. } 86}{=} \mathbf{r}_i \times m_i\ddot{\mathbf{r}}_i;$$

$$\mathbf{r}_i \times \mathbf{F}_{ji} + \mathbf{r}_j \times \mathbf{F}_{ij} \overset{(7.14)}{=} (\mathbf{r}_i - \mathbf{r}_j) \times \mathbf{F}_{ji} = \mathbf{r}_{ij} \times \mathbf{F}_{ji} = 0,$$

так как $\mathbf{r}_{ij} \parallel \mathbf{F}_{ji}$, значит $\sum\limits_{i,j=1}^{k} \mathbf{r}_i \times \mathbf{F}_{ji} = 0;$

$$\dot{\mathbf{L}} = \sum_{i=1}^{k} \mathbf{N}_i = \sum_{i=1}^{k} \mathbf{r}_i \times \mathbf{F}_i \overset{(7.17)}{=} \sum_{i=1}^{k} \mathbf{r}_i \times \mathbf{F}_i^{\text{внеш}} + \sum_{i,j=1}^{k} \mathbf{r}_i \times \mathbf{F}_{ji} = \sum_{i=1}^{k} \mathbf{r}_i \times \mathbf{F}_i^{\text{внеш}}.$$

Пусть $\mathbf{r}'_i = \mathbf{r}_i - \mathbf{r}$ — радиус-вектор, идущий из центра масс системы в i-ю материальную точку а $\dot{\mathbf{r}}'_i = \mathbf{v}'_i = \mathbf{v}_i - \mathbf{v}$ — скорость i-й материальной точки относительно центра масс. Тогда:

$$\sum_{i=1}^{k} m_i\mathbf{r}'_i = 0, \quad \sum_{i=1}^{k} m_i\dot{\mathbf{r}}'_i = \sum_{i=1}^{k} \mathbf{p}'_i = 0. \tag{7.35}$$

Доказательство

$$\mathbf{r} \overset{(7.15)}{=} \frac{1}{m}\sum_{i=1}^{k} m_i\mathbf{r}_i + \frac{1}{m}\sum_{i=1}^{k} m_i\mathbf{r}'_i = \mathbf{r} + \frac{1}{m}\sum_{i=1}^{k} m_i\mathbf{r}'_i,$$

$$\mathbf{p} \overset{(7.20)}{=} \sum_{i=1}^{k} m_i\dot{\mathbf{r}}_i + \sum_{i=1}^{k} m_i\dot{\mathbf{r}}'_i = \mathbf{p} + \sum_{i=1}^{k} m_i\dot{\mathbf{r}}'_i.$$

Кинетический момент с учётом центра масс записывается как

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} + \sum_{i=1}^{k} \mathbf{r}'_i \times \mathbf{p}'_i \tag{7.36}$$

Доказательство

$$\begin{aligned}
\mathbf{L} &= \sum_{i=1}^{k} \mathbf{r}_i \times \mathbf{p}_i = \\
&\overset{\text{стр. 86}}{=} \mathbf{r} \times \mathbf{p} + \mathbf{r} \times \sum_{i=1}^{k} \mathbf{p}'_i + \sum_{i=1}^{k} \mathbf{r}'_i \times \sum_{i=1}^{k} m_i \dot{\mathbf{r}}_i + \sum_{i=1}^{k} \mathbf{r}'_i \times \mathbf{p}'_i = \\
&\overset{\text{стр. 86}}{=} \mathbf{r} \times \mathbf{p} + \mathbf{r} \times \sum_{i=1}^{k} \mathbf{p}'_i + \sum_{i=1}^{k} m_i \mathbf{r}'_i \times \sum_{i=1}^{k} \dot{\mathbf{r}}_i + \sum_{i=1}^{k} \mathbf{r}'_i \times \mathbf{p}'_i = \\
&\overset{(7.35)}{=} \mathbf{r} \times \mathbf{p} + \sum_{i=1}^{k} \mathbf{r}'_i \times \mathbf{p}'_i.
\end{aligned}$$

Таким образом, кинетический момент системы относительно начала координат складывается из двух частей: из кинетического момента системы в предположении, что вся её масса сосредоточена в центре масс, и из кинетического момента, возникающего вследствие движения этой системы относительно центра масс.

Кинетическая энергия с учётом центра масс записывается как

$$T = \frac{m\mathbf{v}^2}{2} + \sum_{i=1}^{k} \frac{m_i \mathbf{v}'^2_i}{2} = T_{\text{перемещение}} + T_{\text{вращение}}. \tag{7.37}$$

Доказательство

$$\begin{aligned}
\sum_{i=1}^{k} \frac{m_i(\mathbf{v} + \mathbf{v}'_i)^2}{2} &= \sum_{i=1}^{k} \frac{m_i \mathbf{v}^2}{2} + \mathbf{v} \sum_{i=1}^{k} m_i \mathbf{v}'_i + \sum_{i=1}^{k} \frac{m_i \mathbf{v}'^2_i}{2} = \\
&\overset{(7.35)}{=} \frac{m\mathbf{v}^2}{2} + \sum_{i=1}^{k} \frac{m_i \mathbf{v}'^2_i}{2}.
\end{aligned}$$

Таким образом, кинетическая энергия системы, подобно кинетическому моменту, складывается из двух частей: из кинетической энергии, получающейся в предположении, что вся масса системы сосредоточена в её центре масс, и из кинетической энергии системы в её движении относительно центра масс. Этот вывод сформулирован в виде *теоремы Шаля*: произвольное перемещение твёрдого тела можно разбить на поступательное и вращательное.

Потенциальную энергию тоже часто удаётся разделить на две подобные части, из которых одна содержит только координаты, соответствующие поступательному движению $U(x, y, z)$, а другая — только угловые координаты $U(\varphi, \theta, \psi)$, где φ, θ, ψ — углы Эйлера.

В общем случае кинетический момент зависит от выбора начала координат, и только в случае, когда центр масс неподвижен относительно начала координат, кинетический момент не зависит от выбора этой точки. Поэтому при вычислениях всегда удобно взять систему координат, связанную с центром масс. В этом случае $\mathbf{r} \times \mathbf{p} = 0$ и $\mathbf{L} = \sum_{i=1}^{k} \mathbf{r}_i \times \mathbf{p}_i$. Тогда при помощи угловой скорости ω вращения системы вокруг центра масс, для которой справедливо уравнение вращения (3.113) $\dot{\mathbf{r}}_i = \omega \times \mathbf{r}_i$ мы можем записать выражение для момента количества движения как

$$\mathbf{L} = \mathbf{I}\,\omega = \sum_{i=1}^{k} m_i(\mathbf{r}_i \times (\omega \times \mathbf{r}_i)) \overset{\text{стр. } 86}{=} \sum_{i=1}^{k} m_i(\omega \mathbf{r}_i^2 - \mathbf{r}_i(\mathbf{r}_i \cdot \omega)). \tag{7.38}$$

или

$$\left\| \begin{matrix} L_x \\ L_y \\ L_z \end{matrix} \right\| = \left\| \begin{matrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{matrix} \right\| \left\| \begin{matrix} \omega_x \\ \omega_y \\ \omega_z \end{matrix} \right\|.$$

Симметрическая матрица $\mathbf{I}$ ($I_{xy} = I_{yx}$) называется *тензор инерции* системы (например, твёрдого тела). Диагональные элементы I_{xx}, I_{yy}, I_{zz} этой матрицы преобразования называются *осевыми*, а остальные I_{xy}, I_{xz}, I_{yz} — *центробежными моментами инерции*, например,

$$I_{xx} = \sum_{i=1}^{k} m_i(r_i^2 - x_i^2), \quad I_{xy} = -\sum_{i=1}^{k} m_i x_i y_i \tag{7.39}$$

Следовательно, вектор кинетического момента получается из угловой скорости посредством линейного преобразования.

Обратим внимание, что векторы $\mathbf{I}$ и ω суть два вектора различной физической природы и размерности, а не один и тот же вектор, выраженный в различных системах координат. Элементы матрицы оператора $\mathbf{I}$ также являются размерными величинами.

В общем случае мы можем записать следующее уравнение (3.114)

$$\left(\frac{d\mathbf{L}}{dt}\right) = \left(\frac{d\mathbf{L}}{dt}\right)_{\text{перемещение}} + (\omega \times \mathbf{L})_{\text{вращение}} = \mathbf{N}^{\text{внеш}}. \tag{7.40}$$

которое также может быть записано в виде так называемых *уравнений Эйлера* для движения твёрдого тела с одной неподвижной точкой:

$$\begin{aligned}
&\dot{L}_x + \omega_y L_z - \omega_z L_y = N_x^{\text{внеш}} \quad \text{или} \quad I_x\dot{\omega}_x - \omega_y\omega_z(I_y - I_z) = N_x^{\text{внеш}},\\
&\dot{L}_y + \omega_z L_x - \omega_x L_z = N_y^{\text{внеш}} \quad \text{или} \quad I_y\dot{\omega}_y - \omega_z\omega_x(I_z - I_x) = N_y^{\text{внеш}},\\
&\dot{L}_z + \omega_x L_y - \omega_y L_x = N_z^{\text{внеш}} \quad \text{или} \quad I_z\dot{\omega}_z - \omega_x\omega_y(I_x - I_y) = N_z^{\text{внеш}}.
\end{aligned}$$

Кинетическую энергию вращения мы можем записать как

$$T_{\text{вращение}} = \frac{\omega \cdot \mathbf{L}}{2} = \frac{\omega \cdot \mathbf{I} \cdot \omega}{2} = \frac{I\omega^2}{2}, \tag{7.41}$$

где $I = \mathbf{n}\cdot\mathbf{I}\cdot\mathbf{n} = \sum_{i=1}^{k} m_i[r_i^2 - (\mathbf{r}_i\cdot\mathbf{n})]$ *момент инерции* системы относительно оси вращения с единичным вектором $\mathbf{n} = \frac{\omega}{|\omega|}$.

Доказательство

$$\sum_{i=1}^{k} \frac{m_i \mathbf{v'}_i^2}{2} = \sum_{i=1}^{k} \frac{m_i \mathbf{v}'_i(\omega \times \mathbf{r}'_i)}{2} = \frac{\omega \cdot \mathbf{L}}{2}.$$

Уравнения кинетического момента, момента сил и тензора инерции также могут быть распространены на сплошную среду:

$$\mathbf{L} = \iiint\limits_V (\mathbf{r} \times \mathbf{j})\, dV, \tag{7.42}$$

$$\dot{\mathbf{L}} = \mathbf{N} = \iiint\limits_V (\mathbf{r} \times \mathbf{f})\, dV, \tag{7.43}$$

$$\mathbf{L} = \mathbf{I}\,\omega = \iiint\limits_V \rho(\mathbf{r})(\mathbf{r} \times (\omega \times \mathbf{r}))\, dV = \iiint\limits_V \rho(\mathbf{r})(\omega\mathbf{r}^2 - \mathbf{r}(\mathbf{r}\cdot\omega))\, dV. \tag{7.44}$$

где, например,

$$I_{xx} = \iiint\limits_V \rho(\mathbf{r})(r^2 - x^2)\, dV, \quad I_{xy} = -\iiint\limits_V \rho(\mathbf{r}) xy\, dV. \tag{7.45}$$

Симметрическую матрицу тензора инерции можно привести к диагональному виду. Собственные векторы матрицы будут *главными осями инерции*, а собственные значения I_x, I_y, I_z (диагональные элементы) — *главными моментами инерции*. Тогда

$$L_x = I_x\omega_x, \quad L_y = I_y\omega_y, \quad L_z = I_z\omega_z; \tag{7.46}$$

$$T_{\text{вращение}} = \frac{I\omega^2}{2} = \frac{I_x\omega_x^2 + I_y\omega_y^2 + I_z\omega_z^2}{2} \tag{7.47}$$

Лагранжиан вращения записывается в виде:

$$L = T - U = \frac{1}{2}(I_x\omega_x^2 + I_y\omega_y^2 + I_z\omega_z^2) - U(\varphi, \theta, \psi). \tag{7.48}$$

В завершение параграфа рассмотрим задачу, для решения которой применяются полученные дифференциальные уравнения. Одной из задач, к которой можно применить уравнения Эйлера, является задача о движении твёрдого тела, не подверженного действию никаких сил:

$$N_x^{\text{внеш}} = N_y^{\text{внеш}} = N_z^{\text{внеш}} = 0.$$

Но вместо их прямого интегрирования можно дать изящное геометрическое описание вращения твёрдого тела, не требующее полного решения задачи, известное как *геометрическая интерпретация Пуансо*.

Введём вектор $\rho = \frac{\omega}{|\omega|\sqrt{I}}$, тогда из (7.47) и из $I = \mathbf{n}\cdot\mathbf{I}\cdot\mathbf{n}$ (7.41) получим уравнение поверхности, называемой *эллипсоидом инерции*:

$$1 = \rho\cdot\mathbf{I}\cdot\rho = I_x\rho_x^2 + I_y\rho_y^2 + I_z\rho_z^2. \tag{7.49}$$

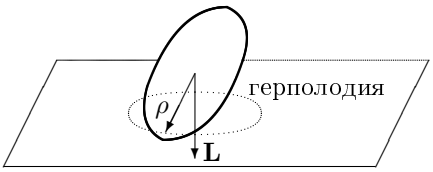


Рис. 7.1: Эллипсоид инерции

Так как направление оси вращения ω изменяется со временем, то вектор ρ также будет менять своё направление (рис. 7.1).

В том частном случае, который мы здесь рассматриваем, направление вектора $\mathbf{L}$ остаётся неизменным, и поэтому эллипсоид инерции (жёстко связанный с телом) должен двигаться в пространстве таким образом, чтобы сохранялась связь между ω и $\mathbf{L}$, при этом расстояние от центра эллипсоида инерции до плоскости, касающейся его в точке ρ, остаётся постоянным, так как

$$\frac{\rho \cdot \mathbf{L}}{L} = \frac{\omega \cdot \mathbf{L}}{\omega L\sqrt{I}} = \frac{2T_{\text{вращение}}}{L\sqrt{I\omega^2}} = \frac{\sqrt{2T_{\text{вращение}}}}{L}. \tag{7.50}$$

Так как нормаль к этой плоскости направлена вдоль $\mathbf{L}$ и, следовательно, имеет неизменное направление, то плоскость является неподвижной. Поэтому рассматриваемое движение можно реализовать посредством качения эллипсоида инерции по некоторой неподвижной плоскости; центр эллипсоида при этом находится в фиксированной точке пространства. Это качение происходит без скольжения, так как точка касания эллипсоида инерции с неподвижной плоскостью определяется вектором ρ, который направлен по мгновенной оси вращения, т.е. по прямой, все точки которой находятся в данный момент в покое.

Кривая, описываемая на эллипсоиде инерции точкой касания, называется *полодией*, а аналогичная кривая на неподвижной плоскости называется *герполодией*. Можно сказать, что полодия катится без скольжения по неподвижной герполодии.

В случае симметричного твёрдого тела эллипсоид инерции становится эллипсоидом вращения, а полодия превращается в окружность с центром на оси симметрии. Вектор угловой скорости ω описывает в этом случае поверхность конуса или совершает *прецессию* вокруг оси симметрии тела. В этом случае уравнения Эйлера имеют вид:

$$I_x\dot{\omega}_x = \omega_y\omega_z(I_x - I_z), \quad I_x\dot{\omega}_y = \omega_x\omega_z(I_z - I_x), \quad I_z\dot{\omega}_z = 0. \tag{7.51}$$

Из третьего уравнения следует, что

$$\omega_z = const. \tag{7.52}$$

Продифференцируем по времени первое уравнение

$$I_x\ddot{\omega}_x = \dot{\omega}_y\omega_z(I_x - I_z)$$

и подставим в него второе. В результате получим:

$$\ddot{\omega}_x = -\Omega^2\omega_x, \quad \text{где} \quad \Omega = \frac{(I_x - I_z)\omega_z}{I_z}. \tag{7.53}$$

Решение этого дифференциального уравнения можно записать в виде

$$\omega_x = A \sin \Omega t \tag{7.54}$$

где A — амплитуда прецессии, а Ω — частота прецессии, что проверяется простой подстановкой в уравнение.

Тогда решение для второго уравнения имеет вид

$$\omega_y = A \cos \Omega t. \tag{7.55}$$

7.4 Принцип Д'Аламбера

На основании сказанного до сих пор может сложиться впечатление, что все задачи механики сводятся к решению дифференциальных уравнений (7.17), т.е. к подстановке в эти уравнения известных сил, действующих на материальные частицы системы, и выполнению определённых математических операций, дающих решение задачи. Однако, это слишком упрощённое представление. Дело в том, что нужно учесть *связи* или функции вида $f = const$, ограничивающие движение системы и задающие соответствующее конфигурационное пространство V.

Если связь задаётся голономной функцией (которую можно восстановить по её дифференциалу), например, вида

$$f(t, \mathbf{r}_1, \ldots, \mathbf{r}_k) = 0$$

то такую связь называют *голономной*. Простейшим примером голономных связей могут служить связи в твёрдом теле, которые выражаются уравнениями вида

$$(\mathbf{r}_i - \mathbf{r}_j)^2 - c_{ij} = 0, \quad \text{где } c_{ij} \text{ — постоянные величины.}$$

Связь иного типа называют *неголономной*. Кроме того, если связь зависит явно от времени t, её называют *реономной*, а если нет — *склерономной*. Неголономные связи могут иметь вид различных неравенств, например, частица, лежащая на поверхности шара радиусом a, ограничена связью

$$\mathbf{r}^2 - a^2 \geqslant 0.$$

Также, неголономные связи могут содержать производные координат по времени. Неголономные связи могут быть вообще неизвестными, подлежащими определению, величинами данной задачи. Ведь в сущности задать в системе связи означает просто указать, что имеются силы, которые непосредственно не известны, но которые определённым образом влияют на движение системы. Кроме того, на обобщённые координаты не следует смотреть, как на декартовы, это могут быть, например, криволинейные координаты и т.д.

Пример. Пусть вертикальный диск без скольжения катится по горизонтальной плоскости с координатами x, y, причём направление качения диска

составляет угол θ с осью y (рис. 7.2). В силу связи «качения» скорость центра диска v пропорциональна скорости угла поворота диска вокруг своей оси φ: $v = r\dot{\varphi}$, где r — радиус диска.

Координаты x, y центра диска на плоскости задаются уравнениями динамической системы:

$$\dot{x} = v\sin\theta, \quad \dot{y} = -v\cos\theta.$$

С учётом $v = r\dot{\varphi}$, получаем *два дифференциальных уравнения связи*:

$$dx - r\sin\theta\, d\varphi = 0, \quad dy + r\cos\theta\, d\varphi = 0.$$

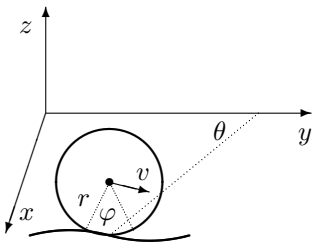


Рис. 7.2: Качение диска

Эти уравнения, очевидно, не могут быть проинтегрированы, пока вся задача не решена полностью. Неинтегрируемые связи «качения» являются частными случаями неголономных связей.

Задача с голономными связями всегда может быть решена, однако, для задач с неголономными связями общего метода решения не существует, и для каждой задачи требуется индивидуальный подход.

Один из методов решения задачи со связями предложил Д'Аламбер. В «Трактате о динамике» (1743), т.е. раньше, чем был сформулирован принцип Мопертюи, Д'Аламбер сформулировал свой фундаментальный принцип, сводящий динамику несвободной системы (со связями) к статике. В результате были получены уравнения, очень похожие на уравнения Лагранжа. Но в отличие от *метода вариации* (принципа наименьшего действия), который является «интегральным принципом», *принцип Д'Аламбера* является аналогичным методу вариаций «дифференциальным принципом», который использует *метод виртуальных перемещений*. Он позволяет применить к решению задач динамики более простые методы статики, и им широко пользуются в инженерной практике.

Возможным или *виртуальным* (бесконечно малым) перемещением системы называется произвольное бесконечно малое изменение её конфигурации $\delta\mathbf{r}_i$, согласующееся со связями, наложенными на неё в данный момент времени t. Виртуальным это перемещение называют для того, чтобы отличить его от действительного перемещения, происходящего за некоторый промежуток времени dt, в течение которого силы и связи могут измениться, поэтому $\delta t = 0$, т.е. время не варьируется.

Тогда запишем

$$\delta\mathbf{r}_i \overset{(3.28)}{=} \frac{\partial\mathbf{r}_i}{\partial\mathbf{q}}\,\delta\mathbf{q}. \tag{7.56}$$

Во-первых, при наличии связей не все координаты материальных точек будут независимыми. Если в системе заданы только голономные связи, то

следовательно, будут независимыми не все уравнения, вытекающие из второго закона Ньютона. При наличии m голономных связей нам достаточно просто уменьшить количество обобщённых координат $\mathbf{q} = (q_1, \ldots, q_n)$, которых теперь будет $n = 3k - m$. Уравнения $\mathbf{r}_i = \mathbf{r}_i(t, \mathbf{q})$, где $i = 1, \ldots, k$, будут уравнениями преобразования переменных.

Тогда мы можем записать

$$\dot{\mathbf{r}}_i \overset{(3.106)}{=} \frac{\partial \mathbf{r}_i}{\partial \mathbf{q}} \dot{\mathbf{q}} + \frac{\partial \mathbf{r}_i}{\partial t}. \tag{7.57}$$

Во-вторых, для учёта связей Д'Аламбер предложил ввести в системе соответствующие наперёд неизвестные силы, которые называются реакциями связей. Тогда сила $\mathbf{F}_i$, действующая на каждую из k материальных точек, будет состоять из *активной силы* $\mathbf{F}_i^\alpha$ и *реакции связи* $\mathbf{f}_i$:

$$\dot{\mathbf{p}}_i = \mathbf{F}_i = \mathbf{F}_i^\alpha + \mathbf{f}_i, \quad i = 1, \ldots, k. \tag{7.58}$$

Тогда получим

$$0 \overset{(7.58)}{=} \sum_{i=1}^{k} (\mathbf{F}_i - \dot{\mathbf{p}}_i)\, \delta\mathbf{r}_i = \sum_{i=1}^{k} (\mathbf{F}_i^\alpha - \dot{\mathbf{p}}_i)\, \delta\mathbf{r}_i + \sum_{i=1}^{k} \mathbf{f}\, \delta\mathbf{r}_i. \tag{7.59}$$

Мы будем рассматривать лишь такие системы, для которых виртуальная работа реакций связей равна нулю $\sum_{i=1}^{k} \mathbf{f}\delta\mathbf{r}_i = 0$. Такие связи называют *идеальными*. Это условие справедливо для любого твёрдого тела, а также для большого числа других систем. Следует отметить, что это утверждение перестаёт быть справедливым при наличии, например, силы трения. Однако это ограничение не является очень сильным, так как в отличие от бесконечно малого перемещения трение представляет, в сущности, макроскопическое явление.

В результате, получаем *уравнение Д'Аламбера–Лагранжа*:

$$\sum_{i=1}^{k} (\mathbf{F}_i^\alpha - \dot{\mathbf{p}}_i)\, \delta\mathbf{r}_i = 0. \tag{7.60}$$

Выразим первый член уравнения (7.60) через $\mathbf{q}$:

$$\sum_{i=1}^{k} \mathbf{F}_i^\alpha \delta\mathbf{r}_i \overset{(7.56)}{=} \sum_{i=1}^{k} \mathbf{F}_i^\alpha \frac{\partial \mathbf{r}_i}{\partial \mathbf{q}}\, \delta\mathbf{q} = \mathbf{Q}\, \delta\mathbf{q}, \tag{7.61}$$

где $Q_j = \sum_{i=1}^{k} \mathbf{F}_i^\alpha \frac{\partial \mathbf{r}_i}{\partial q_j}$ $(j = 1, \ldots, n)$ — обобщённые силы Д'Аламбера.

Для преобразования второго члена уравнения (7.60) запишем

$$\frac{d}{dt}\left(\frac{\partial \mathbf{r}_i}{\partial \mathbf{q}}\right) \overset{(3.87)}{=} \frac{\partial \dot{\mathbf{r}}_i}{\partial \mathbf{q}}, \tag{7.62}$$

$$\frac{d}{dt}\left(\dot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial \mathbf{q}}\right) \overset{(3.8)}{=} \ddot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial \mathbf{q}} + \dot{\mathbf{r}}_i \frac{d}{dt}\left(\frac{\partial \mathbf{r}_i}{\partial \mathbf{q}}\right) \overset{(7.62)}{=} \ddot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial \mathbf{q}} + \dot{\mathbf{r}}_i \frac{\partial \dot{\mathbf{r}}_i}{\partial \mathbf{q}}, \tag{7.63}$$

$$\sum_{i=1}^{k} m_i \dot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial \mathbf{q}} \overset{(3.146)}{=} \sum_{i=1}^{k} m_i \dot{\mathbf{r}}_i \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{\mathbf{q}}} = \frac{\partial}{\partial \dot{\mathbf{q}}}\left(\sum_{i=1}^{k} \frac{m_i \dot{\mathbf{r}}_i^2}{2}\right) \overset{(7.23)}{=} \frac{\partial T}{\partial \dot{\mathbf{q}}}, \tag{7.64}$$

$$\sum_{i=1}^{k} m_i \dot{\mathbf{r}}_i \frac{\partial \dot{\mathbf{r}}_i}{\partial \mathbf{q}} = \frac{\partial}{\partial \mathbf{q}}\left(\sum_{i=1}^{k} \frac{m_i \dot{\mathbf{r}}_i^2}{2}\right) \overset{(7.23)}{=} \frac{\partial T}{\partial \mathbf{q}}. \tag{7.65}$$

Теперь получим выражение для второго члена уравнения (7.60):

$$\begin{aligned} \sum_{i=1}^{k} \dot{\mathbf{p}}_i \, \delta \mathbf{r}_i &\overset{(7.56)}{=} \sum_{i=1}^{k} \dot{\mathbf{p}}_i \frac{\partial \mathbf{r}_i}{\partial \mathbf{q}} \, \delta \mathbf{q} \overset{(7.21)}{=} \sum_{i=1}^{k} m_i \ddot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial \mathbf{q}} \, \delta \mathbf{q} = \\ &\overset{(7.63)}{=} \sum_{i=1}^{k} \left[m_i \frac{d}{dt}\left(\dot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial \mathbf{q}}\right) - m_i \dot{\mathbf{r}}_i \frac{\partial \dot{\mathbf{r}}_i}{\partial \mathbf{q}} \right] \delta \mathbf{q} = \\ &\overset{\substack{(7.64)\\(7.65)}}{=} \left[\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial T}{\partial \mathbf{q}} \right] \delta \mathbf{q}. \end{aligned} \tag{7.66}$$

Запишем уравнение Д'Аламбера–Лагранжа (7.60) в виде:

$$\left[\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial T}{\partial \mathbf{q}} - \mathbf{Q} \right] \delta \mathbf{q} = 0. \tag{7.67}$$

Если связи в системе голономные, то обобщённые координаты $\mathbf{q}$, а значит и вариации $\delta \mathbf{q}$, независимые, и все члены при δq_j должны обращаться в нуль.

В результате получаем n *уравнений Д'Аламбера–Лагранжа*:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_j}\right) - \frac{\partial T}{\partial q_j} = Q_j \quad (j = 1, \ldots, n). \tag{7.68}$$

В консервативной системе для обобщённой силы запишем:

$$Q_j = \sum_{i=1}^{k} \mathbf{F}_i^{\alpha} \frac{\partial \mathbf{r}_i}{\partial q_j} \overset{(7.24)}{=} -\sum_{i=1}^{k} \frac{\partial U}{\partial \mathbf{r}_i} \frac{\partial \mathbf{r}_i}{\partial q_j} = -\frac{\partial U}{\partial q_j}. \tag{7.69}$$

Тогда с учётом $\frac{\partial U}{\partial \dot{q}_j} = 0$ и $L = T - U$ можно записать уравнения Эйлера–Лагранжа (5.47) или (7.3)

$$0 = \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_j}\right) - \frac{\partial T}{\partial q_j} + \frac{\partial U}{\partial q_j} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right) - \frac{\partial L}{\partial q_j}.$$

7.5 Механические инварианты

Согласно теоремы Эмми Нётер (5.48) для механической системы из k точек всякой однопараметрической группе диффеоморфизмов ϱ^s многообразия лагранжевой системы, сохраняющих функцию Лагранжа ($\frac{dL}{ds} = 0$), соответствует первый интеграл уравнений движения вида

$$\mathcal{I} = \mathbf{p} \cdot \mathbf{q}'_s \overset{(7.20)}{=} \sum_{i=1}^{k} m_i \dot{\mathbf{r}}_i \frac{d\mathbf{r}_i}{ds} = const \quad \text{при условии} \quad \frac{dL}{ds} = 0. \tag{7.70}$$

Непрерывная группа преобразований предполагает наличие определённых симметрий в пространстве рассматриваемой системы.

Применительно к механике теорема Нётер утверждает наличие определённых механических инвариантов или инвариантность определённых физических законов относительно присущих системе симметрий:

1. *Однородность* или *гомогенность* (от греч. *ομός* — равный, одинаковый и *γένω* — рождать) — одинаковость свойств.

   Однородность пространства означает, что мы можем произвольно совершать параллельный перенос системы координат, и это не влияет на результат измерений.

   Для i-й материальной точки запишем $\mathbf{r}_i \overset{(2.21)}{=} e_x x_i + e_y y_i + e_z z_i$.

   Пусть система допускает сдвиги вдоль оси x: $\varrho^s \colon \mathbf{r}_i \to \mathbf{r}_i + e_x s$.

   Тогда

   $$\frac{d\mathbf{r}_i}{ds} = e_x \quad \text{для всех } i, \text{ значит согласно (7.70)} \tag{7.71}$$

   $$\mathcal{I} = \sum_{i=1}^{k} m_i \dot{\mathbf{r}}_i \cdot e_x = \sum_{i=1}^{k} m_i \dot{x}_i = p_x = const. \tag{7.72}$$

   Аналогичный результат получается при сдвиге вдоль осей y и z.

   В результате получаем три инварианта: $\mathbf{p} = (p_x, p_y, p_z) = const$.

   Согласно теореме Нётер из *однородности пространства* следует *закон сохранения импульса* в механической системе, точнее, каждой из трёх компонент импульса.

   В самом простом случае механической системы, состоящей из k точек, сохранение импульса в отсутствие внешних сил сразу следует из уравнения (7.22): если сумма внешних сил, действующих на систему, равна нулю, то полное количество движения системы остаётся неизменным.

2. *Изотропность* (от др.-греч. *ἴσος* — равный, одинаковый, подобный и *τρόπος* — оборот, поворот, характер) — инвариантность по отношению к выбору направления.

Изотропность пространства означает инвариантность относительно поворотов: мы можем произвольно поворачивать систему координат, и это не влияет на результат измерений.

Пусть система допускает вращение вокруг оси e_x, тогда согласно уравнению вращения (3.113):

$$\frac{d\mathbf{r}_i}{ds} = e_x \times \mathbf{r}_i, \quad \text{значит согласно (7.70)} \tag{7.73}$$

$$\mathcal{I} = \sum_{i=1}^{k} m_i \dot{\mathbf{r}}_i (e_x \times \mathbf{r}_i) \overset{\text{стр. } 86}{=} \sum_{i=1}^{k} m_i e_x (\mathbf{r}_i \times \dot{\mathbf{r}}_i) = L_x = const. \tag{7.74}$$

Аналогичный результат мы можем получить при вращении вокруг вокруг e_y и e_z. В результате получаем: $\mathbf{L} = (L_x, L_y, L_z) = const.$

Согласно теореме Нётер из *изотропности пространства* следует *закон сохранения кинетического момента* в механической системе.

В самом простом случае механической системы, состоящей из k точек, сохранение кинетического момента в отсутствие внешнего момента сил сразу следует из уравнения моментов (7.34): если момент внешних сил относительно какого-либо центра или оси, действующих на систему, равен нулю, то кинетический момент относительно этого центра или оси будет постоянной величиной.

Системы, на которые не действуют как внешние силы, так и внешние моменты сил, называются *замкнутыми* или *изолированными системами*. При этом на силы внутри системы никаких ограничений не накладывается.

3. *Однородность времени* означает, что мы можем произвольно выбирать точку отсчёта на оси времени или, другими словами, система допускает в уравнениях движения сдвиг вдоль оси t: $\varrho^s : t \to t + s$. При этом согласно (7.70) $\frac{dL}{ds} = \frac{dL}{dt} = 0$ или $L = const$.

   Тогда

$$\frac{d\mathbf{r}_i}{ds} \overset{(3.5)}{=} \frac{d\mathbf{r}_i}{dt}\frac{dt}{ds} = \dot{\mathbf{r}}_i, \quad \text{значит} \quad \mathcal{I} = \sum_{i=1}^{k} m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i = 2T = const. \tag{7.75}$$

   В результате получаем: $E \overset{(7.5)}{=} 2T - L = const.$

   Согласно теореме Нётер из *однородности времени* следует *закон сохранения энергии* в механической системе.

   Согласно уравнению (7.2) полная энергия $E = T + U$ механической системы сохраняется, если она представляет собой автономную динамическую систему (5.3): $\dot{\mathbf{p}} = \varphi(\mathbf{p}, \mathbf{q}), \dot{\mathbf{q}} = \psi(\mathbf{p}, \mathbf{q})$. Это означает, что если в системе действуют только консервативные силы, то энергия внутри системы не может самопроизвольно возникать и исчезать. Система,

в которой справедлив закон сохранения энергии $E = T + U = const$, называется *консервативной*.

Консервативная система — любая механическая система, движущаяся во внешнем стационарном (не изменяющемся со временем) потенциальном силовом поле при условии, что система свободна (без связей) или наложенные на неё связи являются идеальными.

Консервативную систему не следует смешивать с замкнутой (изолированной) системой, для которой имеет место закон сохранения количества движения и кинетического момента, т.е. замкнутая система вообще может не быть консервативной, если внутренние силы не являются потенциальными. В свою очередь, при наличии потенциального силового поля консервативная система не замкнута.

Мы получили семь *интегралов движения* механической системы:

$$\mathbf{L} = (L_x, L_y, L_z) = const, \quad \mathbf{p} = (p_x, p_y, p_z) = const, \quad E = const. \tag{7.76}$$

7.6 Релятивистская механика

Преобразование Галилея представлялось незыблемым до тех пор, пока не был сформулирован принцип близкодействия теории электромагнетизма Максвелла, утверждающий, что передача электромагнитных возмущений от точки к точке происходит с конечной скоростью, а именно со скоростью света. Целый ряд экспериментов, в особенности известные опыты А. Майкельсона (1881) и У. Морли (1885–87), указывали на то, что Дж. Максвелл теоретически показал в 1865 — скорость света *инвариантна*, т.е. одинакова во всех направлениях и не зависит от движения наблюдателя или источника света, а также от среды, в которой распространяется свет. Если отправиться в погоню за светом со скоростью света, то интуиция, основанная на преобразовании Галилея, подсказывает, что мы догоним свет, и он будет казаться нам неподвижным. Но согласно теории Максвелла и не вызывающим сомнения экспериментальным данным, такого явления как неподвижный свет, не существует. Например, В. де Ситтер (1913) предположил, что используя двойные звёзды можно измерить влияние движения источника на скорость света. Однако скорость света, испускаемого движущейся звездой, оказалась равна скорости света от неподвижной звезды.

Рассмотрим сферические волны света, распространяющиеся от точечного источника, расположенного в начале координат. Скорость любой точки волнового фронта равна

$$\dot{\mathbf{r}} = c\mathbf{n} \quad \text{или} \quad d\mathbf{r} = c\mathbf{n}\, dt, \tag{7.77}$$

где $\mathbf{n}$ — единичный вектор нормали к фронту ($|\mathbf{n}|^2 = 1$),
c — скорость света.

В соответствии с преобразованием Галилея (7.13) в произвольной инерциальной системе скорость выбранной точки волнового фронта должна

быть $c\mathbf{n} + \mathbf{v}$, где вектор $\mathbf{n}$ для разных точек фронта имеет разное направление, а $\mathbf{v}$ — одинаковое. Следовательно, в произвольной инерциальной системе скорость волны не будет уже равна c. Кроме того, она будет зависеть от направления, т.е. волна уже не будет сферической. Значит, преобразование Галилея должно быть *уточнено* при помощи дополнительного преобразования, сохраняющего скорость света постоянной во всех инерциальных системах.

Запишем *уточнённое преобразование Галилея* в следующем виде:

$$\frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}'}{dt'} + \mathbf{v}. \tag{7.78}$$

Обратим внимание, что, в отличие от классической записи (7.13), в правой и левой части уравнения используется *разное* (!) время.

При всей кажущейся малости произведённого уточнения, оно имеет принципиальное значение. Мы отошли от постулата *абсолютного времени*. Согласно Галилею реальность у всех одна, а при таком уточнении каждая система имеет своё собственное время.

Постулат, что свойства времени считаются независимыми от системы отсчёта — время одно для всех систем отсчёта, настолько нам привычен, что даже не оговаривался при введении в уравнениях времени в качестве параметра. Мы не удивились этому обстоятельству, так как даже не заметили его. Оказывается, однако, что в разных системах отсчёта время течёт по-разному. Следовательно, утверждение, что между двумя данными событиями прошёл определённый промежуток времени, теперь приобретает смысл только тогда, когда указано, к какой системе отсчёта это утверждение относится.

Отказавшись от абсолютного времени мы потеряли опору. Чтобы сравнивать различные инерциальные системы нам необходимо найти нечто абсолютное, одинаковое во всех этих системах.

Оказывается, что величина

$$s^2 = c^2t^2 - \mathbf{r}^2, \tag{7.79}$$

называемая *интервал* и является таким инвариантом.

Инвариантность интервала является математическим выражением постоянства скорости света. Действительно, уравнение фронта сферической волны света в произвольной инерциальной системе легко вычисляется $\mathbf{r}^2 \overset{(7.77)}{=} c^2t^2$ или $c^2t^2 - \mathbf{r}^2 = 0$. При этом для световой волны также справедливо уравнение $d\mathbf{r}^2 = c^2\,dt^2$. Тогда

$$c^2t^2 - \mathbf{r}^2 = c^2t'^2 - \mathbf{r}'^2, \tag{7.80}$$

$$c^2\,dt^2 - d\mathbf{r}^2 = c^2\,dt'^2 - d\mathbf{r}'^2. \tag{7.81}$$

Выражения (7.80) и (7.81) для света справедливы в любой инерциальной системе. Однако, если скорость света одинакова во всех инерциальных системах $c = const$, то эти выражения должны быть справедливы не только

для света, но и для любого другого объекта. В противном случае окажется, что скорость света не является инвариантом, а является функцией координат и времени.

Допустим теперь, мы наблюдаем из некоторой инерциальной системы отсчёта произвольным образом движущиеся относительно нас со скоростью $v = |\mathbf{v}|$ часы. Спрашивается, какой промежуток времени dt' покажут при этом движущиеся часы.

Если наша инерциальная система неподвижна, то $d\mathbf{r} = 0$. Тогда

$$c^2dt^2 \overset{(7.81)}{=} c^2dt'^2 - d\mathbf{r}'^2 = c^2dt'^2 - v^2dt'^2 \quad \text{или} \quad dt' = \frac{dt}{\sqrt{1-\left(\frac{v}{c}\right)^2}}. \tag{7.82}$$

Для удобства выберем координаты таким образом, чтобы оси x и x' инерциальных систем были направлены вдоль вектора скорости $\mathbf{v}$.

Тогда $dy' = dz' = 0$, и мы можем записать

$$y = y', \quad z = z'; \tag{7.83}$$

$$\frac{dx}{dt} \overset{(7.78)}{=} \frac{dx'}{dt'} + v; \tag{7.84}$$

$$c^2t^2 - x^2 \overset{(7.80)}{=} c^2t'^2 - x'^2. \tag{7.85}$$

Откуда получим

$$dx' \overset{(7.84)}{=} \frac{dt'}{dt}(dx - v\,dt) \overset{(7.82)}{=} \frac{dx - v\,dt}{\sqrt{1-\left(\frac{v}{c}\right)^2}}$$

и после интегрирования запишем

$$x' = \frac{x - vt}{\sqrt{1-\left(\frac{v}{c}\right)^2}}. \tag{7.86}$$

Кроме того, из уравнения (7.85) получим:

$$c^2t^2 - x^2 \overset{(7.86)}{=} c^2t'^2 - \frac{(x-vt)^2}{1-\left(\frac{v}{c}\right)^2}$$

Откуда можно записать:

$$t' = \frac{t - \frac{xv}{c^2}}{\sqrt{1-\left(\frac{v}{c}\right)^2}}. \tag{7.87}$$

Уравнения (7.86) и (7.87) есть преобразование плоскости (ct, x) с матрицей преобразования ω:

$$\omega = \frac{1}{\sqrt{1-\left(\frac{v}{c}\right)^2}} \begin{Vmatrix} 1 & -\frac{v}{c} \\ -\frac{v}{c} & 1 \end{Vmatrix} \quad \text{или} \quad \begin{Vmatrix} ct' \\ x' \end{Vmatrix} = \omega \begin{Vmatrix} ct \\ x \end{Vmatrix}. \tag{7.88}$$

Легко убедиться, что $\det\omega = 1$.

Обратное преобразование немногим отличается от прямого, а именно знаком при величине относительной скорости

$$\omega^{-1} = \frac{1}{\sqrt{1-\left(\frac{v}{c}\right)^2}} \begin{Vmatrix} 1 & \frac{v}{c} \\ \frac{v}{c} & 1 \end{Vmatrix}, \quad \det\omega^{-1} = 1; \tag{7.89}$$

что вполне логично, так как в противоположной системе вектор относительной скорости видится зеркально перевёрнутым. Таким образом, обе системы оперируют *одним и тем же* преобразованием.

Уравнения (7.83), (7.86) и (7.87) образуют систему, которую называют *преобразование Лоренца*.

Отметим два известных следствия из преобразования Лоренца:

- Из преобразования Лоренца следует, что невозможна относительная скорость больше c, так как при $v > c$ преобразование Лоренца даёт мнимые координаты.

- В движущейся системе будет происходить кажущееся «растяжение времени» и, следовательно, кажущееся «сжатие длины» относительно неподвижной системы, названные эффектом Лоренца–Фицджеральда.

  На самом деле все лоренцовы системы эквивалентны, и мы не можем различить неподвижную и равномерно движущуюся системы. Поэтому будет неверным вслед за многочисленными фантастами утверждать, например, что если космический корабль движется относительно якобы «неподвижной» Земли, то на корабле время течёт медленнее земного. Подобные неверные представления (кто относительно кого движется) являются примером того, как популяризация науки приводит к устойчивым, но ошибочным мифам.

Введём обозначения:

$$\cos\varphi = \frac{1}{\sqrt{1-\left(\frac{v}{c}\right)^2}}, \quad \sin\varphi = \frac{\frac{iv}{c}}{\sqrt{1-\left(\frac{v}{c}\right)^2}},$$

где i — мнимая единица.

Тогда уравнения (7.87), (7.86) и (7.83) можно записать в виде

$$\begin{aligned} ict' &= ict\cos\varphi - x\sin\varphi, \\ x' &= ict\sin\varphi + x\cos\varphi, \\ y' &= y, \\ z' &= z; \end{aligned} \quad \text{или} \quad \begin{Vmatrix} ict' \\ x' \\ y' \\ z' \end{Vmatrix} = \begin{Vmatrix} \cos\varphi & -\sin\varphi & 0 & 0 \\ \sin\varphi & \cos\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix} \begin{Vmatrix} ict \\ x \\ y \\ z \end{Vmatrix}.$$

Таким образом, уточнённое преобразование Галилея (7.78) при соблюдении условий (7.80) и (7.81) есть *собственное ортогональное преобразование* (2.87) четырёхмерного радиус-вектора с координатами ct, x, y, z, а именно представляет собой поворот в плоскости (ct, x) на угол φ. Очевидно, что угол φ мнимый, так как $\cos\varphi > 1$.

В привычном нам мире время и пространство представляют собой различные сущности. Теория относительности использовала геометрические идеи многомерного пространства, присоединив к трём пространственным осям ещё и ось времени. В результате прошлое, настоящее и будущее объекта сосуществуют вместе.

Обратим внимание ещё раз. Парадоксальное сохранение скорости света относительно всех инерциальных систем, даже когда мы пытаемся догнать свет, двигаясь с такой же скоростью, с формальной математической точки зрения оказывается обычным вращением системы координат в четырёхмерном пространстве.

Четырёхмерное пространство (ct, x, y, z) называют *пространство Минковского*. Инерциальную систему в пространстве Минковского называют *лоренцовой системой* или *системой Лоренца*.

Бесконечно малый интервал $c^2\,dt^2 - d\mathbf{r}^2$ (7.81) является инвариантом для всех инерциальных систем и поэтому может быть выбран в качестве *метрики* в пространстве Минковского. Но если для света это выражение равно нулю, то для произвольного свободного объекта метрика отлична от нуля и называется *мировым скаляром*:

$$ds^2 = c^2 d\tau^2 = c^2\,dt^2 - d\mathbf{r}^2. \tag{7.90}$$

Допустим, что мы наблюдаем из движущейся инерциальной системы за неподвижными часами. В связанной с часами системе координат часы покоятся, т.е. $d\mathbf{r}' = 0$ и $dt' = d\tau$. Тогда

$$c^2 d\tau^2 \overset{(7.81)}{=} c^2 dt^2 - d\mathbf{r}^2 = c^2 dt^2 - v^2 dt^2 \quad \text{или} \quad d\tau = dt\sqrt{1 - \left(\frac{v}{c}\right)^2}. \tag{7.91}$$

Время, отсчитываемое по часам, движущимся вместе с данным объектом, называется *собственным временем* τ объекта, которое, как мы видим, всегда меньше, чем соответствующий промежуток времени в неподвижной системе. Другими словами, нам всегда кажется, что движущиеся относительно нас часы идут медленнее неподвижных.

Полученный результат формулируют следующим образом: *точка пространства Минковского* описывает *событие*, происходящее с объектом (частицей) в данный момент времени τ. Событие, происходящее по собственному времени объекта τ, называется *мировой точкой*, а совокупность мировых точек, или траектория движения в пространстве Минковского — *мировой линией* объекта, или геометрическим местом всех событий существования объекта.

Всякой частице соответствует мировая линия, точки которой определяют координаты частицы во все моменты времени. Например, промежуток

времени, показываемый часами, равен интегралу

$$\tau_2 - \tau_1 \overset{(7.90)}{=} \frac{1}{c}\int ds \overset{(7.91)}{=} \int\limits_{t_1}^{t_2} dt\sqrt{1-\left(\frac{v}{c}\right)^2}, \tag{7.92}$$

взятому вдоль мировой линии этих часов.

Если часы неподвижны, то их мировая линия является прямой, параллельной оси времени; если же часы совершают неравномерное движение по замкнутому пути, то их мировая линия будет четырёхмерной кривой, проходящей через две точки на прямой мировой линии неподвижных часов, соответствующих началу и концу движения. С другой стороны, покоящиеся часы показывают всегда больший промежуток времени, чем движущиеся. Следовательно, интеграл $\int ds$, взятый между двумя заданными мировыми точками, имеет максимальное значение, если он берется вдоль мировой линии, соединяющей эти точки. Предполагается, разумеется, что эти точки и соединяющие их линии таковы, что все элементы ds вдоль линий временноподобные.

Таким образом, интеграл ds имеет максимум вдоль мировой линии, а действие S должно иметь минимум. Следовательно, действие можно представить, как величину, пропорциональную указанному интегралу и взятую с обратным знаком:

$$\int\limits_{t_1}^{t_2} L\,dt \overset{(5.75)}{=} S = -\alpha\int ds \overset{(7.92)}{=} -\alpha c\int\limits_{t_1}^{t_2} dt\sqrt{1-\left(\frac{v}{c}\right)^2}, \tag{7.93}$$

где α — нормировочный коэффициент.

Тогда для $v \ll c$:

$$L = -\alpha c\sqrt{1-\left(\frac{v}{c}\right)^2} \approx -\alpha c + \frac{\alpha v^2}{2c}.$$

Первый член разложения не зависит от скорости, а значит, не вносит никаких изменений в уравнения движения. Тогда, сравнивая второй член с классическим выражением функции Лагранжа для свободно движущегося объекта (с нулевой потенциальной энергией): $L = T \overset{(7.11)}{=} \frac{mv^2}{2}$, нетрудно определить нормировочный коэффициент: $\alpha = mc$.

Таким образом, окончательно получаем вид функции Лагранжа свободно движущегося объекта:

$$L = -mc^2\sqrt{1-\left(\frac{v}{c}\right)^2}. \tag{7.94}$$

Импульсом называется, как известно, вектор

$$\mathbf{p} \overset{(7.3)}{=} \frac{\partial L}{\partial \dot{\mathbf{q}}} \overset{(7.94)}{=} \frac{m\mathbf{v}}{\sqrt{1-\left(\frac{v}{c}\right)^2}}. \tag{7.95}$$

При $v \ll c$ это выражение переходит в классическое $\mathbf{p} = m\mathbf{v}$.

Энергия вычисляется как

$$E \overset{(5.45)}{=} \mathbf{pv} - L \overset{\substack{(7.94)\\(7.95)}}{=} \frac{mc^2}{\sqrt{1 - \left(\frac{v}{c}\right)^2}}. \tag{7.96}$$

При $v \ll c$ получаем $E \approx mc^2 + \frac{mv^2}{2}$, а при $v = 0$ в релятивистской механике энергия свободной частицы не обращается в нуль, а остаётся равной $E = mc^2$. Эту величину называют *энергией покоя* частицы.

Из выражений (7.95) и (7.96) вытекает следующее соотношение между энергией, импульсом и скоростью свободной частицы

$$\mathbf{p} = \frac{E\mathbf{v}}{c^2}, \quad E^2 = p^2c^2 + m^2c^4. \tag{7.97}$$

При $v = c$ импульс и энергия частицы обращаются в бесконечность. Это значит, что частица с отличной от нуля массой не может двигаться со скоростью света. Однако, световые частицы (кванты) не имеют массы но, двигаясь со скоростью света, имеют импульс $p = \frac{E}{c}$.

Полученные результаты можно интерпретировать следующим образом. Все векторы в пространстве Минковского имеют четвёртую компоненту. Четырёхкомпонентные векторы в пространстве Минковского называют также 4-векторы. Одним из примеров 4-вектора может служить 4-радиус-вектор

$$\mathbf{x} = (ct, \mathbf{r}), \quad \text{где} \quad x^0 = ct,\ x^1 = x,\ x^2 = y,\ x^3 = z. \tag{7.98}$$

Квадрат величины всякого 4-вектора $\mathbf{a} = (a^0, a^1, a^2, a^3)$ определяется как произведение *ковариантного* вектора с компонентами a_i и *контравариантного* вектора с компонентами a^i. При этом

$$a_0 = a^0,\ a_1 = -a^1,\ a_2 = -a^2,\ a_3 = -a^3; \tag{7.99}$$

$$|\mathbf{a}|^2 = a_i\, a^i = a_0\, a^0 + a_1\, a^1 + a_2\, a^2 + a_3\, a^3. \tag{7.100}$$

Компоненту a^0 называют временной, а компоненты a^1, a^2, a^3 — пространственными. По отношению к поворотам в трёхмерном пространстве временная компонента представляет собой скаляр, а три пространственных компоненты a^1, a^2, a^3 по отношению к этим поворотам составляют трёхмерный вектор. По отношению к зеркальным отражениям в трёхмерном пространстве временная компонента, будучи скаляром, не меняется, а пространственные компоненты изменяют знак, т.е. ведут себя как компоненты полярного (истинного) вектора.

Таким образом, пространственные компоненты контравариантного 4-вектора можно понимать как истинный вектор в трёхмерном пространстве стандартной (правой) ориентации, а пространственные компоненты ковариантного 4-вектора можно понимать как зеркальное отражение контравариантного вектора в пространство левой ориентации. Само собой, что временная компонента, будучи скаляром, не меняется при таком зеркальном отражении.

Произведение $a_i\,a^i = a^i\,a_i$ является скаляром — оно инвариантно по отношению к любым поворотам четырёхмерной системы координат, которыми являются, в частности, преобразования Лоренца.

Например, квадрат «длины» 4-радиус-вектора и метрика в пространстве-времени даются выражениями

$$\mathbf{x}^2 = x_i\,x^i = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \overset{(7.79)}{=} s^2, \tag{7.101}$$

$$ds^2 = c^2 d\tau^2 \overset{(7.90)}{=} dx_i\,dx^i. \tag{7.102}$$

В отличие от трёхмерных векторов, у которых компоненты не зависели друг от друга, компоненты 4-векторов не независимы.

Квадрат 4-вектора может быть положительным, отрицательным или равным нулю; в этих случаях говорят соответственно о *временно-подобных*, *пространственно-подобных* и *изотропных* (*нулевых*) векторах.

Рассмотрим два события $(t_1, \mathbf{r}_1)$ и $(t_2, \mathbf{r}_2)$, которые происходят в какой-либо инерциальной системе в одном и том же месте. Тогда $\mathbf{r}_1 = \mathbf{r}_2$ и $s_{21}^2 = c^2(t_2 - t_1)^2 > 0$, т.е. интервал между двумя событиями вещественный. Вещественные интервалы называют *временно-подобными*. Если какие-нибудь два события происходят с одним и тем же телом, то интервал между ними всегда временно-подобный, так как скорость тела не может быть больше c, а пройденное растояние — больше $c\,t_{21}$.

Если два события происходят в какой-либо инерциальной системе в одно и то же время, тогда $t_1 = t_2$ и $s_{21}^2 = -(\mathbf{r}_2 - \mathbf{r}_1)^2 < 0$, т.е. интервал между двумя одновременными событиями мнимый. Мнимые интервалы называют *пространственно-подобными*.

Подразделение интервалов на временно-подобные и пространственно-подобные есть, в силу их инвариантности, понятие абсолютное. Это значит, что свойство интервала быть пространственно-подобным или временно-подобным не зависит от системы отсчёта.

Очевидно, что 4-радиус-вектор $\mathbf{x}$ — временно-подобный.

Вектор 4-*скорости* по определению равен

$$\mathbf{u} = \frac{d\mathbf{x}}{d\tau} \quad \text{или} \quad u^i = \frac{dx^i}{d\tau}: \quad u^0 = \frac{c\,dt}{d\tau},\ u^1 = \frac{dx}{d\tau},\ u^2 = \frac{dy}{d\tau},\ u^3 = \frac{dz}{d\tau}; \tag{7.103}$$

с учётом (7.91) получаем

$$u^0 = \frac{c}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \overset{(7.96)}{=} \frac{E}{mc}, \quad u^{1,2,3} = \frac{v_{x,y,z}}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \tag{7.104}$$

Абсолютная величина 4-скорости является постоянной величиной

$$u^2 = u_i\,u^i = \frac{c^2 - v_x^2 - v_y^2 - v_z^2}{1 - \left(\frac{v}{c}\right)^2} = \frac{c^2 - v^2}{1 - \left(\frac{v}{c}\right)^2} = c^2. \tag{7.105}$$

Отсюда видно, что вектор $\mathbf{u}$ также временно-подобный. Таким образом, все объекты во Вселенной *всегда* движутся в четырёхмерном пространстве

с одной постоянной скоростью света, а скорость движения объекта в обычном пространстве (в заданной лоренцовской системе) является просто отражением соответствующего угла поворота системы координат относительно «неподвижного» наблюдателя.

Для вектора 4-*импульса* μ имеем:

$$|\mu| = m\,|\mathbf{u}| \overset{(7.105)}{=} mc: \quad p^0 = mu^0 \overset{(7.104)}{=} \frac{E}{c}, \quad p^{1,2,3} = mu^{1,2,3}. \tag{7.106}$$

При $v \ll c$ получаем

$$\mathbf{u} = (c, \mathbf{v}); \tag{7.107}$$

$$\mu = \left(\frac{E}{c}, \mathbf{p}\right). \tag{7.108}$$

Таким образом в любой инерциальной системе общее количество движения (импульс) и общая энергия свободного объекта являются компонентами одного 4-вектора μ и образуют в специальной теории относительности один закон — закон постоянства этого 4-вектора. Свет не имеет массы и, следовательно, имеет в любой системе отсчёта только одну, четвёртую, компоненту 4-вектора импульса $p = \frac{E}{c}$.

Аналогично определению 4-скорости, вторую производную

$$\mathbf{w} = \frac{d^2\mathbf{x}}{d\tau^2} \overset{(7.103)}{=} \frac{d\mathbf{u}}{d\tau} \quad \text{или} \quad w^i = \frac{du^i}{d\tau} \tag{7.109}$$

можно назвать 4-*ускорением*.

Дифференцируя соотношение (7.105), получаем

$$u_i\,w^i + w_i\,u^i = 2\,w_i\,u^i \overset{\text{(п. 1 на с. 107)}}{=} 0, \tag{7.110}$$

т.е. 4-векторы скорости и ускорения взаимно ортогональны.

Сохранение *релятивистской энергии* $E \overset{(7.97)}{=} \sqrt{p^2c^2 + m^2c^4}$ при изменении величины p возможно только в том случае, если происходит изменение *энергии покоя* mc^2, т.е. *массы покоя* m. Наиболее ярким примером перехода массы в энергию является взрыв атомной бомбы. Следует отметить, что при взрыве уменьшается лишь масса покоя и значительно увеличивается кинетическая энергия движения, но полная энергия и полная масса системы остаются неизменными. Количество движения при взрыве также сохраняется. Несмотря на фантастическое количество выделяющейся кинетической энергии, потеря массы бомбы не превышает $0,1\%$ от первоначальной массы покоя.

Эйнштейн предложил сохранить для лоренцовых систем постулат эквивалентности, т.е. известные физические законы должны быть инвариантны относительно поворотов (собственных ортогональных преобразований). Таким образом, законы физики должны сохранять инвариантность их формы при преобразованиях Лоренца, или, другими словами, инвариантность физического закона относительно поворота требует аналогичности («ко»!) или

ковариантности выражающего его уравнения изменению базиса векторного пространства.

В классической механике понятие силового поля является лишь некоторым способом описания физического явления — взаимодействия частиц. Вместо того, чтобы говорить о том, что одна частица действует на другую, можно сказать, что частица создаёт вокруг себя поле; на всякую другую частицу, находящуюся в этом поле, действует некоторая сила.

В теории же относительности благодаря конечности скорости распространения взаимодействия ситуация существенным образом меняется. Изменение положения одной из частиц отражается на других частицах лишь спустя некоторый промежуток времени, т.е. мы не можем говорить о непосредственном взаимодействии. Это значит, что поле само по себе становится физической реальностью, и мы должны говорить о взамодействии одной частицы с полем и о последующем взаимодействии поля с другой частицей.

Принцип относительности, включающий явно все электродинамические и оптические явления, был, по-видимому, впервые введен Анри Пуанкаре начиная с 1889 года, когда им впервые высказано предположение о принципиальной ненаблюдаемости движения относительно эфира. Вывод преобразований Лоренца (1904) в статье Лоренца взамен преобразований Галилея содержал некоторые технические ошибки, также исправленные Пуанкаре. Дальнейшее развитие принцип получил в статье Пуанкаре «О динамике электрона» (1905).

Двадцатишестилетний скромный чиновник патентного бюро из швейцарского города Берна Альберт Эйнштейн послал в немецкий журнал Annalen der Physik свою статью «К электродинамике движущихся тел» в июне 1905 г. Эта статья не содержала ни одной формулы, но, перевернув последнюю страницу рукописи Эйнштейна, редактор журнала Макс Планк понял, что общепринятые научные представления низвергнуты, и возникла новая физическая дисциплина — *релятивистская механика* (от лат. relativus — относительный), которая рассматривает законы движения при скоростях, близких к скорости света. Смысл работы Эйнштейна заключался в том, что представление о конечности скорости распространения взаимодействий было добавлено к принципу относительности Галилея, исходящего из бесконечной скорости их распространения.

Математические выкладки для новой теории были сделаны не Эйнштейном, а немецким математиком Германом Минковским, разработавшим до этого геометрическую теорию чисел. В 1907–1909 годах Минковский выступил с рядом статей и лекций, где предложил так называемую «геометродинамику» — четырёхмерную математическую модель релятивистской механики. В 1909 году вышла его книга «Пространство и время», которой суждено было стать его научным завещанием. Минковский ввёл четырёхмерное псевдоевклидово пространство (известное сейчас как пространство Минковского).

Название дисциплины *специальная теория относительности* возникло в результате неудачного буквального перевода немецкого слова speziell — частный, когда стало ясно, что релятивистская механика рассматривает

частный случай — свойства пространства–времени в областях, где полями тяготения можно с желаемой точностью пренебречь.

Глава 8

Электродинамическая система

8.1 Первоначальные определения

Электродинамика — раздел физики, изучающий взаимодействие объектов, имеющих электрический заряд, в электромагнитном поле.

Таким образом, основной величиной, характеризующей свойства вещества в изучаемой системе, является электрический заряд, а в качестве векторного поля изучаемой системы выступает электромагнитное поле.

Запись формул электродинамики зависит (нормировочными коэффициентами) от выбранной системы мер. Мы будем пользоваться так называемой гауссовой абсолютной системой единиц (СГС).

Электрический заряд — это связанное с телом свойство, позволяющее ему быть источником электромагнитного поля и участвовать в электромагнитных взаимодействиях. В 1729 году Шарль Дюфе установил, что существует два рода зарядов. Один образуется при трении стекла о шёлк, а другой — смолы о шерсть. Поэтому Дюфе назвал заряды «стеклянным» и «смоляным». Чисто условное понятие о положительном (стеклянном) и отрицательном (смоляном) заряде ввёл Бенджамин Франклин. Заряды одного знака отталкиваются друг от друга, заряды разных знаков притягиваются создаваемым ими полем.

Под *свободными зарядами* понимают, во-первых, все электрические заряды, которые под влиянием электромагнитного поля могут перемещаться на макроскопические расстояния в пространстве (электроны в металлах, ионы в газах и электролитах и т.д.) и, во-вторых, заряды, нанесённые извне на поверхность, нарушающие её нейтральность и способные перемещаться по поверхности на макроскопические расстояния. Поэтому в зависимости от конфигурации пространства электрической системы в ней могут присутствовать объёмные и поверхностные заряды.

Объёмные заряды характеризуются *объёмной плотностью* ρ, которая

в трёхмерном евклидовом пространстве является скаляром (представляет собой скалярное поле).

Поверхностные заряды характеризуются *поверхностной плотностью* ϱ, которая на двумерной евклидовой поверхности является вектором, направленным по нормали к рассматриваемой точке поверхности.

Если ρ и ϱ не зависят от времени t явно, то распределение зарядов в пространстве остаётся стационарным. Такие системы являются предметом *электростатики* — учения о взаимодействии неподвижных электрических зарядов.

Интегральной характеристикой электрического заряда (аналогично массе (7.27) $m = \iiint_V \rho\, dV$) является *полный заряд* q:

$$q = \iiint\limits_V \rho\, dV + \iint\limits_S \varrho\, d\sigma. \tag{8.1}$$

Для системы из k *точечных зарядов* $e_1, \ldots, e_k$ плотность заряда записывается при помощи δ-функции Дирака:

$$\rho(\mathbf{r}) \overset{(7.9)}{=} \sum_{i=1}^{k} e_i\, \delta(\mathbf{r} - \mathbf{r}_i). \tag{8.2}$$

Электрический ток — направленное движение электрических зарядов, хотя более корректно говорить о движении носителей электрических зарядов (электронов, ионов, дырок и т.д.). Таким образом, электрический ток — это движущийся заряд, который характеризуется *плотностью тока* в данной точке пространства:

$$\mathbf{j} = \rho\, \mathbf{v}, \tag{8.3}$$

где $\mathbf{v}$ — *скорость* свободного заряда.

Плотность тока — это поток заряженного вещества.

Записав для этого потока уравнение неразрывности (4.65) мы получаем *закон сохранения количества электричества*:

$$\operatorname{div} \mathbf{j} + \frac{\partial \rho}{\partial t} = 0 \quad \text{или} \quad \oiint\limits_S \mathbf{j}\, d\sigma = -\frac{\partial}{\partial t} \iiint\limits_V \rho\, dV. \tag{8.4}$$

Интегральной характеристикой электрического тока является *сила тока* J, определяемая как поток (4.57) вектора $\mathbf{j}$ через поверхность S:

$$J = \iint\limits_S \mathbf{j}\, d\sigma \overset{(4.65)}{=} -\frac{\partial q}{\partial t}. \tag{8.5}$$

Поэтому количественно плотность тока $\mathbf{j}$ — это количество электрического заряда, протекающее в единицу времени через единицу перпендикулярного току сечения проводника.

В электродинамике используется понятие *линейного тока*, размеры любого сечения S которого достаточно малы по сравнению с расстоянием от

этого сечения до рассматриваемых точек поля. При переходе от рассмотрения токов конечного сечения к линейным токам и обратно уравнения с интегрированием плотности тока по объёму всегда можно заменить уравнениями с интегрированием линейного тока по длине:

$$J \oint\limits_C \Psi \, d\mathbf{r} \Leftrightarrow \iiint\limits_V \Psi \, \mathbf{j} \, dV, \tag{8.6}$$

где Ψ может быть любой скалярной или векторной функцией точки.

Под плотностью $\mathbf{i}$ *поверхностных токов*, в отличие от плотности $\mathbf{j}$ токов объёмных, понимают количество электричества, протекающего в единицу времени через единицу длины перпендикулярного току отрезка, расположенного на поверхности, по которой течёт ток. Соответственно, при переходе от объёмных токов к поверхностным и обратно уравнения с интегрированием плотности тока по объёму всегда можно заменить уравнениями с интегрированием тока по поверхности:

$$\iint\limits_S \Psi \, \mathbf{i} \, d\sigma \Leftrightarrow \iiint\limits_V \Psi \, \mathbf{j} \, dV, \tag{8.7}$$

где Ψ может быть любой скалярной или векторной функцией точки.

Электромагнитное поле имеет две составляющих — электрическое поле и магнитное поле. Вектор электрического поля называют *напряжённостью* электрического поля и обозначают $\mathbf{E}$. Вектор магнитного поля называют *напряжённостью* магнитного поля и обозначают $\mathbf{H}$.

Электрическое поле — векторное поле, действующее на заряды независимо от их скоростей. *Магнитное поле* — векторное поле, действующее только на движущиеся заряды (токи) и на тела, обладающие магнитным моментом (см. 8.19, 8.20). На неподвижные заряды магнитное поле не действует. Термин магнитное поле введён М. Фарадеем (1845).

Область пространства с определёнными свойствами принято называть *средой*. Обычно рассматривают пространственно однородные среды. Это означает, что в пределах одной среды вещество находится в одном физическом состоянии.

Обычно в электродинамике различают четыре основных вида среды: вакуум, проводник, диэлектрик и магнетик.

Вакуум (от лат. vacuum — пустота) — в рассматриваемом случае, среда, содержащая только точечные объекты рассматриваемой электродинамической системы.

Проводник — вещество, проводящее *электрический ток*. Возникновение электрического тока обусловлено действием электрического поля на имеющиеся в проводнике *свободные* электрические заряды. Возникновение свободных зарядов в проводнике объясняется тем, что часть электронов, входящих в состав его атомов, отщепляется от этих атомов.

Плотность тока $\mathbf{j}$ в данной точке проводника будет пропорциональна напряжённости поля $\mathbf{E}$ в этой точке:

$$\mathbf{j} = \phi \, \mathbf{E}, \tag{8.8}$$

где ϕ — *электропроводность* (*удельная проводимость*) проводника,
$\lambda = \frac{1}{\phi}$ — *удельное сопротивление* проводника.

Если в какой-либо точке внутри проводника напряжённость внешнего (по отношению к зарядам проводника) электрического поля $\mathbf{E}$ будет отлична от нуля, то согласно (8.8) в проводнике возникнет электрический ток, т.е. движение свободных зарядов. Это движение будет продолжаться до тех пор, пока поле собственных зарядов проводника, перераспределившихся по объёму проводника, не скомпенсирует внешнего поля. При этом в одной части проводника оказываются заряды одного знака, а в другой — противоположного знака.

Заряды одного знака, благодаря взаимному отталкиванию, сосредотачиваются в тонком поверхностном слое, который можно считать бесконечно тонким. При этом на поверхности проводников существуют некоторые силы неэлектростатического происхождения, препятствующие выходу свободных зарядов за пределы поверхности проводника.

Поэтому в состоянии *электростатического равновесия* — с постоянной конфигурацией электрических зарядов и поля — электрическое поле внутри проводника равно нулю, а заряды проводника сосредоточены в его поверхностном слое.

В уравнении (8.8) к напряжённости электрического поля нужно добавить напряжённость поля сил неэлектростатического происхождения $\mathbf{E}^{\text{стр}}$, обусловленных химической и физической неоднородностью проводников. Таковы силы, возникающие при соприкосновении проводников различного химического состава (гальванический элемент, аккумулятор) или различной температуры (термоэлемент), при наличии градиента концентрации в растворе электролита (концентрационный гальванический элемент) и т.д.

$$\mathbf{j} = \phi\,(\mathbf{E} + \mathbf{E}^{\text{стр}})\,. \tag{8.9}$$

Диэлектрик — вещество, плохо проводящее или совсем не проводящее электрический ток (не содержащее свободные заряды).

В диэлектрике под воздействием электрического поля заряды не срываются, как в проводнике, полем со своих мест, а лишь смещаются в противоположные стороны — молекула деформируется. Говорят, что происходит *поляризация* диэлектрика.

Поэтому, чтобы определить воздействие поля на диэлектрик, нужно прежде всего найти удобную количественную характеристику распределения зарядов в любой, в целом нейтральной системе. Такой характеристикой нейтральной системы зарядов служит *электрический момент* этой системы, который вычисляется аналогично умноженному на массу радиус-вектору механической системы (7.28) $\mathbf{r} = \frac{1}{m}\iiint_V \rho(\mathbf{r})\,\mathbf{r}\,dV$ по формуле:

$$\mathbf{d} = \iiint\limits_V \rho(\mathbf{r})\,\mathbf{r}\,dV. \tag{8.10}$$

Система предполагается в целом электронейтральной

$$\iiint\limits_V \rho(\mathbf{r})\,dV = 0, \tag{8.11}$$

ибо лишь при этом условии значение вектора $\mathbf{d}$ однозначно определяется распределением зарядов ρ и не зависит от выбора начала координат.

Действительно, если сместить начало координат на произвольный отрезок $\mathbf{a}$, то новый радиус-вектор имеет значение $\mathbf{r}' = \mathbf{r} - \mathbf{a}$, тогда

$$\iiint\limits_V \rho\,\mathbf{r}'\,dV = \iiint\limits_V \rho\,(\mathbf{r}-\mathbf{a})\,dV = \iiint\limits_V \rho\,\mathbf{r}\,dV - \mathbf{a}\iiint\limits_V \rho\,dV \overset{(8.11)}{=} \iiint\limits_V \rho\,\mathbf{r}\,dV.$$

Смещённые заряды электронейтральной молекулы можно рассматривать как *диполь* — идеализированную электронейтральную систему, состоящую из точечных и равных по абсолютной величине положительного и отрицательного электрических зарядов e и $-e$, которые находятся на расстоянии $r_{\pm e}$ друг от друга, причём вектор $\mathbf{r}_{\pm e}$ направлен от отрицательного заряда к положительному (рис. 8.1).

Для в целом электронейтральной системы из $2k$ точечных зарядов $e_1, -e_1, \ldots, e_k, -e_k$ электрический момент вычисляется по формуле

$$\mathbf{d} \overset{(8.2)}{\underset{(4.110)}{=}} \sum_{i=1}^{k} \{e_i\,\mathbf{r}_{e_i} - e_i\,\mathbf{r}_{-e_i}\}. \tag{8.12}$$

Такие заряды нейтральных молекул и ионов твёрдых диэлектриков, закреплённые вблизи определённых положений равновесия и смещённые под действием электрического поля, называют *связанными зарядами*.

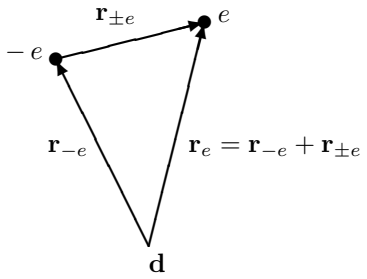


Рис. 8.1: Диполь

Для связанных зарядов используют индекс $_{\text{связ}}$, а для свободных зарядов индекса не используют.

При рассмотрении поверхностей раздела между различными диэлектриками и различнми средами (на границе двух разных сред с разными величинами диэлектрической проницаемости) помимо объемных возникают поверхностные связанные заряды. Поэтому связанные заряды, по аналогии со свободными зарядами, разделяют на объёмные заряды с плотностью $\rho_{\text{связ}}$ и поверхностные заряды с плотностью $\varrho_{\text{связ}}$.

При этом заряды, нанесённые извне на поверхность диэлектриков и нарушающие их нейтральность, являются свободными зарядами.

Электрический момент диполя (рис. 8.1) равен:

$$\mathbf{d}_{\pm e} = e\,\mathbf{r}_{\pm e}. \tag{8.13}$$

Доказательство

$$\mathbf{d}_{\pm e} \overset{(8.12)}{=} e\,\mathbf{r}_e - e\,\mathbf{r}_{-e} \overset{\text{рис. 8.1}}{=} e(\mathbf{r}_{-e} + \mathbf{r}_{\pm e}) - e\,\mathbf{r}_{-e} = e\,\mathbf{r}_{\pm e}. \tag{8.14}$$

Поле поляризованных диэлектриков, как и всякое электрическое поле, создаётся возникшими в диэлектрике диполями, поэтому аналогично системе диполей диэлектрик характеризуется своим электрическим моментом $\mathbf{d}$.

Поляризацией диэлектрика $\mathcal{P} = \frac{\partial \mathbf{d}}{\partial V}$ называется электрический момент единицы его объёма.

Для большинства диэлектриков уравнение состояния диэлектрика (соотношение, связывающее между собой поляризацию диэлектрика и напряжённость внешнего электрического поля) записывается в виде

$$\mathcal{P} = \alpha \mathbf{E}, \tag{8.15}$$

где α — *удельная диэлектрическая восприимчивость*.

Электрическое поле в среде диэлектрика обычно характеризуется двумя истинными или полярными векторами — *электрической индукцией* $\mathbf{D}$ и напряжённостью электрического поля $\mathbf{E}$.

Это различие возникает именно из-за среды. В среде без диэлектриков (в вакууме) $\mathbf{D} = \mathbf{E}$.

Электрическая индукция складывается из напряжённости электрического поля и поляризации с коэффициентом размерности 4π:

$$\mathbf{D} = \mathbf{E} + 4\pi\mathcal{P}. \tag{8.16}$$

Очень часто, для практического удобства в расчётах диэлектриков диэлектрическая восприимчивость выражается через *диэлектрическую проницаемость* среды ε

$$\varepsilon = 1 + 4\pi\alpha, \quad \text{тогда} \quad \mathbf{D} = \varepsilon \mathbf{E}. \tag{8.17}$$

Магнетик — вещество, способное намагничиваться под действием магнитного поля.

Подобно тому, как внесение диэлектриков в поле свободных электрических зарядов вызывает изменение этого поля, обуславливающееся поляризацией диэлектрика, так и внесение магнетиков в магнитное поле токов вызывает изменение этого поля, обусловливаемое *намагничиванием* магнетика. Под действием магнитного поля в магнетике возникает циркуляция молекулярных токов, хотя плотность тока в такой среде равна нулю, и никакого переноса заряда на макроскопические расстояния не происходит.

Но если диэлектрики деполяризуются при исчезновении внешнего электрического поля, то по исчезновении магнитного поля размагничивается лишь часть магнетиков. Существует класс магнетиков, так называемые *ферромагнетики* или постоянные магниты, способные оставаться намагниченными и после исчезновения внешнего магнитного поля.

Чтобы определить воздействие магнитного поля на магнетик нужно, подобно диэлектрикам, прежде всего найти удобную количественную характеристику распределения токов.

Магнитный момент $\mathbf{m}$ системы токов вычисляется аналогично кинетическому моменту механической системы (7.42) $\mathbf{L} = \iiint_V (\mathbf{r} \times \mathbf{j})\, dV$. Запишем формулы магнитного момента для случая точечных зарядов, когда

замкнутый электрический ток плотности $\mathbf{j}$ течёт в ограниченном объёме V (который соответствует в том числе и магнетику), а также и для случая элемента тока $J\,d\mathbf{r}$, соответствующего участку $d\mathbf{r}$ тонкого замкнутого проводника C с постоянной силой тока J:

$$\mathbf{m} = \frac{1}{2\,c}\sum_{i=1}^{n} e_i(\mathbf{r}_i \times \mathbf{v}_i), \tag{8.18}$$

$$\mathbf{m} = \frac{1}{2\,c}\iiint\limits_V (\mathbf{r}\times\mathbf{j})\,dV, \tag{8.19}$$

$$\mathbf{m} \overset{(8.6)}{=} \frac{J}{2\,c}\oint\limits_C \mathbf{r}\times d\mathbf{r}. \tag{8.20}$$

Таким образом, магнитный момент — это аксиальный вектор.

Поле намагниченных магнетиков, как и всякое магнитное поле, создаётся циркулирующими в магнетике электрическими токами, поэтому аналогично системе токов магнетик характеризуется своим магнитным моментом.

Намагниченностью магнетика $\mathcal{M} = \frac{\partial \mathbf{m}}{\partial V}$ называется магнитный момент единицы его объёма.

Магнитное поле в среде магнетика обычно характеризуется двумя псевдовекторами или аксиальными векторами — магнитной индукцией $\mathbf{B}$ и напряжённостью магнитного поля $\mathbf{H}$.

Это различие возникает именно из-за среды. В среде без магнетиков (в вакууме) $\mathbf{B} = \mathbf{H}$.

Магнитная индукция $\mathbf{B}$ складывается из напряжённости магнитного поля и намагниченности с коэффициентом размерности 4π:

$$\mathbf{B} = \mathbf{H} + 4\pi\mathcal{M}. \tag{8.21}$$

В большинстве неферромагнитных сред уравнение состояния магнетика (соотношение, связывающее между собой намагниченность магнетика и напряжённость внешнего магнитного поля) записывается в виде

$$\mathcal{M} = \chi\mathbf{H}, \tag{8.22}$$

где χ — *удельная магнитная восприимчивость*.

Очень часто, для практического удобства в расчётах магнетиков магнитная восприимчивость выражается через *магнитную проницаемость* среды μ

$$\mu = 1 + 4\pi\chi, \quad \text{тогда} \quad \mathbf{B} = \mu\mathbf{H}. \tag{8.23}$$

Характеризующие свойства среды величины ε, μ и ϕ считаются заданными функциями точки, от времени не зависящими, сторонние же силы среды $\mathbf{E}^{\text{стр}}$ считаются заданными функциями точки и времени.

8.2 Потенциалы поля

Экспериментально установлено, что свойства частицы в отношении её взаимодействия с электромагнитным полем определяются всего одним параметром — её *зарядом* e. Свойства же поля характеризуются двумя величинами, скалярной и векторной: скалярную величину φ называют *скалярным потенциалом поля*, а векторную величину $\mathbf{A} = (A_x, A_y, A_z)$, называют *векторным потенциалом поля*.

Скалярный потенциал φ создают имеющиеся в системе скалярные величины — заряды, а векторный потенциал $\mathbf{A}$ создают имеющиеся в системе векторные величины — токи. При этом вычисление потенциалов как одиночного заряда e в вакууме ($\varepsilon = \mu = 1$), так и распределённых зарядов и токов в электромагнитной среде нужно производить в соответствии с выражениями (4.90, 4.100):

$$\varphi = \frac{e}{r} \qquad \text{или} \qquad \varepsilon\varphi = \iiint\limits_V \frac{\rho}{r}\, dV + \iint\limits_S \frac{\varrho}{r}\, d\sigma; \tag{8.24}$$

$$\mathbf{A} = \frac{e\,\mathbf{v}}{c\,r} \qquad \text{или} \qquad \mathbf{A} = \frac{\mu\, J}{c} \oint\limits_C \frac{d\mathbf{r}}{r}, \quad \mathbf{A} = \frac{\mu}{c} \iiint\limits_V \frac{\mathbf{j}}{r}\, dV; \tag{8.25}$$

где c — скорость света, нормировочный коэффициент, введённый, чтобы компенсировать размерность скорости в уравнении $\mathbf{j} = \rho\,\mathbf{v}$;
$\mathbf{r}$ — радиус-вектор, построенный из точки нахождения заряда или тока, до рассматриваемой точки пространства, $r = |\mathbf{r}|$.

Потенциалы связаны следующим соотношением:

$$\operatorname{div} \mathbf{A} = -\frac{\varepsilon\mu}{c} \frac{\partial \varphi}{\partial t}. \tag{8.26}$$

Доказательство

$$\begin{aligned}
\operatorname{div} \mathbf{A} &\overset{(8.25)}{=} \frac{\mu}{c} \iiint\limits_V \operatorname{div}\left(\frac{\mathbf{j}}{r}\right) dV = \\
&\overset{(3.178)}{=} \frac{\mu}{c} \iiint\limits_V \frac{\operatorname{div}\mathbf{j}}{r}\, dV + \frac{\mu}{c} \iiint\limits_V \mathbf{j} \operatorname{grad}\left(\frac{1}{r}\right) dV = \\
&\overset{(8.4)}{\underset{(3.172)}{=}} -\frac{\mu}{c} \frac{\partial}{\partial t} \iiint\limits_V \frac{\rho}{r}\, dV - \frac{\mu}{c} \iiint\limits_V \frac{\mathbf{j}\cdot\mathbf{r}}{r^3}\, dV \overset{(8.24)}{\underset{(4.93)}{=}} -\frac{\varepsilon\mu}{c} \frac{\partial \varphi}{\partial t}.
\end{aligned}$$

Важность понятия электрического момента системы зарядов $\mathbf{d}$ обуславливается тем, что система неподвижных зарядов при условии достаточной малости её размеров и потенциал φ поля, возбуждаемого произвольной, в целом нейтральной системой зарядов на расстояниях больших по сравнению с размерами этой системы, определяется её электрическим моментом. Аналогично электрическому моменту, важность понятия магнитного момента $\mathbf{m}$ обуславливается тем, что система замкнутых токов при условии

достаточной малости её размеров и векторный потенциал $\mathbf{A}$ поля, возбуждаемого произвольной системой замкнутых токов на расстояниях больших по сравнению с размерами этой системы, однозначно характеризуется её магнитным моментом.

Для обоснования этих утверждений введём систему координат с началом где-нибудь внутри системы неподвижных зарядов e_i или токов $e_i\,\mathbf{v}_i$. Обозначим радиус-векторы отдельных зарядов как $\mathbf{r}_i$, а радиус-вектор точки, в которой мы ищем потенциал, как $\mathbf{R}$.

Скалярный потенциал неподвижных зарядов e_i диэлектрика с электрическим моментом $\mathbf{d}$, на расстояниях $R = |\mathbf{R}| \gg |\mathbf{r}_i|$, больших по сравнению с размерами системы, равен

$$\varphi \overset{(8.24)}{=} \sum_{i=1}^{2n} \frac{e_i}{|\mathbf{R}-\mathbf{r}_i|} \approx -\mathbf{d}\cdot\operatorname{grad}\left(\frac{1}{R}\right). \tag{8.27}$$

Доказательство. С учётом $\sum_{i=1}^{2n} e_i = 0$, $\mathbf{d} = \sum_{i=1}^{2n} e_i\,\mathbf{r}_i$ получаем

$$\sum_{i=1}^{2n} \frac{e_i}{|\mathbf{R}-\mathbf{r}_i|} \overset{(1.33)}{=} \frac{\sum_{i=1}^{2n} e_i}{R} + \sum_{i=1}^{2n} e_i\,\mathbf{r}_i\,\frac{\mathbf{R}}{R^3} + \ldots \overset{(3.172)}{\approx} -\mathbf{d}\cdot\operatorname{grad}\left(\frac{1}{R}\right).$$

Векторный потенциал системы стационарных токов магнетика с магнитным моментом $\mathbf{m}$, на расстояниях больших по сравнению с размерами системы равен

$$\mathbf{A} \overset{(8.25)}{=} \frac{1}{c}\sum_{i=1}^{n} \frac{e_i\,\mathbf{v}_i}{|\mathbf{R}-\mathbf{r}_i|} \approx -\mathbf{m}\times\operatorname{grad}\left(\frac{1}{R}\right). \tag{8.28}$$

Доказательство. Если рассмотреть систему стационарно движущихся зарядов, то радиус-вектор каждого заряда относительно начала координат при усреднении за достаточно большой промежуток времени можно считать постоянным ($\overline{\mathbf{r}}_i = const$). Тогда

$$\sum_{i=1}^{n} e_i\,\overline{\mathbf{v}}_i = \frac{d}{dt}\sum_{i=1}^{n} e_i\,\overline{\mathbf{r}}_i = 0, \tag{8.29}$$

$$\frac{d}{dt}\sum_{i=1}^{n}\overline{e_i\,\mathbf{r}_i(\mathbf{r}_i\,\mathbf{R})} = \sum_{i=1}^{n}\overline{e_i\,\{\mathbf{v}_i(\mathbf{r}_i\,\mathbf{R}) + \mathbf{r}_i(\mathbf{v}_i\,\mathbf{R})\}} = 0, \tag{8.30}$$

или

$$\overline{e_i\,\mathbf{v}_i(\mathbf{r}_i\,\mathbf{R})} = -\overline{e_i\,\mathbf{r}_i(\mathbf{v}_i\,\mathbf{R})}. \tag{8.31}$$

В результате получаем

$$\frac{1}{c}\sum_{i=1}^{n}\overline{\frac{e_i\,\mathbf{v}_i}{|\mathbf{R}-\mathbf{r}_i|}} \overset{(1.33)}{=} \frac{\sum_{i=1}^{n} e_i\,\overline{\mathbf{v}}_i}{c\,R} + \frac{1}{c}\sum_{i=1}^{n}\overline{\frac{e_i\,\mathbf{v}_i(\mathbf{r}_i\,\mathbf{R})}{R^3}} + \ldots \approx$$

$$\overset{\substack{(8.29)\\(8.31)}}{\approx} \frac{1}{2\,c}\sum_{i=1}^{n}\overline{\frac{e_i\,\{\mathbf{v}_i(\mathbf{r}_i\,\mathbf{R}) - \mathbf{r}_i(\mathbf{v}_i\,\mathbf{R})\}}{R^3}} =$$

$$\overset{\text{(п. 5, с. 86)}}{=} \frac{1}{2\,c}\sum_{i=1}^{n}\overline{e_i(\mathbf{r}_i\times\mathbf{v}_i)}\times\frac{\mathbf{R}}{R^3} \overset{\substack{(8.18)\\(3.172)}}{=} -\mathbf{m}\times\operatorname{grad}\left(\frac{1}{R}\right).$$

8.3 Уравнения Лагранжа в электродинамике

Как уже отмечалось, законы Ньютона — не самый глубокий уровень формулирования классической механики. В рамках лагранжевой механики имеется одна-единственная формула (запись действия) и один-единственный постулат или принцип наименьшего действия (тела движутся так, чтобы действие было стационарным), и из этого можно вывести все законы Ньютона. Более того, в рамках лагранжева формализма можно рассмотреть гипотетические системы, в которых действие имеет какой-либо другой вид, а объекты имеют другую природу. При этом уравнения движения станут уже непохожими на законы Ньютона, но сама лагранжева механика будет по-прежнему применима.

Одной из таких систем является электродинамика.

Действие для частицы, движущейся в заданном электромагнитном поле, складывается из двух частей: из действия свободной частицы (7.93) $S = -mc\int ds$ и из члена, описывающего взаимодействие частицы с полем. Последний должен содержать как величины, характеризующие частицу, так и величины, характеризующие поле.

Величина ds является 4-вектором, значит, для добавления в действие члена отвечающего за взаимодействие частицы с электромагнитным полем, нам нужно в электродинамике перейти к 4-векторам.

Вместо скалярной величины — плотности заряда и векторной величины — плотности тока можно ввести одну общую величину — 4-вектор тока ξ, как произведение плотности заряда ρ на 4-скорость $\mathbf{u} = (c, \mathbf{v})$ (7.107), где временная компонента есть $c\rho$, а три пространственных компоненты образуют трёхмерную плотность тока $\mathbf{j}$

$$\xi = \rho\,\mathbf{u} = (c\rho, \mathbf{j}). \tag{8.32}$$

Вместо скалярного и векторного потенциалов можно ввести общий 4-*потенциал* $\Phi = (\varphi, A_x, A_y, A_z)$, где φ является временной, A_x, A_y, A_z — пространственными компонентами. Эти четыре компоненты являются функциями координат $\mathbf{r}$ и времени t.

Чтоб учесть влияние электромагнитной среды запишем контрвариантный вектор 4-потенциала как $\Phi^{0,1,2,3} = (\varepsilon\varphi, \frac{\mathbf{A}}{\mu})$, а ковариантный вектор

4-потенциала как $\Phi_{0,1,2,3} = (\varphi, -\mathbf{A})$. Тогда выражения (8.24) и (8.25) для 4-потенциала можно записать в виде

$$\Phi^{0,1,2,3} = \frac{e\,\mathbf{u}}{c\,r} \quad \text{или} \quad \Phi^{0,1,2,3} = \iiint\limits_V \frac{\xi}{c\,r}\,dV. \tag{8.33}$$

Если мы умножим ковариантный вектор 4-потенциала $\Phi_{0,1,2,3} = (\varphi, -\mathbf{A})$ на 4-ток (8.32) или на произведение заряда частицы e на 4-скорость $\mathbf{u} = (c, \mathbf{v})$ (7.107) с коэффициентом $-\frac{1}{c}$, то получим выражение

$$-\frac{e}{c}\,\Phi_i\,u^i = \frac{e}{c}\,\mathbf{A}\cdot\mathbf{v} - e\varphi \quad \text{или} \quad -\frac{\Phi_{0,1,2,3}\,\xi}{c} = \frac{\mathbf{A}\cdot\mathbf{j}}{c} - \rho\,\varphi, \tag{8.34}$$

аналогичное преобразованию Лежандра (5.45) $L = \mathbf{p}\cdot\mathbf{v} - H$, где величина $\frac{e}{c}\,\mathbf{A}$ аналогична механическому импульсу $\mathbf{p}$, а величина энергии заряда в электромагнитном поле $e\varphi$ аналогична механической энергии H.

Выражение (8.34) дополняет действие (7.93), которое можно записать как для точечных зарядов e, так и для распределённых зарядов ρ:

$$S = \int\limits_{t_1}^{t_2} L\,dt = \int\limits_{t_1}^{t_2} \left\{ -mc^2\sqrt{1 - \left(\frac{v}{c}\right)^2} + \frac{e}{c}\,\mathbf{A}\cdot\mathbf{v} - e\varphi \right\} dt, \tag{8.35}$$

$$S \overset{(7.93)}{\underset{(7.103)}{=}} -\int \left\{ mc\,ds + \frac{\rho}{c}\Phi_i\,u^i\,dV\,dt \right\} \overset{(8.32)}{=} -\int \left\{ mc\,ds + \frac{\Phi_i\,\xi^i}{c}\,d\Omega \right\}, \tag{8.36}$$

где $\mathbf{A}, \varphi$ — берутся в точках мировой линии,

$dV\,d(ct) = d\Omega \overset{(3.98)}{=} dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3$ — элемент объёма четырёхмерного пространства–времени.

При этом

$$\frac{\partial L}{\partial \mathbf{v}} \overset{(7.95)}{=} \mathbf{p} + \frac{e}{c}\,\mathbf{A}, \tag{8.37}$$

$$E \overset{(7.97)}{=} \sqrt{p^2c^2 + m^2c^4} + e\varphi. \tag{8.38}$$

Заряд, находящийся в поле, не только подвергается воздействию со стороны поля, но в свою очередь сам влияет на поле, изменяя его. Однако если заряд e не велик, то его действием на заданное поле можно пренебречь и считать, что поле не зависит от заряда.

Уравнения движения заряда в заданном электромагнитном поле вычисляются из уравнений Эйлера–Лагранжа (5.47):

$$\dot{\mathbf{p}} = -\frac{e}{c}\frac{\partial \mathbf{A}}{\partial t} - e\,\operatorname{grad}\varphi + \frac{e}{c}\,(\mathbf{v}\times\operatorname{rot}\mathbf{A})\,. \tag{8.39}$$

Доказательство. Так как дифференцирование по $\mathbf{r}$ производится при

постоянном $\mathbf{v}$, находим:

$$\frac{d}{dt}\frac{\partial L}{\partial \mathbf{v}} \overset{(8.37)}{\underset{(3.106)}{=}} \dot{\mathbf{p}} + \frac{e}{c}(\mathbf{v}\nabla)\,\mathbf{A} + \frac{e}{c}\frac{\partial \mathbf{A}}{\partial t} = \overset{(5.47)}{=} \frac{\partial L}{\partial \mathbf{r}} \overset{(8.35)}{=} \frac{e}{c}\operatorname{grad}(\mathbf{A}\cdot\mathbf{v}) - e\operatorname{grad}\varphi = \overset{(3.180)}{=} \frac{e}{c}(\mathbf{v}\nabla)\,\mathbf{A} + \frac{e}{c}(\mathbf{v}\times\operatorname{rot}\mathbf{A}) - e\operatorname{grad}\varphi.$$

Первый и второй члены в правой части (8.39) не зависят от скорости частицы и могут быть интерпретированы как электрическое поле, действующее на заряд e:

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \operatorname{grad}\varphi. \tag{8.40}$$

Третий член в правой части (8.39) зависит от скорости частицы — пропорционален величине скорости и перпендикулярен к ней. Этот член может быть интерпретирован как магнитное поле (точнее, магнитная индукция) действующее на движущийся со скоростью $\mathbf{v}$ заряд e:

$$\mathbf{B} = \operatorname{rot}\mathbf{A}, \tag{8.41}$$

Обратим внимание, что согласно этим определениям $\mathbf{E}$ — *полярный* вектор, а $\mathbf{B}$ — *аксиальный* вектор.

В левой части уравнения (8.39) стоит производная от импульса частицы по времени. Следовательно, для консервативной системы выражение в правой части есть сила $\mathbf{F} = \dot{\mathbf{p}}$, действующая на заряд в заданном электромагнитном поле, которую называют *силой Лоренца*.

В результате сила Лоренца — сила воздействия электрического поля $\mathbf{E}$ на неподвижный точечный заряд e и магнитной индукции $\mathbf{B}$ на точечный заряд e, движущийся со скоростью $\mathbf{v}$ (относительно системы координат, в которой вычисляются величины $\mathbf{F}$, $\mathbf{E}$ и $\mathbf{B}$), а также на *распределённые заряды и токи*, определяется следующими уравнениями:

$$\mathbf{F} = e\,\mathbf{E} + \frac{e}{c}(\mathbf{v}\times\mathbf{B}), \quad \mathbf{F} = \iiint\limits_V \left\{\rho\,\mathbf{E} + \frac{1}{c}(\mathbf{j}\times\mathbf{B})\right\} dV. \tag{8.42}$$

Выражение для силы Лоренца было получено Х. А. Лоренцом путём обобщения опытных данных в конце 19 в. Из уравнений (8.42) видно, что различие между электрическим и магнитным полями построено на различии тока и заряда: электрическое поле воздействует на любые заряды, а магнитное — только на движущиеся.

Если заданы потенциалы $\mathbf{A}$ и φ, то этим, согласно (8.40) и (8.41), вполне однозначно определены $\mathbf{E}$ и $\mathbf{H}$, а значит и электромагнитное поле. При этом следует понимать, что скалярный и векторный потенциалы поля являются лишь вспомогательными понятиями, и что непосредственный физический

смысл имеют только напряжённости поля, так как именно они входят в выражение для силы Лоренца (8.42). Поэтому два поля с одними и теми же значениями напряжённостей, но разными значениями потенциалов, физически тождественны.

Выпишем силы воздействия на заряд электрического и магнитного полей по отдельности:

$$\mathbf{F} = e\,\mathbf{E}, \qquad \mathbf{F} = \iiint\limits_V \rho\,\mathbf{E}\,dV; \tag{8.43}$$

$$\mathbf{F} = \frac{e}{c}(\mathbf{v}\times\mathbf{B}), \qquad \mathbf{F} = \frac{1}{c}\iiint\limits_V (\mathbf{j}\times\mathbf{B})\,dV, \qquad \mathbf{F} = \frac{J}{c}\oint\limits_C d\mathbf{r}\times\mathbf{B}. \tag{8.44}$$

Выражения (8.44) называют *закон Ампера* — закон взаимодействия постоянных токов, определяющий силу, с которой магнитное поле действует движущийся заряд. Направление вектора силы определяется по правилу вычисления векторного произведения. Закон установлен Андре Мари Ампером в 1820 году. Из закона Ампера следует, что параллельные проводники с постоянными токами, текущими в одном направлении, притягиваются, а в противоположном — отталкиваются.

Рассмотрим теперь вопрос о том, насколько однозначно определены потенциалы поля. Возможный произвол в определении потенциалов следует из уравнения (3.73) $\operatorname{rot}(\operatorname{grad}\psi) = 0$. Это означает, что если мы прибавим к $\mathbf{A}$ градиент произвольного скаляра ψ:

$$\mathbf{A}' = \mathbf{A} + \operatorname{grad}\psi, \tag{8.45}$$

то новому значению векторного потенциала будет соответствовать прежнее значение магнитной индукции $\mathbf{B} = \operatorname{rot}\mathbf{A}$. Если при этом ψ не зависит от времени, то значение $\mathbf{E}$ не изменится, а если зависит, то из (8.40) $\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t} - \operatorname{grad}\varphi$ следует, что при замене $\mathbf{A}$ на $\mathbf{A}'$ также необходимо заменить φ на:

$$\varphi' = \varphi - \frac{1}{c}\frac{\partial\psi}{\partial t}, \tag{8.46}$$

Легко убедиться в том, что при подстановке в уравнения (8.40) и (8.41) вместо $\mathbf{A}$ и φ потенциалов $\mathbf{A}'$ и φ' электрическое и магнитное поля действительно не изменятся.

При такой замене в интеграле действия (8.35) появиться дополнительный член, представляющий собой полный дифференциал

$$\left(\frac{e}{c}\operatorname{grad}\psi\cdot\mathbf{v} + \frac{e}{c}\frac{\partial\psi}{\partial t}\right)dt \overset{(3.104)}{=} d\left(\frac{e}{c}\,\psi\right), \tag{8.47}$$

что также не изменит уравнения движения, так как на варьирование (5.77) полный дифференциал по определению никак не влияет.

Таким образом, записанное преобразование потенциалов не изменяет поля. Следовательно, потенциалы определены не однозначно — векторный определён с точностью до градиента произвольной функции и скалярный

— с точностью до производной по времени от той же функции. В частности, к векторному потенциалу можно добавить любой постоянный вектор, а к скалярному потенциалу — любую постоянную.

Инвариантность поля по отношению к этому классу преобразований называется *калибровочной* или *градиентной инвариантностью*.

Описанная неоднозначность потенциалов даёт всегда возможность выбрать их так, чтобы они удовлетворяли одному произвольному дополнительному условию, — одному, так как мы можем произвольно выбрать одну функцию ψ. Обычно скалярный потенциал выбирают равным нулю на бесконечности. Сделать же векторный потенциал равным нулю, вообще говоря невозможно, так как для его трёх компонент понадобиться выполнение трёх дополнительных условий.

Действие (8.35) для заряда в заданном электромагнитном поле перепишем в следующем виде

$$S \overset{\substack{(7.93)\\(7.103)}}{=} -\int \left\{ mc\, ds + \frac{e}{c}\Phi_j\, dx^j \right\}. \tag{8.48}$$

При варьировании экстремали по координатам δx^i получаем

$$\delta S \overset{(3.63)}{=} -\int \left\{ mc\, \delta(ds) + \frac{e}{c} \sum_{i,j=1}^{n} \left(\frac{\partial \Phi_j}{\partial x^i} - \frac{\partial \Phi_i}{\partial x^j} \right) \delta x^i \wedge dx^j \right\} = 0. \tag{8.49}$$

Компоненты $\omega_{ij} = \frac{\partial \Phi_j}{\partial x^i} - \frac{\partial \Phi_i}{\partial x^j} = -\omega_{ji}$ второго члена запишем в виде кососимметрической матрицы (3.64)

$$\|\omega_{ij}\| \overset{\substack{(8.40)\\(8.41)}}{=} \begin{Vmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{Vmatrix}, \tag{8.50}$$

которая называется *тензор электромагнитного поля*.

Таким образом, если рассматривать 4-потенциал как 1-форму на пространстве–времени, то тензор электромагнитного поля есть полученная внешним дифференцированием потенциала 2-форма

$$\omega \overset{(3.58)}{=} \frac{\partial \Phi}{\partial x^i}\, dx^i = d\Phi \quad \text{или} \quad \omega_{ij}\, dx^j = d\Phi_i. \tag{8.51}$$

Тензор электромагнитного поля (8.50), полученный дифференцированием ковариантных компонент 4-потенциала, является *ковариантным* тензором, составленным из компонент полярного вектора $\mathbf{E}$ и аксиального вектора $\mathbf{B}$. При помощи компонент векторов $\mathbf{D} \overset{(8.17)}{=} \varepsilon\mathbf{E}$ и $\mathbf{H} \overset{(8.23)}{=} \frac{\mathbf{B}}{\mu}$ можно записать *контравариантный* тензор зеркального пространства, поменяв знак

компонент полярного вектора $\mathbf{E}$ и оставив неизменным знак компонент аксиального вектора $\mathbf{H}$:

$$\left\|\omega^{ij}\right\| = \begin{Vmatrix} 0 & -D_x & -D_y & -D_z \\ D_x & 0 & -H_z & H_y \\ D_y & H_z & 0 & -H_x \\ D_z & -H_y & H_x & 0 \end{Vmatrix}. \tag{8.52}$$

Запишем зависимость тензоров от двух векторов как поливектор

$$\|\omega_{ij}\| = (\mathbf{E}, \mu\mathbf{H}), \quad \left\|\omega^{ij}\right\| = (-\varepsilon\mathbf{E}, \mathbf{H}). \tag{8.53}$$

Такая запись удобна тем, что позволяет при помощи скалярного и кососкалярного произведения компонент $\mathbf{E}$ и $\mathbf{H}$ легко составить два инварианта, характерных для электромагнитного поля

$$\left(\omega_{ij}, \omega^{ij}\right) \overset{(2.68)}{=} \mu H^2 - \varepsilon E^2 = \text{inv}, \tag{8.54}$$

$$\left[\omega_{ij}, \omega^{ij}\right] \overset{(2.103)}{=} (1 + \mu\varepsilon)\mathbf{E} \cdot \mathbf{H} = \text{inv}. \tag{8.55}$$

Первая из этих величин истинный скаляр, вторая — псевдоскаляр. Псевдоскалярность второй величины очевидна из того, что она представляет собой произведение полярного и аксиального векторов.

Из инвариантности (8.54) следует, что если $E > H$, $E < H$ или $E = H$, то эти соотношения будут сохраняться и во всякой другой системе отсчёта. Из инвариантности (8.55) следует, что если векторы $\mathbf{E}$ и $\mathbf{H}$ образуют острый, прямой или тупой угол, то этот угол будет сохраняться и во всякой другой системе. Когда оба инварианта равны нулю, то в этом случае векторы $\mathbf{E}$ и $\mathbf{H}$ во всех системах отсчёта равны по величине и перпендикулярны друг другу.

8.4 Уравнения Максвелла

Рассмотрим теперь действие S когда координаты, заряды и токи заданы, а изменяется само электромагнитное поле. Однако, так как электромагнитное поле уже не задано, в действие нужно добавить член для электромагнитного поля в отсутствие зарядов.

Оказывается, что таким членом является выражение

$$\alpha \int \omega_{ij}\, \omega^{ij}\, d\Omega. \tag{8.56}$$

где α — нормировочный коэффициент.

В этом случае вместо вариации координат нужно использовать вариацию 4-потенциала $\delta\Phi_i$. Тогда в уравнении (8.36) вариация первого (для

свободной частицы) члена равна нулю, ввиду заданности её траектории. Во втором члене ввиду заданности не варьируется ток

$$\delta S = \delta \int \left\{ \frac{\Phi_i\, \xi^i}{c^2} + \alpha\, \omega_{ij}\, \omega^{ij} \right\} d\Omega = \int \left\{ \frac{\xi^i\, \delta\Phi_i}{c^2} + 2\,\alpha\, \omega_{ij}\, \delta\omega^{ij} \right\} d\Omega = 0.$$

При варьировании во втором члене учтено, что $\omega^{ij}\, \delta\omega_{ij} = \omega_{ij}\, \delta\omega^{ij}$.
В результате получаем следующее выражение

$$\delta S = \int \left\{ \frac{\xi^i}{c^2} + 2\,\alpha\, \frac{\partial \omega^{ij}}{\partial x^j} \right\} d\Omega\, \delta\Phi_i = 0. \tag{8.57}$$

Доказательство

$$\omega_{ij}\, \delta\omega^{ij} = \omega_{ij}\, \frac{\partial \omega^{ij}}{\partial x^j}\, \delta x^j = \frac{\partial \omega^{ij}}{\partial x^j}\, \delta\Phi_i.$$

По смыслу принципа наименьшего действия вариации $\delta\Phi_i$ произвольны, а нулю должен равняться коэффициент при $\delta\Phi_i$, т.е.

$$\frac{1}{2\,\alpha c^2} \xi^i \overset{(8.57)}{=} -\frac{\partial \omega^{ij}}{\partial x^j} \quad (i = 0, 1, 2, 3). \tag{8.58}$$

Чтобы в дальнейшем получить «правильные» коэффициенты в уравнениях (8.91) и (8.90), нормировочный коэффициент α примем равным $\frac{1}{8\pi c}$. Тогда, подставив соответствующие значения ξ^i и ω^{ij} в (8.58)

$$\frac{4\pi}{c}\, \rho = \frac{1}{c}\frac{\partial D_x}{\partial x} + \frac{1}{c}\frac{\partial D_y}{\partial y} + \frac{1}{c}\frac{\partial D_z}{\partial z}, \qquad \frac{4\pi}{c}\, j_x = -\frac{\partial D_x}{\partial t} + \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z},$$
$$\frac{4\pi}{c}\, j_y = -\frac{\partial D_y}{\partial t} - \frac{\partial H_z}{\partial x} + \frac{\partial H_x}{\partial z}, \qquad \frac{4\pi}{c}\, j_z = -\frac{\partial D_z}{\partial t} + \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y};$$

мы получим два уравнения

$$\operatorname{div} \mathbf{D} = 4\pi\rho, \tag{8.59}$$

$$\operatorname{rot} \mathbf{H} = \frac{4\pi}{c}\, \mathbf{j} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}. \tag{8.60}$$

Полученные уравнения не совсем симметричны, поэтому договорились считать, что помимо *токов проводимости* $\mathbf{j}$ в системах нестационарного поля $\left(\frac{\partial \mathbf{D}}{\partial t} \neq 0\right)$ возникают *токи смещения* $\mathbf{j}_{\text{см}}$, которые возбуждают магнитное поле по тем же законам, что и токи проводимости:

$$\operatorname{rot} \mathbf{H} = \frac{4\pi}{c} \left(\mathbf{j} + \mathbf{j}_{\text{см}} \right), \quad \mathbf{j}_{\text{см}} = \frac{1}{4\pi}\frac{\partial \mathbf{D}}{\partial t}. \tag{8.61}$$

Согласно этому определению, ток смещения может иметь место не только в проводниках, но и в диэлектриках и даже в вакууме. Плотность тока смещения в каждой точке поля пропорциональна скорости изменения

вектора электрической индукции $\mathbf{D}$, однако в стационарном поле ток смещения всегда будет равен нулю.

При наличии в системе диэлектрика будут справедливы уравнения

$$\operatorname{div}\mathbf{D} \overset{(8.16)}{=} \operatorname{div}\mathbf{E} + 4\pi \operatorname{div}\mathcal{P} \overset{(8.59)}{=} 4\pi\rho. \tag{8.62}$$

Тогда с учётом объёмных связанных зарядов диэлектрика $\rho_{\text{связ}}$

$$\operatorname{div}\mathcal{P} = -\rho_{\text{связ}}, \tag{8.63}$$

$$\operatorname{div}\mathbf{E} = 4\pi(\rho + \rho_{\text{связ}}). \tag{8.64}$$

Запишем ещё два уравнения

$$\operatorname{div}\mathbf{B} \overset{\substack{(8.41)\\(3.74)}}{=} 0, \tag{8.65}$$

$$\operatorname{rot}\mathbf{E} \overset{\substack{(8.40)\\(8.41)\\(3.189)}}{=} -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}. \tag{8.66}$$

Уравнения движения электромагнитной системы (8.59, 8.60, 8.65, 8.66) образуют систему *уравнений Максвелла*, которую он сформулировал (правда, в другой записи) в 1864 году. Полученные уравнения функционально связывают электрические и магнитные поля с зарядами и токами, охватывают собой все известные закономерности электромагнетизма и играют в нём такую же роль, как законы Ньютона в механике.

Тем самым Максвелл сумел объединить понятия электричества и магнетизма в единую теорию электромагнитного поля. Сам Максвелл считал при этом, что его заслуга состоит лишь в математическом оформлении идей Фарадея.

К максвелловым уравнениям причисляют также и соотношения, не являющиеся уравнениями движения, но связывающие между собой значения основных векторов электромагнитного поля: (8.9) $\mathbf{j} = \phi\,(\mathbf{E} + \mathbf{E}^{\text{стр}})$, (8.17) $\mathbf{D} = \varepsilon\mathbf{E}$ и (8.23) $\mathbf{B} = \mu\mathbf{H}$.

Система уравнений Максвелла *полна*, т.е. при задании конкретных начальных условий электромагнитное поле в каждой точке пространства и в каждый момент времени *однозначно* определяется этой системой.

Запишем уравнения (8.60, 8.66, 8.59, 8.65) в интегральном виде

$$\oint\limits_C \mathbf{H}\,d\mathbf{r} \overset{(4.59)}{=} \iint\limits_S \operatorname{rot}\mathbf{H}\,d\sigma \overset{(8.60)}{=} \frac{4\pi}{c}\iint\limits_S \mathbf{j}\,d\sigma + \frac{1}{c}\frac{\partial}{\partial t}\iint\limits_S \mathbf{D}\,d\sigma, \tag{8.67}$$

$$\oint\limits_C \mathbf{E}\,d\mathbf{r} \overset{(4.59)}{=} \iint\limits_S \operatorname{rot}\mathbf{E}\,d\sigma \overset{(8.66)}{=} -\frac{1}{c}\frac{\partial}{\partial t}\iint\limits_S \mathbf{B}\,d\sigma, \tag{8.68}$$

$$\oiint\limits_S \mathbf{D}\,d\sigma \overset{(4.64)}{=} \iiint\limits_V \boxed{\operatorname{div}\mathbf{D}}\,dV \overset{(8.59)}{=} \iiint\limits_V \boxed{4\pi\rho}\,dV = 4\pi q, \tag{8.69}$$

$$\oiint\limits_S \mathbf{B}\,d\sigma \overset{(4.64)}{=} \iiint\limits_V \boxed{\operatorname{div}\mathbf{B}}\,dV \overset{(8.65)}{=} 0. \tag{8.70}$$

Уравнение (8.67) было получено математическими методами, а остальные три являются экспериментально установленными законами.

Уравнение (8.68) называют *закон Фарадея* — электродвижущая сила, индуктированная на замкнутом контуре C, помещенном в магнитное поле, равна изменению потока магнитной индукции $\mathbf{B}$ через площадь S, ограниченную этим контуром. Оказвается, что магнитное поле может не только воздействовать на проводник с электрическим током и возникать при движении электрических зарядов, но может, в свою очередь, создавать (возбуждать, индуцировать) электрический ток в проводнике, движущемся в магнитном поле. Это явление — возникновение электрического тока в проводнике, движущемся в магнитном поле, называемое электромагнитной индукцией, было открыто в 1831 году (год рождения Максвелла) Майклом Фарадеем, который в 1821 году написал в дневнике: «Превратить магнетизм в электричество». Основываясь на своём открытии, Фарадей в дальнейшем ввёл понятие электрических и магнитных полей как самостоятельных физических субстанций.

Уравнение (8.69) является содержанием *теоремы Гаусса* — поток вектора электрической индукции через замкнутую поверхность S пропорционален полному свободному заряду q, заключённому внутри объёма V, охватываемого этой поверхностью. Данное уравнение получено К. Ф. Гауссом (1830) для чисто электростатических полей и связано, по существу, с установленным ранее (1785) законом взаимодействия неподвижных электрических зарядов (законом Кулона). Постулировано Дж. К. Максвеллом (1864) в качестве одного из фундаментальных уравнений электродинамики.

Уравнение (8.70) есть поток векторов магнитной индукции и магнитного поля через замкнутую поверхность, который равен нулю ввиду отсутствия в природе магнитных зарядов.

Запишем два дифференциальных уравнения второго порядка:

$$\nabla^2\varphi - \frac{\varepsilon\mu}{c^2}\frac{\partial^2\varphi}{\partial t^2} = -\frac{4\pi}{\varepsilon}\rho, \quad \nabla^2\mathbf{A} - \frac{\varepsilon\mu}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\frac{4\pi\mu}{c}\mathbf{j}. \tag{8.71}$$

Доказательство

$$\nabla^2\varphi \overset{\substack{(3.188)\\(8.40)}}{=} -\operatorname{div}\mathbf{E} - \frac{1}{c}\frac{\partial \operatorname{div}\mathbf{A}}{\partial t} \overset{\substack{(8.59)\\(8.26)}}{=} -\frac{4\pi}{\varepsilon}\rho + \frac{\varepsilon\mu}{c^2}\frac{\partial^2\varphi}{\partial t^2},$$

$$\nabla^2\mathbf{A} \overset{(3.192)}{=} \operatorname{grad}\operatorname{div}\mathbf{A} - \operatorname{rot}\operatorname{rot}\mathbf{A} \overset{\substack{(8.41)\\(8.60)}}{=} \operatorname{grad}\operatorname{div}\mathbf{A} - \frac{4\pi\mu}{c}\mathbf{j} - \frac{\mu}{c}\frac{\partial\mathbf{D}}{\partial t} =$$

$$\overset{(8.40)}{=} \operatorname{grad}\operatorname{div}\mathbf{A} - \frac{4\pi\mu}{c}\mathbf{j} + \frac{\varepsilon\mu}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} + \frac{\varepsilon\mu}{c}\operatorname{grad}\frac{\partial\varphi}{\partial t} =$$

$$\overset{(8.26)}{=} \operatorname{grad}\operatorname{div}\mathbf{A} - \frac{4\pi\mu}{c}\mathbf{j} + \frac{\varepsilon\mu}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} - \operatorname{grad}\operatorname{div}\mathbf{A} = -\frac{4\pi\mu}{c}\mathbf{j} + \frac{\varepsilon\mu}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}.$$

Уравнения (8.71) дают возможность определить значения скалярного и векторного потенциалов по заданному распределению зарядов и токов проводимости.

Введя обозначение $v = \frac{c}{\sqrt{\varepsilon\mu}}$, $\chi(x, y, z, t)$ — заданная функция координат и времени, S — одна из величин φ, A_x, A_y, A_z, можно записать уравнения (8.71) в виде так называемого *уравнения Д'Аламбера*:

$$\nabla^2 S - \frac{1}{v^2}\frac{\partial^2 S}{\partial t^2} = -4\pi\chi. \tag{8.72}$$

При независимости S от времени (стационарное поле) уравнение Д'Аламбера вырождается в известное *уравнение Пуассона* (4.97)

$$\nabla^2 S = -4\pi\chi. \tag{8.73}$$

Наконец, при $\chi = 0$ уравнение Д'Аламбера принимает вид известного *волнового уравнения* (5.65)

$$\nabla^2 S = \frac{1}{v^2}\frac{\partial^2 S}{\partial t^2}. \tag{8.74}$$

Таким образом, из уравнений Максвелла непосредственно вытекает существование электромагнитных волн, предсказанных ещё Фарадеем (1832), т.е. электромагнитных полей, существующих в пустоте при отсутствии зарядов. В октябре 1861 г. Максвелл сообщил Фарадею о своём теоретическом открытии: свет — это электромагнитное возмущение, распространяющееся в непроводящей среде, то есть разновидность электромагнитных волн. Таким образом, появилась физическая основа уравнений Максвелла — принцип близкодействия, утверждающий, что передача электромагнитных возмущений от точки к точке происходит с конечной скоростью, а именно, со скоростью света. Этот завершающий этап теоретических исследований изложен в работе Максвелла «Динамическая теория электромагнитного поля» (1864), а итог его работ подвёл знаменитый «Трактат об электричестве и магнетизме» (1873).

Теория электромагнитного поля и, в особенности, следующий из неё вывод о существовании электромагнитных волн при жизни Максвелла оставались чисто теоретическими положениями, не имевшими никакого экспериментального подтверждения, и современниками зачастую воспринимались как «игра ума». Г. Герц предпринял опыты для опровержения «ошибочной и слабо обоснованной гипотезы Максвелла» и, к своему удивлению, открыл электромагнитные волны, в результате которых теория получила экспериментальное подтверждение, что сыграло решающую роль для её утверждения.

Математическим аппаратом теории Максвелла послужил векторный анализ, представленный в инвариантной форме через кватернионы Гамильтона. Каноническая форма записи, принятая ныне, принадлежит опять же Г. Герцу и О. Хевисайду. После изучения законов отражения и преломления, установления поляризации и измерения скорости электромагнитных волн Герц доказал их полную аналогию со световыми. Благодаря своим опытам Герц пришёл к следующим выводам: волны Максвелла «синхронны» (теория Максвелла, что скорость распространения радиоволн равна

скорости света, справедлива); энергию электрического и магнитного поля можно передавать без проводов. Всё это было изложено в работе «О лучах электрической силы», вышедшей в декабре 1888 г., который считается годом открытия электромагнитных волн и экспериментального подтверждения теории Максвелла. Герц также нашёл, что электромагнитные волны проходили через одни виды материалов и отражались другими, что привело в будущем к появлению радаров. Кроме того, Герц заметил, что заряженный конденсатор теряет свой заряд быстрее при освещении его пластин ультрафиолетовым излучением. Полученные результаты явились открытием нового явления в физике, названного фотоэффектом. Герц считал, что его открытия были не практичнее максвелловских: «Это абсолютно бесполезно. Это только эксперимент, который доказывает, что маэстро Максвелл был прав. Мы всего-навсего имеем таинственные электромагнитные волны, которые не можем видеть глазом, но они есть». «И что же дальше?» — спросил его один из студентов. Герц пожал плечами, он был скромный человек, без претензий и амбиций: «Я предполагаю — ничего».

8.5 Энергия электромагнитного поля

Запишем два уравнения

$$W_E = \frac{1}{8\pi} \iiint\limits_V (\mathbf{D} \cdot \mathbf{E})\, dV = \frac{1}{2} \iiint\limits_V \rho\varphi\, dV = W_\rho, \tag{8.75}$$

$$W_H = \frac{1}{8\pi} \iiint\limits_V (\mathbf{B} \cdot \mathbf{H})\, dV = \frac{1}{2c} \iiint\limits_V (\mathbf{A} \cdot \mathbf{j})\; dV = W_{\mathbf{j}}. \tag{8.76}$$

Доказательство уравнения (8.75)

$$\begin{aligned}
\frac{1}{8\pi} \iiint\limits_V (\mathbf{D} \cdot \mathbf{E})\, dV &\overset{(8.85)}{=} -\frac{1}{8\pi} \iiint\limits_V (\mathbf{D} \cdot \operatorname{grad} \varphi)\, dV = \\
&\overset{(3.178)}{=} \frac{1}{8\pi} \iiint\limits_V \varphi \operatorname{div} \mathbf{D}\, dV - \frac{1}{8\pi} \iiint\limits_V \operatorname{div} (\varphi\, \mathbf{D})\, dV = \\
&\overset{\substack{(4.64)\\(8.59)}}{=} \frac{1}{2} \iiint\limits_V \rho\varphi\, dV - \frac{1}{8\pi} \oiint\limits_S \varphi\, (\mathbf{D} \cdot d\sigma)\,.
\end{aligned}$$

При интегрировании по всему объёму поля интеграл $\oiint_S \varphi\, (\mathbf{D} \cdot d\sigma)$ по пограничной поверхности всего объема этого поля обратиться на бесконечности в нуль.

Доказательство уравнения (8.76)

$$\frac{1}{8\pi}\iiint\limits_V (\mathbf{B}\cdot\mathbf{H})\,dV \overset{(8.41)}{=} \frac{1}{8\pi}\iiint\limits_V (\mathrm{rot}\mathbf{A}\cdot\mathbf{H})\,dV =$$

$$\overset{(3.181)}{=} \frac{1}{8\pi}\iiint\limits_V \mathrm{div}\,(\mathbf{A}\times\mathbf{H})\,dV + \frac{1}{8\pi}\iiint\limits_V (\mathbf{A}\cdot\mathrm{rot}\mathbf{H})\,dV =$$

$$\overset{\substack{(4.64)\\(8.91)}}{=} \frac{1}{8\pi}\oiint\limits_S (\mathbf{A}\times\mathbf{H})\,dV + \frac{1}{2c}\iiint\limits_V (\mathbf{A}\cdot\mathbf{j})\;dV.$$

Если мы распространим интегрирование на весь объём поля, то интеграл $\oiint_S (\mathbf{A}\times\mathbf{H})\,dV$ по пограничной поверхности этого поля обращается в нуль на бесконечности.

Величина W_ρ называется *энергией взаимодействия зарядов* в электрическом поле, а величина W_E называется *электрической энергией поля*. При этом величина

$$w_E = \frac{1}{8\pi}(\mathbf{D}\cdot\mathbf{E}) \tag{8.77}$$

называется *объёмной плотностью электрической энергии*.

В частности энергия точечного заряда равна $W = e\varphi$, где φ — потенциал внешнего поля в точке, занимаемой зарядом e, а энергия диполя равна $W = -\mathbf{d}\cdot\mathbf{E}$, где $\mathbf{d}$ — момент диполя, а $\mathbf{E}$ — напряжённость внешнего поля в точке расположения диполя.

Величина $W_\mathbf{j}$ называется *энергией магнитного взаимодействия токов* в отсутствии ферромагнетиков, а величина W_H называется *магнитной энергией поля*. При этом величина

$$w_H = \frac{1}{8\pi}(\mathbf{B}\cdot\mathbf{H}) \tag{8.78}$$

называется *объёмной плотностью магнитной энергии*.

В случае линейного тока в тонком замкнутом контуре C с постоянной силой тока J величина W_J называется *энергия тока* и вычисляется как:

$$W_J \overset{(8.6)}{=} \frac{J}{2c}\oint\limits_C \mathbf{A}\,d\mathbf{r} \overset{(4.59)}{=} \frac{J}{2c}\iint\limits_S \mathbf{B}\,d\sigma. \tag{8.79}$$

Это уравнение означает, что потенциальная функция замкнутого тока равна сумме «энергий» отдельных его элементов и пропорциональна потоку магнитной индукции.

Величина

$$W = \frac{1}{8\pi}\iiint\limits_V (\mathbf{D}\cdot\mathbf{E} + \mathbf{B}\cdot\mathbf{H})\,dV \tag{8.80}$$

называется *энергией электромагнитного поля*, а величина

$$w = \frac{1}{8\pi}(\mathbf{D}\cdot\mathbf{E} + \mathbf{B}\cdot\mathbf{H}) \tag{8.81}$$

интерпретируется как *плотность энергии электромагнитного поля.*

Теперь, с учётом уравнения (8.34), можно построить дополнительные аналогии между электродинамикой и механикой. Энергию взаимодействия зарядов W_ρ и электрическую энергию поля W_E можно интерпретировать как аналог потенциальной энергии U, а энергию взаимодействия токов $W_{\mathbf{j}}$ и магнитную энергию поля W_H можно интерпретировать как аналог кинетической энергии T.

Все полученные аналогии между функциями механики и электродинамики представлены в виде таблицы 8.1.

Электродинамика	**Механика**
Полный заряд q	Масса m
Плотность тока $\mathbf{j}$	Плотность потока $\mathbf{j}$
Электрический момент $\mathbf{d}$	Радиус-вектор $\mathbf{r}$
Магнитный момент $\mathbf{m}$	Кинетический момент $\mathbf{L}$
$W_{\mathbf{j}}$, W_H	Кинетическая энергия T
W_ρ, W_E	Потенциальная энергия U
$W_H + W_E$, $W_{\mathbf{j}} + W_\rho$	Гамильтониан H
$W_H - W_E$, $W_{\mathbf{j}} - W_\rho$	Лагранжиан L

Таблица 8.1: Аналогии механики и электродинамики

Рассмотрим теперь изменение во времени количества энергии электромагнитного поля W (8.80), находящегося внутри объёма V, ограниченного поверхностью S:

$$\frac{\partial W}{\partial t} = \iiint\limits_V \frac{\partial w}{\partial t}\, dV = P - Q - \oiint\limits_S \mathbf{S} \cdot d\sigma. \tag{8.82}$$

где $P = \iiint_V (\mathbf{jE}^{\text{стр}})\, dV$ — работа, совершаемая сторонними электродвижущими силами в единицу времени, над токами проводимости $\mathbf{j}$,

$Q = \iiint_V \frac{j^2}{\phi}\, dV$ — джоулево тепло, выделяемое токами в единицу времени,

$\oiint_S \mathbf{S} \cdot d\sigma = \frac{c}{4\pi} \oiint_S (\mathbf{E} \times \mathbf{H}) \cdot d\sigma$ — поток электромагнитной энергии (количество энергии в единицу времени), вытекающий через поверхность S из рассматриваемого объёма V.

Вывод уравнения (8.82)

$$\frac{\partial W}{\partial t} \overset{\substack{(8.80)\\(8.17)\\(8.22)}}{=} \frac{1}{4\pi} \iiint\limits_V \left\{ \mathbf{E}\, \frac{\varepsilon\, \partial \mathbf{E}}{\partial t} + \mathbf{H}\, \frac{\mu\, \partial \mathbf{H}}{\partial t} \right\} dV =$$

$$\overset{\substack{(8.66)\\(8.60)}}{=} \iiint\limits_V \left\{ -\mathbf{j}\mathbf{E} + \frac{c}{4\pi} \left(\mathbf{E}\, \mathrm{rot}\, \mathbf{H} - \mathbf{H}\, \mathrm{rot}\, \mathbf{E} \right) \right\} dV =$$

$$\overset{\substack{(8.9)\\(3.181)}}{=} \iiint\limits_V \left\{ \mathbf{j}\mathbf{E}^{\mathrm{стр}} - \frac{j^2}{\phi} - \frac{c}{4\pi}\, \mathrm{div}\, (\mathbf{E} \times \mathbf{H}) \right\} dV =$$

$$\overset{(4.64)}{=} \iiint\limits_V \left\{ \mathbf{j}\mathbf{E}^{\mathrm{стр}} - \frac{j^2}{\phi} \right\} dV - \frac{c}{4\pi} \oiint\limits_S (\mathbf{E} \times \mathbf{H}) \cdot d\sigma.$$

Если поверхность S заключает в себе *полное поле*, то поверхностный интеграл обращается в нуль. Тогда общее приращение электромагнитной энергии равно избытку работы P сторонних электродвижущих сил (химического, термического и тому подобного происхождения) над выделением джоулева тепла Q (при предполагаемой неподвижности материальных тел, механическая работа равна нулю). Таким образом, уравнение (8.82) выражает собой *закон сохранения энергии*. Вместе с тем мы убеждаемся, что токи смещения никакого тепла не выделяют и сторонние электродвижущие силы при их прохождении работы не совершают, ибо в выражения для P и Q плотность токов смещения не входит.

Если поверхность S не обнимает собой полного поля, то выражение есть поток электромагнитной энергии через эту поверхность. Этот положение называется *теорема Пойнтинга*, а величина:

$$\mathbf{S} = \frac{c}{4\pi} (\mathbf{E} \times \mathbf{H}) \overset{(8.9)}{=} \frac{c}{4\pi} (\mathbf{j} \times \mathbf{H}) - \frac{c}{4\pi} (\mathbf{E}^{\mathrm{стр}} \times \mathbf{H}) \tag{8.83}$$

носит название *вектор Пойнтинга* и есть плотность потока электромагнитной энергии в рассматриваемой точке пространства.

Первый член правой части (8.83) представляет собой поток энергии, *притекающий* в проводник из окружающего его пространства внутрь провода; второй член снабжён отрицательным знаком и имеет поэтому *обратное* направление, т.е. представляет собой поток энергии, вытекающей из провода через его поверхность.

Вообще говоря, можно сказать, что в проводнике, по которому циркулирует ток, происходит, в сущности, превращение электромагнитной энергии в тепло; локализована же эта энергия преимущественно во внешнем окружающем проводник пространстве и поступает в проводник через его внешнюю поверхность.

Формулировка закона сохранения энергии с помощью понятия потока энергии (8.82) впервые дана в общей форме Н. А. Умовым (1874).

8.6 Стационарное поле

Получив основные уравнения движения электромагнитной системы, рассмотрим частный, но очень важный случай стационарного поля.

В уравнениях (8.60) и (8.66) знак частной производной по времени означает, что имеется ввиду скорость изменения вектора в фиксированной точке пространства. Таким образом, нестационарность электрической $\frac{\partial \mathbf{D}}{\partial t}$ либо магнитной $\frac{\partial \mathbf{B}}{\partial t}$ индукции в рассматриваемой точке приводит к возникновению, соответственно, магнитного либо электрического вихря в этой точке. Стационарное электрическое поле начинает скручиваться в пространстве, если происходят изменения во времени поля магнитного. И наоборот, магнитное поле скручивается под влиянием изменения во времени поля электрического.

Если же электромагнитное поле и его потенциалы стационарны (нестационарные члены $\frac{\partial \varphi}{\partial t}$, $\frac{\partial \mathbf{A}}{\partial t}$, $\frac{\partial \mathbf{D}}{\partial t}$ и $\frac{\partial \mathbf{B}}{\partial t}$ равны нулю), то получаем

$$\operatorname{div} \mathbf{A} \overset{(8.26)}{=} 0, \tag{8.84}$$

$$\mathbf{E} \overset{(8.40)}{=} -\operatorname{grad} \varphi, \tag{8.85}$$

$$\operatorname{rot} \mathbf{E} \overset{(8.66)}{=} 0, \tag{8.86}$$

$$\operatorname{rot} \mathbf{H} \overset{(8.60)}{=} \frac{4\pi}{c}\mathbf{j}. \tag{8.87}$$

Для последних двух выражений запишем соответствующие интегральные уравнения

$$\oint\limits_C \mathbf{E}\, d\mathbf{r} \overset{(4.59)}{=} \iint\limits_S \operatorname{rot} \mathbf{E}\, d\sigma = 0, \tag{8.88}$$

$$\oint\limits_C \mathbf{H}\, d\mathbf{r} \overset{(4.59)}{=} \frac{4\pi}{c} \iint\limits_S \mathbf{j}\, d\sigma \overset{(8.5)}{=} \frac{4\pi}{c}\, J. \tag{8.89}$$

Теорема (8.89) — циркуляция вектора магнитного поля по замкнутому контуру пропорциональна силе тока через трубку, охватываемую этим контуром — впервые получена Дж. К. Максвеллом.

Исходя из (8.71), запишем уравнения, определяющие потенциалы стационарных полей

$$\varepsilon\, \Delta\varphi = -4\pi\, \rho, \tag{8.90}$$

$$\nabla^2 \mathbf{A} = -\frac{4\pi\mu}{c}\mathbf{j}. \tag{8.91}$$

Выражение (8.90) для φ носит название *уравнение Пуассона*. В тех участках поля, где нет электрических зарядов, выражение $\Delta\varphi = 0$ носит название *уравнение Лапласа*.

Введение скалярного электрического потенциала φ значительно облегчает изучение электрического поля неподвижных электрических зарядов, а

введение векторного магнитного потенциала $\mathbf{A}$ значительно облегчает изучение магнитного поля постоянных токов, ибо задача определения векторов $\mathbf{E}$ и $\mathbf{H}$ сводится к определению полей их потенциалов, которые, в свою очередь, определяются из конфигурации ρ и $\mathbf{j}$.

Например, для точечных и объёмных зарядов и токов (поверхностные заряды и токи не учитываются) запишем:

$$\mathbf{E} \overset{\substack{(8.85)\\(8.24)\\(3.172)}}{=} \frac{e\,\mathbf{r}}{r^3}, \qquad \mathbf{D} = \iiint\limits_V \frac{\rho\,\mathbf{r}}{r^3}\,dV; \tag{8.92}$$

$$\mathbf{H} \overset{\substack{(8.41)\\(8.25)\\(3.183)}}{=} \frac{e}{c\,r^3}(\mathbf{v}\times\mathbf{r}), \qquad \mathbf{H} = \frac{J}{c}\oint\limits_C \frac{d\mathbf{r}\times\mathbf{r}}{r^3}, \qquad \mathbf{H} = \frac{1}{c}\iiint\limits_V \frac{\mathbf{j}\times\mathbf{r}}{r^3}\,dV. \tag{8.93}$$

Возникновение магнитного поля при движении заряда (8.93) называется *закон Био–Савара–Лапласа*. Закон установлен экспериментально Био и Саваром (1820). Лаплас проанализировал данное выражение и показал, что с его помощью путём интегрирования, в частности, можно вычислить магнитное поле движущегося точечного заряда, если считать движение одной заряженной частицы током.

Обратим внимание ещё раз. Различие между электрическим и магнитным полем построено на различии тока и заряда. Из уравнений (8.43) и (8.44) следует, что электрическое поле действует на любые заряды, а магнитное — только на движущиеся. Из уравнений (8.92) и (8.93) следует, что электрические заряды и токи могут сами создавать поле. Но если электрическое поле могут создавать любые заряды, то магнитное — опять же, только движущиеся.

Легко заметить, что уравнения (8.90) и (8.91) соответствуют уравнениям (4.97) и (4.101) для скалярного и векторного потенциала.

Уравнения (8.41) $\mathbf{B} = \operatorname{rot}\mathbf{A}$ и (8.65) $\operatorname{div}\mathbf{B} = 0$ соответствуют уравнениям (3.77, 3.78) и свидетельствуют о том, что векторный потенциал $\mathbf{A}$ *всегда* задаёт в системе соленоидальное магнитное поле $\mathbf{B}$. Уравнения (8.85, 8.86) соответствуют уравнениям (3.75, 3.76) и свидетельствуют о том, что в случае *стационарного* электромагнитного поля скалярный потенциал φ задаёт в системе потенциальное электрическое поле $\mathbf{E}$.

Таким образом, векторные линии потенциального электрического поля возникают из точек расположения электрических зарядов и не имеют завихрений, а векторные линии соленоидального магнитного поля представляют собой вихри вокруг линий электрического тока и не имеют источников. При этом магнитный вихрь вокруг линии тока существенно отличается, например, от водяного или воздушного вихря. В водяном или воздушном вихре соседние слои движутся с разной скоростью, поэтому их векторы скорости имеют разную длину, а ротор скорости равен удвоенной угловой скорости вращения (3.70). В магнитном вихре поле устроено так, что вне линии тока соседние векторы магнитного поля всегда одинаковы ($\operatorname{rot}\mathbf{H} = 0$), и толь-

ко внутри проводника с током ($\operatorname{rot}\mathbf{H} = \frac{4\pi}{c}\mathbf{j}$) ротор поля отличен от нуля. Поэтому только пространство, занятое током, вихревое, а всё остальное пространство магнитного поля — гармоническое. Подобно тому, как потенциальное электростатическое поле полностью определяется заданием силы его истоков (т.е. заданием его дивергенции, как функции координат), так вихревое магнитное поле полностью определяется заданием мощности его токов.

Стационарному электромагнитному полю соответствует постоянная конфигурация электрических зарядов и токов. Очевидно, что в поле *постоянных* (не меняющихся со временем) токов, распределение зарядов в пространстве должно оставаться *стационарным*, ибо если бы имело место какое бы то ни было перераспределение зарядов, то напряжённость поля неизбежно изменилась бы, и ток перестал бы быть постоянным. То обстоятельство, что в данной точке пространства одни элементы заряда сменяются другими, не изменяет их плотности и, соответственно, их поля, которое также остаётся *стационарным*.

Из стационарности распределения зарядов в поле постоянных токов вытекает, что токи эти необходимо должны быть либо *замкнутыми*, т.е. циркулировать по замкнутому контуру, либо уходить в *бесконечность*, ибо, в противном случае, в месте начала (истоков) и окончания (стоков) тока происходило бы накопление или убывание зарядов.

Закон Кулона (закон постоянных точечных зарядов). Если мы рассмотрим в вакууме взаимодействие двух точечных зарядов e_i и e_j, то сила воздействия одного заряда на другой $\mathbf{F}_{ij}$ вычисляется при помощи формул (8.43) и (8.92):

$$\mathbf{F}_{ij} = \frac{e_i\, e_j\, \mathbf{r}_{ij}}{r_{ij}^3}, \quad |\mathbf{F}_{ij}| = F_{ij} = \frac{e_i\, e_j}{r_{ij}^2}, \tag{8.94}$$

причём

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji} \tag{8.95}$$

аналогично третьему закону Ньютона (7.14).

Весьма существенно, что закон Кулона *справедлив только для точечных зарядов*, т.е. достаточно малых, по сравнению с расстоянием между ними. Только в этом случае понятие расстояния имеет однозначный смысл, а взаимодействие между зарядами не зависит от их формы.

Закон Ома (закон постоянных линейных токов).

Запишем уравнение (8.88) для незамкнутого участка проводника C, вдоль которого производится интегрирование

$$\int\limits_C \mathbf{E}\, d\mathbf{r} = \varphi_1 - \varphi_2 = V, \tag{8.96}$$

где φ_1 и φ_2 — значения потенциала на концах участка проводника C.

Ввиду потенциальности поля $\mathbf{E}$ криволинейный интеграл не зависит от пути и равен разнице потенциалов на концах участка, которая называется

напряжением V. Пусть теперь мы имеем проводник длиной L и площадью S. Тогда

$$V = \varphi_1 - \varphi_2 = JR, \tag{8.97}$$

где $R = \lambda \frac{L}{S}$ — *сопротивление* рассматриваемого участка проводника.

Доказательство. Усредним уравнение (8.96) по площади и длине проводника, для этого проинтегрируем левую часть уравнения по площади и длине, а затем разделим на площадь проводника

$$V = \int\limits_C \mathbf{E}\, d\mathbf{r} \overset{(8.8)}{=} \frac{1}{S} \iint\limits_S \mathbf{j}\, d\sigma \int\limits_C \lambda\, d\mathbf{r} \overset{(8.5)}{=} J\lambda \frac{L}{S} = JR.$$

Феноменологический закон постоянного тока (8.97) носит название *закон Ома* — напряжение на концах проводника равно произведению силы тока в проводнике на сопротивление проводника.

Выражение (8.97) есть интегральная форма закона Ома, а выражение (8.8) — дифференциальная форма, в которой исчезает зависимость от геометрических размеров проводника.

Закон Ома открыт Георгом Омом в 1827 году.

Как мы видим, электрические и магнитные поля не являются силами в обычном, употребляемом в механике смысле этого слова. Механическая, или, как её иногда называют *пондеромоторная* (от лат. pondus, род. падеж ponderis — тяжесть и motor — движущий) *сила* $\mathbf{F}$ имеет другую размерность. В отличие от работы пондеромоторной силы $\int_C \mathbf{F}\, d\mathbf{r}$ работа электрического поля $\int_C \mathbf{E}\, d\mathbf{r}$ называется *электродвижущей силой* (ЭДС) $\mathcal{E}$ или *напряжением* V.

Работа, совершаемая электрическим током при переносе электрических зарядов, равна:

$$\int\limits_C \mathbf{F}\, d\mathbf{r} \overset{(8.43)}{=} q \int\limits_C \mathbf{E}\, d\mathbf{r}. \tag{8.98}$$

При прохождении тока по проводнику, произведённая работа будет выделяться в проводнике в виде тепла. Выделение энергии (например, тепла) в единицу времени называется мощностью Q. Мощность выделения тепла в единице объёма называется удельной мощностью w:

$$w = \mathbf{j} \cdot \mathbf{E} \overset{(8.8)}{=} \phi E^2, \tag{8.99}$$

$$Q = \iiint\limits_V \mathbf{j} \cdot \mathbf{E}\, dV, \tag{8.100}$$

$$Q = J \int\limits_C \mathbf{E}\, d\mathbf{r} \overset{(8.96)}{=} JV \overset{(8.97)}{=} J^2 R. \tag{8.101}$$

Полученное уравнение (8.101) называется *закон Джоуля–Ленца*, открытый независимо Джеймсом Джоулем и Эмилем Ленцем в 1840.

8.7 Граница электромагнитных сред

Граница между различными средами в трёхмерном пространстве обычно представляет собой двумерную поверхность. С одной стороны границы находится среда с одними свойствами, а с другой стороны границы — с другими, поэтому на границе между различными средами могут возникать изломы и разрывы функций рассматриваемой системы.

Относительно электрического и магнитного потенциалов можно сказать, что при переходе через границу между различными электромагнитными средами они всегда остаются непрерывными функциями. Электрические и магнитные векторные поля при переходе через границу сред могут испытывать разрывы. В этом случае непрерывные функции потенциалов имеют излом (не являются дифференцируемыми функциями).

Обозначим векторы с одной (внутренней) стороны границы раздела индексом $^{(1)}$, а с другой (внешней) — индексом $^{(2)}$.

Мы должны учесть в каждой точке поверхности две составляющие любых векторов — нормальные составляющие с индексом n, направленные по нормали к поверхности, и тангенциальные составляющие с индексом t, направленные по касательной к поверхности.

В отличие от вектора электрического поля $\mathbf{E}$ вектор электрической индукции $\mathbf{D}$ представляет собой сумму двух совершенно разных физических величин: напряжённости поля $\mathbf{E}$ и умноженной на 4π поляризации единицы объёма среды $\mathcal{P}$. Тем не менее введение в рассмотрение этого вектора чрезвычайно упрощает изучение поля в диэлектриках.

Рассмотрим границу разрыва между двумя средами с различной диэлектрической проницаемостью $\varepsilon^{(1)}$ и $\varepsilon^{(2)}$.

Внешние и внутренние тангенциальные (касательные) составляющие вектора электрического поля на границе всегда будут непрерывны $\mathbf{E}_t^{(2)} = \mathbf{E}_t^{(1)}$. Действительно, если тангенциальные составляющие поля будут испытывать скачок, то при обходе по замкнутому контуру вдоль границы с двух её сторон будет нарушаться условие (8.88) $\oint_C \mathbf{E}\,d\mathbf{r} = 0$. Напротив, тангенциальные слагающие электромагнитной индукции на границе всегда будут испытывать скачок $\frac{\mathbf{D}_t^{(2)}}{\varepsilon^{(2)}} = \frac{\mathbf{D}_t^{(1)}}{\varepsilon^{(1)}}$:

$$\frac{\mathbf{D}_t^{(2)}}{\varepsilon^{(2)}} \overset{(8.17)}{=} \mathbf{E}_t^{(2)} \overset{(8.88)}{=} \mathbf{E}_t^{(1)} \overset{(8.17)}{=} \frac{\mathbf{D}_t^{(1)}}{\varepsilon^{(1)}} \tag{8.102}$$

Для нормальных составляющих запишем:

$$\mathbf{D}_n^{(2)} - \mathbf{D}_n^{(1)} \overset{(8.17)}{=} \varepsilon^{(2)}\mathbf{E}_n^{(2)} - \varepsilon^{(1)}\mathbf{E}_n^{(1)} \overset{\substack{(8.24)\\(4.99)}}{=} 4\pi\varrho, \tag{8.103}$$

$$\mathcal{P}_n^{(2)} - \mathcal{P}_n^{(1)} = -\varrho_{\text{связ}}, \tag{8.104}$$

$$\mathbf{E}_n^{(2)} - \mathbf{E}_n^{(1)} \overset{(8.16)}{=} 4\pi(\varrho + \varrho_{\text{связ}}). \tag{8.105}$$

Если на границе нет свободных зарядов ($\varrho = 0$), то $\mathbf{D}_n^{(2)} = \mathbf{D}_n^{(1)}$ и $\varepsilon^{(2)}\mathbf{E}_n^{(2)} = \varepsilon^{(1)}\mathbf{E}_n^{(1)}$. Следовательно, на незаряженной границе раздела двух

различных сред $\varepsilon^{(2)} \neq \varepsilon^{(1)}$ нормальные слагающие электрической индукции остаются непрерывными, а нормальные же слагающие напряжённости поля испытывают скачок $\frac{\mathbf{E}_n^{(2)}}{\varepsilon^{(2)}} = \frac{\mathbf{E}_n^{(1)}}{\varepsilon^{(1)}}$.

Из уравнений (8.64) и (8.105) следует, что электрическое поле *при наличии диэлектриков* совпадает с полем, которое возбуждалось бы в *отсутствии* диэлектриков теми же свободными зарядами при добавлении к ним зарядов $\rho_{\text{связ}}$ и $\varrho_{\text{связ}}$.

Уравнение (8.24) для φ можно записать в следующем виде:

$$\varphi = \iiint\limits_V \frac{\rho + \rho_{\text{связ}}}{r}\, dV + \iint\limits_S \frac{\varrho + \varrho_{\text{связ}}}{r}\, d\sigma. \tag{8.106}$$

Если рассматривать частный случай на границе двух сред: вакуума и проводника, то ситуация намного проще. Так как в состоянии электростатического равновесия движения зарядов не происходит, то внутри проводника электрическое поле равно нулю ($\mathbf{E} = 0$), потенциал поля имеет постоянное значение на всём протяжении каждого отдельного проводника ($\varphi = const$), а все свободные заряды располагаются на поверхности проводника (внутри проводника заряды отсутствуют, вернее, положительные и отрицательные заряды взаимно компенсируются).

Следовательно, на границе проводника возникает скачок электрического поля, т.е. поверхность проводника образует поверхность разрыва вектора $\mathbf{E}$. Значит, расположение поверхностных свободных зарядов ϱ (вектор *плотности поверхностных зарядов*, имеющий направление нормали к поверхности в рассматриваемой точке) проводника определяется расположением поверхностей разрыва вектора $\mathbf{E}$.

На внешней поверхности проводников вектор электрического поля всегда направлен нормально (перпендикулярно) к их поверхности $\mathbf{E} = \mathbf{E}_n$, так как тангенциальные (касательные к поверхности) слагающие поля внутри проводника равны нулю ввиду отсутствия поля, соответственно, внешние тангенциальные слагающие также всегда нулевые $\mathbf{E}_t = 0$. Для поверхности проводников запишем

$$\mathbf{E}_n^{(2)} - \mathbf{E}_n^{(1)} \overset{(4.99)}{=} 4\pi\varrho, \quad \mathbf{E}_t^{(2)} \overset{(8.88)}{=} \mathbf{E}_t^{(1)} = 0 \tag{8.107}$$

Рассмотрим теперь поведение магнитного поля на границе раздела двух разных сред в присутствии поверхностных магнитных токов:

$$\frac{4\pi}{c} \iint\limits_S \mathbf{i}\, d\sigma \overset{(8.7)}{=} \frac{4\pi}{c} \iiint\limits_V \mathbf{j}\, dV \overset{(8.91)}{=} \iiint\limits_V \mathbf{H}\, dV \overset{(4.70)}{=} \oiint\limits_S d\sigma \times \mathbf{H}.$$

Стянув интегрирование по замкнутой поверхности к *двукратному* интегрированию по двум поверхностям внешней и внутренней на границе раздела двух сред (причём на внутренней стороне поверхности величина $d\sigma$ меняет знак), запишем

$$\oiint\limits_S d\sigma \times \mathbf{H} = \iint\limits_S d\sigma \times \left(\mathbf{H}^{(2)} - \mathbf{H}^{(1)}\right) = \frac{4\pi}{c} \iint\limits_S \mathbf{i}\, d\sigma.$$

В результате, получим

$$\mathbf{n} \times \left(\mathbf{H}^{(2)} - \mathbf{H}^{(1)} \right) = \frac{4\pi}{c} \mathbf{i}, \tag{8.108}$$

где $\mathbf{n}$ — нормаль границы раздела сред.

Это уравнение и представляет собой то пограничное условие, которому при наличии поверхностных токов на поверхности раздела должны удовлетворять касательные (тангенциальные) слагающие вектора $\mathbf{H}$ (ибо лишь эти слагающие и входят в выражение $\mathbf{n} \times \mathbf{H}$).

При отсутствии поверхностных токов это пограничное условие принимает вид: $\mathbf{n} \times \left(\mathbf{H}^{(2)} - \mathbf{H}^{(1)} \right) = 0$ или $\mathbf{H}_t^{(2)} = \mathbf{H}_t^{(1)}$. Это уравнение означает, что при отсутствии поверхностных токов касательные слагающие напряженности поля $\mathbf{H}$ непрерывны при пересечении произвольной поверхности. Относительно нормальных слагающих вектора магнитной индукции $\mathbf{B}$ можно сразу сказать, что они непрерывны всегда $\mathbf{B}_n^{(2)} = \mathbf{B}_n^{(1)}$, что следует из уравнения (8.65) $\operatorname{div} \mathbf{B} = 0$. Действительно, как мы видели при рассмотрении поверхностных электрических зарядов, скачок нормальной слагающей вектора электрической индукции возникает только при наличии свободных поверхностных электрических зарядов. Следовательно, ввиду отсутствия в природе свободных магнитных зарядов, скачок нормальной слагающей вектора магнитной индукции при пересечении границы невозможен.

Глава 9

Термодинамическая система

9.1 Представляющий ансамбль

Уравнения произвольной динамической системы (5.2) предполагаются точными. При этом полагают, что эти уравнения в принципе можно проинтегрировать и вычислить их точные решения.

На первый взгляд, в предположении, что значения функций вычисляются точно, нет ничего необычного. С другой стороны, при рассмотрении реального явления становится ясно, что это очень трудно сделать. Предположим, что мы наблюдаем динамическую систему, состоящую из множества однородных частиц, в течение достаточно длительного промежутка времени. В каждый из моментов наблюдения рассматриваемая система изображается точкой в фазовом пространстве, у которой нужно вычислить точные значения координат. Во-первых, достаточно сложно определить полный перечень этих координат, так как частиц в системе может быть слишком много. Допустим, нам это удалось. Последовательность полученных фазовых точек образует чрезвычайно запутанную траекторию в фазовом пространстве. Для ее нахождения необходимо составить и решить все дифференциальные уравнения движения, что представляется очень трудной задачей. Допустим, нам каким-то образом задать все начальные и граничные условия и проинтегрировать эти уравнения в общем виде. Очевидно, что через достаточно короткий период времени полученные уравнения движения будут нарушаться. В сложной и запутанной системе параллельно всегда происходят разнообразные процессы, кроме того, оказывают влияние внешние условия и различные «шумы».

В результате следует признать, что задача интегрирования гамильтоновых уравнений движения при наличии достаточно большого числа частиц в общем виде не решается. Но именно наличие большого числа однородных частиц позволяет подойти к решению задачи с другой стороны.

Мы можем рассмотреть не последовательность, а совокупность точек фазовой траектории. Вместо того чтобы рассматривать точки, изображаю-

щие состояние *одной системы в различные моменты времени*, можно формальным образом ввести в рассмотрение *одновременно очень большое* (*в пределе — бесконечное*) *число состояний системы*.

Вместо одной динамической системы, которая движется по очень сложной и запутанной траектории фазового пространства, мы рассматриваем множество систем в виде «облака» точек в фазовом пространстве. Такую воображаемую совокупность одинаковых систем Гиббс и Эйнштейн назвали *представляющим ансамблем* систем.

Гиббс определил это понятие следующим образом: «Мы можем представить себе огромное число систем одинаковой природы, но различных по конфигурациям и скоростям, которыми они обладают в данный момент, и различных не только бесконечно мало, но так, что охватывается каждая мыслимая комбинация конфигураций и скоростей. При этом мы можем поставить себе задачей, не прослеживать определенную систему через всю последовательность ее конфигураций, а установить, как будет распределено все число систем между различными возможными конфигурациями и скоростями в любой требуемый момент, если такое распределение было задано для какого-либо момента времени. Основным уравнением при таком исследовании является уравнение, дающее скорость изменения числа систем, заключенных внутри определенных малых границ конфигурации и скорости» [12, с. 12].

Малые границы конфигурации и скорости в фазовом пространстве, о которых говорит Гиббс, называют бесконечно малым элементом фазового объема, и записывают в виде (5.44):

$$dV = dp_1 \wedge \ldots \wedge dp_n \wedge dq_1 \wedge \ldots \wedge dq_n.$$

Элемент фазового объема dV хоть и считается бесконечно малым, но все-таки он достаточно большой (макроскопический) по отношению к размеру точек. Число же точек предполагается столь большим, что фазовое пространство можно считать непрерывной средой, которая, в силу условия (4.67), оказывается подобной несжимаемой жидкости. Действительно, дивергенция поля фазовой скорости оказывается равной нулю:

$$\operatorname{div}(\dot{\mathbf{q}}, \dot{\mathbf{p}}) \overset{(3.83)}{=} \sum \left(\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) \overset{(5.39)}{=} \sum \left(\frac{\partial^2 H}{\partial q_i \partial p_i} - \frac{\partial^2 H}{\partial p_i \partial q_i} \right) = 0.$$

«Несжимаемость фазовой жидкости» произвольной динамической системы, состоящей из множества однородных частиц, не является случайным результатом, а прямо вытекает из теоремы Лиувилля (5.44) о сохранении объёма фазового пространства.

Обратим внимание ещё раз. Рассматриваемая динамическая система состоит из множества однородных частиц, а фазовое простанство этой системы состоит из множества состояний (фаз), которые различаются значениями конфигураций и скоростей. Эта совокупность состояний рассматривается нами как непрерывная среда, подобная несжимаемой (фазовой)

жидкости. Такую динамическую систему принято называть *статистической системой*.

Аналогично всякой непрерывной среде, фазовая жидкость должна иметь соответствующую «плотность» ρ, но в отличие от обычной плотности, которая является функцией пространственных координат x, y, z и, возможно, времени, плотность фазовой жидкости является функцией фазовых координат и, согласно уравнениям (4.66, 4.67) является интегралом движения (см. с. 224):

$$\rho(t, \mathbf{p}, \mathbf{q}) = const. \tag{9.1}$$

Тогда, вычислив полную производную функции ρ, получим

$$0 \overset{\substack{(9.1)\\(\text{с. }106)}}{=} \dot{\rho} \overset{(3.102)}{=} \sum_{i=1}^{n} \left(\dot{q}_i \frac{\partial \rho}{\partial q_i} + \dot{p}_i \frac{\partial \rho}{\partial p_i} \right) + \frac{\partial \rho}{\partial t} \overset{(5.49)}{=} \{H, \rho\} + \frac{\partial \rho}{\partial t}. \tag{9.2}$$

Состояние динамической системы, при котором «плотность фазовой жидкости» при движении вдоль фазовых траекторий принимает стационарное значение $\frac{\partial \rho}{\partial t} = 0$ (функция ρ не зависит от времени t явно), дает следующее уравнение:

$$\{H, \rho\} = 0, \tag{9.3}$$

которое является *условием статистического равновесия*.

Чтобы функция ρ была интегралом движения, она должна быть функцией энергии: $\rho = f(E)$, где $E = H$.

Функция распределения системы является произведением функций распределения ее частей. Допустим, рассматриваемая система состоит из двух независимых частей A и B. Тогда $\rho = \rho_A \cdot \rho_B$. В то же время интеграл движения системы должен являться суммой интегралов движения частей системы, например $E = E_A + E_B$. Кроме того функции распределения для отдельных частей системы должны обладать одинаковой природой, т.е. одинаково выражаться через интеграл движения.

Вспоминая уравнение $\ln AB = \ln A + \ln B$ заметим, что распределение, представляемое выражением

$$\rho = e^{\frac{\psi - E}{\theta}}, \tag{9.4}$$

где θ и ψ — постоянные,

E — энергия;

по-видимому, является наиболее простым мыслимым случаем, когда выполняются все вышеперечисленные условия. Делитель θ (величина, обладающая одинаковой размерностью с E) не усложняет дела, а, наоборот, упрощает его, поскольку наличие его делает распределение независимым от употребляемых единиц.

Если теперь проинтегрировать функцию ρ по всему фазовому пространству (при условии, конечно, что такой интеграл существует) и ввести

нормировочный коэффициент таким образом, чтобы функция ρ удовлетворяла условию нормировки (6.49):

$$\underbrace{\int \dots \int}_{2n} \rho \, dV = 1, \tag{9.5}$$

то мы получим функцию, играющую роль «плотности» распределения вероятности в фазовом пространстве. Такую функцию ρ называют *функцией статистического распределения* (или просто функцией распределения) данной статистической системы.

Отметим одно чрезвычайно важное обстоятельство. При переходе к функции распределения состояний статистической системы мы можем любое из этих состояний принять за начальное состояние.

Следовательно, *состояние статистической системы не зависит от начальных условий.*

9.2 Термодинамические параметры

В замкнутой статистической системе полная энергия системы является её первым интегралом: $E(\mathbf{p}, \mathbf{q}) = const$.

Допустим, что система больше не является изолированной, и произошло некое бесконечно малое изменение пространственных координат $d\mathbf{q}$ системы. При этом взаимодействии с внешней средой изменилась полная энергия системы. Это означает переход от одной поверхности энергии (поверхности уровня) к другой. Причем градиент энергии в соответствии с первым из уравнений (5.39) равен силе $\mathbf{F}$, с которой система воздействует на внешнюю среду

$$\dot{\mathbf{p}} \overset{(7.2)}{=} -\frac{\partial E}{\partial \mathbf{q}} = \mathbf{F}. \tag{9.6}$$

Запишем дифференциал уравнения (9.5) с учётом уравнения (9.4), где модуль распределения θ, являясь нормирующим множителем для энергии, по прежнему остаётся константой, в отличие от величины ψ, которая становится переменной величиной:

$$\frac{d\psi}{\theta} \int e^{\frac{\psi - E}{\theta}} dV - \int \left(\frac{e^{\frac{\psi - E}{\theta}}}{\theta} \frac{\partial E}{\partial \mathbf{q}} \, d\mathbf{q} \right) dV = 0. \tag{9.7}$$

или

$$d\psi = \int \left(\frac{\partial E}{\partial \mathbf{q}} e^{\frac{\psi - E}{\theta}} \, d\mathbf{q} \right) dV \overset{(9.6)}{=} -\left(\int \mathbf{F} \rho \, dV \right) d\mathbf{q} \overset{(6.50)}{=} -\overline{\mathbf{F}} \, d\mathbf{q}. \tag{9.8}$$

Обозначим показатель экспоненты в формуле (9.4) через η, т.е.

$$\eta = \frac{\psi - E}{\theta}, \quad \rho = e^{\eta} \quad \text{или} \quad \eta = \ln \rho. \tag{9.9}$$

Запишем дифференциал уравнения (9.9):

$$\theta\, d\eta = d\psi - dE \quad \text{или} \quad dE \overset{(9.8)}{=} -\theta\, d\eta - \overline{\mathbf{F}}\, d\mathbf{q}. \tag{9.10}$$

Воздействие усреднённой силы системы на внешнюю среду приводит к бесконечно малому изменению пространственных координат $d\mathbf{q}$ и называется *работой* dA, которую система производит над внешней средой. Работа по общему правилу вычисляется как произведение силы на вызываемое силой перемещение (изменение координаты):

$$dA \overset{(7.7)}{=} \overline{\mathbf{F}}\, d\mathbf{q}. \tag{9.11}$$

При взаимодействии системы с внешней средой изменится также и объем V, который занимает система в координатном пространстве. Тогда

$$dV = \frac{\partial V}{\partial q_1}\, dq_1 + \ldots + \frac{\partial V}{\partial q_n}\, dq_n = \operatorname{grad} V \cdot d\mathbf{q}. \tag{9.12}$$

В данном случае $\operatorname{grad} V$ — это вектор на поверхности, разделяющей систему и внешнюю среду, имеющий в каждой точке направление нормали (см. стр. 109) к этой поверхности.

Скалярная величина P, численно равная силе, действующей на единицу площади поверхности по направлению нормали к этой поверхности называется *давлением*. Направление результирующей силы $\overline{\mathbf{f}}$ воздействия системы на внешнюю среду в каждой точке поверхности, очевидно, совпадает с направлением градиента в этой точке, поэтому давление вычисляется по формуле

$$P = \frac{\overline{\mathbf{F}}}{\operatorname{grad} V} \tag{9.13}$$

В результате мы можем записать

$$dA = PdV. \tag{9.14}$$

Рассмотрим теперь член $\theta\, d\eta$. Вообще говоря, выбор модуля распределения θ достаточно произволен. Самым известным модулем распределения для энергии является $\theta = kT$, где k — *постоянная Больцмана*, T — *абсолютная температура* (от лат. temperatura — надлежащее смешение, нормальное состояние), т.е. величина модуля распределения зависит от состояния системы. Тогда выражение, найденное в 1901 году американским физиком–теоретиком Джозайя Уиллардом Гиббсом (1839–1903)

$$\rho(E) \overset{(9.4)}{=} e^{\frac{\psi - E}{\theta}} \overset{(9.5)}{=} Ae^{-\frac{E}{kT}}, \tag{9.15}$$

где A — нормирующий множитель,
называют *распределением Гиббса*.

Обратим внимание еще раз. Распределение Гиббса определяет вероятность различных состояний статистической системы. При этом

1) равновесию соответствует наиболее вероятное распределение;

2) вероятность некого состояния определяется только его энергией.

Выражение $-k\eta$ называется *энтропией* (от греч. entropia — поворот, превращение) системы и обычно обозначается $S = -k\eta$.

Полную энергию замкнутой статистической системы E называют *внутренней энергией* и обозначают U.

Тогда получаем

$$dU \overset{(9.10)}{=} TdS - PdV, \tag{9.16}$$

где

$$T = \left(\frac{\partial U}{\partial S}\right)_V, \quad P = -\left(\frac{\partial U}{\partial V}\right)_S. \tag{9.17}$$

Физический смысл величин T, V и P вполне понятен из личного опыта каждого человека. Физический смысл энтропии S требует дополнительных пояснений.

Допустим, мы рассматриваем не ансамбль систем со средней плотностью N, а одну систему, которая может находиться в одном из N равновероятных состояний. Если принять величину ρ как вероятность системы находиться в каком либо определенном состоянии, то величину $N = \rho^{-1}$ можно интерпретировать как общее число состояний, в которых может находиться система. Тогда из определения показателя экспоненты (9.9) запишем выражение для энтропии

$$S = -k\,\eta = k \ln \rho^{-1} = k \ln N, \tag{9.18}$$

которое называется её статистическим истолкованием.

Статистическое истолкование энтропии дал Людвиг Больцман (1872), и формула (9.18) запечатлена на его могиле в Вене. Однако насколько известно, сам Больцман никогда эту формулу в явном виде не писал. Эту формулу дал Планк (1900) в своих знаменитых лекциях по теории теплового излучения, подчеркнув тогда, что формула определяет энтропию однозначно, не оставляя неопределённости в выборе постоянной k. Постоянная Больцмана k понадобилась для соблюдения размерности $\frac{\text{Дж}}{K}$, где Дж — «джоуль» единица измерения энергии, K — «кельвин» единица измерения абсолютной температуры.

Исторически более раннее понятие энтропии (точнее изменения энтропии системы dS) было введено Р. Клаузиусом (1865) как результат деления изменения количества *теплоты* в системе dQ, на абсолютную температуру системы T:

$$dS = \frac{dQ}{T}. \tag{9.19}$$

Таким образом, Р. Клаузиус определил энтропию при помощи теплоты. До конца 18 в. теплоту считали материальной субстанцией, полагая, что температура тела определяется количеством содержащейся в нем «калорической жидкости», или «теплорода». Позднее Б. Румфорд, Дж. Джоуль и другие физики того времени путем остроумных опытов и рассуждений

опровергли «калорическую» теорию, доказав, что теплота невесома и является просто кинетической частью внутренней энергии вещества, определяемая интенсивным хаотическим движением молекул и атомов, из которых это вещество состоит.

Понятие энтропии прямо противоположно понятию информации. В случае N равновероятных вариантов состояния величина энтропии (9.18) с точностью до постоянной Больцмана совпадает с величиной информации (6.35) $I = \ln N$. При этом информацию называют мерой упорядоченности системы, а энтропию — мерой ее неупорядоченности.

Так два понятия, имеющих прямо противоположные смыслы, имеют, тем не менее, равные значения их величин. Хотя величины энтропии и информации равны, их изменения в системе равны только по абсолютной величине, но противоположны по знаку. Когда система полностью неупорядочена, ее энтропия максимальна, а информация равна нулю. Когда состояние системы полностью определено (система полностью упорядочена) ее энтропия равна нулю, а информация, соответственно, максимальна. Следовательно (с точностью до константы размерности)

$$dS = -dI. \tag{9.20}$$

Пары переменных T, S и P, V называют *термодинамически взаимными величинами* или, в более общем смысле, — *сопряжёнными переменными*. При этом величины S, V называют *координатами*, величины P, T — *силами*. Координаты и силы называют также *термодинамическими параметрами* или *параметрами состояния* системы.

Статистическую систему, зависящую от указанных параметров, называют *термодинамической системой* или *телом*.

Термодинамика (от греч. *θέρμη* — тепло, *δύναμις* — сила) — наука о наиболее общих тепловых свойствах макроскопических тел (термодинамических систем), первая начала заниматься изучением этих параметров.

Термодинамические параметры определяют *состояние* термодинамической системы, которое характеризуется поверхностью состояния или *уравнением состояния* термодинамической системы:

$$f(P, V, T) = const, \quad \text{или} \quad f(P, V, S) = const. \tag{9.21}$$

Если хотя бы один из параметров системы изменяется, то происходит *термодинамический процесс*. Процессы термодинамической системы делят на равновесные и неравновесные. *Равновесный процесс* происходит при выполнении условия статистического равновесия (9.3) $\dot{\rho} = 0$ или $\rho = const$. В соответствии с определением энтропии (9.18), это означает, что при равновесном процессе энтропия системы сохраняется $dS = 0$ или $S = const$. Если условие статистического равновесия нарушается, то говорят, что в системе происходит *неравновесный процесс*. Как оказалось для всех исследованных систем, неравновесные процессы могут идти только с возрастанием энтропии, т.е. при условии $dS > 0$. Поэтому в реальных системах *энтропия*

никогда не уменьшается, т.е. число возможных состояний системы может только увеличиваться $dN > 0$.

Равновесные процессы называют еще *обратимыми*, так как система может пройти обратно через все промежуточные состояния процесса. Неравновесный процесс называется *необратимым*, так как процесс изменения энтропия может проходить только в направлении ее возрастания, обратный процесс убывания энтропии невозможен.

Условие $dU = dA = dQ = 0$ задает замкнутую (изолированную) систему. Если в системе происходит изменение внутренней энергии $dU \neq 0$, то система называется неизолированной, а при изменении количества вещества в системе $dG \neq 0$ — открытой. С введением понятия теплоты вводится понятие термически изолированной системы. Неизолированная система, для которой $dU \neq 0$, называется *термически изолированной*, если для нее выполняется условие $dQ = 0$ или $Q = const$, т.е. в системе существуют так называемые адиабатные поверхности. Экспериментально установлено, что *адиабатные поверхности* (от греч. $\alpha\delta\iota\alpha\beta\alpha\tau o\varsigma$ — непереходимый) являются поверхностями уровня, т.е. как поверхности энергии или любые другие поверхности уровня не пересекаются между собой.

Нарушение статистического равновесия или, что почти то же самое, *термодинамического равновесия* в системе возникает в том случае, если изменение координат системы произошло скачком или, по крайней мере, быстрее, чем статистическое равновесие успело наступить. Если после этого оставить систему в покое, т.е. сделать ее замкнутой или, по крайней мере, изолированной (т.е. находящейся в постоянных внешних условиях), то через некоторое время система обязательно придет в состояние равновесия. Промежуток времени, в течение которого должен обязательно произойти переход к статистическому равновесию, называют *временем релаксации*. Теорию неравновесных процессов, связанных с переходом в состояние равновесия, называют *кинетикой*. Таким образом, равновесные процессы в системе, это просто такие процессы, при которых происходит достаточно медленное изменение координат, чтобы в системе успевало устанавливаться статистическое равновесие.

Рассмотрим теперь замкнутую систему ($dU = dA = 0$), которая находится в неравновесном состоянии. Через некоторое время система придёт в равновесное состояние. При этом, ввиду замкнутости системы, у неё не изменится ни объём, ни внутренняя энергия, однако увеличится энтропия. Процесс установления равновесия завершится, когда энтропия достигнет максимального значения. Это значит, что нам нужно уточнить уравнение (9.16), которое мы вычисляли в предположении, что система находится в статистическом равновесии. Так как все члены уравнения кроме TdS нулевые, а $dS \geqslant 0$, то для расширения этого уравнения на неравновесные системы нам нужно записать неравенство

$$dU \leqslant TdS - PdV. \tag{9.22}$$

Если система замкнута, то она и термически изолирована ($dQ = 0$). Это значит, что для учёта неравновесных процессов нам необходимо также

уточнить и определение (9.19) энтропии при помощи теплоты

$$dS \geqslant \frac{dQ}{T}. \tag{9.23}$$

9.3 Термодинамические потенциалы

Рассмотрим термодинамическую систему (тело), помещённую во внешнюю среду. Это означает, что в рассматриваемая система уже не изолирована и может взаимодействовать с внешней средой. Если система не находится в равновесном состоянии, то в ней происходит термодинамический процесс установления равновесия. Процесс установления термодинамического равновесия произвольной (незамкнутой) системы зависит от *условий взаимодействия* (*сопряжения*) системы с внешней средой.

Процесс при постоянном объеме $V = const$, когда система не совершает внешней работы расширения, называется *изохорическим*; при $P = const$ — *изобарическим*; при $T = const$ — *изотермическим*. Система совершает *адиабатический* или *изоэнтропийный* процесс, если он обратим, и если система термически изолирована, т.е. $Q = const, S = const$.

Диаграммами состояния вещества называются *изолинии* на координатной плоскости полученные сечением поверхности состояния (9.21) соответствующей плоскостью $V, T, P, S = const$ (см. рис. 3.3).

Например, *изохора* — изолиния при постоянном объеме на плоскости P, T; *изобара* — при постоянном давлении на плоскости V, T; *изотерма* — при постоянной температуре на плоскости P, V, *изоэнтропа* (*адиабата*) — при постоянной энтропии на плоскости P, V.

Обратим внимание ещё раз. Процесс установления термодинамического равновесия в системе может протекать равновесно (адиабатически) при $dS = 0$ или неравновесно при $dS > 0$, но условия сопряжения системы с внешней средой при этом могут быть различными, и критерии установления равновесия системы будут зависеть от этих условий.

Наибольший практический интерес представляют четыре типа условий взаимодействия (сопряжения с внешней средой):

$$V = const, \quad S = const; \tag{9.24}$$

$$V = const, \quad T = const; \tag{9.25}$$

$$P = const, \quad S = const; \tag{9.26}$$

$$P = const, \quad T = const. \tag{9.27}$$

Каждый из этих четырёх случаев установления равновесия удобно описывать при помощи специальной термодинамической функции.

В неравновесной системе при условиях (9.24) из неравенства (9.22) следует, что $dU \leqslant 0$. Таким образом, при условиях сопряжения с внешней средой (9.24) внутренняя энергия неравновесной системы будет уменьшаться и достигнет минимального значения в состоянии равновесия.

Такое поведение внутренней энергии системы очень напоминает поведение потенциальной энергии механической системы. Термодинамическую функцию, которая при соответствующих условиях взаимодействия с внешней средой будет вести себя подобным образом называют *термодинамическим потенциалом* или *характеристической функцией*.

Преобразование Лежандра (3.152) $F = U - TS$ даёт уравнение

$$dF = dU - TdS - SdT$$

или неравенство с учётом неравновесного состояния

$$dF \overset{(9.22)}{\leqslant} -S\,dT - PdV, \tag{9.28}$$

из которого следует, что функция F — *энергия Гельмгольца*, *свободная энергия* или *изохорно-изотермический потенциал* при условиях сопряжения с внешней средой (9.25) будет уменьшаться ($dF \leqslant 0$) и достигнет минимального значения в состоянии равновесия.

Преобразование Лежандра $H = U + PV$ даёт с учётом неравновесного состояния неравенство

$$dH \leqslant TdS + V\,dP, \tag{9.29}$$

из которого следует, что функция H — *энтальпия* или *теплосодержание* при условиях (9.26) будет уменьшаться ($dH \leqslant 0$) и достигнет минимального значения в состоянии равновесия.

Преобразование Лежандра $\Phi = U + PV - TS$ даёт с учётом неравновесного состояния неравенство

$$d\Phi \leqslant -S\,dT + V\,dP, \tag{9.30}$$

из которого следует, что функция Φ — *энергия Гиббса* или *изобарно-изотермический потенциал* при условиях (9.27) будет уменьшаться ($d\Phi \leqslant 0$) и достигнет минимального значения в состоянии равновесия.

Поведение четырёх рассмотренных термодинамических потенциалов аналогично поведению потенциала механической системы — каждый из потенциалов при соответствующих условиях взаимодействия с внешней средой стремится к минимуму, производя *за счёт собственной убыли* работу над внешней средой. *Максимальная работа*, возможная *за счёт неравновесия* исходного состояния, достигается при равновесном процессе, если система (тело) в каждый момент процесса находится в равновесном состоянии, но, конечно, не в равновесии со средой. Например, максимальная работа при условиях (9.24) (за вычетом работы расширения, так как $dV = 0$) равна убыли внутренней энергии ΔU, при условиях (9.25) равна убыли ΔF и т.д. Таким образом, оказывается, что максимальная работа за счет неравновесности системы, которая может быть совершена телом (системой) над объектом при выбраных условиях сопряжения со средой равна убыли соответствующего термодинамического потенциала. Этот вывод есть *основной признак потенциала* и обуславливает название характеристической функции — потенциал.

Каждый термодинамический потенциал зависит от своего набора переменных: $U(S,V)$, $F(T,V)$, $H(S,P)$ и $\Phi(T,P)$. Эти переменные называются *естественными независимыми переменными*. Если термодинамический потенциал задан как функция естественных независимых переменных, например $U(S,V)$, то термодинамические свойства системы определены полностью, а график (S,V,U) представляет собой *двумерную поверхность состояния системы* в трёхмерном пространстве, которую можно использовать вместо уравнения состояния (9.21). Вместо координаты P в первом случае мог быть использован потенциал F, а во втором — потенциал U. Если же потенциал задан как функция другого набора переменных, то для полного определения термодинамических свойств этого недостаточно.

Обратим внимание ещё раз. Равновесие термодинамической системы характеризуется максимальным значением энтропии при любых условиях взаимодействия системы с внешней средой. При определённых условиях сопряжения с внешней средой равновесие системы также характеризуется минимум соответствующего термодинамического потенциала. В § 9.5 мы увидим, как с помощью термодинамического потенциала исследовать устойчивость термодинамического равновесия системы относительно малых вариаций её термодинамических параметров.

Заметим, что выражения (9.28, 9.29) при равновесном процессе (когда вместо неравенств можно записать уравнения) аналогичны соответствующим уравнениям (5.38, 5.39, 5.46, 5.47) или (7.2, 7.3) механики:

$$
\begin{aligned}
dU &= T\,dS - P\,dV, & T &= \left(\frac{\partial U}{\partial S}\right)_V, & P &= -\left(\frac{\partial U}{\partial V}\right)_S,\\
dH &= \dot{\mathbf{q}}\,d\mathbf{p} - \dot{\mathbf{p}}\,d\mathbf{q}, & \dot{\mathbf{q}} &= \left(\frac{\partial H}{\partial \mathbf{p}}\right)_{\mathbf{q}}, & \dot{\mathbf{p}} &= -\left(\frac{\partial H}{\partial \mathbf{q}}\right)_{\mathbf{p}},\\
d[-F] &= S\,dT + P\,dV, & S &= \left(\frac{\partial [-F]}{\partial T}\right)_V, & P &= \left(\frac{\partial [-F]}{\partial V}\right)_T,\\
dL &= \mathbf{p}\,d\dot{\mathbf{q}} + \dot{\mathbf{p}}\,d\mathbf{q}, & \mathbf{p} &= \left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right)_{\mathbf{q}}, & \dot{\mathbf{q}} &= \left(\frac{\partial L}{\partial \mathbf{p}}\right)_{\dot{\mathbf{q}}}.
\end{aligned}
$$

Аналогии здесь возможны, если имеют место соответствия, представленные в таблице 9.1.

Тот факт, что величины $\mathbf{p}, \mathbf{q}, \dot{\mathbf{p}}, \dot{\mathbf{q}}$ — векторы, а величины S, V, P, T — скаляры, в данном случае не имеет значения, так как во всех дифференциальных уравнениях присутствует скалярное произведение векторов.

Таким образом, переход от динамической системы к статистической системе можно еще интерпретировать и как замену векторных переменных скалярными переменными. С другой стороны, все последующие уравнения термодинамики справедливы и для незамкнутой динамической системы при обратной замене переменных.

Координаты S, V и потенциалы U, F, H, Φ зависят от количества вещества в системе и поэтому являются *аддитивными* (*экстенсивными*) *величинами*. Силы T, P не зависят от количества вещества в системе, и поэтому

Термодинамика	**Механика**
Объём V	Радиус-вектор $\mathbf{q}$
Энтропия S	Импульс $\mathbf{p}$
Температура T	Скорость $\mathbf{v} = \dot{\mathbf{q}}$
Давление P	Сила $\mathbf{f} = \dot{\mathbf{p}}$
Внутренняя энергия $U(S,V)$	Гамильтониан $H(\mathbf{p},\mathbf{q})$
Свободная энергия $[-F(T,V)]$	Лагранжиан $L(\dot{\mathbf{q}},\mathbf{q})$
Аналога нет	Кинетическая энергия $T(\dot{\mathbf{q}})$
Аналога нет	Потенциальная энергия $U(\mathbf{q})$
Энтальпия $H(S,P)$	Аналога нет
Энергия Гиббса $\Phi(T,P)$	Аналога нет

Таблица 9.1: Аналогии механики и термодинамики

являются *интенсивными величинами*. Значения экстенсивных величин для всей термодинамической системы в целом могут быть представлены в виде произведения *удельной* (отнесенной к единице вещества) *величины* на количество вещества в системе

$$S = sG,\ V = vG,\ U = uG,\ F = fG,\ H = hG,\ \Phi = \varphi G, \tag{9.31}$$

где G — количество вещества в системе,
v — удельный объем,
s — удельная энтропия,
u — удельная внутренняя энергия,
f — удельная свободная энергия,
h — удельная энтальпия,
φ — удельный изобарно-изотермический потенциал или *химический потенциал*.

Выражения для удельных потенциалов имеют вид:

$$f = u - Ts, \quad h = u + Pv, \quad \varphi = u + Pv - Ts. \tag{9.32}$$

Если количество вещества в системе неизменно $G = const$, то при равновесном процессе вместо неравенств мы можем записать дифференциалы удельных термодинамических потенциалов:

$$du = Tds - Pdv, \qquad T = \left(\frac{\partial u}{\partial s}\right)_v, \qquad P = -\left(\frac{\partial u}{\partial v}\right)_s; \tag{9.33}$$

$$df = -s\,dT - Pdv, \qquad s = -\left(\frac{\partial f}{\partial T}\right)_v, \qquad P = -\left(\frac{\partial f}{\partial v}\right)_T; \tag{9.34}$$

$$dh = Tds + v\,dP, \qquad T = \left(\frac{\partial h}{\partial s}\right)_P, \qquad v = \left(\frac{\partial h}{\partial P}\right)_s; \tag{9.35}$$

$$d\varphi = -s\,dT + v\,dP, \qquad s = -\left(\frac{\partial \varphi}{\partial T}\right)_P, \qquad v = \left(\frac{\partial \varphi}{\partial P}\right)_T. \tag{9.36}$$

Точно такой же вид будут иметь уравнения для полных значений V, S, U, F, H, Φ, относящихся ко всей термодинамической системе в целом.

Запишем теперь соотношения

$$\left(\frac{\partial s}{\partial v}\right)_u \overset{(3.161)}{=} -\left(\frac{\partial u}{\partial v}\right)_s \left(\frac{\partial s}{\partial u}\right)_v \overset{(9.33)}{=} \frac{P}{T}, \tag{9.37}$$

$$\left(\frac{\partial s}{\partial P}\right)_h \overset{(3.161)}{=} -\left(\frac{\partial h}{\partial P}\right)_s \left(\frac{\partial s}{\partial h}\right)_P \overset{(9.35)}{=} -\frac{v}{T}. \tag{9.38}$$

Примером процесса при $u = const$ является расширение газа в пустоту. Примером изоэнтальпийного процесса (при $h = const$) является *процесс Джоуля–Томсона*. Рассмотрим теплоизолированный сосуд с газом, разделенный пористой перегородкой на две части. В одной части поддерживается давление P_1, а в другой P_2. Газ будет перетекать в часть сосуда с меньшим давлением. Если отверстия в перегородке достаточно малы, то скорость макроскопического течения газа можно считать равной нулю. Процесс перетекания газа необратим, что видно уже из наличия перегородки с маленькими отверстиями, которая создает большое трение, уничтожающее скорость газа. Энтальпия или теплосодержание $h = u + Pv$ газа при перетекании не изменяется $h_1 = h_2$.

Величина (9.37) всегда положительна, а величина (9.38) всегда отрицательна, как это и должно быть. Переход к большему объему при $u = const$ или меньшему давлению при $h = const$ является необратимым процессом и сопровождается увеличением энтропии.

По теореме о дифференцировании сложной функции получим

$$\frac{\partial}{\partial x}\left(\frac{f}{x}\right)_y \overset{(3.13)}{=} \frac{1}{x}\left(\frac{\partial f}{\partial x}\right)_y - \frac{f}{x^2} \quad \text{или} \quad f - x\left(\frac{\partial f}{\partial x}\right)_y = -x^2 \left(\frac{\partial\left(\frac{f}{x}\right)}{\partial x}\right)_y.$$

Тогда один потенциал можно вычислять из другого:

$$u = f + Ts = f - T\left(\frac{\partial f}{\partial T}\right)_v = -T^2\left(\frac{\partial\left(\frac{f}{T}\right)}{\partial T}\right)_v, \tag{9.39}$$

$$f = \varphi - Pv = \varphi - P\left(\frac{\partial \varphi}{\partial P}\right)_T = -P^2\left(\frac{\partial\left(\frac{\varphi}{P}\right)}{\partial P}\right)_T, \tag{9.40}$$

$$h = u + Pv = u - v\left(\frac{\partial u}{\partial v}\right)_s = -v^2\left(\frac{\partial\left(\frac{u}{v}\right)}{\partial v}\right)_s, \tag{9.41}$$

$$\varphi = h - Ts = h - s\left(\frac{\partial h}{\partial s}\right)_P = -s^2\left(\frac{\partial\left(\frac{h}{s}\right)}{\partial s}\right)_P, \tag{9.42}$$

и так далее.

Количество тепла, при получении которого температура тела повышается на единицу, носит название *общей теплоёмкости* тела C_Z, а отнесенное к единице количества вещества — *теплоёмкости* c_Z:

$$C_Z = \left(\frac{\partial Q}{\partial T}\right)_Z = T\left(\frac{\partial S}{\partial T}\right)_Z, \quad c_Z = \frac{1}{G}\left(\frac{\partial Q}{\partial T}\right)_Z = T\left(\frac{\partial s}{\partial T}\right)_Z, \tag{9.43}$$

где Z — условие измерения теплоёмкости,

$G = const$ — количество вещества в системе.

В зависимости от условий $Z = const$, при которых происходит процесс нагревания системы, мы будем вычислять различные величины теплоемкости. Наиболее распространенными являются *изохорная теплоёмкость* c_v и *изобарная теплоёмкость* c_P:

$$c_Z = T\left(\frac{\partial s}{\partial T}\right)_Z, \quad c_v = T\left(\frac{\partial s}{\partial T}\right)_v, \quad c_P = T\left(\frac{\partial s}{\partial T}\right)_P. \tag{9.44}$$

Изохорная и изобарная теплоёмкости вычисляются как частные производные соответствующих потенциалов:

$$c_v = \left(\frac{\partial u}{\partial T}\right)_v, \quad c_P = \left(\frac{\partial h}{\partial T}\right)_P. \tag{9.45}$$

Доказательство

$$c_v \overset{(9.44)}{=} T\left(\frac{\partial s}{\partial T}\right)_v \overset{(9.33)}{=} \left(\frac{\partial u}{\partial s}\right)_v \left(\frac{\partial s}{\partial T}\right)_v \overset{(3.157)}{=} \left(\frac{\partial u}{\partial T}\right)_v,$$
$$c_P \overset{(9.44)}{=} T\left(\frac{\partial s}{\partial T}\right)_P \overset{(9.35)}{=} \left(\frac{\partial h}{\partial s}\right)_P \left(\frac{\partial s}{\partial T}\right)_P \overset{(3.157)}{=} \left(\frac{\partial h}{\partial T}\right)_P.$$

Также, запишем выражения:

$$c_v = \left(\frac{\partial h}{\partial T}\right)_v - v\left(\frac{\partial P}{\partial T}\right)_v, \quad c_P = \left(\frac{\partial u}{\partial T}\right)_P + P\left(\frac{\partial v}{\partial T}\right)_P. \tag{9.46}$$

Доказательство

$$\left(\frac{\partial u}{\partial T}\right)_P \overset{(3.159)}{=} \left(\frac{\partial u}{\partial s}\right)_v \left(\frac{\partial s}{\partial T}\right)_P + \left(\frac{\partial u}{\partial v}\right)_s \left(\frac{\partial v}{\partial T}\right)_P \overset{\substack{(9.33)\\(9.44)}}{=} c_P - P\left(\frac{\partial v}{\partial T}\right)_P,$$
$$\left(\frac{\partial h}{\partial T}\right)_v \overset{(3.159)}{=} \left(\frac{\partial h}{\partial s}\right)_P \left(\frac{\partial s}{\partial T}\right)_v + \left(\frac{\partial h}{\partial P}\right)_s \left(\frac{\partial P}{\partial T}\right)_v \overset{\substack{(9.35)\\(9.44)}}{=} c_v + v\left(\frac{\partial P}{\partial T}\right)_v.$$

9.4 Законы термодинамики

Нулевое начало термодинамики или закон транзитивности теплового равновесия — при неизменных внешних условиях температура всех тел системы выравнивается и наступает *термодинамическое равновесие*. Если какое

либо тело A находится в тепловом равновесии с телом C, а тело B также находится в тепловом равновесии с телом C, то тела A и B также находятся в состоянии равновесия. Именно поэтому мы можем сравнивать состояния теплового равновесия тел, не приводя их в непосредственный контакт, а измеряя их температуру.

Уравнение

$$dU = dQ - dA \tag{9.47}$$

называется *первым законом термодинамики*, который с учетом (9.19) записывается в уже известном нам виде (9.33).

Первый закон термодинамики является разновидностью закона сохранения энергии и означает, что внутренняя энергия системы в *каждый момент движения* системы должна возрастать на количество подводимого к системе тепла dQ и уменьшаться на количество производимой системой работы dA. Тогда изменение внутренней энергии ΔU определяется разностью $\Delta Q - \Delta A$ при *любом* процессе, при котором система переходит из исходного состояния в рассматриваемое. Это означает, что интеграл $\int dU$ не зависит от пути интегрирования, а *дифференциал dU является полным*. Из этого следует, что невозможно построить машину, которая могла бы создавать энергию, но возможно превратить энергию из одного вида в другой, в частности внутренняя энергия системы может переходить в теплоту или в механическую работу.

Условия для равновесных процессов $dS = 0$ и для неравновесных процессов $dS > 0$ вместе с определением энтропии через теплоту (9.19) дают уже полученный ранее (9.23) *второй закон термодинамики*

$$TdS \geqslant dQ. \tag{9.48}$$

Второй закон термодинамики означает, что изменение энтропии либо пропорционально подводимому к системе теплу либо больше этой величины и ни в коем случае не меньше нуля, даже если тепло отводится от системы. Исторически второй закон термодинамики существует в нескольких формулировках. Постулат Кельвина: невозможен процесс, единственным конечным результатом которого будет превращение в работу теплоты, извлеченной из источника, имеющего всюду одинаковую температуру. Постулат Клаузиуса: невозможен процесс, единственным конечным результатом которого был бы переход теплоты от более холодного тела к более горячему телу. Постулат К. Каратеодори (1909) или принцип адиабатной недостижимости: адиабатные поверхности не пересекаются между собой, следовательно, в любой близости от всякого состояния системы существуют смежные состояния, которые не могут быть достигнуты адиабатным путем. Может показаться, что Каратеодори сформулировал второй закон термодинамики чисто теоретически, без привлечения каких-либо опытных фактов (т.е. не постулировал, а «доказал» этот закон). В действительности это не так. Принцип адиабатной недостижимости является постулатом (т.е. предложением, принимаемым без доказательства) в такой же мере, в какой являются постулатами принципы Клаузиуса, Кельвина и неубывания энтропии.

На математическом языке все формулировки второго закона термодинамики означают, что *дифференциал функции теплоты*

$$dQ = dU + dA = dU + PdV, \tag{9.49}$$

как правило, *не является полным* $\oint dQ \neq 0$.

Действительно, согласно (3.87) для пфаффовой формы (9.49) должно выполняться условие $\left(\frac{\partial 1}{\partial V}\right)_U = \left(\frac{\partial P}{\partial U}\right)_V$.

Однако, как правило, это условие не выполняется

$$\left(\frac{\partial 1}{\partial V}\right)_U = 0 \quad \neq \quad \left(\frac{\partial P}{\partial U}\right)_V \overset{(3.157)}{=} \left(\frac{\partial P}{\partial T}\right)_V \left(\frac{\partial T}{\partial U}\right)_V \overset{(9.45)}{=} \frac{1}{C_v}\left(\frac{\partial P}{\partial T}\right)_V.$$

Так как $\oint dU = 0$, то тогда $\oint dA \neq 0$. Таким образом теплота Q и работа A являются *функциями процесса* в отличие от функции состояния U. Поэтому можно говорить о внутренней энергии системы в данном состоянии, но нельзя говорить, например, о количестве тепла, которым обладает в данном состоянии система. Другими словами, внутреннюю энергию нельзя делить на тепловую и механическую.

Рассмотрим теплоизолированную систему $dQ = 0$, состоящую из нескольких тел, не находящихся друг с другом в тепловом равновесии. В течение процесса установления равновесия система может совершать работу над какими-либо внешними объектами. При этом переход в равновесие может совершаться различным образом. В зависимости от пути системой будет совершена различная работа, и, следовательно, конечная энергия и энтропия системы будут различными.

Вычислим, каким образом должен произойти переход в равновесное состояние, для того, чтобы система произвела *наибольшую возможную работу*. При этом мы интересуемся именно той работой R, которая производится за счет неравновесности системы; это значит, что надо исключить работу A, которая могла бы быть произведена системой, находящейся в равновесии, например, за счет общего расширения системы, т.е. мы рассматриваем процесс при $dV = 0$, а также при $TdS \geqslant dQ = 0$ для теплоизолированной системы. Получаем

$$dU \overset{(9.47)}{=} -dR \overset{(9.16)}{=} TdS \quad \text{или} \quad \left(\frac{\partial R}{\partial S}\right)_V = -\left(\frac{\partial U}{\partial S}\right)_V = -T.$$

Мы видим, что эта производная отрицательна, т.е. R уменьшается с увеличением S. Поэтому максимальное значение R может быть достигнуто, если S останется в течение всего процесса неизменной, т.е. переход тел в состояние теплового равновесия совершается обратимым образом.

Рассмотрим два тела не находящихся в тепловом равновесии и имеющих, поэтому, различные температуры T_1 и T_2, пусть $T_2 > T_1$. Если передача энергии будет происходить непосредственно при соприкосновении тел, то никакой работы вообще не будет произведено. Процесс будет необратимым, и энтропия обоих тел увеличится на $\Delta S = \Delta U\left(\frac{1}{T_1} - \frac{1}{T_2}\right)$.

Для обратимой передачи энергии и совершении при этом максимальной работы необходимо добавить в систему ещё одно вспомогательное или *рабочее тело*, совершающее обратимый круговой процесс. Процесс этот должен происходить следующим образом:

- рабочее тело при температуре T_2 изотермически получает от второго тела определенную энергию,
- рабочее тело адиабатически охлаждается до температуры T_1,
- рабочее тело при температуре T_1 изотермически отдает определенную энергию первому телу,
- рабочее тело адиабатически нагревается до температуры T_2.

Описанный круговой процесс максимально возможного использования внутренней энергии для получения работы называют *цикл Карно*.

При вычислении максимальной работы рабочее тело можно не учитывать, поскольку оно возвращается в исходное состояние. Более нагретое тело теряет энергию $\Delta U_2 = T_2 \Delta S_2$. Более холодное тело получает энергию $\Delta U_1 = -T_1 \Delta S_1$. Ввиду обратимости процесса сумма энтропий обоих тел остается неизменной, т.е. $\Delta S_2 = -\Delta S_1$. Произведенная работа равна уменьшению энергии обоих тел

$$\Delta R = \Delta U_2 - \Delta U_1 = T_2 \Delta S_2 + T_1 \Delta S_1 = T_2 \Delta S_2 - T_1 \Delta S_2 = \\ = \frac{T_2 - T_1}{T_2} \Delta U_2 = \eta \Delta U_2$$

где $\eta = \frac{T_2 - T_1}{T_2}$ — *коэффициент полезного действия* при переходе энергии от более нагретого тела к менее нагретому.

Коэффициент полезного действия η является максимально возможным. Все остальные возможные процессы преобразования внутренней энергии системы в работу будут иметь меньший к.п.д.

Дифференциал dQ можно превратить в полный дифференциал при помощи интегрирующего множителя, который в случае функции двух переменных всегда можно найти. Интегрирующий множитель $\frac{1}{T}$ (9.19) превращает dQ (9.49) в полный дифференциал dS:

$$dS = \frac{1}{T}\, dU + \frac{P}{T}\, dV, \quad \frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)_V, \quad \frac{P}{T} = \left(\frac{\partial S}{\partial V}\right)_U. \tag{9.50}$$

Выражение (9.50) является полным дифференциалом, так как

$$\left(\frac{\partial \left(\frac{1}{T}\right)}{\partial V}\right)_U = \left(\frac{\partial \left(\frac{P}{T}\right)}{\partial U}\right)_V = 0.$$

Таким образом, энтропия S — *функция состояния* системы, и изменение энтропии не зависит от того, каким путем произошло это изменение,

а зависит только от начального и конечного состояния системы. Величины T, P и V также являются *функциями состояния*. Любая функция, образованная простой алгебраической комбинацией функций состояния, сама является функцией состояния, и, следовательно, ее дифференциал является полным дифференциалом. Так как U — *функция состояния*, следовательно и все остальные термодинамические потенциалы: свободная энергия системы F, энтальпия системы H и изобарно-изотермический потенциал Φ — также *функции состояния*.

Из вывода о том, что термодинамические потенциалы есть функции состояния вытекают уравнения, впервые сформулированные Дж .Максвеллом, которые служат чрезвычайно важным инструментом термодинамики. Эти *уравнения Максвелла* получаются из уравнений (9.33–9.36) и условий на полные дифференциалы (3.87) (когда перекрестные вторые производные равны между собой). Уравнения также могут быть записаны в «перевёрнутом» виде (см. 3.157):

$$\left(\frac{\partial T}{\partial v}\right)_s = -\left(\frac{\partial P}{\partial s}\right)_v, \qquad \left(\frac{\partial v}{\partial T}\right)_s = -\left(\frac{\partial s}{\partial P}\right)_v, \tag{9.51}$$

$$\left(\frac{\partial s}{\partial v}\right)_T = \left(\frac{\partial P}{\partial T}\right)_v, \qquad \left(\frac{\partial v}{\partial s}\right)_T = \left(\frac{\partial T}{\partial P}\right)_v, \tag{9.52}$$

$$\left(\frac{\partial T}{\partial P}\right)_s = \left(\frac{\partial v}{\partial s}\right)_P, \qquad \left(\frac{\partial P}{\partial T}\right)_s = \left(\frac{\partial s}{\partial v}\right)_P, \tag{9.53}$$

$$\left(\frac{\partial v}{\partial T}\right)_P = -\left(\frac{\partial s}{\partial P}\right)_T, \qquad \left(\frac{\partial T}{\partial v}\right)_P = -\left(\frac{\partial P}{\partial s}\right)_T. \tag{9.54}$$

Точно такой же вид будут иметь уравнения Максвелла для значений V и S, относящихся к термодинамической системе в целом.

Приведём примеры использования уравнений Максвелла.

Покажем, каким образом можно установить зависимость $T = T(\tau)$ между абсолютной шкалой температуры и некоторой чисто условной шкалой τ, определяемой произвольно градуированным термометром. В этом случае, по крайней мере, в принципе, мы построим термодинамическую шкалу температуры, используя для этого произвольное тело, уравнение состояния которого заранее не предполагается известным.

$$\left(\frac{\partial Q}{\partial P}\right)_T \overset{(9.19)}{=} T\left(\frac{\partial S}{\partial P}\right)_T \overset{(9.51)}{=} -T\left(\frac{\partial V}{\partial T}\right)_P \overset{(3.157)}{=} -T\left(\frac{\partial V}{\partial \tau}\right)_P \left(\frac{\partial \tau}{\partial T}\right)_P.$$

Поскольку τ и T связаны друг с другом взаимно однозначно, то

$$\left(\frac{\partial Q}{\partial P}\right)_T = -T\left(\frac{\partial V}{\partial \tau}\right)_P \frac{d\tau}{dT} = \left(\frac{\partial Q}{\partial P}\right)_\tau \text{ или } \frac{d\ln T}{d\tau} = -T\left(\frac{\partial V}{\partial \tau}\right)_P \left(\frac{\partial P}{\partial Q}\right)_\tau,$$

где в правой части равенства стоят величины, которые могут быть непосредственно измерены как функции условной температуры τ.

Экспериментальное определение теплоёмкости c_v сопряжено со значительными трудностями, поэтому для её вычисления по известным значениям c_P используют уравнение

$$c_P - c_v = T\left(\frac{\partial P}{\partial T}\right)_v \left(\frac{\partial v}{\partial T}\right)_P. \tag{9.55}$$

Доказательство

$$\begin{aligned}\frac{c_P}{T} &\overset{(9.44)}{=} \left(\frac{\partial s}{\partial T}\right)_P \overset{(3.160)}{=} \left(\frac{\partial s}{\partial T}\right)_v + \left(\frac{\partial s}{\partial v}\right)_T \left(\frac{\partial v}{\partial T}\right)_P = \\ &\overset{(9.44)}{=} \frac{c_v}{T} + \left(\frac{\partial s}{\partial v}\right)_T \left(\frac{\partial v}{\partial T}\right)_P \overset{(9.52)}{=} \frac{c_v}{T} + \left(\frac{\partial P}{\partial T}\right)_v \left(\frac{\partial v}{\partial T}\right)_P.\end{aligned}$$

Для идеального газа, состояние которого, как известно, описывается *уравнением Клапейрона* $PV = RT$, находим *уравнение Майера*

$$c_P - c_V = R. \tag{9.56}$$

Доказательство

$$c_P - c_V \overset{(9.55)}{=} T\left(\frac{\partial P}{\partial T}\right)_V \left(\frac{\partial V}{\partial T}\right)_P = T\left(\frac{\partial \frac{RT}{V}}{\partial T}\right)_V \left(\frac{\partial \frac{RT}{P}}{\partial T}\right)_P = T\frac{R}{V}\frac{R}{P} = R.$$

Для u запишем

$$\begin{aligned}\left(\frac{\partial u}{\partial v}\right)_T &\overset{(3.160)}{=} \left(\frac{\partial u}{\partial s}\right)_v \left(\frac{\partial s}{\partial v}\right)_T + \left(\frac{\partial u}{\partial v}\right)_s = \\ &\overset{(9.33)}{=} T\left(\frac{\partial s}{\partial v}\right)_T - P \overset{(9.52)}{=} T\left(\frac{\partial P}{\partial T}\right)_v - P.\end{aligned} \tag{9.57}$$

Тогда получаем

$$u(v,T) \overset{(4.7)}{=} u(v_0,T) + \int\limits_{v_0}^{v} \left[T\left(\frac{\partial P}{\partial T}\right)_v - P\right] dv. \tag{9.58}$$

Соответственно, для h запишем

$$\begin{aligned}\left(\frac{\partial h}{\partial P}\right)_T &\overset{(3.160)}{=} \left(\frac{\partial h}{\partial P}\right)_s + \left(\frac{\partial h}{\partial s}\right)_P \left(\frac{\partial s}{\partial P}\right)_T = \\ &\overset{(9.35)}{=} v + T\left(\frac{\partial s}{\partial P}\right)_T \overset{(9.54)}{=} v - T\left(\frac{\partial v}{\partial T}\right)_P.\end{aligned} \tag{9.59}$$

Тогда получаем

$$h(P,T) \overset{(4.7)}{=} h(P_0,T) + \int\limits_{P_0}^{P} \left[v - T\left(\frac{\partial v}{\partial T}\right)_P\right] dP. \tag{9.60}$$

Заметим, что в обоих случаях (9.58) и (9.60) интегрирование ведётся вдоль изотермы.

Располагая экспериментальными данными о термических (P, v, T) свойствах вещества можно вычислить интегралы, стоящие в правых частях уравнений (9.58) и (9.60), и теплоёмкость c_v в уравнении (9.55). При этом нужно предварительно вычислить производные $\left(\frac{\partial P}{\partial T}\right)_v$ и $\left(\frac{\partial v}{\partial T}\right)_P$.

Запишем уравнения

$$\left(\frac{\partial c_P}{\partial P}\right)_T = -T\left(\frac{\partial^2 v}{\partial T^2}\right)_P, \quad \left(\frac{\partial c_v}{\partial v}\right)_T = T\left(\frac{\partial^2 P}{\partial T^2}\right)_v. \tag{9.61}$$

Доказательство

$$\left(\frac{\partial c_P}{\partial P}\right)_T \overset{(9.45)}{=} \frac{\partial^2 h}{\partial T \partial P} \overset{(9.59)}{=} \frac{\partial}{\partial T}\left[v - T\left(\frac{\partial v}{\partial T}\right)_P\right] \overset{(3.6)}{=} -T\left(\frac{\partial^2 v}{\partial T^2}\right)_P.$$

$$\left(\frac{\partial c_v}{\partial v}\right)_T \overset{(9.45)}{=} \frac{\partial^2 u}{\partial T \partial v} \overset{(9.57)}{=} \frac{\partial}{\partial T}\left[T\left(\frac{\partial P}{\partial T}\right)_v - P\right] \overset{(3.6)}{=} T\left(\frac{\partial^2 P}{\partial T^2}\right)_v.$$

Уравнения (9.61) определяют зависимости на изотерме — теплоёмкости c_P от давления, а теплоёмкости c_v от объёма.

Тогда, при интегрировании вдоль изотермы, получаем

$$c_P(P,T) \overset{(4.7)}{=} c_P(P_0,T) - T\int\limits_{P_0}^{P}\left(\frac{\partial^2 v}{\partial T^2}\right)_P dP. \tag{9.62}$$

$$c_v(v,T) \overset{(4.7)}{=} c_v(v_0,T) + T\int\limits_{v_0}^{v}\left(\frac{\partial^2 P}{\partial T^2}\right)_v dv. \tag{9.63}$$

Так как величина энтропии определяется во втором законе только с точностью до постоянной интегрирования, то полезно сформулировать *третий закон термодинамики* (теорему Нернста–Планка): энтропия химически однородного тела конечной плотности при стремлении температуры к абсолютному нулю стремится к предельному значению, не зависящему от давления, плотности или фазы. Удобно поэтому принимать состояние при 0°K за некоторое начальное состояние и считать

$$\lim_{T\to 0} S = S_0 \equiv 0. \tag{9.64}$$

Тогда энтропия произвольного состояния определяется однозначным образом. Найденную таким образом энтропию называют иногда *абсолютной энтропией*.

Из теоремы Нернста–Планка можно вывести несколько важных следствий. Во-первых, никакими способами нельзя достичь абсолютного нуля температуры (недостижимость абсолютного нуля) и, во-вторых, теплоёмкость C_Z и некоторые другие аналогичные величины должны стремиться к нулю, при температуре, стремящейся к нулю.

Рассмотрим произвольный обратимый процесс Z, связывающий состояние системы при абсолютном нуле с состоянием A, энтропию которого мы вычисляем. Тогда

$$S(A) \overset{(9.43)}{=} \int_0^{T_A} C_Z \frac{dT}{T}, \quad \text{тогда} \quad \lim_{T_A \to 0} C_Z \overset{(9.64)}{=} 0,$$

что экспериментально подтвердилось для всех исследованных систем.

Теорема Нернста–Планка имеет также большое значение для изучения химических реакций.

С точки зрения статистической механики, в которой используется атомистический подход, третий закон является следствием квантовомеханической природы микроскопического движения материи. Если говорить об истории вопроса, то эта теорема была выведена из целого ряда экспериментальных фактов и поэтому представляет собой, как и другие законы термодинамики, эмпирически установленный принцип.

Нулевой, первый, второй и третий законы образуют систему основных аксиом, на которых покоится логическая схема термодинамики.

9.5 Принцип Ле-Шателье–Брауна

При рассмотрении термодинамических потенциалов в § 9.3 мы считали, что процесс установления термодинамического равновесия начинается, если температура и (или) давление внешней среды и погружённого в неё тела различаются между собой. Нулевой закон термодинамики говорит о том, что в системе устанавливается термодинамическое (тепловое) рвновесие, если во всех её частях температура одинакова.

В настоящем параграфе мы попытаемся более строго (формально) сформулировать необходимые и достаточные условия установления термодинамического равновесия.

Рассмотрим систему, состоящую из среды и погруженного в нее тела. Роль «тела» может играть произвольно выбранная подсистема замкнутой системы. Пусть x и y некие термодинамические величины, а $z(x, y)$ есть некая функция состояния (обычно это соответствующий переменным x, y термодинамический потенциал). Тогда

$$dz = X\,dx + Y\,dy, \quad X = \left(\frac{\partial z}{\partial x}\right)_y, \quad Y = \left(\frac{\partial z}{\partial y}\right)_x. \tag{9.65}$$

Для полного дифференциала функции состояния обязательно равенство вторых частных производных (см. 3.87)

$$\frac{\partial^2 z}{\partial x \partial y} = \left(\frac{\partial X}{\partial y}\right)_x = \left(\frac{\partial Y}{\partial x}\right)_y. \tag{9.66}$$

В состоянии равновесия термодинамической системы её потенциал достигает своего минимального значения при соответствующих условиях сопряжения системы с внешней средой.

Пусть функции $A = const$ и $B = const$ представляют собой геометрические связи (условия), для отыскания условного экстремума (см. с. 152) — точки равновесия на поверхности $z(x, y)$. Причём векторы $\operatorname{grad} A$ и $\operatorname{grad} B$ линейно независимы. Например:

$$\left(\frac{\partial A}{\partial x}\right)_y = X_0, \quad \left(\frac{\partial A}{\partial y}\right)_x = 0, \quad \left(\frac{\partial B}{\partial x}\right)_y = 0, \quad \left(\frac{\partial B}{\partial y}\right)_x = Y_0.$$

Пусть $B = const$ означает, что система сама по себе находится в равновесии, не находясь при этом обязательно в равновесии со средой, а $A = const$ и $B = const$ означает, что система находится не только в своем внутреннем равновесии, но также и в равновесии со средой.

Запишем функцию Лагранжа

$$L = z + \lambda A + \mu B, \tag{9.67}$$

где λ и μ — множители Лагранжа.

Необходимое условие экстремума в этом случае имеет вид:

$$\left(\frac{\partial L}{\partial x}\right)_y = \left(\frac{\partial z}{\partial x}\right)_y + \lambda\left(\frac{\partial A}{\partial x}\right)_y + \mu\left(\frac{\partial B}{\partial x}\right)_y = X + \lambda X_0 = 0,$$
$$\left(\frac{\partial L}{\partial y}\right)_x = \left(\frac{\partial z}{\partial y}\right)_x + \lambda\left(\frac{\partial A}{\partial y}\right)_x + \mu\left(\frac{\partial B}{\partial y}\right)_x = Y + \mu Y_0 = 0,$$

или

$$X = X_0, \quad Y = Y_0, \quad \mu = -1, \quad \lambda = -1. \tag{9.68}$$

Если тело находится в своем внутреннем равновесии, то для него необходимо условие $Y = Y_0$. Если же тело находится не только в своем внутреннем равновесии, но также и в равновесии со средой, то для него необходимы условия $X = X_0$ и $Y = Y_0$. Заметим, что на термодинамические величины x и y функцией Лагранжа никаких условий не накладывается.

В качестве примера рассмотрим какую-нибудь изолированную термодинамическую систему, например, газ и жидкость, заключённые в термостате (сосуде с нетеплопроводящими стенками).

Независимо от своего начального состояния термически изолированная ($dQ = 0$) система в конечном итоге приходит в состояние, которое в дальнейшем не меняется ($dS = 0$). Это внутреннее состояние равновесия в данном случае, очевидно, означает, что температура во всех частях сосуда стала одинаковой $T = const$.

Если же теперь необходимо, чтобы система находилась еще и в равновесии с внешней средой, то, во-первых, наш сосуд тогда не будет теплоизолирован, и температура системы должна сравняться с температурой внешней среды $T = T_0$, во-вторых, не только температура, но и давление системы должно уравняться с давлением внешней среды $P = P_0$, и тогда система перестанет совершать работу над средой $dA = 0$. При этом функции $Q = const$ и $A = const$ являются необходимыми условиями равновесия

системы со средой, их градиенты $\frac{\partial Q}{\partial S} = T_0$ и $\frac{\partial A}{\partial V} = P_0$ — линейно независимы.

Заметим, что равенства удельных энтропий s и объёмов v тела и внешней среды для установления равновесия не требуется. Таким образом, *для равновесия требуется равенство термодинамических сил, но не требуется равенства термодинамических координат.* Этот факт используется при изучении фазовых переходов.

Рассмотрим теперь достаточные условия условного экстремума при установлении равновесия системы со средой.

Пусть $Y = const$ означает, что тело само по себе находится в равновесии, не находясь при этом обязательно в равновесии со средой, а $X = X_0 = const$ и $Y = Y_0 = const$ означает, что тело находится не только в своем внутреннем равновесии, но также и в равновесии со средой.

Вторая квадратичная форма (3.132) функции $z(x, y)$ (9.65) имеет вид

$$d^2 z = \left(\frac{\partial Y}{\partial y}\right)_x dy^2 + 2\left(\frac{\partial X}{\partial y}\right)_x dx\, dy + \left(\frac{\partial X}{\partial x}\right)_y dx^2. \tag{9.69}$$

Достаточные условия существования равновесия системы предполагает, что система находится в эллиптической точке на поверхности состояния, где дискриминант уравнения (9.69) больше нуля (см. 3.135):

$$\left(\frac{\partial X}{\partial x}\right)_y \left(\frac{\partial Y}{\partial y}\right)_x - \left(\frac{\partial X}{\partial y}\right)_x^2 > 0. \tag{9.70}$$

Предположим теперь, что путем какого-либо внешнего воздействия нарушается равновесие тела (подсистемы) со средой, причем величина x изменяется на Δx и нарушается условие $X = const$. О величине же y предполагаем, что она данным воздействием непосредственно не затрагивается. Тогда изменение величины X в момент воздействия будет

$$(\Delta X)_y = \left(\frac{\partial X}{\partial x}\right)_y \Delta x. \tag{9.71}$$

Изменение x при постоянном y приводит, конечно, к нарушению также и условия $Y = const$, т.е. внутреннего равновесия тела. После того, как это внутреннее равновесие снова восстановится, величина ΔX будет иметь значение

$$(\Delta X)_Y = \left(\frac{\partial X}{\partial x}\right)_Y \Delta x. \tag{9.72}$$

Сравним оба значения ΔX при помощи свойств якобианов:

$$\left(\frac{\partial X}{\partial x}\right)_Y \overset{(3.164)}{=} \frac{\partial(X,Y)}{\partial(x,Y)} \overset{(3.165)}{=} \frac{\frac{\partial(X,Y)}{\partial(x,y)}}{\frac{\partial(x,Y)}{\partial(x,y)}} \overset{(3.162)}{=} \frac{\left(\frac{\partial Y}{\partial y}\right)_x \left(\frac{\partial X}{\partial x}\right)_y - \left(\frac{\partial X}{\partial y}\right)_x^2}{\left(\frac{\partial Y}{\partial y}\right)_x}.$$

В результате получаем:

$$\left(\frac{\partial X}{\partial x}\right)_Y \left(\frac{\partial Y}{\partial y}\right)_x = \left(\frac{\partial X}{\partial x}\right)_y \left(\frac{\partial Y}{\partial y}\right)_x - \left(\frac{\partial X}{\partial y}\right)_x^2 \overset{(9.70)}{>} 0$$

При условии $\left(\frac{\partial Y}{\partial y}\right)_x > 0$ это даёт

$$\left(\frac{\partial X}{\partial x}\right)_y > \left(\frac{\partial X}{\partial x}\right)_Y > 0 \quad \text{или} \quad |(\Delta X)_y| > |(\Delta X)_Y|, \tag{9.73}$$

где пары величин X, x и Y, y термодинамически взаимны.

Неравенство (9.73) составляет содержание так называемого *принципа Ле-Шателье–Брауна*: внешнее воздействие, выводящее систему из равновесия, стимулирует в ней процессы, стремящиеся ослабить результаты этого воздействия.

Пусть, например, $z(x, y) = u(v, s)$, $x = v$, $y = s$, $X = -P$ и $Y = T$.

При очевидном условии $\left(\frac{\partial T}{\partial s}\right)_v \overset{(9.44)}{=} \frac{T}{C_v} > 0$ получаем известный в термодинамике результат $\left(\frac{\partial(-P)}{\partial v}\right)_s > \left(\frac{\partial(-P)}{\partial v}\right)_T > 0$ или

$$\left(\frac{\partial P}{\partial v}\right)_T < \left(\frac{\partial P}{\partial v}\right)_s < 0 \quad \text{или} \quad |(\Delta P)_s| < |(\Delta P)_T|. \tag{9.74}$$

Смысл этих неравенств заключается в следующем. Если тело выводится из равновесия путем изменения его объема (при неизменной температуре), то меняется, в частности, его давление. Восстановление равновесия в теле приводит к уменьшению абсолютной величины изменения давления. Имея в виду, что уменьшения объема тела увеличивает его давление, можно сказать, что уменьшение объема тела стимулирует в нем процессы, стремящиеся уменьшить его давление и наоборот. Производная $\left(\frac{\partial P}{\partial v}\right)_T$ всегда отрицательна — при изотермическом расширении тела его давление всегда падает.

Если теперь принять $z(x, y) = u(s, v)$, $x = s$, $y = v$, $X = T$ и $Y = -P$, то при $\left(\frac{\partial P}{\partial v}\right)_s < 0$ получаем

$$\left(\frac{\partial T}{\partial s}\right)_v > \left(\frac{\partial T}{\partial s}\right)_P > 0 \quad \text{или} \quad |(\Delta T)_v| > |(\Delta T)_P|. \tag{9.75}$$

В данном случае изменение удельной энтропии тела означает, что телу сообщается или от тела отнимается некоторое количество тепла (при неизменном объеме). В результате нарушается равновесие самого тела и, в частности, изменяется его температура. Восстановление равновесия в теле приведет к тому, что изменение его температуры по абсолютной величине уменьшится. Можно сказать, что нагревание тела стимулирует в нем процессы, стремящиеся понизить его температуру и наоборот.

С учетом (9.44) получим из (9.75) очень известное неравенство:

$$c_P > c_v > 0. \tag{9.76}$$

Принцип смещения равновесия при изменении температуры установил Я. Вант-Гофф (J. van't Hoff) в 1884 году. Общий принцип, отражающий влияние различных факторов на положение термодинамического равновесия, сформулировали Анри Ле Шателье (1884) и Карл Фердинанд Браун (1887). Они применяли главным образом индуктивный метод, рассмотрев большое число примеров термодинамических равновесий, которые, по их мнению, являются частными случаями сформулированного ими общего правила, и пытались представить эти примеры в форме, похожей на правило Ленца в электродинамике. Данная ими формулировка была, однако, столь туманной, что не допускала в каждом конкретном случае однозначного применения правила.

9.6 Сложная и открытая системы

Допустим теперь, что система воздействует на окружающую среду таким образом, что изменяются не только пространственные координаты самой системы, но и *дополнительные* величины (*координаты*) $W_1, W_2, \ldots$ внешней среды. Тогда в дополнение к уравнениям (9.6) возникают дополнительные компоненты силы, соответствующие этим координатам

$$\xi_1 = -\frac{\partial U}{\partial W_1}, \quad \xi_2 = -\frac{\partial U}{\partial W_2}, \quad \ldots. \tag{9.77}$$

Пары величин ξ_i, W_i по аналогии с T, S и P, V, также называют термодинамически взаимными величинами.

Термодинамическая система, единственным видом работы которой является работа расширения PdV, называется *простой системой*. Термодинамическая система, совершающая помимо работы расширения другие виды работы $\xi_i\, dW_i$, называется *сложной системой*.

При рассмотрении сложной системы в уравнении (9.8) возникают дополнительные члены, соответствующие дополнительным координатам:

$$d\psi \overset{(9.77)}{=} -\sum_i \left(\int \xi_i \rho\, dV\right) dW_i - \left(\int \mathbf{f}\rho\, dV\right) d\mathbf{q} = -\sum_i \overline{\xi}_i\, dW_i - \overline{\mathbf{F}}\, d\mathbf{q}.$$

В результате уравнение (9.16) будет иметь вид

$$dU = Td - PdV - \overline{\xi}_1\, dW_1 - \overline{\xi}_2\, dW_2 - \ldots. \tag{9.78}$$

Обычно без потери общности достаточно рассмотрения одной внешней координаты, чем мы и будем пользоваться

$$dU = TdS - PdV - \xi\, dW = TdS - dA, \quad dA = PdV + dA^*, \tag{9.79}$$

где $dA^* = \xi\, dW$ — любой вид работы, совершаемой системой, за исключением работы расширения.

Для сложных систем вводят дополнительные потенциалы:

$$H^* = H + \xi W, \quad \Phi^* = \Phi + \xi W, \tag{9.80}$$

а также удельные дополнительные координаты и потенциалы:

$$W = wG, \quad H^* = h^*G, \quad \Phi^* = \varphi^*G. \tag{9.81}$$

Полные дифференциалы и частные производные для термодинамических потенциалов сложных систем для равновесных процессов при $G = const$ легко получить при помощи преобразования Лежандра (3.152) из уравнений (9.33–9.36):

$$du = Tds - Pdv - \xi\, dw, \tag{9.82}$$

$$T = \left(\frac{\partial u}{\partial s}\right)_{v,w}, \qquad P = -\left(\frac{\partial u}{\partial v}\right)_{s,w}, \quad \xi = -\left(\frac{\partial u}{\partial w}\right)_{s,v}; \tag{9.83}$$

$$df = -s\, dT - Pdv - \xi\, dw, \tag{9.84}$$

$$s = -\left(\frac{\partial f}{\partial T}\right)_{v,w}, \qquad P = -\left(\frac{\partial f}{\partial v}\right)_{T,w}, \quad \xi = -\left(\frac{\partial f}{\partial w}\right)_{T,v}; \tag{9.85}$$

$$dh^* = Tds + v\, dP + w\, d\xi, \tag{9.86}$$

$$T = \left(\frac{\partial h^*}{\partial s}\right)_{P,\xi}, \qquad v = \left(\frac{\partial h^*}{\partial P}\right)_{s,\xi}, \quad w = \left(\frac{\partial h^*}{\partial \xi}\right)_{s,P}; \tag{9.87}$$

$$d\varphi^* = -s\, dT + v\, dP + w\, d\xi, \tag{9.88}$$

$$s = -\left(\frac{\partial \varphi^*}{\partial T}\right)_{P,\xi}, \qquad v = \left(\frac{\partial \varphi^*}{\partial P}\right)_{T,\xi}, \quad w = \left(\frac{\partial \varphi^*}{\partial \xi}\right)_{T,P}. \tag{9.89}$$

Из уравнений (9.82–9.88) и условий на полные дифференциалы (3.87) (когда перекрестные вторые производные равны между собой) выводятся *уравнения Максвелла* для сложных систем:

$$\left(\frac{\partial T}{\partial v}\right)_{s,w} = -\left(\frac{\partial P}{\partial s}\right)_{v,w}, \qquad \left(\frac{\partial T}{\partial w}\right)_{s,v} = -\left(\frac{\partial \xi}{\partial s}\right)_{v,w}, \tag{9.90}$$

$$\left(\frac{\partial P}{\partial w}\right)_{s,v} = \left(\frac{\partial \xi}{\partial v}\right)_{s,w}, \qquad \left(\frac{\partial s}{\partial w}\right)_{T,v} = \left(\frac{\partial \xi}{\partial T}\right)_{v,w}, \tag{9.91}$$

$$\left(\frac{\partial s}{\partial v}\right)_{T,w} = \left(\frac{\partial P}{\partial T}\right)_{v,w}, \qquad \left(\frac{\partial P}{\partial w}\right)_{T,v} = \left(\frac{\partial \xi}{\partial v}\right)_{T,w}, \tag{9.92}$$

$$\left(\frac{\partial T}{\partial P}\right)_{s,\xi} = \left(\frac{\partial v}{\partial s}\right)_{P,\xi}, \qquad \left(\frac{\partial T}{\partial \xi}\right)_{s,P} = \left(\frac{\partial w}{\partial s}\right)_{P,\xi}, \tag{9.93}$$

$$\left(\frac{\partial v}{\partial \xi}\right)_{s,P} = \left(\frac{\partial w}{\partial P}\right)_{s,\xi}, \qquad \left(\frac{\partial v}{\partial \xi}\right)_{T,P} = \left(\frac{\partial w}{\partial P}\right)_{T,\xi}, \tag{9.94}$$

$$\left(\frac{\partial s}{\partial P}\right)_{T,\xi} = -\left(\frac{\partial v}{\partial T}\right)_{P,\xi}, \qquad \left(\frac{\partial s}{\partial \xi}\right)_{T,P} = -\left(\frac{\partial w}{\partial T}\right)_{P,\xi}. \tag{9.95}$$

Уравнения также могут быть записаны в «перевёрнутом» виде (см. 3.157). Такой же вид будут иметь уравнения Максвелла для значений V, S и W, относящихся к термодинамической системе в целом.

Если сложная система находится неравновесном состоянии, тогда уравнения для термодинамических потенциалов нужно переписать в виде неравенств аналогично тому, как уравнение (9.16) для равновесного состояния простой системы нужно записывать в виде неравенства (9.22) для неравновесного состояния этой системы:

$$dU \leqslant TdS - PdV - \xi\, dW, \tag{9.96}$$

$$dF \leqslant -S\, dT - PdV - \xi\, dW, \tag{9.97}$$

$$dH^* \leqslant TdS + V\, dP - W\, d\xi, \tag{9.98}$$

$$d\Phi \leqslant -S\, dT + V\, dP - W\, d\xi. \tag{9.99}$$

Из полученных неравенств следует, что термодинамические потенциалы сложной системы при установлении равновесия ($dS = 0$) смогут достигнуть минимального значения при следующих условиях сопряжения системы с внешней средой:

$$V = const, \quad S = const, \quad W = const \qquad \text{для } U; \tag{9.100}$$

$$V = const, \quad T = const, \quad W = const \qquad \text{для } F; \tag{9.101}$$

$$P = const, \quad S = const, \quad \xi = const \qquad \text{для } H^*; \tag{9.102}$$

$$P = const, \quad T = const, \quad \xi = const \qquad \text{для } \Phi^*. \tag{9.103}$$

Рассмотрим примеры сложных систем:

- Для термодинамической системы «магнетик в магнитном поле» обобщенной силой является отрицательная напряженность $-H$ *магнитного поля*, а обобщенной координатой — намагниченность $\mathcal{M}$ *магнетика*. Тогда

$$du = Tds - Pdv + H\, d\mathcal{M}. \tag{9.104}$$

Уравнения Максвелла для системы в магнитном поле имеют вид

$$\left(\frac{\partial \mathcal{M}}{\partial T}\right)_{s,v} = \left(\frac{\partial s}{\partial H}\right)_{\mathcal{M},v}, \qquad \left(\frac{\partial \mathcal{M}}{\partial s}\right)_{H,p} = -\left(\frac{\partial T}{\partial H}\right)_{s,P}, \tag{9.105}$$

$$\left(\frac{\partial \mathcal{M}}{\partial s}\right)_{T,v} = -\left(\frac{\partial T}{\partial H}\right)_{\mathcal{M},v}, \qquad \left(\frac{\partial \mathcal{M}}{\partial T}\right)_{H,P} = \left(\frac{\partial s}{\partial H}\right)_{T,P}, \tag{9.106}$$

и т.д.

Изобарная теплоёмкость магнетика при постоянных H и j определяется как

$$c_{H,P} = T\left(\frac{\partial s}{\partial T}\right)_{H,P}, \quad c_{\mathcal{M},P} = T\left(\frac{\partial s}{\partial T}\right)_{\mathcal{M},P}. \tag{9.107}$$

Магнитокалорический эффект — изменение температуры магнетика при изменении напряжённости внешнего магнитного поля $\frac{\partial T}{\partial H}$ вычисляется как

$$\left(\frac{\partial T}{\partial H}\right)_{s,P} = -\frac{TH}{c_{H,P}}\left(\frac{\partial \chi}{\partial T}\right)_{H,P}. \tag{9.108}$$

Так как теплоёмкость $c_{H,P}$ всегда положительна, то знак адиабатного магнитокалорического эффекта определяется знаком $\left(\frac{\partial \chi}{\partial T}\right)_{H,P}$.

Доказательство

$$\left(\frac{\partial T}{\partial H}\right)_{s,P} \overset{(3.161)}{=} -\left(\frac{\partial T}{\partial s}\right)_{H,P}\left(\frac{\partial s}{\partial H}\right)_{T,P} = \overset{\substack{(9.107)\\(9.106)}}{=} -\frac{T}{c_{H,P}}\left(\frac{\partial \mathcal{M}}{\partial T}\right)_{H,P} \overset{(8.22)}{=} -\frac{TH}{c_{H,P}}\left(\frac{\partial \chi}{\partial T}\right)_{H,P}.$$

- Для термодинамической системы «диэлектрик в электрическом поле» обобщенной силой является отрицательная напряженность $-E$ *электрического поля*, а обобщенной координатой — поляризация $\mathcal{P}$ *диэлектрика*. Тогда

$$du = Tds - Pdv + E\,d\mathcal{P}.$$

Уравнения Максвелла для системы в электрическом поле имеют вид

$$\left(\frac{\partial \mathcal{P}}{\partial T}\right)_{s,v} = \left(\frac{\partial s}{\partial E}\right)_{\mathcal{P},v}, \qquad \left(\frac{\partial \mathcal{P}}{\partial s}\right)_{E,p} = -\left(\frac{\partial T}{\partial E}\right)_{s,P}, \tag{9.109}$$

$$\left(\frac{\partial \mathcal{P}}{\partial s}\right)_{T,v} = -\left(\frac{\partial T}{\partial E}\right)_{\mathcal{P},v}, \qquad \left(\frac{\partial \mathcal{P}}{\partial T}\right)_{E,P} = \left(\frac{\partial s}{\partial E}\right)_{T,P}, \tag{9.110}$$

$$\left(\frac{\partial v}{\partial E}\right)_{s,P} = -\left(\frac{\partial \mathcal{P}}{\partial P}\right)_{s,E}, \qquad \left(\frac{\partial v}{\partial E}\right)_{T,P} = -\left(\frac{\partial \mathcal{P}}{\partial P}\right)_{T,E}, \tag{9.111}$$

и т.д.

Известно, что для диэлектриков в электрическом поле характерны следующие термодинамические эффекты:

Пьезоэлектрический эффект — изменение поляризации диэлектрика при изменении внешнего давления $\frac{\partial \mathcal{P}}{\partial P}$.

Электрострикционный эффект — изменение объёма диэлектрика при изменении внешнего давления $\frac{\partial v}{\partial E}$.

Пироэлектрический эффект — изменение поляризации диэлектрика при изменении его температуры $\frac{\partial \mathcal{P}}{\partial T}$.

Электрокалорический эффект — изменение температуры диэлектрика при изменении напряжённости электрического поля $\frac{\partial T}{\partial E}$.

Уравнения (9.111) устанавливают связь между пьезоэлектрическим и электрострикционным эффектами при адиабатных и изотермических условиях.

Изобарная теплоёмкость диэлектрика при постоянных E и $\mathcal{P}$ определяется как

$$c_{E,P} = T\left(\frac{\partial s}{\partial T}\right)_{E,P}, \quad c_{\mathcal{P},P} = T\left(\frac{\partial s}{\partial T}\right)_{\mathcal{P},P}. \tag{9.112}$$

Связь между электрокалорическим эффектом и пироэлектрическим эффектом вычисляется как

$$\left(\frac{\partial T}{\partial E}\right)_{s,P} = -\frac{T}{c_{E,P}}\left(\frac{\partial \mathcal{P}}{\partial T}\right)_{E,P}. \tag{9.113}$$

Доказательство

$$\left(\frac{\partial T}{\partial E}\right)_{s,P} \overset{(3.161)}{=} -\left(\frac{\partial T}{\partial s}\right)_{E,P}\left(\frac{\partial s}{\partial E}\right)_{T,P} \overset{\substack{(9.112)\\(9.106)}}{=} -\frac{T}{c_{E,P}}\left(\frac{\partial \mathcal{P}}{\partial T}\right)_{E,P}.$$

- Для твердого тела (стержня) находящегося в условиях упругой деформации под действием растягивающей (сжимающей) силы Ψ, обобщенной силой является величина Ψ, а обобщенной координатой — длина стержня l.
- Для термодинамической системы в поле тяготения обобщенной силой является масса G, а обобщенной координатой — высота z центра тяжести системы.
- Для обратимого гальванического элемента обобщенной силой является электродвижущая сила $\mathcal{E}$, а обобщенной координатой — заряд элемента Z.

Рассмотрим теперь *открытую систему* или систему с переменным количеством вещества $G \neq const$, т.е. такую систему, которая может обмениваться с внешней средой не только энергией, но и веществом.

Часто можно встретиться с заблуждением, что из определения удельных величин (9.31) и (9.81), как может показаться, должен следовать тривиальный вывод о том, что $\left(\frac{\partial U}{\partial G}\right)_{V,S} = u$, $\left(\frac{\partial F}{\partial G}\right)_{V,T} = u$, $\left(\frac{\partial H}{\partial G}\right)_{P,S} = u$ и т.д. Однако этот, на первый взгляд, тривиальный вывод ошибочен. Если бы производные вычислялись при неизменных удельных координатах системы ($v = const, s = const, w = const$), то удельные потенциалы при этом также оставались бы неизменными, например, $\left(\frac{\partial U}{\partial G}\right)_{v,s} = u$.

Дело, однако, в том. что интересующие нас производные вычисляются при неизменности координат системы в целом ($V = const, S = const, W = const$). Между тем, если при этих условиях $G \neq const$, то меняются удельные величины координат, а следовательно, меняются и удельные значения

потенциалов. Отсюда ясно, что, например $\left(\frac{\partial U}{\partial G}\right)_{V,S} \neq \left(\frac{\partial U}{\partial G}\right)_{v,s}$. Необходимо отчётливо представить себе глубокую, принципиальную разницу между этими производными

$$u = \left(\frac{\partial U}{\partial G}\right)_{v,s} \neq \left(\frac{\partial U}{\partial G}\right)_{V,S} = \left(\frac{\partial (uG)}{\partial G}\right)_{V,S} = u + G\left(\frac{\partial u}{\partial G}\right)_{V,S}. \qquad (9.114)$$

Запишем

$$\left(\frac{\partial u}{\partial G}\right)_{V,S} \overset{(3.160)}{=} \left(\frac{\partial u}{\partial G}\right)_{V,s} + \left(\frac{\partial u}{\partial s}\right)_{V,G}\left(\frac{\partial s}{\partial G}\right)_{V,S} =$$
$$\overset{(3.160)}{=} \left(\frac{\partial u}{\partial G}\right)_{v,s} + \left(\frac{\partial u}{\partial v}\right)_{G,s}\left(\frac{\partial v}{\partial G}\right)_{V,s} + \left(\frac{\partial u}{\partial s}\right)_{V,G}\left(\frac{\partial s}{\partial G}\right)_{V,S}.$$

Рассмотрим частные производные в правой части уравнения:

$\left(\frac{\partial u}{\partial G}\right)_{v,s} \overset{(9.33)}{=} 0$, так как $u = const$ при $v, s = const$.

$\left(\frac{\partial u}{\partial v}\right)_{G,s} = \left(\frac{\partial u}{\partial v}\right)_{s} \overset{(9.33)}{=} -P.$

$\left(\frac{\partial v}{\partial G}\right)_{V,s} \overset{(9.31)}{=} \left(\frac{\partial \left(\frac{V}{G}\right)}{\partial G}\right)_{V,s} \overset{(3.9)}{=} -\frac{V}{G^2} \overset{(9.31)}{=} -\frac{v}{G}$, при $V = const$.

$\left(\frac{\partial v}{\partial s}\right)_{V,G} = \left(\frac{\partial v}{\partial s}\right)_{v} \overset{(9.33)}{=} T$, так как $v = const$ при $V, G = const$.

$\left(\frac{\partial s}{\partial G}\right)_{V,S} \overset{(9.31)}{=} \left(\frac{\partial \left(\frac{S}{G}\right)}{\partial G}\right)_{V,S} \overset{(3.9)}{=} -\frac{S}{G^2} \overset{(9.31)}{=} -\frac{s}{G}$, при $S = const$.

В результате получим

$$\left(\frac{\partial U}{\partial G}\right)_{V,S} \overset{(9.114)}{=} u + G\frac{Pv - Ts}{G} = u + Pv - Ts = \varphi.$$

Теперь запишем для потенциала F

$$\left(\frac{\partial F}{\partial G}\right)_{V,T} = \left(\frac{\partial (fG)}{\partial G}\right)_{V,T} = f + G\left(\frac{\partial f}{\partial G}\right)_{V,T}, \qquad (9.115)$$

где

$$\left(\frac{\partial f}{\partial G}\right)_{V,T} \overset{(3.160)}{=} \left(\frac{\partial f}{\partial G}\right)_{v,T} + \left(\frac{\partial f}{\partial v}\right)_{G,T}\left(\frac{\partial v}{\partial G}\right)_{V,T}.$$

Рассмотрим частные производные в правой части уравнения:

$$\left(\frac{\partial f}{\partial G}\right)_{v,T} \overset{(9.34)}{=} 0, \text{ так как } f = const \text{ при } v, T = const.$$

$$\left(\frac{\partial f}{\partial v}\right)_{G,T} = \left(\frac{\partial f}{\partial v}\right)_{T} \overset{(9.34)}{=} -P.$$

$$\left(\frac{\partial v}{\partial G}\right)_{V,T} \overset{(9.31)}{=} \left(\frac{\partial \left(\frac{V}{G}\right)}{\partial G}\right)_{V,T} \overset{(3.9)}{=} -\frac{V}{G^2} \overset{(9.31)}{=} -\frac{v}{G}, \quad \text{при } V = const.$$

В результате получим

$$\left(\frac{\partial F}{\partial G}\right)_{V,T} \overset{(9.115)}{=} f + G\frac{Pv}{G} = f + Pv = \varphi.$$

Теперь запишем для потенциала H

$$\left(\frac{\partial H}{\partial G}\right)_{P,S} = \left(\frac{\partial (hG)}{\partial G}\right)_{P,S} = h + G\left(\frac{\partial h}{\partial G}\right)_{P,S}, \tag{9.116}$$

где

$$\left(\frac{\partial h}{\partial G}\right)_{P,S} \overset{(3.160)}{=} \left(\frac{\partial h}{\partial G}\right)_{P,s} + \left(\frac{\partial h}{\partial s}\right)_{P,G}\left(\frac{\partial s}{\partial G}\right)_{P,S}.$$

Рассмотрим частные производные в правой части уравнения:

$$\left(\frac{\partial h}{\partial G}\right)_{P,s} \overset{(9.35)}{=} 0, \text{ так как } h = const \text{ при } P, s = const.$$

$$\left(\frac{\partial h}{\partial s}\right)_{P,G} = \left(\frac{\partial h}{\partial s}\right)_{P} \overset{(9.35)}{=} T.$$

$$\left(\frac{\partial s}{\partial G}\right)_{P,S} \overset{(9.31)}{=} \left(\frac{\partial \left(\frac{S}{G}\right)}{\partial G}\right)_{P,S} \overset{(3.9)}{=} -\frac{S}{G^2} \overset{(9.31)}{=} -\frac{s}{G}, \quad \text{при } V = const.$$

В результате получим

$$\left(\frac{\partial H}{\partial G}\right)_{P,S} \overset{(9.116)}{=} h - G\frac{Ts}{G} = h - Ts = \varphi.$$

И наконец

$$\left(\frac{\partial \Phi}{\partial G}\right)_{P,T} = \left(\frac{\partial (\varphi G)}{\partial G}\right)_{P,T} = \varphi + G\left(\frac{\partial \varphi}{\partial G}\right)_{P,T} = \varphi,$$

потому что

$$\left(\frac{\partial \varphi}{\partial G}\right)_{P,T} \overset{(9.36)}{=} 0, \text{ так как } \varphi = const \text{ при } P, T = const.$$

Таким образом, мы пришли к несколько неожиданному выводу о том, что производная от соответствующего потенциала по количеству вещества в системе равна φ:

$$\left(\frac{\partial U}{\partial G}\right)_{V,S} = \left(\frac{\partial F}{\partial G}\right)_{V,T} = \left(\frac{\partial H}{\partial G}\right)_{P,S} = \left(\frac{\partial \Phi}{\partial G}\right)_{P,T} = \varphi. \tag{9.117}$$

Величина φ (или φ^* в случае сложных систем) обладает замечательным свойством: она позволяет вычислять изменение характеристической функции любой системы при изменении количества вещества в системе; именно поэтому эта величина и получила название *химического потенциала*.

Полезно также вычислить производную энтропии по количеству вещества в системе

$$\left(\frac{\partial S}{\partial G}\right)_{U,V} = \left(\frac{\partial (sG)}{\partial G}\right)_{U,V} = s + G\left(\frac{\partial s}{\partial G}\right)_{U,V}, \tag{9.118}$$

где

$$\left(\frac{\partial s}{\partial G}\right)_{U,V} \overset{(3.160)}{=} \left(\frac{\partial s}{\partial G}\right)_{u,V} + \left(\frac{\partial s}{\partial u}\right)_{G,V}\left(\frac{\partial u}{\partial G}\right)_{U,V} = \overset{(3.160)}{=} \left(\frac{\partial s}{\partial G}\right)_{u,v} + \left(\frac{\partial s}{\partial v}\right)_{u,G}\left(\frac{\partial v}{\partial G}\right)_{u,V} + \left(\frac{\partial s}{\partial u}\right)_{G,V}\left(\frac{\partial u}{\partial G}\right)_{U,V}.$$

Рассмотрим частные производные в правой части уравнения:

$\left(\frac{\partial s}{\partial G}\right)_{u,v} \overset{(9.33)}{=} 0$, так как $s = const$ при $u, v = const$.

$\left(\frac{\partial s}{\partial v}\right)_{u,G} \overset{(9.31)}{=} \left(\frac{\partial S}{\partial V}\right)_{U} \overset{(9.50)}{=} \frac{P}{T}$.

$\left(\frac{\partial s}{\partial u}\right)_{G,V} \overset{(9.31)}{=} \left(\frac{\partial S}{\partial U}\right)_{V} \overset{(9.50)}{=} \frac{1}{T}$.

$\left(\frac{\partial u}{\partial G}\right)_{U,V} \overset{(9.31)}{=} \left(\frac{\partial \left(\frac{U}{G}\right)}{\partial G}\right)_{U,V} \overset{(3.9)}{=} -\frac{U}{G^2} \overset{(9.31)}{=} -\frac{u}{G}$, при $U = const$.

$\left(\frac{\partial v}{\partial G}\right)_{u,V} \overset{(9.31)}{=} \left(\frac{\partial \left(\frac{V}{G}\right)}{\partial G}\right)_{u,V} \overset{(3.9)}{=} -\frac{V}{G^2} \overset{(9.31)}{=} -\frac{v}{G}$, при $V = const$.

В результате получим

$$\left(\frac{\partial S}{\partial G}\right)_{U,V} \overset{(9.118)}{=} s + G\frac{-u - Pv}{TG} = \frac{sT - u - Pv}{T} = -\frac{\varphi}{T}.$$

Это соотношение относится к простой системе при $U, V = const$, $G \neq const$. Аналогичное соотношение может быть получено и для сложной системы при условиях $U, V, W = const$, $G \neq const$.

В результате мы можем записать

$$\left(\frac{\partial S}{\partial G}\right)_{U,V} = -\frac{\varphi}{T}, \quad \left(\frac{\partial S}{\partial G}\right)_{U,V,W} = -\frac{\varphi^*}{T}. \tag{9.119}$$

Полный дифференциал для энтропии сложной открытой системы в равновесном состоянии имеет вид

$$dS = \frac{1}{T}\,dU + \frac{P}{T}\,dV + \frac{\xi}{T}\,dW - \frac{\varphi^*}{T}\,dG;$$
$$\frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)_{V,W,G}, \frac{P}{T} = \left(\frac{\partial S}{\partial V}\right)_{U,W,G}, \frac{\xi}{T} = \left(\frac{\partial S}{\partial W}\right)_{U,V,G}, \frac{\varphi^*}{T} = \left(\frac{\partial S}{\partial G}\right)_{U,V,W}.$$

Полные дифференциалы термодинамических потенциалов простой и сложной открытой системы в равновесном состоянии имеют вид:

$$dU = TdS - PdV + \varphi\,dG, \tag{9.120}$$
$$dU = TdS - PdV - \xi\,dW + \varphi^* dG, \tag{9.121}$$
$$dF = -S\,dT - PdV + \varphi\,dG, \tag{9.122}$$
$$dF = -S\,dT - PdV - \xi\,dW + \varphi^* dG. \tag{9.123}$$
$$dH = -S\,dT + V dP + \varphi\,dG, \tag{9.124}$$
$$dH^* = TdS + V\,dP + W\,d\xi + \varphi^* dG. \tag{9.125}$$
$$d\Phi = -S\,dT + V\,dP + \varphi\,dG, \tag{9.126}$$
$$d\Phi^* = -S\,dT + V\,dP + W\,d\xi + \varphi^* dG. \tag{9.127}$$

Дифференциал dU (9.120) и (9.121) называют *фундаментальным уравнением Гиббса*. Если открытая система находится в неравновесном состоянии, то вместо уравнений надо записывать неравенства.

Уравнения Максвелла для открытой системы имеют вид

$$\left(\frac{\partial T}{\partial G}\right)_{S,V} = \left(\frac{\partial \varphi}{\partial S}\right)_{G,V}, \qquad \left(\frac{\partial P}{\partial G}\right)_{V,S} = -\left(\frac{\partial \varphi}{\partial V}\right)_{G,S}, \tag{9.128}$$
$$-\left(\frac{\partial S}{\partial G}\right)_{T,V} = \left(\frac{\partial \varphi}{\partial T}\right)_{G,V}, \qquad -\left(\frac{\partial P}{\partial G}\right)_{V,T} = \left(\frac{\partial \varphi}{\partial V}\right)_{G,T}, \tag{9.129}$$
$$\left(\frac{\partial T}{\partial G}\right)_{S,P} = \left(\frac{\partial \varphi}{\partial S}\right)_{G,P}, \qquad \left(\frac{\partial V}{\partial G}\right)_{P,S} = \left(\frac{\partial \varphi}{\partial P}\right)_{G,S}, \tag{9.130}$$

и т.д.

Запишем выражения:

$$S\,dT - V\,dP + G\,d\varphi = 0, \quad S\,dT - V\,dP - W\,d\xi + G\,d\varphi = 0. \tag{9.131}$$

Доказательство для первого уравнения

$$0 = d\Phi - d(U + PV - TS) \overset{(9.31)}{=} d(\varphi G) - d(U + PV - TS) =$$
$$\overset{(3.151)}{=} G\,d\varphi + \varphi\,dG - dU - PdV - V\,dP + TdS + S\,dT =$$
$$\overset{(9.120)}{=} G\,d\varphi - V dV - P\,dP + S\,dT.$$

Важные термодинамические соотношения (9.131) называются *уравнением Гиббса–Дюгема*. Уравнение получено Дж. У. Гиббсом в 1875 и широко применялось П. Дюгемом.

9.7 Фазовые переходы

Рассмотрим теперь случаи, когда термодинамическая система находится в равновесии, а её поверхность состояния, тем не менее, не является гладкой, т.е. испытывает изломы и разрывы.

При различных значениях внешних параметров вещество может находиться в различных *физических состояниях* — твёрдое, жидкое, газообразное, различные аллотропические модификации вещества в твердом состоянии (например, алмаз, графит и аморфный уголь), в сверхпроводнике (нормальное и сверхпроводящее состояние), в ферромагнетике (ферромагнитное и парамагнитное состояния), в сегнетоэлектрике (сегнетоэлектрическое и обычное диэлектрическое состояния) и т.д.

Различные равновесные состояния одного и того же вещества, различающиеся по физическим свойствам от других возможных равновесных состояний называются *фазами*. Обычно рассматривают пространственно однородные фазы. Это означает, что в пределах фазы вещество находится в одном физическом состоянии, и термин фаза в термодинамике означает по сути какое-либо физическое состояние вещества. Поверхность состояния (9.21) (например, $f(P,v,T) = const$ или $f(u,v,s) = const$) внутри каждой фазы представляет собой гладкую поверхность, называемую *однофазной областью*.

Переход вещества из одной фазы в другую называется *фазовый переход*. Фазовый переход связан с качественными изменениями свойств вещества. Обычно вещество может находиться в газовой, жидкой или твёрдой фазе; при этом некоторые вещества в твёрдом состоянии имеют несколько фаз. Фазовый переход жидкость→газ называется *испарением*, твёрдая фаза→газ — *сублимацией* (от лат. sublimo — высоко поднимаю, возношу), твёрдая фаза→жидкость — *плавлением*.

Граница однофазной области на поверхности состояния называется *пограничной кривой*, при пересечении которой (при фазовом переходе) термодинамическая поверхность может терять гладкость — испытывать изломы и разрывы. Каждая пограничная кривая на термодинамической поверхности данного вещества (фазы) однозначно фиксирована и является *одномерным пространством*, поэтому *любая термодинамическая величина в каждой точке пограничной линии является функцией только одной переменной*.

Будем считать, что при фазовом переходе термодинамическая система всегда находится в равновесии, т.е. мы рассматриваем только *равновесные фазовые переходы*.

Область между пограничными кривыми однофазных областей на поверхности состояния называется *двухфазной областью*. В двухфазной области часть вещества перешла из одной фазы в другую, например, часть

жидкости испарилась или замерзла. В результате такого перехода в системе стало одновременно существовать две фазы, которые тем не менее находятся в термодинамическом равновесии. Границей двухфазной области служат пограничные линии однофазных областей.

Таким образом, фазовый переход от одной фазы к другой включает в себя последовательное пересечение первой пограничной линии, двухфазной области и второй пограничной линии.

У сосуществующих фаз всегда равны температура и давление, что следует из необходимых условий существования термодинамического равновесия (см. с. 347), а также другие обобщённые силы. Соответственно, при фазовом переходе температура T и давление P и обобщённая сила ξ не испытывают изменений

$$T^{(1)} = T^{(2)}, \quad P^{(1)} = P^{(2)}, \quad \xi^{(1)} = \xi^{(2)}. \tag{9.132}$$

Очевидно, что если фазовый переход в простой системе происходит при постоянном давлении, то и температура в этом процессе сохраняется постоянной. Температура фазового перехода называется, соответственно, *температурой плавления* или *температурой кипения*.

Координатная плоскость P, T является диаграммой состояния термодинамической системы при фазовых переходах. Двухфазные области вместе с их пограничными кривыми проецируются на P, T-диаграмму в виде *кривых фазовых переходов*. Область проецируется в кривую, так как при постоянном давлении температура не меняется в течение всего фазового перехода. Кривые фазовых переходов представлены на рис. 9.1.

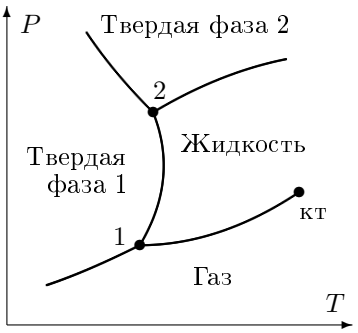


Рис. 9.1: P, T-диаграмма

Важной термодинамической величиной является *полная производная* $\frac{dP}{dT}$, характеризующая наклон кривой фазового перехода на P, T-диаграмме. При испарении и сублимации величина $\frac{dP}{dT}$ всегда положительна, а при плавления эта величина у одних веществ положительна, у других — отрицательна.

Кривая фазового перехода при испарении называется *кривая насыщения* (кривая между точками 1 и кт на рис. 9.1).

Тройная точка — точка пересечения кривых фазового равновесия на P, T-диаграмме состояния вещества (точки 1 и 2 на рис. 9.1), соответствующая устойчивому равновесию трёх фаз.

Критическая точка — точка на линии насыщения, в которой исчезает различие между жидкой и газовой фазами (точка кт на рис. 9.1). Критическая точка является конечной точкой линии насыщения, начинающейся в тройной точке.

Пограничную кривую со стороны жидкости принято называть *левой*, а относящиеся к ней величины снабжать индексом $'$ или $^{(1)}$. Пограничную кривую со стороны газа принято называть *правой*, а относящиеся к ней величины снабжать индексом $''$ или $^{(2)}$. В том случае, когда не указывается, к какой конкретно пограничной кривой относится та или иная величина на кривой насыщения, её записывают с индексом $^{\sigma}$. Величины однофазной области принято снабжать индексом ^{оф}, величины двухфазной области — индексом ^{дф}, величины критической точки — индексом ^{кр}.

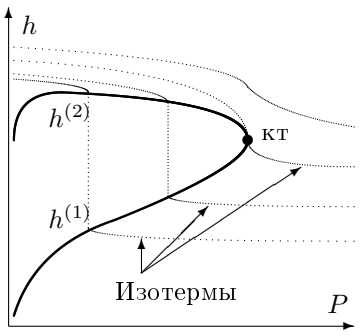


Рис. 9.2: h, P-диаграмма

Двухфазную область «не видно» на P,T-диаграмме, но можно «увидеть» на других координатных плоскостях. На рис. 9.2 представлено типичное сечение $s = const$ поверхности $h(s,P)$ с проекциями линий изотерм $(T = const)$, проходящими по этой поверхности. Крайняя правая точка сечения является критической точкой. Нижняя часть сечения является пограничной кривой $h^{(1)}$ со стороны газа, верхняя часть сечения является пограничной кривой $h^{(2)}$ со стороны жидкости. Между ними расположена двухфазная область.

При фазовом переходе термодинамические функции часто претерпевают излом, скачок или разрыв второго рода.

Например, удельные энтропия s и объём v при фазовом переходе испытывают разрыв первого рода или скачок. Скачок энтропии и удельного объёма связан с тем, что изменилось физическое состояние вещества, например, жидкость испарилась, а газ при одном и том же давлении и температуре занимает гораздо больший объём, чем жидкость.

У сосуществующих фаз, помимо равенства температур и давлений, всегда одинаковы значения химического потенциала. Это обусловлено возможностью перераспределения частиц, приводящего к выравниванию химического потенциала. Поэтому, химический потенциал φ при фазовом переходе также не претерпевает разрыва

$$\varphi^{(1)} = \varphi^{(2)}. \tag{9.133}$$

Однако при пересечения кривой фазового перехода химический потенциал φ будет иметь *излом*. Это вытекает из следующих несложных рассуждений: s и v являются первыми частными производными от химического потенциала, раз они испытывают скачок, значит φ имеет излом.

Действительно, проанализируем изменение химического потенциала вдоль линии $\varphi = const$ при пересечении этой линии с кривой фазового перехода (например, в точке τ на рис. 3.7). Иными словами, следует вычислить частные производные $\left(\frac{\partial P}{\partial T}\right)_{\varphi}$ с двух сторон от кривой фазового перехода.

Запишем

$$\left(\frac{\partial P}{\partial T}\right)_{\varphi} \overset{(3.161)}{=} -\left(\frac{\partial P}{\partial \varphi}\right)_{T}\left(\frac{\partial \varphi}{\partial T}\right)_{P} \overset{(9.36)}{=} \frac{s}{v}.$$

В точке пересечения линии $\varphi = const$ с кривой фазового перехода в P,T-диаграмме с одной стороны от кривой перехода (в области фазы 1) $\left(\frac{\partial P}{\partial T}\right)_{\varphi} = \frac{s^{(1)}}{v^{(1)}}$, а с другой стороны от кривой (в области фазы 2) $\left(\frac{\partial P}{\partial T}\right)_{\varphi} = \frac{s^{(2)}}{v^{(2)}}$. В общем случае $\frac{s^{(1)}}{v^{(1)}} \neq \frac{s^{(2)}}{v^{(2)}}$, поэтому излом линии $\varphi = const$ очевиден.

Величина скачка производной $\left(\frac{\partial f}{\partial x}\right)_y$ на пограничной кривой описывается уравнением (3.139):

$$\Delta\left(\frac{\partial f}{\partial x}\right)_y = -\Delta\left(\frac{\partial f}{\partial y}\right)_x \frac{dy}{dx}, \tag{9.134}$$

где $\Delta\left(\frac{\partial f}{\partial x}\right)_y = \left(\frac{\partial f}{\partial x}\right)_y^{(1)} - \left(\frac{\partial f}{\partial x}\right)_y^{(2)}$.

Подставим $f(x,y) = \varphi(T,P)$ в уравнение (9.134)

$$\Delta\left(\frac{\partial \varphi}{\partial T}\right)_P \overset{(9.36)}{=} -s^{(1)} - s^{(2)} = -\left(v^{(1)} - v^{(2)}\right)\frac{dP}{dT}$$

или

$$\frac{dP}{dT} = \frac{s^{(2)} - s^{(1)}}{v^{(2)} - v^{(1)}}. \tag{9.135}$$

Полученное *уравнение Клапейрона–Клаузиуса* (9.135) связывает величины скачков энтропии и удельного объёма в точке перехода с величиной наклона $\frac{dP}{dT}$ кривой фазового перехода на P,T-диаграмме.

Величина изменения (скачка) энтальпии при фазовом переходе называется *теплотой парообразования* и обычно обозначается

$$r = h^{(2)} - h^{(1)}. \tag{9.136}$$

Так как при фазовом переходе $T = const$ и $P = const$, то из (9.35) $Tds = dh - v\,dP$ следует, что

$$T(s^{(2)} - s^{(1)}) = h^{(2)} - h^{(1)} = r. \tag{9.137}$$

Тогда уравнение Клапейрона–Клаузиуса для h имеет вид

$$\frac{dP}{dT} = \frac{r}{T(v^{(2)} - v^{(1)})} \quad \text{или} \quad v^{(2)} - v^{(1)} = \frac{r}{T}\frac{dT}{dP}. \tag{9.138}$$

Получим также уравнения для u и f:

$$\begin{aligned} u^{(2)} - u^{(1)} &\overset{(9.32)}{=} h^{(2)} - h^{(1)} - P(v^{(2)} - v^{(1)}) = \\ &\overset{(9.136)}{=} r - P(v^{(2)} - v^{(1)}) \overset{(9.138)}{=} r\left(1 - \frac{P}{T}\frac{dT}{dP}\right), \\ f^{(2)} - f^{(1)} &\overset{(9.32)}{=} u^{(2)} - u^{(1)} - T(s^{(2)} - s^{(1)}) = \\ &\overset{(9.137)}{=} r\left(1 - \frac{P}{T}\frac{dT}{dP}\right) - r \overset{(9.138)}{=} -r\frac{P}{T}\frac{dT}{dP}. \end{aligned}$$

Как видно из приведённых выше соотношений, значения при скачке v, s, u и h $^{(2)}>^{(1)}$ — больше, а значение $f^{(2)} < f^{(1)}$ — меньше.

Этот результат имеет ясный физический смысл: фазовый переход связан с затратой работы, а работа системы в изотермическом процессе равна убыли её свободной энергии.

Если рассматривается фазовый переход в термодинамической системе, совершающей не работу расширения, а работу какого-либо другого вида, то обобщённой силой будет не давление P, а сила ξ. Соответственно, обобщённой координатой будет не удельный объём v, а координата w. Тогда можно легко получить уравнение, аналогичное по смыслу уравнению Клапейрона–Клаузиуса:

$$\frac{d\xi}{dT} = \frac{s^{(2)} - s^{(1)}}{w^{(2)} - w^{(1)}}, \quad \frac{d\xi}{dT} = \frac{r}{T(w^{(2)} - w^{(1)})}, \tag{9.139}$$

где r — теплота фазового перехода.

Рассмотрим, например, фазовый переход сверхпроводника из сверхпроводящего в нормальное состояние в магнитном поле (9.104). Тогда уравнение (9.139) будет иметь вид

$$\frac{dH_{\text{кр}}}{dT} = \frac{q}{T(\mathcal{M}_{\text{н}} - \mathcal{M}_{\text{с}})}. \tag{9.140}$$

где $\frac{dH_{\text{кр}}}{dT}$ — полная производная, взятая вдоль кривой фазового перехода на H, T-диаграмме ($H_{\text{кр}}$ — напряжённость критического магнитного поля данного сверхпроводника);

q — теплота фазового перехода сверхпроводника из сверхпроводящего в нормальное состояние;

$\mathcal{M}_{\text{н}}$ и $\mathcal{M}_{\text{с}}$ — удельные намагниченности соответственно нормальной и сверхпроводящей фаз.

Удельная магнитная восприимчивость сверхпроводника в нормальном состоянии пренебрежимо мала, а в сверхпроводящем состоянии выражается формулой

$$\chi_{\text{с}} = -\frac{v_{\text{с}}}{4\pi}, \quad \chi_{\text{н}} = 0,$$

где $v_{\text{с}}$ — удельный объём сверхпроводника в сверхпроводящем состоянии.

Тогда, с учётом (8.22) и $q > 0$ получаем, что напряжённость критического магнитного поля сверхпроводника возрастает с уменьшением температуры

$$\frac{dH_{\text{кр}}}{dT} = -\frac{4\pi q}{T v_{\text{с}} H_{\text{кр}}} < 0. \tag{9.141}$$

Это уравнение, впервые полученное В. Кеезомом (1924) для фазового перехода в сверхпроводнике, аналогично уравнению Клапейрона–Клаузиуса для обычных систем.

9.8 Кривая фазового перехода

Величины s, v, u, f и h аддитивные, и в двухфазной области равны суперпозиции величин однофазных областей. Их значения зависят от соотношения фаз или *степени сухости* x *двухфазной смеси*

$$\begin{aligned}
s^{\text{дф}} &= (1-x)s^{(1)} + xs^{(2)}, \\
v^{\text{дф}} &= (1-x)v^{(1)} + xv^{(2)}, \\
u^{\text{дф}} &= (1-x)u^{(1)} + xu^{(2)}, \\
f^{\text{дф}} &= (1-x)f^{(1)} + xf^{(2)}, \\
h^{\text{дф}} &= (1-x)h^{(1)} + xh^{(2)}.
\end{aligned}$$

Таким образом, величины s, v, u, f и h испытывают излом при пересечении левой и правой пограничных кривых и скачок в целом при фазовом переходе между однофазными областями.

Температура $T^{\text{дф}}$ и давление $P^{\text{дф}}$ одинаковы во время всего фазового перехода (9.132), в то время величины $s^{\text{дф}}, v^{\text{дф}}, u^{\text{дф}}, f^{\text{дф}}$ и $h^{\text{дф}}$ меняются при переходе через двухфазную область. Тогда

$$\left(\frac{\partial P}{\partial y}\right)_T^{\text{дф}} = 0, \qquad \text{где } y = s, v, u, f, h. \tag{9.142}$$

$$\left(\frac{\partial T}{\partial y}\right)_P^{\text{дф}} = 0, \qquad \text{где } y = s, v, u, f, h. \tag{9.143}$$

Уравнения (9.142, 9.143) означают, что в двухфазной области изотерма ($T^{\text{дф}} = const$) и изобара ($P^{\text{дф}} = const$) *имеют горизонтальные касательные* как на поверхности состояния, так и в проекциях на соответствующие P, y- и T, y-диаграммы.

При фазовом переходе оказывается удобным рассматривать термодинамическую функцию двух переменных $f(x, y)$ при фиксированном значении одной из переменных (например, при $y = const$). Поэтому частным, но важным является вопрос о том, как идут внутри двухфазной области *изолинии*, такие как $s, v, u, f, h = const$. Иными словами, следует вычислить частные производные $\left(\frac{\partial P}{\partial T}\right)_y^{\text{дф}}$ $(y = s, v, u, f, h)$ с двухфазной стороны пограничных кривых.

Эта задача решается следующим образом. Полная производная от давления по температуре вдоль линии насыщения имеет вид

$$\frac{dP}{dT} \overset{(3.105)}{=} \left(\frac{\partial P}{\partial T}\right)_y^{\text{дф}} + \left(\frac{\partial P}{\partial y}\right)_T^{\text{дф}} \frac{dy^\sigma}{dT}. \tag{9.144}$$

Здесь $\left(\frac{\partial P}{\partial T}\right)_y^{\text{дф}}$ — частная производная в точке пересечения линии $y = const$ с пограничной кривой, взятая с двухфазной стороны этой кривой; $\left(\frac{\partial P}{\partial y}\right)_T^{\text{дф}}$ — частная производная в точке пересечения изотермы с пограничной кривой, взятая с двухфазной стороны кривой; $\frac{dy^\sigma}{dT}$ — полная производная вдоль кривой насыщения; $y = s, v, u, f, h$.

Так как $\frac{dy^\sigma}{dT}$ на обеих пограничных кривых нигде, кроме критической точки, не обращается в бесконечность, то

$$\frac{dP}{dT} = \left(\frac{\partial P}{\partial T}\right)_y^{\text{дф}} + \left(\frac{\partial P}{\partial y}\right)_T^{\text{дф}} \frac{dy^\sigma}{dT} \overset{(9.142)}{=} \left(\frac{\partial P}{\partial T}\right)_y^{\text{дф}}.$$

В результате запишем

$$\left(\frac{\partial P}{\partial T}\right)_y^{\text{дф}} = \frac{dP}{dT}, \quad (y = s, v, u, f, h). \tag{9.145}$$

Таким образом, оказывается, что *внутри двухфазной области изолинии $s, v, u, f, h = const$ сливаются с кривой фазового перехода.*

Это обстоятельство не вызывает удивления, ибо двухфазная область проецируется на координатную плоскость P, T в виде кривой фазового перехода — кривой насыщения при испарении. С этой кривой сливаются любые линии, проходящие внутри двухфазной области в других координатных плоскостях.

Рассмотрим, как при фазовом переходе меняется *изохорная теплоёмкость* c_v. Запишем выражение

$$\Delta c_v = -T\Delta\left(\frac{\partial P}{\partial T}\right)_v \frac{dv}{dT}. \tag{9.146}$$

Доказательство

$$\frac{\Delta c_v}{T} \overset{(9.44)}{=} \Delta\left(\frac{\partial s}{\partial T}\right)_v \overset{(9.134)}{=} -\Delta\left(\frac{\partial s}{\partial v}\right)_T \frac{dv}{dT} \overset{(9.52)}{=} -\Delta\left(\frac{\partial P}{\partial T}\right)_v \frac{dv}{dT}.$$

При пересечении пограничных кривых c_v испытывает скачок

$$c_v^{\sigma\text{дф}} - c_v^{\sigma\text{оф}} \overset{(9.145)}{\underset{(9.146)}{=}} T\left[\left(\frac{\partial P}{\partial T}\right)_v^{\text{оф}} - \frac{dP}{dT}\right]\frac{dv^\sigma}{dT}. \tag{9.147}$$

В этом нет ничего удивительного, если функция s испытывает излом при пересечении пограничной кривой между однофазной и двухфазной областями, то её производная $\left(\frac{\partial s}{\partial T}\right)_v$ испытывает скачок.

С учётом $\frac{dP}{dT} = \left(\frac{\partial P}{\partial T}\right)_y^{\text{оф}} + \left(\frac{\partial P}{\partial y}\right)_T^{\text{оф}} \frac{dy^\sigma}{dT}$ запишем уравнение (9.147) для левой и правой пограничных кривых фазового перехода при испарении

$$c_v^{(1)\text{дф}} - c_v^{(1)\text{оф}} = -T \left(\frac{\partial P}{\partial v}\right)_T^{\text{оф}} \left(\frac{dv^{(1)}}{dT}\right)^2 \overset{(9.74)}{>} 0, \tag{9.148}$$

$$c_v^{(2)\text{дф}} - c_v^{(2)\text{оф}} = -T \left(\frac{\partial P}{\partial v}\right)_T^{\text{оф}} \left(\frac{dv^{(2)}}{dT}\right)^2 \overset{(9.74)}{>} 0. \tag{9.149}$$

Таким образом, при переходе из однофазной в двухфазную область изохорная теплоёмкость скачкообразно возрастает.

Аналогичным образом ведут себя при фазовом переходе некоторые термические коэффициенты.

Для *коэффициента адиабатной сжимаемости* $\beta_s = -\frac{1}{v}\left(\frac{\partial v}{\partial P}\right)_s$ справедливо выражение

$$v\Delta\beta_s = \Delta\left(\frac{\partial T}{\partial P}\right)_s \frac{ds}{dP}. \tag{9.150}$$

Доказательство

$$v\Delta\beta_s = -\Delta\left(\frac{\partial v}{\partial P}\right)_s \overset{(9.134)}{=} -\Delta\left(\frac{\partial v}{\partial s}\right)_P \frac{ds}{dP} \overset{(9.53)}{=} \Delta\left(\frac{\partial T}{\partial P}\right)_s \frac{ds}{dP}.$$

При пересечении пограничных кривых β_s испытывает скачок

$$\beta_s^{\sigma\text{дф}} - \beta_s^{\sigma\text{оф}} \overset{\substack{(9.145)\\(9.150)}}{=} \frac{1}{v^\sigma}\left[\frac{dT}{dP} - \left(\frac{\partial T}{\partial P}\right)_s^{\text{оф}}\right] \frac{ds^\sigma}{dP}. \tag{9.151}$$

С учётом $\frac{dT}{dP} = \left(\frac{\partial T}{\partial P}\right)_y^{\text{оф}} + \left(\frac{\partial T}{\partial y}\right)_P^{\text{оф}} \frac{dy^\sigma}{dP}$ запишем уравнение (9.151) для левой и правой пограничных кривых фазового перехода при испарении

$$\beta_s^{(1)\text{дф}} - \beta_s^{(1)\text{оф}} = \frac{1}{v^{(1)}} \left(\frac{\partial T}{\partial s}\right)_P^{\text{оф}} \left(\frac{ds^{(1)}}{dP}\right)^2 \overset{(9.75)}{>} 0, \tag{9.152}$$

$$\beta_s^{(2)\text{дф}} - \beta_s^{(2)\text{оф}} = \frac{1}{v^{(2)}} \left(\frac{\partial T}{\partial s}\right)_P^{\text{оф}} \left(\frac{ds^{(2)}}{dT}\right)^2 \overset{(9.75)}{>} 0. \tag{9.153}$$

Для *коэффициента Джоуля–Томсона* $\mu = \left(\frac{\partial T}{\partial P}\right)_h$, называемого также *коэффициентом адиабатного дросселирования*, справедливо выражение

$$\Delta\mu = \Delta\left(\frac{\partial T}{\partial P}\right)_h \overset{(9.134)}{=} -\Delta\left(\frac{\partial T}{\partial h}\right)_P \frac{dh}{dP}. \tag{9.154}$$

В силу определения коэффициента Джоуля–Томсона

$$\mu^{\text{дф}} = \left(\frac{\partial T}{\partial P}\right)_h^{\text{дф}} \overset{(9.145)}{=} \frac{dT}{dP}. \tag{9.155}$$

Следовательно, значения μ и на левой, и на правой пограничных кривых с их двухфазной стороны одинаковы:

$$\mu^{(1)\text{дф}} = \mu^{(2)\text{дф}} = \mu^{\text{дф}}. \tag{9.156}$$

Учитывая это обстоятельство, запишем уравнение для левой и правой пограничных кривых фазового перехода при испарении

$$\mu^{\text{дф}} - \mu^{(1)\text{оф}} = \frac{1}{c_P^{\text{оф}}} \frac{dh^{(1)}}{dP}, \tag{9.157}$$

$$\mu^{\text{дф}} - \mu^{(2)\text{оф}} = \frac{1}{c_P^{\text{оф}}} \frac{dh^{(2)}}{dP}. \tag{9.158}$$

Доказательство

$$\mu^{\sigma\text{дф}} - \mu^{\sigma\text{оф}} \overset{(9.154)}{=} -\left[\left(\frac{\partial T}{\partial h}\right)_P^{\text{дф}} - \left(\frac{\partial T}{\partial h}\right)_P^{\text{оф}}\right] \frac{dh^\sigma}{dP} \overset{\substack{(9.143)\\(9.45)}}{=} \frac{1}{c_P^{\sigma\text{оф}}} \frac{dh^\sigma}{dP}.$$

Рассмотрение h, P-диаграммы (рис. 9.2) показывает, что величина $\frac{dh^{(1)}}{dP}$ всегда положительна; следовательно, всегда

$$\mu^{\text{дф}} > \mu^{(1)\text{оф}}. \tag{9.159}$$

Величина $\frac{dh^{(2)}}{dP}$, как видно из h, P-диаграммы, в области давлений, далёких от критического, положительна, а с ростом давления меняет знак. Следовательно. при невысоких давлениях

$$\mu^{\text{дф}} > \mu^{(2)\text{оф}}, \tag{9.160}$$

а при повышенных давлениях

$$\mu^{\text{дф}} < \mu^{(2)\text{оф}}. \tag{9.161}$$

Это неравенство меняет знак в той точке кривой насыщения, где энтальпия сухого насыщенного пара $h^{(2)}$ достигает наибольшего значения.

Рассмотрим теперь, как при фазовом переходе меняется *изобарная теплоёмкость* c_P. Запишем выражение

$$\Delta c_P = T\Delta\left(\frac{\partial v}{\partial T}\right)_P \frac{dP}{dT}. \tag{9.162}$$

Доказательство

$$\frac{\Delta c_P}{T} \overset{(9.44)}{=} \Delta\left(\frac{\partial s}{\partial T}\right)_P \overset{(9.134)}{=} -\Delta\left(\frac{\partial s}{\partial P}\right)_T \frac{dP}{dT} \overset{(9.54)}{=} \Delta\left(\frac{\partial v}{\partial T}\right)_P \frac{dP}{dT}.$$

Согласно уравнению (9.162) изобарная теплоёмкость при подходе к пограничной кривой со стороны однофазной области всюду, за исключением критической точки имеет предел. Согласно (9.143) при пересечении пограничных кривых изобарная теплоёмкость в двухфазной области обращается в бесконечность, так как $\left(\frac{\partial v}{\partial T}\right)_P^{\text{дф}} = \infty$. Значит на пограничных кривых происходит разрыв второго рода.

Аналогичным образом ведут себя при фазовом переходе *температурный коэффициент объёмного расширения* $\alpha = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_P$

$$\Delta\alpha = -v^{-1}\Delta\left(\frac{\partial v}{\partial T}\right)_P \overset{(9.162)}{=} -(vT)^{-1}\Delta c_P \frac{dT}{dP} \tag{9.163}$$

и *коэффициент изотермической сжимаемости* $\beta_T = -\frac{1}{v}\left(\frac{\partial v}{\partial P}\right)_T$

$$\Delta\beta_T = -v^{-1}\Delta\left(\frac{\partial v}{\partial P}\right)_T \overset{(9.134)}{=} v^{-1}\Delta\left(\frac{\partial v}{\partial T}\right)_P \frac{dT}{dP}. \tag{9.164}$$

Таким образом, мы можем записать

$$c_P^{\text{дф}} = \left(\frac{\partial h}{\partial T}\right)_P^{\text{дф}} = \infty, \tag{9.165}$$

$$\alpha^{\text{дф}} = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_P^{\text{дф}} = \infty, \tag{9.166}$$

$$\beta_T^{\text{дф}} = -\frac{1}{v}\left(\frac{\partial v}{\partial P}\right)_T^{\text{дф}} = \infty. \tag{9.167}$$

Специфической величиной, относящейся только к линии фазового перехода, является *теплоёмкость вдоль пограничной кривой*, определяемая соотношением

$$c^\sigma = T\frac{ds^\sigma}{dT}, \tag{9.168}$$

где $\frac{ds^\sigma}{dT}$ — полная производная от энтропии по температуре вдоль пограничной кривой (одномерному пространству). Тогда с учётом (9.35) $TdS = dh - v\,dP$ находим

$$c^\sigma = \frac{dh^\sigma}{dT} - v^\sigma\frac{dP}{dT}. \tag{9.169}$$

В 1933 г. П. Эренфестом было введено представление о фазовых переходах второго рода.

Обычный фазовый переход (по классификации Эренфеста, фазовый переход первого рода), характеризуется тем, что в точке перехода первые производные от химического потенциала (9.36):

$$v = \left(\frac{\partial\varphi}{\partial P}\right)_T, \quad s = -\left(\frac{\partial\varphi}{\partial T}\right)_P.$$

претерпевают разрыв и изменяются скачком от значения в одной из сосуществующих фаз $v^{(1)}$ и $s^{(1)}$ до значения в другой фазе $v^{(2)}$ и $v^{(2)}$.

Фазовым переходом второго рода Эренфест назвал такой переход, при котором первые производные от химического потенциала в точке перехода непрерывны $v^{(1)} = v^{(2)}$ и $s^{(1)} = s^{(2)}$, а скачком изменяются лишь вторые производные от потенциала: $\left(\frac{\partial^2\varphi}{\partial P^2}\right)_T$, $\left(\frac{\partial^2\varphi}{\partial T^2}\right)_P$ и другие.

С формальной математической точки зрения фазовый переход второго рода эквивалентен переходу из однофазной области в двухфазную через пограничную кривую (левую или правую) при обычном фазовом переходе.

На P,T-диаграмме в случае фазового перехода второго рода, как и в случае обычного фазового перехода, области существования двух фаз будут разделяться кривой фазового перехода. Что же касается P,v-, v,T-, P,s- и s,T-диаграмм, то в случае фазового перехода второго рода области существования двух фаз будут разделяться не двухфазной областью, как при обычном фазовом переходе, а кривой фазового перехода.

Рассмотрение физической сущности фазовых переходов второго рода выходит за пределы книги. Нас интересует лишь величина скачка термодинамических величин при фазовом переходе второго рода

$$\left(\frac{\partial^2\varphi}{\partial T^2}\right)_P = -\left(\frac{\partial s}{\partial T}\right)_P = -\frac{c_P}{T}, \qquad \left[\frac{\partial}{\partial T}\left(\frac{\partial\varphi}{\partial T}\right)_P\right]_v = -\left(\frac{\partial s}{\partial T}\right)_v = -\frac{c_v}{T},$$
$$\left(\frac{\partial^2\varphi}{\partial P^2}\right)_T = \left(\frac{\partial v}{\partial P}\right)_T, \qquad \left[\frac{\partial}{\partial P}\left(\frac{\partial\varphi}{\partial P}\right)_T\right]_s = \left(\frac{\partial v}{\partial P}\right)_s;$$

а также

$$\left[\frac{\partial}{\partial T}\left(\frac{\partial\varphi}{\partial P}\right)_T\right]_P = \left(\frac{\partial v}{\partial T}\right)_P = -\left(\frac{\partial s}{\partial P}\right)_T,$$
$$\left[\frac{\partial}{\partial T}\left(\frac{\partial\varphi}{\partial P}\right)_T\right]_s = \left(\frac{\partial v}{\partial T}\right)_s = -\left(\frac{\partial s}{\partial P}\right)_v,$$

и так далее.

Уравнения, определяющие величину скачков всех этих величин легко получить при помощи уравнения (9.134), например

$$\Delta\left(\frac{\partial v}{\partial T}\right)_P = -\Delta\left(\frac{\partial v}{\partial P}\right)_T \frac{dP}{dT}. \tag{9.170}$$

Мы уже получали такие величины (см. 9.146, 9.150, 9.162).

При помощи уравнений (9.162, 9.170) можно получить аналог уравнения Клапейрона–Клаузиуса для обычных фазовых переходов — *уравнение Эренфеста*, вычисляющее наклон $\frac{dP}{dT}$ фазовой кривой второго рода на P,T-диаграмме

$$\frac{dP}{dT} = \frac{\Delta c_P}{T\Delta\left(\frac{\partial v}{\partial T}\right)_P} = -\frac{\Delta\left(\frac{\partial v}{\partial T}\right)_P}{T\Delta\left(\frac{\partial v}{\partial P}\right)_T}. \tag{9.171}$$

Рассмотрим теперь термодинамические величины в *критической точке*, где исчезает разница между жидкой и газовой фазой.

В соответствии с определением критической точки

$$v^{\text{кр}} = v^{(1)} = v^{(2)}, \quad s^{\text{кр}} = s^{(1)} = s^{(2)}. \tag{9.172}$$

В смысле непрерывности первых производных химического потенциала критическая точка напоминает фазовый переход второго рода, хотя и отличается от него по физической сущности.

Так как критическая точка является конечной точкой кривой фазового перехода, для неё справедливы выражения, полученные для двухфазной области, что подтверждается экспериментальными данными.

Во-первых, в критической точке *изолинии* гладко, без излома *сливаются с линией насыщения*, аналогично (9.145)

$$\left(\frac{dP}{dT}\right)^{\text{кр}} = \left(\frac{\partial P}{\partial T}\right)^{\text{кр}}_{y} \tag{9.173}$$

Для критической изохоры ($v^{\text{кр}} = const$) этот результат носит название *уравнение* (или *правило*) *Планка–Гиббса*.

Во-вторых, критическая изотерма ($T^{\text{кр}} = const$) и критическая изобара ($P^{\text{кр}} = const$) имеют в критической точке *горизонтальные касательные*, как на поверхности состояния, так и в проекциях на соответствующие P, y- и T, y-диаграммы, аналогично (9.142, 9.143):

$$\left(\frac{\partial P}{\partial y}\right)^{\text{кр}}_{T} = 0, \qquad \text{где } y = s, v, u, f, h. \tag{9.174}$$

$$\left(\frac{\partial T}{\partial y}\right)^{\text{кр}}_{P} = 0, \qquad \text{где } y = s, v, u, f, h. \tag{9.175}$$

При записи этих выражений в «перевёрнутом» виде мы получим *вертикальные касательные* изотерм и изобар.

Отсюда сразу следует, что, в-третьих, в критической точке изобарная теплоёмкость и некоторые термические коэффициенты становятся бесконечно большими, аналогично (9.165–9.167)

$$c_P^{\text{кр}} = \left(\frac{\partial h}{\partial T}\right)^{\text{кр}}_{P} = \infty, \tag{9.176}$$

$$\alpha^{\text{кр}} = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)^{\text{кр}}_{P} = \infty, \tag{9.177}$$

$$\beta_T^{\text{кр}} = -\frac{1}{v}\left(\frac{\partial v}{\partial P}\right)^{\text{кр}}_{T} = \infty. \tag{9.178}$$

Запишем

$$r \overset{(9.138)}{=} T(v^{(2)} - v^{(1)})\frac{dP}{dT}. \tag{9.179}$$

Величина $\frac{dP}{dT}$ не может быть бесконечно большой, поэтому в критической точке $v^{(1)} = v^{(2)}$, и теплота парообразования обращается в нуль

$$r^{\text{кр}} = 0. \tag{9.180}$$

Так как непрерывностью величин s, v, u, f, h, и f критическая точка напоминает фазовый переход второго рода, а первыми производными этих величин по P и T — двухфазную область, то имеет смысл исследовать вторые производные этих термодинамических величин.

Принципиальное формальное отличие критической точки от всех других точек фазовой диаграммы заключается в том, что *критические изотерма и изобара имеют касательную второго порядка*, т.е. не только горизонтальную (вертикальную) касательную, но и перегиб в критической точке, что подтверждается экспериментально (см. рис. 9.2)

$$\left(\frac{\partial^2 P}{\partial y^2}\right)_T^{\text{кр}} = 0, \qquad \text{где } y = s, v, u, f, h. \tag{9.181}$$

$$\left(\frac{\partial^2 T}{\partial y^2}\right)_P^{\text{кр}} = 0, \qquad \text{где } y = s, v, u, f, h. \tag{9.182}$$

Второе принципиальное формальное отличие критической точки от двухфазной области заключается в том, что полные производные вдоль кривой насыщения $\frac{dy^\sigma}{dT}$ (где $y = s, v, u, f, h$) в критической точке становятся бесконечно большими. При этом

$$\lim_{T^\sigma \to T^{\text{кр}}} \frac{dy^{(1)}}{dT} = \infty, \quad \lim_{T^\sigma \to T^{\text{кр}}} \frac{dy^{(2)}}{dT} = -\infty, \quad \text{где } y = s, v, u, f, h. \tag{9.183}$$

С формальной математической точки зрения это объясняется тем, что угол наклона кривой насыщения в критической точке теряет смысл, так как кривая вырождается в точку. Удивительно то, что эти формальные математические выводы, которые очевидно следуют из обычных геометрических соображений, имеют экспериментальное подтверждение.

Таким образом, с формальной математической точки зрения критическая точка — это точка перегиба изолиний, где кривая насыщения вырождается в точку. Других экспериментально установленных особенностей поведения термодинамических величин в критической точке, кроме тех, что следуют из формальных математических выкладок, к настоящему времени неизвестно.

Глава 10

Экономическая система

10.1 Экономические термины

Экономика (от греч. *οίκοςνομος* — букв. искусство ведения домашнего очага) — наука, изучающая законы взаимодействия ценных объектов (объектов, имеющих цену) в поле экономических отношений.

Деньги — общепризнанный универсальный эквивалент при обмене объектами в экономических отношениях.

Цена (англ. price) — коэффициент обмена объекта на деньги.

Термин «объект» в экономике не принят, вместо него применяется множество других терминов, например, товар и благо. Разница между этими терминами заключается, прежде всего, в представлениях о том, как образуется цена объекта.

Труд (англ. labor) — целесообразная деятельность человека. Другими словами, деятельность человека, преследующего свои цели.

Товар — в первую очередь продукт труда. Другими словами, объект, в котором овеществлён абстрактный труд, называемый также *стоимостью* товара, благодаря которой товар и имеет свою цену.

Полезность (utility) *объекта* — способность объекта удовлетворять какой-нибудь человеческой потребности (приносящий определенное удовлетворение потребителю).

Таким образом, полезный объект, это в первую очередь объект потребления. Блага имеют цену благодаря своей полезности. Потребители сравнивают полезность объектов при их обмене.

Благо — полезный (имеющий полезность) объект.

Понятия благо и товар часто смешивают и забывают от разнице между ними. Аналогично, смешивают понятия полезность и стоимость.

Так называемая классическая школа сформировалась в конце 18-го века. Достижением этой школы была *трудовая теория стоимости*, заложенная Адамом Смитом и Давидом Рикардо, и развитая в дальнейшем Карлом Марксом в направлении, называемом обычно *марксизмом*.

В качестве противовеса теории трудовой стоимости Маркса в последней трети 19-го века возникла *теория предельной полезности* или предельных издержек, называемая также *маржинализмом* (от англ. marginal — предельный).

Согласно трудовой теории стоимости рабочая сила является товаром и, как всякий товар, имеет свою стоимость, равную стоимости средств, необходимых для поддержания жизни рабочего и членов его семьи (труду, заключённому в этих средствах).

Рабочая сила — способность к труду или совокупность физических и духовных способностей, которыми обладает человек.

Прибавочная стоимость — по марксизму разница между стоимостью, заключённой в труде рабочего, и стоимостью его рабочей силы.

Обратим внимание ещё раз. В отличие от рабочей силы труд рабочего не является товаром и не имеет стоимости, так как он сам и есть мера стоимости. Труд не может иметь своей особой стоимости, как тяжесть не может иметь особого веса, теплота — особой температуры, электричество — особой силы тока. Однако этот тезис экономисты часто отвергают и называют трудом то, что является рабочей силой.

Согласно теории полезности цена блага определяется не вложенным в него трудом, а его способностью удовлетворять потребности.

Предельный анализ (англ. marginal analysis) — применение дифференциального исчисления в экономической науке.

Предельная полезность (нем. Grenznutzen, англ. marginal utility, MU) — дополнительная полезность, получаемая от потребления дополнительной единицы блага (при неизменности потребления других благ).

Общая полезность — совокупность предельных полезностей благ.

Не все блага создавались трудом. В первую очередь это относится к различным природным ресурсам. Поэтому каждый товар является благом, но не каждое благо является товаром. Таким образом, существуют блага, которые не являются продуктами труда и поэтому не являются товарами, но они имеют цену, и их можно продать или купить.

Обратим внимание, что благо или товар — это не обязательно материальный объект (вещь).

Услуга — нематериальный полезный объект или целесообразная деятельность человека, удовлетворяющая какую-нибудь человеческую потребность (полезный труд).

Экономические переменные — совокупность количеств $q_1, \ldots, q_n$ и цен $p_1, \ldots, p_n$ благ, которые мы можем рассматривать в качестве координат экономической системы, полностью определяющих её состояние.

Экономический процесс — процесс взаимодействия благ.

Экономические законы — законы поля экономических отношений и законы взаимодействия благ.

Экономические отношения — объективно складывающиеся отношения между людьми по поводу взаимодействия благ.

Экономические субъекты — люди, вступающие в экономические отношения, а также устойчивые группы людей, объединённые экономическими

отношениями. Само собой разумеется, что устойчивые группы могут вступать в экономические отношения с другими экономическими субъектами (группами или отдельными людьми).

Блага не могут взаимодействовать друг с другом самостоятельно или, другими словами, не могут самостоятельно вступать в экономические отношения. Поэтому каждый экономический субъект обладает определённым набором благ и вступает в экономические отношения по поводу их взаимодействия. Другими словами, все блага принадлежат экономическим субъектам и вступают во взаимодействие посредством экономических отношений субъектов.

Экономическая система (англ. economic system) — совокупность всех экономических субъектов и экономических отношений.

Если теперь посмотреть шире, то отношения между людьми по любому поводу называются *социальными* (*общественными*) *отношениями*; совокупность всех общественных отношений называется *общественной формой* или *формой общения*; вступающие в общественные отношения люди и устойчивые группы людей, объединенные социальными отношениями, называются *социальными субъектами*; а совокупность всех социальных субъектов и социальных отношений называется *социальной системой* или *обществом*. Таким образом, мы пришли к очевидному выводу: экономика — это часть общества, а экономические отношения — часть общественных отношений.

Экономические субъекты делят на *покупателей* — обменивающих деньги на благо, и *продавцов* — обменивающих благо на деньги.

Рынок — совокупность покупателей и продавцов, взаимодействие которых приводит в итоге к возможности обмена.

Производство — экономический процесс преобразования одних благ в другие. По отношению к производству блага делят на две группы:

1. *Средства* или *факторы производства* (от лат. factor — делающий, производящий) — участвующие в производстве блага.

2. *Потребительские блага* — не участвующие в дальнейшем производстве блага, которые удовлетворяют потребности людей непосредственно.

Производственная функция (англ. production function) в микроэкономике — уравнение, связывающее количество используемых факторов производства с максимально возможным при этом выпуском продукции (объёмом произведённых товаров). При этом имеют ввиду связь между скоростью выпуска товара $\dot{q}_i$ и скоростью производственного потребления факторов производства $\dot{q}_1, \ldots, \dot{q}_k$:

$$\dot{q}_i = f_i(\dot{q}_1, \ldots, \dot{q}_k). \tag{10.1}$$

Обратим внимание ещё раз. Количество произведённого товара, это количество, произведённое за определенное время. Следует сказать, что в экономической литературе существует определенная путаница в этом вопросе. Дело в том, что очень часто говорят о выпуске как о количестве товара

и обозначают это количество как q_i, забывая упомянуть, что подразумевают количество товара, произведенное в единицу времени, т.е. скорость производства товара $\dot{q}_i$.

Потребление — использование благ в процессе удовлетворения экономических потребностей общества.

Распределение и *обмен* — экономические процессы, связывающие производство и потребление, а также покупателей и продавцов.

Таким образом, среди экономических процессов взаимодействия благ выделяют: производство, потребление, распределение (присвоение) и обмен (торговля) благ.

Потребление бывает производительным (производственным) и непроизводственным. В процессе производственного потребления потребляются факторы (средства) производства, называемые также производственными ресурсами. Непроизводственное потребление осуществляется за пределами процесса производства, поскольку потребляются конечные предметы потребления или потребительские блага.

Указанное многообразие терминов, когда производственные блага называют также ресурсами, средствами или факторами производства, является обычным делом для экономики.

По отношению к процессам производства и потребления экономические субъекты разделяют *потребителей* и *производителей* (функциональное деление).

Отрасль — совокупность экономических субъектов, производящих и торгующих одним и тем же или родственными видами благ.

Неоклассическая школа (основоположник — Альфред Маршалл), развивая маржинализм, постаралась увязать его идеи с классической школой, за что и получила своё название. Маршалл предположил, что полезность благ отвечает на рынке за спрос на эти блага, а стоимость товаров отвечает на рынке за их предложение.

Спрос D (англ. demand) — представленная на рынке потребность в благе, ограниченная действующими ценами и платежеспособностью потребителей.

Предложение S (англ. supply) — представленное к продаже на рынке количество товара, ограниченное действующими ценами и возможностями производства.

Цена спроса — это максимальная цена, по которой покупатели намерены и в состоянии приобрести определенное количество блага за определенный период (единицу) времени.

Цена предложения — это минимальная цена, по которой продавцы согласны продать определенное количество товара, произведённое за определенный период (единицу) времени.

Равновесная цена — цена на конкурентном рынке, при которой величина спроса и величина предложения равны.

Совокупный спрос AD — это общий объем благ, на который может быть предъявлен спрос (или который может быть куплен за единицу времени) при различных уровнях цен.

Совокупное предложение AS (англ. aggregate supply) — это общее количество товаров, которое может быть предложено (произведено за единицу времени) при разных уровнях цен.

Богатство W (англ. wealth) — накопленные экономическими субъектами блага. Из самого определения богатства ясно, что эта функция зависит от количества благ $q_1, \ldots, q_n$ в экономической системе, не зависит от скоростей их изменения $\dot{q}_1, \ldots, \dot{q}_n$.

Доходы (англ. income) экономического субъекта — текущее (измеренное за единицу времени) количество денег и благ, полученное субъектом в результате его экономических отношений с другими субъектами.

В отношении отдельного производителя, отрасли или совокупного производства экономической системы эту величину вычисляют как *общую выручку* TR (англ. total revenue) или произведение объёма произведённой продукции на её цену

$$TR = \sum_{i=1}^{n} p_i \dot{q}_i = \mathbf{p}\dot{\mathbf{q}}. \tag{10.2}$$

Издержки (англ. cost) экономического субъекта — текущее (измеренное за единицу времени) количество денег и благ, затраченное субъектом в результате его экономических отношений с другими субъектами.

В отношении отдельного производителя, отрасли или совокупного производства экономической системы эту величину вычисляют как *общие издержки* TC (англ. total cost) зависимость между объемом произведенной продукции и минимально необходимыми затратами (факторов производства) на её производство и реализацию

$$TC = \sum_{j=1}^{k} p_j \dot{q}_j = \mathbf{p}\dot{\mathbf{q}}. \tag{10.3}$$

Прибыль (англ. profit) π экономического субъекта — текущая (измеренная за единицу времени) разница между его доходами и издержками.

В отношении отдельного производителя, отрасли или совокупного производства экономической системы эту величину вычисляют как

$$\pi = TR - TC. \tag{10.4}$$

С точки зрения марксизма совокупная прибыль экономической системы равна прибавочной стоимости, которая распределяется между владельцами факторов. Однако экономисты отвергают эту простую и понятную трактовку путём манипулирования с величиной издержек.

Различают производственные (бухгалтерские) издержки, которые вычисляются сложением рыночной ценности всех израсходованных в процессе изготовления товара факторов, и экономические издержки, которые вычисляют, прибавляя к производственным издержкам *ожидаемый доход* от *инвестиций* (от нем. Investition или от лат. investio — одеваю) — долгосрочных вложений в производство. При этом ожидаемый доход не должен опускаться ниже уровня, установленного текущей нормой процента от вложений.

Другими словами, экономический подход к определению издержек основан на выявлении того, какой максимальный доход можно было бы получить при альтернативном применении использованных для изготовления товара факторов. В результате прибыль производителей при равновесных ценах оказывается равной нулю, а то, что распределено между владельцами факторов, называется ожидаемым доходом от инвестиций, который включён в экономические издержки как дополнение к производственным издержкам.

В экономике выделяют две обширные области:

Микроэкономика — изучает экономическую систему на уровне деятельности отдельных экономических субъектов.

Макроэкономика — изучает экономическую систему в целом.

Если рассмотреть совокупное производство экономической системы на макроэкономическом уровне, то всё разнообразие факторов производства можно свести в четыре агрегированных группы ресурсов, называемые *первичными факторами производства*:

1. *Земля* или природные (созданные природой) ресурсы.
2. *Капитал* или производственные (созданные трудом) ресурсы.
3. *Труд* (точнее, рабочая сила) или человеческие ресурсы.
4. *Предпринимательская способность* (разновидность человеческих ресурсов) — сознательная человеческая деятельность по приобретению (производственному потреблению), а в дальнейшем по координации и комбинированию остальных трёх факторов в процессе производства.

Первичные факторы можно назвать *макроэкономическими переменными* или *параметрами экономической системы*.

Первичные факторы также являются благами, как и факторы в микроэкономике. Но у них есть очень существенные особые свойства.

С одной стороны, в этом очевидном с первого взгляда делении содержится определённое противоречие, так как капитал есть «накопленный труд», и они в определённом смысле тождественны.

Предпринимательская способность также есть вид труда. Особые функции, которые приходится исполнять предпринимателю как таковому и которые выпадают на его долю как раз в отличие от рабочих и в противоположность рабочим, представляются как чисто трудовые функции. Он создает прибавочную стоимость не потому, что работает как предприниматель, а потому, что несмотря на свое качество предпринимателя, он тем не менее тоже работает.

С этой точки зрения «у нас остаются только две стороны: природная, объективная — земля, и человеческая, субъективная — труд, включающий в себя капитал» [36, с. 559].

С другой стороны, производственные и природные ресурсы в процессе производства не изменяют своей стоимости, в отличие от человеческих ресурсов, которые в процессе производства изменяют свою стоимость и производят ещё и некоторый избыток — прибыль (прибавочную стоимость). По-

этому землю и капитал можно назвать *постоянным капиталом* или *капиталом* K, а рабочую силу и предпринимательскую способность *переменным капиталом* или *трудом* L. «Переменный капитал является простым показателем рабочей силы, а постоянный капитал — простым показателем массы средств производства, приводимых в движение этой рабочей силой» [24, с. 151].

Производственная функция в макроэкономике — есть *уравнение состояния* или поверхность состояния экономической системы в пространстве параметров:

$$Y = Y(K, L) \quad \text{или} \quad f(Y, K, L) = const, \tag{10.5}$$

которое выражает технологическую взаимосвязь между *уровнем производства* (максимально возможным совокупным выпуском в единицу времени) Y, трудом L и капиталом K, из которых при желании можно выделить землю Z и предпринимательскую способность G, что увеличит количество параметров в уравнении состояния.

Уравнение состояния экономической системы определяет *способ производства* — способ присвоения обществом природы.

Это определение есть ещё у Маркса, но его часто забывают: «Всякое производство есть присвоение индивидуумом предметов природы в пределах определенной общественной формы и посредством нее» [25, с. 713]. Более того, Маркс считал что способ производства (совокупность производства и экономических отношений) определяет все остальные общественные отношения (форму общения): «Сами они (люди) начинают отличать себя от животных, как только начинают производить необходимые им средства к жизни ... Способ, каким люди производят необходимые им средства к жизни это ... их определенный образ жизни. Какова жизнедеятельность индивидов, таковы и они сами. То, что они собой представляют, совпадает, следовательно, с их производством.... Само оно (производство) опять-таки предполагает общение индивидов между собой. Форма этого общения, в свою очередь, обусловливается производством» [26, с. 19].

Переход от микроэкономического описания экономической системы к макроэкономическому аналогичен переходу от микроскопического описания явлений природы к феноменологическому (макроскопическому). Феноменологические теории описывают явление в наиболее общем виде, устанавливая основные закономерности явления. В этом — главное достоинство феноменологических теорий [14, стр. 7].

Домашнее хозяйство (*потребитель*) — экономическая единица, состоящая из одного или более лиц, снабжающая экономику первичными факторами производства, и использующая полученные за них доходы для удовлетворения своих потребностей, т.е. для приобретения потребительских благ. Термин «домохозяйства» применяется в макроэкономическом анализе (в анализе движения национального дохода), в то время как термин «потребители» используется в основном в микроэкономике (в анализе спроса и предложения).

Предприятие или *фирма* (*производитель*) — экономический субъект

(состоящий из группы лиц) с правами юридического лица, который осуществляет процесс производства товаров и потребляет при этом первичные факторы производства. Если термин «предприятие» используется в макроанализе, то в микроанализе применяется термин «фирма».

Используется еще одно (также функциональное) деление экономических субъектов на *отдельных физических лиц* и *организаторов бизнеса* (*бизнесменов*). Таким образом, из всей массы физических лиц иногда выделяют бизнесменов, которые проявляют предпринимательскую способность. Это деление весьма условно, так как практически любой человек может в той или иной степени проявлять предпринимательскую способность, хотя ее может и не хватить для организации отдельного предприятия или для руководства фирмой. Таким образом, при функциональном делении экономических субъектов используют три набора терминов: домохозяйства и предприятия, потребители и фирмы, физические лица и бизнесмены.

Домохозяйства (потребители) и физические лица выступают в роли покупателей потребительских благ и продавцов первичных факторов производства, а предприятия (фирмы) и бизнесмены выступают, соответственно, в роли покупателей первичных факторов производства и продавцов товаров.

Производственные ресурсы делят на *предметы труда* (сырьё), которые в процессе производства переносят свою стоимость на производимые блага полностью, и *средства* (*орудия*) *труда*, с помощью которых обрабатывают предметы труда и которые переносят свою стоимость на производимые блага по частям (изнашиваются в процессе производства).

Амортизация (от лат. amortisatio — погашение) — экономический процесс перенесения стоимости изношенных средств труда на произведенный с их помощью продукт. Скорость перенесения стоимости средств труда на произведённый продукт определяет период их использования.

Многократность использования характерна таже для природных и человеческих ресурсов. Поэтому средства труда, природные и человеческие ресурсы имеют две цены капитальную и прокатную.

Капитальная цена фактора — общая сумма денег, уплачиваемая за весь период его использования.

Прокатная цена фактора — сумма денег, уплачиваемая за его использование в течение определённого срока.

Обычно определяют прокатную цену не средств труда, а распространяют эту цену на все производственные ресурсы, которые необходимо приобретать для производства. Прокатная цена человеческих ресурсов называется — *заработная плата* p_L, природных ресурсов — *рента* p_Z, производственных ресурсов — *ссудный процент* p_K.

Добавленная стоимость — в микроэкономике та часть стоимости продукта, которая создаётся в данном процессе производства (добавляется к предметам труда). С точки зрения трудовой теории стоимости добавленная стоимость состоит из труда (по марксизму — стоимости рабочей силы), амортизации и прибыли (по марксизму — прибавочной стоимости), которые были перенесены на предмет труда в процессе производства.

Добавленная стоимость экономической системы (в данном случае — страны) вычисляется за определённый период, например за год, и имеет специальные названия: *валовой внутренний продукт*, ВВП (gross domestic product, GDP) и *валовой национальный доход*, ВНД.

ВВП — статья системы национальных счетов, которую вычисляют в рыночных ценах как совокупную стоимость конечного продукта (продукции, товаров и услуг), созданного в течение года (квартала, месяца) внутри страны с использованием факторов производства, принадлежащих как данной стране, так и другим странам.

ВНД — балансирующая статья системы национальных счетов, которую вычисляют, добавляя к ВВП сальдо первичных доходов (разницы между полученными доходами из-за границы и выплаченными за границу). К таким первичным доходам обычно относят оплату труда, доходы от собственности в виде дивидендов и т. п.

Чистый национальный доход, ЧНД — валовой национальный доход за вычетом амортизации, поскольку износ средств труда является элементом издержек производства. Другими словами, ЧНД — сумма конечных товаров и услуг, произведенных и приобретенных нацией за определенный период за вычетом той части инвестиций, которая пошла на замену устаревшего и износившегося оборудования.

С теоретической точки зрения ЧНД является более совершенным показателем, так как амортизация, по существу, лишь компенсирует убыль средств труда, созданных в предыдущие периоды, а не создает новую стоимость. Таким образом, вычет амортизации очищает ВНД от двойного счета, но расчёт амортизации затруднителен.

До 1993 года вместо ВНД рассчитывался *валовой национальный продукт*, ВНП (gross national product, GNP) — совокупная стоимость объёма конечных товаров и услуг, произведённых на территории данной страны и за её пределами с использованием факторов производства, принадлежащих данной стране; а вместо ЧНД — *чистый национальный продукт*, ЧНП (net national product, NNP).

Научно–технический прогресс, НТП (от лат. progressus — движение вперед) — качественное изменение экономической системы, которое определяется уровнем развития техники (технологиями) и общественных отношений (организацией производства, квалификацией кадров и т.д.).

10.2 Аналогии экономической системы

Обратим внимание ещё раз. В отличие от природных систем, где свойства объектов и векторное поле их взаимодействия могут существовать без участия субъекта, в экономической системе субъект обязательно участвует как в определении цены блага, так и в экономических отношениях.

Однако рассуждения о субъективной цене не могут объяснить механизм рыночного ценообразования, где несмотря на многообразие субъективных оценок существует единая цена на благо, которая возникает в ре-

зультате конкуренции экономических субъектов.

Аналогично, считается, что и экономические законы носят объективный характер, т.е. воля субъекта подчиняется объективным законам взаимодействия благ.

С объективностью рыночной цены и экономических законов экономисты соглашаются уже потому, что в противном случае исчез бы сам предмет исследований экономики — науки, которой они занимаются. Однако, указанное противоречие между объективностью экономических законов и субъективностью экономических отношений вызывает определённые проблемы. Субъективность экономических отношений часто побуждает экономистов мистифицировать эти законы.

В частности, существует миф, что для описания экономических законов не подходит имеющийся математический аппарат, созданный по сути не как некая абстракция, а для описания окружающей человека природы. Мудрецы, которые так решили, не предлагают нового математического аппарата, а просто отвергают то, что было сделано до них.

Автор настоящей книги исходит из предположения, что общество, как и природа, есть часть материального мира. Следствием этого очевидного предположения, будет вывод о том, что к общественным отношениниям применимы законы материального мира, которые должны быть аналогичны другим, уже известным законам природы.

Действительно, на протяжении всей книги автором подчёркивалась простая мысль: все известные науке законы выведены из геометрических свойств пространства, и никаких дополнительных условий, кроме этих свойств, для их вывода не требовалось. Если мы считаем, что экономическая система существует в пространстве материального мира (а не, например, потустороннего), то её законы также, как и остальные, должны вытекать из геометрических свойств этого пространства.

Более простое объяснение состоит в следующем. Ввиду большого количества экономических субъектов их субъективные действия усредняются и приобретают объективный характер, как усредняется неупорядоченное движение отдельных частиц в общем потоке. Конечно, экономические субъекты могут по тем или иным причинам в определённый момент действовать согласованно, но эти действия носят кратковременный характер и представляют собой аналог флуктуаций и неравновесных процессов в природе. Целенаправленное влияние на экономическую систему может оказывать только такой субъект как государство.

Если взять за основу рыночные экономические отношения, и не учитывать роль государства в управлении экономической системой, то можно выстроить определённый перечень аналогий между переменными и функциями механики и микроэкономики (таб. 10.1). Аналогии, как обычно, будут носить формальный математический характер без учёта специфических свойств соответствующих функций и переменных.

Микроэкономика	**Механика**
Количества благ $\mathbf{q}$	Обощённые координаты $\mathbf{q}$
Цены благ $\mathbf{p}$	Обобщённые импульсы $\mathbf{p}$
Производственные функции $\dot{\mathbf{q}}$	Обобщённые скорости $\dot{\mathbf{q}}$
Богатство $W(t, \mathbf{q})$	Действие $S(t, \mathbf{q})$
Общие издержки $TC(\mathbf{p}, \mathbf{q})$	Полная энергия $E(\mathbf{p}, \mathbf{q})$
Совокупное предложение $AS(\dot{\mathbf{q}}, \mathbf{q})$	Кинетическая энергия $T(\dot{\mathbf{q}}, \mathbf{q})$
Совокупный спрос $AD(\mathbf{q})$	Потенциальная энергия $U(\mathbf{q})$

Таблица 10.1: Аналогии микроэкономики и механики.

Запишем следующие уравнения

$$AS \overset{(7.1)}{=} \sum_{i,j=1}^{n} \alpha_{ij}(q_1, \ldots, q_n)\, \dot{q}_i \dot{q}_j = (\dot{\mathbf{q}}, \alpha\dot{\mathbf{q}}), \tag{10.6}$$

$$\mathbf{p}^S \overset{(5.47)}{\underset{(7.5)}{=}} \frac{\partial AS}{\partial \dot{\mathbf{q}}} \quad \text{— цены предложения,} \tag{10.7}$$

тогда

$$\mathbf{p}^S \overset{(10.7)}{\underset{(3.149)}{=}} 2\alpha\dot{\mathbf{q}} \quad \text{или} \quad p_i^S = 2\sum_{j=1}^{n} \alpha_{ij}(q_1, \ldots, q_n)\, \dot{q}_j. \tag{10.8}$$

Таким образом, цена предложения товара есть частная производная от функции совокупного предложения AS по соответствующей производственной функции (так как совокупное предложение товара на рынке должно зависеть от скорости его производства) и вычисляется как суперпозиция различных производственных функций.

Цены спроса благ есть частные производные от богатства системы по количеству соответствующего блага в системе

$$\mathbf{p}^D \overset{(5.63)}{=} \frac{\partial W}{\partial \mathbf{q}} \quad \text{или} \quad p_i = \frac{\partial W}{\partial q_i} \qquad \text{— цены спроса.} \tag{10.9}$$

В точке τ, где спрос на рассматриваемый товар будет равен предложению, цена данного товара будет равновесной

$$p_i(\tau) = p_i^S(\tau) = p_i^D(\tau). \tag{10.10}$$

Кроме того, запишем уравнения

$$\dot{W} \overset{(5.62)}{\underset{(7.4)}{=}} AS - AD \overset{(5.45)}{=} TR - TC \overset{(10.4)}{=} \pi, \tag{10.11}$$

$$TC \overset{(5.64)}{=} -\frac{\partial W}{\partial t}. \tag{10.12}$$

где $TR \overset{(10.2)}{=} \sum\limits_{i=1}^{n} p_i^S \dot{q}_i$ — функция общей выручки,

$TC \overset{(10.3)}{=} \sum\limits_{i=1}^{k} p_j^D \dot{q}_j$ — функция общих издержек.

Скорость накопления богатства системы $\dot{W}$, очевидно, равна разнице между совокупным предложением и совокупным спросом. С другой стороны накопление богатства, очевидно, есть разница между общей выручкой TR (переносной скоростью изменения богатства) и общими издержками TC (локальной скоростью изменения богатства).

Дадим теперь интерпретацию принципа наименьшего действия для экономической системы. В принципе наименьшего действия постулируется, что функционал W (в нашем случае богатство системы) должен принимать свое минимальное значение. На языке экономики это означает, что накопление богатства системы не является оптимальным. Производимые в экономической системе блага должны потребляться и, тем самым, участвовать в процессе воспроизводства. При этом переход экономической системы из одного заданного состояния в другое будет происходить по оптимальному пути, т.е. с наименьшей затратой средств.

Когда экономическая система движется по оптимальному пути, то скорость изменения цены спроса для каждого блага равно скорости изменения цены его предложения

$$\dot{\mathbf{p}}^S \overset{(5.47)}{\underset{(7.3)}{=}} \dot{\mathbf{p}}^D. \tag{10.13}$$

Оптимальное движение экономической системы означает *экономическое равновесие*, когда спрос приблизительно равен предложению, все произведенные блага потребляются, богатство не накапливается, и справедливы следующие уравнения

$$AS \approx AD, \quad TR \approx TC, \quad W \approx const. \tag{10.14}$$

Тогда имеют место два случая

$$\delta W = 0 \quad \text{— для оптимального (равновесного) состояния,} \tag{10.15}$$

$$\delta W > 0 \quad \text{— для неоптимального (неравновесного) состояния.} \tag{10.16}$$

Богатство может как накапливаться, так и расходоваться, но процесс уменьшения богатства носит частный (индивидуальный) характер, наблюдаемый для отдельных экономических субъектов, а для экономической системы в целом на всем наблюдаемом историческом периоде характерно непрерывное увеличение богатства. Указанное утверждение легко продемонстрировать на примере количества потребляемой человечеством энергии [2], которое определяет объем мирового промышленного производства. На протяжении 140 лет мировое производство энергии выросло в 19 раз, с 0,68 ТВт (10^{12} Вт) в 1850 году до 13,2 ТВт в 1990 году, т.е. средний темп роста составил около 2% в год.

Рассмотрим теперь аналогию при каноническом преобразовании переменных $\mathbf{p}, \mathbf{q}$ к переменным действие–угол $\mathbf{I}, \varphi$ (см. § 5.7).

Величина

$$W_i \overset{(5.80)}{=} \oint\limits_{C_i} \mathbf{p}\, d\mathbf{q} \tag{10.17}$$

есть богатство, образованное за один цикл C_i производства i-ого товара.

Пусть T_i период одного производственного цикла для i-го товара, тогда *доходность* ν_i соответствующего производства равна скорости оборота капитала (частоте с единицей измерения, например, год$^{-1}$):

$$\nu_i \overset{(5.79)}{=} \frac{1}{T_i}. \tag{10.18}$$

В этом случае функция общей выручки будет иметь вид

$$TR = \sum_{i=1}^{n} \nu_i W_i. \tag{10.19}$$

Доказательство

$$TR \overset{(10.2)}{=} \mathbf{p}\dot{\mathbf{q}} = \mathbf{I}\dot{\varphi} \overset{\substack{(5.80)\\(5.82)}}{=} \sum_{i=1}^{n} \frac{\omega_i}{2\pi} \oint\limits_{C_i} \mathbf{p}\, d\mathbf{q} \overset{(5.79)}{=} \sum_{i=1}^{n} \nu_i W_i.$$

Пример. Если рассмотреть в качестве капитала выданные банком кредитные ресурсы W, то соответствующая частота даст нам банковскую ставку r (ссудный процент). Цена денег p равна денежной единице, а скорость $\dot{q}$ сумме денежных единиц, уплаченных банку в виде процентов по кредиту за единицу времени, например, за год

$$p\,\dot{q} = r \oint p\, dq = rW. \tag{10.20}$$

Интерпретируем переход от микроэкономики к макроэкономике аналогично переходу от механики к термодинамике (см. таб. 9.1). В этом случае, во-первых, можно выстроить определённый перечень формальных математических аналогий между параметрами и функциями термодинамики и макроэкономики (таб. 10.2), во-вторых, можно выстроить определённый перечень аналогий с соответствующими законами термодинамики (см. § 9.4), например, уравнение состояния экономической системы (10.5) аналогично уравнению состояния термодинамической системы (9.21), в-третьих, экономическая система должна рассматриваться во взаимодействии с другими экономическими системами (в определённой внешней среде).

Закон 1 (Аналог нулевого начала термодинамики). При неизменных условиях взаимодействия с другими системами уровень производства Y в различных отраслях экономической системы выравнивается.

Закон 2 (Аналог первого закона термодинамики). Чистый национальный доход (ЧНД) по определению в каждый момент движения экономической системы должен возрастать на количество создаваемой в системе

Макроэкономика	**Термодинамика**
Капитал K	Объём V
Труд L	Энтропия S
Уровень производства Y	Температура T
Технический прогресс P	Давление P
Чистый национальный доход $N(K,L)$	Внутренняя энергия $U(S,V)$
Добавленная стоимость Q	Теплота Q
Амортизация A	Работа A

Таблица 10.2: Аналогии макроэкономики и термодинамики.

добавленной стоимости и уменьшаться на количество амортизационных отчислений, направленных на восстановление изношенного капитала, что отражено в уравнении

$$dN \overset{(9.47)}{=} dQ - dA. \tag{10.21}$$

Изменение ЧНД ΔN определяется разностью $\Delta Q - \Delta A$ при любом процессе, при котором система переходит из исходного состояние в рассматриваемое. Это означает, что интеграл $\int N$ не зависит от пути интегрирования, дифференциал dN является полным, а ЧНД N есть одна из функций состояния экономической системы.

Закон 3 (Аналог второго закона термодинамики). Уровень производства Y есть максимально возможный совокупный выпуск продукции в единицу времени, при соответствующих затратах первичных факторов производства. Если этот уровень достигается, то добавленная стоимость связана с затраченным трудом следующим соотношением

$$dQ = Y dL, \tag{10.22}$$

которое отражает оптимальный (равновесный) процесс производства в экономической системе. Оптимальный процесс производства является обратимым, т.е. экономическая система способна без потерь пройти обратно к исходному состоянию через все промежуточные состояния.

Однако эффективный уровень производства достигается не всегда, и в этом случае процесс производства не является оптимальным (равновесным). При этом труд не обязательно используется рационально, а добавленная стоимость не полностью соответствует тем затратам труда, которые были произведены

$$dQ < Y dL. \tag{10.23}$$

На математическом языке выражение (10.23) означает, что дифференциал добавленной стоимости dQ не является полным.

Амортизационные отчисления в каждый момент движения экономической системы могут быть записаны в виде уравнения

$$dA \overset{(9.14)}{=} P dK, \tag{10.24}$$

где дифференциал dA также не является полным.

Общее выражение для оптимальных и неоптимальных процессов производства будет иметь вид

$$dN \leqslant YdL - PdK. \tag{10.25}$$

Размер экономической системы определяется количеством находящихся в ней ресурсов, а именно, количеством трудовых ресурсов L и производственных ресурсов K, которые в данном случае будут экономическими кординатами.

Уровень производства Y и технический прогресс P — экономические силы, взаимные с экономическими координатами L и K.

Чистый национальный доход N является функцией состояния и однозначно определяется размерами трудовых и производственных ресурсов. Экономические параметры K, L, Y и P также являются функциями состояния. Величина валового внутреннего продукта (добавленной стоимости) Q и величина амортизационных отчислений A являются функциями процесса и зависят от избранного пути экономической системы.

Труд, являясь самовоспроизводимым ресурсом при нормальных условиях (если исключить войны, эпидемии и т.д.) может только возрастать, что вполне согласуется с его энтропийной аналогией.

Термин «перегрев экономики», когда темпы экономического роста (увеличения N) слишком велики и поглощают все ресурсы, при указанных аналогиях имеет не образный а вполне конкретный смысл.

Действительно, $dN \sim Y$, а Y аналогичен температуре экономической системы, которая в этом случае слишком велика.

Для оптимального процесса вместо неравенства (10.25) мы можем записать дифференциал ЧНД, выражения для уровня производства Y и научно–технического прогресса P:

$$dN = YdL - PdK, \quad Y = \left(\frac{\partial N}{\partial L}\right)_K, \quad P = -\left(\frac{\partial N}{\partial K}\right)_L, \tag{10.26}$$

а также аналог уравнений Максвелла (9.51–9.54):

$$MP_K = \left(\frac{\partial Y}{\partial K}\right)_L = -\left(\frac{\partial P}{\partial L}\right)_K, \tag{10.27}$$

где MP_K — *предельная производительность капитала*.

Величина

$$C_Z = \left(\frac{\partial Q}{\partial Y}\right)_Z = Y\left(\frac{\partial L}{\partial Y}\right)_Z \tag{10.28}$$

аналогична общей теплоёмкости (9.43), а величина

$$MP_L = \left(\frac{\partial Y}{\partial L}\right)_K \overset{(10.28)}{=} \frac{Y}{C_K} \tag{10.29}$$

есть *предельная производительность труда* (англ. marginal product).

Величина, аналогичная (9.37) имеет вид

$$\left(\frac{\partial K}{\partial L}\right)_N \overset{(3.161)}{=} -\left(\frac{\partial N}{\partial L}\right)_K \left(\frac{\partial K}{\partial N}\right)_L \overset{(10.26)}{=} \frac{Y}{P}. \tag{10.30}$$

Если из капитала K выделить природные ресурсы Z, а вместо НТП или функции P ввести две функции P_K и P_Z, то мы получим *сложную* экономическую систему, аналогичную системе (9.96) термодинамики

$$dN \leqslant Y dL - P_K\, dK - P_Z\, dZ. \tag{10.31}$$

Заинтересованный читатель может самостоятельно выписать уравнения, аналогичные уравнениям термодинамики для сложных систем.

Закон 4 (Аналог теоремы Нернста–Планка). Если в экономической системе ничего не производится ($Y = 0$), то в такой системе люди воспроизводить свою жизнь не могут ($L = 0$).

Закон 5 (Аналог принципа Ле-Шателье–Брауна). Пусть научно–технический прогресс P нарушает равновесие системы со средой, а изменение уровня производства Y нарушает внутреннее равновесие экономической системы. Тогда в соответствии с § 9.5 внешнее воздействие, выводящее экономическую систему из равновесия, стимулирует в ней процессы, стремящиеся ослабить результаты этого воздействия.

По аналогии с (9.74) при $\left(\frac{\partial Y}{\partial L}\right)_K > 0$ получаем

$$\left(\frac{\partial K}{\partial P}\right)_L < \left(\frac{\partial K}{\partial P}\right)_Y < 0 \quad \text{или} \quad |(\Delta K)_Y| < |(\Delta K)_L|. \tag{10.32}$$

Смысл этих неравенств заключается в следующем, если научно–технический прогресс приводит к изменению размера производственных ресурсов в системе, то меняется, в частности, уровень производства. Восстановление оптимального уровня производства приводит некоторому восстановлению производственных ресурсов.

По аналогии с (9.75) при $\left(\frac{\partial K}{\partial P}\right)_L < 0$ получаем

$$\left(\frac{\partial Y}{\partial L}\right)_K > \left(\frac{\partial Y}{\partial L}\right)_P > 0 \quad \text{или} \quad |(\Delta Y)_K| > |(\Delta Y)_P|. \tag{10.33}$$

Это означает, что производительность труда при интенсивном развитии (когда рост происходит за счёт научно–технического прогресса при постоянных производственных ресурсах) оказывается выше, чем при экстенсивном развитии (когда рост происходит за счёт расширения производства при неизменной технологии).

В данном случае если научно–технический прогресс приводит к тому, что системе требуется меньше производственных ресурсов, то система выводится из равновесия, и в частности вырастает уровень производства. Это стимулирует процессы, стремящиеся понизить избыточное предложение товаров, и в системе сокращается количество необходимых производственных ресурсов.

Обсудим теперь причины, которые могут служить причиной нарушения равновесия в связи с изменением количества капитала в системе.

В первую очередь такой причиной может быть изменение функции A (амортизации). Научно-технический прогресс приводит к периодическому обновлению оборудования на более современное и эффективное. Период обновления оборудования является неравновесным промежуточным состоянием системы, хотя в течение всего периода может наблюдаться локальное равновесие, связанное с достаточно быстрым перераспределением трудовых ресурсов.

Еще одной причиной могут служить колебания котировок акций фондовой биржи, которые приводят к изменению цены используемого капитала, и, как следствие, к его перераспределению. Колебания цены на товарных и сырьевых рынках могут приводить к изменению доходности различных отраслей экономики (или отраслей промышленности) и вызывать перелив капитала из одной отрасли в другую.

Принцип Ле-Шателье в сформулированном нами виде говорит о том, что указанные изменения будут происходить не так быстро, как бы этого хотелось участникам рынка. Паровоз под названием промышленное производство трудно разогнать, но и остановить его тоже непросто.

Закон 6 (Аналог открытой системы). Если теперь предположить, что функция предпринимательской способности G определяет количество экономических субъектов (люди, не проявляющие предпринимательскую способность, включены в группы в качестве рабочей силы), то эта функция есть аналог количества вещества в термодинамике (9.31).

Тогда запишем уравнения для *открытой* экономической системы, аналогичные уравнениям термодинамики

$$\left(\frac{\partial N}{\partial G}\right)_{K,L} \overset{(9.117)}{=} \varphi, \quad \left(\frac{\partial L}{\partial G}\right)_{N,K} \overset{(9.119)}{=} -\frac{\varphi}{Y}. \tag{10.34}$$

Величина φ, аналогичная химическому потенциалу, в экономике означает среднюю величину ЧНД, приходящуюся на одного экономического субъекта или, что то же самое, среднюю прибыль π (10.4).

Кроме того, запишем уравнения

$$dN \overset{(9.120)}{=} Y\,dL - P dK + \varphi\, dG, \tag{10.35}$$

$$0 \overset{(9.131)}{=} L\,dY - K\,dP + G\,d\varphi, \tag{10.36}$$

а также аналог уравнений Максвелла (9.128)

$$\left(\frac{\partial Y}{\partial G}\right)_{L,K} = \left(\frac{\partial \varphi}{\partial L}\right)_{G,K}, \qquad \left(\frac{\partial P}{\partial G}\right)_{K,L} = -\left(\frac{\partial \varphi}{\partial K}\right)_{G,L}. \tag{10.37}$$

10.3 Функция полезности

Теория предельной полезности блага имеет два основных направления. Первое направление предполагает, что полезность можно оценить количествен-

но, и называется концепцией количественной или кардиналистской (от англ. cardinal — количественный) предельной полезности.

Кардиналистская концепция основана на следующих принципах:

1. Полезность представляет собой психическое явление, которое потребитель способен оценить *количественно*, например, как расстояние или температуру. Единица полезности называется *ютила*.

2. Принцип снижающейся полезности.
   В истории экономической мысли прочно поселился сформулированный Адамом Смитом во время лекции в университете города Глазго парадокс о воде и алмазе. Несмотря на то, что вода для человека обычно куда полезнее, чем алмаз, цена на последний гораздо выше.
   Для разрешения этого парадокса без привлечения понятия стоимости (овеществлённого в алмазе большого количества труда, затраченного на его поиск или добычу, в сравнении с отсутствием труда для получения воды) К. Менгер сформулировал *принцип снижающейся полезности*. *Общая полезность* U растёт с увеличением количества потребляемого i-го блага q_i, но *предельная полезность*

   $$MU_i \quad \text{или} \quad u_i = \frac{\partial U}{\partial q_i}, \tag{10.38}$$

   т.е. полезность каждой последующей единицы блага при этом снижается, пока не достигнет насыщения ($u_i = 0$), как это случается с водой и никогда не происходит с алмазами из-за их редкости.

3. Потребитель составляет набор благ таким образом, чтобы получить при этом максимум полезности. Это состояние называют *равновесием потребителя*.

Принципы количественной полезности сформулированы в виде двух, так называемых, *законов Госсена* (англ. Gossen laws):

1. Первый закон Госсена. При последовательном потреблении полезность каждой последующей единицы продукта ниже предыдущей

   $$\frac{\partial u_i}{\partial q_i} < 0. \tag{10.39}$$

2. Второй закон Госсена. Лицо максимизирует свою полезность, когда распределяет имеющиеся у него денежные средства между различными благами так, что достигает равного удовлетворения от последней единицы денег, истраченной на каждое из благ

   $$\frac{u_i}{p_i} = \frac{u_j}{p_j}, \tag{10.40}$$

   для любых благ i, j.

   При этом считается, что потребитель не может влиять на цены, заданные рынком.

Доказательство. Цена благ и количество денег (бюджет $M = \sum_{i=1}^{n} p_i q_i$) — главные факторы, ограничивающие потребление. Оптимальный вариант потребления мы можем вычислить при помощи функции Лагранжа $L = U + \lambda M$ для условного экстремума (см. с. 152):

$$\frac{\partial L}{\partial q_i} = u_i + \lambda p_i = 0, \tag{10.41}$$

$$\frac{\partial^2 L}{\partial q_i^2} = \frac{\partial u_i}{\partial q_i} < 0. \tag{10.42}$$

Из (10.42) следует первый закон Госсена, а из (10.41) — второй.

Общая полезность U есть скалярная функция, возрастающая при удалении от начала координат в пространстве с координатами $q_1, \ldots, q_n$. Предельные полезности благ образуют градиент общей полезности

$$\mathbf{u} = (u_1, \ldots, u_n) = \operatorname{grad} U, \tag{10.43}$$

который указывает в направлении наибольшего возрастания общей полезности и является нормалью в каждой точке произвольной поверхности уровня $U = const$, образованной равнополезными наборами благ.

Поверхности уровня ($U = const$) в случае пространства двух благ x и y есть изолинии, называемые *кривые безразличия* (рис. 10.1).

Величина наклона кривой безразличия в любой точке есть отношение предельной полезности блага, отсчитывающегося по горизонтальной оси, к предельной полезности блага, отсчитывающегося по вертикальной оси, и называется *предельная норма замещения* MRS (англ. marginal rate of substitution) одного блага другим:

$$MRS_{xy} = -\frac{u_x}{u_y}. \tag{10.44}$$

Доказательство

$$U = const, \quad \text{значит} \quad dU = u_x\,dx + u_y\,dy = 0, \quad \text{тогда} \quad \frac{dy}{dx} = -\frac{u_x}{u_y}.$$

Характерные свойства кривых безразличия:

1. Кривые безразличия имеют отрицательный наклон.
2. Кривые безразличия не пересекаются.
3. Кривые безразличия выпуклы к началу координат, что следует из (10.39). Их наклон уменьшается по мере движения вниз и вправо вдоль этих кривых.
4. Множество кривых безразличия образует *карту безразличия*. Чем дальше от начала координат расположена кривая безразличия, тем большему уровню благосостояния она соответствует. Всегда можно построить кривую безразличия, проходящую через любой набор благ (точку в в пространстве с координатами $q_1, \ldots, q_n$).

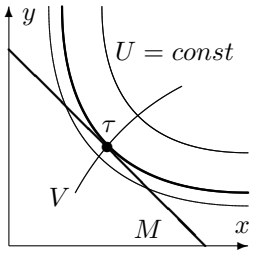


Рис. 10.1: Карта безразличия.

В случае пространства двух благ x и y прямая $M = p_x x + p_y y$ называется *бюджетная линия* (англ. budget line) (рис. 10.1). Точка касания τ бюджетной линии с самой отдалённой кривой безразличия дает *равновесие потребителя*, где он получает максимум полезности:

$$\frac{u_x}{u_y} \overset{(10.40)}{=} \frac{p_x}{p_y}. \tag{10.45}$$

Множество всех точек равновесия потребителя τ образует кривую $V = \{\max U\}$, которая получила название *опосредованная функция полезности*. Эта функция показывает, какой набор благ при заданных ценах и бюджете даёт максимум полезности.

Способность потребителя количественно оценивать психическое явление, называемое полезностью, вызывала сомнения, что в дальнейшем привело к гипотезе, что количественно полезность оценить нельзя, но можно сравнить любые две полезности между собой и таким образом упорядочить все полезности.

Второе направление теории предельной полезности называется концепцией порядковой или ординалистской (от англ. ordinal — порядковый) полезности, основанной на следующих гипотезах:

1. Гипотеза ненасыщения. При прочих равных условиях потребитель предпочитает большее количество блага меньшему его количеству.

2. Гипотеза полной упорядоченности. При наличии двух различных наборов благ потребитель всегда предпочитает один другому или признает их равнополезными.

3. Гипотеза транзитивности. Если потребитель предпочитает набор A набору B, а набор B набору C, то он предпочитает набор A набору C; соответственно, если набор A равнополезен набору B, а набор B набору C, то наборы A и C равнополезны.

4. Гипотеза рефлексивности. При наличии двух одинаковых наборов благ потребитель считает их равнополезными.

Впервые понятие «полезность» ввел в науку швейцарский математик Даниэль Бернулли. Понятие полезности в контексте социальных наук первым употребил Иеремия Бентам. Термин «предельная полезность» впервые использовал Фридрих фон Визер.

Теория предельной полезности или предельных издержек, называемая также маржинализмом (от англ. marginal — предельный) возникла в качестве противовеса марксизму в последней трети 19-го века трудами представителей австрийской школы, Карла Менгера, Евгения фон Бём-Баверка,

Фридриха фон Визера, Йозефа Шумпетера, а также Леона Вальраса (Лозаннская школа). Основные положения теории предельной полезности были сформулированы еще Германом Генрихом Госсеном в надолго забытой работе (1844), но начало массированного проникновения маржиналистских идей в экономическую литературу следует отнести только к середине 1880-х годов.

Представителями концепции порядковой (ординалистской) полезности являются Вильфредо Парето и Джон Хикс.

10.4 Производственная функция

Пространство выпуска (англ. output space) — пространство товаров с координатами $\dot{q}_i\,(i = 1, \ldots, n)$. *Пространство факторов* (англ. input space) — пространство с координатами $\dot{q}_j\,(j = 1, \ldots, k)$.

Производственная поверхность (англ. production possibility surface) есть k-мерная поверхность графика производственной функции (10.1) в $(k+1)$-мерном пространстве с координатами $\dot{q}_i, \dot{q}_1, \ldots, \dot{q}_k$. Все точки производственной поверхности соответствуют технически эффективным вариантам, точки ниже поверхности — неэффективным вариантам, а точки выше поверхности — недостижимым вариантам.

Производственная граница (англ. production frontier) есть граница области в пространстве выпуска, описывающая все возможные комбинации рассматриваемых товаров, которые можно произвести при полном использовании имеющихся факторов производства. Точки внутри области с производственной границей, соответствуют неэффективным вариантам, а точки вне области — недостижимым вариантам. При рассмотрении двумерного пространства выпуска производственная граница называется *кривая производственных возможностей*.

Предельный выпуск фактора производства (англ. marginal product of inputs) есть величина

$$MP_j = \frac{\partial \dot{q}_i}{\partial \dot{q}_j} = \frac{\partial TP}{\partial \dot{q}_j} \quad (j = 1, \ldots, k), \tag{10.46}$$

где $TP = \dot{q}_i$ — *общий выпуск* (англ. total product).

Предельные издержки (англ. marginal cost) вычисляются как

$$MC_i = \frac{\partial TC}{\partial \dot{q}_i} \quad (i = 1, \ldots, n). \tag{10.47}$$

Средний выпуск (англ. average product) есть величина

$$AP_j = \frac{TP}{\dot{q}_j}, \tag{10.48}$$

тогда величина

$$E_j \overset{(3.12)}{=} \frac{MP_j}{AP_j} \tag{10.49}$$

есть *эластичность производственной функции* по фактору $\dot{q}_j$.

Понятие «эластичности» в экономике ввел А. Маршалл (Великобритания), его идеи были развиты Дж. Р. Хиксом (Великобритания), П. Самуэльсоном (США) и др.

Средние общие издержки (англ. average total cost) вычисляются как

$$ATC_i = \frac{TC}{\dot{q}_i}. \tag{10.50}$$

Изокванта (англ. isoproduct curve, isoquant) (от греч. $\iota\sigma o\varsigma$ — одинаковый и лат. quantum — сколько) есть $(k-1)$-мерная поверхность уровня $\dot{q}_i = const$ производственной поверхности (графика производственной функции). Изокванту можно понимать как проекцию производственной поверхности на пространство затрат вдоль оси выпуска $\dot{q}_i$.

Градиент $\left(\frac{\partial \dot{q}_i}{\partial \dot{q}_1}, \dots \frac{\partial \dot{q}_i}{\partial \dot{q}_k}\right)$ является нормалью изокванты в каждой её дифференцируемой точке и указывает направление наибольшего возрастания производственной функции.

Изокоста (англ. isocost) есть $(k-1)$-мерная плоскость $TC = const$ в пространстве затрат. Вектор $(p_1, \dots, p_k)$ есть нормаль изокосты.

Пример. Рассмотрим двухфакторную производственную функцию

$$\dot{q} = f(x, y), \tag{10.51}$$

где $\dot{q}_1 = x$ и $\dot{q}_2 = y$.

Пространство выпуска двухфакторной функции представляет собой полуось $\dot{q} = TP \geqslant 0$, пространство факторов — участок плоскости между полуосями $x \geqslant 0$ и $y \geqslant 0$. Производственная поверхность есть двумерная поверхность в пространстве $\{TP, x, y\}$.

Изокванта двухфакторной функции представляет собой изолинию $\dot{q} = const$, которая называется *кривая замещения* факторов производства или *кривая равного выпуска* (рис. 10.2).

В каждой точке кривой замещения различные сочетания x и y дают одно и то же количество выпускаемой продукции.

Величина наклона кривой замещения называется *предельная норма технической замены* $MRTS$ (marginal rate of technical substitution):

$$MRTS_{xy} = -\left(\frac{\partial y}{\partial x}\right)_{\dot{q}} \overset{(3.161)}{=} \left(\frac{\partial \dot{q}}{\partial x}\right)_y \left(\frac{\partial y}{\partial \dot{q}}\right)_x \overset{(10.46)}{=} \frac{MP_x}{MP_y}. \tag{10.52}$$

Эластичность замещения факторов производства есть величина

$$E_{xy} = \left(\frac{\partial y}{\partial x}\right)_{\dot{q}} \frac{y}{x} = \frac{MP_x}{MP_y}\frac{AP_y}{AP_x} = \frac{E_x}{E_y}. \tag{10.53}$$

Эластичность замещения может быть постоянной величиной (англ. Constant Elasticity of Substitution, CES) и переменной величиной (англ. Variable Elasticity of Substitution, VES).

Основные свойства изоквант (кривых замещения):

1. Изокванты имеют отрицательный наклон.

2. Изокванты не пересекаются.

3. Множество изоквант образует *карту изоквант*. Большему выпуску продукции соответствует более удаленная от начала координат изокванта.

4. Если все виды ресурсов абсолютно необходимы для производства, то изокванты не имеют общих точек с осями координат.

Изокоста двухфакторной функции (10.51) $TC = p_x x + p_y y = const$ образует *бюджетную линию* производителя (рис. 10.2). Точка касания τ бюджетной линии с самой отдалённой кривой замещения дает *равновесие производителя*, где он получает максимальный выпуск при фиксированных издержках:

$$\frac{MP_x}{MP_y} \overset{(10.52)}{=} \frac{p_x}{p_y}. \tag{10.54}$$

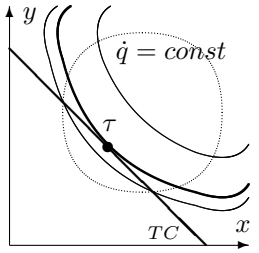


Рис. 10.2: Карта изоквант.

Нетрудно заметить, что производственная функция во многом похожа на функцию полезности, кривая замещения — на кривую безразличия, карта изоквант — на карту безразличия. И дело тут не в простом сходстве. По отношению к ресурсам фирма ведет себя как потребитель, и производственная функция характеризует именно эту сторону производства — производство как потребление.

Тот или иной набор ресурсов полезен для производства постольку, поскольку он позволяет получить соответствующий объем выпуска продукта. Можно сказать, что значения производственной функции выражают полезность для производства соответствующего набора ресурсов. В отличие от потребительской полезности эта «полезность» имеет вполне определенную количественную меру — она определяется объемом производимой продукции.

Типичная зависимость общего, предельного и среднего выпуска от фактора x показана на рис. 10.3, где изображена проекция производственной поверхности на плоскость $\{TP, x\}$, когда $y = const$.

Точка 1 есть точка перегиба TP и максимум MP_x:

$$\frac{\partial MP_x}{\partial x} = \frac{\partial^2 TP}{\partial x^2} \overset{(3.127)}{=} 0. \tag{10.55}$$

Уменьшение MP_x после достижения точки 1 называется *законом снижения предельной производительности переменного фактора*.

Точка 2 геометрически вычисляется при помощи касательной к TP, построенной из начала координат. Из данного построения видно, что в этой

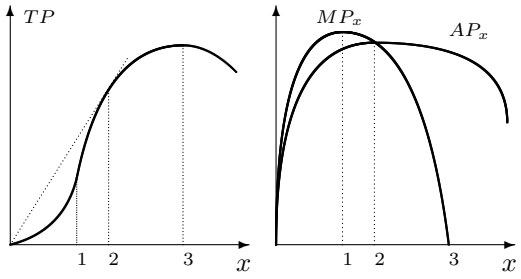


Рис. 10.3: Общий средний и предельный выпуск.

точке, во-первых, средний выпуск достигает максимального значения, и, во-вторых, линия MP_x пересекает линию AP_x

$$\frac{\partial AP_x}{\partial x} = 0, \quad AP_x = MP_x. \tag{10.56}$$

Область между точками 1 и 2 называется эффективной областью. В случае двухфакторной функции $\dot{q} = f(x, y)$ эта область в проекции на плоскость факторов $\{x, y\}$ (рис. 10.2) выделена пунктирной линией.

Максимум среднего выпуска означает, что точка 2 является *технически оптимальной*.

Точка 3 вычисляется как максимум TP:

$$MP_x = \frac{\partial TP}{\partial x} = 0. \tag{10.57}$$

Для области правее точки 2 характерно наличие одинаковой величины TP для двух различных значений x (или y), что приводит к характерному загибанию концов изоквант (рис. 10.2).

Таким образом, двухфакторная производственная функция представляет собой «холм», но, ввиду ограниченности ресурсов, дальний от начала координат склон холма обычно «не виден».

Изокванта $Y = const$ уравнения состояния или поверхности состояния (10.5) также представляет собой кривую замещения, аналогичную рис. 10.2. Величина наклона этой кривой имеет вид

$$MRTS_{LK} = -\left(\frac{\partial K}{\partial L}\right)_Y = \frac{MP_L}{MP_K} = -\left(\frac{\partial Y}{\partial P}\right)_K, \tag{10.58}$$

где MP_L — предельная производительность труда,

MP_K — предельная производительность капитала.

Доказательство

$$-\left(\frac{\partial K}{\partial L}\right)_Y \overset{(3.161)}{=} \left(\frac{\partial Y}{\partial L}\right)_K \left(\frac{\partial K}{\partial Y}\right)_L \overset{\substack{(10.27)\\(10.29)}}{=} \frac{MP_L}{MP_K} =$$
$$\overset{(10.27)}{=} -\left(\frac{\partial Y}{\partial L}\right)_K \left(\frac{\partial L}{\partial P}\right)_K \overset{(3.157)}{=} -\left(\frac{\partial Y}{\partial P}\right)_K$$

Изокоста имеет вид $TC = p_L L + p_K K$. Тогда в точке τ имеем

$$\frac{MP_L}{MP_K} = \frac{p_L}{p_K}. \tag{10.59}$$

Аналитически производственные функции могут быть представлены в различных формах — от простых, как линейная зависимость результата производства от одного исследуемого фактора, до весьма сложных систем уравнений, включающих рекуррентные соотношения, которыми связываются состояния изучаемого объекта в разные периоды времени.

Самая простая производственная функция — линейная, для которой факторы производства замещаемы в любых пропорциях:

$$\dot{q} = a_1 x_1 + \ldots + a_n x_n, \tag{10.60}$$

где $a_1, \ldots, a_n$ — оцениваемые параметры модели.

Наиболее широко распространены мультипликативно–степенные формы представления производственных функций. Их особенность состоит в следующем: если один из сомножителей равен нулю, то результат обращается в нуль. Это реалистично отражает тот факт, что в большинстве случаев в производстве участвуют все анализируемые факторы и без любого из них выпуск продукции оказывается невозможным.

В самой общей форме (она называется канонической) эта функция записывается так:

$$\dot{q} = A x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_k^{\alpha_k} \tag{10.61}$$

Здесь коэффициент размерности A зависит от избранной единицы измерений затрат и выпуска. Сомножители от первого до k-го могут иметь различное содержание в зависимости от того, какие факторы оказывают влияние на общий выпуск.

Степенные коэффициенты $\alpha_1, \ldots, \alpha_k$ показывают ту долю в процентном приросте конечного продукта, которую вносит каждый из сомножителей; они являются коэффициентами эластичности производственной функции относительно затрат соответствующего ресурса.

Сумма коэффициентов

$$\alpha_1 + \ldots + \alpha_k = \alpha, \tag{10.62}$$

означает *степень однородности* функции относительно количества ресурсов. Если $\alpha = 1$, то выпуск пропорционален использованным ресурсам. Когда сумма коэффициентов больше или меньше единицы, возникает *эффект масштаба* (англ. returns to scale) — непропорционально большой или непропорционально малый рост выпуска при увеличении затрат. Постоянная отдача от масштаба означает, что если ресурсы увеличиваются в λ раз, то и продукции производится в λ раз больше.

Первым случаем построения производственной функции была нелинейная производственная функция, предложенная Митчерлихом (США, 1909):

удобрения — урожайность. Независимо от него Спиллман (США) предложил показательное уравнение урожайности.

Производственная функция Кобба–Дугласа (Cobb–Douglas production function) применена американскими исследователями Ч. Коббом и П. Дугласом при анализе развития экономики США в 20–30-х гг. XX века:

$$Y = AL^{\alpha}K^{\beta}, \tag{10.63}$$

где α и β — константы (коэффициенты эластичности производства по труду L и капиталу K).

Если сумма $\alpha + \beta$ равна единице — функция линейно однородная; если больше или меньше единицы, имеет место эффект масштаба (соответственно положительный или отрицательный).

Эластичность замещения функции Кобба–Дугласа постоянна (CES)

$$E_{LK} = \frac{\alpha}{\beta}. \tag{10.64}$$

Неоклассическая модель экономического роста Роберта Солоу (1956) основывается на производственной функции Кобба–Дугласа

$$Y = PK^{\alpha}L^{1-\alpha}, \tag{10.65}$$

где $P = P_0 e^{rt}$ — нейтральный (одинаково влияющий на все задействованные для выпуска продукции ресурсы) технический прогресс,

t — время,

α, r — параметры производственной функции.

Число моделей производственной функции растет. Так Э. Денисон (США) предложил классификацию темпов экономического роста, состоящую из 23 факторов. Наряду с традиционными факторами, он вычленил факторы, влияющие на рост объема производства в расчете на единицу затрат, — прогресс в знаниях, уменьшение возраста капитала, прогресс в распределении ресурсов и др.

Типичная зависимость общих издержек TC от выпуска $\dot{q}$, которая представляет собой функцию по сути обратную функции TP, показана на рис. 10.4. Функции предельных MC и средних общих издержек ATC также изображены на рис. 10.4.

Общие издержки делятся на переменные издержки TVC (англ. total variable coast), зависящие от объема выпуска, и постоянные издержки TFC (англ. total fixed coast) — не зависящие от объема выпуска:

$$TC = TVC + TFC. \tag{10.66}$$

Трактовка постоянных и переменных издержек различается в экономической литературе. Наиболее распространенное, но с нашей точки зрения не совсем правильное определение смысла разделения постоянных и переменных издержек основано на их зависимости от объема выпуска, т.е.

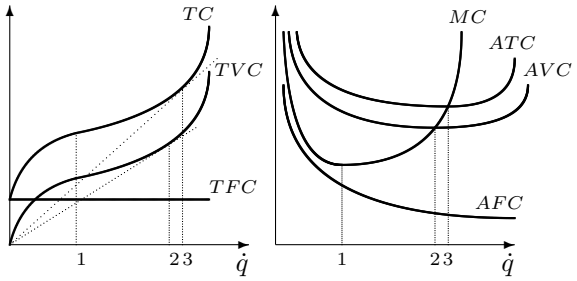


Рис. 10.4: Кривые затрат.

постоянные издержки не зависят от объема выпуска, а переменные — зависят. В связи с этим к переменным издержкам относят затраты на сырье, материалы, топливо, энергию, транспортные услуги, заработную плату рабочим и служащим и тому подобные переменные ресурсы. К постоянным издержкам относят расходы на содержание производственных зданий, машин, оборудования, рентные платежи, страховые взносы, жалование высшему управляющему персоналу и будущим специалистам фирмы. Подобное деление вносит определенную путаницу, так как все факторы производства присутствуют как в переменных издержках, так и в постоянных. Подобная традиция, по-видимому, связана с делением капитала на основной — переносящий свою стоимость на готовую продукцию по частям, и оборотный — переносящий свою стоимость на готовый продукт в течение одного производственного цикла.

Представляется более правильным исходить из деления капитала на постоянный — не изменяющий свою стоимость в процессе производства, и переменный — изменяющий свою стоимость в этом процессе, определение которых дано у Маркса.

Чтобы получить постоянный капитал из основного, нужно добавить к последнему все переменные материальные ресурсы и исключить жалование высшему управляющему персоналу и будущим специалистам фирмы. При различных соотношениях труда (переменного капитала) и капитала (постоянного капитала) структура постоянного капитала, т.е. соотношение между содержащимся в нем основным и оборотным капиталом, будет меняться.

Точка 1 есть точка перегиба TC и максимум MC:

$$\frac{\partial MC}{\partial \dot{q}} = \frac{\partial^2 TC}{\partial \dot{q}^2} \overset{(3.127)}{=} 0. \tag{10.67}$$

Точка 2 геометрически вычисляется при помощи касательной к TVC, построенной из начала координат. Из построения видно, что в этой точке, во-первых, средние переменные издержки AVC достигают минимального значения, и, во-вторых, линия MC пересекает линию AVC

$$\frac{\partial AVC}{\partial \dot{q}} = 0, \quad AVC = MC. \tag{10.68}$$

Точка 3 геометрически вычисляется при помощи касательной к TC, построенной из начала координат. Из данного построения видно, что в этой точке, во-первых, средние общие издержки ATC достигают минимального значения, и, во-вторых, линия MC пересекает линию ATC

$$\frac{\partial ATC}{\partial \dot{q}} = 0, \quad ATC = MC. \tag{10.69}$$

Предельные издержки MC есть цена предложения — минимальная цена, за которую производитель готов продать продукцию. При этом должно выполняться условие $MC \geqslant ATC$. На непродолжительное время, пока не нужно возобновлять постоянные затраты, кривая предложения фирмы удлиняется на отрезок кривой MC, лежащий между кривыми AC и AVC. Увеличение MC после достижения точки 2 называется *законом предложения* — предложение увеличивается при повышении и уменьшается при снижении цены товара.

Функция совокупного предложения для n товаров имеет вид

$$AS = \sum_{i=1}^{n} MC_i. \tag{10.70}$$

10.5 Формула Блэка–Шоулза–Мертона

В октябре 1997 г. профессорам Р. Мертону (Гарвардский университет) и М. Шоулзу (Стенфордский университет) была присуждена Нобелевская премия по экономике за их труд в области оценки опционов (1973). Шоулз работал совместно с Ф. Блэком, умершим в 1995 г., и их совместный результат известен под названием модели Блэка–Шоулза [1, 28]. Мертон сделал значительный вклад в модель и ее расширения и был награжден Нобелевской премией наравне с Шоулзом. Модель очень полезна при принятии инвестиционных решений, но не гарантирует прибыль на опционных торгах.

Формула Блэка–Шоулза–Мертона оценивает справедливую стоимость европейского опциона «колл». Под справедливой ценой понимается такая цена опциона, которая исключает возможность арбитража, т.е. исключает проведение сделок на рынке, позволяющих получить прибыль лишь за счет неправильной оценки опциона. Запишем формулу Блэка-Шоулза-Мертона в обозначениях, принятых в экономической литературе:

$$C_0 = \rho_0 p_0 - \rho_T K e^{-rT} \tag{10.71}$$

где T — время до эспирации (конца действия) опциона в годах;

r — постоянный дрифт (непрерывно начисляемая безрисковая процентная ставка) в единицах измерения год$^{-1}$;

p_0 — цена актива в начальный момент времени;

K — цена исполнения опциона (страйк опциона);

ρ_0, ρ_T — функции распределения вероятности цены актива в начальный момент времени и в момент времени T (доли сделок на рынке, совершенные по соответствующим ценам в указанные периоды времени);
C_0 — теоретическая цена опциона в начальный период времени (которую также называют премией).

Модель предполагает, что статистическое равновесие в системе еще не наступило и идет некий кинетический процесс, при котором изменяется функция распределения доходности актива.

Вывод формулы Блэка–Шоулза–Мертона, как и сама модель, овеяны ореолом таинственности и недоступности из-за сложности применяемого математического аппарата, понятного лишь физикам–теоретикам. При классическом выводе используется весьма сложная формула Ито, которая носит также название формулы замены переменных в стохастическом интеграле. Кроме того, приходится интегрировать уравнение в частных производных второго порядка.

Если изложить модель в общих чертах, то считается, что *доходность актива* ν подчиняется модели стандартного броуновского движения, т.е. происходит по *закону диффузии*, который гласит, что квадрат смещения броуновской частицы от начального положения, усредненный по многим броуновским частицам, пропорционален времени наблюдения.

Функция распределения $\rho(t, \nu)$ доходности может быть описана при помощи уравнения Фоккера–Планка, которое в одномерном случае сводится к уравнению диффузии:

$$\frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial \nu^2}; \tag{10.72}$$

где t — время;
D — коэффициент диффузии;
ν — доходность актива в качестве пространственной координаты.
Решение уравнения диффузии записывается в виде

$$\rho_t = \rho(t, \nu) = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{(\nu - \overline{\nu})^2}{4\pi D t}}, \tag{10.73}$$

где ν — совокупность координат броуновской частицы.

Вид решения совпадает с плотностью вероятности нормального закона распределения, если дисперсию в момент времени T приравнять следующей величине (по соотношению Смолуховского–Энштейна):

$$\overline{(\nu - \overline{\nu})^2} = (\Delta \nu)^2 = 2DT = \sigma^2 T, \tag{10.74}$$

где σ — постоянная волатильность цены актива (мера неопределенности относительно доходности ν данного актива).

Таким образом, изменение доходности актива ν по закону броуновского движения предполагает, что доходность распределена нормально, но дисперсия увеличивается пропорционально времени наблюдения, т.е. функция распределения «расплывается» со временем.

Модель геометрического броуновского движения величины доходности актива была предложена П. Самуэльсоном (1965) и именно она легла в основу модели Блэка–Шоулза–Мертона.

При выводе формулы Блэка-Шоулза существенными являются следующие предположения

1. Отсутствуют транзакционные издержки при совершении купли–продажи активов.
2. Не существует ограничений на короткие продажи.
3. Торговля происходит в непрерывном времени.
4. Базовый актив не приносит дивидендов.
5. На рынке существует постоянная непрерывная безрисковая процентная ставка r.
6. Волатильность базового актива постоянна во времени.
7. Доходность актива ν распределена нормально. Распределение $\rho(t,p)$ цены базового актива при этом является логнормальным.
8. Доходности актива в различные моменты времени являются независимыми.

Очевидно, что эти предположения не отражают реального положения на рынке. Во-первых, безрисковая процентная ставка и волатильность непостоянны во времени. Во-вторых, известно, что цены активов не имеют логнормального распределения, их распределения более островершинные и имеют более протяженные хвосты, чем логнормальное.

В то же время формулу (10.71) можно легко получить, если вполне обоснованно предположить, что изменение доходности за период времени dt, равное относительному изменению цены актива dS равно безрисковой процентной ставке r за вычетом изменения доли сделок на рынке $d\rho$, совершенных по текущей цене p в рассматриваемый период времени. Однако эти формулы для полного дифференциала, вообще говоря, не имеют отношения к броуновскому движению и являются более общими:

$$\frac{dp}{p} = r\,dt - \frac{d\rho}{\rho}. \tag{10.75}$$

Мы записали формулу (10.75) из общих соображений, аналогичных обычным рассуждениям в термодинамике, но ее можно получить и иным способом, если предположить, что доходность актива равна безрисковой процентной ставке r: $\nu = r$. Тогда богатство S, образованное рассматриваемым активом, и общая выручка TR от реализации актива за единицу времени связаны соотношением:

$$p\,\dot{q} = TR \overset{\substack{(10.17)\\(10.19)}}{=} r \oint_C p\,dq, \tag{10.76}$$

где $\dot{q}$ — количество актива, которое производится и потребляется в системе в единицу времени (т.е. скорость изменения актива на рынке);

p — текущая цена актива;

q — текущее количество актива на рынке.

Если мы запишем дифференциал соотношения (10.76), а затем умножим и разделим правую часть уравнения на dt, то получаем:

$$\dot{q}\,dp + p\,d\dot{q} = r\,p\,dq = rp\,\dot{q}\,dt.$$

Интеграл по замкнутому контуру (10.76) не был равен нулю, т.е. дифференциал $p\,dq$ не являлся полным дифференциалом. Если разделить левую и правую части полученного уравнения на интегрирующий множитель $TR = p\,\dot{q}$, который превращает пфаффову форму в полный дифференциал, то в результате получаем выражение:

$$\frac{dp}{p} = r\,dt - \frac{d\dot{q}}{\dot{q}}. \tag{10.77}$$

которое является полным аналогом уравнения (10.75), так как относительное изменение плотности распределения актива (изменение доли данного актива в общем количестве сделок на рынке), очевидно, пропорционально относительной скорости изменения его количества.

Будем считать, что цена актива p_T в момент эспирации T (в конце действия опциона) должна складываться из страйка (цены исполнения опциона) K и премии C_T в этот же момент:

$$p_T = K + C_T. \tag{10.78}$$

В связи с тем, что опцион приобретается в начальный период его действия, весьма разумно, что теоретическая цена опциона C_0, которую мы хотим найти, должна быть дисконтирована на безрисковую процентную ставку относительно премии C_T с учетом доли сделок на рынке, совершаемых на рынке в указанный период времени:

$$C_0 = \rho_T C_T e^{-rT}. \tag{10.79}$$

Интегрируя уравнение (10.75), являющееся полным дифференциалом, получаем уравнение

$$\ln\frac{p_T}{p_0} \overset{\text{(с. 163)}}{\underset{(4.7)}{=}} \int\limits_{p_0}^{p_T}\frac{dp}{p} = \int\limits_0^T r\,dt - \int\limits_{\rho_0}^{\rho_T}\frac{d\rho}{\rho} \overset{\text{(с. 163)}}{\underset{(4.7)}{=}} rT - \ln\frac{\rho_T}{\rho_0},$$

после преобразования которого (с учётом 10.78 и 10.79) и потенцирования получим формулу (10.71).

Заметим, что:

1. Можно считать безрисковую процентную ставку r и волатильность σ функциями времени. В этом случае в формуле (10.71) вместо степени экспоненты rT будет стоять интеграл $\int_0^T r(t)\,dt$, а в формуле (10.74) будет иметь место интеграл $(\Delta\nu)^2 = \int_0^T \sigma^2(t)\,dt$.

2. Мы никак не использовали предположение, что изменение функции распределения происходит по закону броуновского движения, т.е. является нормальным, и можем применять альтернативные распределения, которые представляются нам более правильными.

10.6 Распределение дохода

Предположим, что в экономической системе существует равновесное распределение дохода между экономическими субъектами. Сформулируем эту задачу таким образом, чтобы определить, как устанавливается это равновесие за счёт конкуренции между экономическими субъектами, и отбросим дополнительные факторы: изменение размера экономической системы от времени и роль государства в принудительном перераспределении дохода:

1. Будем считать, что величина совокупного дохода меняется значительно медленнее, чем происходит распределение этого дохода между экономическими субъектами. Другими словами, распределение дохода между субъектами (т.е. установление равновесия) происходит значительно быстрее, чем общий рост или спад производства, которые зависят от количества экономических субъектов и величины совокупного дохода. В этом случае величина совокупного дохода может считаться постоянной.

2. Будем искать распределение дохода между субъектами в отсутствии налоговой политики, которая предусматривает принудительное изъятие части дохода, и социальной политики, которая предусматривает адресное распределение части изъятого дохода. Если мы получим функцию равновесного распределения без учета указанного воздействия со стороны государства, то в дальнейшем мы попробуем сделать определенные выводы, как это воздействие должно влиять на смещение равновесия.

При этих допущениях мы можем представить экономическую систему как совокупность экономических субъектов, состояние которых определяется уровнем их дохода. Согласно таблице аналогий между термодинамикой и экономикой 10.2 доход системы интерпретирован как внутренняя энергия. Тогда доход экономического субъекта TR можно интерпретировать как энергию его состояния E. Следовательно, *функция распределения экономических субъектов по величине их дохода будет аналогична функции распределения Гиббса* (9.15).

В этом случае вероятность того, что значение дохода при измерении попадет в интервал от нуля до определенной величины TR, определяется уравнением

$$P = \int_0^{TR} \frac{e^{-\frac{E}{\theta}}}{\theta}\, dE \overset{(4.7)\ (\text{с. } 163)}{=} 1 - e^{-\frac{TR}{\theta}}, \tag{10.80}$$

где θ — нормировочный множитель, равный в данном случае среднему размеру дохода.

Откуда

$$TR = -\theta \ln(1 - P). \tag{10.81}$$

Доля дохода, входящая в этот интервал, соответственно равна

$$D = P + (1 - P)\ln(1 - P). \tag{10.82}$$

Доказательство

$$D = \frac{\int_0^{TR} \frac{E}{\theta} e^{-\frac{E}{\theta}} dE}{\theta} = 1 - e^{-\frac{TR}{\theta}} - \frac{TR}{\theta} e^{-\frac{TR}{\theta}} \overset{\substack{(10.80)\\(10.81)}}{=} P + (1 - P)\ln(1 - P).$$

Вероятность P эквивалентна доле граждан, располагающих соответствующей долей дохода D. Полученная зависимость (10.82) при построении дает так называемую «кривую Лоренца», которая демонстрирует неравенство в распределении доходов.

Когда рассматривается интервал $\Delta P = P_2 - P_1$, доля дохода определяется как разница $\Delta D = D_2 - D_1$.

Полученные результаты практически совпадают с реальным распределением доходов в обществе в отсутствие государственного вмешательства (табл. 10.3).

Группа	1	2	3	4	5
Равновесное распределение D	2,1	7,2	14	24,5	52,2
Россия (Росстат, 2002)	5,6	10,4	15,4	22,8	45,8
Россия (Березин, 2002)	4,5	9	12	17,3	57,3
Россия (Росстат, 2000)	5,8	10,4	15,1	21,9	46,8
Япония (1994)	8,7	13,2	17,5	23,1	37,5
Южная Корея (1994)	7,4	12,3	16,3	21,8	42,2
Китай (1994)	6,4	11,0	16,4	24,4	41,8
США (1994)	4,7	11,0	17,4	25,0	41,9
Великобритания (1994)	4,6	10,0	16,8	24,3	44,3
Мексика (1994)	4,1	7,8	12,3	19,9	55,9
Бразилия (1994)	2,1	4,9	8,9	16,8	67,5

Таблица 10.3: Распределение совокупного дохода населения в различных странах мира по долям в пяти доходных группах, расположенных в порядке возрастания дохода, %.

В таблице все семьи в экономике, распределенные в соответствии с их годовым доходом, разделены на пять равных групп (квинтилей) по 20% в каждой, и для каждой группы рассчитана доля совокупного дохода. По России взяты данные Росстата [30], а также из работы [9, Березин], где автор приводит как данные Госкомстата России, так и экспертные оценки,

которые он считает более объективными. Данные по другим странам взяты из [27, с. 430].

На рис. 10.5 изображена теоретическая кривая Лоренца, а также кривая Лоренца для России (2002), Японии (1994) и Бразилии (1994) построенные по данным табл. 10.3.

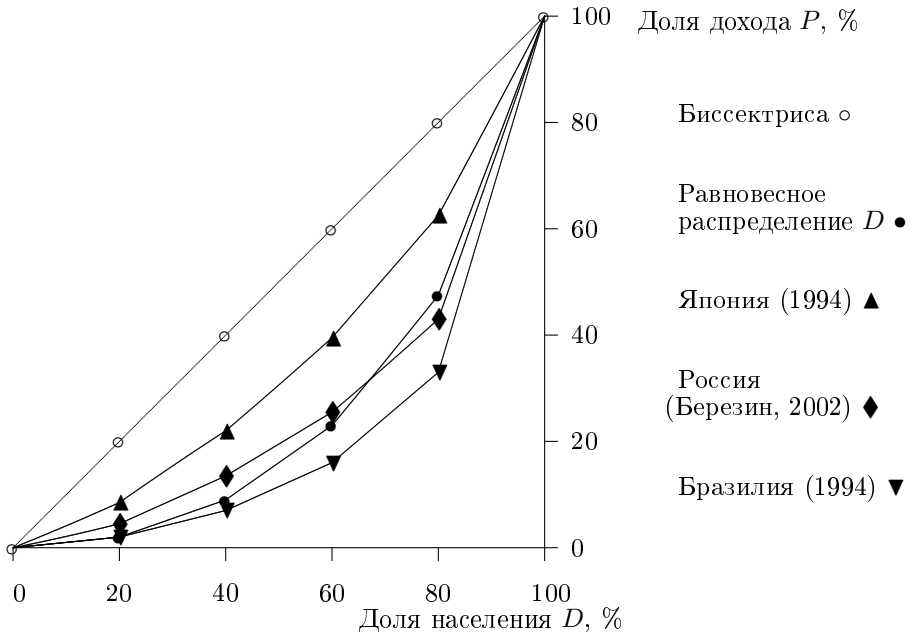


Рис. 10.5: Кривая Лоренца: значение удельных весов доходов групп и соответствующих групп по численности населения нарастающим итогом.

Очевидно, если бы распределение доходов было абсолютно равномерным, то точки кривой Лоренца лежали бы на биссектрисе. Смещение этой кривой вправо–вниз свидетельствует об усилении неравномерности распределения. В результате площадь фигуры, ограниченной с одной стороны биссектрисой, а с другой стороны – кривой Лоренца, увеличивается. Отношение площадей этой фигуры и треугольника, образованного биссектрисой и осями (D и P), называется *коэффициент Джини* J.

При абсолютно равномерном распределении доходов кривая Лоренца сливается с биссектриссой и коэффициент Джини будет равен нулю. Максимальная площадь фигуры между биссектриссой и кривой Лоренца может достигать площади треугольника, в этом случае коэффициент Джини будет равен единице, что будет означать абсолютное неравенство в распределении доходов — все доходы достаются самому богатому человеку.

Исходя из уравнения (10.82) мы можем вычислить равновесный коэффициент Джини, который оказывается равным $0,5$.

Доказательство

$$J = \frac{0,5 - \int_0^1 D\,dP}{0,5} \overset{(10.82)}{=} \frac{0,5 - \int_0^1 [P + (1-P)\ln(1-P)]\,dP}{0,5} = 0,5.$$

Если государство не проводит активной социальной политики, то распределение совокупного дохода будет близко к равновесному, а если элита общества при помощи государства использует власть для перераспределения доходов в свою пользу либо, наоборот, в пользу малоимущих слоев населения, то распределение может смещаться в ту или иную сторону. Однако, любое отклонение от этого равновесия может быть только искусственно вызванным и неоптимальным с точки зрения принципа наименьшего действия. Промежуток времени, в течение которого должен обязательно произойти переход к статистическому равновесию, называют *временем релаксации*.

Из рис. 10.5 видно, что в России (2002) распределение доходов было близко к равновесному, в Японии (1994) — смещено в пользу бедных (в сторону более равномерного распределения), а в Бразилии (1994) — в пользу богатых (более неравномерно).

Согласно [16], близкое к равновесному распределение доходов ($J \approx 0,5$) наблюдается, кроме России, например, в Малайзии, Перу и Киргизской республике, смещённое в пользу бедных ($J \approx 0,2 - 0,3$) — в некоторых европейских странах, и смещённое в пользу богатых ($J \approx 0,6 - 0,7$) — в Уганде, Боливии и Гондурасе.

Для демонстрации того, что уравнение (10.82) с $J = 0,5$ является равновесным, отклонение экономической системы от указанного распределения дохода является искуственно вызванным, и система всегда пытается вернуться к равновесию, проследим распределение совокупного дохода населения России в течение 34 лет (таб. 10.4)

Значения коэффициента Джини за указанный период представлены графически на рис. 10.6.

Снижение J в течение 1970–1990 годов отражает, по-видимому, тот факт, что борьба за сближение города и деревни, рабочих и интеллигенции приводила в СССР к элементарной уравниловке в заработной плате.

После 19 августа 1991 г. произошла смена способа производства и наступил период «дикого капитализма». Система распределения доходов населения вышла из под жесткого контроля государства, и коэффициент Джини двинулся к своему равновесному значению.

Процесс установления равновесия, приведенный на рис. 10.6, хорошо описывается кинетическим уравнением

$$J_{\text{к}} = J_0 \left(1 - e^{-\frac{t}{\tau}}\right), \tag{10.83}$$

где $J_0 = 0,418$,

$\tau = 1$ год,

t — время в годах отсчитывается от 1990 года (начало отсчета $t_0 = 0$).

Группа	1	2	3	4	5	J	$J_к$
1970	7,8	14,8	18,0	22,6	36,8	0,305	
1975	9,5	14,8	18,6	23,3	33,8	0,272	
1980	10,1	14,8	18,6	23,1	33,4	0,264	
1985	10,0	14,6	18,3	23,1	34,0	0,270	
1990	9,8	14,9	18,8	23,8	32,7	0,263	0,000
1991	11,9	15,8	18,8	22,8	30,7	0,225	0,264
1992	6,0	11,6	17,6	26,5	38,3	0,357	0,361
1993	5,8	11,1	16,7	24,8	41,6	0,378	0,397
1994	5,3	10,2	15,2	23,0	46,3	0,414	0,410
1995	5,5	10,2	15,0	22,4	46,9	0,415	0,415
1996	6,2	10,7	15,1	21,6	46,4	0,401	0,417
1997	5,9	10,2	14,8	21,6	47,5	0,414	0,418
1998	6,0	10,4	14,8	21,2	47,6	0,411	0,418
1999	6,1	10,4	14,7	20,9	47,9	0,412	0,418
2000	6,0	10,4	14,8	21,2	47,6	0,411	0,418
2001	5,6	10,4	15,4	22,8	45,8	0,407	0,418
2002	5,6	10,4	15,4	22,8	45,8	0,407	0,418
2003	5,5	10,3	15,3	22,7	46,2	0,411	0,418
2004	5,5	10,2	15,2	22,7	46,4	0,412	0,418

Таблица 10.4: Динамика изменения коэффициента Джини J и распределения совокупного дохода населения России по долям в пяти доходных группах, % (Группы расположены в порядке возрастания дохода).

За *время релаксации порядка* $3\,\tau$, т.е. где-то за 3 года, система пришла в равновесие. Примерно в это же время можно было подводить первые итоги приватизации.

Значения J_0 и τ рассчитаны методом наименьших квадратов для нелинейной регрессии (рис. 10.7).

Тот факт, что коэффициент Джини достиг значения 0,4, а не теоретического значения 0,5, может иметь два объяснения. Во-первых, какая-то социальная политика, направленная на снижение социального неравенства, государством все-таки ведется. Во-вторых, официальные данные отражают картину распределения только «наблюдаемых»доходов, в то время как «ненаблюдаемые» с очевидностью распределяются существенно более неравномерно.

Распределение доходов населения в России на основании экспертных оценок [9] даёт коэффициент Джини около 0,48, т.е. значительно ближе к теоретическому равновесному значению.

Сформулируем несколько выводов из проведённого исследования:

1. В экономической системе существует весьма устойчивое равновесное распределение совокупного дохода с коэффициентом Джини $J = 0,5$ и временем релаксации около трёх лет. Таким образом, распределе-

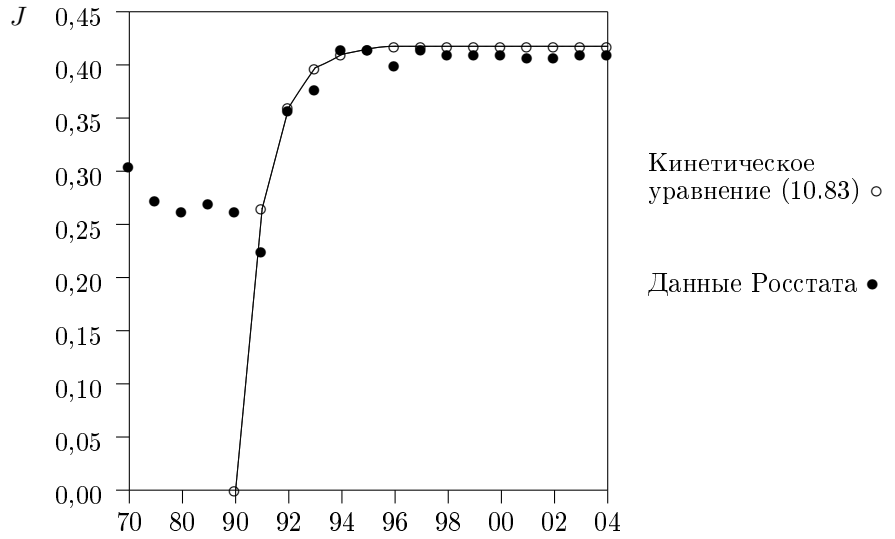


Рис. 10.6: Динамика изменения коэффициента Джини.

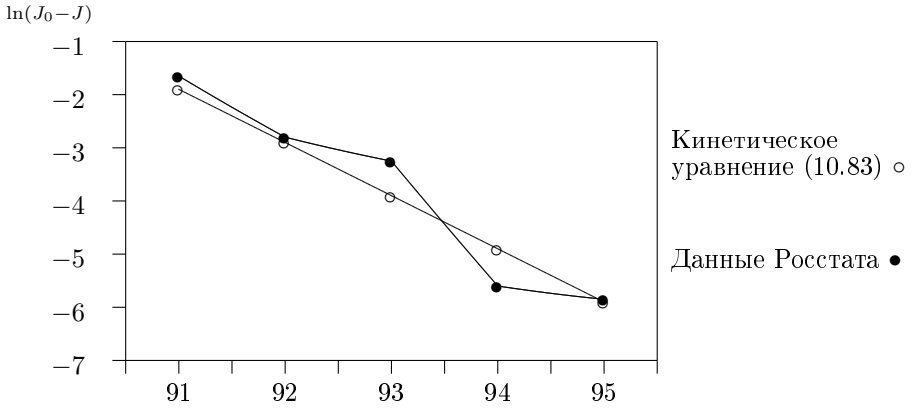


Рис. 10.7: Сравнение теоретической кинетики и кривой изменения J по данным Росстата (табл. 10.4) в логарифмическом виде: $\ln(J_0 - J)$. Тангенс угла наклона прямой равен $-\frac{1}{\tau} = -1$ (год$^{-1}$).

ние доходов в обществе демонстрирует высокую устойчивость, и ни дефолты, ни «удвоения ВВП» не приводят к его изменению.

2. Любое отклонение от равновесного распределения совокупного дохода может быть только искусственно вызванным и приводит к некоторому замедлению экономического развития, так как перераспределение доходов сопровождается чистыми потерями общества. Однако прямая зависимость между неравенством и экономическим развитием

при этом отсутствует.
Сделанный вывод не соответствует теории С. Кузнеца (1956), что экономическое развитие вначале ведет к увеличению неравенства, а затем к его уменьшению. Эта зависимость получила название кривой Кузнеца, или обратной U-кривой. За последние 40 лет предпринимались многочисленные попытки подтвердить эту зависимость. Тем не менее, она остается неясной, так как разные исследования демонстрируют прямо противоположные результаты.

3. Равновесное распределение дохода в принципе коррелирует с известным эмпирическим законом Парето, полученным на основании обширного статистического материала и имеющим степенную зависимость. Парето пришел к выводу, что параметры этого распределения примерно одинаковы и не различаются принципиально в разных странах и в разное время. «Кривая распределения доходов отличается замечательной устойчивостью, она меняется незначительно, хотя сильно преображаются обстоятельства времени и места, при которых ее наблюдают», — писал Парето в «Социалистических системах». Парето считал, что форма этой кривой зависит от биологически заданного распределения психологических особенностей людей.
Приведенные здесь рассуждения показывают, что полученная закономерность никакого отношения к психологии не имеет, а зависит от общих статистических закономерностей, которые имеют место, например в газах и жидкостях. Независимо от психологии человеческих взаимоотношений индивид всего лишь песчинка в игре законов больших чисел.

4. В равновесном распределении дохода имеет место ненулевая плотность вероятности при нулевом доходе, что было отмечено, например, в работе [16], но не следует из обычно применяемой в статистических исследованиях модели логнормального распределения дохода.

5. Равновесное распределение дохода с $J = 0,5$ возникает за счёт конкуренции между экономическими субъектами без государственного вмешательства. Следовательно, значения $J > 0,5$ не возникают даже при наличии олигархической структуры и одновременном отсутствии серьезной государственной социальной политики. Столь глубокое неравенство возможно только при целенаправленных административных или силовых методах (неэкономических действиях) государства.

6. Уменьшение неравенства также возможно только при целенаправленных законодательных, административных и силовых методах, например, при соответствующей налоговой и социальной политике государства. Но при этом должно произойти существенное подавление предпринимательской активности населения. В развитой экономической системе самый эффективный путь для снижения уровня бедности состоит в увеличении среднего дохода на душу населения. Это означает, что за счет уничтожения богатых искоренить бедность не удастся.

Заключение

Нельзя сказать, что попытки применить аппарат теоретической физики в экономической науке не предпринимались. Действительно, существуют, хотя и весьма неудачные, попытки интерпретации экономических переменных при помощи аналогичных физических величин.

В данной работе ставка была сделана не на аналогии между соответствующими функциями и переменными (хотя такие аналогии присутствуют и выписаны явно), а на аналогии при рассмотрении интегро-дифференциальных уравнений определенного вида, которые применяются в естественных науках и имеют аналитические решения, и, частично, на рассмотрении новых уравнений, для которых автором также найдены решения.

Вывод, который по мнению автора можно сделать после прочтения книги, заключается в подтверждении следующей гипотезы:

Любая природная система, как физическая, так, например, и экономическая характеризуется следующими свойствами, которые прямо следуют из геометрических свойств пространства

- система состоит из совокупности объектов с определённым свойством, находящихся в соответствующем поле;
- система представляет собой симплектическое многообразие (фазовое пространство);
- состояние системы задаётся точкой на этом многообразии;
- движение системы (точки) всегда происходит по геодезической линии, т.е. по наикратчайшему пути или «прямой» на многообразии.

Отличие между системами заключается в первом пункте этого списка — в выбранном для рассмотрения поле и свойстве объектов:

- механическая система изучает взаимодействие объектов, имеющих массу, в силовом поле;
- электродинамическая система изучает взаимодействие объектов, имеющих заряд, в электромагнитном поле;
- термодинамическая система изучает взаимодействие объектов, имеющих теплоёмкость, в поле термодинамических сил;

- экономическая система изучает взаимодействие объектов, имеющих цену, в поле экономических отношений.

Если в отношении физических систем данная гипотеза не вызывает больших сомнений, то в экономической науке существует «миф», что для описания экономической системы нельзя применять законы, аналогичные законам физики, так как данная система намного сложнее, и ее функционирование основано на других принципах. Этот миф базируется на очевидных представлениях от том, что законы общественных наук субъективны, в отличие от объективных законов физики.

Но, как известно, очевидные выводы обычно оказываются неправильными. Правильными обычно бывают простые, но неочевидные умозаключения. Данная закономерность отражает известную истину от том, что сложные задачи как правило имеют простые решения, однако эти решения сложно найти.

Физика (от др.-греч. $\varphi\acute{\upsilon}\sigma\iota\varsigma$ — природа) — область естествознания, изучающая наиболее общие и фундаментальные закономерности, определяющие структуру и эволюцию материального мира.

Термин «физика» впервые появился в сочинениях одного из величайших мыслителей древности — Аристотеля, жившего в IV веке до нашей эры. В русский язык слово «физика» было введено Михаилом Васильевичем Ломоносовым, когда он издал первый в России учебник физики в переводе с немецкого языка.

Но ведь никакой «физики» и «законов физики» в природе не существует, есть законы природы, часть которых выделена в область знания, называемую «физикой». Если мы считаем, что общество является частью природы и принадлежит материальному (а не потустороннему) миру, то законы, сформулированные в различных областях знания, изучающих этот материальный мир, должны быть аналогичны. В свою очередь и экономическая система не может быть сложнее самой природы.

Литература

[1] Black F., Scholes M. The pricing of optionsand corporate liabilities // Journal of political economy. — 1973. — № 81. — P. 637–659.

[2] Holdren J. Population and the energy probem // Population and Environment. J. Interdiscipl. Stud. 1991. Vol. 12, № 3.

[3] Андре Анго Математика для электро- и радиоинженеров. — М.: Наука, 1965. — 780 с.

[4] Андронов А.А., Витт А.А., Хайкин С.Э. Теория колебаний. - М.: Физматгиз, 1959. - 915 с.

[5] Арнольд В. И. Гюйгенс и Барроу, Ньютон и Гук — первые шаги математического анализа и теории катастроф, от эвольвент до квазикристаллов — М.: Наука. Гл. ред. физ.-мат. лит. — 1989. — 96 с.

[6] Арнольд В. И. Математические методы классической механики. — М.: Наука. Гл. ред. физ.-мат. лит. 1974. — 472 с.

[7] Арнольд В. И. Обыкновенные дифференциальные уравнения. — Ижевск: Ижевская республиканская типография. 2000. — 368 с.

[8] Арнольд В. И. Теория катастроф. — М.: Наука. Гл. ред. физ.-мат. лит. 1990. — 128 с.

[9] Березин И. С. Распределение доходов населения России — 2003, 2002, 2001. [http://www.marketologi.ru/publ.html].

[10] Больцман Л. Лекции по теории газов. — М., 1953. — 554 с.

[11] Гантмахер Ф.Р. Лекции по аналитической механике. М.: Наука, 1966. - 300 с.

[12] Гиббс Дж. В. Основные принципы статистической механики. — М.: ОГИЗ «ГОСТЕХИЗДАТ», 1946. — 203 с.

[13] Голдстейн Г. Классическая механика. — М.: ОГИЗ «ГОСТЕХИЗДАТ», 1967. — 408 с.

[14] Гуров К. П. Феноменологическая термодинамика необратимых процессов (физические основы). — М.: Наука, 1978. — 128 с.

[15] Зоммерфельд А. Механика. - Ижевск: НИЦ "Регулярная и хаотическая динамика 2001. - 368 с.

[16] Кислицына О. А. Неравенство доходов в России в переходный период: чем оно объясняется? — М.: EERC, 2003. — 79 с. [www.eerc.ru/details/download.aspx?file_id=3741]

[17] Куснер Ю. С., Царев И. Г. Принципы движения экономической системы. — М.: ФИЗМАТЛИТ, 2008. — 200 с.

[18] Лагранж Ж. Аналитическая механика. - М. - Л.: Гостехиздат, 1950.

[19] Ландау Л.Д., Лифшиц Е.М. Теоретическая физика. Т. I. Механика. - М.: Наука, 1988. - 216 с.

[20] Ландау Л.Д., Лифшиц Е.М. Теоретическая физика. Т. II. Теория поля. - М.: Наука, 1988. - 509 с.

[21] Ландау Л.Д., Лифшиц Е.М. Теоретическая физика. Т. V. Статистическая физика. М.: Наука, 1976. - 584 с.

[22] Ландис Е. Е. Тангенциальные особенности // Функцион. анализ и его прил. — 1981. — Т. 15, вып. 2. — С. 36–49.

[23] Левантовский Л. В. Особенности границы области устойчивости // Функцион. анализ и его прил. — 1982. — Т. 16, вып. 1. — С. 44–48.

[24] Маркс К. Капитал. Критика политической экономии. Т. III. — М.: Политиздат, 1949. — 932 с.

[25] Маркс К., Энгельс Ф. Введение (из экономических рукописей 1857–1858 годов) Соч., т. 12.

[26] Маркс К., Энгельс Ф. Немецкая идеология. Т. I.

[27] Мэнкью Н. Г. Принципы экономикс. — СПб.: Питер Ком, 1999. — 784 с.

[28] Первозванский А. А., Первозванская Т. Н. Финансовый рынок: расчет и риск. — М.: ИНФРА-М, 1994. — 192 с.

[29] Платонова О. А. Особенности взаимного расположения поверхности и прямой // Успехи мат. наук. — 1981. — Т. 36, вып. 1. — С. 248–249.

[30] Российский статистический ежегодник. 2004. М.: Стат.сб./Росстат, 2004. — 725 с.

[31] Спивак М. Математический анализ на многообразиях. — М.: Мир, 1968. — 164 с.

[32] Сычев В. В. Дифференциальные уравнения термодинамики. — М.: Высш. шк., 1991. — 224 с.

[33] Тамм И. Е. Основы теории электричества. — М.: Гос. изд. тех.-теор. лит., 1956. — 620 с.

[34] Ферми Э. Термодинамика. Изд. Харьковского университета: Харьков, 1969. — 139 с.

[35] Царев И. Г. Физико–математические аналогии в экономике. — М.: ФГУП ЦПП, 2005. — 215 с.

[36] Энгельс Ф. Наброски к критике политической экономии. Соч., т. 1.

Предметный указатель

Аддитивность, 271
Адиабата, 335
Аффинор, 97
Алгебра
  Ли, 146
  универсальная, 46
Амортизация, 380
Аналитическое продолжение, 41
Анализ
  функциональный, 235
  предельный, 374
Аппроксимация, 103
Аргумент
  комплексного числа, 37
Асимптотическая
  касательная, 151
  линия, 151
Асимптотическое
  направление, 151
Ассоциативность, 17
Атлас, 136
Базис, 52
  косоугольный, 22
  сопряжённый, 54
  стандартный, 53
Бином, 26
Биссектриса, 12
Благо, 373
  потребительское, 375
Близость, 7
Богатство, 377
Брахистохрона, 267
Цена, 373
  капитальная, 380
  предложения, 376
  прокатная, 380
  равновесная, 376
  спроса, 376
Центр масс, 271
Цепь
  сингулярная, 34
Цикл, 36, 177
  Карно, 343
Цилиндр, 11, 150
Циркуляция
  ротора, 178
  вектора, 180
Частота, 238
  круговая, 238
Число
  алгебраическое, 39
  комплексное, 36
Давление, 331
Действие, 236
  укороченное, 238
Деньги, 373
Диаграмма
  состояния, 335, 361
Диэлектрик, 300, 354
Диффеоморфизм, 115
  однопараметрический, 202
Дифференциал, 107, 113
  полный, 131, 141, 156, 178
Дифференцирование, 107
  внешнее, 125
Диполь, 301
Дискретность, 9
Дискриминант, 29
Дисперсия, 255
Дистрибутивность, 17
Дивергенция, 106, 127, 131, 209
Движение, 141

финитное, 238
инфинитное, 238
Эффект
Лоренца–Фицджеральда, 288
электрокалорический, 354
электрострикционный, 354
магнитокалорический, 354
пироэлектрический, 354
пьезоэлектрический, 354
Эйлера
углы, 78
Экспонента, 208
Экстремаль, 235
Экстремум, 153
условный, 153
Эластичность, 110
Электродинамика, 297
Электропроводность (проводимость), 300
Элемент
длины, 122
объёма, 123, 140
площади, 123
Эллипс, 11
Эллипсоид, 11, 150
инерции, 277
трёхосный, 11
вращения, 11
Энергия
Гельмгольца, 336
Гиббса, 336
кинетическая, 264
покоя, 291
полная, 264
поля
электрическая, 317
магнитная, 317
потенциальная, 264
релятивистская, 293
свободная, 336
тока, 317
внутренняя, 332
взаимодействия
токов, 317
зарядов, 317
Энтальпия, 336
Энтропия, 332
абсолютная, 346
Эвольвента, 147
Эволюта, 147
Факторы
производства, 375
Фаза, 202, 360
Ферромагнетик, 302
Фигура, 34
Форма, 26
билинейная, 73
дифференциальная, 119, 121
эрмитова, 74
кососимметрическая, 75
квадратичная, 74, 84
неособая, 93
общественная, 375
пфаффова, 121
полилинейная, 69
точная, 131
вырожденная, 93
внешняя, 90
замкнутая, 131
Формула
Бернулли, 250
Эйлера, 37
Функционал, 235
Функция, 19
Дирака δ-функция, 197
Гамильтона, 225
Лагранжа, 154
алгебраическая, 29
аналитическая, 28, 40, 116
целая, 43
чётная, 25
дифференцируемая (гладкая), 104, 108, 119
единичная, 198
элементарная, 24
гармоническая, 129
голоморфная, 40
голономная, 132
характеристическая, 336
интегрируемая, 167
колебание, 32
комплексная, 40

линейная, 25
мероморфная, 172
нечётная, 25
однородная, 157
однозначная, 40
ограниченная, 32
полилинейная, 69
процесса, 178
производственная, 375
производящая, 233
распределения, 330
скалярная, 47
сопряжённая, 173
состояния, 178
трансцендентная, 29
вероятности, 245
Функтор, 21
Геометрия
внутренняя, 152
Герполодия, 278
Гессиан, 118
Гомеоморфизм, 30
Гомология, 179
Гомотопия, 223
Градиент, 110
График
функции, 24
отображения, 16
Граница
множества, 8, 33
Грань
множества, 33
нижняя, 32
верхняя, 32
Группа, 18
абелева, 18
ортогональная, 77
преобразований, 18
симметрии, 18
симплектическая, 97
унитарная, 77
Импульс, 268
Импульсы
обобщённые, 264
Индекс, 6
суммирования, 52
Индукция, 11
электрическая, 302
магнитная, 303
Информация, 253
Интеграл, 167
Стильтьеса, 198
движения, 226, 285
кратный, 168
криволинейный, 179
неопределённый, 165
определённый, 167
первый, 219
по Риману, 169
поверхностный, 179
повторный, 169
расходящийся, 186
сходящийся, 186
сингулярный, 186
собственный, 186
Интервал, 10, 286
несобственный, 10
Инвариант, i, 53
адиабатический, 241
интегральный, 177, 224
Пуанкаре, 230
Пуанкаре-Картана, 231
калибровочный, 310
топологический, 31
Инволюция, 16, 228
Испарение, 360
Излом, 30, 103, 154
Изобара, 335
Изоэнтропа, 335
Изохора, 335
Изокоста, 394
Изокванта, 394
Изолиния, 111, 335, 365
Изоморфизм, 16
Изотерма, 335
Изотропность, 283
пространства, 284
Капитал, 378
Карта, 133, 136
безразличия, 391
Категория, 4
Кинетика, 334

Класс, 5
Коэффициент
  Джини, 406
  Джоуля–Томсона, 367
  расширения
    объёмного , 369
  сжимаемости
    адиабатной, 367
    изотермической, 369
Колебание, 238
Кольцо, 18
Коммутативность, 17
Коммутатор, 67
Конфигурация, 132
Константа, 19
Континуум, 13, 245
Конус, 150
Координаты, 21
  цилиндрические, 23
  декартовы, 23, 53, 77
  канонические, 225
    действие–угол, 241
  криволинейные, 22
  обобщённые, 264
  параболические, 23, 24
  полярные, 23
  сферические, 24
Кортеж, 6, 62
Ковариантность, 294
Ковектор, 54
Край
  многообразия, 133
  поверхности, 32
Кратность, 79
Кривая, 31
  фазовая, 202
  фазового перехода, 361
  интегральная, 203
  комплексная, 142
  плоская, 146
  пограничная, 154, 360
  замещения, 394
Кривизна, 147
  Гауссова (внешняя), 150
  нормальная, 150
Кручение, 148
Круг, 10
Куб, 11
  сингулярный, 31
    ориентированный, 176
  стандартный, 31
Кусок
  цепи, 34
Квадрат, 11
Квадратура, 166
Лагранжиан, 226
Лапласиан, 129, 130
Лемма
  Эйлера, 158
Линия
  геодезическая, 234
  параболическая, 150
  ротора, 127, 181
  уровня (изолиния), 111
  векторная (тока), 143
Магнетик, 302, 353
Макроэкономика, 378
Масса
  покоя, 293
Математическое ожидание, 253
Матрица, 56
  Грама, 75
  Якоби, 114
  альтернирующая, 75
  эрмитова, 74
  кососимметрическая, 75
  невырожденная, 58
  обратная, 64
  ортогональная, 77
  особая, 58
  преобразования, 61, 76
  расширенная, 57
  симметрическая, 74
  сопряжённая, 73
  транспонированная, 64
  унитарная, 77
Механика, 263
  Гамильтона, 264
  Лагранжа, 264
  Ньютона, 268
  релятивистская, 294
Мера, 14

Лебега, 14
нуль, 15
Метрика, 13
касательного пространства, 118
пространства операторов, 63
риманова, 137
Микроэкономика, 378
Минор, 58
Мировая
линия, 289
Многочлен, 25
однородный, 26
Многообразие, 132
аналитическое, 136
ориентируемое, 138
риманово, 137
симплектическое, 222
вложенное, 135
Множество, 4
числовое, 9
измеримое, 14
компактное, 13
комплексное, 36
несчётное, 6
открытое, 8
производное, 9
пустое, 5
счётное, 6
совершенное, 9
уровня, 112, 219, 239
всюду плотное, 9
замыкание, 8
замкнутое, 7, 9
Множитель
интегрирующий, 132
Модель, 261
Блэка–Шоулза, 400
математическая, 261
Модуль, 14
комплексного числа, 37
Момент
дипольный (электрический), 300
импульса, 273
инерции, 276
кинетический, 273
количества
движения, 273
магнитный, 302
силы, 274
вращающий, 274
Морфизм, 4
Намагниченность, 303
Напряжение, 323
Непрерывность, 6, 7
функции, 29
Независимая
функция, 219
Норма, 14
линейного оператора, 62
вектора, 46
Нормаль, 71, 111
Объект, 4, 261
Объём, 13
многообразия, 168
нуль, 15
параллелепипеда, 50
Область, 9
двухфазная, 360
однофазная, 360
определения, 15
Обмен, 376
Общая выручка, 377
Общество, 375
Общие издержки, 377
Один-форма, 54
Однородность
пространства, 283
времени, 284
Огибающая, 196
Окрестность, 8
Окружность, 10
соприкасающаяся, 147
Операция
алгебраическая, 17
линейной суперпозиции, 45
Оператор, 21
$\mathbb{C}$-линейный, 62
Лапласа, 129
дифференцирования, 105, 108, 113, 119
линейный, 61
набла, 130

по направлению, 144
производящий, 209
сдвига, 120
симплектический, 97
сопряжения
комплексного, 96
сопряжённый, 73
Определитель, 58
Грама, 88
Ориентация
индуцированная, 139
пространства, 49, 86, 88
согласованная, 138
Орт, 53
внешней нормали, 139
Осциллятор, 238
Ось числовая, 6
Отношения
экономические, 374
общественные, 375
Отображение, 15
антиконформное, 42
биективное, 16
изометрическое, 152
конформное, 42
обратное, 16
сопряжённое, 100
Отрасль, 376
Параболоид
гиперболический, 150
вращения, 150
Параллелепипед, 12
Параллелограмм, 12
Параметр, 140
Переменная, 14
функции, 19
сопряжённая, 265, 333
Переменные
экономические, 374
Перестановка, 26
Первообразная, 165
Плавление, 360
Плоскость, 6, 11
касательная, 107
мнимая, 96
расширенная, 41
соприкасающаяся, 147
вещественная, 96
Площадь, 12
Плотность, 21
объёмная, 189, 266
энергии электрической, 317
энергии магнитной, 317
потока, 273
поверхностная, 189, 266
распределения, 129
силы, 273
тока, 298
вероятности, 257
вещества, 273
заряда
объёмная, 297
поверхностная, 298, 325
Покупатель, 375
Поле, 19
числовое, 19
электрическое, 299, 354
гармоническое, 106
магнитное, 299, 353
направлений, 205
потенциальное, 106, 127, 265
силовое, 265
скалярное, 104
векторное, 104, 139
гамильтоново, 225
вихревое, 106, 127
Полезность, 373
предельная, 374, 390
Полодия, 278
Положение равновесия, 204
Полупространство, 11
Поляризация, 302
Полюс, 42
Портрет
фазовый, 202
Последовательность, 20
фундаментальная, 20
испытаний, 250
расходящаяся, 20
сходящаяся, 20
Постоянная
Больцмана, 331

Потенциал, 265, 338
объёма, 189
поверхности, 189
скалярный, 129, 304
термодинамический, 336
векторный, 129, 304
Поток
фазовый, 203
ротора (вихря), 178, 181
вектора, 180
Потребление, 376
Поверхность, 31
адиабатная, 334
компактная, 31
ориентированная, 49
производственная, 393
риманова, 41
состояния, 337
уровня, 112
Правило
Лебница, 158
Лопиталя, 157
цепное, 110
Прецессия, 278
Предел, 14
функции, 21
последовательности, 20
Предложение
совокупное, 377
Предпринимательская
способность, 378
Предприятие, 379
Преобразование, 16
Галилея, 270
Лежандра, 158
Лоренца, 288
каноническое, 224, 231
свободное, 233
ортогональное, 76
подобия, 65
симплектическое, 96
сопряжённое, 73
тождественное, 61
унитарное, 76
Прибыль, 377
Принцип
Д'Аламбера, 280
Ле-Шателье–Брауна, 350
детерминированности
Ньютона, 266
наименьшего действия
Гамильтона, 267
Мопертюи, 267
относительности, 270
продолжения
аналитического, 37
Процесс
Джоуля–Томсона, 339
адиабатический, 335
экономический, 374
изобарический, 335
изоэнтропийный, 335
изохорический, 335
изотермический, 335
необратимый, 334
неравновесный, 333
обратимый, 334
равновесный, 333
термодинамический, 333
Продавец, 375
Прогресс
научно
–технический, 381
Произведение
эрмитово, 48
кососкалярное, 49
невырожденное, 94
прямое, 5
множеств, 5
скалярное
невырожденное, 94
смешанное, 87
тензорное, 98
векторное, 85
внешнее, 49, 88
внешнее форм, 91
внутреннее (скалярное), 47
Производная, 108
частная, 110
по направлению, 140, 144
по нормали, 156
полная, 141, 156

тензорная, 114
векторная, 113
Производство, 375
Проницаемость
диэлектрическая, 302
магнитная, 303
Пространство, 6
Минковского, 289
аффинное, 22
бесконечномерное, 53
евклидово, 71
факторов, 393
фазовое, 202
функциональное, 235
касательное, 112
кокасательное, 221
комплексное, 48
конечномерное, 53
конфигурационное, 132, 264
полное, 21
симплектическое, 93
событий, 245
сопряжённое, 54
унитарное, 73
векторное, 45
вероятностное, 245
выпуска, 393
Проводник, 299
Прямая
комплексная, 36
Псевдоскаляр, 86
Пуассона
скобка, 146, 228
Работа, 331
силы, 265
Радикал, 25
Ранг матрицы, 59
Распределение, 376
Гиббса, 331
Пуассона, 252
биноминальное, 250
нормальное, 257
Расслоение, 113
кокасательное, 221
многообразия, 137
Равновесие
экономическое, 384
потребителя, 392
производителя, 395
термодинамическое, 334, 340
Разбиение, 166
параллелепипеда, 166
Размерность
множеств, 7
величин, 7
Размещение, 26
Разрыв
первого рода, 30, 154
второго рода, 30
Реакция
связи, 281
Резонанс
внутренний, 242
Рынок, 375
Ротор, 106, 127, 131
Ряд
Лорана, 42
Тейлора, 120
функциональный, 27
сходящийся, 27
Сегмент, 10
Сфера, 10
риманова, 41
Сходимость, 14
абсолютная, 20
условная, 20
Сила
Лоренца, 308
активная, 281
электродвижущая, 323
пондеромоторная, 323
тока, 298
Силы
обобщённые, 264
потенциальные, 264
Символ
Кронекера, 54
Леви–Чивиты, 143
Система, 261
Лоренца, 289
автономная, 203
динамическая, 203

единиц Гаусса, 7
экономическая, 375
гамильтонова, 225
инерциальная, 269
консервативная, 285
потенциальная, 264
социальная, 375
термодинамическая, 333
вырожденная, 241
замкнутая (изолированная), 265, 284
Скачок, 30
Скаляр, 47
мировой, 289
Скорости
обобщённые, 264
Скорость, 141
локальная, 141
переносная, 141
угловая, 143
вещества, 273
заряда, 298
След, 79
Слой, 113
Событие, 289
достоверное, 246
случайное, 245
Сочетание, 27
Сопротивление, 300, 323
Спектр
матрицы, 79
Спрос
совокупный, 376
Среда
дискретная, 265
электромагнитная, 299
несжимаемая, 266
сплошная, 266
внешняя, 265
Средства
производства, 375
Степень, 6
свободы, 133, 203
Стоимость
добавленная, 380
прибавочная, 374
товара, 373
Структура
евклидова, 71, 91
эрмитова, 73
симплектическая, 93
Субъект, 261
экономический, 375
социальный, 375
Сублимация, 360
Связность, 9
многообразия, 134
Связь, 279
геометрическая, 112
голономная, 279
идеальная, 281
реономная, 279
склерономная, 279
Свёртка, 54
Тело, 333
рабочее, 343
Температура
абсолютная, 331
Тензор, 97
электромагнитного поля, 310
инерции, 276
контрвариантный, 310
ковариантный, 310
Теорема
Байеса, 248
Бернули, 247
Бореля–Лебега, 13
Эмми Нётер, 227
Фубини, 169
Гаусса, 314
Гаусса–Остроградского, 183
Гельмгольца, 128
Лиувилля, 222, 223
Муавра–Лапласа, 252
Нернста–Планка, 346
Пойнтинга, 319
Пуанкаре, 225
Пуассона, 228
Стокса, 177
Шаля, 275
Якоби, 234
о выпрямлении, 207

существования
и единственности, 206
вероятностей
сложения, 247
умножения, 247
Теория
относительности
специальная, 294
Теплота, 332
парообразования, 363
Теплоёмкость, 340
изобарная, 340
изохорная, 340
магнетика, 353
пограничной кривой, 369
Термодинамика, 333
Точка, 6
эллиптическая, 149
фазовая, 202
гиперболическая, 150
граничная, 8
изолированная, 9, 42
критическая, 361
материальная, 265, 268
мировая, 289
нерегулярная, 149
несобственная, 41
особая, 42, 204
параболическая, 150
перегиба, 147
правильная, 42
предельная, 9
прикосновения, 9
разрыва, 30
существенно особая, 42
тройная, 361
угловая, 103
внешняя, 8
внутренняя, 8
Ток
электрический, 298
линейный, 298
Топология, 8
Тор, 11
Товар, 373
Траектория, 141, 268
движения, 264, 267
фазовая, 202
Трансверсальность, 220
Треугольник, 12
Трубка
ротора, 178, 181
векторная, 181
Труд, 373
Угол, 11
Уравнение, 28
Д'Аламбера, 315
Эренфеста, 370
Гамильтона–Якоби, 234
Гиббса
фундаментальное, 359
Гиббса–Дюгема, 360
Кеезома, 365
Клапейрона, 345
Клапейрона–Клаузиуса, 363
Лапласа, 129, 191, 320
Майера, 345
Планка–Гиббса, 371
Пуассона, 191, 315, 320
дифференциальное, 201
характеристическое, 79
характеристик, 220
квадратное, 29
линейное, 208
моментов, 274
неразрывности, 184
прямой, 51
состояния, 333
векторной линии, 143, 204
волновое, 232, 315
вращения, 143
Уравнения
Д'Аламбера–Лагранжа, 282
Эйлера, 276
Эйлера–Лагранжа, 227
Максвелла, 313, 344, 352
канонические, 225
Ускорение, 141, 269
Условие
несжимаемости, 184
Условия
Коши–Римана, 116

Устойчивость, 216
асимптотическая, 216
Вариационное исчисление, 234
Вариация функционала, 235
Вектор, 6, 46
–функция, 113
Пойнтинга, 319
аксиальный, 86
фазовой скорости, 203
изотропный, 292
касательный, 112
кокасательный, 221
компоненты, 46
направления, 140
направляющий, 51
нулевой, 93
полярный, 86
пространственно-подобный, 292
собственный, 79
временно-подобный, 292
Векторы
коллинеарные, 51
компланарные, 52
линейно независимые, 51
полярные, 51
Величина
бесконечно малая, 107
экстенсивная, 271, 337
интенсивная, 271, 338
удельная, 338
взаимная, 265, 333
Вероятность, 246
полная, 248
условная, 248
Вычет, 172
Вихрь, 106
Восприимчивость
диэлектрическая, 302
магнитная, 303, 364
Время
релаксации, 334, 407
собственное, 289
Ядро
оператора, 65
Якобиан, 114
Закон
Ампера, 309
Био–Савара–Лапласа, 321
Джоуля–Ленца, 323
Фарадея, 314
Кулона, 322
Ома, 323
диффузии, 401
сохранения
энергии, 284
импульса, 283
количества электричества, 298
момента, 284
Законы
Госсена, 390
экономические, 374
Заряд
электрический, 297
полный, 298
свободный, 297
связанный, 301
точечный, 298
Земля, 378
Значение
собственное, 79
среднее, 168, 253