Курс компьютерной геометрии

Д. П. Ильютко, Г. В. Носовский, А. Т. Фоменко.

Содержание

I ОСНОВЫ ДИФФЕРЕНЦИАЛЬНОЙ ГЕОМЕТРИИ 13

1 Введение в дифференциальную геометрию 15

1.1 Криволинейные системы координат. Простейшие примеры . . 15

1.1.1 Мотивировка 15

1.1.2 Декартовы и криволинейные координаты 16

1.1.3 Простейшие примеры криволинейных систем координат 21

1.2 Длина кривой 23

1.2.1 Длина кривой в евклидовых координатах 23

1.2.2 Длина кривой в криволинейных координатах 26

1.2.3 Понятие римановой метрики в области евклидова пространства 29

1.2.4 Индефинитные (знаконеопределенные) метрики 31

1.3 Геометрия на плоскости и на сфере 33

1.4 Псевдосфера и геометрия Лобачевского 39

2 Кривые и поверхности 53

2.1 Теория кривых. Формулы Френе 53

2.1.1 Теория кривых на плоскости. Кривизна 53

2.1.2 Теория пространственных кривых. Кривизна и кручение 58

2.2 Поверхности. Первая и вторая квадратичные формы 62

2.2.1 Определение поверхностей 62

2.2.2 Первая квадратичная форма 64

2.2.3 Вторая квадратичная форма 67

2.2.4 Локальная теория гладких кривых на поверхности. Теорема Менье. Формула Эйлера 71

2.2.5 Гауссова и средняя кривизны в точке на поверхности . 77

3 Общая топология 89

3.1 Простейшие свойства топологических пространств 89

3.1.1 Метрические пространства 89

3.1.2 Топологические пространства 91

3.1.3 Непрерывные отображения 92

3.1.4 Гомотопия и гомотопическая эквивалентность 95

3.1.5 Фактортопологии 96

3.2 Связность. Аксиомы отделимости 97
3.3 Компактные топологические пространства 101
3.4 Функциональная отделимость. Разбиение единицы 103
3.4.1 Функциональная отделимость 104
3.4.2 Разбиение единицы 106

4 Гладкие многообразия (обобщение поверхностей) 109
4.1 Понятие многообразия 111
4.1.1 Основные определения 111
4.1.2 Функции склейки. Определение гладкого многообразия 114
4.1.3 Гладкие отображения многообразий. Диффеоморфизмы 118
4.2 Задание многообразий уравнениями в $\mathbb{R}^n$ 120
4.3 Касательные векторы и касательное пространство 125
4.3.1 Простейшие примеры 125
4.3.2 Общее определение касательного вектора 128
4.3.3 Касательное пространство $T_{P_0}(M)$ 129
4.3.4 Производная функции по направлению 130
4.4 Подмногообразия 133
4.4.1 Дифференциал гладкого отображения 133
4.4.2 Локальные свойства отображений и дифференциал . . . 135
4.4.3 Вложение многообразий в евклидово пространство . . . 138
4.4.4 Римановы метрики на многообразии 139
4.5 Классификация двумерных поверхностей 141
4.5.1 Многообразия с краем 142
4.5.2 Ориентируемые многообразия 143
4.5.3 Классификация двумерных гладких компактных связных многообразий без края 145
4.6 Изометрии 157
4.7 Римановы поверхности алгебраических функций 159

5 Тензорный анализ и риманова геометрия 171
5.1 Понятие гладкого тензорного поля 171
5.2 Простейшие примеры тензорных полей 175
5.2.1 Примеры 175
5.2.2 Алгебраические операции над тензорами 179
5.2.3 Кососимметричные тензоры. Внешние дифференциальные формы 181
5.3 Связности и ковариантное дифференцирование на многообразиях 189
5.3.1 Аффинные связности 189
5.3.2 Римановы связности 195
5.4 Параллельный перенос и геодезические на многообразиях . . . 198
5.4.1 Предварительные замечания 198
5.4.2 Уравнение параллельного переноса на многообразии . . 200
5.4.3 Геодезические 201

5.5 Дифференцирование и интегрирование внешних дифференциальных форм. Формула Стокса 210
5.5.1 Дифференцирование внешних дифференциальных форм 210
5.5.2 Поведение форм при отображениях 215
5.5.3 Интеграл дифференциальной формы по многообразию . 216
5.5.4 Формула Стокса 217

II КОМПЬЮТЕРНАЯ ГЕОМЕТРИЯ 223

6 Гладкие кривые с вычислительной точки зрения 225
6.1 Приблизительная локальная форма кривой, определяемая кривизной и кручением 225
6.2 Формулы для кривизны и кручения кривой относительно произвольного параметра в координатах, задаваемых репером Френе 227
6.3 Восстановление пространственной кривой по ее проекциям на координатные плоскости 229
6.4 Приведение параметрического уравнения кривой к неявному виду 230

7 Сплайны и кривые Безье 235
7.1 Сплайны 235
7.1.1 Примеры сплайнов 235
7.1.2 Построение сплайнов Эрмита 236
7.1.3 Псевдоупругие сплайны Эрмита 238
7.1.4 Случай, когда на концах кривой заданы направления касательных векторов 241
7.1.5 Кубические сплайны. Построение кубического сплайна 242
7.1.6 Сплайн Лагранжа 244
7.1.7 Сплайн Ньютона 245
7.2 Кривые Безье 247
7.2.1 Алгоритм де Кастелье 250
7.2.2 Операторная форма кривой Безье 251
7.2.3 Годографы кривых Безье 253
7.2.4 Деление кривой Безье на две кривые Безье того же порядка в отношении $t^* : (1 - t^*)$ 256
7.2.5 Условия сохранения гладкости сопряжения при делении кривой Безье 257
7.2.6 Увеличение числа опорных точек без изменения формы кривой Безье 258

8 Поверхности Безье 261
8.1 Геометрический смысл поверхности Безье 261
8.2 Формулы вычисления координат точек на поверхности Безье . 262
8.3 Деление поверхности Безье 263

8.4 Геометрические свойства поверхности Безье в угловой точке . 263
8.5 Измельчение сетки при сохранение поверхности Безье 264

9 Проективные (рациональные) кривые Безье 267
9.1 Операция рационального деления отрезка 267
9.2 Свойства рациональных кривых Безье 269
9.3 Деление рациональной кривой Безье 270
9.4 Увеличение числа опорных точек рациональной кривой Безье 273
9.5 Производные на концах рациональной кривой Безье 274
9.6 Рациональные поверхности Безье 275
9.7 Представление кривых второго порядка рациональными кривыми Безье порядка 2 276

10 B-сплайны (бета-сплайны), B-кривые (бета-кривые) и B-поверхности (бета-поверхности) 281
10.1 Постановка задачи 281
10.2 Разделенные разности 282
10.3 Свойства разделенных разностей 287
10.4 Усеченная степенная функция 288
10.5 Рекуррентные соотношения для B-сплайнов 291
10.6 Алгоритм вычисления радиус-вектора B-кривой 294
10.7 Алгоритм вычисления производных B-кривой при условии, что $t_1 = \cdots = t_m$, $t_{n+1} = \cdots = t_{n+m}$ 298
10.8 Алгоритм Де Бура вычисления радиус-вектора B-кривой . . . 301
10.9 Интерполяция с помощью B-кривых 304
10.10 Представление кубического сплайна в виде B-кривой 305
10.11 B-поверхности 307

11 Другие способы представления поверхностей в компьютерной геометрии 309
11.1 Линейчатые поверхности 309
11.2 Секториальные поверхности 309
11.3 Поверхности Кунса 310
11.3.1 Линейные поверхности Кунса 310
11.3.2 Матричный вид уравнения поверхности Кунса 312
11.3.3 Обобщенные поверхности Кунса 312
11.4 Поверхности, построенные по остову из кривых 314
11.4.1 Поверхности Эрмита 314
11.4.2 Применение поверхностей Эрмита: поверхность перехода 316
11.4.3 Поверхности Лагранжа 317
11.4.4 Поверхности Гордона 317
11.4.5 Поверхности, затягивающие сетку кривых заплатами Кунса 319
11.4.6 Поверхности тензорного произведения 320
11.5 Поверхности с треугольной параметрической областью 321
11.5.1 Барицентрические координаты 321

11.5.2 Билинейная треугольная поверхность 323
11.5.3 Треугольная поверхность на трех кривых 323

12 Компьютерная геометрия проективно преобразованных изображений (Г. В. Носовский, Е. С. Скрипка, А. А. Толченников) 325
12.1 Введение 325
12.2 Основные понятия проективной геометрии 326
12.3 Метод распознавания сопряженных точек по разнесениям связанных признаков (Г. В. Носовский) 328
12.4 Устойчивость проективного преобразования к возмущениям конфигурации сопряженных точек (Г. В. Носовский, Е. С. Скрипка) 335
12.5 Вычисление матрицы проективного преобразования и оценка ее устойчивости (Е. С. Скрипка) 339
12.5.1 Прямой алгоритм вычисления проективного преобразования 339
12.5.2 Оценка погрешности матрицы проективного преобразования при возмущенных начальных данных 341
12.6 Математическая модель камеры и пары камер 343
12.6.1 Определение финитной камеры 343
12.6.2 Восстановление характеристик финитной камеры по ее матрице 346
12.6.3 Геометрия двух камер. Фундаментальная матрица . . . 347
12.6.4 Свойства фундаментальной матрицы 350
12.6.5 Восстановление пары камер по фундаментальной матрице 350
12.6.6 Канонические пары камер, порожденные фундаментальной матрицей 352
12.7 ПЛА и НЛА алгоритмы приближенного вычисления проективного преобразования. Оценка их устойчивости (Г. В. Носовский, А. А. Толченников) 354
12.7.1 Вычисление проективного преобразования по точно заданным сопряженным точкам 354
12.7.2 Простой линейный алгоритм (ПЛА) вычисления проективного преобразования по приближенно заданным сопряженным точкам 356
12.7.3 Вычисление вектора, соответствующего минимальному собственному значению матрицы 357
12.7.4 Неинвариантность ПЛА относительно движений плоскости 358
12.7.5 Нормализованный линейный алгоритм (НЛА) 362
12.7.6 Оценка ошибки для алгоритмов ПЛА и НЛА 364

Часть I

ОСНОВЫ ДИФФЕРЕНЦИАЛЬНОЙ ГЕОМЕТРИИ

Глава 1

Введение в дифференциальную геометрию

1.1 Криволинейные системы координат. Простейшие примеры

1.1.1 Мотивировка

Рассмотрим *евклидово пространство* размерности n, которое мы в дальнейшем будем обозначать через $\mathbb{R}^n$. Будем считать, что в нем заданы *декартовы координаты* $x^1, \ldots, x^n$ относительно выбранного и фиксированного ортонормированного базиса $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. Напомним, что с декартовыми координатами тесно связано понятие евклидова скалярного произведения, сопоставляющего каждой паре векторов $\xi, \eta \in \mathbb{R}^n$ вещественное число $\langle \xi, \eta \rangle$, определяемое формулой:

$$\langle \xi, \eta \rangle = \xi^1 \eta^1 + \cdots + \xi^n \eta^n, \qquad \xi = (\xi^1, \ldots, \xi^n), \quad \eta = (\eta^1, \ldots, \eta^n).$$

Однако декартовых координат недостаточно для многих конкретных задач. Весьма часто, например, в физических задачах встречаются траектории, явное выражение которых в декартовых координатах затруднительно. Например, следующее уравнение определяет в декартовых координатах спираль: $\sqrt{x^2 + y^2} - e^{\lambda(\operatorname{arctg} \frac{y}{x})} = 0$ (рис. 1.1). Конечно, эта запись не слишком сложна, но данная кривая запишется значительно проще в другой, так называемой *полярной системе координат* (r, φ), связанной с декартовой системой соотношениями:

$$x = r \cos \varphi, \qquad y = r \sin \varphi$$

(рис. 1.2). В полярных координатах уравнение спирали принимает вид $r = e^{\lambda \varphi}$, что позволяет сразу оценить характер траектории.

Рис. 1.1:

Рис. 1.2:

При решении различных задач механики и физики возникли и другие криволинейные координаты — цилиндрические, сферические и т.д. Изучая подобные способы задания точек пространства наборами вещественных чисел, можно заметить, что в их основе лежит общая идея, которую мы сейчас и опишем.

1.1.2 Декартовы и криволинейные координаты

Рассмотрим произвольную область в $\mathbb{R}^n$. Напомним, что *областью* называется произвольное множество C в евклидовом пространстве, каждая точка P которого входит в C вместе с некоторым шаром достаточно малого радиуса, имеющим точку P своим центром.

Рассмотрим еще один экземпляр евклидова пространства той же размерности, который обозначим через $\mathbb{R}^n_1$. Задать координаты точки P в области C — значит сопоставить ей набор чисел, точку в $\mathbb{R}^n_1$. Ясно, что это соответствие должно удовлетворять некоторым требованиям. В первую очередь, нужно потребовать, чтобы различным точкам отвечали различные наборы координат.

Сопоставляя каждой точке P области C набор из n вещественных чисел, мы получаем n функций $y^1(P), \ldots, y^n(P)$, имеющих областью определения область C; здесь $y^1, \ldots, y^n$ — координаты в пространстве $\mathbb{R}^n_1$. Обычно требуют, чтобы эти функции были непрерывны и даже гладки.

Определение 1.1. *Непрерывной системой координат* в области C евклидова пространства $\mathbb{R}^n$ с декартовыми координатами $x^1, \ldots, x^n$ называется набор функций $y^1(x^1, \ldots, x^n)$, $\ldots$, $y^n(x^1, \ldots, x^n)$, задающих взаимно однозначное и непрерывное в обе стороны отображение области C на некоторую область A в евклидовом пространстве $\mathbb{R}^n_1$ с декартовыми координатами $y^1, \ldots, y^n$. Иными словами, этот набор функций задает *гомеоморфизм* области C на область A (гомеоморфизмом мы будем называть любое взаимно однозначное и непрерывное в обе стороны отображение).

Функции $y^1(P), \ldots, y^n(P)$ будем называть *координатами* точки P относительно координатного отображения $f\colon C \to A$.

Например, в качестве отображения $f\colon C \to A$ можно взять тождественное отображение: $y^1 = x^1, \ldots, y^n = x^n$ (в случае, когда $\mathbb{R}^n$ и $\mathbb{R}^n_1$ совпадают).

Иногда мы будем записывать точку P с ее координатами $y^1(P), \ldots, y^n(P)$ в виде $P(y^1, \ldots, y^n)$.

Среди всех непрерывных координатных отображений выделим такие, которые задают гладкое отображение области C на область A, т.е. все функции $y^1(x^1, \ldots, x^n), \ldots, y^n(x^1, \ldots, x^n)$ являются гладкими функциями. Напомним понятие матрицы Якоби гладкого отображения.

Пусть $f\colon C \to A$ — гладкое отображение, задаваемое функциями $y^1(x^1, \ldots, x^n), \ldots, y^n(x^1, \ldots, x^n)$.

Определение 1.2. *Матрицей Якоби* отображения f называется матрица

$$df = \left(\frac{\partial y}{\partial x}\right) = \begin{pmatrix} \dfrac{\partial y^1}{\partial y^x} & \cdots & \dfrac{\partial y^1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial y^n}{\partial y^x} & \cdots & \dfrac{\partial y^n}{\partial x^n} \end{pmatrix},$$

составленная из частных производных от координатных функций. Ее определитель будем обозначать через $J(f)$ и называть *якобианом* отображения f.

Замечание. Обозначение df для матрицы Якоби не вызовет путаницы с дифференциалом функции f, поскольку дифференциал гладкой функции в данном случае совпадает с матрицей Якоби. Матрица Якоби является переменной матрицей, т.к. зависит от точки P из области C.

Определение 1.3. *Регулярной системой координат* в области C пространства $\mathbb{R}^n$ называется набор гладких функций $y^1(x^1, \ldots, x^n)$, $\ldots$, $y^n(x^1, \ldots, x^n)$, задающих взаимно однозначное отображение области C на некоторую область A в пространстве $\mathbb{R}^n_1$, такой, что якобиан отображения $J(f)$ отличен от нуля во всех точках области C.

Отличие от нуля якобиана отображения f во всех точках области C означает, что отображение f^{-1}, обратное к f, является гладким. (Это следует

из теоремы о системе неявных функций.) Итак, регулярная система координат задается двумя гладкими, взаимно обратными отображениями, устанавливающими гомеоморфизм между областями C и A двух экземпляров евклидова пространства одной и той же размерности.

Можно было бы считать, что с самого начала в области C пространства $\mathbb{R}^n$ были введены декартовы координаты с помощью тождественного отображения C на A при естественном отождествлении обоих пространств $\mathbb{R}^n$ и $\mathbb{R}^n_1$. Тогда введение в области C еще одной системы координат, задаваемой регулярным отображением f, можно рассматривать как замену координат.

Определение 1.4. Регулярную систему координат в области C будем называть также *криволинейными координатами* в C.

Рассмотрим теперь в области C две произвольные криволинейные системы координат: $y^1(P), \ldots, y^n(P)$ и $z^1(P), \ldots, z^n(P)$. Это означает, что заданы два регулярных отображения:

$$f\colon C \to A \subset \mathbb{R}^n_1(y^1, \ldots, y^n); \qquad g\colon C \to B \subset \mathbb{R}^n_2(z^1, \ldots, z^n),$$

т.е. отображения f и g устанавливают взаимно однозначные и гладкие в обе стороны соответствия между областями C, A и C, B соответственно. Иными словами, каждой точке P из области C сопоставлены два набора ее криволинейных координат, $\{y^i(P)\}$ и $\{z^i(P)\}$. Тогда можно сопоставить координатам $y^i(P)$ точки P ее координаты $z^i(P)$, что определяет отображение $\psi_{y,z}\colon A \to B$, т.е. $\psi_{y,z}\colon y^i(P) \to z^i(P)$, $1 \leqslant i \leqslant n$.

Отображение $\psi_{y,z}$ называется *заменой координат* в области C. При этом точка P получает вместо исходных координат $\big(y^1(P), \ldots, y^n(P)\big)$ новые координаты $\big(z^1(P), \ldots, z^n(P)\big)$.

Лемма 1.1. *Отображение $\psi_{y,z}$ является взаимно однозначным, гладким в обе стороны отображением области A на область B с ненулевым якобианом.*

Доказательство. Взаимная однозначность отображения $\psi_{y,z}$ сразу следует из определения 1.3. Гладкость отображения $\psi_{y,z}$ следует из того, что композиция двух гладких отображений является гладким отображением. Осталось проверить, что отображение $\psi_{y,z}$ имеет ненулевой (в каждой точке области A) якобиан.

Заметим, что отображение $\psi_{y,z}$ распадается в композицию двух отображений: $g \circ f^{-1}\colon A \to B$ (рис. 1.3). При этом, матрица Якоби отображения $\psi_{y,z}$ распадается в произведение матриц Якоби отображения f^{-1} и отображения g. В самом деле, так как

$$z^i = z^i(x^1(y^1, \ldots, y^n), \ldots, x^n(y^1, \ldots, y^n)),$$

где функции $\{x^\alpha(y^1, \ldots, y^n)\}$, задают гладкое отображение $f^{-1}\colon A \to C$, то по формуле дифференцирования сложной функции получаем $\dfrac{\partial z^i}{\partial y^j} =$

Рис. 1.3:

$\sum\limits_{k=1}^{n} \dfrac{\partial z^i}{\partial x^k}\dfrac{\partial x^k}{\partial y^j}$, что и означает, что матрица Якоби $d\psi_{y,z}$ распалась в произведение двух матриц, dg и $d(f^{-1})$. Осталось выяснить, как связаны матрицы Якоби df и $d(f^{-1})$. Так как композиция $f^{-1}\circ f$ является тождественным отображением области C на себя, то получаем, что $d(f^{-1}\circ f) = d(f^{-1})\circ df = E$, где E — единичная матрица, т.е. окончательно, $d(f^{-1}) = (df)^{-1}$. Тем самым, $d\psi_{y,z} = (dg)\cdot(df)^{-1}$, т.е. $J(\psi_{y,z}) = \dfrac{J(g)}{J(f)}$, и так как оба якобиана $J(g)$ и $J(f)$ отличны от нуля, то и якобиан $J(\psi_{y,z})$ отличен от нуля. ■

Пусть на области C задан набор гладких функций $\{y^i(P)\}$, $1 \leqslant i \leqslant n$. Как узнать, задает ли он регулярную систему координат?

Лемма 1.2. *Пусть гладкие функции $y^i(P)$, $1 \leqslant i \leqslant n$, определяющие отображение $f\colon C \to A$, $f(P) = \big(y^i(P)\big)$, таковы, что якобиан $J(f)$ отличен от нуля. Тогда для каждой точки P из области C существует такая открытая окрестность, что в ней функции $y^i(P)$ задают регулярные координаты.*

Такие функции, задающие координаты лишь в некоторой окрестности данной точки, называются *локальной системой координат.*

Доказательство. В условии леммы не было предположено, что функции $y^i(P)$ определяют взаимно однозначное отображение области C на область A пространства $\mathbb{R}^n_1$. Однако по теореме о неявной функции из отличия от нуля якобиана вытекает существование (по крайней мере локально) обратного гладкого отображения. Лемма доказана. ■

Отметим, что функции, удовлетворяющие условию леммы 1.2, не обязательно определяют регулярную систему координат сразу во всей области C, т.е. гладкое отображение f^{-1} области A на область C может и не существовать. Действительно, рассмотрим простой пример. Возьмем в качестве области C двумерную плоскость, из которой выколота одна точка — начало координат O, и рассмотрим гладкое отображение

$$f(x) = \big((y^1(x), y^2(x)\big), \quad y^1(x) = (x^1)^2 - (x^2)^2, \quad y^2(x) = 2x^1x^2.$$

Рис. 1.4:

Рис. 1.5:

Если положить $z = x^1 + ix^2$, $w = y^1 + iy^2$ (где i — мнимая единица), то это отображение может быть задано формулой $w = z^2$. Оно переводит комплексное число z в его квадрат. В полярных координатах данной отображение запишется так: $f(r, \varphi) = (r^2, 2\varphi)$; см. рис. 1.4. Матрица Якоби df имеет вид

$$df = \begin{pmatrix} 2x^1 & -2x^2 \\ 2x^2 & 2x^1 \end{pmatrix}, \qquad J(f) = 4\big((x^1)^2 + (x^2)^2\big) > 0.$$

Мы видим, что якобиан положителен во всех точках области C (так как начало координат было выколото). Следовательно, по лемме 1.2 данное отображение устанавливает локальные регулярные координаты в некоторой открытой окрестности каждой точки из области C (т.е. в $\mathbb{R}^2$ с выколотым началом координат). В то же время отображение f не имеет обратного отображения f^{-1}, поскольку f не взаимно однозначно. Каждая точка $w = y^1 + iy^2$, не являющаяся началом координат, всегда имеет ровно два прообраза при отображении f — точки (r, φ) и $(r, \varphi + \pi)$.

Каждая система криволинейных координат в области определяет так называемые координатные линии. А именно: i-я координатная линия задается уравнением

$$\begin{aligned} &y^1(P) = c_1, \quad \ldots, \quad y^{i-1}(P) = c_{i-1}, \quad y^i(P) = t, \\ &y^{i+1}(P) = c_{i+1}, \quad \ldots, \quad y^n(P) = c_n, \end{aligned}$$

где все c_i — постоянные, а t — непрерывный параметр. С изменением t

Рис. 1.6:

точка P прочерчивает гладкую траекторию в области. Таким образом, из каждой точки P области C выходит n гладких траекторий, которые и называются координатными линиями данной системы координат. Для другой точки P будут другие координатные линии, причем они гладко деформируются при изменении точки P. Если координаты — декартовы, то координатные линии являются прямыми. При наглядном изображении криволинейных координат полезно изображать координатные линии. На рис. 1.5 показана гладкая замена координат, "превращающая" жирафа в бегемота.

1.1.3 Простейшие примеры криволинейных систем координат

Для начала отметим, что полярные координаты (r, φ) на плоскости не задают регулярных координат на всей плоскости $\mathbb{R}^2$. В самом деле, переход от полярной системы к декартовой задается функциями: $x^1 = r\cos\varphi$, $x^2 = r\sin\varphi$. Прямым вычислением легко получить, что якобиан замены J равен r. Значит якобиан равен нулю в начале координат и, следовательно, полярные координаты не являются регулярными на всей плоскости. Более того, они не задают взаимно однозначного отображения всей плоскости на себя, так как точки вида (r, φ) и $(r, \varphi + 2\pi)$ переходят в одну и ту же точку.

Выделим область C, где полярные координаты регулярны.

Рассмотрим плоскость $\mathbb{R}^2(x^1, x^2)$. В качестве области C возьмем всю плоскость, за исключением луча $x^1 \geqslant 0$, $x^2 = 0$. Тогда в качестве области A в плоскости $\mathbb{R}^2(y^1, y^2)$, где $y^1 = r$, $y^2 = \varphi$, следует взять бесконечную полуполосу, определяемую неравенствами $0 < \varphi < 2\pi$, $0 < r < +\infty$. Определим отображение $f\colon C \to A$ с помощью обратного к нему отображения $f^{-1}\colon (r, \varphi) \to (x^1, x^2)$, заданного формулами $x^1 = r\cos\varphi$, $x^2 = r\sin\varphi$. На рис. 1.6 показано, что происходит с координатными линиями при отображении f^{-1}. Прямоугольная сетка превращается в полярную сетку. Взаимная однозначность и регулярность отображения очевидны.

Рассмотрим теперь трехмерное евклидово пространство и изучим *цилиндрическую систему координат*. Формулы замены следующие: $x^1 = r\cos\varphi$, $x^2 = r\sin\varphi$, $x^3 = z$. Рассмотрим $\mathbb{R}^3(y^1, y^2, y^3)$, где $y^1 = r$, $y^2 = \varphi$,

Рис. 1.7:

$y^3 = z$, и в качестве A возьмем область $0 < r$, $0 < \varphi < 2\pi$, $-\infty < z < +\infty$. Приведенные выше формулы определяют гладкое отображение обратное к $f\colon C \to A$, где область $C \subset \mathbb{R}^3(x^1, x^2, x^3)$ получается из $\mathbb{R}^3(x^1, x^2, x^3)$ выбрасыванием полуплоскости $x^2 = 0$, $x^1 \geqslant 0$ (рис. 1.7). Матрица Якоби отображения f^{-1} имеет следующий вид:

$$\begin{pmatrix} \cos\varphi & -r\sin\varphi & 0 \\ \sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Якобиан этого отображения равен r. Таким образом, в области C цилиндрические координаты регулярны. Дело в том, что якобиан равен нулю только в точках оси z, но полуплоскость, проходящая через эту ось, была исключена.

Теперь рассмотрим n-мерное евклидово пространство и введем в нем *сферическую систему координат*. Формулы замены таковы:

$$f_n{}^{-1}\colon A(r, \theta_1, \theta_2, \ldots, \theta_{n-1}) \to C(x^1, x^2, \ldots, x^n);$$

$$\begin{cases} x^1 = r\cos\theta_1, \\ x^2 = r\sin\theta_1\cos\theta_2, \\ x^3 = r\sin\theta_1\sin\theta_2\cos\theta_3, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x^{n-1} = r\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-2}\cos\theta_{n-1}, \\ x^n = r\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-2}\sin\theta_{n-1}. \end{cases}$$

Предоставляем читателю вычисление матрицы Якоби и якобиана.

Для трехмерного пространства сферические координаты обычно обозначаются через (r, θ, φ). Тогда формулы замены (после соответствующей перенумерации координат x^i) приобретают вид:

$$x^1 = r\sin\theta\cos\varphi, \quad x^2 = r\sin\theta\sin\varphi, \quad x^3 = r\cos\theta;$$

$$0 < \theta < \pi, \quad 0 < \varphi < 2\pi, \quad r \geqslant 0.$$

В этих координатах $J = r^2\sin\theta$.

Рис. 1.8:

Области C и A показаны на рис. 1.8. Якобиан равен нулю только в точках оси x^3. Полуплоскость $x^2 = 0$, $x^1 \geqslant 0$ удалена для обеспечения регулярности координатной системы. При фиксированном r координатные линии параметров θ, φ показаны на рис. 1.9. Они задают координатную сеть на глобусе. Матрица Якоби отображения f_3^{-1} в трехмерном случае имеет вид

$$\begin{pmatrix} \sin\theta\cos\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix}.$$

Задачи

Рис. 1.9:

1. Доказать, что набор функций $u = x + \sin y$, $v = y - \dfrac{1}{2}\sin x$ задает на плоскости регулярные координаты.

2. Показать, что на окружности S^1 нельзя задать единой системы координат, регулярной на всей окружности.

3. Записать уравнение Лапласа $\Delta u = \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} = 0$ в полярных координатах.

1.2 Длина кривой

1.2.1 Длина кривой в евклидовых координатах

Рассмотрим пространство $\mathbb{R}^n$ и зададим в нем евклидово скалярное произведение $\langle \xi, \eta \rangle = \sum \xi^i \eta^i$. Тогда для каждого вектора ξ можно вычислить его *модуль* или *длину* по формуле $|\xi| = \sqrt{\langle \xi, \xi \rangle}$. Если же мы хотим найти расстояние между двумя произвольными точками $\xi, \eta \in \mathbb{R}^n$, то следует вычислить длину вектора $\xi - \eta$. Как известно из аналитической геометрии, угол φ между двумя векторами можно выразить через скалярное произведение по формуле $\cos\varphi = \dfrac{\langle \xi, \eta \rangle}{|\xi||\eta|}$. Мы видим, что такие важные метрические

понятия, как длины векторов и углы между ними, тесно связаны со скалярным произведением. При построении других важных понятий геометрии мы будем опираться на различные формы скалярных произведений.

Дадим теперь определение длины гладкой кривой. Зададим кривую $\gamma(t)$ в параметрическом виде, т.е. будем считать, что *гладкая кривая* задана в евклидовом пространстве n гладкими функциями $x^1(t), \ldots, x^n(t)$, где параметр пробегает либо всю вещественную ось, либо конечный отрезок $[a, b]$. Здесь $x^1, \ldots, x^n$ — это декартовы координаты в $\mathbb{R}^n$.

Определение 1.5. *Длиной кривой* от точки $\gamma(a)$ до точки $\gamma(b)$ (или от значения параметра $t = a$ до $t = b$) называется число $l(\gamma)_a^b = \int\limits_a^b \sqrt{\langle \dot\gamma(t), \dot\gamma(t)\rangle} dt$, где $\dot\gamma(t)$ — вектор с координатами $\left(\dfrac{dx^1(t)}{dt}, \ldots, \dfrac{dx^n(t)}{dt}\right)$, называемый *вектором скорости* кривой $\gamma(t)$ или *касательным вектором.*

В явном виде формула имеет вид

$$l(\gamma)_a^b = \int\limits_a^b \sqrt{\sum_{i=1}^n \left(\frac{dx^i(t)}{dt}\right)^2} dt.$$

Лемма 1.3. *Пусть на гладкой кривой $\gamma(t)$ заданы две точки $\gamma(a)$ и $\gamma(b)$. Пусть $t = t(\tau)$ — произвольная гладкая замена параметра t на новый параметр τ, причем $\dfrac{dt}{d\tau} > 0$. Тогда длина кривой $l(\gamma(t))_a^b$ не изменится, т.е. имеет место равенство $l(\gamma(t))_a^b = l(\gamma(\tau))_\alpha^\beta$, где $a = t(\alpha)$, $b = t(\beta)$.*

Доказательство. Обозначим вектор $\left(\dfrac{dx^1(t)}{dt}, \ldots, \dfrac{dx^n(t)}{dt}\right)$ через $\dot\gamma_t$. Прямое вычисление дает

$$l(\gamma(t))_a^b = \int\limits_\alpha^\beta \sqrt{\langle \dot\gamma_\tau, \dot\gamma_\tau\rangle \left(\frac{d\tau}{dt}\right)^2} \frac{dt}{d\tau} d\tau = \int\limits_\alpha^\beta \sqrt{\langle \dot\gamma_\tau, \dot\gamma_\tau\rangle} d\tau. \quad \blacksquare$$

Пусть даны две гладкие кривые, $\gamma_1(t)$ и $\gamma_2(\tau)$, пересекающиеся в точке P евклидова пространства, т.е. существуют такие значения параметров $t = a$ и $\tau = b$, что $P = \gamma_1(a) = \gamma_2(b)$. Определим угол между кривыми в точке их пересечения.

Определение 1.6. *Углом* между пересекающимися гладкими траекториями $\gamma_1(t)$ и $\gamma_2(\tau)$ в точке их пересечения называется угол φ, определяемый равенством:

$$\cos\varphi = \frac{\langle \dot\gamma_1(a), \dot\gamma_2(b)\rangle}{|\dot\gamma_1(a)|\,|\dot\gamma_2(b)|},$$

если оба вектора скорости, $\dot\gamma_1(a)$ и $\dot\gamma_2(b)$, отличны от нуля.

Рис. 1.10:

Рис. 1.11:

Замечание. Строго говоря, это равенство определяет не один угол, а два угла, дающие в сумме угол, равный π. Однако поскольку кривые занумерованы, то возникает понятие ориентированного угла, который определяется приведенной выше формулой однозначно. Подчеркнем также существенность требования, что оба вектора скорости отличны от нуля в точке пересечения. Дело в том, что в точках, где вектор скорости обращается в нуль, кривая может претерпевать излом, скачком меняя направление движения, см. рис. 1.10. Наличие на гладкой кривой точки излома не противоречит гладкости кривой (см. определение гладкой кривой).

Пример гладкой кривой, имеющей излом, см. на рис. 1.11. Здесь угол излома равен $\pi/2$. Легко построить гладкую кривую, имеющую в особой точке угол излома, равный π (рис. 1.12).

Упражнение. Напишите параметрическое уравнение кривой, изображенной на рис. 1.12. Можно ли задать эту кривую аналитическими функциями $x(t)$, $y(t)$?

Предоставляем читателю убедиться, что длина отрезка и длина окружности, подсчитанные по указанной общей формуле, совпадают с известными

Рис. 1.12:

из элементарной геометрии величинами.

1.2.2 Длина кривой в криволинейных координатах

Рассмотрим теперь криволинейные координаты в области C евклидова пространства, и пусть $\gamma(t)$ — гладкая кривая в этой области. Как запишется длина кривой в криволинейных координатах? Обозначим криволинейные координаты через $z^1, \ldots, z^n$, т.е. $x^i = x^i(z)$. По правилу дифференцирования сложной функции имеем

$$\frac{dx^i(t)}{dt} = \sum_k \frac{\partial x^i}{\partial z^k} \frac{dz^k}{dt},$$

т.е.

$$l(\gamma)_a^b = \int\limits_a^b \sqrt{\sum_{i=1}^n \left(\frac{dx^i(z(t))}{dt}\right)^2}\, dt = \int\limits_a^b \sqrt{\sum_{i=1}^n \sum_{m,p} \frac{\partial x^i}{\partial z^m} \frac{\partial x^i}{\partial z^p} \frac{dz^m}{dt} \frac{dz^p}{dt}}\, dt =$$

$$= \int\limits_a^b \sqrt{\sum_{m,p} g_{mp}(z) \frac{dz^m}{dt} \frac{dz^p}{dt}}\, dt,$$

где функции $g_{mp}(z)$ имеют следующий вид: $g_{mp}(z) = \sum\limits_{i=1}^n \frac{\partial x^i}{\partial z^m} \frac{\partial x^i}{\partial z^p}$. Очевидно, они симметричны по индексам m и p, т.е. $g_{mp} = g_{pm}$. Следовательно, эти функции можно организовать в симметричную матрицу $G = (g_{mp})$.

Поскольку $d\psi_{z,x} = \left(\frac{\partial x}{\partial z}\right)$, то матрица G является произведением двух матриц: $G = A^\top A$, где $A = d\psi_{z,x}$. Каков закон изменения матрицы $G(z)$ при замене координат?

Заменим переменные, перейдя от $\{z^i\}$ к переменным $\{y^k\}$, т.е. рассмотрим регулярную замену вида $z^i = z^i(y^1, \ldots, y^n)$, $1 \leqslant i \leqslant n$. При этом мы считаем $\{y^k\}$ новыми криволинейными координатами в области C.

Тогда коэффициенты $g_{mp}(z)$ меняются по закону

$$g_{kl}(y) = \sum_{i=1}^n \frac{\partial x^i(y)}{\partial y^k} \frac{\partial x^i(y)}{\partial y^l} = \sum_{m,p} \frac{\partial z^m}{\partial y^k} \left(\sum_{i=1}^n \frac{\partial x^i}{\partial z^m} \frac{\partial x^i}{\partial z^p}\right) \frac{\partial z^p}{\partial y^l} =$$

$$= \sum_{m,p} \frac{\partial z^m}{\partial y^k} g_{mp}(z) \frac{\partial z^p}{\partial y^l},$$

т.е. $G(y) = (d\psi_{y,z})^\top G(z) d\psi_{y,z}$.

Замечание. Для упрощения обозначений в выражениях типа $\sum\limits_i a^i b^i$, где суммирование ведется по повторяющимся индексам, мы иногда будем опускать знак $\sum$ и писать просто $a^i b^i$. Так, например, приведенная выше формула для $g_{kl}(y)$ запишется с учетом этого соглашения в виде $g_{kl}(y) =$

Рис. 1.13:

$\dfrac{\partial x^i(y)}{\partial y^k}\dfrac{\partial x^i(y)}{\partial y^l}$. Иногда мы будем обозначать новые координаты через $z^{i'}$, обозначая компоненты новой координатной системы той же буквой, что и старой, ставя штрих у индексов, записанных в новой системе координат. В этом случае штрих у индексов будет подразумевать штрих у компоненты, к которой этот индекс относится.

Функции $g_{mp}(z)$ имеют прозрачный смысл. Рассмотрим точку P в области C и координатные линии криволинейных координат $z^1, \ldots, z^n$, проходящие через точку P. Каждую линию можно задать параметрически:

$$z^1 = c_1, \quad \ldots, \quad z^{i-1} = c_{i-1}, \quad z^i = t, \quad z^{i+1} = c_{i+1}, \quad \ldots, \quad z^n = c_n,$$

где c_α, $1 \leqslant \alpha \leqslant n$, $\alpha \neq i$, — такие постоянные, что точка P имеет координаты $z^\alpha = c_\alpha$, $1 \leqslant \alpha \leqslant n$. Обозначим m-ю координатную линию через $\gamma_m(t)$. Тогда в системе координат x она запишется так: $x^i(c_1, \ldots, c_{m-1}, z^m = t, c_{m+1}, \ldots, c_n)$, $1 \leqslant i \leqslant n$.

Вектор скорости этой кривой в точке P имеет координаты $\mathbf{e}_m = \left(\dfrac{\partial x^1}{\partial z^m}, \ldots, \dfrac{\partial x^n}{\partial z^m}\right)$. Так как $g_{mp}(z) = \dfrac{\partial x^i}{\partial z^m}\dfrac{\partial x^i}{\partial z^p}$, то $g_{mp}(z) = \langle \mathbf{e}_m, \mathbf{e}_p \rangle$, т.е. функции g_{mp} являются евклидовыми скалярными произведениями векторов, касательных к соответствующим координатным линиям (рис. 1.13).

Мы видим, что матрица $G(z)$ при заменах координат преобразуется как матрица квадратичной формы в $\mathbb{R}^n$. Если исходные координаты были декартовыми, то матрица G является единичной и, следовательно, в других, криволинейных, координатах ее можно записать так: $G(z) = A^\top E A$, где $A = d\psi_{z,x}$ и $\{x^i\}$ — декартовы координаты.

Рассмотрим простейшие примеры формул для длины кривой в различных криволинейных координатах. Попутно мы вычислим и матрицу G для этих координат.

1) Полярные координаты на плоскости. Матрица $G(x)$ в декартовых координатах имеет вид

$$G(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \text{т.е.} \qquad g_{ij} = \delta_{ij} = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases}$$

Матрица Якоби была уже вычислена нами выше. Отсюда получаем

$$G(r,\varphi) = (d\psi)^{\top} d\psi = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.$$

Следовательно, в полярных координатах длина кривой $\gamma(t) = \big(r(t), \varphi(t)\big)$, выражается по формуле

$$l(\gamma)_a^b = \int\limits_a^b \sqrt{\left(\frac{dr}{dt}\right)^2 + r^2 \left(\frac{d\varphi}{dt}\right)^2}\, dt.$$

2) Цилиндрические координаты в трехмерном евклидовом пространстве $\mathbb{R}^3(r,\varphi,z)$. Матрица $G(x)$ в декартовых координатах имеет вид $G(x) = E$. Матрица Якоби была уже вычислена нами выше. Отсюда получаем

$$G(r,\varphi,z) = (d\psi)^{\top} d\psi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Следовательно, в цилиндрических координатах длина кривой $\gamma(t) = \big(r(t), \varphi(t), z(t)\big)$, выражается по формуле

$$l(\gamma)_a^b = \int\limits_a^b \sqrt{\left(\frac{dr}{dt}\right)^2 + r^2 \left(\frac{d\varphi}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt.$$

Пример 1. Длина винтовой линии на прямом круговом цилиндре, заданной параметрически уравнениями: $r(t) = R$, $\varphi(t) = \omega t$, $z = qt$ вычисляется так:

$$l(\gamma)_a^b = \int\limits_a^b \sqrt{R^2\omega^2 + q^2}\, dt = \sqrt{R^2\omega^2 + q^2}(b-a).$$

3) Сферические координаты в трехмерном евклидовом пространстве $\mathbb{R}^3(r,\theta,\varphi)$. Матрица $G(x)$ в декартовых координатах имеет вид $G(x) = E$. Матрица Якоби была уже вычислена нами выше. Отсюда получаем

$$G(r,\theta,\varphi) = (d\psi)^{\top} d\psi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix}.$$

Следовательно, в сферических координатах длина кривой $\gamma(t) = \big(r(t), \theta(t), \varphi(t)\big)$, выражается формулой

$$l(\gamma)_a^b = \int\limits_a^b \sqrt{\left(\frac{dr}{dt}\right)^2 + r^2 \left(\frac{d\theta}{dt}\right)^2 + r^2\sin^2\theta \left(\frac{d\varphi}{dt}\right)^2}\, dt.$$

Иногда удобно выписать явную формулу для дифференциала длины дуги dl. В частности, в разобранных примерах эти дифференциалы имеют вид: $(dl)^2 = (dr)^2 + r^2(d\varphi)^2$ в полярных координатах; $(dl)^2 = (dr)^2 + r^2(d\varphi)^2 + (dz)^2$ в цилиндрических координатах; $(dl)^2 = (dr)^2 + r^2(d\theta)^2 + r^2\sin^2\theta(d\varphi)^2$ в сферических координатах.

1.2.3 Понятие римановой метрики в области евклидова пространства

Мы сопоставили каждой криволинейной системе координат z в области C гладкую матричную функцию $G(z)$, преобразующуюся при замене координат, как квадратичная форма.

Определение 1.7. Мы скажем, что задана *риманова метрика* в области C евклидова пространства, если в каждой регулярной системе координат $z^1, \ldots, z^n$ определен набор гладких функций $g_{mp}(z^1, \ldots, z^n)$ таких, что:

1) $g_{mp}(z) = g_{pm}(z)$, т.е. матрица $G(z)$ симметрична;

2) матрица $G(z) = (g_{mp})$ невырождена и положительно определена;

3) при замене координат $z \to y$ матрица $G(z)$ преобразуется по правилу $G(y) = (d\psi)^{\top} G(z) d\psi$. Здесь через $d\psi$ обозначена матрица Якоби замены координат $d\psi_{y,z}$.

Определение 1.8. Если в области C задана риманова метрика $G(z) = (g_{ij})$, и в системе координат (z^i) задана гладкая кривая $\gamma(t) = \big(z^i(t)\big)$, то ее *длиной* от точки $\gamma(a)$ до точки $\gamma(b)$ называется число

$$l = \int\limits_a^b \sqrt{g_{ij}(z)\left(\frac{dz^i(t)}{dt}\right)^2}\, dt.$$

Определение римановой метрики можно сформулировать в более инвариантных терминах. А именно, риманова метрика позволяет определить квадратичную форму $\langle\, ,\, \rangle_g$, определенную в каждой точке области на множестве всех векторов, касательных к гладким траекториям, проходящим через эту точку. Если $\gamma_1(0) = P = \gamma_2(0)$, то для векторов $\xi = \dot\gamma_1(0)$ и $\eta = \dot\gamma_2(0)$, где $\xi = (\xi^1, \ldots, \xi^n)$, $\eta = (\eta^1, \ldots, \eta^n)$, положим: $\langle \xi, \eta \rangle_g = g_{ij}\xi^i\eta^j$.

Лемма 1.4. *Отображение $\xi, \eta \mapsto \langle \xi, \eta \rangle_g$ задает невырожденную, положительно определенную квадратичную форму, гладко зависящую от точки.*

Доказательство. Симметрия и билинейность построенного отображения следует из определения 1.7. Убедимся в том, что это соответствие определяет билинейную форму. Сделаем регулярную замену координат $z^i \to z^{i'}$. Тогда

$$\gamma_1 = \big(z_1^1(t), \ldots, z_1^n(t)\big), \quad \gamma_2 = \big(z_2^1(t), \ldots, z_2^n(t)\big),$$

$$\xi^{i'} = \frac{dz_1^{i'}}{dt} = \frac{\partial z^{i'}}{\partial z^i}\xi^i, \qquad \eta^{i'} = \frac{\partial z^{i'}}{\partial z^i}\eta^i, \qquad g_{i'j'} = \frac{\partial z^i}{\partial z^{i'}}\frac{\partial z^j}{\partial z^{j'}} g_{ij}.$$

Отсюда

$$\langle \xi', \eta' \rangle_{g'} = g_{i'j'} \xi^{i'} \eta^{j'} =$$

$$= \left(\frac{\partial z^i}{\partial z^{i'}} \frac{\partial z^{i'}}{\partial z^k} \right) \left(\frac{\partial z^j}{\partial z^{j'}} \frac{\partial z^{j'}}{\partial z^p} \right) g_{ij} \xi^k \eta^p = = \delta^i_k \delta^j_p g_{ij} \xi^k \eta^p = \langle \xi, \eta \rangle_g,$$

т.е. $\langle \xi, \eta \rangle_g$ действительно является билинейной формой. При доказательстве мы воспользовались тем, что $\dfrac{\partial z^i}{\partial z^{i'}} \dfrac{\partial z^{i'}}{\partial z^k} = \delta^i_k$. ■

Итак, можно говорить, что в области C евклидова пространства задана риманова метрика, если в каждой точке области задана билинейная форма (скалярное произведение), определенная на векторах, касательных к гладким кривым, проходящим через эту точку, причем форма невырожденная и положительно определенная.

Из леммы 1.4 вытекает эквивалентность этого определения и определения 1.7. В частности, из леммы 1.4 вытекает независимость длины гладкой кривой от выбора криволинейных координат.

Если в точке P пересекаются две гладкие кривые γ_1 и γ_2 (такие, что $\gamma_1(0) = \gamma_2(0) = P$, $\dot\gamma_1(0) \neq 0$, $\dot\gamma_2(0) \neq 0$), то углом между ними, в данной римановой метрике, называется такое число φ, что

$$\cos\varphi = \frac{\langle \dot\gamma_1, \dot\gamma_2 \rangle}{|\dot\gamma_1|\,|\dot\gamma_2|}.$$

Существуют ли римановы метрики? Да, и первый пример мы уже привели. В самом деле, если в области C в декартовых координатах, $x^1, \ldots, x^n$, задать матрицу $G(x) = (\delta_{ij})$, то тогда в любых криволинейных координатах z, получающихся из декартовых регулярной заменой, можно по определению положить $G(z) = (d\psi)^\top G(x) d\psi = (d\psi)^\top d\psi$, где $d\psi$ — матрица Якоби данной замены. Тем самым мы определили риманову метрику, заданную в области евклидова пространства. Эта метрика является евклидовой, и квадрат дифференциала дуги гладкой кривой в декартовых координатах записывается так: $(dl)^2 = \sum\limits_{i=1}^n (dx^i)^2$. Однако если задана произвольная риманова метрика, то не следует думать, что путем подходящей замены координат в области C эту метрику можно привести к виду $\sum\limits_{i=1}^n (dx^i)^2$. В общем случае это неверно.

Определение 1.9. Риманова метрика G, заданная в области C, называется *евклидовой*, если существуют координаты y (вообще говоря, криволинейные), в которых матрица $G(y)$ становится единичной в области C.

Если задана евклидова метрика, то можно описать все системы координат, в которых она также имеет евклидов вид, то есть задается единичной матрицей. Такое описание мы дадим позже. Здесь отметим только, что все такие системы координат получаются из одной путем вращений, сдвигов и отражений в евклидовом пространстве.

Существуют неевклидовы метрики, т.е. такие, которые не могут быть приведены к виду $\sum\limits_{i=1}^{n}(dx^i)^2$ никакой регулярной заменой координат. Пока мы не можем предъявить такую метрику, но вскоре мы это сделаем.

Евклидова метрика, записанная в криволинейных координатах z, теряет свой простой "евклидов" вид и задается матрицей $G(z)$, распознать в которой евклидову метрику иногда не просто. Посмотрим, как записывается евклидова метрика в простейших криволинейных координатах, указанных выше. Часто бывает удобно записывать не матрицу $G(z)$, а квадрат дифференциала длины дуги гладкой кривой: $(dl)^2 = g_{ij}(z)dz^i dz^j$.

1) В полярных координатах на плоскости: $dl^2 = dr^2 + r^2 d\varphi^2$.

2) В цилиндрических координатах в трехмерном пространстве: $dl^2 = dr^2 + r^2 d\varphi^2 + dz^2$.

3) В сферических координатах в трехмерном пространстве: $dl^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2)$.

1.2.4 Индефинитные (знаконеопределенные) метрики

До сих пор мы имели дело только с положительно определенными метриками, которые мы называли римановыми. Однако в приложениях часто встречаются так называемые индефинитные метрики.

Определение 1.10. Мы скажем, что задана *индефинитная метрика* в области C евклидова пространства, если в каждой регулярной системе координат $z^1, \ldots, z^n$ определен набор гладких функций $\{g_{mp}(z^1, \ldots, z^n)\}$, удовлетворяющих всем требованиям, наложенным на риманову метрику (см. определение 1.7), кроме требования положительной определенности, т.е. соответствующая квадратичная форма является индефинитной.

Простейший пример: псевдоевклидовы метрики индекса s в псевдоевклидовых пространствах $\mathbb{R}^n_s$. Для построения этой метрики возьмем обычное евклидово пространство $\mathbb{R}^n$ с декартовыми координатами $x^1, \ldots, x^n$, и зададим в каждой точке P следующую билинейную форму (с постоянными, т.е. не зависящими от точки коэффициентами): $\langle \xi, \eta \rangle_s = -\sum\limits_{i=1}^{s} \xi^i \eta^i + \sum\limits_{j=s+1}^{n} \xi^j \eta^j$. Тогда для гладкой кривой $\gamma(t) = \big(x^i(t)\big)$ длина дуги выражается по формуле

$$l = \int\limits_a^b \sqrt{-\sum_{i=1}^{s}\left(\frac{dx^i}{dt}\right)^2 + \sum_{j=s+1}^{n}\left(\frac{dx^j}{dt}\right)^2}\, dt.$$

При $n = 4$ псевдоевклидово пространство индекса 1 называется иногда *пространством Минковского* (в специальной теории относительности). Мы также будем рассматривать пространства $\mathbb{R}^3_1$ и $\mathbb{R}^2_1$. Отметим, что псевдоевклидово пространство индекса 0 совпадает с обычным евклидовым пространством.

Рис. 1.14:

Как и в обычном евклидовом пространстве, длина вектора ξ в пространстве $\mathbb{R}^n_s$ определяется по формуле $|\xi|_s = \sqrt{\langle \xi, \xi \rangle_s}$, однако в $\mathbb{R}^n_s$, в отличие от $\mathbb{R}^n$, длины векторов могут быть нулевыми и мнимыми. В самом деле, так как форма $\langle\ ,\ \rangle_s$ не является положительно определенной, то множество всех векторов разбивается на следующие три непересекающихся класса: $\langle \xi, \xi \rangle_s < 0$ (*времениподобные* векторы), $\langle \xi, \xi \rangle_s = 0$ (*световые* или *изотропные* векторы), $\langle \xi, \xi \rangle_s > 0$ (*пространственноподобные* векторы). Таким образом, имеются векторы нулевой, вещественной и чисто мнимой длины. Времениподобные векторы имеют чисто мнимую длину, световые (изотропные) — нулевую, а пространственноподобные — вещественную длины. Изотропные векторы, выходящие, например, из начала координат, заполняют конус

$$-\sum_{i=1}^{s} (x^i)^2 + \sum_{j=s+1}^{n} (x^j)^2 = 0$$

с вершиной в начале координат. Времениподобные векторы расположены внутри конуса, т.е. в его полости, определяемой координатной плоскостью $(x^1, \ldots, x^s)$, а пространственноподобные векторы расположены вне светового конуса (рис. 1.14).

Замечание. В пространстве Минковского $\mathbb{R}^4_1$ (в специальной теории относительности) изотропный конус состоит из световых векторов и называется световым конусом. Луч света, выпущенный из начала координат, будет распространяться по одной из образующих этого конуса.

Мы определим *длину* гладкой кривой в псевдоевклидовом пространстве $\mathbb{R}^n_s$ индекса s, положив $l_s = \int\limits_a^b \sqrt{\langle \dot\gamma, \dot\gamma \rangle_s}\, dt$. В отличие от евклидова случая, некоторые кривые могут иметь нулевую, чисто мнимую или комплексную длину.

Пространство Минковского $\mathbb{R}^4_1$ было введено для удобной записи некоторых эффектов специальной теории относительности. Отнесем пространство $\mathbb{R}^4_1$ к координатам $x^0 = ct$, $x^1 = x$, $x^2 = y$, $x^3 = z$, т.е. $dl^2 = -c^2\, dt^2 + dx^2 + dy^2 + dz^2$. Здесь t — время, c — скорость света. Рассмотрим так называемую *мировую линию* какой-нибудь материальной частицы

$\gamma(\tau)$. Эта линия является гладкой траекторией в $\mathbb{R}^4_1$. Если координаты x, y, z трактовать как пространственные, то тогда движение частицы по траектории $\gamma(\tau)$ можно интерпретировать как эволюцию во времени материальной точки, перемещающейся в трехмерном евклидовом пространстве. Пусть $\dot\gamma$ — как и раньше, касательный вектор к траектории $\gamma(\tau)$. В специальной теории относительности принимается постулат, что никакой сигнал не может распространяться со скоростью, превышающей скорость света c. Следовательно, $c\,dt > \sqrt{dx^2+dy^2+dz^2}$, где $(c\,dt, dx, dy, dz)$ — координаты бесконечно малого вектора смещения вдоль траектории γ в направлении касательного вектора $\dot\gamma$.

Рис. 1.15:

Отсюда мы получаем, что вдоль мировой линии материальной частицы всегда выполнено соотношение $-c^2\,dt^2 + dx^2 + dy^2 + dz^2 < 0$, т.е. $\langle\dot\gamma(\tau), \dot\gamma(\tau)\rangle_1 < 0$. Это означает, что каждый касательный вектор к мировой линии является времениподобным. Отсюда следует, что мировая линия всегда имеет чисто мнимую длину. В частности, мировая линия целиком распространяется строго внутри светового (изотропного) конуса, имеющего своей осью ось t. Это условие должно выполняться в каждой точке мировой линии (рис. 1.15).

Задачи

1. Проверить, что длину кривой можно вычислять как предел длин ломаных, которые состоят из отрезков, соединяющих последовательно конечное число точек на кривой, когда максимальная длина отрезков стремится к нулю.

2. Доказать, что в евклидовом пространстве прямолинейный отрезок имеет минимальную длину среди длин кривых, соединяющих две точки.

1.3 Геометрия на плоскости и на сфере

Рассмотрим двумерную плоскость, отнесенную к декартовым координатам x, y и снабженную евклидовой метрикой $dl^2 = dx^2 + dy^2$. Иногда для обозначения бесконечно малого элемента длины дуги кривой используют наряду с dl^2 также символ ds^2. Мы будем пользоваться обоими этими обозначениями. Окружности с центром в точке O окажутся в полярных координатах координатными линиями вида $r(t) = \text{const}$. Здесь бесконечно малый элемент дуги окружности равен $r\,d\varphi$.

Рассмотрим теперь стандартное вложение двумерной сферы в трехмерное пространство, отнесенное к декартовым координатам x, y, z в виде множества точек, являющихся концами векторов длины R, выходящих из точки O. Сделаем следующее общее замечание. Пусть на сфере S^2 задана гладкая

кривая $\gamma(t)$, и нужно найти ее длину. Для этого можно рассмотреть объемлющую трехмерную евклидову метрику $ds^2 = dx^2 + dy^2 + dz^2$, записать кривую в параметрическом виде $\gamma(t) = \big(x(t), y(t), z(t)\big)$ и вычислить длину по правилу, изложенному ранее. Точно так же можно поступить, если нужно измерить угол между двумя лежащими на сфере и пересекающимися кривыми.

Введем в $\mathbb{R}^3$ сферические координаты r, θ, φ. Тогда сфера радиуса R задается одним уравнением $r = R$.

Вычислим скалярное произведение двух векторов, касательных к кривым, целиком лежащим на сфере S^2. Пусть $\gamma_1(t) = \big(R, \theta_1(t), \varphi_1(t)\big)$, $\gamma_2(t) = \big(R, \theta_2(t), \varphi_2(t)\big)$. Тогда $\dot\gamma_1(t) = (0, \dot\theta_1, \dot\varphi_1)$, $\dot\gamma_2(t) = (0, \dot\theta_2, \dot\varphi_2)$, т.е.

$$\langle \dot\gamma_1, \dot\gamma_2 \rangle = R^2\big(\dot\theta_1\dot\theta_2 + \sin^2\theta(t)\dot\varphi_1\dot\varphi_2\big).$$

Формула записана в точке пересечения двух кривых.

Видно, что получившееся скалярное произведение совпадает со скалярным произведением двух векторов, $(\dot\theta_1, \dot\varphi_1)$ и $(\dot\theta_2, \dot\varphi_2)$, относительно новой билинейной формы $R^2(\dot\theta_1\dot\theta_2 + \sin^2\theta(t)\dot\varphi_1\dot\varphi_2)$. Она определяет квадратичную форму $R^2(d\theta^2 + \sin^2\theta\, d\varphi^2)$, получающуюся из соответствующей квадратичной формы в евклидовом пространстве $dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2)$ подстановкой в нее вместо переменных r, θ, φ новых функций от θ, φ вида $r = R$, $\theta = \theta$, $\varphi = \varphi$. Будем говорить, что полученная в результате риманова метрика на сфере S^2

$$R^2(d\theta^2 + \sin^2\theta\, d\varphi^2)$$

является *индуцированной* объемлющей евклидовой метрикой трехмерного пространства.

Положение точки на сфере можно задавать двумя параметрами, θ и φ, (широта и долгота). Поэтому радиус-вектор точки на сфере записывается так:

$$x = R\sin\theta\cos\varphi, \quad y = R\sin\theta\sin\varphi, \quad z = R\cos\theta.$$

Подставляя эти три функции от двух параметров в квадрат дифференциала длины дуги в трехмерном пространстве $dx^2 + dy^2 + dz^2$, мы получаем

$$R^2(d\theta^2 + \sin^2\theta\, d\varphi^2).$$

Этот пример будет нами в дальнейшем сильно обобщен. Он окажется частным случаем индуцированных римановых метрик на поверхностях в евклидовом пространстве.

На сфере S^2 можно ввести и другие криволинейные координаты, иногда используемые при вычислениях. Приведем основные примеры.

Рассмотрим стереографическую проекцию сферы на плоскость. Поместим центр сферы радиуса R в начало координат и рассмотрим координатную плоскость $\mathbb{R}^2(x, y)$, проходящую через точку O. Отметим также на сфере северный полюс N и южный полюс S. Пусть P — точка сферы, отличная от N. Соединим ее с северным полюсом и продолжим этот отрезок

Рис. 1.16:

Рис. 1.17:

до пересечения с плоскостью $\mathbb{R}^2(x, y)$ в точке Q. Сопоставим точке P точку Q. Мы получим некоторое отображение $\nu\colon S^2 \to \mathbb{R}^2$, которое и называется *стереографической проекцией* сферы.

Отображение ν определено во всех точках сферы, кроме северного полюса N. Можно условно считать, что северный полюс изображает бесконечно удаленные точки двумерной плоскости (рис. 1.16). Запишем отображение ν аналитически.

Рассмотрим сферические координаты r, θ, φ в $\mathbb{R}^3$. Они индуцируют на сфере координаты (θ, φ), а на плоскости — координаты (r, φ), т.е. полярные координаты. Так как отображение ν сохраняет координату φ, то для определения ν достаточно найти зависимость радиуса r от угла θ. Рассмотрим плоское сечение S^2 плоскостью, проходящей через точки P, O, N (рис. 1.17). Так как угол ONT равен $\dfrac{\pi}{2} - \dfrac{\theta}{2}$, то получаем формулы замены координат: $\varphi = \varphi$; $r = R\operatorname{ctg}(\theta/2)$.

Несложно проверить, что якобиан такой замены имеет вид $J = -\dfrac{R}{2\sin^2\dfrac{\theta}{2}}$. Следовательно, замена является регулярной во всех точках сферы, кроме северного полюса. Итак, на сфере можно ввести координаты, заимствованные с полярных координат на плоскости. Какой вид принимает в них метрика сферы? Легко проверить, что метрика $ds^2 = R^2(d\theta^2 + \sin^2\theta\, d\varphi^2)$

записывается теперь так:

$$ds^2 = \frac{4R^4}{(R^2+r^2)^2}(dr^2 + r^2\,d\varphi^2).$$

Полученная нами метрика на сфере отличается от евклидовой метрики на плоскости $(dr^2 + r^2\,d\varphi^2)$, записанной в полярных координатах. Отличие — в множителе $\dfrac{4R^4}{(R^2+r^2)^2}$. Такие метрики называются конформными.

Определение 1.11. Риманова метрика $g_{ij}(z)$, заданная в области C евклидова пространства в криволинейных координатах $z^1, \ldots, z^n$, называется *конформной*, если она имеет вид $g_{ij}(z) = \lambda(z) g^l_{ij}(z)$, где $\lambda(z)$ — гладкая функция, а $g^l_{ij}(z)$ — евклидова метрика, записанная в координатах $z^1, \ldots, z^n$. Иными словами, метрика $g_{ij}(z)$ называется конформной, если существует такая система координат x, в которой $g_{ij}(x) = \lambda(x) \sum\limits_{k=1}^{n} (dx^k)^2$.

Таким образом, на евклидовой плоскости возникли две метрики: $dr^2 + r^2\,d\varphi^2$, т.е. евклидова, и $\dfrac{4R^4}{(R^2+r^2)^2}(dr^2 + r^2\,d\varphi^2)$, т.е. метрика сферы. Вернемся к вопросу об эквивалентности метрик. Существует ли такая регулярная замена координат на плоскости, при которой первая метрика перейдет во вторую?

Мы приведем наглядное обоснование того факта, что эти две метрики неэквивалентны. Подсчитаем длину окружности $x^2 + y^2 = a^2$ в евклидовой и в сферической метриках как функцию от радиуса окружности. Длина в евклидовой метрике нам известна: $l_e = 2\pi a$, где a — величина радиуса, подсчитанная в евклидовой метрике. Рассмотрим теперь сферическую метрику. В ней соотношение между евклидовой величиной радиуса a и его величиной ρ в сферической метрике, а также длина окружности имеют вид:

$$\rho = 2\int\limits_0^a \frac{R^2}{R^2+r^2}\,dr = 2R\,\mathrm{arctg}(a/R),$$

$$l_c = 2\int\limits_0^{2\pi} \frac{R^2 a\,d\varphi}{R^2+a^2} = \frac{4\pi a R^2}{R^2+a^2} = 2\pi R \sin\rho.$$

Геометрически величина ρ изображается как длина меридиана, соединяющего северный полюс с точкой на окружности (рис. 1.18).

В частности, при $\rho \to 0$, т.е. для окружностей малого радиуса, получаем, что $l_c \sim 2\pi\rho$, и формула переходит в выражение для длины плоской окружности. Итак, сравнивая две формулы: “евклидова длина окружности евклидова радиуса ρ равна $2\pi\rho$” и “сферическая длина окружности сферического радиуса ρ равна $2\pi R \sin\rho$”, мы видим, что они различны. В частности, одна линейная, а вторая периодическая.

Мы видим, что формулы для длины окружности одного и того же радиуса в евклидовой и сферической метрике — разные. Это указывает на то, что метрики не эквивалентны, или, как часто говорят, не изометричны. Строгое доказательство их неизометричности мы получим позже.

Рис. 1.18:

Тот факт, что выпуклая поверхность сферы не деформируется с сохранением длин всех кривых на ней в область на евклидовой плоскости, можно осознать, представив себе, какое требуется усилие, чтобы сплющить сферический сегмент на плоскость, по сравнению с усилием, которое мы прикладываем при изгибании прямого кругового цилиндра, накладывая его на плоскость.

Длину окружности на сфере можно вычислить непосредственно в сферических координатах (θ, φ), в которых метрика сферы имеет вид: $R^2(d\theta^2 + \sin^2\theta\, d\varphi^2)$. Областью определения метрики можно считать диск радиуса π на плоскости переменных θ, φ. В этих координатах очевидно, что длина окружности радиуса θ равна $2\pi R \sin\theta$ (рис. 1.19). При этом точка O отождествляется с северным полюсом сферы, а граничная окружность диска (радиуса π) склеивается в одну точку, отождествляясь с южным полюсом сферы.

Разберем другой пример вычисления длин кривых на сфере. Рассмотрим так называемую *локсодрому* — кривую на сфере, пересекающую меридианы под одним и тем же углом α. Эта кривая хорошо известна в теории навигации, по ней иногда удобно прокладывать маршруты самолетов.

Рис. 1.19:

Найдем уравнение локсодромы. При стереографической проекции сферы на плоскость меридианы перейдут в лучи, выходящие из начала координат. Образом локсодромы будет кривая на плоскости, пересекающая эти лучи под тем же углом α. Это утверждение следует из более общего факта: стереографическая проекция сохраняет углы между пересекающимися кривыми. Такие преобразования называются *конформными*. Более точно, мы имеем в виду следующее: рассмотрим на сфере метрику, индуцированную объемлющей евклидовой метрикой, и пусть две кривые на сфере пересекаются под углом $\alpha(\dot\gamma_1, \dot\gamma_2)$ (мы предполагаем углы ориентированными) (рис. 1.20).

Кривые γ_1 и γ_2 переходят при стереографической проекции в кривые q_1 и q_2. Угол между ними, подсчитанный в евклидовой метрике на плоскости, мы обозначим через $\beta(\dot q_1, \dot q_2)$.

Лемма 1.5. *Имеет место равенство* $\alpha(\dot\gamma_1, \dot\gamma_2) = \beta(\dot q_1, \dot q_2)$.

Доказательство. Достаточно сравнить явные формулы для углов α и β,

Рис. 1.20:

а также использовать запись сферической метрики после отображения ν в полярных координатах на плоскости $\mathbb{R}^2$. ■

В действительности имеет место еще более общее утверждение.

Лемма 1.6. *Пусть $g_{ij}(z)$ и $q_{ij}(z)$ — две метрики, заданные в области C евклидова пространства в криволинейных координатах $(z^1, \dots, z^n)$. Тогда если в каждой точке $z \in C$ выполнено тождество $g_{ij} = \lambda q_{ij}$, где $\lambda = \lambda(z)$ — гладкая функция, то углы между пересекающимися кривыми, вычисленные в этих метриках, совпадают.*

Доказательство дословно повторяет рассуждение из леммы 1.5.

Вернемся к локсодроме. В силу леммы 1.5 достаточно найти ее уравнение на евклидовой плоскости. Условие сохранения угла α означает, что $\langle(\dot r, \dot\varphi), (1,0)\rangle = \cos\alpha$. Здесь $(\dot r, \dot\varphi)$ — касательный вектор к локсодроме, $(1,0)$ — вектор скорости луча $\varphi = \mathrm{const}$, $r = t$. Отсюда $\dfrac{\dot r}{\sqrt{\dot r^2 + r^2\dot\varphi^2}} = \cos\alpha$. Решая это простое уравнение, получаем:

$$\dot r/r = \dot\varphi \operatorname{ctg}\alpha, \quad r = c\, e^{\varphi \operatorname{ctg}\alpha},$$

где $c = \mathrm{const}$. Вычисляя длину дуги, получаем

$$l(\gamma)_{\varphi_0}^{\varphi} = c' e^{\varphi t \operatorname{ctg}\alpha} + c'',$$

где c'', $c' = \mathrm{const}$. Задача: найдите значение постоянных c', c''.

Задачи

1. Показать, что у треугольника на сфере S^2, составленного из дуг больших окружностей, сумма углов больше 2π.

2. Выразить сумму углов треугольника на сфере S^2 через его площадь (треугольник составлен из дуг больших окружностей).

3. Показать, что на сфере S^2 преобразование подобия возможно только с единичным коэффициентом.

4. Показать, что для любой римановой метрики найдется такая система координат, в которой в данной точке матрица римановой метрики единична.

Рис. 1.21:

Рис. 1.22:

1.4 Псевдосфера и геометрия Лобачевского

В евклидовом пространстве $\mathbb{R}^n$ сфера S^{n-1} (гиперсфера) может быть определена как множество точек, удаленных от начала координат на расстояние ρ. В псевдоевклидовом пространстве $\mathbb{R}^n_s$ индекса s также можно рассмотреть множество точек, удаленных от некоторой точки на расстояние ρ. Но здесь число ρ может быть не только вещественным, но также чисто мнимым или нулем. Это множество мы назовем *псевдосферой* индекса s и обозначать через S^{n-1}_s. Будем различать псевдосферы вещественного, мнимого и нулевого радиусов. Псевдосфера нулевого радиуса описывается уравнением второго порядка $-\sum\limits_{i=1}^{s}(x^i)^2 + \sum\limits_{j=s+1}^{n}(x^j)^2 = 0$, где $x^1, \ldots, x^n$ — декартовы координаты в $\mathbb{R}^n$, на котором мы моделируем псевдоевклидово пространство. Псевдосфера нулевого радиуса совпадает с изотропным конусом.

Рассмотрим примеры. Пусть $n = 2$, $s = 1$. Изотропный "конус" на двумерной плоскости состоит из двух прямых: $x^1 = \pm x^2$. Он разбивает плоскости на две области: в одной из них $\langle\xi,\xi\rangle_1 > 0$, т.е. $|x^2| > |x^1|$, а в другой $\langle\xi,\xi\rangle_1 < 0$, т.е. здесь $|x^2| < |x^1|$ (рис. 1.21).

Псевдоокружности вещественного радиуса — это гиперболы $-(x^1)^2 + (x^2)^2 = \alpha^2$, где α — вещественно. Псевдоокружности мнимого радиуса — это гиперболы $-(x^1)^2 + (x^2)^2 = -\alpha^2$ (рис. 1.22).

Пусть теперь $n = 3$, $s = 1$. Изотропный конус, т.е. псевдосфера нулевого

Рис. 1.23:

Рис. 1.24:

радиуса, — это обычный конус второго порядка, задаваемый уравнением $-(x^1)^2 + (x^2)^2 + (x^3)^2 = 0$, с осью x^1. Он разбивает пространство на две области — внутреннюю и внешнюю (рис. 1.23). Псевдосферы вещественного радиуса — это однополостные гиперболоиды $-(x^1)^2 + (x^2)^2 + (x^3)^2 = \alpha^2$. Псевдосферы мнимого радиуса — это двухполостные гиперболоиды $-(x^1)^2 + (x^2)^2 + (x^3)^2 = -\alpha^2$, (рис. 1.24). Напомним, что для удобства мы моделируем псевдоевклидово пространство $\mathbb{R}^3_1$ с помощью точек обычного евклидова пространства $\mathbb{R}^3$. Через x, y, z обозначаются декартовы координаты. Тогда $\langle\xi,\xi\rangle_1 = -x^2 + y^2 + z^2$.

Рассмотрим псевдосферу мнимого радиуса — двухполостный гиперболоид $-\alpha^2 = -x^2 + y^2 + z^2$. Геометрия (метрика) пространства $\mathbb{R}^3_1$ индуцирует определенную геометрию (метрику) на псевдосфере мнимого радиуса.

Для простоты ограничимся изучением только одной полости этой псевдосферы, например, определенной неравенством $x > 0$). Точками геометрии псевдосферы объявим обычные точки гиперболоида, а "прямыми" — линии на гиперболоиде, являющиеся пересечениями с плоскостями $ax+by+cz = 0$, проходящими через начало координат (рис. 1.25).

Далее, рассмотрим преобразование, аналогичное стереографической проекции обычной сферы на плоскость. Центром псевдосферы $S^2_1 = \{-\alpha^2 = -x^2 + y^2 + z^2\}$ мы считаем начало координат; северным полюсом N — точку с координатами $(-\alpha, 0, 0)$; южным полюсом S — точку с координатами $(\alpha, 0, 0)$. В качестве плоскости проекции возьмем плоскость YOZ, проходя-

Рис. 1.26:

щую через центр псевдосферы. Кстати, ограничение скалярного произведения $\langle \xi, \eta \rangle_1$ на эту плоскость имеет вид $\xi^2\eta^2+\xi^3\eta^3$, т.е. является евклидовым.

Рис. 1.25:

Рассмотрим переменную точку P на правой полости гиперболоида и соединим ее с точкой N, т.е. с северным полюсом. Отрезок $\overrightarrow{PN}$ встречает плоскость YOZ в точке $f(P)$, которую мы и назовем образом точки P при стереографической проекции $f\colon S_1^2 \to \mathbb{R}^2$. Точно так же определяется стереографическая проекция левой полости гиперболоида на ту же плоскость, если за центр проектирования взять южный полюс S. На рис. 1.26 показано сечение псевдосферы плоскостью, проходящей через ось OX. Образ правой полости гиперболоида покрывает не всю плоскость YOZ, а только внутренность диска радиуса α : $y^2+z^2 < \alpha^2$. Образ левой полости гиперболоида покрывает внешность окружности $y^2+z^2 = \alpha^2$. В отличие от обычной сферы, образ псевдосферы при стереографической проекции покрывает не всю плоскость YOZ, так как окружность $y^2+z^2 = \alpha^2$ не принадлежит образу проекции. Северный полюс N переходит в бесконечно удаленную точку плоскости YOZ.

Пусть точка P имеет координаты (x, y, z), где $x > 0$, и пусть (u^1, u^2) — декартовы координаты точки $f(P)$ на плоскости YOZ, где отображение f — стереографическая проекция. Найдем явные формулы для стереографической проекции.

Лемма 1.7. *Пусть $P = (x, y, z)$, $f(P) = (u^1, u^2)$. Тогда*

$$x = \alpha\frac{|\mathbf{u}|^2+\alpha^2}{\alpha^2-|\mathbf{u}|^2}, \quad y = \frac{2\alpha^2u^1}{\alpha^2-|\mathbf{u}|^2}, \quad z = \frac{2\alpha^2u^2}{\alpha^2-|\mathbf{u}|^2},$$

где $|\mathbf{u}|^2 = (u^1)^2 + (u^2)^2$, $\mathbf{u} = (u^1, u^2)$.

Доказательство. Из рис. 1.26 следует

$$\frac{y}{u^1} = \frac{x+\alpha}{\alpha}, \quad \frac{z}{u^2} = \frac{x+\alpha}{\alpha}.$$

Так как $-\alpha^2 = -x^2 + y^2 + z^2$, то

$$-\alpha^2 = -x^2 + \big((u^1)^2 + (u^2)^2\big)\frac{(x+\alpha)^2}{\alpha^2}.$$

Отсюда

$$x = \alpha\frac{|\mathbf{u}|^2 + \alpha^2}{\alpha^2 - |\mathbf{u}|^2}. \qquad \blacksquare$$

Лемма 1.8. *Координаты (u^1, u^2), меняющиеся в открытом диске $(u^1)^2 + (u^2)^2 < \alpha^2$, задают регулярную систему координат на правой полости гиперболоида, т.е. стереографическая проекция задает регулярные координаты $f\colon (x, y, z) \to (u^1, u^2)$.*

Доказательство. Можно считать, что правая полость псевдосферы задается уравнением $x = \sqrt{\alpha^2 + y^2 + z^2}$. Тем самым стереографическую проекцию f можно понимать как замену координат $(y, z) \to (u^1, u^2)$. Осталось найти матрицу Якоби и убедиться, что якобиан отличен от нуля. Прямой подсчет дает

$$J(f) = 4\alpha^4\frac{\alpha^2 + |\mathbf{u}|^2}{(\alpha^2 - |\mathbf{u}|^2)^3} = \frac{4\alpha^3 x}{(\alpha^2 - |\mathbf{u}|^2)^2} > 0.$$

Итак, якобиан положителен в диске $y^2 + z^2 < \alpha^2$. $\blacksquare$

Вернемся к обычной сфере. Какая геометрия возникнет на ней, если в качестве точек взять обычные точки сферы, а в качестве “прямых” — сечения сферы плоскостями, проходящими через центр O, т.е. экваторы? Каким аксиомам будет удовлетворять множество таких точек и прямых? Ясно, что через каждую пару не диаметрально противоположных точек проходит одна и только одна прямая. Однако если точки диаметрально противоположны, то через них проходит бесконечное множество прямых. Кроме того, через точку, расположенную вне прямой, нельзя провести ни одной прямой, не пересекающейся с исходной прямой. В получившейся геометрии на сфере не существует параллельных, т.е. не пересекающихся, прямых.

Кое-какие улучшения сделать можно. Объявим “точками” новой геометрии пары диаметрально противоположных точек, P и $-P$, на сфере. Теперь через любую пару точек проходит одна и только одна “прямая” (если точки не совпадают). Легко проверяется, что в получившейся геометрии выполнены все классические аксиомы Евклида, кроме так называемого пятого постулата о том, что через точку вне данной прямой можно провести одну и только одну прямую, параллельную данной. Он не выполняется в построенной таким образом геометрии, так как, согласно приведенным определениям “точек” и “прямых”, через “точку”, взятую вне “прямой”, нельзя провести ни одной “прямой”, параллельной данной, т.е. любые две “прямые” либо пересекаются в одной точке, либо совпадают. Эта геометрия называется иногда *эллиптической*.

Рис. 1.27:

Описанное отождествление точек вида P и $-P$ эквивалентно факторизации сферы по действию симметрии отражения в точке O. Так как каждая пара $(P, -P)$ определяет ровно одну прямую в трехмерном пространстве (куда вложена сфера), то можно сопоставить каждой "прямой" эллиптической геометрии, т.е. экватору, ортогональную прямую, проходящую через O. Таким образом, эллиптическая геометрия моделируется на двумерном вещественном проективном пространстве.

Вернемся к псевдоевклидовой геометрии и к геометрии, индуцируемой на псевдосфере мнимого радиуса. Рассмотрим стереографическую проекцию $f\colon {}^+S_1^2 \to \{y^2 + z^2 < \alpha^2\} = D^2$. Через ${}^+S_1^2$ обозначена правая полость гиперболоида. Ее точки переходят во внутренние точки диска D^2 радиуса α. В какие кривые на круге D^2 перейдут "прямые" геометрии на гиперболоиде, т.е. линии пересечения его с плоскостями, проходящими через центр псевдосферы — точку O (т.е. аналоги экваторов на сфере).

Лемма 1.9. *Каждая линия пересечения ${}^+S_1^2$ с плоскостью вида $ax + by + cz = 0$ переходит при отображении f в дугу окружности, пересекающую окружность $y^2 + z^2 = \alpha^2$ под прямым углом* (*рис.* 1.27).

Доказательство. В силу леммы 1.7 достаточно подставить в уравнение плоскости $ax + by + cz = 0$ выражения переменных x, y, z как функций от переменных u^1, u^2. Пусть, например, $a \neq 0$. Тогда уравнение

$$a\alpha\frac{|\mathbf{u}|^2 + \alpha^2}{\alpha^2 - |\mathbf{u}|^2} + \frac{2b\alpha^2 u^1}{\alpha^2 - |\mathbf{u}|^2} + \frac{2c\alpha^2 u^2}{\alpha^2 - |\mathbf{u}|^2} = 0$$

после элементарных преобразований приводится к виду

$$\left(u^1 + \frac{b\alpha}{a}\right)^2 + \left(u^2 + \frac{c\alpha}{a}\right)^2 = \frac{\alpha^2}{a^2}(b^2 + c^2 - a^2),$$

т.е. определяет окружность с центром в точке $\left(-\dfrac{b\alpha}{a}, -\dfrac{c\alpha}{a}\right)$ радиуса $\dfrac{\alpha}{a}\sqrt{b^2 + c^2 - a^2} = r$, пересекающую окружность $y^2 + z^2 = \alpha^2$ в точках A и

Рис. 1.28:

B под прямым углом (рис. 1.28). Очевидно, что угол в точках пересечения равен $\pi/2$. ■

Итак, геометрия, индуцированная на псевдосфере мнимого радиуса в $\mathbb{R}^3_1$ совпадает с геометрией, возникающей в открытом евклидовом диске радиуса α, если в качестве точек взять обычные точки диска, а в качестве "прямых" взять дуги окружностей, пересекающих границу круга под прямым углом. В частности, "прямыми" являются все диаметры круга, поскольку их можно рассматривать как дуги окружностей бесконечно большого радиуса. Эта геометрия называется *гиперболической или геометрией Лобачевского*, а ее модель в круге радиуса α называется *моделью Пуанкаре* геометрии Лобачевского.

Рис. 1.29:

На модели Пуанкаре легко проверить, что здесь все аксиомы Евклида, кроме пятого постулата, справедливы. На рис. 1.29 видно, что через любую точку, взятую вне прямой, можно провести бесконечное число прямых, параллельных данной, т.е. не пересекающихся с ней. Если устремить параметр α к бесконечности, то в любой конечной области геометрия Лобачевского будет "стремиться" к евклидовой, так как дуги окружностей начнут распрямляться и превращаться в евклидовы прямые. Граница модели Пуанкаре, окружность, называется *абсолютом*. Здесь расположены бесконечно удаленные точки плоскости Лобачевского. При изучении плоскости Лобачевского часто полагают для простоты $\alpha = 1$.

Замечание. Можно рассмотреть геометрию, возникающую на псевдосферах вещественного радиуса, т.е. на однополостных гиперболоидах. Она совпадает с геометрией Лобачевского.

Вычислим риманову метрику, индуцированную на псевдосфере мнимого радиуса объемлющей индефинитной метрикой. Введем в $\mathbb{R}^3_1$ аналог сферических координат. В плоскости YOZ зададим полярные координаты (r, φ),

где φ — угол с осью y. Введем еще параметр θ' и сделаем замену переменных

$$y = \alpha \operatorname{sh} \theta' \cos \varphi, \quad z = \alpha \operatorname{sh} \theta' \sin \varphi, \quad x = \alpha \operatorname{ch} \theta'.$$

В этой "псевдосферической" системе координат уравнение псевдосферы очевидно запишется в виде $\alpha = \mathrm{const}$.

Вычислим риманову метрику на псевдосфере в координатах u^1, u^2 на модели Пуанкаре. Подставляя формулы для стереографической проекции в квадрат дифференциала длины дуги в $\mathbb{R}^3_1$, получаем (проверьте!)

$$4\alpha^4 \frac{(du^1)^2 + (du^2)^2}{\left(\alpha^2 - (u^1)^2 - (u^2)^2\right)^2}.$$

В полярных координатах (пусть $\alpha = 1$) на модели Пуанкаре эта же метрика запишется так:

$$ds^2 = 4\,\frac{dr^2 + r^2\,d\varphi^2}{(1 - r^2)^2}.$$

Видно, что получившаяся метрика конформна, т.е. отличается от евклидовой множителем $\lambda(r) = 4(1 - r^2)^{-2}$. Перепишем метрику в псевдосферических координатах. Заменим координаты (r, φ) на параметры (χ, φ) по формулам: $r = \operatorname{cth}(\chi/2)$, $\varphi = \varphi$. Непосредственное вычисление дает (проверьте!) $ds^2 = d\chi^2 + \operatorname{sh}^2\chi\, d\varphi^2$. Этот вид метрики полностью аналогичен записи метрики сферы в координатах (θ, φ), но только обычные тригонометрические функции заменены здесь на гиперболические.

Найдем геометрический смысл параметра χ. Рассмотрим плоскость XOZ с индуцированной на ней псевдоевклидовой метрикой $-dx^2 + dz^2$. Псевдосфера пересекает эту плоскость по гиперболе, которая в псевдосферических координатах имеет вид (мы считаем, что $\alpha = 1$) $x = \operatorname{ch}\theta'$, $z = \operatorname{sh}\theta'$ (рис. 1.30). В качестве параметра θ возьмем евклидово значение угла POS (рис. 1.30). Тогда, очевидно, $\operatorname{tg}\theta = \operatorname{th}\theta'$. Найдем длину отрезка гиперболы в псевдоевклидовой метрике от значения 0 до θ'. Имеем

$$l = \int\limits_0^{\theta'} \sqrt{-\operatorname{sh}^2\theta + \operatorname{ch}^2\theta}\, d\theta = \theta'.$$

Таким образом, θ' совпадает с длиной "меридиана" на псевдосфере, идущего из южного полюса S в точку P, т.е. параметр аналогичен параметру θ на сфере.

Рассмотрим теперь стереографическую проекцию псевдосферы. Для простоты, мы рассматриваем только плоскость XOZ, так как все вычисления сохраняются при вращениях плоскости XOZ вокруг оси OX. Так как $z = \dfrac{2u^2\alpha^2}{\alpha^2 - r^2}$, то получаем: $z = \dfrac{2r}{1 - r^2}$ (так как $u^2 = r$ в плоскости XOZ). Так как $z = \operatorname{sh}\theta' \sin\varphi = \operatorname{sh}\theta'$ (так как $\varphi = \pi/2$), то $\dfrac{2r}{1 - r^2} = \operatorname{sh}\theta'$, отку-

Рис. 1.30:

Рис. 1.31:

да $r = \operatorname{cth}(\theta'/2)$. Тем самым доказано, что $\chi = \theta'$. Итак, мы получили на модели Пуанкаре риманову метрику $4\,\dfrac{dr^2 + r^2\,d\varphi^2}{(1-r^2)^2}$.

Отметим, что она — положительно определенная, хотя объемлющая метрика была псевдоевклидовой. То обстоятельство, что на псевдосфере индуцируется положительно определенная метрика, можно усмотреть из наглядных соображений. Рассмотрим для простоты сечение псевдосферы плоскостью XOZ, и пусть ξ — вектор скорости гиперболы в точке P. Мы хотим убедиться в том, что его псевдоевклидова длина вещественна. Это следует из рис. 1.32, поскольку вектор ξ расположен вне светового конуса с вершиной в точке P.

Рис. 1.32:

Полученная риманова метрика называется *метрикой Лобачевского.* Ее можно рассматривать на евклидовом круге. Мы уже познакомились с другими метриками на круге — евклидовой и сферической. Было показано, что они метрики не эквивалентны.

Докажем, что метрика Лобачевского не эквивалентна ни одной из двух предыдущих метрик. Применим уже использованный прием: найдем длину

Рис. 1.33:

окружности на плоскости Лобачевского как функцию от радиуса (вычисленного в метрике Лобачевского). Пусть окружность имеет центр в точке O и ее евклидов радиус равен a. Найдем длину радиуса в метрике Лобачевского. По определению длины кривой, получаем

$$\chi = 2\int\limits_0^a \frac{dr}{1-r^2} = \ln\frac{1+a}{1-a},$$

т.е. $a = \mathrm{th}(\chi/2)$. Длина окружности:

$$l = 2\int\limits_0^{2\pi} \frac{a d\varphi}{1-a^2} = \frac{4\pi a}{1-a^2} = 2\pi\,\mathrm{sh}\,\chi.$$

Если χ мало, то приближенно можно считать, что $l \sim 2\pi\chi$, т.е. получаем формулу для длины окружности в евклидовой метрике.

Поскольку мы выразили длину окружности (в метрике Лобачевского) в инвариантных относительно замен координат терминах, т.е. через величину радиуса, также подсчитанную в метрике Лобачевского, то полученная формула инвариантна относительно замен координат. Следовательно, метрика Лобачевского не эквивалентна ни одной из двух предыдущих метрик. На рис. 1.33 сравниваются разные записи метрик сферы и псевдосферы.

Укажем еще две полезные записи перечисленных ранее метрик — так называемые комплексные формы.

Введем на евклидовой плоскости комплексную координату $z = x + iy$. В качестве $\bar{z}$ возьмем $x - iy$. Рассмотрим замену $(x, y) \to (z, \bar{z})$ как формальную замену переменных. Тогда якобиан имеет вид: $J = -2i \neq 0$, а потому замена регулярна. Так как $dz = dx + i\,dy$, $d\bar{z} = dx - i\,dy$, то евклидова метрика в новых координатах принимает вид

$$ds^2 = dx^2 + dy^2 = (dx + i\,dy)(dx - i\,dy) = dz\,d\bar{z}.$$

Метрика сферы, следовательно, принимает вид

$$ds^2 = 4\,\frac{dx^2 + dy^2}{(1 + x^2 + y^2)^2} = 4\,\frac{dz d\bar{z}}{(1 + |z|^2)^2},$$

где $|z|^2 = r^2$. Аналогично получается комплексная форма метрики Лобачевского: $ds^2 = 4\,\dfrac{dz d\bar{z}}{(1 - |z|^2)^2}$.

Для метрики Лобачевского существует еще одна полезная форма записи на верхней полуплоскости. Рассмотрим еще один экземпляр плоскости с комплексной координатой w. Выделим верхнюю полуплоскость неравенством $\mathrm{Im}(w) > 0$, где $w = u + iv$, $v = \mathrm{Im}(w)$.

Рассмотрим отображение $\mathbb{R}^2(w) \to \mathbb{R}^2(z)$, задаваемое формулой $z = \dfrac{aw + b}{cw + d}$, где a, b, c, d — комплексные числа такие, что $ad - bc \neq 0$. Такие отображения называются *невырожденными дробно-линейными*. Отметим, что если $ad - bc = 0$, то отображение переводит всю плоскость в одну точку (проверьте!).

Легко найти невырожденное отображение вида $z = \dfrac{aw + b}{cw + d}$, переводящее верхнюю полуплоскость на внутренность единичного круга $|z| < 1$ на плоскости z. При этом вещественная прямая $\mathrm{Im}(w) = 0$ перейдет в окружность $|z| = 1$. Между прочим, любое невырожденное дробно-линейное отображение однозначно задается образом любых трех точек, не лежащих на одной прямой. Мы не будем здесь доказывать этот факт. Итак, найдем такое отображение $z = \dfrac{aw + b}{cw + d}$, чтобы $0 \to 1$, $i \to 0$, $1 \to i$ (рис. 1.34). Решая получающуюся систему уравнений для a, b, c, d, находим: $z = \dfrac{1 + iw}{1 - iw}$. Итак, мы нашли дробно-линейное преобразование, переводящее верхнюю полуплоскость на единичный круг (таких преобразований имеется много).

Преобразование $z = \dfrac{1 + iw}{1 - iw}$ является регулярным. В самом деле, представим его в виде

$$z = -1 - \frac{2}{iw - 1}.$$

Достаточно доказать, что преобразование $z = \dfrac{1}{w}$ является регулярной заменой координат, поскольку исходное отображение является композицией

Рис. 1.34:

преобразований: $z = \dfrac{1}{w}$, сдвига на постоянный вектор, поворота и растяжения. В вещественном виде отображение $z = \dfrac{1}{w}$ записывается так:

$$x = \frac{u}{u^2+v^2}, \quad y = \frac{-v}{u^2+v^2}.$$

Отсюда якобиан $J = (u^2+v^2)^2 > 0$. Утверждение доказано.

Далее:

$$dz = \frac{2i\,dw}{(1-iw)^2}.$$

Производя замену в выражении $ds^2 = \dfrac{4dzd\bar{z}}{(1-|z|^2)^2}$, получаем

$$ds^2 = -4\,\frac{dw\,d\bar{w}}{(w-\bar{w})^2} = \frac{du^2+dv^2}{v^2}.$$

Отсюда еще раз видно, что точки вещественной оси (образ абсолюта модели Пуанкаре) являются бесконечно удаленными точками на плоскости Лобачевского. В самом деле, если мы захотим подсчитать длину отрезка оси Ov, от точки i до точки O, то получим $l = \int\limits_0^1 \dfrac{dv}{v} = \ln v\Big|_0^1 = -\ln(0) \to \infty$.

Выясним, в какие кривые перейдут при отображении на верхнюю полуплоскость прямые плоскости Лобачевского (в модели Пуанкаре). Сначала решим аналогичный вопрос для эллиптической геометрии. Напомним, что здесь "прямыми" являются экваторы на сфере. Рассмотрим стереографическую проекцию сферы на плоскость и найдем образы экваторов.

Лемма 1.10. *При стереографической проекции экваторы сферы переходят либо в окружности на плоскости, либо в прямые. Причем, экватор отображается в прямую тогда и только тогда, когда он проходит через северный полюс сферы.*

Доказательство. Запишем стереографическую проекцию в декартовых координатах. Пусть x, y, z — координаты точки P на сфере, а x', y' — координаты ее образа на плоскости. Из рис. 1.35 получаем: $x' = \dfrac{x}{1-z}$, $y' = \dfrac{y}{1-z}$.

Рис. 1.35:

Окружность на сфере радиуса 1 зададим двумя уравнениями: $ax + by + cz = d$, $x^2 + y^2 + z^2 = 1$. Прямое вычисление дает:

$$(x')^2 + (y')^2 + \frac{2a}{c-d}x' + \frac{2b}{c-d}y' = 1 + \frac{2d}{c-d}.$$

Это уравнение второго порядка определяет на плоскости либо прямую, либо окружность. ■

Аналогичное утверждение имеется и для плоскости Лобачевского.

Лемма 1.11. *Невырожденное дробно-линейное преобразование* $z = \dfrac{aw+b}{cw+d}$ *плоскости в себя, отображает прямые и окружности снова в прямые и окружности, причем прямая может перейти в окружность, и наоборот.*

Доказательство. Если $c = 0$, то утверждение очевидно, так как тогда преобразование представляет собой параллельный сдвиг на вектор и умножение на комплексное число. Пусть $c \neq 0$. Тогда

$$z = \frac{a}{c} - \frac{(ad-bc)}{c(cw+d)},$$

т.е. утверждение осталось доказать лишь для преобразования $z = 1/w$.

Запишем произвольную окружность на плоскости z в виде $|z - z_0|^2 = \varepsilon^2$, т.е. $(z - z_0)(\bar{z} - \bar{z}_0) = \varepsilon^2$. Делая замену $z = 1/w$, получаем

$$w\bar{w}(\varepsilon^2 - z_0\bar{z}_0) + z_0 w + \bar{z}_0\bar{w} - 1 = 0.$$

Ясно, что это уравнение, в зависимости от параметров ε, z_0, определяет либо окружность, либо прямую. ■

Следствие 1.1. *Рассмотрим отображение* $z = \dfrac{1+iw}{1-iw}$, *переводящее верхнюю полуплоскость в единичный круг. Тогда "прямые" модели Пуанкаре, т.е. дуги окружностей, ортогональных абсолюту, переходят либо в прямые на плоскости* $w = u + iv$, *ортогональные вещественной оси, либо в полуокружности, ортогональные вещественной оси* (*рис.*1.36).

Рис. 1.36:

Доказательство. Отображение $z = \dfrac{1+iw}{1-iw}$ можно считать определенным всюду на плоскости w. В силу леммы 1.11 оно переводит прямые либо в прямые, либо в окружности. Аналогично и окружности переходят либо в прямые, либо в окружности. Преобразование имеет обратное: $w = \dfrac{z-1}{i(z+1)}$. Следовательно, "прямые" модели Пуанкаре переходят либо в прямые, либо в окружности на плоскости w. Осталось доказать, что они должны быть ортогональны вещественной оси в точках пересечения с ней. Это будет следовать из леммы 1.12.

Лемма 1.12. *Любое невырожденное дробно-линейное преобразование* $z = \dfrac{aw+b}{cw+d}$, *отображающее плоскость* w *на плоскость* z, *сохраняет углы между гладкими кривыми в точках их пересечения.*

Доказательство. Достаточно вычислить евклидову метрику $dz\,d\bar{z}$ в новых координатах $w = u + iv$. Прямое вычисление показывает, что

$$dz = \frac{ad-bc}{(cw+d)^2}dw,$$

т.е. $dz\,d\bar{z} = \dfrac{|ad-bc|^2}{|cw+d|^4}\,dw\,d\bar{w}$. Таким образом, отображение оказалось конформным, т.е. умножает евклидову метрику на положительный множитель и, как было доказано ранее, сохраняет углы между пересекающимися кривыми. ■

Тем самым полностью доказано и следствие 1.1. ■

Отметим, что через любые две различные точки на плоскости Лобачевского всегда проходит единственная прямая (в смысле Лобачевского). Ее построение см. на рис. 1.37. В том случае, когда обе точки расположены на прямой, ортогональной вещественной оси, прямая плоскости Лобачевского совпадает с этой прямой.

Рис. 1.37:

Найдем формулу для длины отрезка прямой от точки P_1 до точки P_2 в метрике Лобачевского, реализованной, например, на верхней полуплоскости. Применяя горизонтальный параллельный перенос (он сохраняет метрику Лобачевского $ds^2 = (dx^2 + dy^2)/y^2$), можно считать, что центр окружности, изображающей прямую, проходящую через точки P_1 и P_2, находится в точке O. Уравнение прямой имеет вид $r = r_0$, где (r, φ) — полярные координаты. Пусть точки P_1 и P_2 задаются координатами (r_0, φ_1) и (r_0, φ_2) соответственно. Тогда длина дуги от P_1 до P_2 имеет вид (проверьте!)

$$l(P_1, P_2) = \int\limits_{\varphi_1}^{\varphi_2} \frac{r_0 d\varphi}{r_0 \sin\varphi} = \ln\Big(\frac{\operatorname{tg}(\varphi_1/2)}{\operatorname{tg}(\varphi_2/2)}\Big).$$

Глава 2

Кривые и поверхности

2.1 Теория кривых. Формулы Френе

2.1.1 Теория кривых на плоскости. Кривизна

Рассмотрим на плоскости декартовы координаты (x, y). Гладкие плоские кривые $\gamma(t)$ будем задавать радиус-вектором $\mathbf{r}(t) = \big(x(t), y(t)\big)$, выходящим из начала координат. Напомним, что вектором скорости $\mathbf{v}(t)$ кривой $\gamma(t)$ называется вектор с координатами $\big(dx(t)/dt, dy(t)/dt\big)$. Определяемая им прямая касается кривой в точке $\gamma(t)$. Обычно будем предполагать, что $\mathbf{v}(t) \neq 0$. Напомним, что в точках, где вектор скорости обращается в нуль, гладкая кривая может претерпевать излом. Мы будем рассматривать *регулярные* кривые, т.е. с ненулевым вектором скорости. Производную будем иногда обозначать точкой или штрихом.

Через $|d\mathbf{r}(t)/dt| = |\mathbf{v}(t)|$ обозначим модуль вектора скорости в евклидовой метрике. Пусть s — длина дуги кривой от некоторой фиксированной точки до переменной точки на кривой. Поскольку длина дуги монотонно растет при движении точки по кривой в одном направлении, то длину дуги можно взять за параметр вдоль кривой. Он называется *натуральным параметром.* Уравнение кривой $\mathbf{r}(s) = \big(x(s), y(s)\big)$, записанной относительно параметра s, называется *натуральной параметризацией кривой.*

Лемма 2.1. *Модуль вектора скорости кривой, записанной в натуральном параметре, постоянен и равен единице.*

Доказательство. Утверждение следует из формулы для дифференциала dl длины дуги l, т.е. $dl = dt\big|d\mathbf{r}(t)/dt\big|$. ■

Отметим, что в каждой точке регулярной гладкой кривой вектор скорости $\mathbf{v}(s)$ отличен от нуля.

Далее, в каждой точке кривой можно задать еще один вектор, а именно, ускорение кривой: $\ddot{\mathbf{r}}(s) = \dfrac{d\mathbf{v}(s)}{ds}$. Оказывается, ускорение (относительно натурального параметра) ортогонально скорости кривой.

Лемма 2.2. *Пусть задана вектор-функция* $\mathbf{p}(t)$ *такая, что* $|\mathbf{p}(t)| \equiv 1$. *Тогда вектор* $\dfrac{d\mathbf{p}(t)}{dt}$ *ортогонален вектору* $\mathbf{p}(t)$.

Доказательство. Дифференцируя тождество $\langle \mathbf{p}, \mathbf{p} \rangle \equiv 1$, получаем

$$\left\langle \frac{d\mathbf{p}}{dt}, \mathbf{p} \right\rangle \equiv 0. \qquad \blacksquare$$

Итак, в каждой точке гладкой кривой, отнесенной к натуральному параметру, возникают два ортогональных друг другу вектора, один — скорость, другой — ускорение. Вектор ускорения не обязан быть единичным. Предположим, что ускорение отлично от нуля и рассмотрим единичный вектор $\mathbf{n}(s) = \big(d\mathbf{v}(s)/ds\big) \big/ |d\mathbf{v}(s)/ds|$, т.е. "нормированное ускорение". Тем самым мы получаем вдоль кривой гладкое семейство реперов вида: $\big(\mathbf{v}(s), \mathbf{n}(s)\big)$. Они называются реперами или базисами Френе. Вектор $\mathbf{n}(s)$ называется *вектором главной нормали* к кривой в точке s. Каждый репер Френе, после его параллельного переноса в начало координат, однозначно определяет некоторое вращение плоскости вокруг точки O. Тем самым реперное поле вдоль кривой задает гладкое отображение $\gamma(s)$ кривой в группу ортогональных матриц, т.е. в группу вращений плоскости. Ниже мы изучим свойства этого отображения.

Определение 2.1. Пусть кривая отнесена к натуральному параметру. *Кривизной кривой* в точке s называется величина $k(s) = \left|\dfrac{d^2\mathbf{r}(s)}{ds^2}\right|$, т.е. модуль ускорения.

Из определения сразу видим, что $d^2\mathbf{r}/ds^2 = k(s)\mathbf{n}(s)$, где $\mathbf{n}(s)$ — вектор нормали к кривой.

Определение 2.2. *Радиусом кривизны* гладкой кривой в точке s называется число $R(s) = 1/k(s)$.

Рассмотрим простые примеры. Прямая на плоскости задается линейной вектор-функцией: $x(s) = x(0) + \alpha s$; $y(s) = y(0) + \beta s$, где s — натуральный параметр. Ясно, что должно выполняться равенство $\sqrt{\alpha^2 + \beta^2} = 1$, так как $\mathbf{v} = (\alpha, \beta)$ и $|\mathbf{v}(s)| \equiv 1$. Тогда ускорение $d\mathbf{v}(s)/ds$ равно нулю тождественно, а потому кривизна прямой линии равна нулю и радиус кривизны равен бесконечности.

Рассмотрим плоскую окружность радиуса R. Ее параметрическое уравнение таково:

$$x(s) = x(0) + R\cos(s/R), \quad y(s) = y(0) + R\sin(s/R).$$

Прямой подсчет дает, что кривизна окружности постоянна и равна $\dfrac{1}{R}$, а радиус кривизны равен $\dfrac{1}{k} = R$.

Однако во многих задачах уравнения кривой отнесены не к натуральному параметру, а к какому-то произвольному параметру t, а потому полезно уметь вычислять кривизну кривой, отнесенной к произвольному параметру.

Теорема 2.1. *Пусть гладкая кривая $\gamma(t)$ отнесена к некоторому параметру t, не обязательно натуральному. Пусть в точке t вектор скорости $\mathbf{v}(t)$ отличен от нуля. Тогда*

$$k = \frac{|x''y' - y''x'|}{\big((x')^2 + (y')^2\big)^{3/2}},$$

где через x', x'', y', y'' обозначены производные по параметру t.

Доказательство. Пусть $\mathbf{r}(t) = \big(x(t), y(t)\big)$ — радиус-вектор кривой и $\mathbf{v}(t) = \big(x'(t), y'(t)\big)$ — вектор скорости. Если s — натуральный параметр, то для произвольной вектор-функции $\mathbf{q}(t)$ имеем

$$\frac{d}{ds}\mathbf{q}(t) = \frac{d\mathbf{q}(t)}{dt}\frac{dt}{ds}.$$

Из определения кривизны получаем

$$k = \left|\frac{d}{ds}\left(\frac{\mathbf{v}(t)}{|\mathbf{v}(t)|}\right)\right|.$$

Далее:

$$\frac{d}{ds}\left(\frac{\mathbf{v}(t)}{|\mathbf{v}(t)|}\right) = \frac{dt}{ds}\frac{d}{dt}\left(\frac{\mathbf{v}(t)}{|\mathbf{v}(t)|}\right)$$

и

$$\frac{dt}{ds} = \frac{1}{\left|\dfrac{d\mathbf{r}(t)}{dt}\right|} = \frac{1}{|\mathbf{v}(t)|}.$$

Отсюда

$$k = \frac{1}{|\mathbf{v}(t)|}\left|\frac{d}{dt}\left(\frac{\mathbf{v}(t)}{|\mathbf{v}(t)|}\right)\right| = \frac{1}{|\mathbf{r}'|^2}\left|\mathbf{r}'' - \frac{\mathbf{r}'}{2|\mathbf{r}'|^2}\frac{d}{dt}|\mathbf{r}'|^2\right|.$$

Так как

$$\frac{d}{dt}|\mathbf{r}'|^2 = 2\langle \mathbf{r}', \mathbf{r}''\rangle,$$

то

$$k = \frac{1}{|\mathbf{r}'|^2}\left|\mathbf{r}'' - \mathbf{r}'\frac{\langle \mathbf{r}', \mathbf{r}''\rangle}{|\mathbf{r}'|^2}\right|.$$

Далее:

$$\frac{1}{|\mathbf{r}'|^2}\left(\mathbf{r}'' - \mathbf{r}'\frac{\langle \mathbf{r}', \mathbf{r}''\rangle}{|\mathbf{r}'|^2}\right) =$$
$$= \frac{1}{|\mathbf{r}'|^2}\left(\frac{x''(y')^2 - x'y'y''}{(x')^2 + (y')^2}\right)\mathbf{e}_1 + \frac{1}{|\mathbf{r}'|^2}\left(\frac{y''(x')^2 - y'x'x''}{(x')^2 + (y')^2}\right)\mathbf{e}_2.$$

Отсюда $k = \dfrac{|x''y' - y''x'|}{\left((x')^2 + (y')^2\right)^{3/2}}$. ■

Оказывается, производные от векторов репера Френе удовлетворяют простым соотношениям, называемым “формулами Френе” .

Теорема 2.2 (Формулы Френе). *Если кривая отнесена к натуральному параметру, то выполнены равенства*

$$\frac{d\mathbf{v}(s)}{ds} = k(s)\mathbf{n}(s) \qquad \frac{d\mathbf{n}(s)}{ds} = -k(s)\mathbf{v}(s).$$

Доказательство. Первая формула сразу следует из определения кривизны. Далее, в силу определения имеем $\langle \mathbf{n}(s), \mathbf{n}(s)\rangle = 1$. Из леммы 2.2 получаем

$$\left\langle \mathbf{n}, \frac{d}{ds}\mathbf{n} \right\rangle = 0,$$

т.е.

$$\frac{d\mathbf{n}(s)}{ds} = \lambda(s)\mathbf{v}(s),$$

где $\lambda(s)$ — некоторая гладкая функция. Найдем ее. Дифференцируя тождество $\langle \mathbf{v}, \mathbf{n}\rangle = 0$, получаем

$$\left\langle \frac{d\mathbf{v}}{ds}, \mathbf{n} \right\rangle + \left\langle \mathbf{v}, \frac{d\mathbf{n}}{ds} \right\rangle = 0,$$

откуда $k\langle \mathbf{n}, \mathbf{n}\rangle + \langle \mathbf{v}, \lambda\mathbf{v}\rangle = 0$, т.е. $k = -\lambda$. ■

Рис. 2.1:

Формулы Френе можно переписать в виде

$$\begin{pmatrix} d\mathbf{v}/ds \\ d\mathbf{n}/ds \end{pmatrix} = \begin{pmatrix} 0 & k \\ -k & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{n} \end{pmatrix},$$

Это соотношение имеет прозрачный геометрический смысл. Рассмотрим в точке s репер $\omega(s) = \big(\mathbf{v}(s), \mathbf{n}(s)\big)$ и сместимся по кривой $\mathbf{r}(s)$ в бесконечно близкую точку $s + \Delta s$ (рис. 2.1). Перенеся параллельно репер $\omega(s + \Delta s)$ в точку s, получим в ней два репера: $\omega(s)$ и $\omega(s + \Delta s)$, причем $\omega(s + \Delta s)$ получается из репера $\omega(s)$ поворотом на бесконечно малый угол $\Delta\varphi$. Следовательно, реперы $\omega(s + \Delta s)$, $\omega(s)$ связаны ортогональным преобразованием: $\omega(s + \Delta s) = A(\Delta s)\omega(s)$, где $A(\Delta s) = \begin{pmatrix} \cos\Delta\varphi & \sin\Delta\varphi \\ -\sin\Delta\varphi & \cos\Delta\varphi \end{pmatrix}$. Разлагая функции $\cos\Delta\varphi$, $\sin\Delta\varphi$ в ряды по малому приращению $\Delta\varphi$ и пренебрегая членами второго порядка малости по $\Delta\varphi$, получаем

$$A(\Delta\varphi) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & \Delta\varphi \\ -\Delta\varphi & 0 \end{pmatrix} + \ldots,$$

т.е.

$$\omega(s+\Delta s) = \omega(s) + \begin{pmatrix} 0 & \Delta\varphi \\ -\Delta\varphi & 0 \end{pmatrix} \omega(s) + \ldots$$

Отсюда

$$\frac{d}{ds}\omega(s) = \begin{pmatrix} 0 & \dfrac{d\varphi(s)}{ds} \\ -\dfrac{d\varphi(s)}{ds} & 0 \end{pmatrix} \omega(s),$$

где $\varphi(s)$ — угол поворота репера $\omega(s)$ относительно какого-то фиксированного репера на плоскости (скажем, относительно репера $\omega(0)$). В то же время, из формул Френе следует, что $\dfrac{d}{ds}\omega(s) = \begin{pmatrix} 0 & k(s) \\ -k(s) & 0 \end{pmatrix} \omega(s)$. Сравнивая полученные матрицы, видим, что $k(s) = d\varphi(s)/ds$. Итак, кривизна кривой в точке s равна скорости изменения угла $\varphi(s)$ в этой точке (т.е. скорости поворота репера Френе). В случае плоской кривой функция $k(s)$ полностью определяет кривую, если только $k \neq 0$ для всех s. Более точно, имеет место

Теорема 2.3. *Пусть задана гладкая функция $k(s)$, не обращающаяся в нуль для всех таких s, что $a \leqslant s \leqslant b$. Тогда на плоскости существует гладкая кривая $\mathbf{r}(s)$, определяемая однозначно с точностью до параллельного переноса и ортогонального преобразования, для которой $k(s)$ является кривизной, а s — натуральным параметром.*

Доказательство. Рассмотрим систему дифференциальных уравнений

$$\begin{pmatrix} d\mathbf{v}(s)/ds \\ d\mathbf{n}(s)/ds \end{pmatrix} = \begin{pmatrix} 0 & k(s) \\ -k(s) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v}(s) \\ \mathbf{n}(s) \end{pmatrix},$$

где $k(s)$ — заданная функция. Так как $k(s) \neq 0$, то по теореме существования и единственности из теории дифференциальных уравнений, эта система имеет единственное (при фиксированных начальных данных) решение, гладкое на всем интервале $a < s < b$. Следовательно, можно рассмотреть уравнение $\dfrac{d\mathbf{n}(s)}{ds} = -k(s)\mathbf{v}(s)$.

По тем же причинам, это уравнение имеет решение, однозначно определенное при задании начальных данных на всем интервале $a < s < b$. (Проверьте, что решение $\mathbf{r} = \mathbf{r}(s)$ является искомой кривой!) Здесь любые начальные данные на плоскости можно совместить параллельным переносом и поворотом. ■

Какова роль условия $k(s) \neq 0$? Рассмотрим гладкую кривую, кривизна которой в некоторой точке $s = s_0$ обращается в нуль вместе с производными всех порядков. Существование таких кривых вытекает из теоремы 2.3, примененной к гладкой функции $k(s)$, имеющей нуль бесконечного порядка в одной из концевых точек, a или b, где $a \leqslant s \leqslant b$. Сопрягая две такие кривые двумя разными способами в их общей концевой точке, где кривизна обращается в нуль вместе со всеми производными, мы и получаем две гладкие кривые, имеющие совпадающие кривизны, но не конгруэнтные (см. рис. 2.2).

Рис. 2.2:

2.1.2 Теория пространственных кривых. Кривизна и кручение

Рассмотрим гладкую кривую $\mathbf{r}(t)$ в пространстве $\mathbb{R}^3$, отнесенном к декартовым координатам x, y, z, т.е. $\mathbf{r}(t) = \big(x(t), y(t), z(t)\big)$. Как и в плоском случае, с каждой точкой кривой $\mathbf{r}(s)$ можно однозначно связать репер Френе, гладко меняющийся вдоль кривой при изменении натурального параметра s.

Определение 2.3. *Кривизной* пространственной кривой называется модуль ускорения в параметре s: $k(s) = |\ddot{\mathbf{r}}(s)|$. *Радиусом кривизны* называется величина, обратная кривизне: $R = k^{-1}$.

Мы знаем из леммы 2.2, что векторы скорости $\mathbf{v}(s) = \dot{\mathbf{r}}(s)$ и ускорения $\ddot{\mathbf{r}}(s)$ ортогональны, так как $|\mathbf{v}| = 1$. Однако в трехмерном случае векторов $\dot{\mathbf{r}}$ и $\ddot{\mathbf{r}}$ не хватает, чтобы составить полный репер пространства, даже если $\ddot{\mathbf{r}} \neq 0$. Кроме того, одной кривизны в трехмерном пространстве не хватает для того, чтобы охарактеризовать геометрические свойства кривой. Представьте себе, например, кривую, намотанную на цилиндр:

$$x = R\cos t, \qquad y = R\sin t, \qquad z = t$$

(винтовую линию). Кроме кривизны, она еще закручивается по третьему направлению. Третий вектор базиса можно взять ортогональным к $\dot{\mathbf{r}}$ и $\ddot{\mathbf{r}}$.

Определение 2.4. Регулярная кривая называется, *бирегулярной* если ее кривизна всюду отлична от нуля.

Итак, пусть $\mathbf{r}(s)$ — натурально параметризованная бирегулярная кривая в $\mathbb{R}^3$. Обозначим через $\mathbf{n}$ нормированный вектор ускорения $\ddot{\mathbf{r}}(s)/\big|\ddot{\mathbf{r}}(s)\big|$, который также называется *главной нормалью* к $\mathbf{r}$ в точке $\mathbf{r}(s)$; а через $\mathbf{b}$ — векторное произведение $\mathbf{v} \times \mathbf{n}$, называемое *бинормалью* к $\mathbf{r}$ в точке $\mathbf{r}(s)$. Мы получаем ортонормированный репер $(\mathbf{v}, \mathbf{n}, \mathbf{b})$, называемый *репером Френе*.

Продифференцируем бинормаль $\mathbf{b}$. По формуле Лейбница, имеем

$$\dot{\mathbf{b}} = \frac{d}{ds}(\mathbf{v} \times \mathbf{n}) = \dot{\mathbf{v}} \times \mathbf{n} + \mathbf{v} \times \dot{\mathbf{n}} = \mathbf{v} \times \dot{\mathbf{n}},$$

так как $\dot{\mathbf{v}} = \ddot{\mathbf{r}}$ колинеарен $\mathbf{n}$, и, значит, $\dot{\mathbf{v}} \times \mathbf{n} = 0$. Так как $|\mathbf{n}| = 1$, $\dot{\mathbf{n}}$ перпендикулярен $\mathbf{n}$, поэтому он параллелен плоскости, натянутой на векторы $\mathbf{v}$ и $\mathbf{b}$. Следовательно, векторное произведение $\mathbf{v} \times \dot{\mathbf{n}}$ перпендикулярно этой плоскости и, значит, коллинеарно с $\mathbf{n}$. Мы приходим к следующему заключению: векторы $\dot{\mathbf{b}}$ и $\mathbf{n}$ коллинеарны. Положим $\varkappa = -\langle \dot{\mathbf{b}}, \mathbf{n} \rangle$. Таким образом, $\dot{\mathbf{b}} = -\varkappa \mathbf{n}$. Величина $\varkappa$ называется *кручением* кривой $\mathbf{r}$ в точке $\mathbf{r}(s)$. Ясно, что кручение $\varkappa$ является гладкой функцией параметра s.

Теорема 2.4 (Формулы Френе). *Для любой пространственной кривой $\mathbf{r} = \mathbf{r}(s)$, где s — натуральный параметр, имеют место следующие формулы Френе*:

$$\begin{cases} \dot{\mathbf{v}} = k\,\mathbf{n}, \\ \dot{\mathbf{n}} = -k\,\mathbf{v} + \varkappa\,\mathbf{b}, \\ \dot{\mathbf{b}} = -\varkappa\,\mathbf{n}. \end{cases}$$

Доказательство. Первое и третье уравнения — это фактически определения кривизны и кручения. Осталось доказать второе уравнение.

Так как $\dot{\mathbf{n}}$ перпендикулярно $\mathbf{n}$, то $\dot{\mathbf{n}} = \alpha\,\mathbf{v} + \beta\,\mathbf{b}$. В силу ортонормальности репера Френе, имеем: $\alpha = \langle \dot{\mathbf{n}}, \mathbf{v} \rangle$ и $\beta = \langle \dot{\mathbf{n}}, \mathbf{b} \rangle$. Мы должны показать, что $\alpha = -k$, а $\beta = \varkappa$.

Так как $\langle \mathbf{n}, \mathbf{v} \rangle = \langle \mathbf{n}, \mathbf{b} \rangle = 0$, то, дифференцируя эти равенства, заключаем, что

$$0 = \langle \dot{\mathbf{n}}, \mathbf{v} \rangle + \langle \mathbf{n}, \dot{\mathbf{v}} \rangle = \alpha + k, \quad 0 = \langle \dot{\mathbf{n}}, \mathbf{b} \rangle + \langle \mathbf{n}, \dot{\mathbf{b}} \rangle = \beta - \varkappa,$$

что и требовалось. Теорема доказана. ■

Приведем теперь формулы для вычисления кривизны и кручения кривой в произвольной параметризации.

Теорема 2.5. *Пусть $\mathbf{r}(t)$ — бирегулярная кривая. Тогда*

$$k = \frac{|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}|}{|\dot{\mathbf{r}}|^3}, \quad \varkappa = \frac{(\dot{\mathbf{r}}, \ddot{\mathbf{r}}, \dddot{\mathbf{r}})}{|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}|^2},$$

где через (u, v, w) мы обозначили смешенное произведение векторов u, v и w из $\mathbb{R}^3$.

Доказательство. Пусть сначала $t = s$ — натуральный параметр. Тогда

$$\dot{\mathbf{r}} = \mathbf{v}, \quad |\dot{\mathbf{r}}| = 1, \quad \ddot{\mathbf{r}} = k\,\mathbf{n}, \quad |\dot{\mathbf{r}} \times \ddot{\mathbf{r}}| = k,$$
$$\dddot{\mathbf{r}} = \dot{k}\,\mathbf{n} + k(-k\,\mathbf{v} + \varkappa\,\mathbf{b}), \quad (\dot{\mathbf{r}}, \ddot{\mathbf{r}}, \dddot{\mathbf{r}}) = k^2\varkappa,$$

поэтому в этом случае теорема имеет место.

Пусть кривая $\mathbf{r}(t)$ произвольно параметризована. Обозначая натуральный параметр через s, получаем

$$\dot{\mathbf{r}}_s = \mathbf{r}'_t\,t_s, \quad \ddot{\mathbf{r}}_{ss} = \mathbf{r}''_{tt}\,t_s^2 + \mathbf{r}'_t\,t_{ss}, \quad \dddot{\mathbf{r}}_{sss} = \mathbf{r}'''_{ttt}\,t_s^3 + 3\mathbf{r}''_{tt}\,t_s\,t_{ss} + \mathbf{r}'_t\,t_{sss},$$

откуда

$$\dot{\mathbf{r}}_s \times \ddot{\mathbf{r}}_{ss} = \mathbf{r}'_t \times \mathbf{r}''_{tt}\, t_s^3, \quad (\dot{\mathbf{r}}_s, \ddot{\mathbf{r}}_{ss}, \dddot{\mathbf{r}}_{sss}) = (\mathbf{r}'_t, \mathbf{r}''_{tt}, \mathbf{r}'''_{ttt})\, t_s^6,$$

что мгновенно влечет

$$\frac{|\dot{\mathbf{r}}_s \times \ddot{\mathbf{r}}_{ss}|}{|\dot{\mathbf{r}}_s|^3} = \frac{|\mathbf{r}'_t \times \mathbf{r}''_{tt}|}{|\mathbf{r}'_t|^3}, \qquad \frac{(\dot{\mathbf{r}}_s, \ddot{\mathbf{r}}_{ss}, \dddot{\mathbf{r}}_{sss})}{|\dot{\mathbf{r}}_s \times \ddot{\mathbf{r}}_{ss}|^2} = \frac{(\mathbf{r}'_t, \mathbf{r}''_{tt}, \mathbf{r}'''_{ttt})}{|\mathbf{r}'_t \times \mathbf{r}''_{tt}|^2}.$$

Теорема доказана. ■

В заключение следует отметить, что кривизна и кручение — полный набор геометрических инвариантов кривой в евклидовом пространстве. Более точно:

— Если для пространственной кривой известны функции $k = k(s)$ и $\varkappa = \varkappa(s)$, то можно восстановить кривую с точностью до движения всего пространства. Эта пара уравнений называется *натуральным уравнением пространственной кривой.*

Приведем другое доказательство формул Френе. Докажем сначала вспомогательное утверждение о дифференцировании матричных функций.

Рассмотрим гладкую кривую в линейном пространстве матриц, т.е. однопараметрическое семейство $A(t)$, где $A(t)$ — $(n \times n)$-матрица, коэффициенты которой суть гладкие функции по t, $t \in (-a, a)$. Предположим, что все матрицы $A(t)$ ортогональны с определителем $+1$, а также, что $A(0) = E$, где E — единичная матрица.

Лемма 2.3. *Пусть $X = \dot{A}(t)|_{t=0}$ — производная этого однопараметрического семейства ортогональных матриц $A(t)$ при $t = 0$, т.е. X — это матрица, составленная из функций вида $\left.\dfrac{da_{ij}(t)}{dt}\right|_{t=0}$, где $A(t) = \big(a_{ij}(t)\big)$. Тогда X — кососимметрическая матрица.*

Доказательство. Ортогональная матрица $A(t)$, действуя как оператор, сохраняет евклидово скалярное произведение. Поэтому любых двух векторов $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ имеет место тождество $\langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$. Слева стоит гладкая функция от t. Вычислим ее производную при $t = 0$. Получим $\langle \dot{A}\mathbf{x}, A\mathbf{y}\rangle|_{t=0} + \langle A\mathbf{x}, \dot{A}\mathbf{y}\rangle|_{t=0} = 0$, т.е. $\langle X\mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{x}, X\mathbf{y}\rangle = 0$, что и означает косую симметрию матрицы X. ■

Матричнозначную функцию $A(t)$ можно разложить в окрестности точки $t = 0$ по степеням бесконечно малого приращения Δt:

$$A(\Delta t) = E + \left.\frac{dA(t)}{dt}\right|_{t=0} \Delta t + \ldots$$

Тогда матрица X появляется как "коэффициент" при Δt.

Пусть $\mathbf{e}_1(s) = \mathbf{v}(s)$, $\mathbf{e}_2(s) = \mathbf{n}(s)$, $\mathbf{e}_3(s) = \mathbf{b}(s)$ — зависящий от натурального параметра s репер, связанный с данной пространственной кривой. Тогда репер $\mathbf{e}_1(s + \Delta s)$, $\mathbf{e}_2(s + \Delta s)$, $\mathbf{e}_3(s + \Delta s)$ получается из репера $\mathbf{e}_1(s)$,

$\mathbf{e}_2(s)$, $\mathbf{e}_3(s)$ ортогональным преобразованием, так как оба эти репера ортонормированы:

$$\begin{aligned} &\mathbf{e}_i(s+\Delta s) = \sum_{j=1}^{3} a_{ij}(s,\Delta s)\mathbf{e}_j(s) \qquad (i=1,2,3), \\ &\sum_{i=1}^{3} a_{ij}(s,\Delta s)a_{ik}(s,\Delta s) = \delta_{jk}. \end{aligned} \tag{2.1}$$

Дифференцируя равенство (2.1) по Δs и полагая $\Delta s = 0$, мы получаем

$$\dot{\mathbf{e}}_i(s) = \sum_{j=1}^{3} b_{ij}\mathbf{e}_j(s),$$

где матрица (b_{ij}) кососимметрическая в силу леммы 2.3. Поэтому матрица $B = (b_{ij})$ имеет вид

$$B = \begin{pmatrix} 0 & b_{12} & b_{13} \\ -b_{12} & 0 & b_{23} \\ -b_{13} & -b_{23} & 0 \end{pmatrix}.$$

По определению

$$\dot{\mathbf{e}}_1 = \frac{d\,\mathbf{v}}{d\,s} = k\mathbf{n} = k\mathbf{e}_2,$$

поэтому $b_{12} = k$, $b_{13} = 0$. Значит матрица B имеет вид

$$B = \begin{pmatrix} 0 & k & 0 \\ -k & 0 & b_{23} \\ 0 & -b_{23} & 0 \end{pmatrix}.$$

Обозначив здесь $\varkappa = b_{23}$, мы получаем формулы Френе.

Если кривая плоская, то вектор бинормали является постоянным вектором, не меняется при изменении точки на кривой; в частности, кручение кривой равно нулю. Следовательно, плоские кривые могут быть охарактеризованы (с точки зрения $\mathbb{R}^3$) как кривые с нулевым кручением.

Изучим подробнее роль кручения пространственной кривой, когда оно отлично от нуля. Рассмотрим скольжение репера $(\mathbf{v}, \mathbf{n}, \mathbf{b})$ вдоль кривой и будем проектировать векторы $\mathbf{v}(s)$, $\mathbf{n}(s)$ и $\mathbf{b}(s)$ на плоскость, натянутую на векторы $\mathbf{b}(s_0)$ и $\mathbf{n}(s_0)$, где s_0 — некоторое фиксированное значение параметра s, а значения s предполагаются бесконечно близкими к s_0. При этом вектор скорости спроектируется в вектор бесконечно малой длины, а потому можно считать (в первом приближении), что этот вектор проектируется в нуль. Тогда в плоскости $\mathbf{b}(s_0)$, $\mathbf{n}(s_0)$ возникает некоторое движение векторов $\mathbf{b}(s)$, $\mathbf{n}(s)$ спроектированных на эту плоскость. Из формул Френе следует, что это движение описывается формулами $\dfrac{d\mathbf{n}}{ds} = \varkappa\mathbf{b}$, $\dfrac{d\mathbf{b}}{ds} = -\varkappa\mathbf{n}$, т.е. движение определяется кососимметрической матрицей $\begin{pmatrix} 0 & \varkappa \\ -\varkappa & 0 \end{pmatrix}$, что определяет бесконечно малый поворот репера $\mathbf{b}$, $\mathbf{n}$. Следовательно, векторы $\mathbf{b}$,

$\mathbf{n}$ вращаются относительно вектора скорости кривой, причем скорость вращения однозначно задается кручением кривой (отсюда, кстати, происходит и сам термин "кручение"). При этом кривая теряет свою плоскую форму и становится "изогнутой в пространстве", если она первоначально была плоской. Итак, можно считать, что пространственная кривая (в малом) может быть получена из плоской кривой, если двигаться по ней с равномерной скоростью, "подкручивая" в каждый момент времени эту кривую с помощью кручения $\varkappa$ (рис. 2.3).

Задачи

1. Доказать, что если у плоской кривой кривизна тождественно равна нулю, то это — прямая линия.

2. Доказать, что в трехмерном пространстве плоские кривые характеризуются условием $\varkappa \equiv 0$.

3. Доказать, что прямые в трехмерном пространстве характеризуются условиями $k \equiv 0$, $\varkappa \equiv 0$.

4. Описать класс кривых, для которых $k = \mathrm{const}$, $\varkappa = \mathrm{const}$.

Рис. 2.3:

5. Доказать, что траектории движения материальной точки в центральном силовом поле являются плоскими кривыми.

2.2 Поверхности. Первая и вторая квадратичные формы

2.2.1 Определение поверхностей

Определение 2.5. *Непрерывной параметрической поверхностью* V^k размерности k в n-мерном евклидовом пространстве $\mathbb{R}^n$, $n \geqslant k$, называется произвольное непрерывное отображение $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ из некоторой области $\Omega \subset \mathbb{R}^k$ в $\mathbb{R}^n$. Напомним, что *областью* мы называем произвольное открытое связное множество в $\mathbb{R}^n$.

Пусть $x^1, \ldots, x^n$ — стандартные евклидовы координаты в $\mathbb{R}^n$. Тогда наша поверхность V^k задается набором из n *координатных функций* $x^i(u^1, \ldots, u^k)$, где u^i — стандартные евклидовы координаты в $\mathbb{R}^k \supset \Omega$. При этом непрерывность $\mathbf{r}$ равносильна непрерывности всех функций $x^i(u^1, \ldots, u^k)$. Координаты u^i называются *параметрами* для $\mathbf{r}$ или *координатами* на V^k.

Определение 2.6. Непрерывная параметрическая поверхность $\mathbf{r}\colon \Omega \to \mathbb{R}^n$ называется *гладкой*, если все задающие ее координатные функции $x^i(u^1, \ldots, u^k)$ — гладкие. В этом случае определены векторы $\mathbf{r}_{u^i} =$

$\left(x^1_{u^i}, \dots, x^n_{u^i}\right)$, называемые *базисными касательными векторами для* $\mathbf{r}$ *в точке* $\mathbf{r}(u^1, \dots, u^k)$. Здесь через $x^i_{u^j}$ мы обозначили частные производные функций x^i по параметрам u^j. Гладкая параметрическая поверхность называется *регулярной*, если ее базисные касательные векторы всюду линейно независимы. Это равносильно тому, что матрица Якоби $\left(x^i_{u_j}\right)$ имеет во всех точках поверхности максимальный ранг, равный, таким образом, k.

Часто поверхность можно задать и другими способами. Опишем два таких способа. Для простоты мы ограничимся рассмотрением регулярных двумерных поверхностей в $\mathbb{R}^3$, называя их просто *поверхностями*.

Пусть (x^1, x^2, x^3) — стандартные координаты в $\mathbb{R}^3$

1) Первый способ — это задание поверхности V^2 в виде графика гладкой функции $x^3 = f(x^1, x^2)$, где $f\colon \mathbb{R}^2 \supset \Omega \to \mathbb{R}$ — гладкая функция. Будем говорить, что поверхность V^2 *задана графиком функции* $x^3 = f(x^1, x^2)$.

2) Другой способ — это задание поверхности V^2 уравнением $F(x^1, x^2, x^3) = 0$, где $F\colon \mathbb{R}^3 \supset \Omega \to \mathbb{R}$ — неявная функция.

Теорема 2.6. *Для каждой точки P регулярной поверхности V^2 в некоторой окрестности точки P поверхность V^2 можно задать каждым из перечисленных трех способов: в виде параметрической поверхности, в виде графика функции и с помощью неявной функции с отличным от нуля в точке P градиентом.*

Доказательство. а) Пусть $P = (u^1_0, u^2_0)$ и имеется параметрическое задание регулярной поверхности V^2: $x^i = x^i(u^1, u^2)$, $x^i_0 = x^i(u^1_0, u^2_0)$ и $\operatorname{rank}\left(\dfrac{\partial x^i}{\partial u^j}\right) = 2$, $i = 1, 2, 3$, $j = 1, 2$. Без ограничения общности предположим, что в точке P матрица

$$\begin{pmatrix} x^1_{u^1} & x^1_{u^2} \\ x^2_{u^1} & x^2_{u^2} \end{pmatrix}$$

невырождена. По теореме об обратной функции, в некоторой окрестности U точки $\left(x^1(P), x^2(P)\right)$ существуют такие функции $u^1 = u^1(x^1, x^2)$ и $u^2 = u^2(x^1, x^2)$, что

$$x^1 = x^1\left(u^1(x^1, x^2), u^2(x^1, x^2)\right), \quad x^2 = x^2\left(u^1(x^1, x^2), u^2(x^1, x^2)\right), \quad \text{и}$$
$$u^1 = u^1\left(x^1(u^1, u^2), x^2(u^1, u^2)\right), \quad u^2 = u^2\left(x^1(u^1, u^2), x^2(u^1, u^2)\right).$$

Получили локальное представление V^2 в виде графика $x^3 = f(x^1, x^2) = x^3\left(u^1(x^1, x^2), u^2(x^1, x^2)\right)$, $x^3_0 = f(x^1_0, x^2_0)$, около точки (x^1_0, x^2_0, x^3_0).

б) Пусть поверхность V^2 задается графиком некоторой гладкой функции $x^3 = f(x^1, x^2)$. Тогда такую поверхность можно задать с помощью неявной функции $F(x^1, x^2, x^3) = x^3 - f(x^1, x^2)$, причем $\operatorname{grad} F \neq 0$.

в) Покажем теперь, что поверхность $V^2 = \{(x^1, x^2, x^3) | F(x^1, x^2, x^3) = 0\}$, где F — неявная функция с $\operatorname{grad} F\big|_P \neq 0$, можно задать в виде графика функции.

Без ограничения общности, будем предполагать, что $F_{x^3}(P) \neq 0$. Тогда, по теореме о неявной функции, в некоторой окрестности U точки $P = (x_0^1, x_0^2, x_0^3)$ множество $V^2 \cap U$ задается в виде графика некоторой гладкой функции $x^3 = f(x^1, x^2)$, определенной в некоторой окрестности W точки $(x_0^1, x_0^2) \in \mathbb{R}^2$. В частности, для всех $(x^1, x^2) \in W$ выполняется $F\big(x^1, x^2, f(x^1, x^2)\big) = 0$.

г) Осталось рассмотреть случай, когда поверхность представлена в виде графика гладкой функции $x^3 = f(x^1, x^2)$, определенной на области $\Omega \subset \mathbb{R}^2$, а мы хотим задать поверхность параметрически.

Тогда гладкое отображение $\mathbf{r}\colon \Omega \to \mathbb{R}^3$, определенное так:

$$\mathbf{r}\colon (u^1, u^2) \mapsto \big(u^1, u^2, f(u^1, u^2)\big),$$

задает регулярную поверхность, так как векторы $r_{u^1} = (1, 0, f_{u^1})$ и $r_{u^2} = (0, 1, f_{u^2})$ — линейно независимы в каждой точке области Ω. Теорема доказана. ■

2.2.2 Первая квадратичная форма

Рассмотрим в пространстве $\mathbb{R}^3$ регулярную поверхность V^2 размерности 2. Выше мы обсуждали различные способы задания поверхностей. Выберем параметрическое задание. Сейчас мы будем интересоваться лишь локальными свойствами поверхности. Поэтому можно считать, что поверхность V^2 является *вложенной*, т.е. наша поверхность V^2 задается гладким радиус-вектором $\mathbf{r} = \mathbf{r}(u^1, u^2)$, причем $\mathbf{r}$ задает взаимно-однозначное отображение с образом и вектора $\mathbf{r}_{u^1}$, $\mathbf{r}_{u^2}$ — линейно независимы. Вложенность поверхности позволяет отождествить поверхность с ее образом. Поэтому, говоря, что "точка P лежит на поверхности V^2", мы будем под P понимать как точку из Ω, так и точку $\mathbf{r}(P)$ из $\mathbb{R}^n$.

Говорят, что *кривая* $\gamma\colon [0,1] \to \mathbb{R}^3$ *лежит на* V^2, если образ отображения γ принадлежит V^2. Кривая $\big(u^1(t), u^2(t)\big)$, заданная в координатах (u^1, u^2), определяет кривую $\mathbf{r}(t) = \mathbf{r}\big(u^1(t), u^2(t)\big)$, лежащую на нашей поверхности в $\mathbb{R}^3$. Пусть $P \in V^2$ — произвольная точка, и $\gamma(0) = P$, т.е. γ проходит через P. Вектор скорости

$$\dot\gamma(0) = \frac{d}{dt}\mathbf{r}\big(u^1(t), u^2(t)\big)\Big|_{t=0} = \mathbf{r}_{u^1}(P)\,\dot u^1(0) + \mathbf{r}_{u^2}(P)\,\dot u^2(0) \tag{2.2}$$

кривой γ в точке P назовем *касательным вектором к* V^2 *в точке* P. Из (2.2) следует, что любой касательный вектор к V^2 в точки P есть линейная комбинация линейно независимых векторов $\mathbf{r}_{u^1}(P)$, $\mathbf{r}_{u^2}(P)$ (мы рассматриваем регулярные поверхности). Следовательно, векторы, касающиеся поверхности в данной точке, образуют двумерное пространство с базисом $\mathbf{r}_{u^1}(P)$, $\mathbf{r}_{u^2}(P)$, которое и называется *касательной плоскостью к* V^2 *в* P и обозначается через $T_P V^2$. Пара $\big(r_{u^1}(P), r_{u^2}(P)\big)$ называется *каноническим базисом* касательной плоскости $T_P V^2$, а координаты $\dot u^1(0)$ и $\dot u^2(0)$ вектора скорости $\dot\gamma(0)$ — *координатами касательного вектора* $\dot\gamma(0)$ к поверхности V^2 в точке P по отношению к системе координат (u^1, u^2) на V^2.

Пусть $ds^2 = \sum\limits_{i=1}^{3}(dx^i)^2$ — евклидова метрика; тогда возникает следующая квадратичная форма:

$$ds^2(V^2) = \sum_{i=1}^{3}\sum_{k,p=1}^{2}\frac{\partial x^i}{\partial u^k}\frac{\partial x^i}{\partial u^p}du^k du^p = \sum_{k,p=1}^{2} g_{kp}(u)du^k du^p,$$

$$g_{kp} = \left\langle \frac{\partial \mathbf{r}}{\partial u^k}, \frac{\partial \mathbf{r}}{\partial u^p}\right\rangle;$$

здесь через $\langle\ ,\ \rangle$ обозначено скалярное произведение в $\mathbb{R}^3$.

Определение 2.7. *Первой квадратичной формой поверхности* V^2, или *индуцированной метрикой на поверхности* V^2, или *первой фундаментальной формой поверхности* V^2 в $\mathbb{R}^3$ называется форма $ds^2(V^2) = \sum\limits_{k,p} g_{kp}du^k du^p$, где функции $g_{kp}(u^1, u^2)$ определены выше.

Первая квадратичная форма обычно записывается в виде $ds^2(V^2) = E(du^1)^2 + 2Fdu^1du^2 + G(du^2)^2$, где $E = \langle \mathbf{r}_{u^1}, \mathbf{r}_{u^1}\rangle$, $F = \langle \mathbf{r}_{u^1}, \mathbf{r}_{u^2}\rangle$, $G = \langle \mathbf{r}_{u^2}, \mathbf{r}_{u^2}\rangle$ — коэффициенты формы, выраженные через компоненты радиус-вектора $\mathbf{r}$:

$$E = (x^1_{u^1})^2 + (x^2_{u^1})^2 + (x^3_{u^1})^2, \quad F = x^1_{u^1}x^1_{u^2} + x^2_{u^1}x^2_{u^2} + x^3_{u^1}x^3_{u^2},$$
$$G = (x^1_{u^2})^2 + (x^2_{u^2})^2 + (x^3_{u^2})^2.$$

Напомним, что метрика $ds^2(V^2)$ называется *конформно-евклидовой*, если $E = G$ и $F = 0$. Первая квадратичная форма определена на векторах, касательных к V^2. Более точно: если $\mathbf{a}, \mathbf{b} \in T_P V^2$ — два касательных вектора, то определено скалярное произведение $\langle \mathbf{a}, \mathbf{b}\rangle_{ds^2(V^2)} = g_{kp}a^k b^p = \sum\limits_{k,p=1}^{2} g_{kp}a^k b^p$.

Напомним, что для упрощения обозначений мы опускаем знак суммирования, когда оно ведется по совпадающим верхним и нижним индексам. Ясно, что скалярное произведение $\langle \mathbf{a}, \mathbf{b}\rangle_{ds^2(V^2)}$ совпадает с евклидовым скалярным произведением векторов $\mathbf{a}$ и $\mathbf{b}$, рассматриваемых как векторы объемлющего пространства $\mathbb{R}^3$. Симметричная матрица метрического тензора G, составлена из функций g_{kp}.

Длина дуги $\gamma(t)$, расположенной в V^2, выражается по формуле $\int\limits_a^b \sqrt{\langle \dot\gamma, \dot\gamma\rangle}\,dt$, т.е.

$$l = \int_a^b \sqrt{\sum_{k,p=1}^{2} g_{kp}\big(u(t)\big)\frac{du^k(t)}{dt}\frac{du^p(t)}{dt}}\,dt.$$

Через метрику также естественно записывается площадь поверхности. Так как площадь не является локальным свойством поверхности, мы считаем поверхность вложенной. Обозначим через $\Pi(P)$, где $P \in \Omega$ — произвольная точка, параллелограмм, натянутый на векторы $\mathbf{r}_{u^1}(P), \mathbf{r}_{u^2}(P)$, а через $|\Pi(P)|$ — его площадь.

Определение 2.8. *Площадью параметрической поверхности* $\mathbf{r}$ называется следующая величина:

$$\mathrm{vol}(V^2) = \int\limits_{\Omega} |\Pi(P)|\, du^1\, du^2.$$

Если $\mathbf{r}_{u^1}$ и $\mathbf{r}_{u^2}$ — канонический базис в точке P, то $\Pi(P) = |\mathbf{r}_{u^1} \times \mathbf{r}_{u^2}| = \sqrt{\det(g_{ij})} = \sqrt{EG - F^2}$, поэтому

$$\mathrm{vol}(V^2) = \int\limits_{\Omega} |\mathbf{r}_{u^1} \times \mathbf{r}_{u^2}|\, du^1\, du^2 = \int\limits_{\Omega} \sqrt{g}\, du^1\, du^2 =$$
$$= \int\limits_{\Omega} \sqrt{EG - F^2}\, du^1\, du^2,$$

где $g = \det(g_{ij})$.

Какой вид приобретает первая квадратичная форма для различных способов задания поверхности? Пусть V^2 задана в виде графика $x^3 = f(x^1, x^2)$. Имеем:

$$ds^2|_{V^2} = (dx^1)^2 + (dx^2)^2 + \big(dx^3(x^1, x^2)\big)^2 =$$
$$= \sum_{k,p=1}^{2} \left(\delta_{kp} + \frac{\partial f}{\partial x^k}\frac{\partial f}{\partial x^p}\right) dx^k dx^p.$$

Иногда будем обозначать $\dfrac{\partial f}{\partial x^\alpha}$ через f_{x^α}.

Пусть поверхность V^2 задана с помощью неявной функции, т.е. в виде $F(x^1, x^2, x^3) = 0$, где $\dfrac{\partial F}{\partial x^3} \neq 0$. Тогда по теореме о неявной функции существует локальное решение уравнения $F(x^1, x^2, x^3) = 0$ вида $x^3 = f(x^1, x^2)$, причем $\partial f/\partial x^i = -(\partial F/\partial x^i)/(\partial F/\partial x^3)$. Отсюда получаем $\mathrm{G} = (g_{kp})$, где

$$g_{kp} = \left[\left(\frac{\partial F}{\partial x^k}\frac{\partial F}{\partial x^p}\right) \Big/ \left(\frac{\partial F}{\partial x^3}\right)^2\right] + \delta_{kp}.$$

Рассмотрим первую квадратичную форму поверхностей вращения, вложенных в $\mathbb{R}^3$. Введем цилиндрические координаты (r, φ, x^3), и пусть поверхность задана параметрически: $\varphi = u^1$, $x^3 = u^2$, $r = r(u^2)$. Тогда вычисление дает

$$ds^2(V^2) = \big(1 + (r'_{u^2})^2\big)(du^2)^2 + r^2(du^1)^2.$$

Здесь $F = 0$, $E = r^2(u^2)$, $G = 1 + (r'_{u^2})^2$. Тот факт, что $F = 0$, означает, что координаты линии $u^2 = u^2_0 = \mathrm{const}$ и $u^1 = u^1_0 = \mathrm{const}$ ортогональны друг другу в каждой точке (рис. 2.4).

Лемма 2.4. *Пусть V^2 — регулярная поверхность и* G *— первая квадратичная форма. Тогда форма* G *невырождена.*

Рис. 2.4:

Доказательство. В силу определения радиус-вектора $\mathbf{r} = \mathbf{r}(u^1, u^2)$ все векторы $\mathbf{r}_{u^k}$ линейно независимы. Поскольку матрица G составлена из скалярных произведений векторов $\mathbf{r}_{u^k}$ и $\mathbf{r}_{u^p}$, то G невырождена. ■

Определение 2.9. Гладкое взаимно однозначное, имеющее гладкое обратное, отображение (диффеоморфизм) f поверхности V^k на себя называется *движением* римановой метрики g_{ij} или *изометрией*, если риманова метрика при этом отображении переходит в себя, т.е. выполнено тождество

$$g_{kp}(v) = g_{ij}(u(v))\frac{\partial u^i(y)}{\partial v^k}\frac{\partial u^j(y)}{\partial v^p},$$

где $u^1, \ldots, u^k$, $v^1, \ldots, v^k$ — две системы координат на V^k.

2.2.3 Вторая квадратичная форма

Пусть V^2 — поверхность в $\mathbb{R}^3$, заданная радиус-вектором вида $\mathbf{r}(u^1, u^2)$. Выберем направление нормали в точке P, т.е. зафиксируем $\mathbf{n}(P)$ — единичный вектор, ортогональный поверхности V^2 в точке P. Определим квадратичную форму $\mathrm{Q}(\mathbf{a}, \mathbf{a})$, задав ее значения для произвольного вектора $\mathbf{a} \in T_P(V^2)$ следующим образом. Рассмотрим гладкую кривую $\gamma(t)$, лежащую на V^2 и проходящую через точку P, причем так, что $\gamma(0) = P$, $\dot\gamma(0) = \mathbf{a}$. Такая кривая всегда существует, хотя и определена неоднозначно (рис. 2.5). Так как вдоль кривой $\gamma(t)$ радиус-вектор $\mathbf{r}$ является функцией от t, то $\mathbf{a} = d\mathbf{r}(u(t))/dt|_{t=0}$. Рассмотрим вектор-функцию $\dot{\mathbf{r}} = d\mathbf{r}(u(t))/dt$ и ее производную по t, т.е. $\ddot{\mathbf{r}} = d^2\mathbf{r}(u(t))/dt^2$. Обозначим через $\ddot{\mathbf{r}}$ значение $\ddot{\mathbf{r}}$ при $t = 0$. Это — вторая производная $\mathbf{r}$ по направлению вектора $\mathbf{a}$.

Рис. 2.5:

Определение 2.10. Положим $\mathrm{Q}(\mathbf{a}, \mathbf{a}) = \langle \ddot{\mathbf{r}}\ , \mathbf{n} \rangle$.

Определенное нами число является величиной проекции вектора $\ddot{\mathbf{r}}$ на выбранную нормаль $\mathbf{n}$ в точке P (рис. 2.6). Вычислим $\mathrm{Q}(\mathbf{a},\mathbf{a})$ через координаты вектора $\mathbf{r}$. Имеем

$$\frac{d^2\mathbf{r}}{dt^2} = \mathbf{r}_{u^k u^p}\frac{du^k}{dt}\frac{du^p}{dt} + \mathbf{r}_{u^k}\frac{d^2u^k}{dt^2},$$

$$\left\langle \frac{d^2}{dt^2}\mathbf{r}(u(t)), \mathbf{n}\right\rangle = \left\langle \mathbf{r}_{u^k u^p}\frac{du^k}{dt}\frac{du^p}{dt}, \mathbf{n}\right\rangle + 0.$$

Рис. 2.6:

При этом следует помнить, что вектор $\mathbf{a}$ имеет координаты $\left(\dfrac{du^1(0)}{dt}, \dfrac{du^2(0)}{dt}\right)$, т.е. окончательно

$$\mathrm{Q}(\mathbf{a},\mathbf{a}) = \langle \mathbf{n}, \mathbf{r}_{u^k u^p}|_{t=0}\rangle a^k a^p.$$

Эта квадратичная форма однозначно определяет билинейную форму $\mathrm{Q}(\mathbf{a},\mathbf{b})$, значение которой на паре произвольных векторов $\mathbf{a},\mathbf{b} \in T_P(V^2)$ определяется так:

$$\mathrm{Q}(\mathbf{a},\mathbf{b}) = q_{kp}a^k b^p,$$

где $q_{kp} = \langle \mathbf{r}_{u^k u^p}|_{t=0}, \mathbf{n}\rangle$.

Лемма 2.5. *Выражение* $\mathrm{Q}(\mathbf{a},\mathbf{b}) = q_{kp}a^k b^p$, *где* $\mathbf{a},\mathbf{b} \in T_P(V^2)$ *определяет билинейную форму.*

Доказательство. Имеем

$$q_{k'p'} = \langle \mathbf{r}_{u^{k'}u^{p'}}|_{t=0}, \mathbf{n}\rangle = \frac{\partial u^k}{\partial u^{k'}}\frac{\partial u^p}{\partial u^{p'}}\left\langle \frac{\partial^2\mathbf{r}(0)}{\partial u^k \partial u^p}, \mathbf{n}\right\rangle = \frac{\partial u^k}{\partial u^{k'}}\frac{\partial u^p}{\partial u^{p'}}q_{kp},$$

т.е. функции q_{kp} преобразуются при замене координат как коэффициенты билинейной формы, что и требовалось доказать. ■

Определение 2.11. Билинейная форма $\mathrm{Q}(\mathbf{a},\mathbf{b})$ называется *второй квадратичной формой поверхности* $V^2 \subset \mathbb{R}^3$.

Форма Q не является инвариантной при изометриях V^2 в $\mathbb{R}^3$.

Пример 1. Пусть V^2 — двумерная плоскость в $\mathbb{R}^3$; тогда радиус-вектор $\mathbf{r}(u,v)$ можно считать линейной функцией от u, v. Следовательно, первая форма является евклидовой метрикой: $du^2 + dv^2$ (рис. 2.7). Теперь рассмотрим изометрическое преобразование V^2 — свертывание плоскости $\mathbb{R}^2$ в цилиндр, ось которого параллельна оси Ox (рис. 2.8). Ясно, что вторая форма цилиндра отлична от нуля, так как отлично от нуля число $\langle \mathbf{r}_{vv}, \mathbf{n}\rangle$. В то же время вторая форма плоскости $\mathbb{R}^2$ тождественно равна нулю. Следовательно, при изгибании (т.е. при изометрии) вторая форма изменилась.

Рис. 2.7:

Рис. 2.8:

Рассмотрим фиксированную поверхность V^2 в $\mathbb{R}^3$. В каждой ее точке P определена первая форма G и вторая форма Q. С этой парой связан набор инвариантов, позволяющих изучать V^2 независимо от ее параметризации. Пусть G и Q — матрицы форм. Рассмотрим полином по переменной λ: $\det(\mathrm{Q}-\lambda\,\mathrm{G})$. Поскольку форма G невырождена, то существует матрица G^{-1}. Следовательно, уравнение $\det(\mathrm{G}^{-1}\,\mathrm{Q} - \lambda E) = 0$ эквивалентно уравнению $\det(\mathrm{Q} - \lambda\,\mathrm{G}) = 0$.

Обозначим через λ_1, λ_2 собственные числа матрицы $\mathrm{G}^{-1}\,\mathrm{Q}$, т.е. корни уравнения $\det(\mathrm{Q} - \lambda\,\mathrm{G}) = 0$. Вскоре мы докажем, что все они вещественны. Запишем характеристический полином $F(\lambda)$ в виде $\sum\limits_{k=0}^{2} \sigma_k \lambda^k$, где σ_k — симметрические функции от корней λ_1, λ_2.

Лемма 2.6. *Функции σ_k являются инвариантами пары форм* G *и* Q*, т.е. сохраняются при произвольной невырожденной замене координат в окрестности точки $P \in V^2$.*

Доказательство. При регулярной замене координат $x \to x'$ в касательном пространстве $T_P V^2$ возникает индуцированное линейное невырожденное преобразование с помощью матрицы Якоби: J. При этом $\mathrm{G} \to J\,\mathrm{G}\,J^\top = \mathrm{G}'$; $\mathrm{Q} \to J\mathrm{Q}J^\top = \mathrm{Q}'$. Следовательно,

$$\det\bigl((\mathrm{G}')^{-1}\mathrm{Q}' - \lambda E\bigr) = \det\bigl((J^\top)^{-1}(\mathrm{G}^{-1}\,\mathrm{Q} - \lambda E)J^\top\bigr) = \det(\mathrm{G}^{-1}\,\mathrm{Q} - \lambda E),$$

что и требовалось доказать. ■

Нас будут интересовать инварианты

$$\sigma_1 = \lambda_1 + \lambda_2 = \mathrm{Tr}(\mathrm{G}^{-1}\,\mathrm{Q}); \quad \sigma_2 = \lambda_1\lambda_2 = \det(\mathrm{G}^{-1}\,\mathrm{Q}).$$

Определение 2.12. Функция $H(P) = \sigma_1(P) = \sigma_1(\lambda_1, \lambda_2)$ называется *средней кривизной* поверхности V^2. Функция $K(P) = \sigma_2(P) = \sigma_2(\lambda_1, \lambda_2)$ называется *гауссовой кривизной* поверхности V^2.

Теорема 2.7. *Все собственные числа λ_1, λ_2 пары форм* G, Q *вещественны. В том случае, когда эти числа различны, собственные векторы* $\mathbf{e}_1, \mathbf{e}_2$ *матрицы* $\mathrm{G}^{-1}\,\mathrm{Q}$ *взаимно ортогональны, как относительно объемлющей евклидовой метрики в $\mathbb{R}^3$, так и относительно римановой метрики, индуцированной на V^2 вложением $V^2 \to \mathbb{R}^3$.*

Доказательство. По известной теореме из алгебры собственные числа симметрической матрицы вещественны, и все ее собственные векторы взаимно ортогональны для различных собственных чисел. Пока что эта теорема не применима, так как матрица $\mathrm{G}^{-1}\,\mathrm{Q}$, вообще говоря, не симметрична. Так как форма $\mathrm{G}(P)$ симметрична, то в некоторой окрестности точки P существует регулярная замена координат такая, что в одной точке P форма $\mathrm{G}(P)$ приводится к диагональному виду. Приведя G к диагональному виду, можно затем привести ее к единичной матрице, применяя растяжения вдоль главных осей формы. Пусть A — линейный оператор $A\colon T_PV^2 \to T_PV^2$, приводящий G к единичной матрице. Тогда $\mathrm{G} = AEA^{\top} = AA^{\top}$. Матрица E определяет ортобазис φ_1, φ_2 в T_PV^2. Получаем $\det(\mathrm{Q} - \lambda\,\mathrm{G}) = \det\bigl(A(A^{-1}\mathrm{Q}(A^{-1})^{\top} - \lambda E)A^{\top}\bigr)$.

Рассмотрим форму $\widetilde{\mathrm{Q}} = B\mathrm{Q}B^{\top}$, где $B = A^{-1}$. Исходное уравнение $\det(\mathrm{Q} - \lambda\,\mathrm{G}) = 0$ запишется в базисе φ_1, φ_2 в виде $\det(\widetilde{\mathrm{Q}} - \lambda E) = 0$, так как $\det A \neq 0$. При этом $\widetilde{\mathrm{Q}}^{\top} = \widetilde{\mathrm{Q}}$, так как $\mathrm{Q}^{\top} = \mathrm{Q}$. Следовательно, все собственные числа и собственные векторы у формы $\widetilde{\mathrm{Q}}$ и матрицы $\mathrm{G}^{-1}\,\mathrm{Q}$ одни и те же. Так как форма $\widetilde{\mathrm{Q}}$ симметрична, то все ее собственные числа (т.е. λ_1, λ_2) вещественны, и в случае, когда все они попарно различны, все ее собственные векторы взаимно ортогональны. Это следует из известной теоремы алгебры. ■

Определение 2.13. Направления, задаваемые векторами $\mathbf{e}_1$, $\mathbf{e}_2$ (определенными однозначно, если $\lambda_1 \neq \lambda_2$) называются *главными направлениями поверхности* V^2 в точке P, или *главными осями.*

Итак, мы связали с каждой точкой поверхности V^2 однозначно определенный (если $\lambda_1 \neq \lambda_2$) с точностью до выбора знака ортобазис $\mathbf{e}_1$, $\mathbf{e}_2$, гладко зависящий от P.

Поскольку в ортобазисе $\mathbf{e}_1$, $\mathbf{e}_2$ матрица G становится единичной, то λ_1, λ_2 совпадают с собственными числами формы Q, записанной в ортобазисе $\mathbf{e}_1$, $\mathbf{e}_2$.

Рис. 2.9:

Рассмотрим частный случай. Пусть V^2 задана в виде графика $x^3 = f(x^1, x^2)$ и плоскость $T_P V^2$ в точке P совпадает с плоскостью переменных x^1, x^2 (рис. 2.9). Тогда нормаль $\mathbf{n}$ к V^2 в точке P имеет координаты $(0, 0, 1)$; радиус-вектор $\mathbf{r}$, описывающий поверхность V^2, имеет вид $\mathbf{r} = \big(x^1, x^2, f(x^1, x^2)\big)$. Так как плоскость $(x^1, x^2) = T_P V^2$ касается V^2 в точке P, то $\left.\dfrac{\partial f}{\partial x^i}\right|_P = 0$; отсюда $\mathrm{G}(P) = E$, так как $g_{ij} = f_{x^i} f_{x^j} + \delta_{ij}$.

Рассмотрим матрицу $\mathrm{Q} = (q_{ij})$, где

$$q_{ij} = \langle \mathbf{r}_{x^i x^j}, \mathbf{n} \rangle = \frac{\partial^2 f(P)}{\partial x^i \partial x^j}.$$

Итак, $\mathrm{Q} = \big(f_{x^i x^j}(P)\big)$ совпадает с матрицей гессиана функции f в точке P. Средняя кривизна $H(P)$ имеет вид

$$H(P) = f_{x^1 x^1} + f_{x^2 x^2} = \Delta f,$$

где Δ — оператор Лапласа, для гауссовой кривизны $K(P)$ получаем $K(P) = f_{x^1 x^1} f_{x^2 x^2} - f^2_{x^1 x^2}$.

2.2.4 Локальная теория гладких кривых на поверхности. Теорема Менье. Формула Эйлера

Пусть $\mathbf{n}(P)$ — нормаль к поверхности V^2. Рассмотрим двумерную плоскость $\mathbb{R}^2$, проходящую через P и пересекающую V^2 по некоторой гладкой кривой $\gamma(t) = \mathbb{R}^2 \cap V^2$. Мы рассматриваем здесь малую окрестность точки.

Определение 2.14. Гладкая кривая $\gamma(t) = \mathbb{R}^2 \cap V^2$ называется *плоским сечением поверхности.*

Пусть γ — плоское сечение поверхности V^2. Зафиксируем содержащую его плоскость $\mathbb{R}^2$. Предположим, что начало координат (точка O), из которого выходит радиус-вектор $\mathbf{r}(u^1, u^2)$, описывающий поверхность V^2, попало в плоскость $\mathbb{R}^2$. Пусть $\mathbf{m}$ — главная нормаль к плоской кривой $\gamma = \mathbb{R}^2 \cap V^2$, содержащийся в $\mathbb{R}^2$ (рис. 2.10). Нормали $\mathbf{n}$ и $\mathbf{m}$, вообще говоря, не совпадают.

Рис. 2.10:

Введем на кривой $\gamma = \mathbb{R}^2 \cap V^2$ натуральный параметр s. Тогда в плоскости $\mathbb{R}^2$ кривая $\gamma(s)$ задается радиус-вектором $\mathbf{r}(s) = \mathbf{r}\big(u^1(s), u^2(s)\big)$. В силу формул Френе для плоских кривых имеем $\dfrac{d^2\mathbf{r}}{ds^2} = \mathbf{m}k$, где $k = \left|\dfrac{d^2\mathbf{r}}{ds^2}\right|$ — кривизна кривой. С другой стороны, если $\mathbf{a} = \dfrac{d}{ds}\mathbf{r}(s)$ — вектор скорости кривой γ, то, в силу определения второй формы Q, имеем

$$\mathrm{Q}(\mathbf{a}, \mathbf{a}) = \langle k\mathbf{m}, \mathbf{n}\rangle = k\cos\theta,$$

где θ — угол между нормалями $\mathbf{m}$ и $\mathbf{n}$ (рис. 2.10). С другой стороны,

$$\mathrm{Q}(\mathbf{a}, \mathbf{a}) = \left(\frac{ds}{dt}\right)^{-2} \mathrm{Q}\left(\frac{d\mathbf{r}}{dt}, \frac{d\mathbf{r}}{dt}\right),$$

где t — произвольный гладкий параметр вдоль кривой γ. Здесь $\dfrac{d\mathbf{r}}{dt} = \mathbf{a}$ — произвольный касательный вектор к кривой γ в точке P. Если $\mathbf{a} = (a^1, a^2)$, то

$$k\cos\theta = \frac{q_{ij}a^i a^j}{g_{ij}a^i a^j}.$$

Следовательно, нами доказана

Теорема 2.8 (Менье). *Для любого касательного вектора $\mathbf{a} \in T_P V^2$ и любого плоского сечения γ (такого, что $\dot\gamma = \mathbf{a}$) отношение второй формы к первой равно $k\cos\theta$, т.е.*

$$\frac{\mathrm{Q}(\mathbf{a}, \mathbf{a})}{\mathrm{G}(\mathbf{a}, \mathbf{a})} = \frac{q_{ij}a^i a^j}{g_{ij}a^i a^j} = k\cos\theta,$$

где k — кривизна данного плоского сечения.

Среди плоских сечений выделен класс нормальных сечений.

Определение 2.15. Плоское сечение $\gamma = \mathbb{R}^2 \cap V^2$ называется *нормальным*, если $\mathbf{n}(P) \in \mathbb{R}^2$, т.е. $\theta = 0$ или π.

Каждое нормальное сечение в точке P однозначно задается касательным вектором $\mathbf{a} \in T_P V^2$. Плоскость $\mathbb{R}^2$, определяющая сечение, натянута на нормаль $\mathbf{n}(P)$ и вектор $\mathbf{a}$. Вращая плоскость $\mathbb{R}^2$ вокруг $\mathbf{n}(P)$, получаем все нормальные сечения поверхности в точке P.

Определение 2.16. *Кривизной нормального сечения* γ поверхности V^2 в точке P будем называть кривизну кривой γ со знаком плюс, если главная нормаль к γ в точке P сонаправлена с выбранным направлением нормали к поверхности, и со знаком минус в противном случае.

Для кривизн нормального сечения γ доказанная в теореме 2.8 формула принимает вид

$$k = \frac{q_{ij} a^i a^j}{g_{ij} a^i a^j}.$$

Если известна кривизна нормального сечения $k(\mathbf{a})$, то кривизна любого другого плоского сечения γ (вдоль $\mathbf{a}$) находится из формулы $k(\theta, \mathbf{a}) = \dfrac{1}{\cos\theta} k(\mathbf{a})$, где θ — угол между главной нормалью к сечению γ и $\mathbf{n}$. (В последней формуле $k(\mathbf{a})$ — кривизна нормального сечения с приписанным знаком, а $k(\theta, \mathbf{a})$— обычная неотрицательная кривизна плоского сечения.)

Теперь напомним, что в касательной плоскости $T_P V^2$ всегда определены главные направления $\mathbf{e}_1$, $\mathbf{e}_2$, определяющиеся однозначно, когда $\lambda_1 \neq \lambda_2$. Рассмотрим эти "главные оси" $\mathbf{e}_1$, $\mathbf{e}_2$ и построим по каждому из них нормальное плоское сечение $\gamma_i = \mathbb{R}^2_i \cap V^2$, где плоскость $\mathbb{R}^2_i$ натянута на $\mathbf{n}$, $\mathbf{e}_i$. Все главные направления $\mathbf{e}_i$ взаимно ортогональны в евклидовой метрике на $T_P V^2$. Сечения γ_i называются *главными нормальными сечениями.*

Определение 2.17. *Главными кривизнами* поверхности V^2 называются кривизны $k_i = k(\mathbf{e}_i)$ главных нормальных сечений γ_i, т.е. кривизны сечений как кривых с приписанным знаком плюс, если главная нормаль к сечению сонаправлена с выбранным направлением нормали к поверхности, и со знаком минус в противном случае.

Таким образом, главные кривизны поверхности также как и кривизны нормальных сечений к поверхности не обязаны быть неотрицательными. Они могут отличаться знаком от кривизн соответствующих кривых (рассматриваемых как кривые в $\mathbb{R}^3$ без "привязки" к поверхности).

Теорема 2.9. *Собственные числа λ_1, λ_2 совпадают с главными кривизнами k_1, k_2.*

Доказательство. Из теоремы 2.8 имеем $k \cos\theta = \dfrac{q_{ij} a^i a^j}{g_{ij} a^i a^j}$. Для нормального сечения вдоль $\mathbf{a}$ получаем $k(\mathbf{a}) = \dfrac{q_{ij} a^i a^j}{g_{ij} a^i a^j}$. Фиксируем в $T_P V^2$ ортобазис

$\mathbf{e}_1, \mathbf{e}_2$. Тогда $g_{ij} = \delta_{ij}$, $q_{ij} = \delta_{ij}\lambda_i$; следовательно,

$$k(\mathbf{a}) = \frac{\sum\limits_{i=1}^{2} \lambda_i (a^i)^2}{\sum\limits_{i=1}^{2} (a^i)^2}.$$

Если $\mathbf{a} \in T_P V^2$ совпадает с одним из $\mathbf{e}_i$, то $k_i = \lambda_i$, что и требовалось. ■

Рассмотрим в $T_P V^2$ произвольный вектор $\mathbf{a}$ и рассмотрим нормальное сечение вдоль $\mathbf{a}$. Обозначим через φ_i углы, образуемые $\mathbf{a}$ с главными направлениями $\mathbf{e}_1$, $\mathbf{e}_2$.

Предложение 2.1. *Для нормального сечения вдоль произвольного вектора $\mathbf{a} \in T_P V^2$ выполнена формула Эйлера*

$$k(\mathbf{a}) = \sum_{i=1}^{2} \lambda_i \cos^2 \varphi_i.$$

Доказательство. В силу теоремы 2.9 имеем

$$k(\mathbf{a}) = \sum_{i=1}^{2} \lambda_i \frac{(a_i)^2}{\left(\sqrt{\sum\limits_{i=1}^{2} (a^i)^2}\right)^2} = \sum_{i=1}^{2} \lambda_i \cos^2 \varphi_i,$$

где

$$\cos \varphi_i = \frac{a_i}{|\mathbf{a}|}. \qquad ■$$

Формула Эйлера позволяет установить "экстремальные свойства" главных кривизн λ_1, λ_2. Рассмотрим кривизну нормального сечения $k(\mathbf{a})$ как функцию от $\mathbf{a} \in T_P V^2$. Положив $x^i = \cos \varphi_i$, можно считать, что $k(\mathbf{a}) = k(x^1, x^2) = \lambda_1 (x^1)^2 + \lambda_2 (x^2)^2$ является функцией на окружности S^1, так как $(x^1)^2 + (x^1)^2 = 1$. В окрестности каждой точки окружности можно ввести локальную координату y.

Назовем точку $x_0 \in S^1$ *критической* для функции $f(x)$, заданной на S^1, если $\left.\dfrac{\partial f}{\partial y}\right|_{x_0} = 0$.

Вопрос: какие критические точки имеет на сфере гладкая функция $k(\mathbf{a})$? И, кроме того, какие значения принимает кривизна $k(\mathbf{a})$ в этих критических точках?

Теорема 2.10. *Критическими точками кривизны $k(\mathbf{a})$ на сфере $S^1 \in T_P V^2$ являются в точности точки вида $\pm\mathbf{e}_i$, $i = 1, 2$, в том случае, когда $\lambda_1 \neq \lambda_2$. В точках $\pm\mathbf{e}_i$ функция $k(\mathbf{a})$ принимает значения, равные λ_i. В этом смысле главные кривизны являются экстремальными значениями функции кривизны $k(\mathbf{a})$. Если $\lambda_1 = \lambda_2$, то критическими точками функции $k(\mathbf{a})$ является вся окружность S^1 и $k(\mathbf{a}) = \mathrm{const}$.*

Рис. 2.11:

Доказательство. Формула Эйлера принимает вид $k(\alpha) = \lambda_1 \cos^2 \varphi_1 + \lambda_2 \cos^2 \varphi_2$, где $\cos^2 \varphi_1 + \cos^2 \varphi_2 = 1$. Пусть $\lambda_1 \geqslant \lambda_2$, тогда $k(\mathbf{a}) = (\lambda_1 - \lambda_2) \cos^2 \varphi_1 + \lambda_2$. Ясно, что минимум $k(\mathbf{a})$ равен λ_2, максимум $k(\mathbf{a})$ равен λ_1. Если $\lambda_1 = \lambda_2$, то $k(\mathbf{a}) = \lambda_1 = \lambda_2$ (рис. 2.11). Теорема доказана. ■

Квадрикой нормальных кривизн называется следующая поверхность: $\lambda_1 \cos^2 \varphi_1 + \lambda_2 \cos^2 \varphi_2 = \mathrm{const}$. Направления, соответствующие экстремальным (критическим) значениям кривизны, совпадают с главными осями квадрики. Если все главные кривизны неотрицательны, то квадрика является эллипсом.

Гауссова кривизна $K(P)$ описывает важные локальные свойства поверхности. Поскольку $K(P) = \lambda_1 \lambda_2$, то выделяются три случая: а) $K > 0$; б) $K < 0$; в) $K = 0$.

Достаточно рассмотреть только случаи: а) $\lambda_1 > 0, \lambda_2 > 0$; б) $\lambda_1 > 0, \lambda_2 < 0$; в) $\lambda_1 = 0, \lambda_2 \neq 0$; $\lambda_1 = \lambda_2 = 0$. В случае а) поверхность выпукла, то есть локально расположена по одну сторону от касательной плоскости в точке P (рис. 2.12). В случае б) поверхность локально имеет вида седла (рис. 2.13). Такие точки называются седловыми или точками перевала. В случае в), при $\lambda_2 \neq 0, \lambda_1 = 0$, поверхность V^2 имеет (локально, в первом приближении) вид "цилиндра" (рис. 2.14). Если обе главные кривизны нулевые, такая точка называется точкой уплощения поверхности.

Рис. 2.12:

Рис. 2.13:

Рис. 2.14:

2.2.5 Гауссова и средняя кривизны в точке на поверхности

Пусть поверхность $V^2 \subset \mathbb{R}^3$ задана в виде графика $z = f(x, y)$, где (x, y, z) — декартовы координаты. Пусть $f(0, 0) = 0$; тогда координатная плоскость (x, y) — касательная к V^2 в точке $(0, 0)$. Так как

$$\mathrm{G}\,|_{(0,0)} = E, \quad \mathrm{Q}|_{(0,0)} = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix},$$

то $K = f_{xx}f_{yy} - f_{xy}^2$; $H = f_{xx} + f_{yy}$. Теперь вычислим H и K в точках, отличных от $(0, 0)$.

Пусть $\mathrm{G} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}$, $\mathrm{Q} = \begin{pmatrix} L & M \\ M & N \end{pmatrix}$. Тогда прямой подсчет дает:

$$H = \frac{(GL - 2MF + EN)}{EG - F^2}, \qquad K = \frac{LN - M^2}{EG - F^2}.$$

Если $z = f(x, y)$, то $ds^2 = (1 + f_x^2)dx^2 + 2f_xf_y dx\, dy + (1 + f_y^2)dy^2$. Так как радиус-вектор поверхности имеет вид $\mathbf{r} = \big(x, y, f(x, y)\big)$, то

$$\mathbf{r}_{xx} = (0, 0, f_{xx}), \quad \mathbf{r}_{xy} = (0, 0, f_{xy}), \quad \mathbf{r}_{yy} = (0, 0, f_{yy}),$$
$$\mathbf{n} = \frac{\operatorname{grad}(z - f)}{|\operatorname{grad}(z - f)|} = \frac{(-f_x, -f_y, 1)}{\sqrt{1 + f_x^2 + f_y^2}},$$

Отсюда

$$H = \frac{(1 + f_y^2)f_{xx} - 2f_xf_yf_{xy} + (1 + f_x^2)f_{yy}}{(1 + f_x^2 + f_y^2)^{3/2}}, \qquad K = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}.$$

Гауссова и средняя кривизны не зависят от выбора локальных координат на поверхности.

Легко подсчитать, что для сферы $K = \dfrac{1}{R^2}$, $H = \dfrac{2}{R}$, в частности, ее гауссова и средняя кривизны постоянны.

Гауссова и средняя кривизны двумерной плоскости равны нулю.

Определение 2.18. Двумерная поверхность $V^2 \subset \mathbb{R}^3$ называется *поверхностью постоянной кривизны*, если ее гауссова кривизна K постоянна.

Определение 2.19. Двумерная поверхность $V^2 \subset \mathbb{R}^3$ называется *поверхностью положительной, нулевой, отрицательной кривизны*, если соответственно гауссова кривизна поверхности во всех ее точках положительна, равна нулю, отрицательна.

Стандартная двумерная сфера является поверхностью положительной (постоянной) кривизны.

Рис. 2.15:

Задача. Докажите, что эллипсоид $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$ является поверхностью строго положительной кривизны, если полуоси a, b, c отличны от 0 и ∞. Эллипсоид можно записать параметрически так:

$$x = a\cos\theta\cos\varphi, \quad y = b\cos\theta\sin\varphi, \quad z = \sin\theta.$$

Евклидова плоскость является поверхностью нулевой постоянной кривизны . Пример поверхности отрицательной кривизны дает график $z = x^2 - y^2$. Ясно, что

$$K = \frac{-4}{\left(1 + 4(x^2 + y^2)\right)^2} < 0.$$

Здесь гауссова кривизна является переменной функцией. Легко построить поверхность постоянной отрицательной кривизны $V^2 \subset \mathbb{R}^3$. Сейчас мы укажем такую поверхность. Тем самым мы докажем утверждение: *в трехмерном евклидовом пространстве существуют (локально) поверхности постоянной положительной, нулевой и отрицательной кривизны.*

Рассмотрим на плоскости кривую γ, расположенную в первом квадранте, со следующим свойством: длина отрезка касательной от точки касания с кривой до точки пересечения касательной с осью Ox постоянна и равна a (рис. 2.15). При перемещении точки A по кривой γ точка B скользит по оси Ox, и отрезок AB имеет постоянную длину, равную a.

Найдем уравнение кривой γ. Из треугольника ABx (рис. 2.15) имеем $\operatorname{tg}\varphi = -y'_x$, где $y = y(x)$ — график γ; $a\sin\varphi = y$. Отсюда

$$x'_y = -\frac{\sqrt{a^2 - y^2}}{y},$$

где $x = x(y)$ — график γ. Итак,

$$x(y) = -\int\limits_y^a \frac{1}{y}\sqrt{a^2 - y^2}\,dy = -\sqrt{a^2 - y^2} + \frac{a}{2}\ln\left(\frac{a + \sqrt{a^2 - y^2}}{a - \sqrt{a^2 - y^2}}\right).$$

Рассмотрим поверхность вращения, получающуюся вращением кривой γ вокруг оси Ox (рис. 2.16). Она называется *поверхностью Бельтрами*, или

Рис. 2.16:

Рис. 2.17:

псевдосферой. Найдем ее гауссову кривизну. Сделаем это для поверхности вращения общего вида.

Рассмотрим в $\mathbb{R}^3(x, y, z)$ поверхность V^2, образованную вращением вокруг оси Ox кривой $x = x(y)$ (т.е. образующей), расположенной в плоскости Oxy. На V^2 возникает ортогональная координатная сеть: параллели и меридианы (рис. 2.17).

Лемма 2.7. *В каждой точке поверхности вращения главные направления, т.е. отвечающие главным кривизнам λ_1, λ_2, всегда можно считать совпадающими с направлениями меридиана и параллели, проходящих через эту точку.*

Замечание. Если $\lambda_1 \neq \lambda_2$, то главные направления определены однозначно и совпадают с направлениями меридиана и параллели. Если $\lambda_1 = \lambda_2$, то любое направление главное.

Доказательство. Главными направлениями в $T_P V^2$ являются те ортобазисы $\mathbf{e}_1$, $\mathbf{e}_2$, в которых обе формы G и Q диагональны. Ясно, что первая форма диагональна в координатах (u, v) на поверхности, порожденных меридианами и параллелями. Докажем, что вторая форма также диагональна. Надо доказать, что $M = \langle \mathbf{r}_{uv}, \mathbf{n} \rangle \equiv 0$, где $\mathbf{r}(u, v)$ — радиус-вектор V^2, $\mathbf{n}$ — нормаль к V^2. Рассмотрим цилиндрические координаты (r, φ, x) в $\mathbb{R}^3$.

Рис. 2.18:

Пусть V^2 задается образующей $r = r(x)$. Тогда радиус-вектор $\mathbf{r}(x, \varphi)$ поверхности V^2 имеет вид $\mathbf{r} = \big(x, r(x)\cos\varphi, r(x)\sin\varphi\big)$ (рис. 2.18). Нормаль к V^2 имеет вид $\mathbf{n} = (r', -\cos\varphi, -\sin\varphi)/\sqrt{1+(r')^2}$. Ясно, что $\langle \mathbf{r}_{x\varphi}, \mathbf{n}\rangle = 0$, что и требовалось. ■

Лемма 2.8. *Модуль гауссовой кривизны K поверхности вращения имеет вид*

$$|K| = \frac{|r''|}{r\big(1+(r')^2\big)^2},$$

где $r = r(x)$ — образующая поверхности вращения в цилиндрических координатах.

Доказательство. В силу леммы 2.7 главные направления совпадают с направлениями меридиана и параллели в точке P. Поэтому $K(P) = \lambda_1\lambda_2$, где λ_2 и λ_1 — кривизны плоских кривых — меридиана и нормального сечения вдоль направления параллели соответственно. Так как меридиан совпадает с образующей, то $\lambda_2(x) = \dfrac{r''}{\big(1+(r')^2\big)^{3/2}}$, что следует из формул Френе.

Найдем $\lambda_1(x)$ для нормального сечения вдоль направления параллели в точке $P = \big(x, r(x)\big)$. Кривизна $k(\theta, \mathbf{a})$ самой параллели-окружности (вдоль касательного вектора $\mathbf{a}$ к параллелью) очевидно равна $1/r(x)$. Здесь θ — это угол между нормалью $\mathbf{n}$ к V^2 и вектором $\mathbf{m}$, расположенным в плоскости параллели (рис. 2.19). Напомним, что

$$k(\theta, \mathbf{a}) = \frac{1}{\cos\theta}k(\mathbf{a}).$$

Здесь $k(\mathbf{a})$ — интересующая нас кривизна λ_1. Так как в плоскости xOr вектор $\mathbf{m}$ имеет координаты $(0,1)$, а вектор $\mathbf{n}$ — координаты $\dfrac{1}{\sqrt{1+(r')^2}}(-r', 1)$, то $\cos\theta = \dfrac{1}{\sqrt{1+(r')^2}}$. Следовательно, $\lambda_1(x) = \dfrac{1}{r\sqrt{1+(r')^2}}$, что и завершает доказательство. ■

Рис. 2.19:

Лемма 2.9. *Поверхность Бельтрами является поверхностью постоянной отрицательной кривизны.*

Доказательство. Поскольку поверхность Бельтрами является поверхностью вращения, то можно воспользоваться леммой 2.8. Функция $y = y(x)$ является обратной функцией к функции $x = x(y)$, найденной нами ранее. Как было уже подсчитано, $x'_y = \dfrac{-\sqrt{a^2 - y^2}}{y}$. Отсюда $x'' = \dfrac{a^2}{r^2\sqrt{a^2 - y^2}}$. Прямое вычисление дает, что $K = -\dfrac{1}{a^2}$. Знак минус возник потому, что кривая $y = y(x)$ выпукла вниз. ■

Итак, мы указали в $\mathbb{R}^3$ поверхности постоянной положительной, нулевой и отрицательной гауссовой кривизны. Сфера имеет постоянную положительную кривизну. Поверхностью нулевой гауссовой кривизны являются плоскость или конус, образованный прямыми, исходящими из одной точки (конечной или бесконечной) и скользящими по произвольной гладкой кривой. Предъявленная нами поверхность отрицательной постоянной кривизны не является замкнутой поверхностью, поскольку имеет край — окружность (рис. 2.16). Можно показать (мы не будем этого делать), что поверхность Бельтрами нельзя продолжить за эту окружность с сохранением условия $K(P) = -\dfrac{1}{a^2} < 0$. Обычно эту поверхность "достраивают", добавляя поверхность, симметричную исходной относительно плоскости yOz (рис. 2.20). Но эта поверхность имеет "окружность излома", состоящую из особых точек.

Найдем индуцированную риманову метрику на "воронке" Бельтрами. Введем в $\mathbb{R}^3$ цилиндрические координаты (x, r, φ), где $x = x$, $y = r\cos\varphi$, $z = r\sin\varphi$ (т.е. ось Ox — ось вращения). Тогда индуцированная метрика на поверхности вращения с образующей, задаваемой функцией $x = x(r)$, имеет вид

$$ds^2 = \big(1 + (x'_r)^2\big)dr^2 + r^2 d\varphi^2.$$

В нашем примере $x'_r = -\dfrac{\sqrt{a^2 - r^2}}{r}$ (см. выше), т.е. $ds^2 = \dfrac{dr^2}{r^2} + r^2 d\varphi^2$.

Рис. 2.20:

Рис. 2.21:

Предложение 2.2. *Риманова метрика, индуцированная на поверхности Бельтрами объемлющей евклидовой метрикой, является метрикой Лобачевского.*

Доказательство. Рассмотрим замену $u = \varphi$, $v = \dfrac{1}{r}$. Тогда $ds^2 = \dfrac{v^2}{v^4}\,dv^2 + \dfrac{du^2}{v^2} = \dfrac{du^2 + dv^2}{v^2}$, что, очевидно, и доказывает утверждение. ■

Итак, локально поверхность Бельтрами изометрична плоскости Лобачевского, т.е. мы построили изометричное вложение некоторой области плоскости Лобачевского в трехмерное евклидово пространство. Можно описать — какая именно часть плоскости Лобачевского допускает изометричное вложение в $\mathbb{R}^3$ в виде поверхности Бельтрами, разрезанной вдоль ее образующей (рис. 2.21). Соответствующая ей область на плоскости Лобачевского (в модели Пуанкаре) показана на рис. 2.22. Она отмечена символами ∞, A_0, B_0 и заключена между двумя параллельными прямыми геометрии Лобачевского, выходящими из одной точки ∞ на абсолюте, и дугой A_0B_0, являющейся дугой длины 2π (если $a = 1$), частью окружности (в евклидовом смысле), касающейся точки ∞ на абсолюте.

Таким образом, область (∞, A_0, B_0) является бесконечной полосой, заключенной между двумя параллельными прямыми и ограниченной с одного конца дугой A_0B_0.

Рис. 2.22:

Рис. 2.23:

Вопрос. Можно ли изометрично реализовать в $\mathbb{R}^3$, в виде двумерной поверхности постоянной отрицательной кривизны не только полосу (∞, A_0, B_0), но и всю плоскость Лобачевского? Ответ отрицательный (мы не будем здесь этого доказывать).

Рассмотрим теперь среднюю кривизну поверхностей $V^2 \subset \mathbb{R}^3$. Как было отмечено, средняя кривизна зависит от способа вложения $V^2 \subset \mathbb{R}^3$, т.е. определяется как первой, так и второй формами. Как устроены поверхности постоянной средней кривизны? Например, поверхности нулевой средней кривизны — это так называемые *минимальные* поверхности. Они характеризуются тем, что их площадь локально минимальна по сравнению с площадью других поверхностей, отличающихся от исходной только внутри (любого) шара достаточно малого радиуса (рис. 2.23). Физической моделью минимальных поверхностей $V^2 \subset \mathbb{R}^3$ являются "мыльные пленки", возникающие на замкнутых проволочных контурах после их извлечения из мыльной воды. Уравнение $H = 0$ называется уравнением минимальных поверхностей. Найдем его явный вид. Так как $H = \dfrac{GL - 2MF + EN}{EG - F^2}$, то уравнение $H = 0$ имеет вид $GL - 2MF + EN = 0$. Если поверхность задана с помощью графика $z = f(x, y)$, то (см. выше) уравнение приобретает вид:

Рис. 2.24:

$$ds^2 = (1 + f_x^2)\, dx^2 + 2 f_x f_y dx\, dy + (1 + f_y^2)\, dy^2,$$

$$L = \frac{f_{xx}}{\sqrt{1 + f_x^2 + f_y^2}}, \quad M = \frac{f_{xy}}{\sqrt{1 + f_x^2 + f_y^2}}, \quad N = \frac{f_{yy}}{\sqrt{1 + f_x^2 + f_y^2}},$$

следовательно,

$$(1 + f_y^2) f_{xx} - 2 f_{xy} f_x f_y + (1 + f_x^2) f_{yy} = 0.$$

Уравнение довольно сложное и описывать его решения непросто. Конечно, есть простые случаи. Например, евклидова плоскость является минимальной поверхностью, так как $\mathrm{Q} \equiv 0$. Приведем более сложный пример некомпактной минимальной поверхности. Рассмотрим в $\mathbb{R}^3$ две ортогональные, пересекающиеся в точке O прямые l_1 и l_2. Фиксируем прямую l_1 и будем перемещать вдоль нее с постоянной скоростью a прямую l_2, одновременно вращая l_2 с постоянной угловой скоростью ω вокруг оси l_1 (винтовое движение). Прямая l_2 заметает поверхность $V^2 \subset \mathbb{R}^3$, которая называется *прямым геликоидом* (рис. 2.24). Докажите, что прямой геликоид — минимальная поверхность. Отметим, кстати, что любая минимальная поверхность имеет неположительную гауссову кривизну, так как $\lambda_1 + \lambda_2 = 0$.

Приведем еще один пример минимальной поверхности. Рассмотрим некомпактную поверхность, образованную вращением вокруг оси Ox гладкой $\gamma(t)$, задаваемой уравнением $y = a\,\mathrm{ch}(x/a)$. Как известно из курса анализа, эта кривая задает форму провисания тяжелой цепи, закрепленной в двух точках, A и B (рис. 2.25). Соответствующая поверхность называется *катеноидом* (рис. 2.26). Найдем H. Из доказательства леммы 2.8 видно, что

$$H = \frac{1}{y\sqrt{1 + (y')^2}} - \frac{y''}{(1 + (y')^2)^{3/2}} =$$
$$= \left(a\,\mathrm{ch}\,\frac{x}{a}\right)^{-1} \left(1 + \mathrm{sh}^2\,\frac{x}{a}\right)^{-1/2} - \frac{1}{2}\,\mathrm{ch}\,\frac{x}{a}\left(1 + \mathrm{sh}^2\,\frac{x}{a}\right)^{-3/2} = 0.$$

Рис. 2.25:

Рис. 2.26:

Итак, катеноид — минимальная поверхность. Если рассмотреть его часть, заключенную между двумя окружностями, образованными вращением точек A и B вокруг оси Ox, то получим пример минимальной поверхности, натянутой на контур Г, состоящий из этих двух граничных окружностей.

С минимальными поверхностями тесно связаны так называемые гармонические поверхности.

Определение 2.20. Радиус-вектор $\mathbf{r}(u, v)$ называется *гармоническим* относительно координат u, v, если $\dfrac{\partial^2 \mathbf{r}}{\partial u^2} + \dfrac{\partial^2 \mathbf{r}}{\partial v^2} = 0$, т.е. $\Delta \mathbf{r} = 0$, где Δ — оператор Лапласа в координатах u, v.

Радиус-вектор $\mathbf{r}(u, v)$, гармонический в координатах (u, v), не обязан быть гармоническим в других координатах (u', v').

Определение 2.21. Поверхность $V^2 \subset \mathbb{R}^3$ называется *гармонической*, если она задается гармоническим радиус-вектором $\mathbf{r}(u, v)$ в некоторых криволинейных координатах (u, v).

Будем говорить, что радиус-вектор $\mathbf{r}(u, v)$ *минимален*, если средняя кривизна соответствующей поверхности тождественно равна нулю. Так как средняя кривизна не меняется при заменах координат на поверхности, то

радиус-вектор, минимальный относительно одной системы координат, будет минимальным и относительно любой другой регулярной системы координат. Для гармонических поверхностей это уже не так. Поэтому говорят о гармонических отображениях $\mathbf{r}\colon D(u,v) \to \mathbb{R}^3(x,y,z)$, где $D(u,v)$ — область на плоскости (u,v), а $\mathbf{r}(u,v)$ — отображение, задающее поверхность $V^2 \subset \mathbb{R}^3$. Отображение $\mathbf{r}$, гармоническое в одних координатах, уже не будет, вообще говоря, гармоническим в других координатах (приведите пример!). Пример гармонической поверхности: $\mathbf{r}(x,y) = (x, y, x^2 - y^2)$, где x, y, z — декартовы координаты, т.е. поверхность V^2 задается графиком $z = x^2 - y^2$. Ясно, что эта поверхность гармоническая. В то же время она не минимальна: $H = 0$ только в точке $(0,0)$, а в других точках $H \neq 0$.

Напомним, что криволинейные координаты (u,v) на поверхности называются *конформными*, если в них метрика $E\,du^2 + 2F\,dudv + G\,dv^2$ является скалярной, т.е. $E = G$, $F = 0$.

Замечание. Рассмотрим двумерную поверхность, на которой задана риманова метрика ds^2 с вещественно аналитическими коэффициентами E, F, G, рассматриваемыми как функции от локальных криволинейных координат u, v. Тогда для любой точки P существует такая окрестность $U = U(P)$, что в ней можно ввести координаты p, q (являющиеся вещественно аналитическими функциями от исходных координат u, v), в которых ds^2 имеет вид $\lambda(p,q)(dp^2 + dq^2)$, т.е. координаты p, q будут конформными. Доказательство несложно, однако требует ссылки на теорему существования решения специального уравнения в частных производных (уравнения Бельтрами–Лапласа), что выходит за рамки нашего курса.

Как связаны гармонические и минимальные векторы? Пример гармонического, но не минимального вектора мы привели. Минимальный вектор тоже не обязан быть гармоническим. Тем не менее имеет место следующее

Предложение 2.3. *Минимальный вектор, записанный в конформных координатах, является гармоническим.*

Доказательство. Пусть (u,v) — конформные координаты, а $\mathbf{r}(u,v)$ — вектор, задающий минимальную поверхность.

Напомним, что $E = G = \langle \mathbf{r}_v, \mathbf{r}_v \rangle = \langle \mathbf{r}_u, \mathbf{r}_u \rangle$, $F = \langle \mathbf{r}_u, \mathbf{r}_v \rangle$. Требуется доказать, что $\mathbf{r}_{uu} + \mathbf{r}_{vv} = 0$. Положим $\alpha = \mathbf{r}_{uu} + \mathbf{r}_{vv}$. Так как $H = 0$, то $GL - 2MF + EN = 0$, т.е. $L + N = 0$. Отсюда $\langle \alpha, \mathbf{n} \rangle = \langle \mathbf{r}_{uu}, \mathbf{n} \rangle + \langle \mathbf{r}_{vv}, \mathbf{n} \rangle = L + N = 0$, где $\mathbf{n}$ — нормаль к V^2. Осталось доказать, что $\langle \alpha, \mathbf{r}_u \rangle = \langle \alpha, \mathbf{r}_v \rangle = 0$, так как тогда α будет иметь нулевые скалярные произведения с ортогональными векторами $\mathbf{r}_u$, $\mathbf{r}_v$, т.е. α должен быть нулевым вектором. Дифференцируя тождества $E = G$, $F = 0$, получаем $\langle \mathbf{r}_{uu}, \mathbf{r}_u \rangle = \langle \mathbf{r}_{uv}, \mathbf{r}_v \rangle$, $\langle \mathbf{r}_{uv}, \mathbf{r}_u \rangle = \langle \mathbf{r}_{vv}, \mathbf{r}_v \rangle$, $\langle \mathbf{r}_{uu}, \mathbf{r}_v \rangle = -\langle \mathbf{r}_u, \mathbf{r}_{uv} \rangle$, $\langle \mathbf{r}_{vu}, \mathbf{r}_v \rangle = -\langle \mathbf{r}_u, \mathbf{r}_{vv} \rangle$.

Отсюда $\langle \alpha, \mathbf{r}_u \rangle = \langle \mathbf{r}_{uv}, \mathbf{r}_v \rangle - \langle \mathbf{r}_{vu}, \mathbf{r}_v \rangle = 0$. Аналогично, $\langle \alpha, \mathbf{r}_v \rangle = 0$, что и требовалось. ■

Задачи

1. Доказать, что если гауссова и средняя кривизны тождественно равны нулю, то поверхность в $\mathbb{R}^3$ является плоскостью.

2. Пусть поверхность S образована касательными прямыми к кривой. Выразить гауссову и среднюю кривизны поверхности S через кривизну и кручение кривой.

3. В предыдущей задаче показать, что метрика на поверхности зависит только от кривизны кривой.

Глава 3

Общая топология

Топология — это раздел математики, который изучает свойства геометрических объектов, не меняющиеся при "деформации" или при преобразованиях, подобных деформациям.

Первоначально топологические задачи возникли в математическом анализе, где встречается много понятий схожих по своим свойствам и методам исследования. Например, понятия сходимости и предела встречаются в анализе в качестве: а) предела последовательности, б) различных типов пределов функций одной переменной, в) пределов функций многих переменных, г) пределов векторнозначных функций, д) сходимости интегральных сумм. Все эти понятия сходимости и предела основаны на некоторых общих для них приемах исследования, которые интуитивно мы понимаем как близость точек некоторого множества.

Другим важным стимулом для развития топологии было изучение различных типов понятия непрерывности, к которым тесно примыкает понятие сходимости.

Общая топология возникла в результате изучения наиболее общих свойств геометрических пространств и их преобразований, связанных со свойствами сходимости и непрерывности.

3.1 Определения и простейшие свойства метрических и топологических пространств

3.1.1 Метрические пространства

Определение 3.1. *Метрикой* на произвольном абстрактном множестве X называется числовая неотрицательная функция $\rho(x, y)$, зависящая от пары элементов $x, y \in X$, для которой выполнены следующие аксиомы:

а) $\rho(x, y) = 0$ тогда и только тогда, когда $x = y$ (аксиома тождества);

б) $\rho(x, y) = \rho(y, x)$ (аксиома симметрии);

в) для любых трех элементов $x, y, z \in X$ имеет место неравенство $\rho(x, y) \leqslant \rho(x, z) + \rho(z, y)$ (неравенство треугольника).

Множество X, снабженное некоторой метрикой ρ, называется *метрическим пространством*. Элементы множества X называются его *точками*. Значение метрики $\rho(x, y)$ называется *расстоянием между точками* x и y.

Примером метрического пространства является множество вещественных чисел $\mathbb{R}^1$, метрика на котором задается равенством $\rho(x, y) = |x - y|$. Аналогично, n-мерное векторное пространство $\mathbb{R}^n$ с метрикой $\rho(x, y) = \sqrt{\langle x - y, x - y\rangle}$, $x, y \in \mathbb{R}^n$, тоже является примером метрического пространства. Любое подмножество $Y \subset X$ метрического пространства X является метрическим пространством по отношению к ограничению метрики ρ на подмножество Y. Подмножество Y при этом называется подпространством пространства X.

Определение 3.2. *Шаровой окрестностью с центром $x \in X$ радиуса ε* называется множество $O_\varepsilon(x)$ всех точек $y \in X$, для которых $\rho(x, y) < \varepsilon$.

Расстоянием $\rho(Y_1, Y_2)$ между двумя множествами Y_1, $Y_2 \subset X$, называется нижняя грань множества чисел вида $\rho(x, y)$, когда x пробегает множество Y_1, а y пробегает множество Y_2. Если множества Y_1 и Y_2 имеют общую точку, то $\rho(Y_1, Y_2) = 0$.

Шаровой окрестностью подмножества $Y \subset X$ радиуса ε называется множество $O_\varepsilon(Y)$ всех точек $x \in X$, для которых $\rho(x, Y) < \varepsilon$.

Точкой прикосновения подмножества $Y \subset X$ называется всякая точка $x \in X$, для которой $\rho(x, Y) = 0$.

Замыканием множества Y называется множество всех его точек прикосновения. Замыкание множества Y обозначается через $\overline{Y}$.

Подмножество Y метрического пространства называется *замкнутым*, если его замыкание $\overline{Y}$ совпадает с Y: $Y = \overline{Y}$.

Точка x называется *внутренней точкой* множества Y, если ее некоторая шаровая окрестность содержится в Y: $O_\varepsilon(x) \subset Y$. Множество всех внутренних точек множества Y называется *внутренностью* множества Y и обозначается через $\operatorname{Int} Y$. Если множество Y совпадает со своей внутренностью $\operatorname{Int} Y$, то множество Y называется *открытым*.

Утверждение 3.1. *Пусть X — метрическое пространство. Множество $Y \subset X$ является замкнутым тогда и только тогда, когда дополнение $X \setminus Y$ открыто.*

Объединение любого семейства открытых множеств, а также пересечение конечного числа открытых множеств являются открытыми множествами. Пересечение любого семейства и объединение конечного числа замкнутых множеств являются замкнутыми множествами.

Во всех предыдущих утверждениях мы пока не пользовались свойством в) метрики. Сформулируем несколько утверждений о метрических пространствах, основанных на этом неравенстве.

а) шаровая окрестность $O_\varepsilon(x)$ является открытым множеством;

б) внутренность $\operatorname{Int} Y$ произвольного множества Y является открытым множеством;

в) замыкание $\overline{Y}$ произвольного множества Y является замкнутым множеством.

Во многих утверждениях о метрических пространствах используются только свойства, сформулированные выше. Поэтому можно расширить класс пространств, для которых имеет смысл говорить об открытых, замкнутых множествах, внутренних точках и точках прикосновения.

3.1.2 Топологические пространства

Скажем, что на множестве X задана *топология*, если задано некоторое семейство подмножеств множества X, называемых *открытыми* множествами, для которого выполнены следующие условия:

а) все множество X, а также пустое множество являются открытыми множествами;

б) объединение любого семейства и пересечение конечного числа открытых множеств являются открытыми множествами.

Множество X с заданной на нем топологией называют *топологическим пространством*, элементы множества X называются *точками* пространства X. Дополнения к открытым множествам называются *замкнутыми множествами*. В топологическом пространстве X выполнены двойственные условия для замкнутых множеств:

а$'$) все множество X, а также пустое множество являются замкнутыми множествами,

б$'$) пересечение любого семейства и объединение конечного числа замкнутых множеств являются замкнутыми множествами.

Семейство открытых множеств метрического пространства X задает на X некоторую топологию, т.е. превращает X в топологическое пространство.

В топологических пространствах воспроизводятся многие понятия метрических пространств. *Окрестностью точки* x топологического пространства X называется любое открытое множество, содержащее точку x. Открытое множество, содержащее подмножество Y, называется *окрестностью множества* Y. *Точкой прикосновения* множества $Y \subset X$ называется такая точка x, для которой каждая ее окрестность имеет непустое пересечение с множеством Y. Множество всех точек прикосновения множества Y называется *замыканием* множества Y и обозначается через $\overline{Y}$. *Внутренней точкой* множества Y называется такая точка $x \in Y$, которая лежит в Y вместе с некоторой своей окрестностью. Множество всех внутренних точек множества Y называется *внутренностью* и обозначается через $\operatorname{Int} Y$.

Множество $Y \subset X$ замкнуто (т.е. является дополнением к открытому множеству) тогда и только тогда, когда $Y = \overline{Y}$.

Замыкание $\overline{Y}$ произвольного множества Y топологического пространства X является замкнутым множеством, т.е. $\overline{\overline{Y}} = \overline{Y}$.

Пример 1. Рассмотрим множество X, состоящее из одного элемента x.

Тогда на X может быть введена единственная топология, открытыми множествами которой являются множество X и пустое множество.

Пример 2. Рассмотрим множество X, состоящее из двух элементов, x и y, $x \neq y$. На этом множестве можно ввести уже несколько различных топологий. Первая топология задается множеством всех подмножеств в качестве открытых множеств, т.е. $\{\emptyset, \{x\}, \{y\}, X\}$. Вторая топология задается следующим семейством открытых множеств: $\{\emptyset, X\}$. Можно задать и третью топологию, в качестве семейства открытых множеств которой следует взять $\{\emptyset, \{x\}, X\}$ или $\{\emptyset, \{y\}, X\}$. Мы перечислили все различные топологии на одном и том же множестве X, которые задают три различных топологических пространства.

Пример 3. Пусть X — произвольное множество. Введем на X топологию, объявив любое его подмножество открытым. Тогда любое одноточечное подмножество является открытым, а значит, и всякое подмножество как объединение своих точек тоже открыто. Такая топология называется *дискретной*.

Пусть X — топологическое пространство, $Y \subset X$ — его подмножество. Тогда на Y также можно задать топологию, объявив открытым всякое множество вида $Y \cap U$, где U — открытое в X множество. Тогда топологическое пространство Y называется *подпространством* топологического пространства X, а топология в Y — *индуцированной топологией*. Если X — метрическое пространство, а Y — его подпространство, то топология в Y задается независимо от порядка операций: ограничения метрики и перехода к топологии, или перехода к топологии и индуцирования топологии.

3.1.3 Непрерывные отображения

Понятие топологического пространства настолько удобно, что определение непрерывности отображения дословно переносится из математического анализа.

Определение 3.3 (по Коши). Отображение $f\colon X \to Y$ топологических пространств называется *непрерывным в точке* $x_0 \in X$, если для любой окрестности $V\big(f(x_0)\big)$ точки $f(x_0) \in Y$ существует такая окрестность $U(x_0)$ точки $x_0 \in X$, что $f\big(U(x_0)\big) \subset V\big(f(x_0)\big)$. Отображение f называется *непрерывным отображением*, если оно непрерывно в каждой точке пространства X.

Теорема 3.1. *Отображение $f\colon X \to Y$ непрерывно тогда и только тогда, когда выполнено одно из следующих эквивалентных условий*:

а) прообраз любого открытого множества является открытым множеством;

б) прообраз любого замкнутого множества является замкнутым множеством.

Доказательство. Поскольку для прообразов справедливо соотношение $f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)$, то условия а) и б) эквивалентны. Допустим, что f — непрерывное отображение, $V \subset Y$ — открытое множество. Покажем, что его прообраз $f^{-1}(V)$ открыт. Пусть $x \in f^{-1}(V)$; тогда $f(x) \in V$, т.е. V является окрестностью точки $f(x)$. Тогда согласно определению непрерывности отображения f найдется такая окрестность U точки x, что $f(U) \subset V$, т.е. $U \subset f^{-1}(V)$; это означает, что множество $f^{-1}(V)$ открыто.

Обратно, пусть выполнено условие а). Если $V \ni f(x_0)$ — окрестность точки $f(x_0)$, то $U = f^{-1}(V) \ni x_0$ — окрестность точки x_0, причем $f(U) = f(f^{-1}(V)) = V \subset V$. Следовательно, f — непрерывное отображение. ■

Условия а) и б) теоремы 3.1 удобны при проверке непрерывности отображений топологических пространств. В частности, если отображения $f\colon X \to Y$ и $g\colon Y \to Z$ непрерывные отображения топологических пространств. Тогда композиция отображений, $gf\colon X \to Z$, тоже непрерывна.

Пусть топологическое пространство X представлено в виде объединения двух своих замкнутых подмножеств: $X = F_1 \cup F_2$, и пусть $f\colon X \to Y$ — отображение пространства X в топологическое пространство Y. Отображение f непрерывно тогда и только тогда, когда непрерывны ограничения $f|_{F_1}$ и $f|_{F_2}$ отображения f на подмножества F_1 и F_2.

Рассмотрим непрерывное отображение $f\colon X \to Y$ топологического пространства X в топологическое пространство Y. Если отображение f взаимно однозначно, а обратное отображение f^{-1} непрерывно, то отображение f называется *гомеоморфизмом*. При этом топологические пространства X и Y называются *гомеоморфными топологическими пространствами*. При гомеоморфизме устанавливается не только взаимно однозначное соответствие между точками топологических пространств X и Y, но и взаимно однозначное соответствие между самими топологиями, т.е. между семействами открытых множеств (семействами замкнутых множеств).

Пример 4. Важным частным случаем непрерывных отображений являются непрерывные функции, т.е. непрерывные отображения топологического пространства X в пространство $\mathbb{R}^1$ вещественных чисел. Условие непрерывности функции f можно сформулировать следующим образом: *для любой точки $x_0 \in X$ и любого $\varepsilon > 0$ найдется такая окрестность U точки x_0, что при $x \in U$ выполнено неравенство $|f(x_0) - f(x)| < \varepsilon$.* Для функций на топологическом пространстве можно определить равномерный предел последовательности непрерывных функций. *Функция f называется равномерным пределом последовательности непрерывных функций f_n, если для любого $\varepsilon > 0$ найдется натуральное число N, такое что для всех $n > N$ и всех точек $x \in X$ выполнено неравенство $|f_n(x) - f(x)| < \varepsilon$.* Так же, как и в случае функций одной вещественной переменной, справедливо утверждение: *если $f = \lim\limits_{n\to\infty} f_n$ и последовательность $\{f_n\}$ непрерывных функций на топологическом пространстве X сходится к f равномерно, то функция f тоже непрерывна.*

Пример 5. Пусть $f\colon X \to Y$ — непрерывное отображение метрических пространств, ρ_1 и ρ_2 — метрики на пространствах X и Y соответственно. Тогда условие непрерывности отображения f можно сформулировать следующим образом: *для любого $x_0 \in X$ и $\varepsilon > 0$ существует такое $\delta > 0$, что из неравенства $\rho_1(x, x_0) < \delta$ следует неравенство $\rho_2\big(f(x), f(x_0)\big) < \varepsilon$.*

Скажем, что последовательность точек $\{x_n\}$ *сходится к точке* x_0, $\lim\limits_{n\to\infty} x_n = x_0$, если $\lim\limits_{n\to\infty} \rho(x_0, x_n) = 0$. В терминах сходящихся последовательностей метрического пространства формулируются многие свойства пространств и отображений. Например, множество $Y \subset X$ является замкнутым, если для любой сходящейся последовательности точек $\{x_n\} \subset Y$ предел $x_0 = \lim\limits_{n\to\infty} x_n$ также принадлежит множеству Y. Далее, условие непрерывности отображения $f\colon X \to Y$ метрических пространств можно сформулировать по Гейне.

Определение 3.4 (по Гейне). Отображение f *непрерывно* в точке x_0, если из равенства $x_0 = \lim\limits_{n\to\infty} x_n$ следует равенство $\lim\limits_{n\to\infty} f(x_n) = f(x_0)$.

Пример 6. Рассмотрим два топологических пространства X и Y. Образуем новое топологическое пространство $X \times Y$. Множество $X \times Y$ есть множество всех пар вида (x, y), $x \in X$, $y \in Y$, и называется *декартовым произведением множеств* X и Y. Определим теперь топологию в $X \times Y$. Множество $U \subset X \times Y$ назовем открытым, если U представимо в виде объединения $U = \bigcup_\alpha (V_\alpha \times W_\alpha)$, где $V_\alpha \subset X$, $W_\alpha \subset Y$ — открытые множества. Проверка свойств открытых множеств тривиальна. Множество $X \times Y$ с определенной выше топологией называется *декартовым произведением топологических пространств X и Y*. При этом топологические пространства X и Y называются *сомножителями* декартова произведения $X \times Y$.

Пусть Z, X_1 и X_2 — топологические пространства. Для любых двух отображений $f_1\colon Z \to X_1$ и $f_2\colon Z \to X_2$ отображение $F\colon Z \to X_1 \times X_2$, $F(z) = \big(f_1(z), f_2(z)\big)$, непрерывно в том и только том случае, если f_1 и f_2 непрерывны.

Пример 7. Пусть X и Y — метрические пространства; тогда декартово произведение $X \times Y$ допускает метрику, согласованную с топологией декартова произведения. Пусть ρ_1, ρ_2 — метрики пространств X и Y соответственно. Определим метрику ρ в декартовом произведении $X \times Y$ по формуле

$$\rho\big((x_1, y_1), (x_2, y_2)\big) = \max\{\rho_1(x_1, x_2), \rho_2(y_1, y_2)\}. \tag{3.1}$$

Другой способ задавать метрику в декартовом произведении использует аналогию между между сомножителями X и Y декартова произведения с осями координат на плоскости. Метрика задается по формуле

$$\rho'\big((x_1, y_1), (x_2, y_2)\big) = \sqrt{\rho_1(x_1, x_2)^2 + \rho_2(y_1, y_2)^2}. \tag{3.2}$$

Топология, задаваемая метрикой (3.2), совпадает с топологией, задаваемой метрикой (3.1), для чего достаточно найти две такие константы $C_1 > 0$ и $C_2 > 0$, что

$$C_1\rho\big((x_1, y_1), (x_2, y_2)\big) \leqslant \rho'\big((x_1, y_1), (x_2, y_2)\big) \leqslant C_2\rho\big((x_1, y_1), (x_2, y_2)\big).$$

3.1.4 Гомотопия и гомотопическая эквивалентность

Определение 3.5. Пусть X, Y — произвольные топологические пространства и $I = [0, 1]$ — единичный отрезок вещественных чисел. Непрерывные отображения $f\colon X \to Y$ и $g\colon X \to Y$ называются *гомотопными* ($f \sim g$), если существует семейство отображений $\varphi_t : X \to Y$, $t \in I$, такое, что

1) $\varphi_0 = f$, $\varphi_1 = g$,

2) отображение $\Phi\colon X \times I \to Y$, $\Phi(x, t) = \varphi_t(x)$, непрерывно.

Отображение Φ (или иногда семейство φ_t) называют *гомотопией, связывающий f с g.*

Очевидная проверка показывает, что отношение гомотопности рефлексивно, симметрично и транзитивно.

Пример 8. Все отображения произвольного пространства X в отрезок I гомотопны между собой. Действительно, отображение $x \mapsto (1 - t)f(x) + tg(x)$ составляют гомотопию, связывающую произвольные отображения $f, g\colon X \to I$.

Пусть заданы два топологических пространства X и Y и два непрерывных отображения

$$f\colon X \to Y,$$
$$g\colon Y \to X.$$

Суперпозиции $f \circ g : Y \to Y$, $g \circ f : X \to X$ отображают пространства Y и X в себя. Для каждого из них имеется тождественное отображение $\mathrm{Id}_X : X \to X$ и $\mathrm{Id}_Y : Y \to Y$.

Определение 3.6. Топологические пространства X и Y называются *гомотопически эквивалентными друг другу*, если найдутся такие непрерывные отображения $f\colon X \to Y$ и $g\colon Y \to X$, что суперпозиции $f \circ g$ и $g \circ f$ гомотопны тождественным отображениям Id_Y и Id_X соответственно. Если X и Y гомотопически эквивалентны, то пишут $X \sim Y$.

Определение 3.7. Пространство X называется *стягиваемым*, если тождественное отображение $\mathrm{Id}_X : X \to X$ гомотопно отображению $X \to X$, переводящему все X в точку.

Пример 9. Евклидово пространство $\mathbb{R}^n$ (как и любая стягиваемая область в нем) гомотопически эквивалентно одной точке $x_0 \sim \mathbb{R}^n$.

Доказательство. Вложение $f\colon x_0 \to \mathbb{R}^n$ и постоянное отображение $g\colon \mathbb{R}^n \to x_0$ таковы, что $g \circ f = \mathrm{Id}_{x_0}$, а отображение $f \circ g\colon \mathbb{R}^n \to \mathbb{R}^n$, переводящее все $\mathbb{R}^n$ в точку x_0, гомотопно тождественному отображению $\mathrm{Id}_{\mathbb{R}^n}$. Действительно, гомотопия $\Phi(x,t)$ задается формулой $\Phi(x,0) = x$, $\Phi(x,1) = x_0$, $\Phi(x,t) = (1-t)x + tx_0$, $0 \leqslant t \leqslant 1$. Для любого множества $A \subset \mathbb{R}^n$, которое по себе стягивается к точке x_0, доказательство аналогично (проведите его!). Примеры таких областей дает единичный шар и любая гомеоморфная ему область. ■

Пример 10. Область $\mathbb{R}^2 \setminus \{a\}$ гомотопически эквивалентна окружности S^1, а область $\mathbb{R}^3 \setminus \{a\}$ — сфере S^2 (задайте гомотопии формулами!).

Пример 11. Приведем пример гомотопически эквивалентных не гомеоморфных топологических пространств. В качестве X возьмем окружность S^1, а Y — кольцо. Здесь можно взять за $f\colon X \to Y$ естественное вложение и положить $g = f^{-1} \circ h\colon Y \to X$, где h — сжатие по радиусу.

3.1.5 Фактортопологии

Пусть X — топологическое пространство, Y — некоторое множество и $f\colon X \to Y$ — отображение пространства X на Y. *Фактортопология на множестве Y, задаваемая отображением f*, определяется следующим образом: множество $U \subset Y$ открыто тогда и только тогда, когда $f^{-1}(U)$ — открытое подмножество пространства X.

Пример 12. Пусть X — замкнутый интервал $[0, 2\pi]$, Y — единичная окружность $x^2 + y^2 = 1$. Определим f формулой $f(t) = (\cos t, \sin t)$. Тогда топология окружности, индуцированная из $\mathbb{R}^2$ совпадает с фактортопологией, задаваемой отображением f.

Пример 13. Пусть R — отношение эквивалентности на топологическом пространстве X, X/R — множество всех классов эквивалентности и $f\colon X \to X/R$ — естественное отображение, ставящее в соответствие каждой точке из X ее класс эквивалентности. Тогда на X/R вводится фактортопология, задаваемая этим отображением. Наделенное этой топологией пространство X/R называется *факторпространством пространства X по модулю R*. Например, пусть $X = \mathbb{R}^1$ — действительная прямая и R — отношение эквивалентности "$x \equiv y \pmod R$ тогда и только тогда, когда число $x - y$ целое". Тогда факторпространство X/R гомеоморфно окружности.

Пример 14. Пусть X — квадрат $ABCD$ на плоскости $\mathbb{R}^2$. Тогда X является топологическим пространством с топологией, индуцированной из $\mathbb{R}^2$. Введем следующие отношения эквивалентности R_i:

— (R_1) отождествим точки отрезков AB и DC, лежащие на одной и той же прямой, параллельной AD,

— (R_2) отождествим точки отрезков AB и CD, лежащие на одной и той же прямой, проходящей через центр квадрата,

— (R_3) отождествим точки отрезков AB и DC по (R_1) и точки отрезков BC и AD, лежащие на одной и той же прямой, параллельной AB,

— (R_4) отождествим точки отрезков AB и CD по (R_2) и точки отрезков BC и AD, лежащие на одной и той же прямой, параллельной AB,

— (R_5) отождествим точки отрезков AB и CD по (R_2) и точки отрезков BC и DA, лежащие на одной и той же прямой, проходящей через центр квадрата.

Тогда X/R_1 — кольцо, X/R_2 — лист Мебиуса, X/R_3 — двумерный тор, X/R_4 — бутылка Клейна, X/R_5 — проективная плоскость.

Задачи

1. Привести пример метрики на конечном множестве, которая не индуцируется никаким его вложением в евклидово пространство.

2. Показать, что конечное множество на прямой является замкнутым.

3. Показать, что $\rho(x, Y) = \rho(x, \overline{Y})$.

4. Доказать, что функция $f(x) = \rho(x, Y)$ непрерывна для любого подмножества Y.

5. Показать, что любая метрика на конечном множестве индуцирует на нем дискретную топологию.

6. Доказать, что открытый интервал, полуинтервал и замкнутый отрезок на вещественной прямой попарно не гомеоморфны.

7. Доказать, что метрика $\rho(x, y)$ является непрерывной функцией на декартовом квадрате $X \times X$ метрического пространства X.

8. Доказать, что подмножество A топологического пространство X представимо в виде разности двух замкнутых множеств тогда и только тогда, когда $(\overline{A} \setminus A)$ является замкнутым множеством в X.

3.2 Связность. Аксиомы отделимости

Рассмотрим две точки $x, y \in X$ в топологическом пространстве X и такое непрерывное отображение единичного отрезка $\gamma\colon [0,1] \to X$, что $\gamma(0) = x$, $\gamma(1) = y$. Скажем, что γ есть кривая, соединяющая точки x и y. Топологическое пространство X называется *линейно связным*, если любые его точки можно соединить кривой. В этом случае пространство нельзя разбить на два непустых непересекающихся открыто замкнутых подмножества. Этот факт вытекает из того, что единичный отрезок уже обладает этим свойством. В самом деле, допустим, что отрезок $[a, b]$ представлен в виде объединения двух непустых открыто-замкнутых непересекающихся множеств, $[a, b] = A \cup B$. Без ограничения общности будем считать, что $a \in A$. Тогда в силу открытости A найдется такое $\varepsilon > 0$, что полуинтервал $[a, a+\varepsilon) \subset A$. Пусть $\varepsilon_0 = \sup\{\varepsilon\}$, когда ε пробегает такие числа, что $[a, a+\varepsilon) \subset A$. Тогда для любого $\varepsilon < \varepsilon_0$ имеем включение $[a, a+\varepsilon+(\varepsilon_0-\varepsilon)/2) \subset A$, т.е. $a+\varepsilon \in A$

для любого $\varepsilon < \varepsilon_0$. Следовательно, $a + \varepsilon_0 \in A$ в силу замкнутости A. Так как A одновременно открыто, то ε_0 входит в A вместе с некоторой окрестностью. Единственной возможностью для ε_0 может быть только равенство $a + \varepsilon_0 = b$, поскольку в противном случае $\varepsilon_0 < \sup\{\varepsilon\}$. Таким образом, $A = [a, b]$, т.е. $B = \emptyset$, что противоречит предположению.

Вообще, пространство X называется *связным*, если его нельзя разбить на два непустых открыто-замкнутых подмножества.

Теорема 3.2. *Пусть $X = \bigcup_\alpha X_\alpha$, каждое X_α связно и пересечение $\bigcap_\alpha X_\alpha$ непусто. Тогда пространство X связно.*

Доказательство. Допустим противное, т.е. $X = A \cup B$, $A \cap B = \emptyset$, множества A и B открыты и непусты. Тогда $X_\alpha = (X_\alpha \cap A) \cup (X_\alpha \cap B)$, где $X_\alpha \cap A$ и $X_\alpha \cap B$ открыты в X_α. Но поскольку множество X_α связно, то либо $X_\alpha \cap A = \emptyset$, либо $(X_\alpha \cap B) = \emptyset$, т.е. каждое множество X_α целиком лежит либо в A, либо в B. Далее, так как множества A и B непусты, то найдутся точки $a \in A$, $b \in B$. Пусть $a \in X_{\alpha_0}$, тогда $X_{\alpha_0} \subset A$. Пусть $b \in X_{\alpha_1}$, тогда $X_{\alpha_1} \subset B$. Значит, $X_{\alpha_0} \cap X_{\alpha_1} = \emptyset$, что противоречит условию теоремы. ■

Теорема 3.3. *Образ при непрерывном отображении связного пространства связен.*

Доказательство. Пусть $f\colon X \to Y = f(X)$ — непрерывное отображение. Поэтому если Y — несвязное пространство, то $Y = A \cup B$, $A \cap B = \emptyset$, множества A, B — открыты и непусты. Следовательно, $X = f^{-1}(Y) = f^{-1}(A) \cup f^{-1}(B)$, множества $f^{-1}(A), f^{-1}(B)$ открыты, непусты и не пересекаются. ■

Пример 15. Из теоремы 3.3 следует, что всякая непрерывная вещественная функция $y = f(x)$, определенная на отрезке $[a, b]$ вещественной прямой, принимает промежуточные значения. Если бы функция f не принимала какое-либо промежуточное значение y_0, то образ $f([a, b])$ распадался бы в объединение двух открытых непустых множеств: тех значений, которые меньше y_0, и тех значений, которые больше y_0, что противоречит теореме 3.3.

Существуют примеры связных, но не линейно связных топологических пространств.

Пример 16. Пусть X — замыкание графика функции $y = f(x) = \sin(1/x)$ как множества в двумерном евклидовом пространстве $\mathbb{R}^2$ (рис. 3.1). Метрика в $\mathbb{R}^2$ берется классическая, по длине отрезков, соединяющих пары точек в $\mathbb{R}^2$. Тогда множество X состоит из объединения графика функции $y = \sin(1/x)$ и вертикального отрезка $\Gamma_3 = \{(x, y) \mid x = 0, \quad -1 \leqslant y \leqslant +1\}$. График функции f распадается на два подмножества, каждое из которых гомеоморфно интервалу:

$$\Gamma_1 = \{(x, y) \mid 0 < x < \infty, \quad y = f(x)\},$$

Рис. 3.1:

$$\Gamma_2 = \{(x, y) \mid -\infty < x < 0, \quad y = f(x)\}.$$

Поэтому, если $X = A \cup B$, $A \cap B = \emptyset$, A, B — открытые непустые множества, то каждое из подмножеств Γ_1, Γ_2, Γ_3 целиком лежит либо в A, либо B. Пусть $\Gamma_3 \subset B$. Нетрудно проверить, что любая окрестность Γ_3 пересекается как с Γ_1, так и с Γ_2, т.е. $\Gamma_1 \subset B$, $\Gamma_2 \subset B$. Значит, $A = \emptyset$, что противоречит предложению. Таким образом, X — связное пространство.

Покажем теперь, что X не является линейно связным. Рассмотрим две точки $P = (-1/\pi, 0)$ и $Q = (1/\pi, 0)$ в X. Предположим, что существует непрерывное отображение $f\colon [0,1] \to X$, $f(0) = P$, $f(1) = Q$. Отображение f задается двумя непрерывными числовыми функциями: $f(t) = \big(x(t), y(t)\big)$, $y(t) = \sin\big(1/x(t)\big)$ при $x(t) \neq 0$. Поскольку $x(0) = -1/\pi$, то точная нижняя грань t_0 тех значений t, для которых $x(t) = 0$, строго больше нуля, $t_0 > 0$.

Итак, на полуинтервале $[0, t_0)$ выполнены условия: $x(t) < 0$, $y(t) = \sin\big(1/x(t)\big)$.нуля. Так как функция $x(t)$ непрерывна, то $x(t_0) = 0$ (иначе получим противоречие с тем, что t_0 является точной нижней гранью). Но функция $y(t)$ также непрерывна в точке t_0, поэтому для любой последовательности $t_k < t_0$ такой, что $t_k \to t_0$, имеем $y(t_k) \to y(t_0)$. Но для каждого $t_k < t_0$ функция $x(t)$, ограниченная на $[t_k, t_0]$, в силу своей непрерывности принимает все значения между $x(t_k) < 0$ и $x(t_0) = 0$, поэтому можно выбрать последовательности $t'_k < t_0$ и $t''_k < t_0$, стремящиеся к t_0 и такие, что $\sin \frac{1}{t'_k} = 1$, а $\sin \frac{1}{t''_k} = -1$, т.е. $y(t'_k) \to 1$, а $y(t''_k) \to -1$, противоречие с тем, что обе этих последовательности должны сходиться к $y(t_0)$.

Определение 3.8. Топологическое пространство X называется *хаусдорфовым пространством*, если для любых точек x, $y \in X$, $x \neq y$, найдутся такие окрестности $U(x)$ и $U(y)$, что $U(x) \cap U(y) = \emptyset$.

В хаусдорфовом пространстве каждая точка $x \in X$ является замкнутым множеством. В самом деле, если $y \neq x$, то найдется окрестность $U(y)$ точки y, не содержащая x. Тогда множество $X \setminus \{x\} = \bigcup_{y \neq x} U(y)$ является открытым, а его дополнение, состоящее из одной точки x, замкнуто.

Дискретное топологическое пространство X является хаусдорфовым пространством. В самом деле, каждая точка $x \in X$ является открытым

множеством, значит, если $x \neq y$, то окрестности $\{x\}$ и $\{y\}$ не пересекаются.

Если X и Y — хаусдорфовы пространства, то их декартово произведение, несвязная и связная суммы являются хаусдорфовыми топологическими пространствами.

Всякое метрическое пространство хаусдорфово. Более того, если два замкнутых подмножества F_1 и F_2 метрического пространства не пересекаются, то существуют непересекающиеся их окрестности $U_1 \supset F_1$ и $U_2 \supset F_2$.

Действительно, пусть X — метрическое пространство с метрикой ρ, F_1 и F_2 — два непересекающихся замкнутых множества. Пусть $x \in F_1$, $\varepsilon(x) = (1/3)\rho(x, F_2)$. Положим $U_1 = \bigcup_{x \in F_1} O_{\varepsilon(x)}(x)$. Аналогично определим открытое множество $U_2 = \bigcup_{y \in F_2} O_{\varepsilon'(y)}(y)$, где $\varepsilon'(y) = (1/3)\rho(y, F_1)$. Мы получили окрестности множеств F_1 и F_2. Покажем, что множества U_1 и U_2 не пересекаются. Допустим противное, т.е. что существует точка $z \in U_1 \cap U_2$. Тогда для некоторых точек $x \in F_1$ и $y \in F_2$, имеем $z \in O_{\varepsilon(x)}(x)$, $z \in O_{\varepsilon'(y)}(y)$, т.е. $\rho(x, z) < (1/3)\rho(x, F_2)$, $\rho(y, z) < (1/3)\rho(y, F_1)$. В частности,

$$\rho(x, z) < (1/3)\rho(x, y), \qquad \rho(y, z) < (1/3)\rho(y, x).$$

Складывая эти два неравенства, получаем

$$\rho(x, z) + \rho(y, z) < (2/3)\rho(x, y),$$

что противоречит неравенству треугольника.

Система открытых множеств $\{U_\alpha\}$ топологического пространства X называется *открытым покрытием*, если $X = \bigcup_\alpha U_\alpha$. Покрытия — удобное понятие при изучении топологических пространств. Например, если в каждом множестве U_α задана непрерывная функция f_α, а на каждом пересечении $U_\alpha \cap U_\beta$ функции f_α и f_β совпадают, то существует одна непрерывная функция f на пространстве X, совпадающая с f_α в каждом открытом множестве U_α.

Пусть задано два открытых покрытия $\{U_\alpha\}$ и $\{V_\beta\}$ топологического пространства X. Скажем, что покрытие $\{V_\beta\}$ *измельчает покрытие* $\{U_\alpha\}$ или *является более мелким, чем покрытие* $\{U_\alpha\}$, если каждое множество V_β лежит в некотором множестве U_α, $\alpha = \alpha(\beta)$.

Теорема 3.4. *Пусть X — метрическое пространство, $\{U_\alpha\}_{\alpha=1}^N$ — конечное открытое покрытие. Тогда существует более мелкое покрытие $\{V_\alpha\}_{\alpha=1}^N$, причем $\overline{V}_\alpha \subset U_\alpha$.*

Доказательство. Заметим, что в теореме 3.4 утверждается существование не просто более мелкого покрытия, а такого, у которого элементы нумерованы тем же индексом α и включение множеств происходит с одинаковым значением индекса α. Для доказательства теоремы рассмотрим замкнутые непересекающиеся множества $X \setminus \bigcup_{\alpha=2}^N U_\alpha$ и $X \setminus U_1$. Тогда существуют окрестности, разделяющие эти два множества. В качестве окрестности V_1 нужно взять ту из них, что содержит $X \setminus \bigcup_{\alpha=2}^N U_\alpha$. Тогда

$X \setminus \bigcup_{\alpha=2}^{N} U_\alpha \subset V_1 \subset \overline{V}_1 \subset U_1$ и система множеств $\{V_1, U_2, \ldots, U_N\}$ покрывает пространство X. Поэтому мы можем найти множество $V_2 \subset \overline{V}_2 \subset U_2$, причем система $\{V_1, V_2, U_3, \ldots, U_N\}$ тоже покрывает пространство X. Заменяя последовательно множество U_k на $V_k \subset \overline{V}_k \subset U_k$, мы придем к искомому покрытию $\{V_1, \ldots, V_N\}$ за N шагов. ■

Задачи

1. Доказать, что если из $\mathbb{R}^n$, $n \geqslant 2$, выбросить конечное (или счетное) число точек, то оставшееся пространство будет связным.

2. Доказать, что если из $\mathbb{R}^n$ выбросить конечное число подпространств размерности меньше, чем $n-1$, то оставшееся пространство будет связно.

3. Подсчитать, на какое наибольшее число компонент связности разобьется $\mathbb{R}^2$ конечным числом прямых. Каково наименьшее число компонент связности?

4. Пусть $f\colon X \to X$ — непрерывное отображение хаусдорфова пространства. Доказать, что множество неподвижных точек (т.е. точек x, для которых $f(x) = x$) замкнуто.

6. Доказать, что пространство X хаусдорфово тогда и только тогда, когда диагональ $\Delta = \{(x, y) \mid x = y\} \subset X \times X$ замкнута в $X \times X$.

7. Доказать, что отображение $f\colon X \to Y$ в хаусдорфово пространство Y непрерывно тогда и только тогда, когда график $\Gamma_f = \{(x, f(x)) \mid x \in X\} \subset X \times Y$ замкнут в $X \times Y$.

3.3 Компактные топологические пространства

Свойство компактности, которое мы рассмотрим в настоящем параграфе, является одним из важнейших свойств топологических пространств. В частности, это свойство играет фундаментальную роль при исследовании вещественных чисел и непрерывных функций.

Определение 3.9. Хаусдорфово топологическое пространство X называется *компактным*, если из его всякого открытого покрытия $\{U_\alpha\}$ можно выделить конечную часть $\{U_{\alpha_k}\}_{k=1}^{N}$, покрывающую X.

Можно требовать всего лишь, чтобы существовало конечное более мелкое покрытие.

Пример 17. Докажем, что любой конечный отрезок $[a, b]$ вещественной оси $\mathbb{R}$ является компактным пространством. В самом деле, если $\{U_\alpha\}$ — открытое покрытие отрезка $[a, b]$, то без ограничения общности можно считать, что каждый элемент покрытия U_α является интервалом (c_α, d_α) (за исключением двух полуинтервалов, $[a, a')$ и $(b', b]$). Рассмотрим множество P всех таких чисел $x \in [a, b]$, что отрезок $[a, x]$ покрывается конечным числом элементов U_α. Тогда, если $x \in P, y < x$, то y тоже принадлежит P. Поскольку

точка a принадлежит элементу покрытия $[a, a')$, $a' > a$, то множество P состоит более чем из одной точки a.

Пусть x_0 — точная верхняя грань множества P, $x_0 > a$. Если $x_0 < b$, то точка x_0 лежит в некотором интервале $U_{\alpha_0} = (c_{\alpha_0}, d_{\alpha_0})$, т.е. $c_{\alpha_0} < x_0 < d_{\alpha_0}$. Пусть y, z — такие числа, что $c_{\alpha_0} < y < x_0 < z < d_{\alpha_0}$. Тогда $y \in P$ и, значит, конечное число множеств U_β покрывает отрезок $[a, y]$, поэтому отрезок $[a, z]$ тоже покрывается конечным числом множеств U_β. Следовательно, x_0 не является верхней гранью множества P. Таким образом, $x_0 = b$. Тогда $b' < x_0 = b$. Если $b' < y < x_0$, то $y \in P$, и отрезок $[a, y]$ опять-таки покрывается конечным числом множеств U_β. Вместе с ним покрывается конечным числом множеств U_β и весь отрезок $[a, b]$.

Теорема 3.5. *Пусть $F \subset X$ — компактное подпространство в хаусдорфовом топологическом пространстве X. Тогда F — замкнутое множество в пространстве X.*

Доказательство. Пусть x — произвольная точка пространства X, не принадлежащая F. В силу хаусдорфовости пространства X для точек $y \in F$ найдутся непересекающиеся окрестности $U_y \ni y$ и $V_y \ni x$. Тогда семейство $\{U_y\}$ покрывает множество F, а в силу его компактности найдется конечное семейство $\{U_{y_k}\}_{k=1}^N$, покрывающее F. Тогда объединение $\bigcup_{k=1}^N U_{y_k}$ не пересекается с пересечением $\bigcap_{k=1}^N V_{y_k}$ и содержит множество F. Следовательно, точка x имеет окрестность, не пересекающуюся с множеством F, что и означает, что множество F замкнуто. Теорема доказана. ■

Теорема 3.6. *Пусть $f\colon X \to Y$ — непрерывное отображение компактного пространства X в пространство Y. Тогда образ $f(X)$ является компактным пространством.*

Доказательство. Пусть $\{U_\alpha\}$ — открытое покрытие множества $f(X)$. Тогда семейство $\{f^{-1}(U_\alpha)\}$ является открытым покрытием компактного пространства X. Следовательно, некоторое конечное семейство $\{f^{-1}(U_{\alpha_k})\}_{k=1}^N$ покрывает пространство X. Тогда семейство $\{U_{\alpha_k}\}_{k=1}^N$ покрывает образ $f(X)$. ■

Из теорем 3.5 и 3.6 вытекает следующее общее свойство непрерывных функций на компактных пространствах, известное из курса математического анализа.

Теорема 3.7. *Пусть $f\colon X \to \mathbb{R}^1$ — непрерывная функция на компактном пространстве X. Тогда функция f ограничена и принимает максимальное и минимальное значения.*

Доказательство. По теореме 3.6 образ $f(X)$ является компактным подпространством в $\mathbb{R}^1$, а по теореме 3.5 — замкнутым множеством. Если бы образ $f(X)$ не был ограниченным, то система интервалов $U_n = (-n, n)$ покрывала

бы $f(X)$ и из нее нельзя было бы выделить конечного подпокрытия. Положим $A = \sup\limits_{x \in X}\{f(x)\}$, $B = \inf\limits_{x \in X}\{f(x)\}$. Тогда A и B — точки прикосновения множества $f(X)$. Так как множество $f(X)$ замкнуто, то $A, B \in f(X)$. ■

Для метрических пространств свойство компактности можно сформулировать в привычных для математического анализа терминах.

Теорема 3.8. *Пусть X — метрическое пространство. Пространство X является компактным тогда и только тогда, когда выполнено одно из следующих эквивалентных свойств*:

1) *всякая последовательность $\{x_n\}$ имеет сходящуюся подпоследовательность*;

2) *всякая последовательность вложенных непустых замкнутых подмножеств $\{F_n\}$, $F_n \supset F_{n+1}$, имеет непустое пересечение.*

В качестве следствия из теоремы 3.8 приведем следующее фундаментальное свойство вещественных чисел: *последовательность вложенных отрезков вещественной прямой имеет общую точку.*

Если хаусдорфово пространство X является объединением конечного числа своих компактных подмножеств, то X является компактным пространством.

Теорема 3.9. *Декартово произведение $X \times Y$ метрических компактных пространств X и Y является компактным пространством.*

Замечание. В самом деле, согласно теореме 3.8 рассмотрим произвольную последовательность $\{z_n\}$, $z_n = (x_n, y_n)$. Поскольку X — компактное пространство, то существует сходящаяся подпоследовательность $\{x_{n_k}\}$. Тогда в силу компактности пространства Y из последовательности $\{y_{n_k}\}$ выделяется сходящаяся подпоследовательность $\{y_{n_{k_s}}\}$. Таким образом, подпоследовательность $\{z_{n_{k_s}}\}$ также сходится.

Задачи

1. Доказать, что объединение конечного числа компактных пространств компактно.

2. Показать, что если X некомпактное метрическое пространство, то существует непрерывная и не ограниченная на этом пространстве функция.

3. Доказать, что если в компактном метрическом пространстве два замкнутых множества A и B не пересекаются, то $\rho(A, B) > 0$.

3.4 Функциональная отделимость. Разбиение единицы

В этом параграфе мы приведем теоремы, позволяющие при довольно слабых предположениях проводить анализ непрерывных функций так же, как

и в случае функций одной вещественной переменной. Как уже отмечалось ранее, непрерывные функции на топологических пространствах во многих отношениях ведут себя так же, как и функции одной вещественной переменной. Сумма $f+g$, произведение $f\,g$ двух непрерывных функций, а также отношение f/g при $g \neq 0$ являются непрерывными функциями. В классе непрерывных функций на топологическом пространстве X можно переходить к пределу. Последовательность функций f_n *равномерно сходится* к функции f, если для любого $\varepsilon > 0$ существует такой номер N, что при $n > N$ выполнено неравенство $|f(x) - f_n(x)| < \varepsilon$ для любой точки $x \in X$.

Теорема 3.10. *Равномерный предел последовательности непрерывных функций на топологическом пространстве X является непрерывной функцией.*

Доказательство. Пусть $f(x) = \lim\limits_{n\to\infty} f_n(x)$. Покажем, что функция f непрерывна. Фиксируем число $\varepsilon > 0$ и точку $x_0 \in X$. Тогда найдется такой номер n, что для любой точки $x \in X$ выполнено неравенство $|f(x) - f_n(x)| < \varepsilon/3$. Поскольку функция f_n непрерывна, то найдется такая окрестность $O(x_0)$, что при $x \in O(x_0)$ выполнено неравенство $|f_n(x) - f_n(x_0)| < \varepsilon/3$. Тогда при $x \in O(x_0)$ имеем

$$|f(x) - f(x_0)| < |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + \\ + |f_n(x_0) - f(x_0)| < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon. \qquad \blacksquare$$

3.4.1 Функциональная отделимость

Определение 3.10. Хаусдорфово топологическое пространство X называется *нормальным*, если в нем всякие два непересекающихся замкнутых множеств имеют непересекающиеся окрестности.

Если F_1, F_2 — два непересекающихся замкнутых множества в топологическом пространстве X, то для построения непересекающихся окрестностей множеств F_1 и F_2 достаточно построить такую непрерывную функцию f на пространстве X, для которой $f(x) \geqslant a$ при $x \in F_1$, $f(x) \leqslant b$ при $x \in F_2$ и $a > b$. Тогда в качестве окрестностей множеств F_1 и F_2 можно взять прообразы интервалов $f^{-1}(c, \infty)$ и $f^{-1}(-\infty, c)$, где $b < c < a$. В некотором смысле справедливо и обратное утверждение.

Теорема 3.11 (лемма Урысона). *Пусть X — нормальное топологическое пространство, F_0, F_1 — два замкнутых непересекающихся множества. Тогда существует такая непрерывная функция $f\colon X \to [0,1]$, что $f|_{F_0} \equiv 0$, $f|_{F_1} \equiv 1$.*

Доказательство. Построим такую систему открытых множеств Γ_r, занумерованных всеми двоично-рациональными числами $0 \leqslant r \leqslant 1$, что выполнены следующие условия: 1) $F_0 \subset \Gamma_0$; 2) $\overline{\Gamma}_1 \subset X \setminus F_1$, 3) $\overline{\Gamma}_r \subset \Gamma_{r'}$ при $r < r'$.

Тогда систему открытых множеств $\{\Gamma_r\}$ можно расширить, добавив к ним множества $\Gamma_t = \bigcup_{r<t} \Gamma_r$, где t — уже произвольное вещественное число, $0 \leqslant t \leqslant 1$. Система открытых множеств $\{\Gamma_t\}$ удовлетворяет тем же условиям: $\overline{\Gamma}_t \subset \Gamma_{t'}$ при $t < t'$. В самом деле, взяв такие рациональные числа r, r', что $t < r < r' < t'$, получаем $\Gamma_t \subset \Gamma_r$, $\Gamma_{r'} \subset \Gamma_{t'}$, следовательно, $\overline{\Gamma}_t \subset \overline{\Gamma}_r \subset \Gamma_{r'} \subset \Gamma_t$.

Построим теперь непрерывную функцию $f\colon X \to [0,1]$, полагая $f(x) = 0$ при $x \in \Gamma_0$ и $f(x) = \sup\{t \mid x \notin \Gamma_t\}$ при $x \notin \Gamma_0$. Покажем, что f — непрерывная функция. Фиксируем точку x_0 и $\varepsilon > 0$. Положим $t_0 = f(x_0)$. В силу определения функции f имеем $x_0 \notin \Gamma_{(t_0-\varepsilon/2)}$, $x_0 \in \Gamma_{(t_0+\varepsilon/2)}$. Рассмотрим окрестность U точки x_0: $U = \Gamma_{(t_0+\varepsilon/2)} \setminus \overline{\Gamma}_{(t_0-\varepsilon/2)}$. Тогда если $y \in U$, то $y \in \Gamma_{(t_0+\varepsilon/2)}$, $y \notin \overline{\Gamma}_{(t_0-\varepsilon/2)}$. По определению функции f выполнены неравенства $t_0 - \varepsilon/2 \leqslant f(y) \leqslant t_0 + \varepsilon/2$, т.е. $|f(x_0) - f(y)| \leqslant \varepsilon$. Таким образом, функция f непрерывна. Кроме того, если $x \in F_0$, то $f(x) = 0$, а если $x \in F_1$, то $f(x) = 1$.

Чтобы завершить доказательство теоремы 3.11, нам осталось построить систему открытых множеств Γ_r, удовлетворяющих условиям 1), 2) и 3). В силу нормальности пространства X для любого замкнутого множества F и его окрестности U, $F \subset U$, найдется другая окрестность V такая, что $F \subset V \subset \overline{V} \subset U$. Для краткости будем писать $V \ll U$, если $V \subset \overline{V} \subset U$. Итак, $F_0 = \overline{F}_0 \subset X \setminus F_1$. Поэтому $F_0 \ll X \setminus F_1$. Следовательно, найдется открытое множество Γ_0 такое, что $F_0 \ll \Gamma_0 \ll X \setminus F_1$. Аналогично, найдется такое открытое множество Γ_1, что $F_0 \ll \Gamma_0 \ll \Gamma_1 \ll X \setminus F_1$. Предположим, что для всех двоично-рациональных чисел вида $r = p/2^n$, $0 \leqslant p \leqslant 2^n$, уже построены открытые множества Γ_r, причем $\Gamma_{(p/2^n)} \ll \Gamma_{((p+1)/2^n)}$. Определим тогда множество $\Gamma_{(2p+1)/2^{n+1}}$ так, что $\Gamma_{p/2^n} \ll \Gamma_{(2p+1)/2^{n+1}} \ll \Gamma_{(p+1)/2^n}$. По индукции все семейство множеств $\{\Gamma_r\}$ построено. ■

Из теоремы 3.11 следует теорема о продолжении непрерывных функций.

Теорема 3.12. *Пусть X — нормальное топологическое пространство, $F \subset X$ — замкнутое множество, $f\colon F \to \mathbb{R}^1$ — непрерывная функция на множестве F. Тогда функция f продолжается до непрерывной функции $g\colon X \to \mathbb{R}^1$ на всем пространстве X. Если функция f ограничена, $|f(x)| \leqslant A$, то и функцию g можно выбрать ограниченной той же константой, $|g(x)| \leqslant A$.*

Доказательство. Предположим сначала, что функция f ограничена, $|f(x)| \leqslant A$. Положим $\varphi_0(x) = f(x)$ и рассмотрим два замкнутых подмножества: $A_0 = \{x \mid \varphi_0(x) \leqslant -A/3\}$, $B_0 = \{x \mid \varphi_0(x) \geqslant A/3\}$. Поскольку множества A_0 и B_0 не пересекаются, то согласно теореме 3.11 существует непрерывная функция $f_0\colon X \to [-A/3, A/3]$, равная $-A/3$ на множестве A_0 и $A/3$ на множестве B_0. Другими словами, $|f_0(x)| \leqslant A/3$, $x \in X$, $|\varphi_0(x) - f_0(x)| \leqslant 2A/3$.

Положим $\varphi_1(x) = \varphi_0(x) - f_0(x)$. Тогда функция φ_1 ограничена на множестве F константой $2A/3$. Поэтому, повторяя весь процесс, можно построить два непересекающихся замкнутых множества: $A_1 = \{x \mid \varphi_1(x) \leqslant -2A/9\}$,

$B_1 = \{x \mid \varphi_1(x) \geqslant 2A/9\}$, и непрерывную функцию $f_1 \colon X \to [-2A/9, 2A/9]$, равную $-2A/9$ на множестве A_1 и $2A/9$ на множестве B_1. Другими словами, $|f_1(x)| \leqslant 2/3 \cdot A/3$, $|\varphi_1(x) - f_1(x)| \leqslant 4A/9$. Повторяя указанный процесс бесконечное число раз, мы построим две последовательности функций $f_n \colon X \to \mathbb{R}^1$, $\varphi_n \colon F \to \mathbb{R}^1$, удовлетворяющих следующим условиям:

$$\varphi_{n+1}(x) = \varphi_n(x) - f_n(x),$$

$$|f_n(x)| \leqslant \left(\frac{2}{3}\right)^n \frac{A}{3}, \qquad |\varphi_n(x)| \leqslant \left(\frac{2}{3}\right)^n A. \tag{3.3}$$

Далее,

$$f(x) = \varphi_0(x) = f_0(x) + \varphi_1(x) = \ldots$$
$$\cdots = f_0(x) + f_1(x) + \cdots + f_n(x) + \varphi_{n+1}(x) \text{ при } x \in F.$$

В силу неравенства (3.3) ряд $\sum\limits_{k=0}^{\infty} f_k(x)$ сходится равномерно на всем пространстве X. По теореме 3.10 функция $g(x) = \sum\limits_{k=0}^{\infty} f_k(x)$ является непрерывной и по построению $g(x) = f(x)$ при $x \in F$. В силу неравенства (3.3) имеем

$$|g(x)| \leqslant \sum_{k=0}^{\infty} |f_k(x)| \leqslant \sum_{k=0}^{\infty} \left(\frac{2}{3}\right)^n \frac{A}{3} = A.$$

Тем самым теорема доказана для случая ограниченной функции f.

В общем случае рассмотрим гомеоморфизм $h \colon \mathbb{R}^1 \to (-1, 1)$. Тогда композиция $hf \colon F \to (-1, 1)$ является непрерывной ограниченной функцией. Применяя теорему 3.12 для ограниченной функции hf, построим непрерывную функцию $g \colon X \to [-1, 1]$, продолжающую функцию hf. Значения ± 1 функция g принимает на некотором замкнутом множестве F_1, не пересекающемся со множеством F. По теореме 3.11 существует непрерывная функция $\psi \colon X \to [0, 1]$, равная 1 на множестве F и 0 на множестве F_1. Тогда функция $g_1(x) = \psi(x)g(x)$, $x \in X$, совпадает с функцией hf на множестве F и не принимает значений ± 1. Таким образом, функция g_1 отображает пространство X в интервал $(-1, 1)$. Наконец, положим $g(x) = h^{-1}g_1(x)$. Тогда $g \colon X \to \mathbb{R}^1$ непрерывна и при $x \in F$ выполнены равенства $g(x) = h^{-1}g_1(x) = h^{-1}hf(x) = f(x)$. Теорема полностью доказана. ■

3.4.2 Разбиение единицы

Определение 3.11. *Носителем* непрерывной функции f на топологическом пространстве X называется замыкание множества тех точек $x \in X$, где $f(x) \neq 0$. Носитель функции f обозначается через $\operatorname{supp} f$. Таким образом, вне носителя функция f тождественно равна нулю.

Одним из полезных приемов в топологии является разложение функций в сумму так, чтобы каждое из слагаемых имело достаточно малый носитель.

Теорема 3.13. *Пусть X — метрическое пространство, $\{U_\alpha\}$ — конечное открытое покрытие. Тогда существуют такие функции $\varphi_\alpha\colon X \to \mathbb{R}^1$, $0 \leqslant \varphi_\alpha(x) \leqslant 1$, что*: а) $\operatorname{supp}\varphi_\alpha \subset U_\alpha$, б) $\sum\limits_\alpha \varphi_\alpha(x) \equiv 1$.

Система функций $\{\varphi_\alpha\}$, указанных в теореме 3.13, называется *разбиением единицы, подчиненным покрытию* $\{U_\alpha\}$.

Доказательство. Рассмотрим конечное покрытие $\{U_\alpha\}$. Согласно теореме 3.4 найдется более мелкое покрытие $\{V_\alpha\}$, причем $\overline{V}_\alpha \subset U_\alpha$. По теореме 3.11 существует непрерывная функция ψ_α на пространстве X, удовлетворяющая условиям

$$\psi_\alpha|_{\overline{V}_\alpha} \equiv 1, \qquad \psi_\alpha|_{X\setminus U_\alpha} \equiv 0, \qquad 0 \leqslant \psi_\alpha \leqslant 1.$$

Это значит, что $\operatorname{supp}\psi_\alpha \subset U_\alpha$ и $\psi_\alpha(x) = 1$ при $x \in V_\alpha$.

Положим $\psi(x) = \sum\limits_\alpha \psi_\alpha(x)$. Функция ψ непрерывна. Покажем, что $\psi(x) \geqslant 0$ в каждой точке $x \in X$. В самом деле, поскольку система $\{V_\alpha\}$ покрывает пространство X, то найдется такой номер α_0, что $x \in V_{\alpha_0}$, т.е. $\psi_{\alpha_0}(x) > 0$. Тогда $\psi(x) = \sum\limits_\alpha \psi_\alpha(x) \geqslant \psi_{\alpha_0}(x) > 0$.

Наконец, положим $\varphi_\alpha(x) = \psi_\alpha(x)/\psi(x)$. Тогда

$$\operatorname{supp}\varphi_\alpha = \operatorname{supp}\psi_\alpha \subset U_\alpha, \quad 0 \leqslant \varphi_\alpha(x) \leqslant 1,$$

$$\sum_\alpha \varphi_\alpha(x) = \sum \frac{\psi_\alpha(x)}{\psi(x)} = \frac{\sum \psi_\alpha(x)}{\psi(x)} = \frac{\psi(x)}{\psi(x)} \equiv 1. \quad \blacksquare$$

Задачи

1. Доказать, что в теореме 3.11 можно требовать, чтобы функция f была гладкой, если X — евклидово пространство $\mathbb{R}^n$.

2. Доказать, что в теореме 3.13 при $X = \mathbb{R}^n$ можно требовать, чтобы функции φ_α были гладкими.

3. Для окружности построить разбиение единицы системой а) непрерывных, б) гладких функций.

4. Доказать, что для окружности не существует разбиения единицы системой аналитических функций.

5. Обобщить задачи 3, 4 на случай сферы $S^n = \{x_1^2 + \ldots x_{n+1}^2 = 1\}$ в $\mathbb{R}^{n+1}$.

Глава 4

Гладкие многообразия (обобщение поверхностей)

В гл. 1 на различных примерах было показано, что основным инструментом при изучении геометрических объектов является система координат, характеризующая расположение точек в пространстве. С помощью системы координат мы можем использовать для решения многих задач методы дифференциального и интегрального исчисления. Поэтому в геометрии выделяют в отдельный раздел изучение таких пространств, которые допускают такие понятия, как дифференцируемые или гладкие функции, операции дифференцирования и интегрирования.

Полезно рассмотреть несколько примеров таких пространств.

Пример 1. Рассмотрим на плоскости (т.е. в двумерном евклидовом пространстве $\mathbb{R}^2$) окружность S^1 единичного радиуса. Для того чтобы описать точки окружности с помощью системы координат, зададим ее как множество точек, удовлетворяющих уравнению

$$x^2 + y^2 = 1, \tag{4.1}$$

где (x, y) – декартовы координаты точки на плоскости. Таким образом, каждая точка $P \in S^1$ однозначно задается парой чисел – декартовыми координатами x и y. Однако для точек окружности излишне задавать обе координаты x и y. Если нам известна координата x точки P, то вторая координата y может быть получена из уравнения (4.1): $y = \pm\sqrt{1 - x^2}$, т.е. вторая координата y однозначно (с точностью до знака) определяется первой координатой x. Более того, если точка $P_0 = (x_0, y_0)$ не находится на оси абсцисс, т.е. $y_0 \neq 0$, то найдется такая достаточно малая окрестность U точки P_0, что для всех точек $P \in U$ выбор знака координаты y точки P определяется однозначно знаком координаты y_0 точки P_0. В некотором смысле мы можем сказать, что точки на окружности S^1 описываются одним числовым параметром – ее первой декартовой координатой x. Более точно, можно утверждать,

что точки, лежащие в верхней половине окружности, т.е. удовлетворяющие неравенству $y > 0$, однозначно задаются одним числовым параметром x. Аналогично, точки, лежащие в нижней половине окружности, т.е. удовлетворяющие неравенству $y < 0$, тоже однозначно задаются одним числовым параметром x. Как для верхней половины окружности, так и для ее нижней половины область изменения параметра x одна и та же, а именно интервал вещественной оси $(-1, 1)$. Если бы мы захотели подобным же образом параметрически описывать точки окружности, включая и оставшиеся "особые" точки $P_0 = (1, 0)$ и $P_1 = (-1, 0)$, то нам бы пришлось поменять местами координаты x и y и выразить из уравнения (4.1) координату x через координату y.

Возникает вопрос: нельзя ли так ввести параметризацию точек на окружности S^1, чтобы сразу все точки окружности однозначно задавались некоторым значением числового параметра? Наилучшим приближением к ответу на этот вопрос является угловой параметр φ, равный углу между осью абсцисс и радиус-вектором с концом в точке P. Но угловой параметр определен неоднозначно. Если же мы ограничимся значением угла φ в некотором интервале, скажем, $0 < \varphi \leqslant 2\pi$, то функция, сопоставляющая точке $P \in S^1$ значение углового параметра φ, будет претерпевать разрыв в точке $P_0 = (1, 0)$.

Таким образом, можно сформулировать следующее утверждение: *на окружности S^1 (как на топологическом пространстве) не существует такой непрерывной функции, значения которой однозначно определяли бы точки окружности.*

Пример 2. В качестве еще одного примера рассмотрим двумерную сферу S^2 в трехмерном евклидовом пространстве, задаваемую уравнением

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

Так же, как и в случае окружности, для точек сферы S^2 излишне задавать все три декартовы координаты (x, y, z), поскольку, скажем, координату z можно выразить через первые две: $z = \pm\sqrt{1 - x^2 - y^2}$. Ясно, что на верхней (и, аналогично, на нижней) полусфере, координата z точки P однозначно восстанавливается по ее первым двум координатам x и y.

Можно показать, что и в случае двумерной сферы невозможно найти две непрерывные функции на сфере S^2, по значениям которых точка P на сфере восстанавливалась бы однозначно.

Рассмотренные примеры показывают, что мы имеем только один выход – отказаться от попытки построения системы координат, единой для всех точек рассматриваемого пространства, и довольствоваться для различных частей пространства своей системой координат. Строгое описание этой конструкции привело к введению специального понятия в геометрии – многообразия.

4.1 Понятие многообразия

4.1.1 Основные определения

Метрическое пространство M называется *n-мерным многообразием* (или просто *многообразием*), если каждая его точка P содержится в окрестности $U \subset M$, гомеоморфной некоторой области V евклидова пространства $\mathbb{R}^n$. Это условие кратко формулируют следующим образом: n-мерное многообразие M локально гомеоморфно области в евклидовом пространстве $\mathbb{R}^n$. При этом говорят, что *размерность многообразия* M равна n и записывают $\dim M = n$. Таким образом, если M — n-мерное многообразие, то в пространстве M можно указать систему открытых множеств $\{U_\alpha\}$, занумерованных некоторым (конечным или бесконечным) множеством индексов i, и гомеоморфизмов $\varphi_\alpha \colon U_\alpha \to V_\alpha \subset \mathbb{R}^n$ множеств U_α на области V_α. При этом система открытых множеств $\{U_\alpha\}$ должна покрывать пространство $M : M = \cup_\alpha U_\alpha$. Области V_α могут, вообще говоря, пересекаться между собой.

Фиксируем в пространстве $\mathbb{R}^n$ некоторую декартову систему координат $(x^1, \ldots, x^n)$, как систему линейных функций на $\mathbb{R}^n$. Тогда композиции $x^k = x^k(P) = x^k\big(\varphi_\alpha(P)\big)$, $P \in U_\alpha$, задают систему функций на открытом множестве U_α, которая называется локальной системой координат на карте U_α.

Совокупность карт $\{U_\alpha\}$, покрывающая все многообразие M, называется *атласом карт*.

Удобно локальные координаты точки $P \in M$ нумеровать дополнительным индексом, нумерующим саму карту $U_\alpha : x^k_\alpha = x^k_\alpha(P)$. Поскольку точка P может принадлежать одновременно нескольким картам, то она приобретает несколько наборов локальных координат.

Рассмотрим простейшие примеры многообразий.

Пример 3. Во введении была рассмотрена окружность $S^1 \subset \mathbb{R}^2$, заданная уравнением $x^2 + y^2 = 1$. Покроем S^1 атласом, состоящим из четырех карт (рис. 4.1)

$$U_1 = \{(x, y) \in S^1 : y > 0\}, \qquad U_2 = \{(x, y) \in S^1 : y < 0\},$$
$$U_3 = \{(x, y) \in S^1 : x > 0\}, \qquad U_4 = \{(x, y) \in S^1 : x < 0\}.$$

Соответствующие им области V_1, V_2, V_3, V_4 на вещественной прямой $\mathbb{R}^1$ совпадают и равны открытому интервалу $(-1, 1)$. Гомеоморфизмы φ_1 и φ_2 построим как проекции окружности на ось абсцисс: $\varphi_1(x, y) = \varphi_2(x, y) = x$, а гомеоморфизмы φ_3, φ_4 — как проекции на ось ординат: $\varphi_3(x, y) = \varphi_4(x, y) = y$. Чтобы доказать, что отображения φ_k, $k = 1, \ldots, 4$, являются гомеоморфизмами, достаточно в явном виде записать формулы для обратных отображений:

$$\varphi_1^{-1}(x) = (x, \sqrt{1 - x^2}) \in S^1, \qquad \varphi_2^{-1}(x) = (x, -\sqrt{1 - x^2}) \in S^1$$
$$\varphi_3^{-1}(y) = (\sqrt{1 - y^2}, y) \in S^1, \qquad \varphi_4^{-1}(y) = (-\sqrt{1 - y^2}, y) \in S^1$$

Рис. 4.1:

Рис. 4.2:

и убедиться, что обратные отображения непрерывны. Тогда на окружности получаются четыре локальных системы координат, каждая из которых состоит из одной координаты: $x_1 = \varphi_1(x, y) = x$, $x_2 = \varphi_2(x, y) = x$, $x_3 = \varphi_3(x, y) = y$, $x_4 = \varphi_4(x, y) = y$. Некоторые точки снабжены сразу двумя локальными системами координат. Так, например, для точек P пересечения $U_1 \cap U_3$ определены координаты $x_1(P)$ и $x_3(P)$ (рис. 4.2).

Существуют и другие способы введения атласа карт на окружности. В гл. 1 были рассмотрены полярные координаты (r, φ) на плоскости. Уравнение окружности в полярных координатах имеет простой вид: $r = 1$. Строго говоря, полярные координаты на плоскости не являются системой координат (см. гл. 1). Поэтому на окружности S^1 введем карты $U_1 = \{(x, y) \in S^1 : x \neq -1\}$, $U_2 = \{(x, y) \in S^1 : x \neq 1\}$ (рис. 4.3). Положим $\varphi_1(P) = \varphi_1(x, y)$ равным значению угла φ, лежащему в промежутке $(-\pi, \pi)$, а $\varphi_2(P) = \varphi_2(x, y)$ — равным значению угла φ, лежащему в промежутке $(0, 2\pi)$, т.е. $V_1 = (-\pi, \pi)$, $V_2 = (0, 2\pi)$. Ясно, что локальные координаты $\varphi_1 = \varphi_1(P)$ и $\varphi_2 = \varphi_2(P)$ совпадают для точек верхней полуокружности и не совпадают для точек нижней полуокружности, т.е. $\varphi_1(x, y) = \varphi_2(x, y)$ при $y > 0$, $\varphi_1(x, y) = \varphi_2(x, y) - 2\pi$ при $y < 0$ (рис. 4.4).

Пример 4. Окружность S^1, рассмотренная в примере 3, представляет собой уже довольно сложное многообразие. Простейшим примером является само евклидово пространство $\mathbb{R}^n$. Атлас карт можно взять состоящим всего

Рис. 4.3:

Рис. 4.4:

лишь из одной карты $U = \mathbb{R}^n$, координатный гомеоморфизм φ — это тождественное отображение $\varphi : U \to V = \mathbb{R}^n$, локальная система координат — это декартовы координаты точек в $\mathbb{R}^n$. Аналогично, любая область $U \subset \mathbb{R}^n$ является n-мерным многообразием, атлас карт которого тоже состоит из одной карты с декартовой системой координат.

Пример 5. Рассмотрим *n-мерную* сферу S^n единичного радиуса, задаваемую как множество точек в $\mathbb{R}^{n+1}$, удовлетворяющих уравнению $(x^1)^2 + (x^2)^2 + \cdots + (x^{n+1})^2 = 1$. Покажем, что n-мерная сфера является n-мерным многообразием. В качестве атласа карт возьмем открытые множества

$$U_i^+ = \{(x^1, \ldots, x^{n+1}) \in S^n : x^i > 0\},$$
$$U_i^- = \{(x^1, \ldots, x^{n+1}) \in S^n : x^i < 0\}.$$

Координатные гомеоморфизмы φ_i^+, φ_i^- определим как ограничения проекций евклидова пространства $\mathbb{R}^{n+1}$ на $\mathbb{R}^n$ вдоль координаты x^i.

Тогда области V_i^+ и V_i^- совпадают друг с другом и равны шару единичного радиуса.

Пример 6. Рассмотрим проективную плоскость $\mathbb{R}P^2$. Мы ее будем представлять как пространство, точками которого объявлены всевозможные прямые, проходящие через начало координат в $\mathbb{R}^3$. Определим расстояние

между двумя прямыми как меньший угол между ними. Тогда $\mathbb{R}P^2$ превращается в метрическое пространство. Покажем, что $\mathbb{R}P^2$ является 2-мерным многообразием. Для этого удобно каждую прямую $P \in \mathbb{R}P^2$ задавать тремя однородными координатами $(x : y : z)$, которые разрешается умножать на одно и то же число $\lambda \neq 0$. Однородные координаты не обращаются в нуль одновременно, т.е. $x^2 + y^2 + z^2 > 0$. Покроем $\mathbb{R}P^2$ тремя картами: $U_1 = \{(x : y : z) | x \neq 0\}$, $U_2 = \{(x : y : z) | y \neq 0\}$, $U_3 = \{(x : y : z) | z \neq 0\}$. Пусть $V_1 = V_2 = V_3 = \mathbb{R}^2$. В качестве координатных гомеоморфизмов возьмем отображения $\varphi_k \colon U_k \to V_k = \mathbb{R}^2$:

$$\varphi_1(x : y : z) = (y/x, z/x), \qquad \varphi_2(x : y : z) = (x/y, z/y),$$
$$\varphi_3(x : y : z) = (x/z, y/z).$$

Таким образом, мы построили три локальных системы координат:

$$x_1^1 = y/x, \quad x_1^2 = z/x; \quad x_2^1 = x/y, \quad x_2^2 = z/y;$$
$$x_3^1 = x/z, \quad x_3^2 = y/z.$$

Рассмотренные примеры показывают, что в одном и том же многообразии M можно фиксировать различные атласы карт. Даже если карты как открытые множества оставить неизменными, можно менять локальную систему координат в карте путем выбора другого координатного гомеоморфизма. Если $U' \subset U$ — открытое подмножество в карте U, то ограничивая координатные функции на подмножество U' получается новая карта, более мелкая, чем карта U.

Пусть $\{U_\alpha\}$, $\{U'_\beta\}$ — два атласа карт на многообразии M. Тогда существует третий атлас карт, измельчающий как $\{U_\alpha\}$, так и атлас $\{U'_\beta\}$. Этот новый атлас можно, например составить из попарных пересечений карт исходных двух атласов.

4.1.2 Функции склейки. Определение гладкого многообразия

Определение 4.1. Пусть M — n-мерное многообразие, $\{U_\alpha\}$ — его атлас карт, φ_α — координатные гомеоморфизмы, $\{x_\alpha^k\}$ — набор локальных систем координат. В каждом пересечении $U_{\alpha\beta} = U_\alpha \cap U_\beta$ двух карт определены две локальные системы координат, $\{x_\alpha^k\}$ и $\{x_\beta^k\}$, причем координатные функции $x_\alpha^k(P)$ одной карты функционально выражаются через координатные функции $x_\beta^k(P)$ другого атласа, $x_\alpha^k(P) = x_\alpha^k\big(x_\beta^1(P), \ldots, x_\beta^n(P)\big)$, $P \in U_{\alpha\beta}$. Функции $x_\alpha^k = x_\alpha^k(x_\beta^1, \ldots, x_\beta^n)$ называются *функциями замены координат* или *функциями перехода* от координат $\{x_\alpha^k\}$ к координатам $\{x_\beta^k\}$.

Функции замены координат определены не во всей области V_β, а в ее некоторой части $V_{\beta\alpha} = \varphi_\beta(U_{\alpha\beta})$, там, где имеет смысл говорить сразу о двух системах координат.

Рис. 4.5:

На рис. 4.5 для удобства области V_α и V_β в евклидовом пространстве показаны как непересекающиеся множества.

Функции замены координат $x^k_\alpha = x^k_\alpha(x^1_\beta, \ldots, x^n_\beta)$ осуществляют совместное отображение области $V_{\beta\alpha}$ в область $V_{\alpha\beta}$, лежащие в $\mathbb{R}^n$:

$$\mathbf{x}_\alpha = \varphi_\alpha \varphi_\beta^{-1}(\mathbf{x}_\beta) = \varphi_{\alpha\beta}(\mathbf{x}_\beta).$$

Определение 4.2. *Гладким n-мерным многообразием* называется n-мерное многообразие M, на котором фиксирован атлас карт $\{U_\alpha\}$ с локальными системами координат $\{x^k_\alpha\}$, удовлетворяющими условию: функции замены координат $x^k_\alpha = x^k_\alpha(x^1_\beta, \ldots, x^n_\beta)$ являются непрерывно дифференцируемыми функциями для любой пары карт U_α и U_β и во всей области их определения.

Определение 4.3. Функция $f\colon M \to \mathbb{R}^1$, определенная на гладком многообразии M, называется *непрерывно дифференцируемой функцией* в точке $P_0 \in M$, если в любой локальной системе координат $(x^1_\alpha, \ldots, x^n_\alpha)$ (карты $U_\alpha \ni P_0$ из фиксированного атласа карт) функция f представляется в виде непрерывно дифференцируемой функции $h(x^1_\alpha, \ldots, x^n_\alpha)$ от n независимых переменных в окрестности точки $\big(x^1_\alpha(P_0), \ldots, x^n_\alpha(P_0)\big)$.

Пример 7. Рассмотрим следующий пример атласа карт на многообразии M. Пусть $M = \mathbb{R}^1$ — вещественная прямая, атлас карт состоит из двух одинаковых карт $U_1 = U_2 = M = \mathbb{R}^1$, но с различными системами координат. На U_1 зададим координату $x_1 = x$, $x \in \mathbb{R}^1$, а на U_2 координату зададим формулой $x_2 = x^3$. Тогда функции замены координат имеют вид

$$x_2 = x_2(x_1) = (x_1)^3, \tag{4.2}$$

$$x_1 = x_1(x_2) = \sqrt[3]{x_2}. \tag{4.3}$$

Если функция замены координат (4.2) является непрерывно дифференцируемой функцией (многочлен!), то функция (4.3) имеет разрывную производную. Поэтому согласно определению 4.2 многообразие M с атласом карт $\{U_1, U_2\}$ не является гладким многообразием.

Замечание. Если атлас карт на многообразии M состоит из одной карты (т.е. M гомеоморфно области в евклидовом пространстве), то M является гладким многообразием.

Определение 4.4. Пусть на многообразии M заданы два атласа карт, $\{U_\alpha\}$ и $\{U'_\beta\}$, и по отношению к каждому из них M является гладким многообразием. Два атласа карт, $\{U_\alpha\}$ и $\{U'_\beta\}$, называются *эквивалентными*, если функции перехода от каждой локальной системы координат в атласе $\{U_\alpha\}$ в каждую локальную систему координат в атласе $\{U'_\beta\}$ являются непрерывно дифференцируемыми.

Определение 4.4 оправдывается тем, что *всякая функция f на многообразии M является непрерывно дифференцируемой в атласе $\{U_\alpha\}$ тогда и только тогда, когда она непрерывно дифференцируема в атласе $\{U'_\beta\}$*. Таким образом, с точки зрения непрерывно дифференцируемых функций на многообразии M эквивалентные атласы карт равноправны, и для представления функций в виде непрерывно дифференцируемых вещественнозначных функций от независимых переменных (координат точки) можно пользоваться любыми из эквивалентных атласов карт.

Напомним, что вещественнозначная функция $h(x^1, \ldots, x^n)$ имеет класс гладкости C^r в окрестности $(x_0^1, \ldots, x_0^n)$, если в окрестности этой точки существуют и непрерывны все частные производные функции h до порядка r включительно. В случае $r = \infty$ это означает, что существуют и непрерывны частные производные функции h всех порядков.

Определение 4.5. Многообразие M с фиксированным атласом карт $\{U_\alpha\}$ называется *гладким многообразием класса C^r*, если все функции замены координат являются функциями класса гладкости C^r во всех точках области их определения.

В дальнейшем, если не оговорено противное, мы будем рассматривать только многообразия класса C^∞, а функции на многообразии — класса гладкости C^∞.

Пример 8. Модифицируем пример 7. Выберем координату во второй карте U_2 равной $x_2 = x + x|x|$. Тогда многообразие M является гладким многообразием класса C^1, но не является гладким многообразием класса C^2.

В примерах 3– 6 рассмотрены многообразия с такими атласами карт, что эти многообразия автоматически являются гладкими многообразиями класса C^∞.

В геометрии рассматриваются и другие, более сильные условия на атласы карт и их функции замены координат. Так, если все функции замены координат являются вещественно аналитическими функциями, т.е. в окрестности каждой точки их определения разлагаются в сходящиеся ряды Тейлора, то само многообразие называется *вещественно аналитическим многообразием*. Вещественно аналитическое многообразие является гладким многообразием класса C^∞.

Более важный класс многообразий — комплексно аналитические многообразия. Пусть M — $2n$-мерное многообразие, $\{U_\beta\}$ — атлас его карт, $\varphi_\beta\colon U_\beta \to V_\beta \subset \mathbb{R}^{2n}$ — его координатные гомеоморфизмы. Отождествим $2n$-мерное евклидово пространство $\mathbb{R}^{2n}$ с n-мерным комплексным линейным пространством $\mathbb{C}^n$, считая, что комплексные координаты точки $(z^1, \ldots, z^n)$ дают $2n$ вещественных координат $(x^1, \ldots, x^n, y^1, \ldots, y^n)$, $z^k = x^k + iy^k$. Тогда $2n$ координатных функций $x^1_\beta(P), \ldots, x^n_\beta(P), y^1_\beta(P), \ldots, y^n_\beta(P)$ в карте U_β превращаются в n комплекснозначных функций $z^k_\beta(P) = x^k_\beta(P) + iy^k_\beta(P)$. Функции $z^k_\beta(P)$ мы назовем комплексными координатами точки в карте U_β. На пересечении двух карт $U_\beta \cap U_\gamma$ мы имели функции перехода из одной системы координат в другую в виде

$$x^k_\beta = x^k_\beta(x^1_\gamma, \ldots, x^n_\gamma, y^1_\gamma, \ldots, y^n_\gamma),$$
$$y^k_\beta = y^k_\beta(x^1_\gamma, \ldots, x^n_\gamma, y^1_\gamma, \ldots, y^n_\gamma),$$

которые можно представить как комплекснозначные функции от n независимых комплексных переменных:

$$z^k_\beta = z^k_\beta(z^1_\gamma, \ldots, z^n_\gamma), \tag{4.4}$$

Функции (4.4) мы назовем функциями перехода или функциями замены комплексных координат.

Многообразие M с фиксированным атласом карт $\{U_\beta\}$ и локальными системами комплексных координат $(z^1_\beta, \ldots, z^n_\beta)$ называется *комплексно аналитическим многообразием*, если все функции замены комплексных координат (4.4) являются комплексно аналитическими функциями, т.е. разлагаются в сходящиеся ряды Тейлора комплексных переменных в окрестности каждой точки своей области определения.

В качестве примера многообразия, допускающего структуру комплексно аналитического многообразия, рассмотрим двумерную сферу S^2 и специальным образом построим на ней атлас карт. В гл. 1 была построена стереографическая проекция сферы $S^2 = \{x^2 + y^2 + z^2 = 1\}$ из северного полюса $P_0 = (0, 0, 1)$ на координатную плоскость (x, y). Обозначим это отображение через φ_0. Отображение φ_0 отображает все точки сферы S^2, за исключением полюса P_0, т.е. открытое множество $U_0 = S^2 \setminus P_0$ гомеоморфно на всю плоскость $V_0 = \mathbb{R}^2$. В декартовых координатах гомеоморфизм φ_0 имеет вид $\varphi_0(x, y, z) = \big(x/(1 - z), y/(1 - z)\big)$. Поэтому введем в карте U_0 одну комплексную координату $w_0 = (x + iy)/(1 - z)$, выраженную через декартовы координаты точки на сфере. Кроме этого, рассмотрим южный полюс $P_1 = (0, 0, -1)$ и стереографическую проекцию φ_1 из южного полюса на ту же координатную плоскость (x, y). Отображение φ_1 гомеоморфно отображает множество $U_1 = S^2 \setminus P_1$ на всю плоскость $V_1 = \mathbb{R}^2$. В декартовых координатах гомеоморфизм φ_1 имеет вид $\varphi_1(x, y, z) = \big(x/(1+z), y/(1+z)\big)$. Введем в карте U_1 комплексную координату $w_1 = (x - iy)/(1 + z)$. Тогда на пересечении $U_0 \cap U_1$ получаем $w_0 w_1 = (x^2 + y^2)/(1 - z^2) \equiv 1$. Следовательно,

$$w_0 = w_0(w_1) = \frac{1}{w_1}, \quad w_1 = w_1(w_0) = \frac{1}{w_0}. \tag{4.5}$$

Функции (4.5) являются комплексно аналитическими функциями. Значит, сфера S^2 является комплексно аналитическим многообразием.

4.1.3 Гладкие отображения многообразий. Диффеоморфизмы

Пусть M_1 и M_2 — два гладких многообразия, $f\colon M_1 \to M_2$ — непрерывное отображение.

Определение 4.6. Отображение $f\colon M_1 \to M_2$ гладких многообразий называется *гладким отображением* класса C^r, если для любых локальных систем координат $(x^1, \ldots, x^n)$ в окрестности любой точки $P_0 \in M_1$ и $(y^1, \ldots, y^m)$ в окрестности точки $Q_0 = f(P_0) \in M_2$ представление функции f в виде вектор-функции $\mathbf{y} = (y^k) = \big(h^k(x^1, \ldots, x^n)\big) = h(\mathbf{x})$ является вектор-функцией класса гладкости C^r.

Отметим, что определение гладкого отображения класса гладкости C^r имеет смысл только в том случае, когда классы гладкости многообразий M_1 и M_2 не меньше, чем r.

Пусть $f\colon M_1 \to M_2$ — гомеоморфизм многообразий. Если f является гладким отображением класса C^r, то обратное отображение f^{-1} не обязано быть гладким отображением. В случае, когда обратное отображение $f^{-1}\colon M_2 \to M_1$ тоже является гладким отображением класса C^r, то гомеоморфизм f называют *гладким гомеоморфизмом класса* C^r или *диффеоморфизмом класса* C^r. Если существует диффеоморфизм $f\colon M_1 \to M_2$, то многообразия M_1 и M_2 называются *диффеоморфными многообразиями*.

Теорема 4.1. *Пусть $f\colon M_1 \to M_2$ — гладкий гомеоморфизм класса C^r $(r \geqslant 1)$ гладких многообразий. Тогда* $\dim M_1 = \dim M_2$.

Доказательство. Выберем локальные координаты $(x^1, \ldots, x^n)$ на многообразии M_1 в окрестности точки P_0 и локальные координаты $(y^1, \ldots, y^m)$ на многообразии M_2 в окрестности точки $Q_0 = f(P_0)$. Пусть $g = f^{-1}$ — обратное к f отображение. Тогда отображения f и g представляются в виде двух вектор-функций, которые мы обозначим теми же буквами, $\mathbf{y} = f(\mathbf{x})$ и $\mathbf{x} = g(\mathbf{y})$, причем $g\big(f(\mathbf{x})\big) \equiv \mathbf{x}$, $f\big(g(\mathbf{y})\big) \equiv \mathbf{y}$. Другими словами, для функций f и g выполнены равенства

$$gf = \mathrm{Id}_x, \qquad fg = \mathrm{Id}_y,$$

где $\mathrm{Id}_x, \mathrm{Id}_y$ — тождественные отображения векторных пространств. Значит, и для матриц Якоби выполняются аналогичные равенства

$$(dg)(df) = E_x, \qquad (df)(dg) = E_y,$$

где E_x, E_y — единичные матрицы соответствующих размерностей.

Из линейной же алгебры известно, что ранг произведения двух матриц не превосходит ранга каждого из сомножителей. Поскольку $\operatorname{rank} dh \leqslant$

$\min(m,n)$, $\operatorname{rank} dh^{-1} \leqslant \min(m,n)$ (матрицы dh и dh^{-1} — прямоугольные!), то $\operatorname{rank} de_0 \leqslant \min(m,n)$, $\operatorname{rank} de_1 \leqslant \min(m,n)$. Следовательно, $m \leqslant \min(m,n)$, $n \leqslant \min(m,n)$, или $\max(m,n) \leqslant \min(m,n)$, что и означает, что $m = n$. Теорема доказана. ■

В заключение приведем два полезных утверждения об атласах карт.

Лемма 4.1. *В гладком многообразии M существует такой атлас карт $\{U_\alpha\}$, что каждая карта U_α диффеоморфна $\mathbb{R}^n$.*

Доказательство. Сначала покажем, что можно построить такой атлас карт, чтобы каждая карта была диффеоморфна открытому шару некоторого радиуса ε в $\mathbb{R}^n$. Пусть $P_0 \in M$ — произвольная точка, $U_\alpha \ni P_0$, $\varphi_\alpha \colon U_\alpha \to V_\alpha \subset \mathbb{R}^n$ — координатный гомеоморфизм, $Q_0 = \varphi_\alpha(P_0)$. Поскольку V_α — открытое множество в $\mathbb{R}^n$, то найдется такое число ε, что открытый шар радиуса ε с центром в точке Q_0 лежит в V_α. Обозначим этот шар через $O_\varepsilon(Q_0)$, а его прообраз $\varphi_\alpha^{-1}\big(O_\varepsilon(Q_0)\big)$ — через W_P. Семейство открытых множеств $\{W_P\}$ является атласом карт на многообразии M, и каждая карта W_P диффеоморфна открытому шару в $\mathbb{R}^n$. Чтобы завершить доказательство леммы, покажем, что открытый шар радиуса ε диффеоморфен $\mathbb{R}^n$. Достаточно рассмотреть случай $\varepsilon = 1$. Итак, пусть $(x^1, \ldots, x^n)$ — точка в шаре радиуса 1, $(x^1)^2 + \cdots + (x^n)^2 < 1$. Положим

$$\begin{aligned} y^k &= \frac{x^k}{\sqrt{1 - (x^1)^2 - \cdots - (x^n)^2}}, \\ x^k &= \frac{y^k}{\sqrt{1 + (y^1)^2 + \cdots + (y^n)^2}}. \end{aligned} \tag{4.6}$$

Функции (4.6) являются гладкими и осуществляют взаимно обратное отображения шара радиуса 1 и $\mathbb{R}^n$. Лемма доказана. ■

Лемма 4.2. *Пусть M — гладкое компактное многообразие, $\{U_\alpha\}$ — атлас карт на нем. Тогда существует гладкое разбиение единицы ψ_α, подчиненное покрытию $\{U_\alpha\}$, т.е.*

$$0 \leqslant \psi_\alpha \leqslant 1, \quad \sum_\alpha \psi_\alpha = 1, \quad \operatorname{supp} \psi_\alpha \subset U_\alpha.$$

Доказательство. Достаточно считать на основании леммы 4.1, что все карты гомеоморфны шару радиуса 1. Пусть $\varphi_\alpha \colon U_\alpha \to D_1^n \subset \mathbb{R}^n$ — координатные гомеоморфизмы. (Здесь D_r^n — открытый шар радиуса r с центром в начале координат.) Выберем такое достаточно малое $\varepsilon > 0$, чтобы $\varphi_\alpha^{-1}(D^n_{(1-\varepsilon)})$ покрывали многообразие M. Допустим, что существует такая гладкая функция f класса C^∞ на шаре D_1^n, что $\operatorname{supp} f = D^n_{(1-\varepsilon)}$, $0 \leqslant f \leqslant 1$. Положим

$$\overline{\psi}_\alpha(P) = \begin{cases} 0, & \text{если} \quad P \notin U_\alpha, \\ f\big(\varphi_\alpha(P)\big), & \text{если} \quad P \in U_\alpha. \end{cases}$$

Поскольку $f\big(\varphi_\alpha(P)\big) = 0$ при $P \notin \varphi_\alpha^{-1}(D^n_{(1-\varepsilon)})$, то функции $\overline{\psi}_\alpha$ гладкие на многообразии M, причем $\operatorname{supp}\overline{\psi}_\alpha \subset U_\alpha$, $0 \leqslant \overline{\psi}_\alpha \leqslant 1$. Более того, $\operatorname{supp}\overline{\psi}_\alpha \supset \varphi_\alpha^{-1}(D^n_{(1-\varepsilon)})$. Следовательно, сумма функций $\overline{\psi}_\alpha$, $\overline{\psi}(P) = \sum\limits_\alpha \overline{\psi}_\alpha(P)$, строго больше нуля в каждой точке. Тогда положим $\psi_\alpha(P) = \overline{\psi}_\alpha(P)/\overline{\psi}(P)$. Функции $\psi_\alpha(P)$ образуют гладкое разбиение единицы, подчиненное покрытию $\{U_\alpha\}$.

Таким образом, нам осталось построить такую функцию f класса C^∞ в $\mathbb{R}^n$, чтобы ее носитель был шар $D^n_{(1-\varepsilon)}$. Будем искать функцию f в виде $f(x^1,\dots,x^n) = h\big((x^1)^2 + \dots + (x^n)^2\big)$. Следовательно, достаточно построить гладкую функцию одной переменной $h(x)$, чтобы выполнялись условия: $h(x) = 0$ при $x > (1-\varepsilon)^2$ и $h(x) > 0$ при $x < (1-\varepsilon)^2$. В качестве h возьмем функцию

$$h(x) = \begin{cases} e^{-1/(x-(1-\varepsilon)^2)^2}, & x < (1-\varepsilon)^2, \\ 0, & x \geqslant (1-\varepsilon)^2, \end{cases}$$

которая, как известно, является гладкой класса C^∞. Лемма доказана. ■

Задачи

1. Доказать, что пространство расположений жесткого отрезка на плоскости является гладким многообразием.

2. Доказать, что группа SO(3) гомеоморфна трехмерному проективному пространству.

3. Описать конфигурационное пространство системы из двух шарнирно соединенных стержней в трехмерном пространстве.

4. Привести пример гладкого взаимно однозначного отображения, не являющегося диффеоморфизмом.

5. Показать, что на сфере S^n не существует атласа, состоящего из одной карты.

4.2 Задание многообразий уравнениями в $\mathbb{R}^n$

Ранее многие примеры многообразий возникали в виде множества решений некоторого нелинейного уравнения, заданного в евклидовом пространстве. Так, например, n-мерная сфера S^n задается уравнением $(x^1)^2 + (x^2)^2 + \dots + (x^{n+1})^2 = 1$ в евклидовом пространстве $\mathbb{R}^{n+1}$; псевдосфера S^2_1 задается уравнением $x^2 + y^2 - z^2 = -1$. Вообще, если $f(x^1,\dots,x^n)$ — непрерывно дифференцируемая функция, то множество решений уравнения $f(x^1,\dots,x^n) - c = 0$ называется *многообразием уровня c функции f*. Таким образом, все евклидово пространство $\mathbb{R}^n$ разлагается в объединение многообразий уровня функций f. В случае функции двух переменных решения уравнения обычно называют *линиями уровня* функции f, а в случае функции трех переменных — *поверхностями уровня*.

Рис. 4.6:

Чтобы оправдать название "многообразие уровня функции f" необходимо было бы доказать, что многообразие уровня функции f действительно является многообразием. Однако это не всегда так.

Пример 9. Рассмотрим функцию $f(x,y) = x^2 - y^2$. Ее линии уровня описываются уравнением $x^2 - y^2 = c$. Если $c > 0$, то линия уровня состоит из двух компонент связности, каждая из которых может быть описана одним из уравнений $x = \sqrt{c+y^2}$, $x = -\sqrt{c+y^2}$, т.е. является графиком функции одной переменной (рис. 4.6). Аналогично, при $c < 0$ линия уровня состоит из графиков двух функций: $y = \sqrt{x^2 - c}$, $y = -\sqrt{x^2 - c}$. Таким образом, при $c \neq 0$ линии уровня являются одномерными многообразиями. Особый случай представляет линия уровня при $c = 0$. В этом случае линия уровня состоит из пары пересекающихся прямых: $x = y$, $x = -y$, и не является многообразием.

Тем не менее в некотором смысле почти всегда многообразия уровня непрерывно дифференцируемой функции f являются многообразиями.

Теорема 4.2. *Пусть $f = f(x^1, \ldots, x^n)$ — функция класса C^∞, определенная во всем евклидовом пространстве $\mathbb{R}^n$. Положим $M_c = \{(x^1, \ldots, x^n) | f(x^1, \ldots, x^n) = c\}$. Если градиент функции f отличен от нуля в каждой точке множества M_c, то M_c является гладким $(n-1)$-мерным многообразием класса C^∞. При этом в окрестности каждой точки $P_0 \in M_c$ в качестве локальных координат можно взять некоторые $n-1$ декартовых координат объемлющего евклидова пространства $\mathbb{R}^n$.*

Доказательство. По сути дела теорема 4.2 является теоремой о неявной функции, сформулированной в удобных для нас терминах. Фиксируем некоторую точку $P_0 \in M_c$, $P_0 = (x_0^1, \ldots, x_0^n)$. Поскольку

$$\operatorname{grad}_{P_0} f \neq 0, \quad \operatorname{grad} f = \left(\frac{\partial f}{\partial x^1}, \ldots, \frac{\partial f}{\partial x^n} \right),$$

то найдется отличная от нуля в точке P_0 частная производная. Без ограничения общности мы можем считать, что $\dfrac{\partial f}{\partial x^n}(x_0^1, \ldots, x_0^n) \neq 0$. Пусть

$Q_0 = (x_0^1, \ldots, x_0^{n-1})$ — точка в $\mathbb{R}^{n-1}$, являющаяся образом точки P_0 при проекции вдоль координатной оси x^n.

Согласно теореме о неявной функции существуют такая окрестность $V_0 \ni Q_0$ точки Q_0, интервал $(x_0^n - \delta, x_0^n + \delta)$ и такая непрерывная функция $y = y(x^1, \ldots, x^{n-1})$ класса гладкости C^∞, определенная в окрестности V_0, что:

а) $f\big(x^1, \ldots, x^{n-1}, y(x^1, \ldots, x^{n-1})\big) \equiv c$ в области V_0;

б) $x_0^n = y(x_0^1, \ldots, x_0^{n-1})$;

в) $|x_0^n - y(x^1, \ldots, x^{n-1})| < \delta$ в области V_0;

г) всякое решение $(x^1, \ldots, x^n) \in V_0 \times (x_0^n - \delta, x_0^n + \delta)$ уравнения $f(x^1, \ldots, x^n) = c$ имеет вид $x^n = y(x^1, \ldots, x^{n-1})$.

Тогда через U_0 обозначим окрестность точки $P_0 \in M_c$, равную $U_0 = M_c \cap \big(V_0 \times (x_0^n - \delta, x_0 + \delta)\big)$. Окрестность U_0 и будет искомой картой, содержащей точку P_0. В качестве координатного гомеоморфизма φ_0 возьмем ограничение на U_0 проекции $\mathbb{R}^n$ в $\mathbb{R}^{n-1}$, $\varphi_0(x^1, \ldots, x^n) = (x^1, \ldots, x^{n-1}) \in V^0$. Обратное отображение φ_0^{-1} зададим равенством

$$\varphi_0^{-1}(x^1, \ldots, x^{n-1}) = \big(x^1, \ldots, x^{n-1}, y(x^1, \ldots, x^{n-1})\big).$$

Из условия в) следует, что $\varphi_0^{-1}(x^1, \ldots, x^{n-1}) \in V_0 \times (x_0^n - \delta, x_0^n + \delta)$, а из условия а), что $\varphi_0^{-1}(x^1, \ldots, x^{n-1}) \in M_c$. Таким образом, $\varphi_0^{-1}(x^1, \ldots, x^{n-1}) \in U_0$. Отображения φ_0 и φ_0^{-1} непрерывны и взаимно обратны.

Мы доказали, что множество M_c является $(n-1)$-мерным многообразием, и указали в окрестности каждой точки $P_0 \in M_c$ локальную систему координат, составленную из некоторых декартовых координат евклидова пространства $\mathbb{R}^n$. Докажем теперь, что функции замены координат являются гладкими функциями. Пусть точка P_0 содержится еще в одной карте U_1 и в качестве локальных координат в карте U_1 взяты декартовы координаты $x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^n$. Тогда на пересечении карт $U_0 \cap U_1$ координаты $x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^n$ выражаются через координаты $x^1, \ldots, x^{n-1}$ как функции следующим образом:

$$\begin{aligned} &x^1 = x^1, \\ &\ldots\ldots\ldots \\ &x^{n-1} = x^{n-1}, \\ &x^n = y(x^1, \ldots, x^{n-1}). \end{aligned} \tag{4.7}$$

Поскольку функция $y = y(x^1, \ldots, x^{n-1})$ имеет класс гладкости C^∞, то и все функции (4.7) будут класса гладкости C^∞. Теорема 4.2 полностью доказана. ■

Пример 10. Снова рассмотрим n-мерную сферу S^n, заданную уравнением $f(x^1, \ldots, x^{n+1}) = \sum\limits_{k=1}^{n+1} (x^k)^2 = 1$. Градиент функции f равен $\operatorname{grad} f = (2x^1, 2x^2, \ldots, 2x^{n+1})$. Если точка $P = (x^1, x^2, \ldots, x^{n+1})$ лежит на сфере S^n, то не все ее координаты обращаются в нуль; следовательно, одна из координат градиента отлична от нуля. Выполнены условия теоремы 4.2, и, значит, сфера S^n является гладким многообразием класса C^∞.

Пример 11. Рассмотрим евклидово пространство $\mathbb{R}^{n^2}$ размерности n^2. Будем представлять точки пространства $\mathbb{R}^{n^2}$ в виде квадратных матриц A порядка n с координатами $A = (a_{ij})$. Рассмотрим множество $\mathrm{SL}(n, \mathbb{R})$ всех матриц $A \in \mathbb{R}^{n^2}$, детерминант которых равен единице: $\det A = 1$. Множество $\mathrm{SL}(n, \mathbb{R})$ является группой относительно операции умножения матриц и называется *специальной линейной группой*. Покажем, что группа $\mathrm{SL}(n, \mathbb{R})$ является гладким многообразием класса C^∞ и размерности $(n^2 - 1)$.

Рассмотрим функцию f от n^2 переменных $f(a_{ij}) = \det(a_{ij})$. Функция f является многочленом, и, значит, имеет класс гладкости C^∞. Чтобы применить теорему 4.2, следует вычислить градиент функции f во всех точках группы $\mathrm{SL}(n, \mathbb{R})$. Пусть E — единичная матрица. Поскольку $\det E = 1$, то $E \in \mathrm{SL}(n, \mathbb{R})$.

Вычислим градиент функции f в точке E. Для этого разложим сначала $\det A$ по первой строке:

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{n+1} a_{1n} \det A_{1n}. \tag{4.8}$$

В разложении (4.8) стоят детерминанты матриц A_{1k}, которые являются многочленами всех переменных a_{ij}, кроме элементов первой строки. Тогда частная производная функции f по переменной a_{11} имеет вид $\partial f / \partial a_{11} = \partial(a_{11} \det A_{11}) / \partial a_{11} = \det A_{11}$. В точке E получаем

$$\frac{\partial f}{\partial a_{11}}(E) = 1. \tag{4.9}$$

Таким образом, градиент функции f в точке E отличен от нуля.

Покажем теперь, что в произвольной точке $A_0 \in \mathrm{SL}(n, \mathbb{R})$ градиент функции f тоже отличен от нуля. Введем новые переменные b_{ij}, задаваемые равенством $(b_{ij}) = B = A_0^{-1} A = A_0^{-1}(a_{ij})$. Если $A = A_0$, то $B = E$. Тогда

$$f(A) = f(A_0 B) = \det(A_0 B) = \det A_0 \det B = f(B).$$

Тогда, дифференцируя суперпозицию функций, получаем

$$\frac{\partial f}{\partial b_{11}}(E) = \sum_{ij} \frac{\partial f}{\partial a_{ij}}(A_0) \cdot \frac{\partial a_{ij}}{\partial b_{11}}. \tag{4.10}$$

Левая часть равенства (4.10) равна единице согласно формуле (4.9). Следовательно, в правой части хотя бы одно слагаемое отлично от нуля. Значит, одна из частных производных $\dfrac{\partial f}{\partial a_{ij}}(A_0)$, а вместе с ней и градиент функции f отличны от нуля. Таким образом, выполнены условия теоремы 4.2, и, значит, группа $\mathrm{SL}(n, \mathbb{R})$ является гладким многообразием размерности $(n^2 - 1)$.

Теорема 4.2 легко обобщается на случай систем нелинейных уравнений. Заметим, что условие теоремы 4.2 можно сформулировать следующим образом. Градиент функции f представится в виде столбца частных производных функции f, и, значит, является матрицей Якоби функции f. Тогда

условие нетривиальности градиента f в некоторой точке $P_0 \in \mathbb{R}^n$ эквивалентно тому, что ранг матрицы Якоби df функции f равен единице, т.е. максимален.

Пусть задана система уравнений

$$\begin{cases} f^1(x^1, \ldots, x^n) = c^1, \\ \ldots\ldots\ldots\ldots\ldots \\ f^k(x^1, \ldots, x^n) = c^k, \end{cases} \tag{4.11}$$

которую кратко можно записать в виде $f(\mathbf{x}) = \mathbf{c}$, где $\mathbf{x} = (x^1, \ldots, x^n) \in \mathbb{R}^n$, $\mathbf{c} = (c^1, \ldots, c^k) \in \mathbb{R}^k$, а f — отображение, определяемое функциями $(f^1, \ldots, f^k)$. Множество M решений системы (4.11) будем называть многообразием уровня системы функций $(f^1, \ldots, f^k)$.

Теорема 4.3. *Пусть $f\colon \mathbb{R}^n \to \mathbb{R}^k$ — отображение класса гладкости C^∞, M — множество решений системы уравнений $f(\mathbf{x}) = \mathbf{c}$. Если ранг матрицы Якоби отображения f максимален в каждой точке $P_0 \in M$ (т.е. $\operatorname{rank} df(P_0) = k$), то M является $(n-k)$-мерным гладким многообразием класса C^∞. При этом в окрестности каждой точки $P_0 \in M$ в качестве локальных координат можно взять некоторые $(n-k)$ декартовых координат объемлющего евклидова пространства $\mathbb{R}^n$.*

Доказательство дословно повторяет доказательство теоремы 4.2 с той лишь разницей, что вместо одной переменной x^n выделяется k переменных $x^{i_1}, \ldots, x^{i_k}$. Обозначив эту группу переменных одной буквой, скажем, $\mathbf{y} = (x^{i_1}, \ldots, x^{i_k})$, мы будем получать те же формулы, что и в доказательстве теоремы 4.2. ■

Пример 12. Рассмотрим в евклидовом пространстве $\mathbb{R}^4$ с координатами x^1, x^2, x^3, x^4 систему двух уравнений

$$(x^1)^2 + (x^2)^2 = 1, \quad (x^3)^2 + (x^4)^2 = 1. \tag{4.12}$$

Соответствующие им функции f_1 и f_2 имеют вид

$$f^1(x^1, x^2, x^3, x^4) = (x^1)^2 + (x^2)^2,$$
$$f^2(x^1, x^2, x^3, x^4) = (x^3)^2 + (x^4)^2.$$

Чтобы применить теорему 4.3, вычислим матрицу Якоби отображения $f = (f^1, f^2)$:

$$df = \begin{pmatrix} \dfrac{\partial f^1}{\partial x^1} & \dfrac{\partial f^1}{\partial x^2} & \dfrac{\partial f^1}{\partial x^3} & \dfrac{\partial f^1}{\partial x^4} \\[2ex] \dfrac{\partial f^2}{\partial x^1} & \dfrac{\partial f^2}{\partial x^2} & \dfrac{\partial f^2}{\partial x^3} & \dfrac{\partial f^2}{\partial x^4} \end{pmatrix} = \begin{pmatrix} 2x^1 & 2x^2 & 0 & 0 \\ 0 & 0 & 2x^3 & 2x^4 \end{pmatrix}.$$

Ясно, что $\operatorname{rank} df \leqslant 1$ только в том случае, когда одна из строк матрицы Якоби равна нулю, что невозможно в точках решения системы (4.12). Таким

образом, решения системы (4.12) образуют гладкое двумерное многообразие класса C^∞. Поскольку система (4.12) распадается на два уравнения, каждое от своей группы переменных, то и множество решений можно представить в виде декартова произведения решений каждого уравнения в отдельности, т.е. решения системы (4.11) представляются в виде произведения двух экземпляров окружности. Это многообразие называется (двумерным) *тором*.

4.3 Касательные векторы и касательное пространство

В гл. 1 мы видели, что для изучения метрических свойств кривых и поверхностей и вообще метрических свойств областей евклидового пространства важную роль играют так называемые инфинитезимальные свойства пространств. Это такие свойства, которые определяются в очень малой окрестности фиксированной точки P путем пренебрежения величинами большего порядка малости, чем расстояние до точки P. В математическом анализе существуют аналогичные процедуры пренебрежения бесконечно малыми величинами при изучении поведения функции в окрестности некоторой точки. При изучении гладких многообразий тоже возникает естественное желание осуществлять процедуру пренебрежения бесконечно малыми величинами. Один из таких способов заключается во введении специальных понятий, аналогичных касательным векторам к кривым и касательным плоскостям к поверхностям.

4.3.1 Простейшие примеры

Рассмотрим гладкую кривую в трехмерном пространстве $\mathbb{R}^3$, параметризованную некоторым параметром t: $\mathbf{x} = \mathbf{x}(t) = \big(x^1(t), x^2(t), x^3(t)\big)$. Фиксируем значение параметра t_0. Применим к вектор-функции $\mathbf{x} = \mathbf{x}(t)$ разложение по формуле Тейлора в окрестности точки t_0:

$$\mathbf{x}(t_0 + \Delta t) = \mathbf{x}(t_0) + \frac{d\mathbf{x}}{dt}(t_0)\Delta t + O(\Delta t^2). \tag{4.13}$$

Первые два слагаемых в правой части (4.13) можно, с одной стороны, рассматривать как некоторую аппроксимацию вектор-функции $\mathbf{x}(t)$ в окрестности точки t_0 с помощью линейной вектор-функции. С другой стороны, эта линейная функция $\mathbf{y}(\Delta t) = \mathbf{x}(t_0) + \dfrac{d\mathbf{x}}{dt}(t_0)\Delta t$ задает в $\mathbb{R}^3$ параметризованное представление прямой, проходящей через точку $P_0 = \mathbf{x}(t_0)$. Эта прямая называется касательной прямой к кривой $\mathbf{x}(t)$ в точке P_0, а направляющий вектор $\dfrac{d\mathbf{x}}{dt}(t_0)$ называется касательным вектором к кривой $\mathbf{x}(t)$ в точке P_0 (рис. 4.7).

Рассмотрим теперь поверхность M в трехмерном пространстве $\mathbb{R}^3$, заданную в параметрическом виде как вектор-функция $\mathbf{x} = \mathbf{x}(u, v)$ от двух

Рис. 4.7:

Рис. 4.8:

независимых параметров u, v. Поверхность $\mathbf{x}(u, v)$ называется *невырожденной*, если в каждой точке частные производные $\dfrac{\partial \mathbf{x}}{\partial u}(u, v)$ и $\dfrac{\partial \mathbf{x}}{\partial v}(u, v)$ линейно независимы как векторы в $\mathbb{R}^3$.

Разлагая функцию $\mathbf{x}(u, v)$ по формуле Тейлора в точке (u_0, v_0):

$$\mathbf{x}(u_0 + \Delta u, v_0 + \Delta v) = \mathbf{x}(u_0, v_0) + \\ + \frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)\Delta u + \frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)\Delta v + O(\Delta u^2 + \Delta v^2), \quad (4.14)$$

получаем, что линейная часть разложения (4.14) определяет параметрическое представление плоскости, которая называется касательной плоскостью П к поверхности M в точке $P_0 = \mathbf{x}(u_0, v_0)$ (рис. 4.8).

Касательным вектором к поверхности M в точке P_0 называется всякий вектор с началом в точке P_0, лежащий в касательной плоскости П. Касательный вектор ξ разлагается в линейную комбинацию векторов $\dfrac{\partial \mathbf{x}}{\partial u}(u_0, v_0)$ и $\dfrac{\partial \mathbf{x}}{\partial v}(u_0, v_0)$:

$$\xi = \frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)\Delta u + \frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)\Delta v$$

для подходящего выбора параметров Δu и Δv. Таким образом, векторы $\dfrac{\partial \mathbf{x}}{\partial u}(u_0, v_0)$ и $\dfrac{\partial \mathbf{x}}{\partial v}(u_0, v_0)$ образуют базис в касательной плоскости П, а вели-

чины Δu и Δv служат линейными координатами касательного вектора ξ в этом базисе.

Проведем теперь на поверхности M гладкую кривую $\mathbf{x} = \mathbf{x}(t)$, проходящую через точку P_0. Поскольку кривая $\mathbf{x} = \mathbf{x}(t)$ лежит на поверхности M, то ее параметрически можно представить в виде композиции

$$\mathbf{x}(t) = \mathbf{x}\big(u(t), v(t)\big) \tag{4.15}$$

для некоторых функций $u(t)$, $v(t)$. Иначе это можно сказать следующим образом: функции $u(t)$, $v(t)$ являются параметрическим заданием кривой в локальной системе координат (u, v) на поверхности M. Тогда условие прохождения кривой через точку P_0 перепишем в виде условия на координаты $u_0 = u(t_0)$, $v_0 = v(t_0)$. Вычислим теперь касательный вектор к кривой (или, как его иначе называют, вектор скорости кривой):

$$\begin{aligned}\frac{d\mathbf{x}}{dt}(t_0) &= \frac{d}{dt}\mathbf{x}\big(u(t), v(t)\big)\big|_{t=t_0} = \frac{\partial \mathbf{x}}{\partial u}\big(u(t_0), v(t_0)\big)\frac{du}{dt}(t_0) + \\ &+ \frac{\partial \mathbf{x}}{\partial v}\big(u(t_0), v(t_0)\big)\frac{dv}{dt}(t_0) = \frac{\partial \mathbf{x}}{\partial u}(u_0, v_0)\frac{du}{dt}(t_0) + \frac{\partial \mathbf{x}}{\partial v}(u_0, v_0)\frac{dv}{dt}(t_0).\end{aligned}$$

Следовательно, касательный вектор к кривой, лежащей на поверхности M, лежит в касательной плоскости.

Определение 4.7. Пусть $\xi = \dfrac{\partial \mathbf{x}}{\partial u}(u_0, v_0)\xi^1 + \dfrac{\partial \mathbf{x}}{\partial v}(u_0, v_0)\xi^2$ — касательный вектор к поверхности M в точке P_0. Тогда числа (ξ^1, ξ^2) назовем *координатами касательного вектора* ξ к поверхности M в точке P_0 по отношению к локальной системе координат (u, v) на поверхности M.

Если (u', v') — другая система координат, то координаты u и v выражаются как гладкие функции от координат u', v': $u = u(u', v')$, $v = v(u', v')$, $u_0 = u(u'_0, v'_0)$, $v_0 = v(u'_0, v'_0)$. Тогда, рассматривая композиции функций, мы получаем новое параметрическое задание поверхности M:

$$\mathbf{x} = \mathbf{x}\big(u(u', v'), v(u', v')\big).$$

Поэтому кривая (4.15) может быть записана с помощью некоторых функций $u'(t)$, $v'(t)$ так, чтобы $\mathbf{x} = \mathbf{x}\big(u(u'(t), v'(t)), v(u'(t), v'(t))\big)$. Тогда согласно определению координат касательного вектора к кривой по отношению к этой новой локальной системе координат (u', v') пара чисел $\big(du'(t_0)/dt, dv'(t_0)/dt\big)$ является координатами касательного вектора к кривой. Дифференцируя сложные функции, получим зависимость между координатами касательного вектора к кривой в различных локальных системах координат:

$$\begin{aligned}\frac{du}{dt}(t_0) &= \frac{\partial u}{\partial u'}(u'_0, v'_0)\frac{du'}{dt}(t_0) + \frac{\partial u}{\partial v'}(u'_0, v'_0)\frac{dv'}{dt}(t_0), \\ \frac{dv}{dt}(t_0) &= \frac{\partial v}{\partial u'}(u'_0, v'_0)\frac{du'}{dt}(t_0) + \frac{\partial v}{\partial v'}(u'_0, v'_0)\frac{dv'}{dt}(t_0).\end{aligned} \tag{4.16}$$

4.3.2 Общее определение касательного вектора

Рассмотренные примеры позволяют дать следующее общее определение касательного вектора к многообразию.

Определение 4.8. Пусть M — гладкое n-мерное многообразие, $P_0 \in M$ — произвольная точка. *Касательным вектором* ξ *в точке* P_0 *к многообразию* M называется соответствие, которое каждой локальной системе координат $(x^1_\alpha, \dots, x^n_\alpha)$ сопоставляет набор чисел $(\xi^1_\alpha, \dots, \xi^n_\alpha)$, удовлетворяющий следующему соотношению для каждой пары локальных систем координат:

$$\xi^k_\alpha = \sum_{l=1}^{n} \frac{\partial x^k_\alpha}{\partial x^l_\beta}(P_0)\xi^l_\beta. \tag{4.17}$$

Числа $(\xi^1_\alpha, \dots, \xi^n_\alpha)$ называются *координатами касательного вектора* ξ в локальной системе координат $(x^1_\alpha, \dots, x^n_\alpha)$. Соотношение (4.17) называется *тензорным законом преобразования координат* касательного вектора ξ при заменах локальных координат.

Определение 4.8 обобщает понятие координат касательного вектора к кривой на поверхности. Закон (4.16) изменения этих координат является частным случаем тензорного закона (4.17) преобразования координат касательного вектора к многообразию. Более того, всякая гладкая кривая на гладком многообразии снабжается в каждой своей точке касательным вектором в смысле определения 4.8. Это важное свойство мы сформулируем в виде предложения.

Предложение 4.1. *Пусть M — гладкое многообразие, $\gamma\colon (-1,1) \to M$ — гладкое отображение интервала $(-1,1)$ в многообразие M. Тогда соответствие, которое каждой локальной системе координат $(x^1, \dots, x^n)$ в окрестности точки $P_0 = \gamma(0)$ сопоставляет набор чисел $\Big(\dfrac{dx^1}{dt}(\gamma(t)), \dots, \dfrac{dx^n}{dt}(\gamma(t))\Big)\Big|_{t=0}$, является касательным вектором в смысле определения* 4.8.

Доказательство. Для доказательства предложения 4.1 достаточно проверить тензорный закон преобразования координат (4.17). Положим $\xi^k_\beta = \dfrac{dx^k_\beta}{dt}(\gamma(t))\big|_{t=0}$, где $(x^1_\beta, \dots, x^n_\beta)$ — локальная система координат на многообразии M в окрестности точки P_0. Тогда для двух локальных систем координат получим:

$$\xi^k_\alpha = \frac{d}{dt}x^k_\alpha(\gamma(t))\big|_{t=0} = \frac{d}{dt}x^k_\alpha(x^1_\beta(\gamma(t)), \dots, x^n_\beta(\gamma(t)))\big|_{t=0} =$$
$$= \sum_{l=1}^{n} \frac{\partial x^k_\alpha}{\partial x^l_\beta}(\gamma(t))\frac{d}{dt}x^l_\beta(\gamma(t))\big|_{t=0} = \sum_{l=1}^{n} \frac{\partial x^k_\alpha}{\partial x^l_\beta}(P_0)\xi^l_\beta.$$

Это и есть тензорный закон (4.17). Предложение доказано. ■

Таким образом, указанное в предложении 4.1 соответствие естественно называть *касательным вектором к кривой* γ или *вектором скорости кривой* γ. Касательный вектор к кривой будем обозначать через $\dfrac{d\gamma}{dt}(t_0)$ или $\dot{\gamma}(t_0)$.

4.3.3 Касательное пространство $T_{P_0}(M)$

Множество всех касательных векторов в фиксированной точке P_0 к многообразию M называется *касательным пространством к многообразию* M *в точке* P_0. Это множество обозначается через $T_{P_0}(M)$. Каждый касательный вектор $\xi \in T_{P_0}(M)$ однозначно определяется своими компонентами в одной фиксированной системе координат. Действительно, если задан набор чисел $(\mu^1, \ldots, \mu^n)$ и мы считаем этот набор компонент искомого касательного вектора в фиксированной локальной системе координат $(x^1_{\alpha_0}, \ldots, x^n_{\alpha_0})$, т.е. $\mu^k = \xi^k_{\alpha_0}$, то для задания всего касательного вектора необходимо определить его компоненты в каждой локальной системе координат $(x^1_\alpha, \ldots, x^n_\alpha)$. Положим для этого

$$\xi^k_\alpha = \sum_{l=1}^{n} \frac{\partial x^k_\alpha}{\partial x^l_{\alpha_0}}(P_0)\mu^l.$$

Полученные компоненты удовлетворяют тензорному закону преобразования координат (4.17):

$$\sum_{l=1}^{n} \frac{\partial x^k_\alpha}{\partial x^l_{\alpha_0}}(P_0)\mu^l = \sum_{s=1}^{n} \frac{\partial x^k_\alpha}{\partial x^s_\beta}(P_0) \sum_{l=1}^{n} \frac{\partial x^s_\beta}{\partial x^l_{\alpha_0}}(P_0)\mu^l =$$
$$= \sum_{l=1}^{n} \Big(\sum_{s=1}^{n} \frac{\partial x^k_\alpha}{\partial x^s_\beta}(P_0) \frac{\partial x^s_\beta}{\partial x^l_{\alpha_0}}(P_0) \Big) \mu^l.$$

Поскольку $\partial x^k_\alpha(P_0)/\partial x^l_{\alpha_0} = \sum\limits_{s=1}^{n} \partial x^k_\alpha(P_0)/\partial x^s_\beta \cdot \partial x^s_\beta(P_0)/\partial x^l_{\alpha_0}$ (закон изменения матрицы Якоби тройной замены координат), то соотношение (4.17) выполняется тождественно.

Мы установили, таким образом, что множество всех касательных векторов к многообразию M в точке P_0 однозначно описывается своими компонентами в одной фиксированной локальной системе координат. Следовательно, все касательное пространство $T_{P_0}(M)$ отождествляется с арифметическим векторным пространством $\mathbb{R}^n$. Это значит, что касательное пространство $T_{P_0}(M)$ можно снабдить структурой линейного пространства. Структура линейного пространства в $T_{P_0}(M)$ не зависит от выбора локальной системы координат в окрестности точки P_0, т.е. справедливо следующее

Предложение 4.2. *Операции сложения векторов и умножения вектора на число в касательном пространстве* $T_{P_0}(M)$ *не зависят от выбора локальной системы координат на многообразии* M *в окрестности точки* P_0.

Доказательство. Проверка справедливости предложения основана на линейности тензорного закона (4.17) преобразования координат касательных векторов. ■

4.3.4 Производная функции по направлению

Существует способ представлять касательный вектор на многообразии M в виде операции дифференцирования гладких функций.

Определение 4.9. Пусть $P_0 \in M$, $\xi \in T_{P_0}(M)$, $\gamma(t)$ — гладкая кривая, проходящая через точку P_0, $\gamma(t_0) = P_0$ и ее касательный вектор в точке P_0 равен ξ. Пусть f — гладкая функция на многообразии M. Число

$$\frac{d}{dt} f\big(\gamma(t)\big)\big|_{t=t_0} = \xi(f) \tag{4.18}$$

называется *производной функции f по направлению касательного вектора ξ*. Операция взятия производной называется *дифференцированием функции f по направлению вектора ξ*.

Теорема 4.4. *Пусть $(x^1, \ldots, x^n)$ — локальная система координат в окрестности точки $P_0 = (x_0^1, \ldots, x_0^n)$ многообразия M, $\xi = (\xi^1, \ldots, \xi^n)$ — касательный вектор к многообразию M в точке P_0, $f = f(x^1, \ldots, x^n)$ — гладкая функция в окрестности точки P_0, представленная как функция от локальных координат $(x^1, \ldots, x^n)$. Тогда*

$$\xi(f) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\xi^i. \tag{4.19}$$

Следовательно, определение (4.18) *производной не зависит от выбора кривой γ, а правая часть* (4.19) *не зависит от выбора локальной системы координат.*

Доказательство. Представим кривую $\gamma(t)$, проходящую через точку P_0, в координатном виде: $\gamma(t) = \big(x^1(t), \ldots, x^n(t)\big)$. Тогда по определению касательного вектора к кривой $\dfrac{dx^i}{dt}(t_0) = \xi^i$. Следовательно,

$$\xi(f) = \frac{d}{dt} f\big(\gamma(t)\big)\big|_{t=t_0} = \frac{d}{dt} f\big(x^1(t), \ldots, x^n(t)\big)\big|_{t=t_0} =$$
$$= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\frac{dx^i}{dt}(t_0) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)\xi^i. \qquad ■$$

Операция дифференцирования гладкой функции f по направлению касательного вектора ξ удовлетворяет следующим двум свойствам:

а) операция дифференцирования по направлению вектора ξ линейна, т.е. если f и g — две гладких функции, а λ, μ — два произвольных числа, то

$$\xi(\lambda f + \mu g) = \lambda\xi(f) + \mu\xi(g);$$

б) операция дифференцирования по направлению вектора ξ удовлетворяет формуле Ньютона–Лейбница: если f и g — две гладких функции в окрестности точки P_0, то для произведения функций f и g справедлива формула

$$\xi(fg) = f(x_0^1, \ldots, x_0^n)\xi(g) + \xi(f)g(x_0^1, \ldots, x_0^n). \tag{4.20}$$

Дадим общее определение.

Определение 4.10. Операция A, сопоставляющая каждой гладкой функции f класса C^∞ на гладком многообразии M число $A(f)$, удовлетворяющее свойствам а) и б), называется *операцией дифференцирования в точке* $P_0 \in M$.

Теорема 4.5. *Пусть M — гладкое многообразие класса C^∞, $P_0 \in M$ — произвольная точка, A — операция дифференцирования в смысле определения* 4.10. *Тогда операция A совпадает с дифференцированием вдоль некоторого (единственного) касательного вектора ξ в точке P_0.*

Доказательство. Касательный вектор будем искать в виде столбца его компонент в некоторой локальной системе координат $(x^1, \ldots, x^n)$ в окрестности точки P_0. Тогда всякую гладкую функцию будем представлять в виде функции от переменных $(x^1, \ldots, x^n)$. Докажем две вспомогательные леммы.

Лемма 4.3. *Всякую гладкую функцию $f(x^1, \ldots, x^n)$ класса C^∞ можно представить в виде*

$$f(x^1, \ldots, x^n) = f(x_0^1, \ldots, x_0^n) + \sum_{i=1}^n \frac{\partial f}{\partial x^i}(x_0^1, \ldots, x_0^n)(x^i - x_0^i) + \\ + \sum_{i,j=1}^n h_{ij}(x^1, \ldots, x^n)(x^i - x_0^i)(x^j - x_0^j), \tag{4.21}$$

где $h_{ij}(x^1, \ldots, x^n)$ — гладкие функции класса C^∞.

Доказательство. Запишем тождество

$$f(x^1, \ldots, x^n) \equiv f(x_0^1, \ldots, x_0^n) + \\ + \int_0^1 \frac{d}{dt} f\big(x_0^1 + t(x^1 - x_0^1), \ldots, x_0^n + t(x^n - x_0^n)\big)\, dt$$

и произведем дифференцирование по t под знаком интеграла:

$$f(x^1, \ldots, x^n) = f(x_0^1, \ldots, x_0^n) + \sum_{i=1}^n (x^i - x_0^i) h_i(x^1, \ldots, x^n), \tag{4.22}$$

где

$$h_i(x^1,\dots,x^n) \;=\; \int\limits_0^1 \frac{\partial f}{\partial x^i}\big(x_0^1 + t(x^1 - x_0^1),\dots,x_0^n + t(x^n - x_0^n)\big)\,dt. \quad (4.23)$$

Применим теперь формулу (4.22) к самим функциям $h_i(x^1,\dots,x^n)$. Получим

$$h_i(x^1,\dots,x^n) \;=\; h_i(x_0^1,\dots,x_0^n) \;+\; \sum_{j=1}^n (x^j - x_0^j)h_{ij}(x^1,\dots,x^n). \quad (4.24)$$

Подставим (4.24) в (4.22) и учтем (4.23):

$$f(x^1,\dots,x^n) = f(x_0^1,\dots,x_0^n) + \sum_{i=1}^n (x^i - x_0^i)\frac{\partial f}{\partial x^i}(x_0^1,\dots,x_0^n) + \\ + \sum_{i,j=1}^n (x^i - x_0^i)(x^j - x_0^j)h_{ij}(x^1,\dots,x^n). \qquad \blacksquare$$

Лемма 4.4. *Пусть f, g — две такие гладкие функции на многообразии M, что $f(P_0) = g(P_0) = 0$. Тогда для любого дифференцирования A в точке P_0 имеет место равенство $A(fg) = 0$.*

Доказательство следует немедленно из формулы Ньютона–Лейбница (4.20). ■

Перейдем теперь к доказательству теоремы 4.5. Представим функцию f в виде (4.21) и применим к левой и правой частям операцию дифференцирования A. В силу линейности операции получаем

$$A(f) = f(x_0^1,\dots,x_0^n)A(1) + \sum_{i=1}^n \frac{\partial f}{\partial x^i}(x_0^1,\dots,x_0^n)A(x^i - x_0^i) + \\ + \sum_{i,j=1}^n A\big((x^i - x_0^i)(x^j - x_0^j)h_{ij}(x^1,\dots,x^n)\big). \quad (4.25)$$

Заметим, что для постоянной функции, равной единице, $A(1) = A(1\cdot 1) = A(1)\cdot 1 + 1\cdot A(1) = 2A(1) = 0$. Далее, в последней сумме из (4.25) каждое слагаемое представимо в виде произведения двух функций, $(x^i - x_0^i)$ и $(x^j - x_0^j)h_{ij}(x^1,\dots,x^n)$, каждая из которых обращается в нуль в точке P_0. Поэтому в силу леммы 4.4 $A\big((x^i - x_0^i)(x^j - x_0^j)h_{ij}(x^1,\dots,x^n)\big) = 0$. Следовательно,

$$A(f) = \sum_{i=1}^n \frac{\partial f}{\partial x^i}(x_0^1,\dots,x_0^n)A(x^i - x_0^i).$$

Полагая $\xi^i = A(x^i - x_0^i)$, получаем такой вектор $\xi = (\xi^1,\dots,\xi^n)$, что

$$A(f) = \xi(f). \qquad \blacksquare$$

Примером дифференцирования гладких функций служит операция взятия частной производной в некоторой локальной системе координат $(x^1, \ldots, x^n)$. По теореме 4.5 операция $\dfrac{\partial}{\partial x^k}$ является касательным вектором, координаты которого равны $(0, \ldots, 1, \ldots, 0)$, где 1 стоит на месте с номером k. Поэтому касательные векторы $\left\{\dfrac{\partial}{\partial x^k}\right\}$ образуют базис в касательном пространстве $T_{P_0}(M)$, а всякий касательный вектор $\xi = (\xi^1, \ldots, \xi^n)$ разлагается в линейную комбинацию $\xi = \xi^1 \dfrac{\partial}{\partial x^1} + \cdots + \xi^n \dfrac{\partial}{\partial x^n}$.

4.4 Подмногообразия

4.4.1 Дифференциал гладкого отображения

Определение 4.11. Пусть $f\colon M_1 \to M_2$ гладкое отображение гладких многообразий. *Дифференциалом* df_{P_0} *гладкого отображения* f *в точке* $P_0 \in M_1$ называется линейное отображение касательного пространства $T_{P_0}(M_1)$ в касательное пространство $T_{Q_0}(M_2)$, $Q_0 = f(P_0)$, определяемое в локальных системах координат матрицей Якоби отображения f.

Определение 4.11 дифференциала df_{P_0}, не зависит от выбора локальных систем координат в окрестностях точек $P_0 \in M_1$ и $Q_0 \in M_2$. Чтобы показать это, достаточно переформулировать определение 4.11 в терминах дифференцирований гладких функций, в которых вообще не участвуют какие бы то ни было локальные координаты.

Лемма 4.5. *Пусть* $f\colon M_1 \to M_2$ *гладкое отображение,* $f(P_0) = Q_0$, $\xi \in T_{P_0}(M_1)$ *— касательный вектор к многообразию* M_1 *в точке* P_0, $\eta = df_{P_0}(\xi)$ *— касательный вектор к многообразию* M_2 *в точке* Q_0 *в смысле определения* 4.11. *Тогда для любой гладкой функции* g *на многообразии* M_2 *выполнено соотношение*

$$\eta(g) = \xi(g \circ f). \tag{4.26}$$

Проверку соотношения (4.26) достаточно производить в произвольной системе координат. Правая часть (4.26) не зависит от выбора локальной системы координат. Значит от этого выбора не зависит и правая часть.

Лемма 4.6. *Если* $\gamma = \gamma(t)$ *— кривая на многообразии* M_1, *проходящая через точку* P_0, $\gamma(t_0) = P_0$, *а* ξ *— касательный вектор к кривой* γ *в точке* P_0, *то образ* $\eta = df_{P_0}(\xi)$ *вектора* ξ *при отображении* $f\colon M_1 \to M_2$ *является касательным вектором к образу кривой* γ, $f(\gamma(t))$.

Доказательство непосредственно проверяется в некоторых локальных системах координат многообразий M_1 и M_2.

Пример 13. Рассмотрим гладкую функцию $y = f(x)$ как гладкое отображение многообразий $f\colon \mathbb{R}^1 \to \mathbb{R}^1$. Тогда согласно определению 4.11 дифференциал — это линейное отображение $T_x(\mathbb{R}^1) = \mathbb{R}^1$ в $T_{f(x)}(\mathbb{R}^1) = \mathbb{R}^1$,

задаваемое формулой $\eta = f'(x)\xi$. В математическом анализе под дифференциалом функции $f(x)$ понимают линейную часть приращения функции f как функцию от двух независимых аргументов (x и dx): $dy = f'(x)dx$. Так что, полагая $dx = \xi$, $dy = \eta$, получаем совпадение понятий.

Пример 14. Рассмотрим гладкую функцию f от n независимых переменных $y = f(x^1, \ldots, x^n)$. Так же, как и в случае функции одной переменной, представим ее как гладкое отображение многообразий $f\colon \mathbb{R}^n \to \mathbb{R}^1$. Тогда дифференциал отображения f является линейным отображением касательных пространств

$$df_{P_0}\colon T_{P_0}(\mathbb{R}^n) = \mathbb{R}^n \to T_{Q_0}(\mathbb{R}^1) = \mathbb{R}^1,$$
$$P_0 = (x_0^1, \ldots, x_0^n), \quad Q_0 = f(x_0^1, \ldots, x_0^n).$$

Дифференциал df_{P_0} в локальных координатах задается формулой $\eta = \sum\limits_{k=1}^{n} \dfrac{\partial f}{\partial x^k}(P_0)\xi^k$. С другой стороны, дифференциал функции f есть линейная часть приращения функции $(x^1, \ldots, x^n)$ как функция от двух групп независимых переменных, $(x^1, \ldots, x^n)$ и $(dx^1, \ldots, dx^n)$:

$$df(x^1, \ldots, x^n) = \sum_{k=1}^{n} \frac{\partial f}{\partial x^k}(x^1, \ldots, x^n)\, dx^k.$$

Полагая $\xi^k = dx^k$, $\eta = dy$, получаем совпадение понятий дифференциала функции и дифференциала ее же как отображения многообразий. Более того, матрица дифференциала df_{P_0} есть матрица Якоби отображения f, а с другой стороны — градиент $\operatorname{grad} f$ функции f; таким образом, градиент f — это матрица дифференциала df_{P_0} в фиксированной системе координат. Ясно, что при изменении системы координат будут меняться и компоненты градиента функции f.

Пример 15. Рассмотрим гладкую функцию $f = f(x^1, \ldots, x^n)$, пусть $P_0 = (x_0^1, \ldots, x_0^n)$ — экстремальная точка функции f. Тогда одна из теорем анализа гласит, что $\operatorname{grad}_{P_0} f = 0$. В нашей терминологии это значит, что $df_{P_0} = 0$. Это утверждение обобщается на случай произвольных многообразий: *если в точке $P_0 \in M$ гладкая функция f достигает локального максимума, то $df_{P_0} = 0$*

Это обстоятельство получает новое освещение в теории многообразий. В случае областей в евклидовом пространстве всегда существуют такие гладкие функции, для которых $\operatorname{grad} f \neq 0$ в каждой точке. Для гладкого многообразия это уже не так. Например, на двумерной сфере S^2 для всякой гладкой функции f ее дифференциал df обращается в нуль по крайней мере в двух точках: в точке максимума функции f и в точке ее минимума. Вообще, *если M — компактное гладкое многообразие, то для всякой гладкой функции f дифференциал df обращается в нуль не менее чем в двух точках.* На рис. 4.9 изображено поведение дифференциала функции "высоты" на двумерной сфере S^2, а на рис. 4.10 — функция высоты на торе T^2.

Рис. 4.9:

Рис. 4.10:

Пример 16. Пусть $f\colon \mathbb{R}^n \to \mathbb{R}^m$ — линейное отображение, которое в стандартных координатах записывается в матричном виде:

$$Y = AX,$$

$$Y = \begin{pmatrix} y^1 \\ \vdots \\ y^m \end{pmatrix}, \quad X = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}.$$

Тогда, очевидно, матрица Якоби отображения f (а, значит, и дифференциал df_{P_0}) совпадает с матрицей A. Другими словами, $df_{P_0}(\xi) = f(\xi)$. Впрочем, это не удивительно. Если представить дифференциал df_{P_0}, как и в случае функций многих переменных в виде линейной части приращения отображения, то получим $\Delta Y = A(X + \Delta X) - AX = A\Delta X$. Следовательно, само приращение ΔY тождественно совпадает с его линейной частью. Таким образом, дифференциал линейного отображения не зависит от точки $P_0 \in \mathbb{R}^n$.

4.4.2 Локальные свойства отображений и дифференциал

Определение 4.12. Пусть $f\colon M_1 \to M_2$ — гладкое отображение. Точка $P_0 \in M_1$ называется *регулярной точкой отображения* f, если дифференциал

отображения $df_{P_0}\colon T_{P_0}(M_1) \to T_{Q_0}(M_2)$, $Q_0 = f(P_0)$, является эпиморфизмом, т.е. отображением на все пространство $T_{Q_0}(M_2)$. Точка $Q_0 \in M_2$ называется *регулярной точкой отображения* f, если всякая точка P_0 из прообраза $f^{-1}(Q_0)$ является регулярной точкой отображения f.

Следующая теорема является обобщением теоремы 4.3:

Теорема 4.6. *Пусть $f\colon M_1 \to M_2$ — гладкое отображение гладких многообразий, $Q_0 \in M_2$ — регулярная точка отображения f. Тогда прообраз $M_3 = f^{-1}(Q_0)$ является гладким многообразием,* $\dim M_3 = \dim M_1 - \dim M_2$. *Более того, в качестве локальных координат в многообразии M_3 можно взять некоторые из локальных координат многообразия $P_0 \in M_3$.*

Доказательство. Для того чтобы доказать, что M_3 является многообразием, достаточно в окрестности каждой точки $P_0 \in M_3$ применить теорему 4.3. Получаем, что каждая точка $P_0 \in M_3$ допускает окрестность $U \ni P_0$, гомеоморфную области в евклидовом пространстве $\mathbb{R}^{n-m}$, где $n = \dim M_1$, $m = \dim M_2$. Более того, в качестве локальных координат в окрестности U можно взять некоторые $(n-m)$ локальных координат $(x^1, \ldots, x^n)$ многообразия M_1 в окрестности точки P_0. Если это координаты $(x^{i_1}, \ldots, x^{i_{n-m}})$, то остальные локальные координаты (x^j) выражаются на M_3 через $(x^{i_1}, \ldots, x^{i_{n-m}})$ в виде гладких функций. Отсюда следует, что M_3 — гладкое многообразие. В самом деле, пусть $(y^1, \ldots, y^n)$ — другая система координат на многообразии M_1, а $(y^{i_1}, \ldots, y^{i_{n-m}})$ — образуют локальную систему координат на M_3. Тогда

$$y^{i_k} = y^{i_k}(x^1, \ldots, x^n) = y^{i_k}\big(x^1(x^{i_1}, \ldots, x^{i_{n-m}}), \ldots, x^n(x^{i_1}, \ldots, x^{i_{n-m}})\big)$$

— гладкая функция. Теорема доказана. ■

Определение 4.13. Пусть $f\colon M_1 \to M_2$ гладкое отображение. Отображение f называется *погружением*, если в каждой точке $P \in M_1$ дифференциал df_P является мономорфизмом, т.е. взаимно однозначным отображением на свой образ.

Если, кроме того, отображение f взаимно однозначно отображает многообразие M_1 на свой образ $f(M_1)$ и $f(M_1)$ — замкнутое множество, то отображение f называется *вложением*. Образ $f(M_1)$ (равно как и M_1) в этом случае называется *подмногообразием* многообразия M_2.

Предложение 4.3. *Гладкая кривая с ненулевым вектором скорости является гладким одномерным многообразием, гладко вложенным в $\mathbb{R}^n$.*

Доказательство. То, что кривая является гладким многообразием, непосредственно следует из определения гладкого многообразия. Остается проверить, что она является гладким подмногообразием в $\mathbb{R}^n$, для чего нужно изучить дифференциал отображения вложения i, т.е. линейное отображение di. Поскольку оно полностью определяется вектором скорости кривой, то матрица Якоби имеет максимальный ранг в каждой точке кривой, а потому кривая является подмногообразием. Предложение доказано. ■

Рис. 4.11:

Ясно, что любое одномерное гладкое замкнутое (т.е. не имеющее края) многообразие диффеоморфно либо прямой (некомпактное многообразие), либо окружности (компактное многообразие). Таким образом, все одномерные многообразия исчерпываются только двумя различными многообразиями. Они уже не диффеоморфны, так как прямая некомпактна, а окружность компактна.

Пример 17. Прообраз регулярной точки отображения $f\colon M_1 \to M_2$ согласно теореме 4.6 является подмногообразием. В самом деле, так как в качестве локальной системы координат в M_3 можно взять некоторые из локальных координат объемлющего многообразия M_1, то в локальных координатах тождественное отображение $\varphi\colon M_3 \to M_1$ имеет вид

$$\begin{gathered} x^1 = x^1, \\ \cdots\cdots \\ x^{n-m} = x^{n-m}, \\ x^{n-m+1} = x^{n-m+1}(x^1, \ldots, x^{n-m}), \\ \cdots\cdots\cdots\cdots \\ x^n = x^n(x^1, \ldots, x^{n-m}). \end{gathered}$$

Поэтому матрица Якоби $d\varphi$ содержит в себе единичную квадратную матрицу. Следовательно, $\operatorname{rank} d\varphi = n - m$, т.е. $d\varphi$ является мономорфизмом.

Пример 18. Рассмотрим отображение $f\colon S^1 \to \mathbb{R}^2$, заданное формулами $f(\varphi) = (\cos\varphi, \sin 2\varphi)$. Вектор скорости равен $\dfrac{df}{d\varphi} = (-\sin\varphi, 2\cos 2\varphi)$ и ни в какой точке не обращается в нуль, т.е. ранг матрицы Якоби равен единице. Поэтому f является погружением (рис. 4.11). Кривая, изображенная на рис. 4.11, является одной из фигур Лиссажу, и может быть получена на осциллографе при синусоидальных сигналах на вертикальной и горизонтальной развертках.

4.4.3 Вложение многообразий в евклидово пространство

Мы завершим настоящий параграф теоремой Уитни, показывающей, что всякое компактное многообразие можно считать подмногообразием евклидова пространства достаточно большой размерности.

Теорема 4.7 (Слабая теорема Уитни). *Пусть M — гладкое компактное многообразие. Тогда существует вложение $\varphi\colon M \to \mathbb{R}^N$ для подходящего выбора размерности N.*

Доказательство. Пусть $\{U_\alpha\}_{\alpha=1}^M$ — конечный атлас карт, $(x_\alpha^1, \dots, x_\alpha^n)$ — локальная система координат в карте U_α. Без ограничения общности можно считать, что карты U_α гомеоморфны шару $D^n \subset \mathbb{R}^n$ радиуса 1, причем координаты $(x_\alpha^1, \dots, x_\alpha^n)$ осуществляют гомеоморфизм φ_α карты U_α на шар D^n. Далее, можно считать, что шар D^n лежит в $\mathbb{R}^n$ и не содержит начала координат. Этого можно достигнуть с помощью параллельного переноса в $\mathbb{R}^n$. Пусть, далее, $D_1^n \subset D^n$ — шар меньшего радиуса с тем же центром, что и шар D^n, причем открытые множества $U'_\alpha = \varphi_\alpha^{-1}(D_1^n) \subset U_\alpha$ покрывают многообразие M. Пусть f — гладкая функция в $\mathbb{R}^n$, равная тождественно единице на D_1^n, и $\operatorname{supp} f \subset D^n$. Положим тогда:

$$y_\alpha^k(P) = \begin{cases} f\big(\varphi_\alpha(P)\big) x_\alpha^k(P), & \text{если } P \in U_\alpha, \\ 0, & \text{если } P \notin U_\alpha. \end{cases}$$

Ясно, что $y_\alpha^k(P) = x_\alpha^k(P)$, если $P \in U'_\alpha$. Мы получили систему гладких функций $\{y_\alpha(P)\}$ в количестве $N = nM$. Эта система функций совместно осуществляет отображение g многообразия M в евклидово пространство $\mathbb{R}^N$: $g(P) = \big(y_\alpha^k(P)\big) \in \mathbb{R}^N$.

Покажем, что ранг дифференциала отображения g в каждой точке равен n. Пусть $P \in M$ — произвольная точка, $U'_\alpha \ni P$, $(x_\alpha^1, \dots, x_\alpha^n)$ — локальная система координат. Матрица Якоби отображения g в точке P в локальной системе координат $(x_\alpha^1, \dots, x_\alpha^n)$ состоит из частных производных $\big(\partial y_\beta^k / \partial x_\alpha^j(P)\big)$. В частности, если $\beta = \alpha$, то

$$\frac{\partial y_\beta^k}{\partial x_\alpha^j}(P) = \frac{\partial x_\alpha^k}{\partial x_\alpha^j}(P) = \delta_\beta^k,$$

т.е. матрица Якоби отображения g содержит единичную матрицу порядка n. Следовательно, $\operatorname{rank} dg = n$.

Таким образом, отображение g является погружением. Для того чтобы отображение g было вложением, необходимо, чтобы различные точки P и Q переходили в различные точки $g(P)$ и $g(Q)$.

Построим новое отображение

$$\bar{g}(P) = \big(y_\alpha^k(P), f(\varphi_\alpha(P))\big) \in \mathbb{R}^{N+M}.$$

Оно по тем же причинам, что и g, является погружением. Пусть теперь P и Q, $P \neq Q$, — две точки на многообразии. Рассмотрим такой номер α, что

$f(\varphi_\alpha(P)) = 1$. Если $f(\varphi_\alpha(Q)) < 1$, то $\bar{g}(P) \neq \bar{g}(Q)$; если $f(\varphi_\alpha(Q)) = 1$, то $y^k_\alpha(P) = x^k_\alpha(P)$, $y^k_\alpha(Q) = x^k_\alpha(Q)$, и, значит, для некоторого номера k координат $(x^1_\alpha, \ldots, x^n_\alpha)$ имеем: $x^k_\alpha(P) \neq x^k_\alpha(Q)$, т.е. $\bar{g}(P) \neq \bar{g}(Q)$. Таким образом, отображение $\bar{g}\colon M \to \mathbb{R}^{N+M}$ является взаимно однозначным погружением, т.е. вложением. Теорема доказана. ■

Итак, всякое гладкое компактное многообразие M можно считать вложенным в виде подмногообразия в евклидово пространство $\mathbb{R}^N$ некоторой достаточно большой размерности N. На самом деле размерность евклидова пространства $\mathbb{R}^N$ можно существенно понизить. Так, например, сфера S^n вкладывается в $\mathbb{R}^{n+1}$, а тор T^n вкладывается в $\mathbb{R}^{2n}$. Проективная плоскость $\mathbb{R}P^2$ не вкладывается в $\mathbb{R}^3$, однако может быть вложена в $\mathbb{R}^5$. В самом деле, пусть $(x_1 : x_2 : x_3)$ — однородные координаты точки P в $\mathbb{R}P^2$. Положим

$$y^1 = \frac{x_1^2}{x_1^2 + x_2^2 + x_3^2}, \quad y^2 = \frac{x_2^2}{x_1^2 + x_2^2 + x_3^2}, \quad y^3 = \frac{x_3^2}{x_1^2 + x_2^2 + x_3^2},$$
$$y^4 = \frac{x_1 x_2}{x_1^2 + x_2^2 + x_3^2}, \quad y^5 = \frac{x_2 x_3}{x_1^2 + x_2^2 + x_3^2}, \quad y^6 = \frac{x_3 x_1}{x_1^2 + x_2^2 + x_3^2}.$$

Получаем отображение

$$g\colon \mathbb{R}P^2 \to \mathbb{R}^6, \qquad g(P) = g(x_1 : x_2 : x_3) = (y^1, y^2, y^3, y^4, y^5, y^6).$$

На самом деле образ отображения g лежит в линейном подпространстве $\mathbb{R}^5 \subset \mathbb{R}^6$, задаваемом уравнением $y^1 + y^2 + y^3 = 1$.

4.4.4 Римановы метрики на многообразии

Теперь мы можем дать общее понятие римановой метрики на многообразии, частные случаи которого рассматривались выше.

Римановой метрикой на гладком многообразии M назовем семейство положительно определенных скалярных произведений, заданных в каждом касательном пространстве $T_P(M)$. Если в окрестности U точки P фиксирована система координат $(x^1, \ldots, x^n)$, то и в $T_P(M)$ задается система координат $(\xi^1, \ldots, \xi^n)$. Скалярное произведение в системе $(\xi^1, \ldots, \xi^n)$ определяется невырожденной симметрической матрицей $G = (g_{ij})$, зависящей от точки. При переходе к новой системе координат $(y^1, \ldots, y^n)$ матрица меняется по закону:

$$g'_{ij}(P) = \sum_{\alpha,\beta=1}^{n} g_{\alpha\beta}(P) \frac{\partial x^\alpha}{\partial y^i}(P) \frac{\partial x^\beta}{\partial y^j}(P),$$

где $G' = (g'_{ij}(P))$ — матрица скалярного произведения в новой системе координат $(y^1, \ldots, y^n)$.

Определение 4.14. *Римановой метрикой* на гладком многообразии M называется такое семейство невырожденных положительно определенных скалярных произведений в каждом касательном пространстве $T_P(M)$, $P \in M$, что в локальной системе координат матрица скалярного произведения является гладкой функцией от локальных координат.

Метрику можно задавать как соответствие, сопоставляющее каждой системе координат $(x_\alpha^1, \dots, x_\alpha^n)$ в карте U_α матричнозначную гладкую функцию в карте U_α: $G^\alpha(P) = \big(g_{ij}^\alpha(P)\big)$, такую, что:

а) матрица $G^\alpha(P)$ симметрична и положительно определена;

б) коэффициенты $g_{ij}^\alpha(P)$ преобразуются по закону

$$g_{ij}^\alpha(P) = \sum_{k,l=1}^{n} g_{kl}^\beta(P) \frac{\partial x_\beta^k}{\partial x_\alpha^i}(P) \frac{\partial x_\beta^l}{\partial x_\alpha^j}(P) \tag{4.27}$$

для точек $P \in U_\alpha \cap U_\beta$. Если $\xi = (\xi_\alpha^1, \dots, \xi_\alpha^n)$, $\eta = (\eta_\alpha^1, \dots, \eta_\alpha^n)$ — два касательных вектора, то скалярное произведение ξ, η записывается так:

$$\langle \xi, \eta \rangle = \sum_{i,j=1}^{n} g_{ij}^\alpha(P) \xi_\alpha^i \eta_\alpha^j.$$

Ниже мы приведем более общую конструкцию римановой метрики на многообразии.

Пусть $f\colon M_1 \to M_2$ — гладкое вложение многообразия M_1 в многообразие M_2, $\dim M_1 = m$, $\dim M_2 = n$. На многообразии M_2 определена риманова метрика, которая задана набором матричнозначных функций $\big\{G^\alpha(P) = \big(G_{ij}^\alpha(P)\big)\big\}$ на атласе карт $\{U_\alpha\}$.

Тогда вложение f задает на подмногообразии M_1 риманову метрику, порожденную римановой метрикой в объемлющем многообразии M_2. В самом деле, без ограничения общности можно считать, что атлас карт $\{V_\alpha\}$ на подмногообразии M_1 задается пересечениями

$$V_\alpha = M_1 \cap U_\alpha$$

и некоторыми локальными координатами $\{y_\alpha^i\}_{1 \leqslant i \leqslant m}$ в карте V_α. Вложение f в картах V_α и U_α задается набором формул

$$x_\alpha^i = x_\alpha^i(y_\alpha^1, \dots, y_\alpha^m).$$

Матрица Якоби этого отображения имеет максимальный ранг

$$\operatorname{rank}\Big(\frac{\partial x_\alpha^i}{\partial y_\alpha^j}\Big) = m = \dim M_1.$$

Метрика на многообразии M_1 задается формулами

$$H^\alpha(P) = \big(h_{ij}^\alpha(P)\big)_{1 \leqslant i,j \leqslant m}, \quad P \in V_\alpha,$$

где

$$h_{ij}^\alpha(P) = g_{kl}^\alpha(P) \frac{\partial x_\alpha^k}{\partial y_\alpha^i} \frac{\partial x_\alpha^l}{\partial y_\alpha^j} \tag{4.28}$$

Предложение 4.4. *Формулы* (4.28) *задают риманову метрику на многообразии* M_1.

Доказательство. Симметричность и положительная определенность матрицы $H^\alpha(P)$ следует из алгебраических свойств формул (4.28). Осталось проверить только выполнение закона преобразования (4.27). В силу формул (4.28) имеем

$$h^\alpha_{ij}(P) = g^\alpha_{kl}(P)\frac{\partial x^k_\alpha}{\partial y^i_\alpha}\frac{\partial x^l_\alpha}{\partial y^j_\alpha} = g^\beta_{st}(P)\frac{\partial x^s_\beta}{\partial x^k_\alpha}\frac{\partial x^t_\beta}{\partial x^l_\alpha}\frac{\partial x^k_\alpha}{\partial y^i_\alpha}\frac{\partial x^l_\alpha}{\partial y^j_\alpha} =$$
$$= g^\beta_{st}(P)\frac{\partial x^s_\beta}{\partial y^i_\alpha}\frac{\partial x^t_\beta}{\partial y^j_\alpha} = g^\beta_{st}(P)\frac{\partial x^s_\beta}{\partial y^p_\beta}\frac{\partial y^p_\beta}{\partial y^i_\alpha}\frac{\partial x^t_\beta}{\partial y^q_\beta}\frac{\partial y^q_\beta}{\partial y^j_\alpha} = h^\beta_{pq}(P)\frac{\partial y^p_\beta}{\partial y^i_\alpha}\frac{\partial y^q_\beta}{\partial y^j_\alpha}.$$

Тем самым выполнен закон преобразования (4.27). Предложение доказано. ■

Следствие 4.1. *На любом компактном многообразии существует риманова метрика.*

Доказательство. Напомним, что в силу предложения 4.4 риманова метрика определена на подмногообразии M_1 если она она определена на объемлющем многообразии M_2.

В силу теоремы Уитни, задача сводится к построению римановой метрики на евклидовом пространстве $\mathbb{R}^N$. Так как многообразие $\mathbb{R}^N$ покрывается одной картой, то необходимо предъявить лишь одну функцию $G(P)$, $P \in \mathbb{R}^N$. В качестве этой функции можно выбрать $G(P) \equiv E$, где E — единичная матрица. Тем самым, на $\mathbb{R}^N$ имеется хотя бы одна риманова метрика, а, значит, на любом компактном многообразии также существует риманова метрика. Следствие доказано. ■

Задачи

1. Построить вложение тора T^n в $\mathbb{R}^{n+1}$.

Указание. Представить тор T^n как гиперповерхность вращения T^{n-1} вокруг оси.

2. Построить вложение $S^2 \times T^2$ в $\mathbb{R}^5$.

Указание. Найти в $\mathbb{R}^5$ открытую область, гомеоморфную $S^2 \times \mathbb{R}^3$.

4.5 Классификация двумерных поверхностей

Мы начнем с общих определений, затем приведем примеры, а потом докажем теорему классификации поверхностей.

4.5.1 Многообразия с краем

Расширим понятие гладкого многообразия, включив в него подмножества в $\mathbb{R}^n$, задаваемые системами уравнений и неравенств. Обозначим через $\mathbb{R}^n_+ \subset \mathbb{R}^n$ полупространство, задаваемое так: $x^n \geqslant 0$. Границу $\mathbb{R}^n_+$, задаваемую $x^n = 0$, обозначим через $\mathbb{R}^{n-1}_0$. Если $f\colon \mathbb{R}^n_+ \to \mathbb{R}^1$ — непрерывная функция, то во всех точках, включая и граничные, определим дифференцируемость функции f. Если точка $x_0 \in \mathbb{R}^n_+$ внутренняя, т.е. $x^n > 0$, то определение дифференцируемости обычное. Если же $x_0 \in \mathbb{R}^{n-1}_0$, т.е. $x^n = 0$, то скажем, что f *дифференцируема в точке* x_0, если справедливо разложение

$$f(x) = f(x_0) + \sum_{i=1}^{n} a_i(x^i - x_0^i) + o(x - x_0),$$

где $\lim\limits_{x\to x_0}\bigl(o(x - x_0)/|x - x_0|\bigr) = 0$, $x^n \geqslant 0$. Тогда $a_i = \partial f/\partial x^i(x_0)$ при $i = 1, 2, \ldots, n-1$ и $a_n = \lim\limits_{h\to+0}(1/h)\bigl(f(x_0^1, \ldots, x_0^{n-1}, x_0^n + h) - f(x_0^1, \ldots, x_0^n)\bigr)$. Последний предел уместно назвать *частной производной* $\dfrac{\partial f}{\partial x^n}$ функции f в точке x_0. Аналогично определяется класс функций гладкости C^r.

Определение 4.15. Метрическое пространство M называется *гладким многообразием с краем*, если существует такой атлас карт $\{U_\alpha\}$ и координатные гомеоморфизмы $\varphi_\alpha\colon U_\alpha \to V_\alpha \subset \mathbb{R}^n_+$, где V_α — открытое множество в $\mathbb{R}^n_+$, что функции замены координат

$$\varphi_\beta\varphi_\alpha^{-1}\colon V_{\alpha\beta} = \varphi_\alpha(U_\alpha \cap U_\beta) \to V_{\beta\alpha} = \varphi_\beta(U_\alpha \cap U_\beta)$$

— гладкие функции.

Линейные координаты $(x^1, \ldots, x^n)$ в $\mathbb{R}^n_+$ индуцируют локальные координаты в карте U_α : $x_\alpha^k(P) = x^k\bigl(\varphi_\alpha(P)\bigr)$. Тогда в любой локальной системе координат имеем $x_\alpha^n(P) \geqslant 0$. Точка $P \in M$ называется *внутренней* точкой, если $x_\alpha^n(P) > 0$. Точка P называется *граничной*, если $x_\alpha^n(P) = 0$.

Итак, многообразие с краем характеризуется тем, что существует такой атлас карт $\{U_\alpha\}$ с локальными координатами $(x_\alpha^1, \ldots, x_\alpha^n)$, что для любой карты выполнено строгое неравенство $x_\alpha^n > 0$ во внутренних точках и равенство $x_\alpha^n = 0$ в граничных точках. Если карта U_α не содержит граничных точек, то условие $x_\alpha^n > 0$ можно без ограничения общности отбросить. Множество ∂M граничных точек — гладкое многообразие на единицу меньшей размерности. В самом деле, в качестве атласа карт следует взять пересечение $W_\alpha = \partial M \cap U_\alpha$, а в качестве координат — первые $(n-1)$ координат $x_\alpha^1, \ldots, x_\alpha^{n-1}$ карты U_α. Координатные гомеоморфизмы φ_α (см. выше) отображают W_α гомеоморфно на $V_\alpha \cap \mathbb{R}^{n-1}_0$, а функции замены остаются гладкими как ограничения функций замены координат многообразия M.

Если $\partial M = \emptyset$, то приходим к прежнему понятию многообразия, которое отныне будем называть *замкнутым многообразием*.

Пример 19. Рассмотрим шар D^n, задаваемый неравенством $\sum\limits_{i=1}^{n}(x^i)^2 \leqslant 1$ в $\mathbb{R}^n$. Это — многообразие с краем S^{n-1}.

Пример 20. Пусть $f\colon \mathbb{R}^n \to \mathbb{R}^k$ — вектор-функция, удовлетворяющая условиям теоремы о неявных функциях, т.е. во всех точках $P \in M = \{f(P) = 0\}$ дифференциал df имеет максимально возможный ранг. Рассмотрим систему

$$\begin{cases} f^1(x^1, \ldots, x^n) = 0, \\ \ldots\ldots\ldots\ldots\ldots \\ f^{k-1}(x^1, \ldots, x^n) = 0, \\ f^k(x^1, \ldots, x^n) \geqslant 0, \end{cases}$$

и пусть M' — множество решений этой системы.

Если в каждой точке $P \in M'$ ранг $\|\partial f^i/\partial x^s\|$, $1 \leqslant i \leqslant k-1$; $1 \leqslant s \leqslant n$, равен $(k-1)$, то M' — многообразие с краем $\partial M' = M$. В самом деле, в точках $P \in M' \setminus M$ существует атлас карт по теореме о неявной функции для системы $f^1 = 0, \ldots, f^{k-1} = 0$. Рассмотрим $P \in M$. Пусть N' — множество решений системы $f^1 = 0, \ldots, f^{k-1} = 0$ в некоторой окрестности точки P. Тогда f^k можно рассматривать как функцию на N'. Ясно, что дифференциал f^k на N' отличен от нуля. Следовательно, по теореме 4.2 N' диффеоморфно прямому произведению интервала $(-\varepsilon, \varepsilon)$ на многообразие N, являющееся пересечением M с окрестностью точки P : $N' = N \times (-\varepsilon, \varepsilon)$. Поэтому неравенство $f^n \geqslant 0$ выделяет карту вида $N \times [0, \varepsilon)$, в которой в качестве координат можно взять набор $(x^1, \ldots, x^{n-k})$ на многообразии N и функцию $f^k \geqslant 0$.

Пример 21. В предыдущем примере видно, что декартово произведение многообразия M с краем ∂M на замкнутое многообразие N является многообразием $M \times N$ с краем $\partial M \times N$.

4.5.2 Ориентируемые многообразия

Определение 4.16. Многообразие (с краем) M называется *ориентируемым*, если существует такой атлас карт $\{U_\alpha\}$, что якобиан функций перехода от одной локальной системы координат к другой положителен. Будем говорить также, что атлас карт $\{U_\alpha\}$, у которого все якобианы функций перехода положительны, задает на многообразии M ориентацию, а сам атлас карт назовем *ориентированным атласом карт*. Два атласа карт, $\{U_\alpha\}$ и $\{U'_\beta\}$, задают одинаковую ориентацию, если объединение атласов $\{U_\alpha\} \cup \{U'_\beta\}$ снова является ориентированным атласом карт.

Всякий атлас карт на ориентируемом многообразии можно превратить в ориентированный атлас карт путем изменения локальных координат в каждой карте (докажите!).

Предложение 4.5. *На связном ориентированном многообразии существуют ровно две различные ориентации, причем любая карта задает локальную ориентацию, совпадающую с одной из ориентаций M.*

Доказательство предоставляется читателю.

Существует другое представление об ориентации M. Скажем, что базис $(\mathbf{e}_1, \ldots, \mathbf{e}_n) \in \mathbb{R}^n$ задает ориентацию евклидова пространства. Два базиса считаются *одинаково ориентированными*, если детерминант матрицы перехода положителен. Тогда $\mathbb{R}^n$ приобретает ровно две различных ориентации. Задать ориентацию M — значит задать ориентацию касательного пространства $T_P M$ в каждой точке $P \in M$. Пусть $\varphi\colon [0,1] \to M$ — непрерывный путь. Для любого базиса в начальной точке $(\mathbf{e}_1, \ldots, \mathbf{e}_n) \in T_{\varphi(0)}M$ построим непрерывное семейство базисов $\big(\mathbf{e}_1(t), \ldots, \mathbf{e}_n(t)\big) \in T_{\varphi(t)}M$, начинающееся с базиса $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. Любые два таких семейства задают в $T_{\varphi(t)}M$ одинаковую ориентацию. Если M ориентируемо, то ориентация в конечной точке $\varphi(1)$ пути φ не зависит от выбора пути. Таким образом, задать ориентацию на M — значит задать ориентацию касательных пространств $T_P M$ так, чтобы они были согласованы по любому непрерывному пути, соединяющему пары точек. (Докажите эквивалентность двух определений ориентируемости).

Теорема 4.8. *Если многообразие с краем M ориентируемо, то край ∂M тоже является ориентируемым многообразием.*

Доказательство. Пусть $\{U_\alpha, (x_\alpha^1, \ldots, x_\alpha^n)\}$ — атлас на M. Последняя координата x_α^n в любой карте является неотрицательной функцией. Поскольку M ориентируемо, то, меняя в каждой карте U_α координату x_α^1 на $-x_\alpha^1$, если это необходимо, можно считать, что якобиан замены переменных положителен в каждом пересечении карт $U_{\alpha\beta} = U_\alpha \cap U_\beta$. В качестве карт на ∂M, как и раньше, возьмем пересечения $W_\alpha = U_\alpha \cap \partial M$, а в качестве локальных координат — первые $(n-1)$ координат, т.е. $(y_\alpha^1, \ldots, y_\alpha^{n-1}) = (x_\alpha^1, \ldots, x_\alpha^{n-1})$. Покажем, что $\det \|\partial y_\alpha^i / \partial y_\beta^j\| > 0$ в произвольной точке $P \in W_\alpha \cap W_\beta$.

Рассмотрим в объемлющем многообразии M матрицу Якоби $\dfrac{\partial(x_\alpha)}{\partial(x_\beta)}$ замены координат в точке P.

Поскольку на пересечении $W_\alpha \cap W_\beta$ имеем: $x_\alpha^n = x_\beta^n = 0$, то $\dfrac{\partial x_\alpha^n}{\partial x_\beta^i} \equiv 0$, $1 \leqslant i \leqslant n-1$. Поэтому $\det \dfrac{\partial(x_\alpha)}{\partial(x_\beta)} = \det \dfrac{\partial(y_\alpha)}{\partial(y_\beta)} \cdot \dfrac{\partial x_\alpha^n}{\partial x_\beta^n} > 0$. Следовательно, $\dfrac{\partial x_\alpha^n}{\partial x_\beta^n} \neq 0$. Так как $x_\alpha^n \geqslant 0$ в пересечении $U_\alpha \cap U_\beta$, то $\dfrac{\partial x_\alpha^n}{\partial x_\beta^n}(P) > 0$. Итак, $\det \dfrac{\partial y_\alpha}{\partial y_\beta}(P) > 0$. Теорема доказана. ■

Определение 4.17. Пусть M — ориентированное многообразие с краем, $\{U_\alpha, (x_\alpha^1, \ldots, x_\alpha^n)\}$, $x_\alpha^n \geqslant 0$, — атлас, задающий ориентацию на M. Ориентация на ∂M, задаваемая атласом $W_\alpha = U_\alpha \cap \partial M$, $(y_\alpha^1, \ldots, y_\alpha^{n-1}) = (x_\alpha^1, \ldots, x_\alpha^{n-1})$, называется *ориентацией, согласованной с ориентацией на M*.

Доказательство. Рассмотрим диск $D^2 \subset \mathbb{R}^2$. Для задания ориентации на диске достаточно в одной точке фиксировать базис. Если точку P выбрать

Рис. 4.12:

на границе, то базис можно выбрать так, чтобы первый вектор $\mathbf{e}_1$ касался края, а второй $\mathbf{e}_2$ указывал направление внутрь диска. Тогда ориентация края, определяемая $\mathbf{e}_1$, согласована с ориентацией D^2.

Пример 22. Рассмотрим лист Мебиуса, представленный как квадрат с отождествленными противоположными сторонами в разных направлениях (рис. 4.12). Фиксировав базис ($\mathbf{e}_1$, $\mathbf{e}_2$) в точке A, построим непрерывное семейство базисов вдоль кривой ABA. В конце кривой базис изменит ориентацию, т.е. лист Мебиуса неориентируем. Однако, край — ориентируемая окружность.

4.5.3 Классификация двумерных гладких компактных связных многообразий без края

Переходим к описанию двумерных многообразий. Простейшее из них — сфера S^2, диффеоморфна $\mathbb{C}P^1$. Мы знакомы с тором $T^2 = S^1 \times S^1$, допускающим также следующее представление. Пусть $\mathrm{G} = \mathbb{Z}(a) \oplus \mathbb{Z}(b)$ — абелева группа, образующие которой a и b представлены сдвигами на $\mathbb{R}^2$: $a(x, y) = (x, y+1)$; $b(x, y) = (x+1, y)$. Рассмотрев факторпространство $\mathbb{R}^2/\mathbb{Z} \oplus \mathbb{Z}$, получаем T^2. Квадрат $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$ называется *фундаментальной областью.* Ее роль в том, что, применяя к ней элементы G, мы можем "замостить" всю плоскость (рис. 4.13).

Рассмотрим на $\mathbb{R}^2(x, y)$ другую группу $\mathrm{G}(a, b)$, образующие которой задаются на $\mathbb{R}^2$ так: $b(x, y) = (x+1, y)$; $a(x, y) = (1-x, y+1)$. (Найдите соотношения между a и b.) Фундаментальная область показана на рис. 4.14. Стрелками показаны отождествления сторон в соответствии с действием группы. Поверхность, возникающая после факторизации, показана на рис. 4.14 и называется *бутылкой Клейна.* Тор и бутылка Клейна — гладкие многообразия; тор — ориентируемое, а бутылка Клейна — неориентируемое. (Проверьте!)

Рассмотрим склейку на границе квадрата, показанную на рис. 4.15. Так как квадрат гомеоморфен диску, то мы получили проективную плоскость. Попробуем изобразить ее в $\mathbb{R}^3$, выполнив требуемые склейки. На рис. 4.16

Рис. 4.13:

Рис. 4.14:

Рис. 4.15:

Рис. 4.16:

показана склейка, приводящая к модели $\mathbb{R}P^2$ в $\mathbb{R}^3$. Она не является погруженным подмногообразием, поскольку есть две особые точки A и B. Тем не менее, погружение $\mathbb{R}P^2$ в $\mathbb{R}^3$ все же существует. Об этом ниже.

Лемма 4.7. *$\mathbb{R}P^2$ гомеоморфно склейке по общей границе диска и листа Мебиуса.*

Доказательство на рис. 4.17. ■

Эквивалентная формулировка: если из $\mathbb{R}P^2$ удалить диск, то останется лист Мебиуса.

Рассмотрим T^2 и $\mathbb{R}P^2$ с иной точки зрения. Применим к ним гомеоморфизм, показанный на рис. 4.18. Тор получается из S^2 выбрасыванием двух дисков и приклейкой вместо них ручки (рис. 4.18), гомеоморфной $S^1 \times D^1$. Для получения $\mathbb{R}P^2$ следует выбросить из S^2 один диск и приклеить вместо него лист Мебиуса (рис. 4.18). Эту операцию называют приклейкой пленки Мебиуса.

Применим описанные операции: приклейку ручки и пленки Мебиуса по два раза. Какие получим поверхности? Рассмотрим восьмиуголь-

Рис. 4.17:

Рис. 4.18:

Рис. 4.19:

ник (рис. 4.19), на границе которого расставлены буквы a, b, c, d, образующие при последовательном обходе (по часовой стрелке) слово $W = aba^{-1}b^{-1}cdc^{-1}d^{-1}$. Выполним склейки (см. рис. 4.19). Полученное двумерное многообразие называется "кренделем"; оно гомеоморфно сфере с двумя ручками: $S^2 + 2r$ (через r условно обозначаем ручку).

Лемма 4.8. *Бутылка Клейна гомеоморфна сфере, заклеенной двумя пленками Мебиуса.*

Доказательство см. на рис. 4.20. Склейка двух листов Мебиуса по их общей границе эквивалентна вклейке этих листов в сферу с двумя дырками (заклейка цилиндра двумя пленками Мебиуса). Лемма доказана. ■

Предъявим погружение $\mathbb{R}P^2$ в $\mathbb{R}^3$. Рассмотрим погружение листа Мебиуса в $\mathbb{R}^3$ в виде половинки бутылки Клейна (см. рис. 4.21). Поясним, что разрезав бутылку так, как показано на рис. 4.21, получаем два листа Мебиуса (см. лемму 4.8). В силу леммы 4.7 для получения $\mathbb{R}P^2$ достаточно заклеить лист Мебиуса диском. Осуществим эту заклейку в $\mathbb{R}^3$ так, чтобы получилось погружение $\mathbb{R}P^2 \to \mathbb{R}^3$. Поместим границу листа Мебиуса в плоскость (см. рис. 4.22) и начнем перемещать эту плоскость параллельно самой себе по направлению к зрителю. При этом будем в ней гладко деформировать граничную кривую листа Мебиуса (рис. 4.22). Автоматически достраиваем исходное погружение листа Мебиуса поверхностью, заметаемой деформирующейся окружностью. В последний момент времени окружность становится несамопересекающейся, и мы заклеиваем ее диском, что дает искомое погружение $\mathbb{R}P^2$ в $\mathbb{R}^3$.

Рис. 4.20:

Рис. 4.21:

Рис. 4.22:

Рис. 4.23:

Рис. 4.24:

Мы получили два способа построения двумерных многообразий (рис. 4.23):

1) к S^2 нужно приклеить k ручек;

2) к S^2 нужно приклеить s пленок Мебиуса.

Получаем две серии многообразий: $M^2_{g=k} = S^2 + kr$ (ориентируемые), $M^2_{\mu=s} = S^2 + s\mu$ (неориентируемые). Можно приклеивать к S^2 одновременно как ручки, так и пленки Мебиуса. Однако новых многообразий не получается. Оказывается, ручка в присутствии пленки Мебиуса превращается в две пленки Мебиуса (мы докажем это позже). Следовательно, $S^2 + kr + s\mu \cong S^2 + (2k + s)\mu$, т.е. многообразия смешанного типа гомеоморфны неориентируемым многообразиям.

Как было показано, все M^2_g гладко вкладываются в $\mathbb{R}^3$. Докажем, что неориентируемые многообразия гладко погружаются в $\mathbb{R}^3$. Мы показали это для $\mathbb{R}P^2 = S^2 + \mu$. Из рис. 4.24 следует, что M^2_μ есть склейка (так называемая связная сумма) нескольких экземпляров $\mathbb{R}P^2$. Погружая каждый из них в $\mathbb{R}^3$, получаем погружение M^2_μ в $\mathbb{R}^3$.

Теорема 4.9 (теорема классификации). *Любое гладкое, компактное, связное, замкнутое двумерное многообразие гомеоморфно либо сфере S^2 с k ручками, либо сфере S^2 с s пленками Мебиуса.*

Доказательство. Разобьем его на несколько этапов. Будем говорить, что M^2 триангулировано, если на нем отмечено конечное число точек $P_1, \ldots, P_N$, соединенных конечным числом гладких дуг γ_α так, что:

Рис. 4.25:

1) каждая дуга γ_i начинается в некоторой вершине P_α и заканчивается в некоторой вершине P_β, $P_\beta \neq P_\alpha$; причем на ней нет других P_γ, кроме P_α и P_β;

2) совокупность этих дуг разбивает M^2 на конечное число замкнутых треугольников с вершинами в точках из множества $\{P_\alpha\}$;

3) любые два таких треугольника Δ_1 и Δ_2 либо не пересекаются, либо пересекаются по одной общей вершине, либо пересекаются по одной общей стороне (т.е. по какому-то из отрезков γ_i).

Задача. Постройте триангуляции тора, проективной плоскости и бутылки Клейна.

Одно и то же многообразие допускает бесконечно много триангуляций.

Лемма 4.9. *Любое двумерное, гладкое, компактное, связное, замкнутое M^2 допускает конечную триангуляцию.*

Доказательство будет дано в гл. 5.

Рассмотрим триангуляцию M^2. Каждое ребро каждого треугольника отметим какой-либо буквой. Будем считать, что разные отрезки занумерованы разными буквами. На каждой стороне произвольным образом поставим стрелку (ориентацию). Разрежем M^2 вдоль всех отрезков. Тогда M^2 рассыпается на конечное множество треугольников, у которых на каждой стороне стоит буква и стрелка. Разрезая M^2 вдоль какого-то отрезка, мы ставим на обоих берегах разреза ту букву и ту стрелку, которые на нем были (рис. 4.25).

Наша цель — склеить обратно все треугольники, но так, чтобы получился плоский многоугольник. Возьмем какой-то треугольник и обозначим его Δ_1. Берем любую его сторону с буквой и ориентацией. Так как каждая буква входит в общий набор сторон треугольников в точности два раза (напомним, что M^2 замкнуто, т.е. не имеет края), то найдется сторона с той же буквой на другом треугольнике (обозначим его Δ_2). Эта буква найдется на треугольнике, отличном от Δ_1, так как в противном случае исходное разбиение не было бы триангуляцией. Склеим Δ_1 и Δ_2 по этой общей стороне, совмещая стрелки. Получается плоская фигура, на границе которой

Рис. 4.26:

расставлены буквы со стрелками. Снова берем какую-то букву; снова найдется Δ_3, на одной из сторон которого стоит та же буква. Приклеиваем Δ_3 и т.д. Процесс закончится, когда все треугольники исчерпаются.

Мы подклеим к плоской фигуре все треугольники, так как если бы в какой-то момент оказалось, что никакая граничная буква получающейся области не появляется среди букв, оставшихся на еще неприсоединенных треугольниках, то, склеив все буквы-дубликаты, вернемся к исходному M^2 (склейка ликвидирует разрезы). Но тогда M^2 оказалось бы несвязным, что противоречит условию. Итак, получился плоский многоугольник W.

1) Плоский многоугольник W определен неоднозначно, даже при фиксированной триангуляции. 2) Его граница ∂W состоит из четного числа ребер. 3) На каждом из них стоит буква и стрелка; каждая буква встречается ровно два раза.

Фиксируем на W ориентацию и, обходя границу (начиная с произвольной вершины), последовательно выпишем все буквы. Если направление обхода совпадает со стрелкой, отвечающей букве, то буква войдет в слово в степени $+1$; в противном случае запишем букву в степени -1. После обхода получим слово $W = a_{i_1}^{\varepsilon_1} a_{i_2}^{\varepsilon_2} \dots a_{i_k}^{\varepsilon_k}$, где $\varepsilon_\alpha = \pm 1$. Это слово однозначно задает многоугольник W (см. рис. 4.26).

Мы сопоставили каждому M^2 (в некоторой триангуляции) слово W (неоднозначно). Слово W можно рассматривать как код M^2.

Лемма 4.10. *Слово W можно перестроить так (с помощью гомеоморфизма M^2), что все вершины многоугольника W склеятся в одну точку.*

Доказательство. Разобьем вершины W на классы эквивалентных вершин. Назовем две вершины эквивалентными, если они склеиваются в одну точку при отождествлении ребер на ∂W. Если класс эквивалентных вершин только один, то все доказано. Пусть имеются не менее двух классов эквивалентных вершин: $\{P\}$ и $\{Q\}$. Можно считать, что существует ребро a такое, что его начало принадлежит $\{P\}$, а конец $\{Q\}$. Перестроим W как показано на рис. 4.27. Получим многоугольник W', у которого в классе $\{P\}$ стало на одну вершину меньше, а в $\{Q\}$ — на одну вершину больше. Будем уменьшать таким способом число вершин в $\{P\}$, пока в $\{P\}$ не останется одна вершина.

Рис. 4.27:

Рис. 4.28:

На последнем шаге, когда в $\{P\}$ — одна вершина, соседние ребра имеют вид a, a^{-1} (см. рис. 4.28). Лемма доказана. ■

Лемма 4.11. *Пусть $W = Abaa^{-1}cB$. Существует гомеоморфизм M^2, переводящий W в слово $AbcB$.*

Доказательство см. на рис. 4.28. Здесь через A и B обозначены части W, не меняющиеся при перестройке. ■

Лемма 4.12. *Если $W = BaAaC$, то существует гомеоморфизм M^2 такой, что W переходит в $W' = BA^{-1}aaC$.*

Доказательство см. на рис. 4.29. ■

Лемма 4.13. *Если $W = AaRbPa^{-1}Qb^{-1}B$, то существует гомеоморфизм M^2, переводящий W в $W' = AQPaba^{-1}b^{-1}RB$.*

Доказательство см. на рис. 4.30. ■

Лемма 4.14. *Если слово W (где уже выполнены все предыдущие операции) содержит пару букв a и a^{-1} : $W = AaBa^{-1}C$, причем $B \neq \emptyset$, то существует пара b, b^{-1} такая, что $W = AaDbQa^{-1}Rb^{-1}T$, где $B = DbQ$, $C = Rb^{-1}T$, т.е. для любой пары a, a^{-1} (где $B \neq \emptyset$) существует "зацепленная" с ней пара b, b^{-1}.*

Рис. 4.29:

Рис. 4.30:

Доказательство. Допустим противное: пусть для любой $b \in B$ соответствующая ей буква b^ε (где $\varepsilon = \pm 1$) попадает в же самое слово B. Выполним склейку W по a, см. рис. 4.31. Все вершины W разбиваются по крайней мере на два класса неэквивалентных вершин. Так как над W выполнена операция леммы 4.10, то получили противоречие. Осталось доказать, что $\varepsilon = -1$. Если бы $\varepsilon = +1$, то получили бы противоречие с видом W из леммы 4.12. Лемма доказана. ■

Итак, W приведено к произведению выражений вида $aba^{-1}b^{-1}$ (*коммутаторы*) и cc (*квадраты*). Осталось разобрать случай, когда в слове встречаются как коммутаторы, так и квадраты.

Лемма 4.15. *Если $W = Aaba^{-1}b^{-1}BccQ$, то существует гомеоморфизм, приводящий W к виду*

$$W' = Mp^2q^2d^2N = Aabd^{-1}B^{-1}bad^{-1}Q.$$

Рис. 4.31:

Рис. 4.32:

Рис. 4.33:

Доказательство см. на рис. 4.32. Получается слово $\tilde{W} = Aabd^{-1}B^{-1}bad^{-1}Q$. Осталось собрать вместе три квадрата, что можно сделать леммой 4.12. Лемма доказана. ■

Тем самым, многообразия смешанного типа гомеоморфны неориентируемым многообразиям.

Комбинируя доказанные леммы, получаем:

Лемма 4.16. *Пусть W — один из кодов M^2. Тогда существует гомеоморфизм M^2, приводящий W к одному из следующих видов*:

1) $W = aa^{-1}$;

2) $W = a_1b_1a_1^{-1}b_1^{-1}\dots a_gb_ga_g^{-1}b_g^{-1}$;

3) $W = c_1c_1c_2c_2\dots c_kc_k$.

Какие многообразия отвечают этим трем типам? В случае 1) ясно, что M^2 гомеоморфно сфере (рис. 4.33). В случае 2) M^2 представляется в виде S^2 с g ручками. Число g называется *родом поверхности*. В случае 3) M^2 гомеоморфно сфере с k пленками Мебиуса. Теорема 4.9 доказана. ■

Существуют и другие удобные представления M^2.

Теорема 4.10. *Любое гладкое, компактное, связное, замкнутое, двумерное многообразие M^2 можно представить в виде*

$$W = a_1 a_2 \dots a_N a_1^{-1} a_2^{-1} \dots a_{N-1}^{-1} a_N^{\varepsilon},$$

где $\varepsilon = -1$ тогда и только тогда, когда $M^2 = M_g^2$ (ориентируемо), в этом случае N четно; и $\varepsilon = +1$ (N — любое) тогда и только тогда, когда $M^2 = M_\mu^2$ (неориентируемо).

Доказательство несложно и предоставляется читателю.

Мы не располагаем пока аппаратом для доказательства того, что многообразия M_g^2 не гомеоморфны друг другу при $g_1 \neq g_2$, и что M_μ^2 не гомеоморфны при $\mu_1 \neq \mu_2$, а также, что M_g^2 и M_μ^2 не гомеоморфны. Мы сообщаем эти факты без доказательства.

4.6 Изометрии

Определение 4.18. Диффеоморфизм f многообразия M^n на себя называется *движением* римановой метрики g_{ij} или *изометрией*, если риманова метрика при этом отображении переходит в себя, т.е. выполнено тождество

$$g_{kp}(y) = g_{ij}\big(x(y)\big) \frac{\partial x^i(y)}{\partial y^k} \frac{\partial x^j(y)}{\partial y^p},$$

где $y^1, \dots, y^n$ — локальные координаты в некоторой окрестности точки $y \in M^n$; $x^1, \dots, x^n$ — локальные координаты в некоторой окрестности точки $x \in M^n$; $x^i = x^i(y^1, \dots, y^n)$, — функции, задающие (локально) отображение f, причем $x = f(y)$.

Мы дали "координатное" определение изометрии. Иногда удобно пользоваться инвариантным определением, не использующим локальные координаты. При диффеоморфизме f дифференциал df изоморфно отображает $T_y M^n$ на $T_x M^n$. В каждом из касательных пространств $T_y M^n$ и $T_x M^n$ определено скалярное произведение $\langle\, , \rangle_y$ и $\langle\, , \rangle_x$ соответственно, построенное по римановой метрике. Например, $\langle \mathbf{a}, \mathbf{b} \rangle_y = g_{ij}(y) a^i b^j$, где $\mathbf{a}$, $\mathbf{b} \in T_y M^n$.

Определение 4.19. Диффеоморфизм f многообразия M^n на себя называется *изометрией*, если $\langle \mathbf{a}, \mathbf{b} \rangle_y = \big\langle df(\mathbf{a}), df(\mathbf{b}) \big\rangle_x$ для любых $\mathbf{a}, \mathbf{b} \in T_y M^n$, $x = f(y)$.

Лемма 4.17. *Координатное и инвариантное определение изометрий эквивалентны.*

Доказательство. Пусть $\mathbf{a} \in T_y M^n$, $\mathbf{a} = (a^1, \dots, a^n)$ в локальных координатах $y^1, \dots, y^n$; отсюда $df(\mathbf{a}) \in T_x M^n$ имеет вид

$$\big(df(\mathbf{a})\big)^i = \frac{\partial x^i(y)}{\partial y^k} a^k,$$

так как $df\colon T_yM^n \to T_xM^n$ задается матрицей Якоби. Отсюда

$$\big\langle df(\mathbf{a}), df(\mathbf{b})\big\rangle_x = g_{ij}\big(x(y)\big)\frac{\partial x^i(y)}{\partial y^k}\frac{\partial x^j(y)}{\partial y^p}a^k a^p = g_{kp}a^k a^p,$$

что и доказывает лемму. ■

Лемма 4.18. *Множество всех изометрий риманова многообразия M^n образует группу (в алгебраическом смысле).*

Доказательство. То, что композиция изометрий снова изометрия, вытекает из правила дифференцирования сложной функции и закона изменения коэффициентов g_{ij} при замене координат. То, что f^{-1} является изометрией, вытекает из того, что матрица Якоби отображения f^{-1} обратна к матрице Якоби $J(f)$. В качестве единичного элемента группы следует взять тождественное преобразование. Лемма доказана. ■

Группа изометрий риманова многообразия M^n обычно снабжается топологией из пространства отображений и обозначается через $\mathrm{Iso}(M^n)$. Рассмотрим простейшие примеры.

Пример 23. Снабдим вещественную прямую евклидовой метрикой $ds^2 = dx^2$, где x — координата на прямой. Пусть f — диффеоморфизм $\mathbb{R}^1$ на себя, задаваемый, следовательно, строго возрастающей (или убывающей) функцией $x = f(y)$. Поскольку метрика сохраняется, то $ds^2 = (f'_y)^2dy^2 = dy^2$, т.е. $(f'_y)^2 = 1$. Таким образом, либо $f(y) = y + a$, либо $f(y) = -y + b$, где a и b — постоянные. Итак, группа изометрий вещественной прямой гомеоморфна паре вещественных прямых. Собственные изометрии сохраняют ориентацию прямой, а несобственные — меняют.

Пример 24. Найдем группу изометрий евклидовой плоскости, сохраняющих начало координат. Сначала будем искать изометрии среди линейных преобразований (можно показать, что любая изометрия плоскости линейна; но мы не будем здесь на этом останавливаться). Условие инвариантности метрики $dx^2 + dy^2$, записывается в виде матричного уравнения $E = A^\top A$, где $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ — линейное преобразование. Множество решений этого матричного уравнения составляет группу ортогональных матриц $\mathrm{O}(2)$, состоящую из матриц

$$\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$$

— определяющих вращения плоскости и матриц

$$\begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix}$$

— определяющих отражения плоскости. Вращения образуют подгруппу в группе $\mathrm{O}(2)$, обозначаемую $\mathrm{SO}(2)$. Отражения подгруппы не образуют.

Докажите, что группа $\mathrm{O}(2)$ совпадает с группой всех изометрий окружности с метрикой $ds^2 = d\varphi^2$. *Указание*: поступить по аналогии с группой $\mathrm{Iso}(\mathbb{R}^1)$.

Пример 25. Изометрии евклидовой плоскости записываются в виде $\mathbf{y} = A\mathbf{x}+\mathbf{b}$, где $A \in \mathrm{O}(2)$, а вектор $\mathbf{b}$ определяет параллельный перенос (сдвиг) на плоскости. Все такие преобразования сохраняют евклидову метрику (проверьте!). Они исчерпывают собою все изометрии плоскости.

Пример 26. Группа линейных изометрий $\mathrm{Iso}(\mathbb{R}^n)$ евклидова пространства $\mathbb{R}^n$ изоморфна группе преобразований $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, где матрица A — ортогональна, а вектор $\mathbf{b}$ задает сдвиг. Группа $\mathrm{Iso}(\mathbb{R}^n)$ изоморфна группе матриц

$$\begin{pmatrix} & A & & \mathbf{b} \\ 0 & \dots & 0 & 1 \end{pmatrix}.$$

т.е. как топологическое пространство она гомеоморфна прямому произведению $\mathrm{O}(n) \times \mathbb{R}^n$.

4.7 Двумерные многообразия как римановы поверхности алгебраических функций

Отнесем $\mathbb{C}^n$ к координатам $z^1, \dots, z^n$, где $z^\alpha = x^\alpha + iy^\alpha$. Для $n = 1$ имеем

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial \bar z} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right), \quad \frac{\partial z}{\partial \bar z} = \frac{\partial z}{\partial z} = 0.$$

Отождествим $\mathbb{C}^n\{z^k\}$ и $\mathbb{R}^{2n}\{x^k, y^k\}$. Тогда любой полином $f(z^1, \dots, z^n)$ можно записать в виде: $g(x^1, y^1, \dots, x^n, y^n)$. Обратно: любой полином $g(x^1, y^1, \dots, x^n, y^n)$ можно записать как $f(z^1, \bar z^1, \dots, z^n, \bar z^n)$.

Лемма 4.19. *Полином $f(z^1, \bar z^1, \dots, z^n, \bar z^n)$ не зависит от $\bar z^\alpha$ тогда и только тогда, когда $\partial f/\partial \bar z^\alpha \equiv 0$.*

Доказательство. В одну сторону утверждение очевидно. Проверим, что если $\partial f/\partial \bar z^\alpha \equiv 0$, то $f(z^1, \bar z^1, \dots, z^n, \bar z^n)$ не содержит переменной $\bar z^\alpha$. Допустим противное и представим f в виде полинома по степеням переменной $\bar z^\alpha$: $f = \omega \cdot (\bar z^\alpha)^p + \dots$, где коэффициенты $\omega, \dots$ не зависят от $\bar z^\alpha$. Пусть p — максимальная степень $\bar z^\alpha$. Тогда $\partial f/\partial \bar z^\alpha = p \cdot \omega (\bar z^\alpha)^{p-1} + \dots$, что противоречит условию. ■

Напомним, что функция $f(z^1, \bar z^1, \dots, z^n, \bar z^n)$ называется *аналитической*, если все $\partial f/\partial \bar z^\alpha \equiv 0$. При $n = 1$ имеем $f(x, y) = u(x, y) + iv(x, y)$. Условие $\partial f/\partial \bar z = 0$ эквивалентно соотношениям

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Нам потребуется комплексный аналог теоремы о неявной функции. Доказательство см. в теории функций комплексного переменного.

Рис. 4.34:

Утверждение 4.1. *Пусть $f(z,w)$ — комплексно аналитическая функция на $\mathbb{C}^2(z,w)$. Рассмотрим уравнение $f(z,w)=0$ относительно переменной w, и пусть в некоторой точке $P_0 \in \{f=0\}$ выполнено соотношение $\partial f/\partial w \neq 0$.*

Тогда существует открытая окрестность $U(P_0)$ точки P_0 в $\mathbb{C}^2$ такая, что в ней существует функция $w=g(z)$ со свойствами:

1) $g(z)$ — комплексно аналитическая;

2) функция $w=g(z)$ — решение уравнения $f(z,w)=0$ в окрестности $U(P_0)$, т.е. $f(z,g(z))=0$ в $U(P_0)$, и это решение единственно в $U(P_0)$.

Геометрически удобно рассматривать решение $w=g(z)$ в виде графика (рис. 4.34).

Определение 4.20. Пусть $f(z,w)$ — полином по переменным z, w в $\mathbb{C}^2$, и пусть уравнение $f(z,w)=0$ разрешимо относительно w в некоторой открытой окрестности $U(P_0)$ (например, пусть $\partial f/\partial w \neq 0$). Тогда функция $w=g(z)$, являющаяся решением этого уравнения, называется *алгебраической*, а совокупность точек (z,w) таких, что $f(z,w)=0$, называется *римановой поверхностью для алгебраической функции* $w=g(z)$ (в тех точках, где определена $w=g(z)$).

Если в точке P_0 имеем $\partial f/\partial z \neq 0$, то локально существует решение $z=\varphi(w)$, т.е. уравнение $f=0$ можно разрешить относительно z. При этом $\varphi(w)$ удовлетворяет условиям 1), 2) утверждения 4.1. Итак, условие локальной разрешимости уравнения $f(z,w)=0$ имеет вид $\operatorname{grad} f=(\partial f/\partial z, \partial f/\partial w) \neq 0$, где $\operatorname{grad} f$ — комплексный градиент функции f.

Запишем полином $f(z,w)$ в общем виде, разложив его, например, по степеням z: $f=a_0(w)z^n+a_1(w)z^{n-1}+\dots$ Здесь коэффициенты $a_k(w)$ — полиномы от w. Отметим, что римановы поверхности алгебраических функций некомпактны и "уходят на бесконечность" в $\mathbb{C}^2$. В самом деле, для любой точки w_0 возникает уравнение $a_0(w_0)z^n+a_1(w_0)z^{n-1}+\dots=0$ относительно z. По известной теореме алгебры оно всегда имеет комплексный корень z_0.

Так как f — полином, то можно перейти от неоднородных координат в $\mathbb{C}P^2$ к однородным координатам x^1,x^2,x^3, положив $z=x^1/x^3$,

$w = x^2/x^3$. Тогда $f(z,w) = \sum a_{pq} w^p z^q$ превратится в однородный полином $\sum a_{pq}(x^1)^q(x^2)^p(x^3)^{s-(p+q)}$, где s — наибольшая степень мономов $w^p z^q$. Так можно "компактифицировать" риманову поверхность, перенеся ее из $\mathbb{C}^2$ в $\mathbb{C}P^2$. Поскольку уравнение $g(x^1, x^2, x^3) = 0$ на $\mathbb{C}P^2$ полиномиально, от эта поверхность уровня в $\mathbb{C}P^3$ компактна. Мы не будем вдаваться в детали этой компактификации, так как они нам не потребуются.

Изучим топологию римановых поверхностей. С вещественной точки зрения, уравнение $f(z,w) = 0$ распадается на два: $\mathrm{Re}(f) = 0$ и $\mathrm{Im}(f) = 0$ в $\mathbb{R}^4$, т.е. в точках общего положения множество $f = 0$ — двумерная вещественная поверхность.

Теорема 4.11. *Пусть полином $f(z,w)$ имеет вид $f = w^q - P_n(z)$, где полином $P_n(z)$ не имеет кратных корней. Тогда уравнение $f(z,w) = 0$ определяет гладкое двумерное (над $\mathbb{R}$) подмногообразие в $\mathbb{C}^2(z,w)$ (одномерную комплексную алгебраическую кривую).*

Доказательство. Пусть точка $P_0 = (z_0, w_0)$ принадлежит множеству $\{f = 0\}$ и $w_0 \neq 0$. Тогда $\partial f/\partial w\Big|_{P_0} = q w_0^{q-1} \neq 0$, т.е. по теореме о неявной функции поверхность $\{f = 0\}$ является в некоторой окрестности P_0 гладким двумерным подмногообразием — графиком алгебраической функции $w = g(z)$. Решение $w = g(z)$ в нашем случае имеет вид $w = \sqrt[q]{P_n(z)}$.

Осталось рассмотреть точки $P_0 = (z_0, 0)$. Но $\mathrm{grad}\, f(P_0) = (-dP_n(z_0)/dz, 0) \neq 0$. В самом деле, если бы $dP_n(z_0)/dz = 0$, то в точке $(z_0, 0)$ мы имели бы $P_n(z_0) = 0$ и $dP_n(z_0)/dz = 0$. Отсюда z_0 —кратный корень $P_n(z)$, что противоречит условию. Итак, $\mathrm{grad}\, f(P_0) \neq 0$ во всех точках поверхности $\{f = 0\}$. Осталось применить теорему о неявной функции. ■

Опишем основные свойства алгебраических функций.

1) Алгебраические функции обычно многозначны, т.е. почти всем значениям аргумента z отвечает несколько значений функции $w = g(z)$. Пример: $w = \sqrt{z}$. Здесь каждому $z_0 \neq 0$ отвечают ровно два значения $w = \pm\sqrt{z_0}$.

Рассмотрим проекцию $\pi\colon \mathbb{C}^2 \to \mathbb{C}^1(z)$, $\pi(z,w) = (z,0)$. Прообразом любой точки z при проекции $\pi\colon \Gamma \to \mathbb{C}^1(z)$, где $\Gamma = \{f = 0\}$, являются в точности все значения функции $w = g(z)$ в точке z (рис. 4.35). Поскольку f — полином, то образом поверхности Γ при проекции является вся плоскость $\mathbb{C}^1(z)$ (см. выше).

Заменим переменную: запишем $w = g(z)$ в виде $w = \rho(P)$, где P — переменная точка на поверхности Γ (рис. 4.35). Ясно, что $w = \rho(P)$ — однозначная функция на римановой поверхности. Тем самым, алгебраическая функция стала однозначной, но за это "заплачено" усложнением области определения функции: вместо $\mathbb{C}^1(z)$ теперь аргумент пробегает поверхность Γ — двумерное многообразие.

Итак, риманова поверхность $\Gamma = \{f = 0\}$ алгебраической функции $w = g(z)$ является областью однозначности этой функции.

2) Так как для каждого $z_0 \in \mathbb{C}^1(z)$ существуют значения $w_i = g(z_0)$ (их число равно числу разных корней полиномиального уравнения $f(z_0, w) =$

Рис. 4.35:

Рис. 4.36:

0), то в открытой окрестности каждого z_0 возникает набор непрерывных функций $w = \{\varphi_i(z)\}$, $1 \leqslant i \leqslant k$, где k — степень полинома $a_0(z)w^k + a_1(z)w^{k-1} + \cdots + a_k(z) = f(z, w)$ по переменной w. Каждая из них описывает изменение какого-то одного корня уравнения $f(z, w) = 0$ при изменении z. Функции $\varphi_i(z)$ продолжаются на всю плоскость. В некоторых точках они могут совпадать и даже "меняться местами".

Функции $\varphi_i(z)$ называются *ветвями алгебраической функции* $w = g(z)$. Над каждым z_0 "висят" значения функции $w = g(z)$, являющиеся корнями уравнения

$$f(z_0, w) = a_0(z_0)w^k + a_1(z_0)w^{k-1} + \cdots + a_k(z_0) = 0.$$

Из теоремы о неявной функции следует, что каждая ветвь определяет (почти во всех точках) гладкое, даже комплексно аналитическое, подмногообразие в $\mathbb{C}^2$. Оно называется листом римановой поверхности.

3) Сопоставим точке z_0 число $k(z_0)$, равное числу различных корней уравнения $f(z_0, w) = 0$. Ясно, что $k(z_0) \leqslant k$. Если все корни уравнения $f(z_0, w) = 0$ простые, то $k(z_0) = k$. Если же есть кратные корни, то $k(z_0) < k$. В точках z_0, где $k(z_0) < k$, некоторые ветви $\varphi_i(z)$ функции $w = g(z)$ сливаются вместе (рис. 4.36).

Пусть $f(z, w) = w^q - P_n(z)$, где полином $P_n(z)$ не имеет кратных корней. В силу теоремы 4.11 поверхность $\Gamma = \{f = 0\}$ — комплексно аналитическое подмногообразие в $\mathbb{C}^2$. Точки z_0, для которых $k(z_0) < k$, назовем *точками*

Рис. 4.37:

ветвления алгебраической функции $w = g(z)$. В случае $w = \sqrt[q]{P_n(z)}$ точками ветвления в конечной части плоскости $\mathbb{C}^1(z)$ являются корни полинома $P_n(z)$.

4) В точках ветвления сливаются несколько ветвей функции $w = g(z)$. Так как точка ветвления изолирована, то ее можно считать центром малого диска D^2, где нет других точек ветвления. Рассмотрим окружность с центром в z_0, расположенную в D^2 и обойдем ее. Тогда ветви функции $g(z)$, вообще говоря, меняются местами.

Разберем пример. Пусть $f = w^2 - z$. Тогда $w = g(z) = \sqrt{z}$. Ясно, что $\varphi_1(z) = +\sqrt{z}$, $\varphi_2(z) = -\sqrt{z}$. Пусть $z = re^{i\varphi}$. Точка $0 \in \mathbb{C}^1(z)$ — единственная точка ветвления в конечной части плоскости. При обходе вокруг 0 ветви $\varphi_1(z)$ и $\varphi_2(z)$ меняются местами:

$$\varphi_1(z) = \sqrt{r}e^{\frac{i\varphi}{2}} \to -\sqrt{r}e^{\frac{i\varphi}{2}} = \varphi_2(z).$$

Условную схему см. рис. 4.37. На самой римановой поверхности особых точек нет, так как Γ — гладкое подмногообразие в $\mathbb{C}^2$.

Изучим топологические свойства римановых поверхностей. Ограничимся случаем $f(z, w) = w^2 - P_n(z)$, где $P_n(z)$ не имеет кратных корней. Поверхность Γ задается в $\mathbb{C}^2$ как график $w = g(z) = \sqrt{P_n(z)}$. Изучим поведение Γ "на бесконечности". Имеются разные способы придать этому понятию содержательный смысл. Опишем так называемую *компактификацию* позволяющую превращать некомпактные открытые многообразия в компактные многообразия, "присоединяя край". Известно несколько способов компактификации.

Сначала пополним комплексную прямую $\mathbb{C}^1(z)$ бесконечно удаленной точкой. В результате получим $\mathbb{C}P^1$. В самом деле, достаточно рассмотреть $\mathbb{C}P^1$ как множество точек $(\lambda z^1, \lambda z^2)$, где $\lambda \neq 0$ и взять отображение

$$h\colon (\lambda z^1, \lambda z^2) \mapsto \frac{\lambda z^1}{\lambda z^2} = \frac{z^1}{z^2} \in S^2 = \mathbb{C}^1(z) \cup \infty.$$

Напомним, что пополненная плоскость отождествляется со сферой при помощи стереографической проекции.

Теперь компактифицируем $\mathbb{C}^2$. Рассмотрим $\mathbb{C}P^2$ с однородными координатами (x^1, x^2, x^3). Тогда $\mathbb{C}P^2$ покрывается тремя картами, гомеоморфными $\mathbb{C}^2$. Возьмем одну из них, например, $A_3 = \{\lambda(x^1, x^2, x^3), x^3 \neq 0\}$. Положим

$$\alpha\colon A_3 \to \mathbb{C}^2(z, w), \quad \text{где} \quad z = \frac{x^1}{x^3}, \quad w = \frac{x^2}{x^3}.$$

Тогда α — гомеоморфизм между картой A_3 и $\mathbb{C}^2$. Положим $\mathbb{C}P^1 = S^2 = \{\lambda(x^1, x^2, 0)\}$. Тогда $\mathbb{C}P^2 = A \cup S^2$. Итак, $\mathbb{C}^2$ компактифицируется, превращается в в $\mathbb{C}P^2$, путем присоединения сферы S^2. Что происходит с римановой поверхностью Γ, вложенной в $\mathbb{C}^2$? Она также превращается в компактное пространство $\widetilde{\Gamma}$. Докажите, что в случае простых корней $\widetilde{\Gamma}$ — гладкое многообразие.

Вообще, пусть $g(z, w)$ — многочлен степени n по совокупности переменных. При подстановке $z = x^1/x^3$, $w = x^2/x^3$ получаем $g(z, w) = Q_n(x^1, x^2, x^3)/(x^3)^n$, где Q_n — однородный многочлен. На проективную плоскость $\mathbb{C}P^2$ уравнение $g(z, w) = 0$ продолжается в виде $Q_n(x^1, x^2, x^3) = 0$. Точки, где $x^3 = 0$ называются "бесконечно удаленными" точками римановой поверхности. В неособом случае однородное уравнение задает двумерное компактное подмногообразие.

Например, в нашем случае, при $n = 1$ имеем:

$$w^2 - z = 0, \qquad \left(\frac{x^2}{x^3}\right)^2 - \frac{x^1}{x^3} = 0, \qquad (x^2)^2 - x^1 x^3 = 0.$$

Пересечение $\widetilde{\Gamma}$ с бесконечно удаленной сферой S^2 — одна точка, имеющая в карте $A_1 = \{\lambda(x^1, x^2, x^3), x^1 \neq 0\}$ координаты $\alpha = x^2/x^1 = 0$, $\beta = x^3/x^1 = 0$. В окрестности этой точки сделаем замену $\alpha = x^2/x^1$, $\beta = x^3/x^1$, что дает

$$\left\{(x^2)^2 - x^1 x^3 = 0\right\} = \left\{\alpha^2 - \beta = 0\right\}.$$

Ясно, что это уравнение определяет гладкое подмногообразие в окрестности точки $\alpha = \beta = 0$, что и доказывает гладкость поверхности $\widetilde{\Gamma}$ в бесконечно удаленной точке.

Изучим точки ветвления функции $w = \sqrt{P_n(z)}$, определенной уже на всей сфере $S^2 = \mathbb{C}P^1$.

Утверждение 4.2. *Пусть $w^2 - P_n(z) = 0$, где $P_n(z)$ имеет только простые корни $a_1, \ldots, a_n$. Тогда все эти точки — точки ветвления $w = \sqrt{P_n(z)}$, расположенные в конечной части плоскости. На сфере S^2 к ним добавляется еще одна точка ветвления — бесконечность в том случае, когда степень P_n нечетна. Для P_n четной степени точка ∞ не есть точка ветвления. Других точек ветвления функция $w = \sqrt{P_n(z)}$ не имеет.*

Доказательство. Пусть a_k — произвольный корень $P_n(z)$. Рассмотрим малую окружность S^1 с центром в a_k, внутри которой нет других корней

Рис. 4.38:

полинома, и запишем точку $z \in S^1$ в виде $z = a_k + re^{i\varphi}$, где r — радиус окружности. Имеем

$$w(z) = \sqrt{\prod_{p \neq k} (z - a_p)} \cdot r^{1/2} \cdot e^{i\varphi/2}$$

(рис. 4.38). Так как при обходе вокруг a_k аргумент $(z - a_k)$ меняется на 2π, то $\sqrt{z - a_k}$ меняет знак. В то же время при обходе вокруг a_k аргументы чисел $z - a_p$, $p \neq k$, не получают приращения на 2π, т.е. все радикалы $\sqrt{z - a_p}$ возвращаются к прежнему значению.

Итак, $\{a_k\}$ — точки ветвления. Рассмотрим ∞ и сделаем замену $u = 1/z$. Тогда $w = g(z)$ примет вид

$$\frac{1}{\sqrt{u^n} \prod_{k=1}^{n} \sqrt{1 - a_k u}}.$$

Если n четно, то очевидно, что 0 не есть точка ветвления. Если $n = 2p + 1$ нечетно, имеется ветвление. ■

Как мы знаем, поверхность Γ — область однозначности алгебраической функции. Представим Γ как склейку нескольких листов, каждый из которых будет графиком однозначной функции. Начнем с примера.

Рассмотрим функцию $w = \pm\sqrt{z}$, имеющую на сфере две точки ветвления: 0 и ∞. Соединим их гладкой дугой γ и сделаем вдоль γ разрез от 0 до ∞. Тогда Γ получается объединением двух подмножеств: Γ_1 — график ветви $\varphi_1 = \sqrt{z}$, определенной на $S^2 \setminus \gamma$, а Γ_2 — график ветви $\varphi_2 = -\sqrt{z}$, определенной на $S^2 \setminus \gamma$. Ясно, что на $S^2 \setminus \gamma$ нельзя перейти с ветви φ_1 на ветвь φ_2. То есть, каждый из листов Γ_1 и Γ_2 гомеоморфен $S^2 \setminus \gamma$ (рис. 4.39). При проекции $\pi\colon \mathbb{C}^2 \to \mathbb{C}^1(z)$ каждый из листов Γ_1 и Γ_2 гомеоморфно проектируется на $S^2 \setminus \gamma$. Итак, Γ склеена из двух листов Γ_1 и Γ_2. Каким образом? Зададим на S^2 ориентацию и отметим один берег разреза знаком "$-$", а другой — знаком "$+$". Эти метки возникают на берегах разрезов каждого листа: Γ_1 и Γ_2. При склейке двух экземпляров $S^2 \setminus \gamma$, нужно согласовать знаки "$+$" и "$-$". Так как при обходе вокруг точек ветвления ветви меняются

Рис. 4.39:

местами, то берег γ_1^+ листа Γ_1 должен склеиться с берегом γ_2^- листа Γ_2. Соответственно склеиваются и берега γ_1^- и γ_2^+ на листах Γ_1 и Γ_2 (рис. 4.40). Ясно, что эта склейка дает сферу. Итак, нами доказана

Лемма 4.20. *Компактифицированная риманова поверхность $\widetilde{\Gamma}$ алгебраической функции $w = \pm\sqrt{z}$ гомеоморфна сфере S^2.*

Рассмотрим риманову поверхность функции $w = \sqrt{(z-a)(z-b)}$, где $a \neq b$. В силу утверждения 4.2 эта функция имеет ровно две точки ветвления: $z = a$ и $z = b$.

Лемма 4.21. *Компактифицированная риманова поверхность $\widetilde{\Gamma}$ алгебраической функции $w = \pm\sqrt{(z-a)(z-b)}$ гомеоморфна двумерной сфере.*

Доказательство. Достаточно сделать разрез γ от a до b и повторить рассуждения леммы 4.20. ■

Теорема 4.12. *Пусть $f(z,w) = w^2 - P_n(z)$, где полином P_n не имеет кратных корней. Тогда компактифицированная риманова поверхность $\widetilde{\Gamma}$ алгебраической функции $w = \pm\sqrt{P_n(z)}$ гомеоморфна сфере с $\left[\frac{n-1}{2}\right]$ ручками (через $[\,;\,]$ обозначена целая часть).*

Γ_1
0
γ^+
γ^-
∞
$\cup$
Γ_2
0
γ^+
γ^-
∞
Γ_1
0
Γ_2
γ_1^+
γ_1^-
γ_2^+
γ_2^-
Γ_1 0
0 Γ_2
γ_1^+
γ_1^-
γ_2^+
γ_2^-
∞
∞
0
Γ_1
Γ_2
$\gamma_1^- = \gamma_2^+$
$\gamma_1^+ = \gamma_2^-$
∞
0
$S^2 = \Gamma = \Gamma_1 \cup \Gamma_2$
∞

Рис. 4.40:

Рис. 4.41:

Доказательство. Пусть $n = 2p+1$. Тогда в силу утверждения 4.2 точками ветвления являются все корни $a_1, \ldots, a_n$ и точка ∞. Это множество можно разбить на пары, например, $(a_1, a_2), \ldots, (a_n, a_{n+1})$, где $a_{n+1} = \infty$. Соединяя точки a_{2k-1} и a_{2k} гладким отрезком γ_k, $1 \leqslant k \leqslant p+1$, получаем сферу S^2 с $p+1$ разрезами. Тогда на $S^2 \setminus (\gamma_1 \cup \cdots \cup \gamma_{p+1})$ обе ветви, φ_1 и φ_2, являются однозначными функциями. В самом деле, мы запретили обходы вокруг нечетного числа точек ветвления — при нашей системе разрезов мы можем обойти только вокруг четного числа точек ветвления. Следовательно, при любом обходе получается четное число изменений знака перед радикалами.

Отметим берега каждого разреза знаками $+$ и $-$. Тогда каждый лист Γ_1 и Γ_2 гомеоморфен $S^2 \setminus (\gamma_1 \cup \cdots \cup \gamma_{p+1})$.

Следовательно, для восстановления Γ следует склеить два листа, отождествив берега разрезов с учетом их знаков (рис. 4.41). Получаем сферу S^2 с $p = \left[\frac{n-1}{2}\right]$ ручками.

Пусть $n = 2p$. В силу утверждения 4.2 точками ветвления являются все корни $a_1, \ldots, a_{2p}$ и только они. Затем дословно повторяем предыдущие рассуждения. ■

Утверждение 4.3. *Каждое двумерное гладкое компактное связное замкнутое ориентируемое многообразие рода g (т.е. сфера с ручками) является*

комплексно аналитическим многообразием.

Доказательство. Рассмотрим Γ для функции $w = \pm\sqrt{P_n(z)}$, где P_n — имеет простые корни. Так как $\Gamma \subset \mathbb{C}^2$, где задается графиком $w = g(z)$ или графиком $z = \omega(w)$, то имеем две ортогональные проекции, $(z, w) \to \mathbb{C}^1(z)$ и $(z, w) \to \mathbb{C}^1(w)$. В каждой точке Γ определена по крайней мере одна из проекций, так как $\operatorname{grad}(w^2 - P_n(z)) \neq 0$ (теорема 4.11). Получаем покрытие Γ локальными дисками. Функции перехода имеют вид либо $w = g(z)$, либо $z = \omega(w)$. Так как функции $g(z)$ или $\omega(w)$ комплексно аналитические (утверждение 4.1), то утверждение доказано. ■

Неориентируемые двумерные многообразия не являются комплексно аналитическими.

Утверждение 4.4. *Каждое двумерное гладкое компактное связное замкнутое ориентируемое многообразие можно снабдить конформной римановой метрикой.*

Доказательство. Метрика конформна, если существуют координаты, в которых она имеет вид: $g_{ij} = \alpha \cdot \delta_{ij}$. Реализуем M_g^2 в пополненном $\mathbb{C}^2$ как риманову поверхность функции $w = \pm\sqrt{P_n(z)}$. Рассмотрим на M_g^2 координату z. Она обслуживает все M_g^2, кроме точки 0, в окрестности которой нужно ввести координату w, связанную с z комплексно аналитическим преобразованием $w = g(z)$. Зададим в $\mathbb{C}^2 = \mathbb{R}^4$ метрику

$$ds^2 = dz\, d\bar{z} + dw\, d\bar{w} = \sum (dx^k)^2$$

и рассмотрим ее ограничение на Γ. Имеем

$$ds^2(M_g^2) = \left(1 + |g_z'(z)|^2\right) dz\, d\bar{z},$$

или

$$ds^2(M_g^2) = (1 + u_x^2 + v_x^2)(dx^2 + dy^2),$$

где $z = x + iy$, $g = u + iv$. ■

При комплексно аналитических заменах $z = \eta(\xi)$, полученная выше конформная метрика останется конформной (проверьте!).

Глава 5

Тензорный анализ и риманова геометрия

5.1 Общее понятие гладкого тензорного поля на многообразии

Рассмотрим гладкое многообразие M^n; пусть $P \in M^n$, и $x^1, \ldots, x^n$ — регулярные координаты в окрестности P. Мы изучим объекты, инвариантные относительно замен координат. Обозначим новые координаты знаком "штрих" — $x^{1'}, \ldots, x^{n'}$.

1) Пусть $\mathbf{a} \in T_P(M^n)$ — касательный вектор к M^n. Найдем связь между компонентами вектора, вычисленными в разных координатах. Так как $\mathbf{a} = a^i \dfrac{\partial}{\partial x^i} = a^{i'} \dfrac{\partial}{\partial x^{i'}}$ и так как $\left\{\dfrac{\partial}{\partial x^{i'}}\right\}$ образуют базис в $T_P(M^n)$, то $a^{i'} = \dfrac{\partial x^{i'}}{\partial x^i} a^i$. Мы видим, что компоненты вектора $a^1, \ldots, a^n$ преобразуются с помощью матрицы Якоби $\left(\dfrac{\partial x^{i'}}{\partial x^i}\right) = J$. Это условие мы получили из инвариантности вектора как геометрического объекта при замене координат.

Теперь можно определить вектор как объект, задаваемый в каждой системе координат $x^1, \ldots, x^n$ числами $a^1, \ldots, a^n$, преобразующимися при переходе от координат $x^1, \ldots, x^n$ к координатам $x^{1'}, \ldots, x^{n'}$ по правилу $a^{i'} = \dfrac{\partial x^{i'}}{\partial x^i} a^i$. Ясно, что определенный таким образом объект будет инвариантным.

2) Пусть $f(x)$ — гладкая функция на M^n. Рассмотрим $\operatorname{grad} f = (\partial f / \partial x^i)$. Найдем закон преобразования этого набора функций. Имеем

$$\xi_{i'} = \frac{\partial f}{\partial x^{i'}} = \xi_i \frac{\partial x^i}{\partial x^{i'}},$$

где

$$\xi_i = \frac{\partial f}{\partial x^i}, \quad (J^{-1})^\top = \begin{pmatrix} \dfrac{\partial x^1}{\partial x^{1'}} & \cdots & \dfrac{\partial x^n}{\partial x^{1'}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial x^1}{\partial x^{n'}} & \cdots & \dfrac{\partial x^n}{\partial x^{n'}} \end{pmatrix},$$

т.е. набор $(\xi_1, \ldots, \xi_n)$ преобразуется по другому закону, чем компоненты вектора; а именно с помощью матрицы $(J^{-1})^\top$. Тот же закон получится из инвариантности (относительно замен координат) производной функции f по направлению вектора $\mathbf{a}$. В самом деле:

$$\frac{df}{d\mathbf{a}} = \lim_{\varepsilon \to 0} \frac{f\big(\gamma(\varepsilon)\big) - f\big(\gamma(0)\big)}{\varepsilon},$$

где

$$\gamma(t) \in M^n, \quad \gamma(0) = P, \quad \dot\gamma(0) = \mathbf{a} \in T_P(M^n),$$
$$\frac{df}{d\mathbf{a}} = a^i \frac{\partial f}{\partial x^i} = a^{i'} \frac{\partial f}{\partial x^{i'}};$$

отсюда

$$a^{i'} \xi_{i'} = a^i \xi_i, \quad \xi_{i'} = \frac{\partial x^i}{\partial x^{i'}} \xi_i,$$

что и требовалось.

3) Рассмотрим $T_P(M^n)$ и сопряженное к нему пространство $T_p^*(M^n)$ вещественных линейных функционалов $l(\mathbf{a})$, где $\mathbf{a} \in T_P(M^n)$. Как меняются компоненты функционала l при замене координат? Введем в $T_P^*(M^n)$ базис $e^1, \ldots, e^n$ из функционалов: $e^k(\mathbf{e}_\alpha) = \delta^k_\alpha$, где $\mathbf{e}_1, \ldots, \mathbf{e}_n$ — базис в $T_P(M^n)$. Так как скаляр $l(\mathbf{a})$ не зависит от выбора координат, то $l(\mathbf{a}) = l'(\mathbf{a}')$; отсюда

$$\mathbf{a} = a^i \mathbf{e}_i; \quad l(\mathbf{a}) = l_k a^k = l_{k'} a^{k'} = l_{k'} \frac{\partial x^{k'}}{\partial x^k} a^k.$$

Отсюда в силу произвольности $\mathbf{a}$ имеем $l_{k'} = \dfrac{\partial x^k}{\partial x^{k'}} l_k$. Поэтому координаты функционала l преобразуются как координаты $\operatorname{grad} f$. Элементы $T_P^*(M^n)$ называются ковекторами. Итак, $\operatorname{grad} f$ — ковектор.

Вводя в T^* сопряженный базис из ковекторов $e^1, \ldots, e^n$, отождествим T с T^*, положив $\varphi\colon T \to T^*$, $\varphi(\mathbf{a}) = l$, где $\mathbf{a} = a^k \mathbf{e}_k$, $l = \sum\limits_{k=1}^n a^k e^k$, т.е. считая, что l имеет относительно $e^1, \ldots, e^n$ те же координаты, что и $\mathbf{a}$ относительно $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Этот изоморфизм не инвариантен при замене координат $(x) \to (x')$, так как компоненты вектора преобразуются матрицей J, а ковектора — матрицей $(J^{-1})^\top$. Имеются, впрочем, замены, при которых соответствие $\varphi\colon T \to T^*$ сохраняется. Это замены, где $J = (J^{-1})^\top$, т.е. задающиеся в точке P ортогональной матрицей Якоби.

4) Пусть C — линейный однородный оператор, преобразующий в себя $T_P(M^n)$; пусть C — матрица оператора, записанная в базисе $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

При замене $(x) \to (x')$, имеем $C' = JCJ^{-1}$, т.е. $c^{i'}_{j'} = \dfrac{\partial x^{i'}}{\partial x^i}\dfrac{\partial x^j}{\partial x^{j'}}c^i_j$. Закон преобразования матрицы оператора получается из соотношения $\mathbf{b} = C(\mathbf{a})$, инвариантного относительно замен координат.

5) Рассмотрим билинейную симметричную форму B на $T_P(M^n)$: $B(\mathbf{a}, \mathbf{b}) = b_{ij}a^ib^j$. Из алгебры известно, что при замене координат матрица B формы преобразуется так: $B' = (J^{-1})^\top BJ^{-1}$, т.е. $b_{i'j'} = \dfrac{\partial x^i}{\partial x^{i'}}\dfrac{\partial x^j}{\partial x^{j'}}b_{ij}$. Как и выше, закон преобразования B получается из требования сохранения скаляра $B(\mathbf{a}, \mathbf{b})$ при заменах координат. В самом деле,

$$B(\mathbf{a}, \mathbf{b}) = b_{ij}a^ib^j \equiv b_{i'j'}a^{i'}b^{j'} = b_{i'j'}\frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^{j'}}{\partial x^j}a^ib^j$$

т.е.

$$b_{i'j'} = \frac{\partial x^i}{\partial x^{i'}}\frac{\partial x^j}{\partial x^{j'}}b_{ij}.$$

Мы привели простейшие примеры законов преобразования. Различия между ними мы указывали разным положением индексов у компонент объектов: a^i, l_j, c^i_j, b_{ij}. Верхние индексы называются *контравариантными*, нижние — *ковариантными*. Удобно задавать тип подобных объектов парой (p, q), где q — число ковариантных, p — число контравариантных индексов. Набор координат вектора имеет тип $(1, 0)$, ковектора — $(0, 1)$, оператора — $(1, 1)$, билинейной формы на векторах — $(0, 2)$. Наряду с такой формой существует и форма $B(l, m)$ на ковекторах l, m. Вычисление (по аналогии с $B(\mathbf{a}, \mathbf{b})$) дает, что $B' = JBJ^\top$.

Итак, если в каждой системе координат задан набор чисел (функций), преобразующихся по указанным правилам, то они определяют инвариантный объект, который мы изучаем с помощью координатной записи в различных системах координат. Это обстоятельство кладется в основу общего определения тензорного поля на многообразии, т.е. инвариантного геометрического объекта, не зависящего от выбора координат. От координатной системы зависит лишь конкретная запись его компонент.

Определение 5.1. *Гладким тензорным полем* типа (p, q) ранга $p + q$ называется объект, задаваемый в каждой системе координат $(x) = (x^1, \ldots, x^n)$ набором гладких функций $T^{i_1\ldots i_p}_{j_1\ldots j_q}(x)$, преобразующихся при замене координат $(x) \to (x')$ по закону

$$T^{i'_1\ldots i'_p}_{j'_1\ldots j'_q} = \frac{\partial x^{i'_1}}{\partial x^{i_1}}\cdots\frac{\partial x^{i'_p}}{\partial x^{i_p}}\frac{\partial x^{j_1}}{\partial x^{j'_1}}\cdots\frac{\partial x^{j_q}}{\partial x^{j'_q}}T^{i_1\ldots i_p}_{j_1\ldots j_q}.$$

Мы ввели объект, инвариантный относительно замен координат. В самом деле, как известно из курса алгебры, тензор (тензорное поле) можно определить как полилинейное отображение

$$T\colon \underbrace{T_P \times \cdots \times T_P}_{q} \times \underbrace{T^*_P \times \cdots \times T^*_P}_{p} \to \mathbb{R},$$

задаваемое формулой

$$T(\mathbf{a}_1,\dots,\mathbf{a}_q;l^1,\dots,l^p)=T^{i_1\dots i_p}_{j_1\dots j_q}a^{j_1}_1\dots a^{j_q}_q l^1_{i_1}\dots l^p_{i_p},$$

где $\mathbf{a}_k=a^{j_k}_k\mathbf{e}_{j_k}$; $l^k=l^k_{i_k}e^{i^k}$.

Функции $T^{i_1\dots i_p}_{j_1\dots j_q}$ являются коэффициентами отображения T. Если фиксированы базисы $\mathbf{e}_1,\dots,\mathbf{e}_n$ и $e^1,\dots,e^n$, то для $T^{i_1\dots i_p}_{j_1\dots j_q}$ возникает полезная формула

$$T^{i_1\dots i_p}_{j_1\dots j_q}=T(\mathbf{e}_{j_1},\dots,\mathbf{e}_{j_q};e^{i_1},\dots,e^{i_p}).$$

В самом деле

$$\begin{aligned}T(\mathbf{e}_{j_1},\dots,\mathbf{e}_{j_q};e^{i_1},\dots,e^{i_p})&=\\ =T^{\alpha_1\dots\alpha_p}_{\rho_1\dots\rho_q}(\mathbf{e}_{j_1})^{\rho_1}\dots(\mathbf{e}_{j_q})^{\rho_q}(e^{i_1})_{\alpha_1}\dots(e^{i_p})_{\alpha_p}&=\\ =T^{\alpha_1\dots\alpha_p}_{\rho_1\dots\rho_q}\delta^{\rho_1}_{j_1}\dots\delta^{\rho_q}_{j_q}\delta^{i_1}_{\alpha_1}\dots\delta^{i_p}_{\alpha_p}&=T^{i_1\dots i_p}_{j_1\dots j_q}.\end{aligned}$$

Все объекты в формуле $T\colon (T_P)^q\times(T^*_P)^p\to\mathbb{R}$, а также свойство полилинейности, инвариантны. От выбора координат зависит лишь конкретная запись коэффициентов $T^{i_1\dots i_p}_{j_1\dots j_q}$.

Мы будем часто работать с координатной записью, так как она полезна при конкретных вычислениях. Для удобства введем так называемые мультииндексы $(i)=(i_1\dots i_p)$, заменяя целую строку одним индексом. Например, тензорный закон запишется так:

$$T^{(i')}_{(j')}=\frac{\partial x^{(i')}}{\partial x^{(i)}}\frac{\partial x^{(j)}}{\partial x^{(j')}}T^{(i)}_{(j)},\quad\text{где}\quad\frac{\partial x^{(i')}}{\partial x^{(i)}}=\frac{\partial x^{i'_1}}{\partial x^{i_1}}\dots\frac{\partial x^{i'_p}}{\partial x^{i_p}}.$$

Изучим простейшие свойства тензорных полей.

Лемма 5.1. *Пусть $(x)\to(x')$ — регулярная замена. Тогда тензорный закон $T^{(i')}_{(j')}=\dfrac{\partial x^{(i')}}{\partial x^{(i)}}\dfrac{\partial x^{(j)}}{\partial x^{(j')}}T^{(i)}_{(j)}$ обратим, т.е.*

$$T^{(i)}_{(j)}=\frac{\partial x^{(i)}}{\partial x^{(i')}}\frac{\partial x^{(j')}}{\partial x^{(j)}}T^{(i')}_{(j')}.$$

Доказательство следует из тождества $\dfrac{\partial x^i}{\partial x^{k'}}\dfrac{\partial x^{k'}}{\partial x^j}=\dfrac{\partial x^i}{\partial x^j}=\delta^i_j$. ■

Лемма 5.2. *Пусть тензорное поле записано в системе (x) и сделаны замены координат $(x)\to(z)\to(y)$ и $(x)\to(v)\to(y)$. Тогда запись поля T в системе (y) не зависит от способа перехода от (x) к (y).*

Доказательство следует из формулы дифференцирования сложных функций:

$$\frac{\partial y^i}{\partial x^k}=\frac{\partial y^i}{\partial z^q}\frac{\partial z^q}{\partial x^k}=\frac{\partial y^i}{\partial v^\alpha}\frac{\partial v^\alpha}{\partial x^k}.\qquad■$$

Напомним свойства тензоров, известные из курса алгебры.

а) Тензоры ранга $p+q$ типа (p,q) (тензоры рассматриваются в точке) образуют линейное пространство H, причем $\dim H = n^{p+q}$. Доказательство следует из представления тензора как полилинейного отображения: линейная комбинация таких отображений снова полилинейна.

б) В пространстве H всех тензоров ранга $p+q$ типа (p,q) имеется аддитивный базис из следующих полилинейных отображений:

$$\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \dots \otimes e^{j_q}.$$

Число их равно n^{p+q}. Каждое отображение

$$\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \dots \otimes e^{j_q}$$

задается формулами

$$\begin{gathered}
\mathbf{e}_{i_\alpha}(l) = l_{i_\alpha}, \qquad \text{где} \qquad l \in T_P^* M^n; \\
e^{j_\alpha}(\mathbf{a}) = a^{j_\alpha}, \qquad \text{где} \qquad \mathbf{a} \in T_P M^n; \\
(\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \dots \otimes e^{j_q})(\mathbf{a}_1, \dots, \mathbf{a}_q, l^1, \dots, l^p) = \\
= \mathbf{e}_{i_1}(l^1) \dots \mathbf{e}_{i_p}(l^p) e^{j_1}(\mathbf{a}_1) \dots e^{j_q}(\mathbf{a}_q) = l^1_{i_1} \dots l^p_{i_p} a^{j_1}_1 \dots a^{j_q}_q.
\end{gathered}$$

Если $T\colon (T_P M^n)^q \times (T_P^* M^n)^p \to \mathbb{R}$ — произвольное полилинейное отображение ранга (p,q), то

$$T(\mathbf{a}_1, \dots, \mathbf{a}_q; l^1, \dots, l^p) = T^{i_1 \dots i_p}_{j_1 \dots j_q} a^{j_1}_1 \dots a^{j_q}_q l^1_{i_1} \dots l^p_{i_p},$$

т.е.

$$T = T^{i_1 \dots i_p}_{j_1 \dots j_q} \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \dots \otimes e^{j_q}.$$

С помощью мультииндексов это разложение произвольного тензора T по базисным тензорам записывается так: $T = T^{(i)}_{(j)} \mathbf{e}_{(i)} \otimes e^{(j)}$.

Напомним закон преобразования базисных тензоров $\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes e^{j_1} \otimes \dots \otimes e^{j_q}$ при замене координат:

$$\mathbf{e}_{(i')} \otimes e^{(j')} = \frac{\partial x^{(i)}}{\partial x^{(i')}} \frac{\partial x^{(j')}}{\partial x^{(j)}} \mathbf{e}_{(i)} \otimes e^{(j)}.$$

Все эти свойства и определения переносятся с тензоров в точке на случай гладких тензорных полей на многообразии. Нужно рассматривать линейные комбинации базисных полей типа (p,q) с переменными коэффициентами, что превращает пространство тензорных полей в бесконечномерное пространство.

5.2 Простейшие примеры тензорных полей

5.2.1 Примеры

1) Гладкое векторное поле $\mathbf{T}(x)$ на многообразии является тензорным полем типа $(1,0)$. Пространство всех гладких векторных полей бесконечномерно.

2) Гладкое ковекторное поле задается в каждой точке $x \in M^n$ линейным функционалом $l(x) \in T_x^* M^n$, гладко зависящим от точки. В этом бесконечномерном пространстве тензорных полей типа $(0,1)$ лежит бесконечномерное подпространство потенциальных полей $\operatorname{grad} f(x)$.

3) Пусть на M^n задана риманова метрика $g_{ij}(x)$. Так как $g_{i'j'} = (\partial x^i/\partial x^{i'})(\partial x^j/\partial x^{j'})g_{ij}$, то возникает тензорное поле типа $(0,2)$. Этот тензор называется *метрическим тензором.*

4) На комплексно аналитическом многообразии в каждом $T_x M^{2n}$ определен оператор I — "умножение на мнимую единицу", $I^2 = -E$. Он гладко зависит от точки и получаем тензорное поле типа $(1,1)$.

5) *Тензор моментов инерции.* Рассмотрим в $\mathbb{R}^3$ твердое тело с закрепленной точкой 0, вращающееся вокруг нее. Пусть тело состоит из конечного числа N материальных точек, жестко скрепленных между собою. Пусть массы точек — $m_1, m_2, \ldots, m_N$, а их координаты — $x^i_{(1)}, \ldots, x^i_{(N)}$, $i = 1, 2, 3$. Построим матрицу

$$a^{ij} = -\sum_{\alpha=1}^{N} m_\alpha x^i_{(\alpha)} x^j_{(\alpha)} + \delta^{ij} \sum_{\alpha=1}^{N} m_\alpha |\mathbf{x}_{(\alpha)}|^2.$$

Этот симметричный тензор ранга 2 типа $(2,0)$ называется тензором моментов инерции твердого тела.

Поясним его роль. Пусть через 0 проведена прямая с единичным направляющим вектором $\mathbf{l} = (l^1, l^2, l^3)$. Найдем сумму $\sum\limits_{ij} a^{ij} l^i l^j = H(\mathbf{l})$. Вычисления дают

$$H(\mathbf{l}) = \sum_{\alpha=1}^{N} m_\alpha \big(|\mathbf{x}_{(\alpha)}|^2 - \langle \mathbf{x}_{(\alpha)}, \mathbf{l}\rangle^2\big).$$

Величина $|\mathbf{x}_{(\alpha)}|^2 - \langle \mathbf{x}_{(\alpha)}, \mathbf{l}\rangle^2$ — это квадрат расстояния точки с массой m_α до оси $\mathbf{l}$. Мы получили известное выражение для момента инерции тела относительно оси. Собственные векторы матрицы a^{ij} задают главные оси инерции твердого тела.

6) *Тензор деформаций.* Рассмотрим сплошную среду, заполняющую некоторый объем в пространстве $\mathbb{R}^n$, отнесенном к декартовым координатам $x^1, \ldots, x^n$. Рассмотрим малую деформацию среды под действием каких-то сил. Смещения точек среды задаются гладкими функциями $u^i(x^1, \ldots, x^n)$, зависящими от точки. Координаты смещенной точки $\tilde{P}$ выражаются через координаты исходной точки так: $\tilde{x}^i = x^i + u^i(x^1, \ldots, x^n)$. Пусть смещения малы, т.е. малы функции $u^i(x)$.

Как меняются длины гладких кривых при деформации среды? Рассмотрим две близкие точки: $P(x^1, \ldots, x^n)$ и $P'(x^{1'}, \ldots, x^{n'})$. Пусть $(\Delta l)^2 = \sum\limits_{i=1}^{n} (x^{i'} - x^i)^2 = \sum\limits_{i=1}^{n} (\Delta x^i)^2$ — квадрат длины евклидова отрезка, соединяющего P и P'. Найдем квадрат длины $(\Delta l')^2$ отрезка, соединяющего образы

Рис. 5.1:

точек P и P' после их смещения в новые точки $\widetilde{P}$ и $\widetilde{P}'$ (рис. 5.1). Имеем

$$(\Delta l')^2 = \sum_{i=1}^{n}\big(x^{i'} + u^i(x') - x^i - u^i(x)\big)^2 =$$
$$= (\Delta l)^2 + 2\sum_{i=1}^{n}\Delta x^i \Delta u^i + \sum_{i=1}^{n}(\Delta u^i)^2.$$

Так как $\Delta u^i = \dfrac{\partial u^i}{\partial x^k}\Delta x^k$, то

$$(\Delta l')^2 - (\Delta l)^2 = \sum_{i,k}\left(\frac{\partial u^i}{\partial x^k} + \frac{\partial u^k}{\partial x^i}\right)\Delta x^i \Delta x^k + \sum_{i,k,p}\frac{\partial u^i}{\partial x^k}\frac{\partial u^i}{\partial x^p}\Delta x^k \Delta x^p.$$

Итак,

$$(dl')^2 - (dl)^2 = \sum_{i,k}\Big(\frac{\partial u^i}{\partial x^k} + \frac{\partial u^k}{\partial x^i}\Big)dx^i\, dx^k + \sum_{i,k,p}\frac{\partial u^i}{\partial x^k}\frac{\partial u^i}{\partial x^p}\, dx^k\, dx^p.$$

Если деформации $u^i(x)$ малы, то можно считать, что

$$(dl')^2 - (dl)^2 \cong \sum_{i,k}\left(\frac{\partial u^i}{\partial x^k} + \frac{\partial u^k}{\partial x^i}\right)dx^i\, dx^k.$$

Положим $\eta_{ik}(x) = \dfrac{\partial u^i}{\partial x^k} + \dfrac{\partial u^k}{\partial x^i}$; тогда $(dl')^2 - (dl)^2 \cong \eta_{ik}\, dx^i\, dx^k$. Набор функций η_{ik} образует тензор второго ранга типа $(0,2)$, называемый тензором малых деформаций.

7) *Тензор напряжений.* Рассмотрим деформацию упругого тела. В нем появляются “напряжения”. Пусть $d\sigma$ — плоская малая ориентированная площадка в теле, проходящая через точку P. Пусть $\mathbf{n}(P)$ — нормаль к $d\sigma$ в точке P (рис. 5.2). Вблизи площадки упругое тело разбито ею на две части: одна расположена с положительной стороны площадки, другая — с отрицательной. Считается, что первая из этих частей упругого тела действует на

Рис. 5.2:

вторую через площадку $d\sigma$ с силой $\mathbf{F}$, приложенной в P и пропорциональной площади площадки. Площадь тоже обозначим $d\sigma$. Итак, каждой нормали $\mathbf{n}(P)$ в точке P отвечает вектор $\mathbf{F}(\mathbf{n})$; т.е. $\mathbf{F} = \mathbf{F}(\mathbf{n})$. Характер зависимости устанавливается в теории упругости на основе экспериментов. Оказывается, часто можно считать зависимость $\mathbf{F}(\mathbf{n})$ линейной, т.е. $\mathbf{F}(\mathbf{n}) = Q_P(\mathbf{n})d\sigma$, где $F^i = Q^i_j n^j d\sigma$; $\{n^j\}$ — координаты $\mathbf{n}(P)$. Возникает тензорное поле Q_P типа $(1,1)$. Оно и определяет тензор напряжений.

Из механических соображений следует, что тензор напряжений Q_P симметричен. Его можно определить не только в деформированном упругом теле. Он имеется и в идеальной жидкости. Считается, что здесь силы внутреннего трения отсутствуют, а потому сила напряжения $\mathbf{F}(\mathbf{n})$, действующая на $d\sigma$, направлена по нормали к $d\sigma$. Отсюда следует, что оператор Q_P — диагонален, т.е.

$$Q^i_j = q(P)\delta^i_j; \qquad F^i = q(P)n^i(P)d\sigma,$$

где скалярная функция $q(P)$ называется давлением. Оно не зависит от направления $\mathbf{n}$ и определяется только самой точкой P. Если среда такова, что $Q_P = q(P) \cdot E$, то говорят, что выполнен закон Паскаля. Он выполняется не всегда. Например, в вязкой жидкости тензор напряжений может иметь более сложный вид.

Между тензорами деформации и напряжения имеется зависимость. В первом приближении считается, что при малых деформациях тензор напряжений линейно зависит от тензора деформаций, т.е. $Q^i_j = \alpha^{ikl}_j \eta_{kl}$, где функции α^{ikl}_j образуют тензор 4-го ранга типа $(3,1)$. В $\mathbb{R}^3$ тензор α^{ikl}_j имеет 81 компоненту. Указанное линейное соотношение называется законом Гука. Предположим, что среда однородна и изотропна, т.е. вид зависимости $Q = Q(\eta)$, не меняется при ортогональных преобразованиях, т.е. тензор α^{ikl}_j инвариантен относительно $\mathrm{SO}(3)$. Можно показать, что в этом случае $Q^i_j = Q_{ij} = \mu\eta_{ij} + \lambda\,\mathrm{Tr}(\eta) \cdot \delta_{ij}$, где числа μ и λ называются коэффициентами Ламе. Функция $\Sigma\eta_{ii} = \mathrm{Tr}(\eta)$ имеет простой смысл (мы не различаем верхних и нижних индексов, так как события происходят в $\mathbb{R}^3$). Рассмотрим перемещение частиц среды под действием малых деформаций. Тогда функции u^i — это компоненты векторного поля деформации. Обозначим его через

$\mathbf{v}(P)$. Тогда $\mathrm{Tr}(\eta) = \sum\limits_i \left(\dfrac{\partial u^i}{\partial x^i} + \dfrac{\partial u^i}{\partial x^i}\right) = 2\sum\limits_i \dfrac{\partial u^i}{\partial x^i} = 2\,\mathrm{div}(\mathbf{v})$, т.е. $\mathrm{Tr}(\eta)$ — дивергенция поля перемещений. Коэффициент μ для жидкости называется коэффициентом вязкости.

5.2.2 Алгебраические операции над тензорами

1) Сложив два тензорных поля одинакового типа и ранга $T^{i_1\ldots i_p}_{j_1\ldots j_q}$ и $P^{i_1\ldots i_p}_{j_1\ldots j_q}$ получаем новое тензорное поле (проверьте!): $C^{i_1\ldots i_p}_{j_1\ldots j_q} = T^{i_1\ldots i_p}_{j_1\ldots j_q} + P^{i_1\ldots i_p}_{j_1\ldots j_q}$.

2) Если $T^{i_1\ldots i_p}_{j_1\ldots j_q}$ — тензорное поле, а $f(x)$ — гладкая функция на M^n, то $f(x)\cdot T^{i_1\ldots i_p}_{j_1\ldots j_q}$ также тензорное поле. Доказательство очевидно.

3) *Перестановка индексов одного типа.* Рассмотрим, например, тензорное поле $T_{i_1\ldots i_q}$ и построим новое поле по формуле: $P_{i_1\ldots i_s\ldots i_\alpha\ldots i_q} = T_{i_1\ldots i_\alpha\ldots i_s\ldots i_q}$; т.е. перенумеруем компоненты тензора. Эта операция — тензорная (проверьте!). Здесь мы поменяли местами два индекса: i_s и i_α. Ясно, что можно было сделать любую перестановку.

Замечание. Перестановка ковариантного и контравариантного индексов не является, вообще говоря, тензорной операцией, так как верхние индексы преобразуются с помощью матрицы J, а нижние — с помощью $(J^{-1})^\top$. Рассмотрим, например, поле типа $(1,1)$, т.е. поле операторов C. Пусть в координатах (x) оператор симметричен, т.е. $C^i_j = C^j_i$. Другими словами, $C = C^\top$. Допустим, что и в любых других координатах $C' = (C')^\top$. Если A — матрица Якоби, то $C' = ACA^{-1}$, $ACA^{-1} = (ACA^{-1})^\top$, т.е. $BC = CB$, где $B = A^\top A$. Ясно, что матрицы B и C не всегда коммутируют. Если матрица Якоби ортогональна, то перестановка индексов i и j является тензорной операцией.

4) *Операция свертки.* Пусть $T^{i_1\ldots i_p}_{j_1\ldots j_q}$ — тензорное поле. Выделим два индекса разного типа — ковариантный j_s и контравариантный i_α. Построим функции: $P^{i_1\ldots i_{\alpha-1}i_{\alpha+1}\ldots i_p}_{j_1\ldots j_{s-1}j_{s+1}\ldots j_q} = \sum\limits_i T^{i_1\ldots i_{\alpha-1},i_\alpha=i,i_{\alpha+1}\ldots i_p}_{j_1\ldots j_{s-1},j_s=i,j_{s+1}\ldots j_q}$. Эта операция — тензорная (проверьте!). Она переводит тензор типа (p,q) в тензор типа $(p-1,q-1)$. Если $p=q$, то для исходного поля $T^{i_1\ldots i_p}_{j_1\ldots j_q}$ можно определить полную свертку, свернув все верхние индексы со всеми нижними. Получится скалярная функция $\mathrm{Tr}(T)$, являющаяся инвариантом поля T, т.е. не меняющаяся при заменах координат.

5) *Тензорное произведение.* Из двух тензорных полей $T^{i_1\ldots i_p}_{j_1\ldots j_q}$ и $P^{\alpha_1\ldots\alpha_s}_{\beta_1\ldots\beta_t}$ образуем новое поле $C^{i_1\ldots i_p\alpha_1\ldots\alpha_s}_{j_1\ldots j_q\beta_1\ldots\beta_t} = T^{i_1\ldots i_p}_{j_1\ldots j_q}P^{\alpha_1\ldots\alpha_s}_{\beta_1\ldots\beta_t}$. Ясно, что это — тензорное поле типа $(p+s,q+t)$. Оно обозначается так: $C = T\otimes P$. Тензорное умножение, вообще говоря, некоммутативно.

6) *Поднятие и опускание индексов.* Рассмотрим, например, невырожденное тензорное поле a_{ij} типа $(0,2)$, т.е. матрица $A = (a_{ij})$ невырождена. Коэффициенты обратной матрицы A^{-1} обозначим a^{ij}. Отсюда $a^{ik}a_{kj} = \delta^i_j$.

Для произвольного тензорного поля $T^{i_1\dots i_p}_{j_1\dots j_q}$ построим поле вида $P^{i_2\dots i_p}_{\alpha j_1\dots j_q} = a_{\alpha i_1} T^{i_1\dots i_p}_{j_1\dots j_q}$. Аналогично строится поле $C^{\alpha i_1\dots i_p}_{j_2\dots j_q} = a^{\alpha j_1} T^{i_1\dots i_p}_{j_1\dots j_q}$.

Первая операция называется *опусканием индекса*, вторая — *поднятием индекса*. Операции тензорные, так как являются композицией тензорных операций: умножения и свертки тензоров. Операции поднятия и опускания индексов взаимно обратны в силу невырожденности поля a^{ij}. В самом деле: $a_{j_1\alpha} C^{\alpha i_1\dots i_p}_{j_2\dots j_q} = a_{j_1\alpha} a^{\alpha s} T^{i_1\dots i_p}_{s j_2\dots j_q} = \delta^s_{j_1} T^{i_1\dots i_p}_{s j_2\dots j_q} = T^{i_1\dots i_p}_{j_1 j_2\dots j_q}$. Обычно в качестве a_{ij} берется метрический тензор g_{ij}. Если $M^n = \mathbb{R}^n$ и $g_{ij} = \delta_{ij}$, то в декартовых координатах нет разницы между верхними и нижними индексами. Например, можно считать все их нижними. Поднятие и опускание индексов в общем случае позволяет канонически отождествить $T_P M^n$ и $T^*_P M^n$. В самом деле, элементами $T_P M^n$ являются векторы, а элементами $T^*_P M^n$ — ковекторы. Построим линейные отображения:

$$A\colon T \to T^*, \qquad A(\mathbf{a}) = \xi, \qquad \text{где} \qquad \xi_i = g_{i\alpha} a^\alpha;$$
$$B\colon T^* \to T; \qquad B(\eta) = \mathbf{b}; \qquad b^i = g^{i\alpha}\eta_\alpha.$$

Ясно, что

$$B \circ A\colon T \to T, \qquad B \circ A = \mathrm{Id}_T$$

(тождественное отображение). Аналогично: $A \circ B\colon T^* \to T^*$ и $A \circ B = \mathrm{Id}_T$. Это означает, что A и B — изоморфизмы. Такое отождествление T и T^* инвариантно относительно замен координат, поскольку использованные операции — тензорные.

7) *Симметрирование.* Рассмотрим все индексы одного типа у тензорного поля T (например, нижние). Определим операцию так: $P_{j_1\dots j_q} = T_{(j_1\dots j_q)} = \dfrac{1}{q!}\sum\limits_{(\sigma)} T_{\sigma(j_1\dots j_q)}$, где суммирование ведется по всем перестановкам индексов. Операция симметрирования — тензорная (проверьте!).

Определение 5.2. Поле $T^{i_1\dots j_p}_{j_1\dots i_q}$ называется *симметричным*, если оно не меняется при перестановке любых двух индексов одного типа.

Симметрирование симметричного поля очевидно его не меняет.

8) *Альтернирование.* Рассмотрим все индексы одного типа у тензорного поля T (например, нижние). Определим альтернирование так:

$$P_{j_1\dots j_q} = T_{[j_1\dots j_q]} = \frac{1}{q!}\sum_{(\sigma)} (-1)^{\varphi(\sigma)} T_{\sigma(j_1\dots j_q)}.$$

Здесь $\varphi(\sigma)$ — четность перестановки σ (иногда пишут $(-1)^\sigma$). Операция альтернирования — тензорная (проверьте!).

Определение 5.3. Тензорное поле $T^{i_1\dots i_p}_{j_1\dots j_q}$ называется *кососимметричным*, если его компоненты меняют знак при транспонировании (перестановке) любых двух соседних индексов одного типа.

Лемма 5.3. *Альтернирование не меняет кососимметричный тензор и обращает симметричный тензор в нуль.*

Доказательство очевидно следует из определений.

Понятие симметрии и косой симметрии операторов приобретают инвариантный смысл лишь при наличии римановой метрики (более общо: при наличии невырожденного тензорного поля типа $(0,2)$ или $(2,0)$). Риманова метрика g_{ij} порождает в каждом T_PM^n скалярное произведение $\langle \mathbf{a}, \mathbf{b}\rangle_P = g_{ij}(P)a^ib^j$.

Определение 5.4. Оператор T или операторное поле типа $(1,1)$ называется *симметричным* (соответственно, *кососимметричным*), если $\langle T\mathbf{a}, \mathbf{b}\rangle \equiv \langle \mathbf{a}, T\mathbf{b}\rangle$ для любых $\mathbf{a}$, $\mathbf{b} \in T_PM^n$; соответственно: $\langle T\mathbf{a}, \mathbf{b}\rangle \equiv -\langle \mathbf{a}, T\mathbf{b}\rangle$ для кососимметричного случая.

Например, для симметричного оператора имеем: $\langle \mathbf{a}, T\mathbf{b}\rangle = (g_{ij}T^j_k)a^ib^k = T_{ik}a^ib^k$, где $T_{ik} = g_{ij}T^j_k$. Аналогично: $\langle T\mathbf{a}, \mathbf{b}\rangle = T_{ki}a^ib^k$, где $T_{ki} = g_{jk}T^j_i$. Так как $T_{ik}a^ib^k \equiv T_{ki}a^ib^k$, то $T_{ik} = T_{ki}$. Итак, после опускания индексов у тензора T типа $(1,1)$ условие симметрии приобретает привычный вид, т.е. T_{ik} не меняются при перестановке индексов. Аналогично рассматривается и кососимметрический случай.

5.2.3 Кососимметричные тензоры. Внешние дифференциальные формы

Рассмотрим ковариантные кососимметричные тензоры $T_{i_1\ldots i_k}$.

Лемма 5.4. *Кососимметричный тензор $T_{i_1\ldots i_n}$ максимального ранга n на M^n полностью определяется только одной своей компонентой $T_{12\ldots n}$ (существенной компонентой); остальные отличаются от нее только множителем $(-1)^\sigma$, т.е.*

$$T_{i_1\ldots i_n} = (-1)^\sigma T_{12\ldots n}, \quad (i_1, \ldots, i_n) = \sigma(1, 2, \ldots, n).$$

Доказательство. В исходных координатах $x^1, \ldots, x^n$ имеем: $T_{i_1\ldots i_n} = (-1)^\sigma T_{12\ldots n}$ (см. определение косой симметрии). Сделаем замену: $(x) \to (x')$. Тогда:

$$T_{1'\ldots n'} = \frac{\partial x^{i_1}}{\partial x^{1'}} \cdots \frac{\partial x^{i_n}}{\partial x^{n'}} T_{1\ldots n} = \left(\sum_\sigma (-1)^\sigma \frac{\partial x^{i_1}}{\partial x^{1'}} \cdots \frac{\partial x^{i_n}}{\partial x^{n'}}\right) T_{12\ldots n} =$$
$$= (\det J)^{-1} \cdot T_{12\ldots n},$$

где $J = \left(\dfrac{\partial x^{i'}}{\partial x^i}\right)$. Лемма доказана. ■

Для кососимметричных тензоров определена важная операция: *внешнее умножение*. По двум кососимметричным тензорам $T_{i_1\ldots i_k}$ и $P_{j_1\ldots j_q}$ определим новый кососимметричный тензор $R = T \wedge P$, где

$$R_{i_1\ldots i_k j_1\ldots j_q} = T_{[i_1\ldots i_k} P_{j_1\ldots j_q]} = \frac{1}{k!q!}\sum_{\sigma}(-1)^{\sigma} T_{\sigma(i_1\ldots i_k} P_{j_1\ldots j_q)}.$$

Это умножение билинейно. Ранги тензоров суммируются. Рассмотрим интерпретацию на языке внешних дифференциальных форм. Сначала рассмотрим кососимметричные тензоры $T_{i_1\ldots i_k}$ в одной точке M^n. Рассмотрим координаты $x^1, \ldots, x^n$ как гладкие функции в окрестности точки P. Тогда определены их дифференциалы $dx^1, \ldots, dx^n$. Для любой гладкой функции $f(x)$ ее дифференциал df есть элемент пространства, сопряженного к T_*, т.е. линейный функционал на T_*. В самом деле, если $\mathbf{a} \in T_*$, то

$$\frac{df}{d\mathbf{a}} = \frac{df\big(\gamma(t)\big)}{dt}\Big|_{t=0},$$

где $\gamma(0) = P$, $\dot\gamma(0) = \mathbf{a}$, т.е. $\dfrac{df}{d\mathbf{a}} = \dfrac{\partial f}{\partial x^k}a^k$. Если $f = x^i$, то $\dfrac{dx^i}{d\mathbf{a}} = a^i$ и $dx^k = a^k dt$.

Итак, dx^k являются функционалами на T_*, т.е. — элементами T^*. Здесь dt можно рассматривать как коэффициент пропорциональности и опускать его. Если $\mathbf{e}_1, \ldots, \mathbf{e}_n \in T_*$ — базис в T_*, то в качестве базиса в T^* возьмем $dx^1, \ldots, dx^n$, считая, что $dx^k(\mathbf{e}_\alpha) = \delta^k_\alpha$, т.е. $dx^k(\mathbf{a}) = a^k$. Определим внешнюю алгебру $\Lambda(dx^1, \ldots, dx^n)$ с образующими $dx^1, \ldots, dx^n$ и с соотношениями: $dx^i \wedge dx^j + dx^j \wedge dx^i = 0$, $i \neq j$. Умножение $\wedge$ билинейно, алгебра Λ аддитивно порождена мономами: $dx^1, \ldots, dx^n$; $dx^i \wedge dx^j$, $i < j$; $\ldots$; $dx^{i_1} \wedge \ldots \wedge dx^{i_k}$, $i_1 < \ldots < i_k$; $dx^1 \wedge \ldots \wedge dx^n$. Как известно из курса алгебры, между этими мономами нет линейных соотношений с постоянными коэффициентами. Мультипликативными образующими являются $dx^1, \ldots, dx^k$. Алгебра Λ разлагается в прямую сумму подпространств Λ^k; $\Lambda = \bigoplus_{k=0}^n \Lambda^k$, где Λ^k порождено мономами $dx^{i_1} \wedge \ldots \wedge dx^{i_k}$. При этом, $\Lambda^0 \cong \mathbb{R}$ и порождено единицей 1. Размерность алгебры Λ равна 2^n.

Рассмотрим новую алгебру $\Lambda(M^n)$, элементами которой являются линейные комбинации $\omega^{(k)} = T_{i_1\ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ и всевозможные линейные комбинации $\sum\limits_{k=0}^{n} \omega^{(k)}$, где $T_{i_1\ldots i_k}(x)$ — кососимметричное тензорное поле ранга k и индексы $i_1, \ldots, i_k$ упорядочены в порядке возрастания. Умножение в $\Lambda(M^n)$ определим ниже.

Что происходит с $\omega^{(k)}$ при замене $(x) \to (x')$?

Лемма 5.5. *Элементы $\omega^{(k)} = T_{i_1\ldots i_k}(x)\, dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ определены инвариантно, в том смысле, что при замене $(x) \to (x')$ имеем*

$$\sum_{i'_1 < \cdots < i'_k} T_{i'_1\ldots i'_2} dx^{i'_1} \wedge \ldots \wedge dx^{i'_k} \equiv \sum_{i'_1 < \cdots < i'_k} T_{i_1\ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}.$$

Доказательство. Достаточно воспользоваться законами преобразования $T_{i_1\ldots i_k}(x)$ и dx^i. ■

Элементы $\omega^{(k)}$ алгебры $\Lambda(M^n)$ называются *внешними дифференциальными формами.*

Размерность $\Lambda(M^n)$ бесконечна. Каждая внешняя форма $\omega^{(k)}$ определяет компоненты $T_{i_1\ldots i_k}(x)$ кососимметричного тензора. Верно и обратное: любое кососимметрическое тензорное поле однозначно определяет внешнюю форму.

Замечание. При определении внешних форм можно рассматривать линейные комбинации вида $A_{i_1\ldots i_k}dx^{i_1}\wedge\ldots\wedge dx^{i_k}$, где $A_{i_1\ldots i_k}$ — произвольное тензорное поле (не обязательно кососимметричное), и суммирование ведется по всем неупорядоченным наборам $i_1,\ldots,i_k$. Но в силу косой симметрии внешнего умножения имеем

$$A_{i_1\ldots i_k}dx^{i_1}\wedge\ldots\wedge dx^{i_k}=k!\sum_{i_1<\ldots<i_k}A_{[i_1\ldots i_k]}dx^{i_1}\wedge\ldots\wedge dx^{i_k},$$

т.е. любая комбинация $A_{i_1\ldots i_k}dx^{i_1}\wedge\ldots\wedge dx^{i_k}$ порождает некоторую внешнюю дифференциальную форму.

Зададим умножение в $\Lambda(M^n)$. Если $\omega^{(k)}=\sum\limits_{i_1<\ldots<i_k}T_{i_1\ldots i_k}dx^{i_1}\wedge\ldots\wedge dx^{i_k}$ и $\omega^{(s)}=\sum\limits_{j_1<\ldots<j_s}P_{j_1\ldots j_s}dx^{j_1}\wedge\ldots\wedge dx^{j_s}$ — две внешние формы, то их произведение $\omega^{(k+s)}=\omega^{(k)}\wedge\omega^{(s)}$ определим как форму вида

$$\begin{aligned}\omega^{(k+s)}&=\sum_{i_1<\ldots<i_k}T_{i_1\ldots i_k}dx^{i_1}\wedge\ldots\wedge dx^{i_k}\wedge\sum_{j_1<\ldots<j_s}P_{j_1\ldots j_s}dx^{j_1}\wedge\ldots\wedge dx^{j_s}=\\&=T_{i_1\ldots i_k}P_{j_1\ldots j_s}\underset{(i_1<\ldots<i_k)}{dx^{i_1}\wedge\ldots\wedge dx^{i_k}}\wedge\underset{(j_1<\ldots<j_s)}{dx^{j_1}\wedge\ldots\wedge dx^{j_s}}=\\&=\sum_{l_1<\ldots<l_{k+s}}T_{[l_1\ldots l_k}P_{l_{k+1}\ldots l_{k+s}]}dx^{l_1}\wedge\ldots\wedge dx^{l_{k+s}}.\end{aligned}$$

Ясно, что умножение форм совпадает с внешним умножением соответствующих им тензорных полей. Итак, $\Lambda(M^n)$ является алгеброй с единицей. Умножение ассоциативно, но не коммутативно. Алгебра $\Lambda(M^n)$ распадается в прямую сумму подпространств $\Lambda^k(M^n)=\{\sum\limits_{i_1<\ldots<i_k}T_{i_1\ldots i_k}dx^{i_1}\wedge\ldots\wedge dx^{i_k}\}$. Число k называется степенью формы $\omega^{(k)}$. Элементы $\omega^{(k)}$ называются однородными элементами алгебры $\Lambda(M^n)$. Дифференциальные формы $\omega^{(k)}$ удобно трактовать как полилинейные кососимметрические отображения. Если

$\mathbf{a}_1, \ldots, \mathbf{a}_k \in T_* M^n$, то, считая, что $dx^i(\mathbf{a}) = a^i$, получаем

$$\omega^{(k)}(\mathbf{a}_1, \ldots, \mathbf{a}_k) = \Big(\sum_{i_1 < \ldots < i_k} T_{i_1 \ldots i_k}(x) dx^{i_1} \wedge \ldots \wedge dx^{i_k} \Big)(\mathbf{a}_1, \ldots, \mathbf{a}_k) =$$
$$= \sum_{i_1 < \ldots < i_k} T_{i_1 \ldots i_k}(x) dx^{[i_1}(\mathbf{a}_1) \cdot \ldots \cdot dx^{i_k]}(\mathbf{a}_k) =$$
$$= \sum_{i_1 < \ldots < i_k} T_{i_1 \ldots i_k}(x) a_1^{[i_1} \cdot \ldots \cdot a_k^{i_k]} = T_{i_1 \ldots i_k} a_1^{i_1} \ldots a_k^{i_k}.$$

Рассмотрим декартовы координаты $x^1, \ldots, x^n$ в $\mathbb{R}^n$. Пусть $\omega^{(n)} = dx^1 \wedge \ldots \wedge dx^n$ — внешняя форма максимального ранга; $\mathbf{a}_1, \ldots, \mathbf{a}_n$ — векторы в $\mathbb{R}^n$. Обозначим через $\operatorname{vol} \Pi(\mathbf{a}_1, \ldots, \mathbf{a}_n)$, n-мерный объем параллелепипеда $\Pi(\mathbf{a}_1, \ldots, \mathbf{a}_n)$, натянутого на $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Из курса алгебры известна следующая лемма.

Лемма 5.6. *Пусть*

$$A = \begin{pmatrix} a_1^1 & \ldots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^n & \ldots & a_n^n \end{pmatrix}$$

— матрица, построенная из координат $\mathbf{a}_1, \ldots, \mathbf{a}_n$. *Тогда* $\det(A) = \operatorname{vol} \Pi(\mathbf{a}_1, \ldots, \mathbf{a}_n)$.

Вернемся к внешним формам.

Лемма 5.7. *Пусть* $\omega^{(n)} = dx^1 \wedge \ldots \wedge dx^n$ *— форма максимального ранга в декартовых координатах в* $\mathbb{R}^n$. *Тогда* $\omega^{(n)}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \operatorname{vol} \Pi(\mathbf{a}_1, \ldots, \mathbf{a}_n)$.

Доказательство. Так как $dx^i(\mathbf{a}) = a^i$, то

$$\omega^{(n)}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = (dx^1 \wedge \ldots \wedge dx^n)(\mathbf{a}_1, \ldots, \mathbf{a}_n) =$$
$$= a_1^{[i_1} \cdot \ldots \cdot a_n^{i_n]} = \det A. \qquad \blacksquare$$

Итак, объем параллелепипеда (как значение внешней формы) не скаляр. Он меняет знак при нечетной перестановке векторов. Это — "ориентированный объем". Такая точка зрения на объем плодотворна. Аналогичная интерпретация возможна и для форм меньшей степени. Пусть в $\mathbb{R}^n$ задана форма $\omega^{(k)} = dx^{i_1} \wedge \ldots \wedge dx^{i_k}$. Найдем значение $(dx^{i_1} \wedge \ldots \wedge dx^{i_k})(\mathbf{a}_1, \ldots, \mathbf{a}_k)$, где $\mathbf{a}_1, \ldots, \mathbf{a}_k$ — векторы в $\mathbb{R}^n$. Докажите, что $\omega^{(k)}(\mathbf{a}_1, \ldots, \mathbf{a}_k) = \operatorname{vol} \Pi(\mathbf{b}_1, \ldots, \mathbf{b}_k)$ — объем k-мерного параллелепипеда в плоскости $\mathbb{R}^k(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k})$, натянутого на векторы $\mathbf{b}_1, \ldots, \mathbf{b}_k$, являющиеся ортогональными проекциями векторов $\mathbf{a}_1, \ldots, \mathbf{a}_k$ на плоскость $\mathbb{R}^k(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k})$ (рис. 5.3).

Если форма $\omega^{(k)}$ есть комбинация вида $\omega_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}$, то ее значение на $\mathbf{a}_1, \ldots, \mathbf{a}_k$ будет линейной комбинацией объемов $\operatorname{vol} \Pi(\mathbf{b}_1, \ldots, \mathbf{b}_k)$ с "весами" $\omega_{i_1 \ldots i_k}$.

Рис. 5.3:

Опишем связь внешних форм с объемом области на римановом многообразии. Мы считаем, что читатель знаком с понятием многомерного (кратного) интеграла Римана. Пусть M^n — риманово многообразие и D — открытая область в M^n, диффеоморфная области U в $\mathbb{R}^n$, отнесенном к декартовым координатам $x^1, \ldots, x^n$. Пусть $g(x)$ — определитель матрицы $\big(g_{ij}(x)\big)$.

Определение 5.5. *Объемом области* $D \subset M^n$ называется число:

$$V(D) = \mathrm{vol}(D) = \int\limits_{U(x)} \cdots \int \sqrt{g(x)}\, dx^1 \wedge \ldots \wedge dx^n,$$

где $x^1, \ldots, x^n$ — декартовы координаты в области $U(x) \subset \mathbb{R}^n$.

Если область на многообразии "большая", т.е. не покрывается одной картой, то ее объем определяется при помощи разбиения единицы, ассоциированного с данным гладким атласом.

Определение обосновывается тем, что в простейших случаях оно приводит к тем же значениям объема, какие получаются из других соображений.

1) Пусть $M^n = \mathbb{R}^n$, $U = D \subset \mathbb{R}^n$ — ограниченная область в $\mathbb{R}^n$, $g_{ij} = \delta_{ij}$, т.е. $g(x) \equiv 1$, а потому

$$\mathrm{vol}(D) = \int\limits_{U(x)} \cdots \int dx^1 \ldots dx^n,$$

что совпадает с обычным, "евклидовым", определением объема области в $\mathbb{R}^n$. Если считать $dx^1, \ldots, dx^n$ бесконечно малыми величинами, то $d\sigma = dx^1 \cdot \ldots \cdot dx^n$ — это объем бесконечно малого параллелепипеда со сторонами $dx^1, \ldots, dx^n$. Итак, объем области $U \subset \mathbb{R}^n$ представляется как "сумма" объемов бесконечно малых параллелепипедов.

2) Пусть M^n — подмногообразие в $\mathbb{R}^N$ и $g_{ij}(x)$ — индуцированная риманова метрика на M^n. Пусть $x^1, \ldots, x^n$ — криволинейные координаты в окрестности точки P на M^n. Пусть область D "мала", т.е. содержится в этой окрестности. Тогда можно считать, что объем области D есть "сумма" объемов бесконечно малых параллелепипедов (но уже не прямоугольных)

Рис. 5.4:

Рис. 5.5:

$\{\Pi_k\}$, получающихся разбиением области D криволинейными "координатными плоскостями" $x^i = \mathrm{const}$ (с "бесконечно малым шагом") (рис. 5.4). Так как M^n — гладкое подмногообразие, то Π_k хорошо аппроксимируется "линейным параллелепипедом" $\widetilde{\Pi}_k$, содержащимся в $T_P M^n$, где P — вершина Π_k (рис. 5.5). Стороны Π_k направлены вдоль координатных линий $\{x^i\}$, проходящих через P. Пусть $\mathbf{a}_1, \ldots, \mathbf{a}_n$ — векторы скоростей этих линий. Так как g_{ij} индуцирована евклидовой метрикой, то $\mathrm{vol}(\Pi_k) \cong \mathrm{vol}(\widetilde{\Pi}_k)$ совпадает с евклидовым объемом $\widetilde{\Pi}_k$ в $\mathbb{R}^n$. Итак, $\mathrm{vol}(D) \cong \sum\limits_{(k)} \mathrm{vol}(\widetilde{\Pi}_k)$. Пусть около точки P, наряду с криволинейными координатами $x^1, \ldots, x^n$ введены координаты $y^1, \ldots, y^n$, ортогональные в P, т.е. являющиеся декартовыми в $T_P M^n$. Пусть $\mathbf{e}_1, \ldots, \mathbf{e}_n$ — единичные векторы скоростей этих новых координатных линий в точке P. Тогда $g_{ij}(y) = \delta_{ij}$ в точке P. Пусть $J = \left(\dfrac{\partial y^i}{\partial x^j}\right)$ — матрица Якоби замены в точке P. Тогда J переводит $\mathbf{e}_1, \ldots, \mathbf{e}_n$ в $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Докажем, что $\mathrm{vol}(\widetilde{\Pi}_k) = \sqrt{g(x)}\, d\sigma^n$, где $d\sigma^n = dx^1 \cdot \ldots \cdot dx^n$ — евклидов объем прямоугольного параллелепипеда со сторонами $dx^1, \ldots, dx^n$. Рассмотрим параллелепипед Π'_k, натянутый на $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Тогда $\mathrm{vol}\,\Pi'_k(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \sqrt{g(x)}$. В самом деле, так как $g_{ij}(y) = \delta_{ij}$, то $\big(g_{ij}(x)\big) = J^{\top}\big(g_{ij}(y)\big)J = J^{\top}J$, т.е. $\det J = \sqrt{g(x)}$. В силу леммы 5.6: $\sqrt{g(x)} = \det J = \mathrm{vol}\,\Pi'_k(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. Утверждение доказано.

Рис. 5.6:

Рис. 5.7:

Так как $\widetilde{\Pi}_k$ натянут на $(dx^1)\mathbf{a}_1, \ldots, (dx^n)\mathbf{a}_n$ (рис. 5.6), то

$$\operatorname{vol} \widetilde{\Pi}_k = \sqrt{g(x)}\, dx^1 \cdot \ldots \cdot dx^n.$$

Так как из объемов $\widetilde{\Pi}_k$ (при их суммировании) и набирается объем области D, то $\operatorname{vol}(D) = \int\limits_{U(x)} \cdots \int \sqrt{g(x)}\, dx^1 \ldots dx^n$, что и требовалось.

3) Пусть $M^2 \subset \mathbb{R}^3$ — подмногообразие, задаваемое радиус-вектором $\mathbf{r}(u, v)$. Тогда для площади области D на M^2, имеем

$$\operatorname{vol}(D) = \iint \sqrt{EG - F^2}\, dudv = \iint\limits_{U(u,v)} |\mathbf{r}_u \times \mathbf{r}_v|\, dudv.$$

4) Пусть подмногообразие $M^{n-1} \subset \mathbb{R}^n$ задано как график гладкой функции $x^n = f(x^1, \ldots, x^{n-1})$. Найдем объем ограниченной области $D \subset M^{n-1}$. Можно считать, что $\operatorname{vol}(D) = \Sigma(d\sigma)$, где $d\sigma$ — объем бесконечно малого параллелепипеда. Проектируя его на гиперплоскость $\mathbb{R}^{n-1}(x^1, \ldots, x^{n-1})$, получаем бесконечно малый объем $dw = dx^1 \cdot \ldots \cdot dx^{n-1}$, связанный с $d\sigma$ так: $d\sigma = \dfrac{dw}{\cos\alpha}$ (рис. 5.7). Здесь α — угол между нормалью $\mathbf{m}$ к M^{n-1} и вектором $\mathbf{e}_n$. Ясно, что $\cos\alpha = \langle \mathbf{e}_n, \mathbf{m} \rangle$, где $\mathbf{e}_n = (0, \ldots, 0, 1)$; $\mathbf{m} = \dfrac{\operatorname{grad} F}{|\operatorname{grad} F|}$,

$$F(x^1, \ldots, x^n) = x^n - f(x^1, \ldots, x^{n-1}),$$

т.е.

$$d\sigma = \sqrt{1 + \sum_{i=1}^{n-1} (f_{x^i})^2}\, dx^1 \cdot \ldots \cdot dx^{n-1}.$$

Итак,

$$\mathrm{vol}(D) = \int\limits_{U(x^1,\ldots,x^{n-1})} \cdots \int \sqrt{1 + \sum_{i=1}^{n-1} (f_{x^i})^2}\, dx^1 \ldots dx^{n-1}.$$

Эта формула совпадает с общим определением объема (см. выше). В самом деле, как мы знаем, индуцированная метрики на M^{n-1} имеет вид

$$ds^2 = \big(1 + (f_{x^i})^2\big)(dx^i)^2 + 2 f_{x^i} f_{x^j} dx^i dx^j;$$

откуда $g(x) = 1 + \sum\limits_{i=1}^{n-1} f_{x^i}^2$ (проверьте!). Например, если $n = 3$, то $g(x) = 1 + f_x^2 + f_y^2$.

Упражнение. Найдите площадь круга радиуса r на плоскости Лобачевского и на двумерной сфере.

Итак, мы предъявили несколько обоснований общей формулы для объема области на римановом многообразии

$$\mathrm{vol}(D) = \int\limits_{U(x)} \cdots \int \sqrt{g(x)}\, dx^1 \ldots dx^n.$$

В частности, мы доказали, что для $\mathrm{vol}(D)$ существует представление

$$\mathrm{vol}(D) = \int\limits_{U(x)} \cdots \int \sqrt{g(x)} \cdot (dx^1 \wedge \ldots \wedge dx^n)(\mathbf{a}_1 dt^1, \ldots, \mathbf{a}_n dt^n),$$

где $\mathbf{a}_1, \ldots, \mathbf{a}_n$ — касательные векторы скоростей координатных линий $\{x^i\}$ в точке P, а $dt^1, \ldots, dt^n$ — величины бесконечно малых смещений из P вдоль этих линий. Эти смещения образуют бесконечно малый параллелепипед $\widetilde{\Pi}_k(\mathbf{a}_1 dt^1, \ldots, \mathbf{a}_n dt^n)$. Поэтому формула для объема и представляется в виде, который мы только что указали. Обычно ее пишут в сокращенном виде: $\int\limits_{U(x)} \cdots \int \sqrt{g(x)}\, dx^1 \wedge \ldots \wedge dx^n$, подразумевая развернутое выражение, приведенное выше.

Ясно, что этот интеграл не зависит от выбора локальной системы координат. В самом деле, если дана замена: $(x) \to (y)$, то

$$\int\limits_{U(y)} \cdots \int \sqrt{g(y)}\, dy^1 \wedge \ldots \wedge dy^n = \int\limits_{U(y)} \cdots \int (\det J_y) \cdot dy^1 \wedge \ldots \wedge dy^n =$$

$$= \int\limits_{U\left(y(x)\right)} \cdots \int (\det J_x) \cdot \det\left(\frac{\partial x}{\partial y}\right) \cdot \det\left(\frac{\partial y}{\partial x}\right) \cdot dx^1 \wedge \ldots \wedge dx^n =$$

$$= \int\limits_{U(x)} \cdots \int \sqrt{g(x)}\, dx^1 \wedge \ldots \wedge dx^n.$$

5.3 Связности и ковариантное дифференцирование на многообразиях

5.3.1 Аффинные связности

Необходимость дифференцировать тензорные поля возникает в прикладных задачах. При этом нужно, чтобы дифференцирование было тензорной операцией, т.е. чтобы оно превращало тензор в тензор. Первым кандидатом на эту операцию является обычное дифференцирование компонент тензоров в криволинейных координатах. Возьмем, для простоты, ковариантное поле $T_{i_1,\ldots,i_k}$ и рассмотрим функции $P_{\alpha;i_1,\ldots,i_k} = \dfrac{\partial T_{i_1,\ldots,i_k}}{\partial x^\alpha}$, т.е. вычислим обычные частные производные. В другой системе координат мы должны взять производные $P_{\alpha';i'_1,\ldots,i'_k} = \dfrac{\partial T_{i'_1,\ldots,i'_k}}{\partial x^{\alpha'}}$. Вопрос: образуют ли эти наборы тензорное поле P? Другими словами, будут ли они преобразовываться друг в друга при замене координат по тензорному закону? Проверим это. При переходе от (x) к (x') имеем

$$P_{\alpha';i'_1\ldots i'_k} = \frac{\partial T_{i'_1\ldots i'_k}}{\partial x^{\alpha'}} = \frac{\partial}{\partial x^{\alpha'}}\left(\frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i'_k}} T_{i_1\ldots i_k}\right) =$$

$$= P_{\alpha;i_1\ldots i_k} \frac{\partial x^\alpha}{\partial x^{\alpha'}} \frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i'_k}} + T_{i_1\ldots i_k} \frac{\partial}{\partial x^{\alpha'}}\left(\frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x_{i'_k}}\right).$$

Итак,

$$P_{\alpha';i'_1\ldots i'_k} = P_{\alpha;i_1\ldots i_k} \frac{\partial x^\alpha}{\partial x^{\alpha'}} \frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i'_k}} + S_{\alpha';i'_1\ldots i'_k}(T, x, x'),$$

где

$$S_{\alpha';i'_1\ldots i'_k} = T_{i_1\ldots i_k} \frac{\partial}{\partial x^{\alpha'}}\left(\frac{\partial x^{i_1}}{\partial x^{i'_1}} \cdots \frac{\partial x^{i_k}}{\partial x^{i'_k}}\right).$$

При замене общего вида $(x) \to (x')$ выражение $S_{\alpha';i'_1\ldots i'_k}$, вообще говоря, отлично от нуля, т.е. $P_{\alpha;i_1\ldots i_k}$ преобразуются не по тензорному закону.

Итак, обычное дифференцирование $\dfrac{\partial}{\partial x^i}$ — не тензорная операция. Оно будет тензорным лишь в частных случаях, например, при линейных заменах координат, поскольку здесь $S_{\alpha';i'_1\ldots i'_k} \equiv 0$.

Наша цель: научиться инвариантно дифференцировать тензорные поля на многообразиях в произвольных системах координат. Сначала ограничимся евклидовым пространством, где, наряду с декартовыми, рассматриваются всевозможные криволинейные координаты. Обозначим разыскиваемую нами операцию через ∇ ("набла") и потребуем от нее следующего.

1) В декартовых координатах в $\mathbb{R}^n$ операция ∇ должна совпадать с обычным дифференцированием $\left\{\dfrac{\partial}{\partial x^i}\right\}$.

2) Операция ∇ должна быть тензорной, т.е. если T — тензорное поле, то и ∇T должно быть тензорным полем.

Начнем с примеров. Рассмотрим в $\mathbb{R}^n$ векторное поле T^i. Пусть (x) — декартовы координаты, (x') — криволинейные. Запишем условия 1) и 2), наложенные на ∇. В системе (x) имеем: $(\nabla T)^i_j = \dfrac{\partial T^i}{\partial x^j}$. При переходе к (x') должно быть: $(\nabla T)^{i'}_{j'} = \dfrac{\partial x^{i'}}{\partial x^i}\dfrac{\partial x^j}{\partial x^{j'}}(\nabla T)^i_j$. Найдем явный вид ∇ в криволинейных координатах. Имеем

$$(\nabla T)^{i'}_{j'} = \frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial}{\partial x^j}\left(\frac{\partial x^i}{\partial x^{k'}}T^{k'}\right) = \frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial x^i}{\partial x^{k'}}\frac{\partial T^{k'}}{\partial x^{\alpha'}}\frac{\partial x^{\alpha'}}{\partial x^j} +$$
$$+\frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^j}{\partial x^{j'}}T^{k'}\frac{\partial}{\partial x^j}\left(\frac{\partial x^i}{\partial x^{k'}}\right) = \delta^{i'}_{k'}\delta^{\alpha'}_{j'}\frac{\partial T^{k'}}{\partial x^{\alpha'}} + T^{k'}\frac{\partial x^{i'}}{\partial x^i}\frac{\partial^2 x^i}{\partial x^{j'}\partial x^{k'}};$$
$$(\nabla T)^{i'}_{j'} = \frac{\partial T^{i'}}{\partial x^{j'}} + T^{k'}\Gamma^{i'}_{j'k'}; \qquad \Gamma^{i'}_{j'k'} = \frac{\partial x^{i'}}{\partial x^i}\frac{\partial^2 x^i}{\partial x^{j'}\partial x^{k'}}.$$

Итак, появились функции $\Gamma^{i'}_{j'k'}$, измеряющие отклонение ∇ от операции обычного дифференцирования.

Рассмотрим теперь в $\mathbb{R}^n$ ковекторное поле T_i. Чтобы найти явный вид ∇ на T_i, нужно решить систему уравнений: $(\nabla T)_{ij} = \dfrac{\partial T_i}{\partial x_j}$; $(\nabla T)_{i'j'} = \dfrac{\partial x^i}{\partial x^{i'}}\dfrac{\partial x^j}{\partial x^{j'}}(\nabla T)_{ij}$. Аналогично предыдущему случаю получаем (проверьте!):

$$(\nabla T)_{i'j'} = \frac{\partial T_{i'}}{\partial x^{j'}} + T_{k'}\widetilde{\Gamma}^{k'}_{i'j'};$$

где $\widetilde{\Gamma}^{k'}_{i'j'} = \dfrac{\partial^2 x^{k'}}{\partial x^j\partial x^i}\dfrac{\partial x^i}{\partial x^{i'}}\dfrac{\partial x^j}{\partial x^{j'}}$. Итак, возникли функции $\widetilde{\Gamma}^{k'}_{i'j'}$, измеряющие отклонение ∇ от обычного дифференцирования на ковекторных полях.

Лемма 5.8. *Имеет место равенство* $\widetilde{\Gamma}^{k'}_{i'j'} = -\Gamma^{k'}_{i'j'}$.

Доказательство. Дифференцируя очевидное тождество $\dfrac{\partial x^{i'}}{\partial x^{i''}}\cdot\dfrac{\partial x^{i''}}{\partial x^{k'}} = \delta^{i'}_{k'}$ по $x^{p'}$, получаем $\Gamma^{i'}_{p'k'} + \widetilde{\Gamma}^{i'}_{p'k'} = 0$. Лемма доказана. ■

Итак, действие ∇ на векторные и ковекторные поля (в криволинейных координатах в $\mathbb{R}^n$) имеет вид

$$(\nabla T)^{i'}_{j'} = \frac{\partial T^{i'}}{\partial x^{j'}} + T^{k'}\Gamma^{i'}_{k'j'}, \quad (\nabla T)_{i'j'} = \frac{\partial T_{i'}}{\partial x^{j'}} - T_{k'}\Gamma^{k'}_{i'j'}.$$

Теперь рассмотрим действие ∇ на операторных полях, т.е. на полях типа $(1,1)$. Имеем

$$\begin{gathered}
(\nabla T)^i_{jk} = \frac{\partial}{\partial x^k}(T^i_j), \\
(\nabla T)^{i'}_{j'k'} = \frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial x^k}{\partial x^{k'}}\frac{\partial}{\partial x^k}\left(\frac{\partial x^i}{\partial x^{\alpha'}}\frac{\partial x^{p'}}{\partial x^j}T^{\alpha'}_{p'}\right) = \\
= \frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial x^k}{\partial x^{k'}}\frac{\partial x^i}{\partial x^{\alpha'}}\frac{\partial x^{p'}}{\partial x^j}\frac{\partial T^{\alpha'}_{p'}}{\partial x^{q'}}\frac{\partial x^{q'}}{\partial x^k} + \\
+T^{\alpha'}_{p'}\frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial x^k}{\partial x^{k'}}\frac{\partial^2 x^i}{\partial x^{q'}\partial x^{\alpha'}}\frac{\partial x^{q'}}{\partial x^k}\frac{\partial x^{p'}}{\partial x^j} + \\
+T^{\alpha'}_{p'}\frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial x^k}{\partial x^{k'}}\frac{\partial x^i}{\partial x^{\alpha'}}\frac{\partial^2 x^{p'}}{\partial x^k\partial x^j} = \frac{\partial}{\partial x^{k'}}(T^{i'}_{j'}) + T^{p'}_{j'}\Gamma^{i'}_{p'k'} - T^{i'}_{p'}\Gamma^{p'}_{j'k'}.
\end{gathered}$$

Итак, мы нашли действие ∇ на T^i_j.

Теорема 5.1. *Пусть $M^n = \mathbb{R}^n$, (x) — декартовы координаты, (x') — криволинейные координаты. Тогда в $\mathbb{R}^n$ существует тензорная операция ∇, задаваемая на произвольном тензорном поле $T^{i_1\dots i_k}_{j_1\dots j_p}$ формулой*

$$\begin{gathered}
(\nabla T)^{i'_1\dots i'_k}_{j'_1\dots j'_p;\alpha'} = \\
= \frac{\partial}{\partial x^{\alpha'}}(T^{i'_1\dots i'_k}_{j'_1\dots j'_p}) + \sum_{s=1}^{k} T^{i'_1\dots i'_s=q'\dots i'_k}_{j'_1\dots j'_p}\Gamma^{i'_s}_{q'\alpha'} - \sum_{s=1}^{p} T^{i'_1\dots i'_k}_{j'_1\dots j'_s=q'\dots j'_p}\Gamma^{q'}_{j'_s\alpha'},
\end{gathered}$$

где функции $\Gamma^{i'}_{j'q'}$ преобразуются при замене $(x') \to (x'')$ так:

$$\Gamma^{i''}_{j''k''} = \frac{\partial x^{i''}}{\partial x^{i'}}\frac{\partial x^{j'}}{\partial x^{j''}}\frac{\partial x^{k'}}{\partial x^{k''}}\Gamma^{i'}_{j'k'} + \frac{\partial x^{i''}}{\partial x^{i'}}\frac{\partial^2 x^{i'}}{\partial x^{j''}\partial x^{k''}}.$$

Доказательство. Явный вид ∇ на поле $T^{i_1\dots i_k}_{j_1\dots j_p}$ устанавливается дословным повторением проделанных выше вычислений для случая векторных, ковекторных, и операторных полей, повторенных столько раз, каков ранг поля $T^{i_1\dots i_k}_{j_1\dots j_p}$. Выписывание формул оставляется читателю, как обязательное

упражнение. Изучим закон преобразования $\Gamma^{i'}_{j'k'}$. Имеем

$$\nabla_{k''}T^{i''} = \frac{\partial T^{i''}}{\partial x^{k''}} + T^{p''}\Gamma^{i''}_{p''k''} =$$
$$= \frac{\partial x^{k'}}{\partial x^{k''}}\frac{\partial}{\partial x^{k'}}\left(\frac{\partial x^{i''}}{\partial x^{i'}}T^{i'}\right) + \frac{\partial x^{p''}}{\partial x^{p'}}T^{p'}\Gamma^{i''}_{p''k''} =$$
$$= \frac{\partial x^{k'}}{\partial x^{k''}}\frac{\partial x^{i''}}{\partial x^{i'}}\frac{\partial T^{i'}}{\partial x^{k'}} + T^{i'}\frac{\partial x^{k'}}{\partial x^{k''}}\frac{\partial^2 x^{i''}}{\partial x^{k'}\partial x^{i'}} + T^{p'}\frac{\partial x^{p''}}{\partial x^{p'}}\Gamma^{i''}_{p''k''};$$
$$\nabla_{k''}T^{i''} = \frac{\partial x^{k'}}{\partial x^{k''}}\frac{\partial x^{i''}}{\partial x^{i'}}\nabla_{k'}T^{i'} = \frac{\partial x^{k'}}{\partial x^{k''}}\frac{\partial x^{i''}}{\partial x^{i'}}\left(\frac{\partial T^{i'}}{\partial x^{k'}} + T^{p'}\Gamma^{i'}_{p'k'}\right).$$

Сравнивая, получаем

$$T^{p'}\frac{\partial x^{k'}}{\partial x^{k''}}\frac{\partial x^{i''}}{\partial x^{i'}}\Gamma^{i'}_{p'k'} = T^{p'}\frac{\partial x^{k'}}{\partial x^{k''}}\frac{\partial^2 x^{i''}}{\partial x^{k'}\partial x^{p'}} + T^{p'}\frac{\partial x^{p''}}{\partial x^{p'}}\Gamma^{i''}_{p''k''}.$$

Так как это тождество должно иметь место для любого поля T^i, то

$$\Gamma^{i''}_{p''k''} = \Gamma^{i'}_{p'k'}\frac{\partial x^{p'}}{\partial x^{p''}}\frac{\partial x^{k'}}{\partial x^{k''}}\frac{\partial x^{i''}}{\partial x^{i}} - \frac{\partial x^{p'}}{\partial x^{p''}}\frac{\partial x^{k'}}{\partial x^{k''}}\frac{\partial^2 x^{i''}}{\partial x^{k'}\partial x^{p'}}.$$

В силу леммы 5.8

$$\frac{\partial^2 x^{i''}}{\partial x^{k'}\partial x^{p'}}\frac{\partial x^{p'}}{\partial x^{p''}}\frac{\partial x^{k'}}{\partial x^{k''}} = -\frac{\partial^2 x^{i'}}{\partial x^{p''}\partial x^{k''}}\frac{\partial x^{i''}}{\partial x^{i'}}.$$

Теорема доказана. ■

Мы доказали существование “тензорного дифференцирования” пока только для случая $M^n = \mathbb{R}^n$, т.е. “в присутствии” декартовых координат. Это позволило нам в явном виде вычислить функции $\Gamma^{i'}_{j'k'}$.

Рассмотрим теперь произвольные гладкие многообразия. Новую операцию ∇ зададим аксиоматически, положив в основу определения тензорные свойства ∇ на $\mathbb{R}^n$.

Определение 5.6. Будем говорить, что на гладком многообразии M^n задано *ковариантное дифференцирование* ∇, если для каждого гладкого атласа в каждой карте заданы гладкие функции Γ^i_{jk} (называемые символами Кристоффеля), преобразующиеся при замене координат по закону

$$\Gamma^{i'}_{j'k'} = \frac{\partial x^{i'}}{\partial x^{i}}\frac{\partial x^{j}}{\partial x^{j'}}\frac{\partial x^{k}}{\partial x^{k'}}\Gamma^i_{jk} + \frac{\partial x^{i'}}{\partial x^{i}}\frac{\partial^2 x^{i}}{\partial x^{j'}\partial x^{k'}}.$$

Тогда ∇ задается формулой

$$(\nabla T)^{i_1\ldots i_k}_{j_1\ldots j_p;\alpha} =$$
$$= \frac{\partial}{\partial x^\alpha}(T^{i_1\ldots i_k}_{j_1\ldots j_p}) + \sum_{s=1}^{k} T^{i_1\ldots i_s=q\ldots i_k}_{j_1\ldots j_p}\Gamma^{i_s}_{q\alpha} - \sum_{s=1}^{p} T^{i_1\ldots i_k}_{j_1\ldots j_s=q\ldots j_p}\Gamma^{q}_{j_s\alpha}.$$

Существование многообразий, где можно ввести ∇, доказано нами выше: можно взять $M^n = \mathbb{R}^n$. Тогда Γ^i_{jk} задаются полученными выше формулами. Важно, что набор Γ^i_{jk} не образует тензора! Символы Кристоффеля преобразуются как тензор лишь в частных случаях, например, при линейных заменах координат. Иногда говорят, что символы Кристоффеля (или ∇) задают на M^n *аффинную связность*. Если координаты фиксированы, то ∇ представляется как совокупность операций ∇_k — ковариантных дифференцирований по отдельным координатам.

Определение 5.7. *Тензором кручения аффинной связности* Γ^i_{jk} называется тензор, задаваемый в каждой системе координат равенством: $\Omega^i_{jk} = \Gamma^i_{jk} - \Gamma^i_{kj}$.

Лемма 5.9. *Набор функций* Ω^i_{jk} *действительно образует тензор.*

Доказательство. Используя закон преобразования символов Кристоффеля и альтернируя $\Gamma^{i'}_{j'k'}$ по нижним индексам, получаем, в силу симметрии "нетензорного добавка" по j' и k', формулу

$$\Omega^{i'}_{j'k'} = \frac{\partial x^{i'}}{\partial x^i}\frac{\partial x^j}{\partial x^{j'}}\frac{\partial x^k}{\partial x^{k'}}\Omega^i_{jk}. \qquad \blacksquare$$

Определение 5.8. Связность (или ковариантное дифференцирование) ∇ называется *симметричной*, если тензор кручения равен нулю.

Связность, введенная нами в $\mathbb{R}^n$, была симметричной (проверьте!). Кстати, на произвольном M^n символы Кристоффеля уже не обязаны иметь вид вторых частных производных от координат, как это было в $\mathbb{R}^n$.

Предложение 5.1. *Ковариантное дифференцирование (связность)* ∇ *удовлетворяет соотношениям:*

1) *операция* $\nabla = \{\nabla_k\}$ *— линейна;*

2) *для произвольного тензорного поля* $T^{(i)}_{(j)}$ *набор функций* $\nabla_k T^{(i)}_{(j)} = (\nabla T)^{(i)}_{k;(j)}$ *образует тензорное поле;*

3) *если тензорное поле — скалярное (т.е. гладкая функция* f *на* M^n*), то*

$$\nabla f = (\nabla_k f) = \left(\frac{\partial f}{\partial x^k}\right) = \operatorname{grad} f;$$

4) *операция* ∇ *на векторных полях* T^i *имеет вид*

$$\nabla_k T^i = \frac{\partial T^i}{\partial x^k} + T^\alpha \Gamma^i_{\alpha k},$$

операция ∇ *на ковекторных полях* T_i *имеет вид*

$$\nabla_k T_i = \frac{\partial T_i}{\partial x^k} - T_\alpha \Gamma^\alpha_{ik};$$

5) операция ∇ удовлетворяет формуле Лейбница:

$$\nabla_k\big(T^{(i)}_{(j)}\cdot P^{(\alpha)}_{(\beta)}\big)=\big(\nabla_k T^{(i)}_{(j)}\big)\cdot P^{(\alpha)}_{(\beta)}+T^{(i)}_{(j)}\cdot\big(\nabla_k P^{(\alpha)}_{(\beta)}\big),$$

где $T^{(i)}_{(j)}$ и $P^{(\alpha)}_{(\beta)}$ — произвольные тензорные поля.

Доказательство. Свойства 1)–4) сразу следуют из определения ∇. Осталось доказать 5). Рассмотрим простейший случай: одно из полей $T^{(i)}_{(j)}$, $P^{(\alpha)}_{(\beta)}$ — скалярное. Тогда 5) вытекает из формулы Лейбница для скалярных функций. Пусть теперь поля T и P — векторные поля T^i, P^j. Имеем

$$\nabla_k(T^i\cdot P^j)=\frac{\partial}{\partial x^k}(T^i\cdot P^j)+T^\alpha P^j\Gamma^i_{\alpha k}+T^iP^\alpha\Gamma^j_{\alpha k}=$$
$$=\left(\frac{\partial T^i}{\partial x^k}+T^\alpha\Gamma^i_{\alpha k}\right)P^j+T^i\left(\frac{\partial P^j}{\partial x^k}+P^\alpha\Gamma^j_{\alpha k}\right)=(\nabla_kT^i)P^j+T^i(\nabla_kP^j).$$

Доказательство для произвольных полей P и T аналогично предыдущему после замены индексов i,j на мультииндексы $(i),(j)$. Предположение доказано. ■

Теорема 5.2. *Пусть на M^n задана операция $\nabla=\{\nabla_k\}$, удовлетворяющая* 1)–5) (*см. предложение* 5.1). *Тогда для произвольного тензорного поля $T^{(i)}_{(j)}$ имеет место тождество*

$$\nabla_kT^{(i)}_{(j)}=\nabla_kT^{i_1\dots i_p}_{j_1\dots j_q}=$$
$$=\frac{\partial}{\partial x^k}\big(T^{i_1\dots i_p}_{j_1\dots j_q}\big)+\sum_{r=1}^{p}T^{i_1\dots i_r=\alpha\dots i_p}_{j_1\dots j_q}\Gamma^{i_r}_{\alpha k}-\sum_{s=1}^{q}T^{i_1\dots i_p}_{j_1\dots j_s=\beta\dots j_q}\Gamma^{\beta}_{j_sk},$$

т.е. ∇ — ковариантное дифференцирование в смысле нашего определения (см. выше). Свойства 1)–5) *однозначно задают операцию ∇, т.е. ее можно вводить аксиоматически с помощью этих свойств.*

Доказательство. Действие ∇ на скалярных функциях и на тензорах ранга 1 задано согласно 1)–4). Осталось найти ∇_k на тензорах произвольного типа. Докажем лемму: любое тензорное поле разлагается в линейную комбинацию (с гладкими коэффициентами) произведений полей первого ранга. В самом деле, любое тензорное поле есть полилинейное отображение,

$$T=a^{i_1\dots i_p}_{j_1\dots j_q}\mathbf{e}_{i_1}\otimes\dots\otimes\mathbf{e}_{i_p}\otimes e^{j_1}\otimes\dots\otimes e^{j_q},$$

где $\mathbf{e}_{i_\alpha}$, e^{i_α} — поля 1-го ранга и $a^{i_1\dots i_p}_{j_1\dots j_q}$ — гладкие функции. Пусть $\{T^\alpha_{i_s}\}$ компоненты тензора $\mathbf{e}_{i_s}$, а $T^{j_s}_\beta$ — компоненты e^{j_s}. Тогда

$$T=T^{\alpha_1\dots\alpha_p}_{\beta_1\dots\beta_q}\left(\frac{\partial}{\partial x^{\alpha_1}}\right)\otimes\dots\otimes\left(\frac{\partial}{\partial x^{\alpha_p}}\right)\otimes dx^{\beta_1}\otimes\dots\otimes dx^{\beta_q},$$

где

$$T^{\alpha_1\ldots\alpha_p}_{\beta_1\ldots\beta_q} = a^{i_1\ldots i_p}_{j_1\ldots j_q} T^{\alpha_1}_{i_1}\ldots T^{\alpha_p}_{i_p} T^{j_1}_{\beta_1}\ldots T^{j_q}_{\beta_q}.$$

Вернемся к доказательству теоремы. Для простоты рассмотрим случай 2-го ранга (для произвольного поля выкладки аналогичны). В силу линейности ∇ формулу 5) достаточно проверить на мономах $T_{ij} = \alpha(x)T_iP_j$, где $\alpha(x)$ — гладкая функция. По формуле Лейбница 5) имеем

$$\nabla_k(T_{ij}) = \nabla_k(\alpha(x)T_iP_j) = \frac{\partial\alpha}{\partial x^k}T_iP_j + \alpha\left(\frac{\partial T_i}{\partial x^k} - T_p\Gamma^p_{ik}\right)P_j +$$
$$+\alpha T_i\left(\frac{\partial P_j}{\partial x^k} - P_q\Gamma^q_{jk}\right) = \frac{\partial}{\partial x^k}(T_{ij}) - \Gamma^p_{ik}T_{pj} - \Gamma^q_{jk}T_{iq},$$

что и требовалось. ■

Определение 5.9. Пусть ∇ — аффинная связность на M^n. Локальные координаты $x^1,\ldots,x^n$ называются *евклидовыми* для ∇, если в них $\Gamma^i_{jk}(x) = 0$.

Такие координаты могут и не существовать, например, если ∇ не симметрична. В этом случае Ω^i_{jk} отличен от нуля. Если бы существовали координаты, в которых $\Gamma^i_{jk} \equiv 0$, то обратился бы в нуль и Ω^i_{jk}, а тогда он был бы тождественным нулем в любых координатах в силу тензорности.

Общее ковариантное дифференцирование никак не связано с римановой метрикой. Риманова метрика и аффинная связность — это различные структуры на M^n. В частности, евклидовы координаты для аффинной связности и евклидовы координаты для метрики g_{ij} (т.е. такие, в которых $g_{ij} = \delta_{ij}$), это — разные понятия.

Пусть теперь на M^n задана метрика g_{ij}. Тогда среди всех аффинных симметричных связностей выделяется одна (и только одна!) связность, согласованная с этой метрикой и полностью определяемая ею. Такие связности называются римановыми.

5.3.2 Римановы связности

Определение 5.10. Аффинная симметричная связность $\nabla = \{\Gamma^i_{jk}\}$ на римановом многообразии называется *согласованной с метрикой* g_{ij} (или называется *римановой связностью*), если $\nabla(g_{ij}) \equiv 0$.

В силу тензорности ∇ тождество $\{\nabla_k(g_{ij}) = 0\}$ имеет место в любых координатах. Относительно римановой связности тензор g_{ij} — "постоянен" в том смысле, что его ковариантная производная равна нулю. В частности, для любого тензорного поля $T^{(p)}_{(q)}$ имеем $\nabla_k\big(g_{ij}T^{(p)}_{(q)}\big) \equiv g_{ij}\nabla_k\big(T^{(p)}_{(q)}\big)$, см. формулу Лейбница.

Теорема 5.3. *Пусть g_{ij} — риманова метрика на M^n. Тогда существует и единственна симметричная аффинная связность, согласованная с g_{ij}, причем*

$$\Gamma^i_{jk} = \frac{1}{2}g^{i\alpha}\left(\frac{\partial g_{\alpha k}}{\partial x^j} + \frac{\partial g_{j\alpha}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^\alpha}\right).$$

Доказательство. Предположим, что существование доказано. Покажем единственность связности. Согласно определению: $\nabla_k(g_{ij}) = 0$. Достаточно доказать, что отсюда однозначно находятся Γ^i_{jk} как функции g_{ij}. Имеем: $\nabla_k g_{ij} = 0$, $\dfrac{\partial g_{ij}}{\partial x^k} = g_{\alpha j}\Gamma^\alpha_{ik} + g_{\alpha i}\Gamma^\alpha_{jk}$. Циклически переставляя индексы, получаем

$$\begin{array}{c} + \\ + \\ - \end{array} \left\{ \begin{array}{ll} \dfrac{\partial g_{ij}}{\partial x^k} & = \Gamma_{j;ik} + \Gamma_{i;jk} \\ \dfrac{\partial g_{ki}}{\partial x^j} & = \Gamma_{i;kj} + \Gamma_{k;ij} \\ \dfrac{\partial g_{jk}}{\partial x^i} & = \Gamma_{k;ji} + \Gamma_{j;ki} \end{array} \right| \begin{array}{l} (ijk) \\ (kij) \\ (jki), \end{array}$$

где $\Gamma_{i;kj} = g_{i\alpha}\Gamma^\alpha_{kj}$. Складывая первые два тождества и вычитая третье, получаем

$$\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^i} = 2\Gamma_{i;jk} = 2g_{i\alpha}\Gamma^\alpha_{jk}$$

(напомним: $\Gamma^i_{jk} = \Gamma^i_{kj}$). Далее,

$$\Gamma^i_{jk} = \frac{1}{2}g^{i\alpha}\left(\frac{\partial g_{\alpha k}}{\partial x^j} + \frac{\partial g_{j\alpha}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^\alpha}\right).$$

Мы решали здесь систему вида: $g_{\alpha\beta}T^\alpha = Q_\beta$. Так как $g_{\alpha\beta}$ — невырожденный тензор, то существует обратный тензор $g^{\alpha\beta}$, а потому $T^\alpha = g^{\beta\alpha}Q_\beta$.

Итак, если решение исходной системы существует, то оно единственно. Для доказательства существования определим Γ^i_{jk} с помощью полученных выше формул. Повторяя выкладки в обратном порядке, получаем $\nabla_k(g_{ij}) \equiv 0$. Теорема доказана. ■

Замечание. Пусть на M^n существуют координаты, в которых $g_{ij} = \delta_{ij}$. Существование такой системы эквивалентно существованию координат, в которых (g_{ij}) становится постоянной матрицей (не зависящей от точки). Тогда линейной заменой эта матрица приводится (одновременно во всех точках окрестности) к виду δ_{ij}. Поэтому координаты можно называть евклидовыми, если в них метрический тензор постоянен. Вернемся к римановой связности.

Предложение 5.2. *Система координат евклидова с точки зрения римановой связности, согласованной с g_{ij}, тогда и только тогда, когда эти координаты евклидовы с точки зрения g_{ij} (т.е. g_{ij} локально постоянен).*

Доказательство. В самом деле, если g_{ij} локально постоянен, то, в силу теоремы 5.3 $\Gamma^i_{jk} \equiv 0$. Обратно: если $\Gamma^i_{jk} = 0$, то в силу теоремы 5.3: $\dfrac{\partial g_{ij}}{\partial x^k} = g_{\alpha j}\Gamma^\alpha_{jk} + g_{\alpha i}\Gamma^\alpha_{jk} = 0$, т.е. g_{ij} постоянен в этих координатах, что и требовалось. ■

Рассмотрим гладкую гиперповерхность $V^{n-1} \subset \mathbb{R}^n$, заданную графиком: $x^n = f(x^1, \ldots, x^{n-1})$. Пусть $P_0 \in V^{n-1}$ — точка, в которой $T_{P_0}(V^{n-1})$ параллельна плоскости $\mathbb{R}^{n-1}(x^1, \ldots, x^{n-1})$, т.е. $\dfrac{\partial f(P_0)}{\partial x^i} = 0$. Мы вычисляли g_{ij} относительно таких координат. Имеем: $g_{ij} = \delta_{ij} + f_{x^i} f_{x^j}$. Отсюда: $\Gamma^i_{jk}(P_0) = 0$ для связности, согласованной с g_{ij}. В окрестности P_0 функции Γ^i_{jk} могут быть ненулевыми.

Риманова связность позволяет корректно определить дивергенцию потока. Отметим, что понятие дивергенции (изменения бесконечно малого объема) предполагает наличие метрики.

Итак, пусть $\mathbf{T}$ — векторное поле на M^n с римановой связностью. Положим: $\operatorname{div}(\mathbf{T}) = \nabla_i T^i$. Ясно, что $\operatorname{div}(\mathbf{T})$ — скалярная функция.

Предложение 5.3. *Имеет место формула*

$$\operatorname{div}(\mathbf{T}) = \frac{\partial T^i}{\partial x^i} + T^\alpha \frac{\partial}{\partial x^\alpha}(\ln \sqrt{g}),$$

где g — определитель матрицы (g_{ij}).

Доказательство. В силу теоремы 5.3 имеем

$$\operatorname{div} \mathbf{T} = \nabla_i T^i = \frac{\partial T^i}{\partial x^i} + T^\alpha \Gamma^i_{\alpha i} =$$
$$= \frac{\partial T^i}{\partial x^i} + T^\alpha \cdot \frac{1}{2} g^{ip} \left(\frac{\partial g_{\alpha p}}{\partial x^i} + \frac{\partial g_{pi}}{\partial x^\alpha} - \frac{\partial g_{\alpha i}}{\partial x^p} \right),$$

Далее,

$$\operatorname{div} \mathbf{T} = \frac{\partial T^i}{\partial x^i} + \frac{1}{2} T^\alpha g^{ip} \frac{\partial g_{ip}}{\partial x^\alpha},$$

так как

$$g^{ip} \frac{\partial g_{\alpha p}}{\partial x^i} - g^{ip} \frac{\partial g_{\alpha i}}{\partial x^p} = 0.$$

Мы утверждаем далее, что

$$g^{ip} \frac{\partial g_{ip}}{\partial x^\alpha} = 2 \frac{\partial}{\partial x^\alpha}(\ln \sqrt{g}).$$

Напомним, что $g^{ip} = \dfrac{\Delta_{ip}}{g}$, где Δ_{ip} — минор, дополнительный к g_{ip} в матрице (g_{ij}). Отсюда следует, что достаточно проверить соотношение

$$\sum_{(i,p)} \Delta_{ip} \frac{\partial g_{ip}}{\partial x^\alpha} = \frac{\partial g}{\partial x^\alpha}.$$

Определитель g есть сумма однородных мономов степени n. Соберем все члены в сумме g, содержащие g_{ip}. Тогда: $g = \ldots + g_{ip} R_{ip} + \ldots$, где R_{ip} —

полином степени $n-1$, уже не содержащий g_{ip}. При этом, $R_{ip} = \Delta_{ip}$. Вычисляя $\dfrac{\partial g}{\partial x^\alpha}$ видим, что $\dfrac{\partial g_{ip}}{\partial x^\alpha}$ войдет с множителем Δ_{ip}. В другие слагаемые, т.е. в (A_{ip}), где $(A_{ip}) = g - g_{ip}\Delta_{ip}$, функция g_{ip} не входит. Следовательно, $\dfrac{\partial g}{\partial x^\alpha} = \sum\limits_{(i,p)} \dfrac{\partial g_{ip}}{\partial x^\alpha}\Delta_{ip}$, что и требовалось. ■

5.4 Параллельный перенос и геодезические на многообразиях

5.4.1 Предварительные замечания

Рассмотрим гладкое многообразие (не обязательно риманово). Во многих задачах требуется сравнить касательные векторы, приложенные в различных точках. Для произвольного M^n задача сложна по той причине, что T_xM^n и T_yM^n различны. В случае $M^n = \mathbb{R}^n$ имеется операция параллельного переноса, позволяющая сравнивать векторы в разных точках. Формально ее можно определить так. Рассмотрим пару точек P и Q и пусть $\mathbf{a}$ — вектор в точке P. Рассмотрим гладкую кривую $\gamma(t)$ такую, что $\gamma(0) = P$ и $\gamma(1) = Q$. Параллельно перенесем $\mathbf{a}$ вдоль $\gamma(t)$ так, чтобы начало вектора скользило по $\gamma(t)$. В результате получаем вдоль γ векторное поле $\mathbf{a}(t)$, имеющее постоянные компоненты, равные компонентам $\mathbf{a}$ в начальный момент. Производные по t от компонент поля $\mathbf{a}(t)$ равны нулю. Вектор $\mathbf{a}(1)$, возникший в точке Q, не зависит от выбора кривой γ. В этом смысле параллельный перенос в $\mathbb{R}^n$ не зависит от пути.

При переходе к произвольному M^n эта простая схема разрушается, поскольку в общем случае M нельзя покрыть одной картой, т.е. ввести единую систему координат. Предположим сначала, что M^2 гладко вложено в $\mathbb{R}^3$ и пусть P, Q — близкие точки на M^2. Пусть $\mathbf{a} \in T_PM^2$ и γ — кривая на поверхности из P в Q. Можно предложить такой "параллельный перенос" на M^2. Рассмотрим $\mathbf{a}$ как вектор из $\mathbb{R}^3$ и перенесем его параллельно вдоль γ с точки зрения $\mathbb{R}^3$. Получится некоторый вектор $\mathbf{b}$ в точке Q. Но $\mathbf{b}$ уже не обязан лежать в T_QM^2. Поэтому мы спроектируем $\mathbf{b}$ на T_QM^2 и назовем его проекцию $\pi\mathbf{b}$ результатом параллельного переноса из P в Q (рис. 5.8). Эта операция не зависит от пути. Однако она корректно определена лишь в малой окрестности точки P. Если мы захотим перенести $\mathbf{a}$ "далеко", то может оказаться, что вектор $\pi\mathbf{b}$ окажется нулевым. Например, так произойдет на S^2 (рис. 5.9), если мы пройдем половину меридиана. "Параллельный перенос" , дающий в результате нулевой вектор, следует отвергнуть. Впрочем, можно поступить более аккуратно. Будем двигаться по γ бесконечно малыми шагами и каждый раз ортогонально проектировать получающийся вектор на $T_{\gamma(t)}M^2$. Оказывается, возникает корректный "параллельный перенос" на $M^2 \subset \mathbb{R}^3$. Мы не будем здесь изучать эту операцию детальнее.

Обратим внимание, что при определении параллельного переноса мы вынуждены фиксировать кривую, вдоль которой нужно двигаться. На много-

Рис. 5.8:

Рис. 5.9:

образии общего вида заранее неясно: зависит или нет результат переноса от пути. Может оказаться, что зависит. Пример: параллельный перенос вдоль оси листа Мебиуса (рис. 5.10). В данном случае зависимость переноса от пути есть следствие неориентируемости, однако и на ориентируемом многообразии параллельный перенос будет, вообще говоря, зависеть от пути.

Рис. 5.10:

5.4.2 Уравнение параллельного переноса на многообразии

Положим поэтому в основу понятия параллельного переноса на M^n операцию дифференцирования ∇.

Пусть P, Q — две точки на M^n, соединенные гладкой или кусочно-гладкой траекторией $\gamma(t)$ такой, что $\gamma(0) = P$ и $\gamma(1) = Q$. Пусть $\dot\gamma$ поле скоростей вдоль $\gamma(t)$. Компоненты этого поля обозначим $\{\xi^k\}$. Пусть на M^n задана аффинная связность $\nabla = \{\nabla_k\}$. Определим ковариантную производную тензорного поля $T = \left\{T^{(\alpha)}_{(\beta)}\right\}$ вдоль векторного поля $\dot\gamma$ формулой: $\nabla_{\dot\gamma}(T) = \left\{\nabla_{\dot\gamma} T^{(\alpha)}_{(\beta)}\right\}$. Условно будем писать: $\nabla_{\dot\gamma} = \xi^k \nabla_k$. Эту операцию назовем *ковариантным дифференцированием вдоль кривой.*

Определение 5.11. Пусть $\mathbf{T} = \{T^i\}$ — гладкое векторное поле вдоль кривой $\gamma(t)$. Назовем его *параллельным вдоль* $\gamma(t)$ *относительно связности* ∇, если $\nabla_{\dot\gamma}(\mathbf{T}) \equiv 0$.

Другими словами, поле $\mathbf{T}$, "параллельное вдоль $\gamma(t)$", имеет ковариантно постоянные компоненты. В этом — аналогия с евклидовым случаем, поскольку общее определение параллельности превращается для $\{\nabla_k = \partial/\partial x^k\}$ в $M^n = \mathbb{R}^n$ в обычный параллельный перенос.

Запишем условие параллельности поля $\mathbf{T}$ в координатах. Имеем

$$\nabla_{\dot\gamma}(T^i) = \xi^k \nabla_k T^i = 0; \quad \xi^k = \frac{dx^k(t)}{dt},$$
$$\frac{dx^k}{dt}\frac{\partial T^i}{\partial x^k} + T^p \frac{dx^k}{dt}\Gamma^i_{pk} = \frac{dT^i}{dt} + T^p \Gamma^i_{pk}\frac{dx^k}{dt} = 0.$$

Определение 5.12. Уравнение $\dfrac{dT^i}{dt} + T^p\Gamma^i_{pk}\dfrac{dx^k}{dt} = 0$ называется *уравнением параллельного переноса вдоль кривой* $\gamma(t)$.

Изменив γ, мы меняем уравнение переноса.

Итак, пусть $\gamma(t)$ — кривая, идущая из P в Q, и $\mathbf{a} = (a^i) \in T_P M^n$ — вектор, заданный в P. Наша цель: построить в точке Q вектор $\mathbf{b} \in T_Q M^n$, который естественно назвать "параллельным вектору $\mathbf{a}$". Рассмотрим вдоль γ уравнение параллельного переноса. В нем функции Γ^i_{pk} и $\dfrac{dx^i(t)}{dt}$ известны. Ищем функции $\{T^i(t)\}$ — компоненты параллельного поля $\mathbf{T}(t)$. При этом нужно, чтобы в начальный момент выполнялось условие: $T^i(0) = a^i$. Как известно из курса обыкновенных дифференциальных уравнений, решение системы $\dfrac{dT^i}{dt} + T^p\Gamma^i_{pk}\dfrac{dx^k}{dt} = 0$ существует, единственно и продолжается вплоть до Q.

Определение 5.13. Вектор $\mathbf{b} = \mathbf{T}(1) \in T_Q M^n$, возникающий в Q, называется *параллельным* вектору $\mathbf{a} \in T_P M^n$ вдоль кривой $\gamma(t)$.

Ясно, что $\mathbf{b}$ зависит, вообще говоря, от γ. В случае $M^n = \mathbb{R}^n$ вектор $\mathbf{b}$ параллелен $\mathbf{a}$ в обычном смысле, если в качестве ∇ взять евклидову связность: $\Gamma^i_{jk} \equiv 0$.

Рассмотрим свойства параллельного переноса на римановом M^n в римановой связности ∇.

Теорема 5.4. *Пусть* $\mathbf{a}, \mathbf{b} \in T_P M^n$ *и* $\gamma(t)$ *— гладкая кривая из* P *в* Q. *Перенесем векторы* $\mathbf{a}$ *и* $\mathbf{b}$ *параллельно вдоль* $\gamma(t)$. *Тогда скалярное произведение векторов сохраняется, т.е. если* $\mathbf{a}(t)$ *и* $\mathbf{b}(t)$ *— параллельные поля вдоль* $\gamma(t)$, $\mathbf{a}(0) = \mathbf{a}$, $\mathbf{b}(0) = \mathbf{b}$, *то* $\dfrac{d}{dt}\big\langle \mathbf{a}(t), \mathbf{b}(t)\big\rangle \equiv 0$, *где* $\langle\ ,\ \rangle$ *— скалярное произведение в* $T_{\gamma(t)}M^n$, *порожденное* g_{ij}.

Доказательство. Включим $\mathbf{a}$ и $\mathbf{b}$ в параллельные поля: $\mathbf{T} = \mathbf{a}(t)$ и $\mathbf{R} = \mathbf{b}(t)$, $\mathbf{T}(0) = \mathbf{a}$, $\mathbf{R}(0) = \mathbf{b}$. Рассмотрим функцию $f(t) = \big\langle \mathbf{T}, \mathbf{R}\big\rangle_{\gamma(t)} = (g_{ij}T^iR^j)(t)$, т.е. скалярное произведение вдоль $\gamma(t)$. Дифференцируя, получаем

$$\frac{df(t)}{t} = \nabla_{\dot\gamma} f(t) = \xi^k \nabla_k (g_{ij}T^iR^j) =$$
$$= \xi^k g_{ij}\nabla_k(T^iR^j) = g_{ij}R^j(\xi^k\nabla_k T^i) + g_{ij}T^i(\xi^k\nabla_k R^j) \equiv 0,$$

так как $\nabla_{\dot\gamma}(\mathbf{T}) = \nabla_{\dot\gamma}(\mathbf{R}) = 0$. Теорема доказана. ■

Верно и обратное: *если на римановом* M^n *дана симметричная аффинная связность, в которой параллельный перенос вдоль любой кривой сохраняет скалярное произведение, то эта связность — риманова.* В самом деле, обращаясь к доказательству теоремы 5.4, получаем: $\xi^k T^i R^j(\nabla_k g_{ij}) \equiv 0$, т.е. $\nabla_k g_{ij} \equiv 0$.

Хотя мы рассматривали гладкие кривые, однако легко определить перенос и вдоль кусочно-гладких кривых. Пусть на $\gamma(t)$ есть изолированная точка излома (рис. 5.11). Перенося вдоль $\gamma(t)$ вектор $\mathbf{a}$, сначала получим в A вектор $\mathbf{b}$, параллельный $\mathbf{a}$. Примем $\mathbf{b}$ за начальный вектор нового параллельного поля уже вдоль участка AB и повторим процесс.

Рис. 5.11:

5.4.3 Геодезические

Определение 5.14. Пусть M^n снабжено аффинной связностью (наличие метрики не предполагается). Гладкая кривая $\gamma(t)$ называется *геодезической в данной связности* ∇, если $\nabla_{\dot\gamma}(\dot\gamma) = 0$, где $\dot\gamma$ — векторное поле скорости траектории $\gamma(t)$.

Иными словами, геодезическая — это траектория, поле скоростей которой параллельно вдоль нее, т.е. при параллельном переносе ее вектор скорости переходит в вектор скорости. Выведем уравнение геодезических.

Имеем:

$$0 = \nabla_{\dot\gamma}(\dot\gamma) \quad \text{и поскольку} \quad T^i = \frac{dx^i}{dt}, \quad \text{то}$$

$$\frac{dT^i}{dt} + T^\alpha T^k \Gamma^i_{\alpha k} = 0, \quad \text{или} \quad \frac{d^2x^i}{dt^2} + \Gamma^i_{\alpha k}\frac{dx^\alpha}{dt}\frac{dx^k}{dt} = 0.$$

Мы получили *уравнения геодезических*. Их решения — это функции $x^1(t), \ldots, x^n(t)$, задающие геодезическую $\gamma(t)$. Это система n уравнений второго порядка в обыкновенных производных. Ее решение однозначно задается начальными данными: $x^i(0) = P^i$ и $\dfrac{dx^i(0)}{dt} = a^i$, $\mathbf{a} \in T_P M^n$. Итого — $2n$ постоянных: n из них определяют точку P, а другие n задают вектор скорости в этой точке. Из теории обыкновенных дифференциальных уравнений сразу следует

Предложение 5.4. *Пусть $P \in M^n$ и $\mathbf{a} \in T_P M^n$. Тогда существует и единственна геодезическая $\gamma(t)$ такая, что $\gamma(0) = P$, $\dot\gamma(0) = \mathbf{a}$.*

Следствие 5.1. *Две геодезические, коснувшиеся друг друга в некоторой точке, совпадают.*

Рассмотрим теперь риманово M^n и риманову связность ∇. Пусть γ — геодезическая, $\dot\gamma$ — поле скоростей, $\mathbf{T}$ — векторное поле, параллельное вдоль γ. Тогда в каждой точке $\gamma(t)$ возникает число: $\cos\alpha(t) = \dfrac{\langle \mathbf{T}, \dot\gamma\rangle}{|\mathbf{T}|\cdot|\dot\gamma|}$, где $\alpha(t)$ — угол между векторами $\mathbf{T}$ и $\dot\gamma$.

Лемма 5.10. *При параллельном переносе вектора $\mathbf{T}$ вдоль геодезической γ угол α сохраняется*: $\alpha(t) \equiv \mathrm{const}$.

Доказательство. В силу теоремы 5.4 сохраняются скалярные произведения и длины векторов. ■

Рис. 5.12:

В двумерном случае сохранение угла однозначно задает параллельный перенос вдоль геодезической. Пусть γ — геодезическая и $\mathbf{a} \in T_P M^2$ (рис. 5.12). Тогда параллельное поле $\mathbf{T}(t)$ такое, что $\mathbf{T}(0) = \mathbf{a}$, образовано векторами, имеющими ту же длину, что и $\mathbf{a}$ и образующими с вектором $\dot\gamma(t)$ один и тот же угол α, равный углу между $\mathbf{a}$ и $\dot\gamma(0)$. Теперь можно определить параллельный перенос вдоль любой кусочно-гладкой кривой γ. Аппроксимируем γ ломаной геодезической, составленной из малых геодезических отрезков. Затем перенесем вектор вдоль каждого из них, используя постоянство угла. Потом перейдем к пределу, устремляя к нулю длину геодезических отрезков. При параллельном переносе вдоль траектории, не являющейся геодезической, угол,

Рис. 5.13:

Рис. 5.14:

образуемый переносимым вектором с ее вектором скорости, будет, вообще говоря, меняться. Рассмотрим примеры.

1) Если $M^n = \mathbb{R}^n$, то параллельный перенос вдоль кривой совпадает с обычным: параллельное поле имеет постоянные компоненты относительно декартовых координат.

2) Пусть M^2 — прямой круговой конус в $\mathbb{R}^3$ с углом при вершине, равным θ (рис. 5.13). Мы считаем, что вершина "выколота". Пусть $\gamma(t)$ — сечение конуса плоскостью, ортогональной оси. Пусть $OP = r$, $\mathbf{a} \in T_P M^2$ направлен в вершину конуса. Перенесем этот вектор параллельно вдоль $\gamma(t)$ и вернемся в точку P. На какой угол повернется вектор? Используем связность, согласованную с индуцированной метрикой на конусе. Так как она евклидова, то конус можно развернуть на плоскость, разрезав его вдоль образующей. Достаточно найти поворот $\mathbf{a}$ при его параллельном переносе на $\mathbb{R}^2$ вдоль дуги окружности (рис. 5.14). Легко подсчитать, что угол поворота φ равен $2\pi\big(1 - \sin(\theta/2)\big)$.

Укажем геометрический смысл символов (коэффициентов) Кристоффеля с точки зрения параллельного переноса. Найдем $\nabla_{\partial_\alpha}(\partial_\beta)$, где ∂_α и ∂_β — координатные векторные поля координат $x^1, \ldots, x^n$ на M^n. Тензор $\nabla_{\partial_\alpha}(\partial_\beta)$ снова — векторное поле.

Лемма 5.11. *Имеет место тождество* $\nabla_{\partial_\alpha}(\partial_\beta) = \Gamma^k_{\beta\alpha}\partial_k$.

Доказательство. Имеем

$$\left[\nabla_{\partial_\alpha}(\partial_\beta)\right]^k = a^q_\alpha \nabla_q(T^k_\beta),$$

где

$$\partial_\alpha = \{a^q_\alpha = \delta^q_\alpha\}; \qquad \partial_\beta = \{T^k_\beta = \delta^k_\beta\};$$
$$a^q_\alpha \nabla_q(T^k_\beta) = a^q_\alpha \left(\frac{\partial}{\partial x^q} T^k_\beta + T^\omega_\beta \Gamma^k_{\omega q}\right) = \delta^q_\alpha \delta^\omega_\beta \Gamma^k_{\omega q} = \Gamma^k_{\beta\alpha};$$

т.е.

$$\nabla_{\partial_\alpha}(\partial_\beta) = \Gamma^k_{\beta\alpha}\partial_k.$$

Лемма доказана. ■

Это утверждение можно понимать так. Сделаем бесконечно малый перенос ∂_β вектора в направлении ∂_α. При этом ∂_β как-то "повернется", и коэффициенты его разложения по $\partial_1, \ldots, \partial_n$ как раз и равны $\Gamma^k_{\beta\alpha}$.

3) Пусть $M^n = \mathbb{R}^n$ отнесено к декартовым координатам $x^1, \ldots, x^n$. Тогда $\Gamma^i_{jk} \equiv 0$. Уравнения геодезических имеют вид

$$\frac{d^2x^\alpha}{dt^2} \equiv 0, \quad \text{т.е.} \quad x^\alpha = a^\alpha t + b^\alpha.$$

Итак, геодезические — прямые линии и только они.

4) Пусть $M^2 = S^2$ со стандартной метрикой $ds^2 = d\theta^2 + \sin^2\theta\, d\varphi^2$ в сферических координатах. Найдем геодезические. Прямой подсчет дает (проверьте!): $\Gamma^1_{22} = -\sin(2\theta)/2$; $\Gamma^2_{12} = \Gamma^2_{21} = \operatorname{ctg}\theta$; $\Gamma^i_{jk} = 0$ для всех остальных наборов индексов (i, j, k). Здесь $x^1 = \theta$, $x^2 = \varphi$; $g_{11} = 1$, $g_{12} = g_{21} = 0$, $g_{22} = \sin^2\theta$. Отсюда, уравнения геодезических имеют вид: $\dfrac{d^2\theta}{dt^2} - \dfrac{1}{2}\sin 2\theta \left(\dfrac{d\varphi}{dt}\right)^2 = 0$, $\dfrac{d^2\varphi}{dt^2} + \operatorname{ctg}\theta \dfrac{d\varphi}{dt}\dfrac{d\theta}{dt} = 0$. Одно из решений имеет вид: $\varphi = \text{const}$, $\theta = t$. Это — меридиан, выходящий из северного полюса, отнесенный к параметру $\theta = t$ — длине дуги.

Предложение 5.5. *Пусть S^2 снабжена стандартной метрикой. Тогда геодезическими римановой связности являются все центральные плоские сечения сферы (экваторы) и только они.*

Доказательство. Сначала докажем, что любое центральное плоское сечение S^2 — геодезическая (относительно натурального параметра). Для одного меридиана γ_0 это уже доказано. Рассмотрим произвольный "экватор" γ. Очевидно существует вращение, переводящее γ в γ_0.

Лемма 5.12. *Пусть $f\colon M^n \to M^n$ — изометрия риманова многообразия M^n и γ — геодезическая римановой связности. Тогда образ γ при изометрии f — также является геодезической.*

Рис. 5.15:

Доказательство. Изометрия сохраняет риманову связность, а потому сохраняет и уравнения геодезических, т.е. f переводит решение системы снова в решение системы, что и требовалось. ■

Возвращаясь к доказательству предложения 5.5, получаем, что γ — геодезическая. Обратно: пусть γ — геодезическая на S^2. Рассмотрим в произвольной точке γ ее вектор скорости $\dot\gamma$ и проведем через эту точку в направлении вектора экватор (через любую точку S^2 в любом направлении проходит один и только один экватор) (рис. 5.15). Так как γ и экватор — решения одной и той же системы, то, касаясь, они совпадают. Предложение доказано. ■

5). Пусть $M^2 = \mathrm{L}_2$ — плоскость Лобачевского со стандартной метрикой $ds^2 = d\chi^2 + \mathrm{sh}^2\,\chi\, d\varphi^2$. Напомним, что в модели Пуанкаре метрика имеет вид $4(1-r^2)^{-2}(dr^2 + r^2\, d\varphi^2)$. Найдем геодезические.

Предложение 5.6. *Геодезическими плоскости Лобачевского в модели Пуанкаре являются все дуги окружностей, выходящих на абсолют под прямым углом (в частности, все диаметры), и только они.*

Так как метрика L_2 получается из метрики S^2 заменой тригонометрических функций на гиперболические, то: $\Gamma^1_{22} = -\mathrm{sh}(2\chi)/2$; $\Gamma^2_{12} = \Gamma^2_{21} = \mathrm{cth}\,\chi$; $\Gamma^i_{jk} = 0$ для всех остальных индексов (i,j,k); здесь $x^1 = \chi$, $x^2 = \varphi$. Отсюда $\dfrac{d^2\chi}{dt^2} - \dfrac{1}{2}\mathrm{sh}(2\chi)\cdot\left(\dfrac{d\varphi}{dt}\right)^2 = 0$, $\dfrac{d^2\varphi}{dt^2} + \mathrm{cth}\,\chi\cdot\dfrac{d\varphi}{dt}\dfrac{d\chi}{dt} = 0$. Одно из решений имеет вид: $\varphi = \mathrm{const}$, $\chi = t$ — прямая, проходящая через точку 0 на плоскости. Так как (χ,φ) "обслуживают" в то же время правую полость гиперболоида — псевдосферы мнимого радиуса, то при стереографической проекции "прямая" с уравнением $\varphi = \varphi_0$, $\chi = t$ перейдет в диаметр единичного круга. Мы доказали, что один из диаметров γ_0 на модели Пуанкаре — геодезическая.

Докажем, что любая дуга окружности, ортогональная абсолюту — геодезическая. Используем лемму 5.12. Надо доказать, что любая такая окружность переводится изометрией в γ_0. Перейдем к модели на верхней полуплоскости с метрикой $(dx^2 + dy^2)/y^2$. Напомним, что существует изометрия

Рис. 5.16:

Рис. 5.17:

(дробно-линейное преобразование), переводящая единичный круг в полуплоскость. При этом граница круга переходит в вещественную ось, а диаметр γ_0 — в прямую, ортогональную вещественной оси (рис. 5.16). Можно считать, что ось OY на полуплоскости является геодезической, поскольку она — образ диаметра γ_0 при изометрии. Тогда любая прямая, ортогональная вещественной оси, — геодезическая, так как сдвиг $z \to z + \alpha$, $\alpha \in \mathbb{R}$, — изометрия. Отсюда следует, что любая дуга окружности, выходящая под прямым углом на ось OX, — геодезическая, так как переводится в ось OY дробно-линейным преобразованием: сначала сдвиг $z \to z + \alpha$, $\alpha \in \mathbb{R}$, затем $z \to -z/(z-a)$, $a \in \mathbb{R}$. Доказано, что все прямые, ортогональные вещественной оси, и все дуги окружностей, встречающих ось OX под прямым углом, — геодезические. Пусть теперь γ — произвольная геодезическая. Докажем, что она совпадает с уже описанными геодезическими. Возьмем на γ точку P и рассмотрим вектор $\dot{\gamma}$. Проведем через P дугу окружности, встречающую абсолют под прямым углом и имеющую тот же вектор скорости (рис. 5.17). Как мы уже знаем, γ совпадает с этой дугой. Доказательство завершено.

Покажем, как параллельно переносится вектор на плоскости Лобачевского вдоль прямой, параллельной оси OX. Это — не геодезическая. Аппроксимируем γ ломаной геодезической (рис. 5.18). Качественная картина параллельного переноса $\mathbf{a}$ показана на рис. 5.19. Переносимый вектор вращается вокруг своего начала.

Рис. 5.18:

Рис. 5.19:

Замечание. Геодезические на плоскости Лобачевского можно найти, проинтегрировав уравнения геодезических. Это удобно выполнить на верхней полуплоскости. Символы Кристоффеля имеют вид (проверьте!): $\Gamma^1_{12} = \dfrac{1}{y}$; $\Gamma^2_{11} = -\dfrac{1}{y}$; $\Gamma^2_{22} = \dfrac{1}{y}$; остальные равны нулю. Уравнения геодезических $\ddot{x} = \dfrac{2\dot{x}\dot{y}}{y}$, $\ddot{y} = \dfrac{\dot{y}^2 - \dot{x}^2}{y}$. Отсюда

$$\frac{d^2y}{dx^2} = \frac{\ddot{y}\dot{x} - \ddot{x}\dot{y}}{\dot{x}^3} = -\frac{1}{y}({y'_x}^2 + 1);$$
$$yy'' + {y'}^2 = -1; \quad (yy')' = -1;$$
$$yy' = -x + C; \quad \frac{y^2}{2} = -\frac{x^2}{2} + Cx + \frac{D}{2};$$
$$(x - C)^2 + y^2 = C^2 + D.$$

Если $\dot{x} = 0$, то получаем прямые, ортогональные вещественной оси. Если $\dot{x} \neq 0$, то получаем окружности, ортогональные абсолюту.

Задача. Перенести параллельно вектор **a** вдоль плоского (не центрального) сечения S^2 (рис. 5.20).

Рис. 5.20:

6) Пусть $M^2 = T^2$ — двумерный тор с локально евклидовой метрикой. Рассмотрим на окружностях $S^1(\varphi)$ и $S^1(\psi)$ углы φ и ψ. Они будут координатами на торе, в которых метрика примет вид: $d\varphi^2 + d\psi^2$. Ее можно представить как индуцированную евклидовой метрикой из $\mathbb{R}^4$ при вложении T^2 в $\mathbb{R}^4 \cong \mathbb{C}^2$ по формуле: $f(\varphi, \psi) \to (e^{i\varphi}, e^{i\psi})$. Так как $ds^2(\mathbb{R}^4) = dzd\bar{z} + dwd\bar{w}$, то $ds^2(T^2) = d\varphi^2 + d\psi^2$.

Поскольку в евклидовой метрике $\Gamma^i_{jk} = 0$, то геодезические на торе — это образы прямых плоскости $\mathbb{R}^2(\varphi, \psi)$ при ее факторизации $h\colon \mathbb{R}^2(\varphi, \psi) \to T^2(\varphi, \psi)$, $h(\varphi, \psi) = (\varphi, \psi) \pmod{2\pi}$, т.е. $T^2 = \mathbb{R}^2/\mathbb{Z} \oplus \mathbb{Z}$ (рис. 5.21). Других геодезических на торе нет. Геодезические разбиваются на два класса: замкнутые и незамкнутые. Их удобно изображать прямыми на $\mathbb{R}^2(\varphi, \psi)$ с отмеченной решеткой $\mathbb{Z} \oplus \mathbb{Z} = (2\pi m, 2\pi n)$, $m, n \in \mathbb{Z}$. Рассмотрим прямые, выходящие из точки 0. Ясно, что соответствующая геодезическая на торе замкнута тогда и только тогда, когда прямая встречает какую-либо "целую" точку $(2\pi m, 2\pi n)$. Напротив, геодезическая на T^2 незамкнута (гомеоморфна прямой) тогда и только тогда, когда соответствующая прямая не содержит "целых" точек, кроме точки 0. Выразим это в терминах тангенса угла наклона прямой к оси OX. Геодезическая замкнута тогда и только тогда, когда $\operatorname{tg}\alpha$ (рис. 5.22) — рационален. Геодезическая незамкнута, если $\operatorname{tg}\alpha$ иррационален. На рис. 5.22 показана прямая, проходящая через точку $(2\pi \cdot 3,\, 2\pi \cdot 2)$. После факторизации мы получим на квадрате $0 \leqslant \varphi \leqslant 2\pi$, $0 \leqslant \varphi \leqslant 2\pi$ образ этой прямой (рис. 5.23). На рис. 5.23 показана также соответствующая геодезическая на торе (окружность). Если тангенс иррационален, то геодезическая порождает "иррациональную обмотку" тора, т.е. ее замыканием является весь тор.

Следующую теорему мы сообщим без доказательства.

Лемма 5.13. *Для каждой точки P_0 на римановом многообразии M^n существует окрестность U и число $\varepsilon > 0$ такие, что:*

1) *каждые две точки из U соединяет одна и только одна геодезическая длины меньше ε;*

2) *эта геодезическая гладко зависит от своих концов и целиком лежит в этой окрестности.*

Рис. 5.21:

Рис. 5.22:

Рис. 5.23:

Отсюда несложно вывести следующий факт.

Теорема 5.5. *Любое двумерное гладкое, компактное, связное, замкнутое многообразие триангулируемо.*

Доказательство. Снабдим M^2 римановой метрикой (например, вложим M^2 в евклидово пространство). Так как M^2 — компактно и замкнуто, то его можно покрыть конечным числом малых дисков. В силу леммы 5.13 можно считать, что в каждом из них любая пара точек соединяется единственной геодезической длины, не превосходящей ε. Покрыв M^2 достаточно плотной сетью точек $\{P_i\}$, соединим геодезическими точки, попавшие внутрь какого-либо диска покрытия, и в результате разобьем диски на треугольники, удовлетворяющие требованиям триангуляции. Теорема доказана. ■

5.5 Дифференцирование и интегрирование внешних дифференциальных форм. Формула Стокса

5.5.1 Дифференцирование внешних дифференциальных форм

Для внешних дифференциальных форм дифференциальное исчисление несколько упрощается по сравнению с дифференциальным исчислением произвольных тензорных полей. Если на многообразии фиксирована некоторая связность, то ковариантный градиент кососимметрического тензорного поля не является, вообще говоря, кососимметрическим тензорным полем. Поэтому естественно определить градиент внешней дифференциальной формы как композицию ковариантного градиента и альтернирования по всем индексам тензорного поля. При этом, если связность симметрична, то в формуле для градиента внешней дифференциальной формы вообще не участвуют символы Кристоффеля, т.е. определение градиента внешней дифференциальной формы не зависит от выбора симметрической связности на многообразии M.

Определение 5.15. *Градиентом внешней дифференциальной формы* ω называется внешняя дифференциальная форма $d\omega$, компоненты которой в локальной системе координат $(x^1, \ldots, x^n)$ имеют вид

$$(d\omega)_{j_1 \ldots j_{k+1}} = \sum_{s=1}^{k+1} (-1)^{s+1} \frac{\partial}{\partial x^{j_s}} \omega_{j_1 \ldots j_{s-1} j_{s+1} \ldots j_{k+1}}. \tag{5.1}$$

Теорема 5.6. *Градиент внешних дифференциальных форм удовлетворяет следующим условиям*:

1) $d(\omega_1 \wedge \omega_2) = d\omega_1 \wedge \omega_2 + (-1)^{\deg \omega_1} \omega_1 \wedge d\omega_2$,

2) $d(d\omega) = 0$.

Доказательство. Пусть $(x^1, \ldots, x^n)$ — локальная система координат, ω_1, ω_2 — две внешних дифференциальных формы степеней p и q соответственно. Компоненты произвольной дифференциальной формы ω отличны от нуля только для таких наборов индексов, у которых индексы принимают попарно различные значения. Значит для определения компонент формы ω достаточно рассматривать только такие наборы индексов, значения которых попарно различны и упорядочены по возрастанию. Такие наборы индексов соответствуют подмножествам $I \subset [1, n]$ в отрезке всех целых чисел от 1 до n. Соответствующую компоненту формы ω будем обозначать через ω_I. При таких обозначениях формулы умножения и дифференцирования дифференциальных форм примут более простой вид:

$$(\omega_1 \wedge \omega_2)_K = \sum_{I \cup J = K} (-1)^{\sigma(I,J)} \omega_{1,I} \omega_{2,J}, \tag{5.2}$$

$$(d\omega)_K = \sum_{\{i\} \cup J = K} (-1)^{\sigma(i,J)} \frac{\partial \omega_J}{\partial x^i}, \tag{5.3}$$

где суммирование производится по всем подходящим разбиениям множества K на два подмножества $I \cup J = K$, состоящим из, соответственно, p и q элементов, а $\sigma(I, J)$ — такая перестановка, которая переставляет индексы из множества I на первые p мест с сохранением порядка. Во второй формуле смысл всех обозначений аналогичен с тем изменением, что множество I состоит из одного индекса i.

Применим формулу (5.3) к (5.2), получим

$$\big(d(\omega_1 \wedge \omega_2)\big)_K = \sum_{\{i\} \cup I \cup J = K} (-1)^{\sigma(i,I,J)} \frac{\partial}{\partial x^i} (\omega_{1,I} \cdot \omega_{2,J}).$$

В последней сумме суммирование производится по всевозможным разбиениям множества K на три непересекающихся подмножества $\{i\}$,I,J, а $\sigma(i, I, J)$ — перестановка, которая переставляет индекс i на первое место, а далее идут множества I и J с сохранением порядка внутри множеств I и J.

Таким образом,

$$\begin{aligned}\big(d(\omega_1 \wedge \omega_2)\big)_K = &\sum_{\{i\} \cup I \cup J = K} (-1)^{\sigma(i,I,J)} \frac{\partial \omega_{1,I}}{\partial x^i} \omega_{2,J} + \\ &+ (-1)^{\deg \omega_1} \sum_{\{i\} \cup I \cup J = K} (-1)^{\sigma(I,i,J)} \omega_{1,I} \frac{\partial \omega_{2,J}}{\partial x^i} = \\ &= (d\omega_1 \wedge \omega_2)_K + (-1)^{\deg \omega_1} (\omega_1 \wedge d\omega_2)_K.\end{aligned}$$

Свойство 2) достаточно доказать для базисных форм, по которым разлагается любая форма. Без ограничения общности, можно считать, что в локальной системе координат $(x^1, \ldots, x^n)$ форма ω имеет вид: $\omega = f(x^1, \ldots, x^n)dx^1 \wedge \ldots \wedge dx^k$. Заметим, что если свойство 2) выполнено для форм ω_1 и ω_2, то оно справедливо и для их внешнего произведения $\omega_1 \wedge \omega_2$. В самом деле,

$$\begin{aligned} dd(\omega_1 \wedge \omega_2) &= d\big(d\omega_1 \wedge \omega_2 + (-1)^{\deg \omega_1}\omega_1 \wedge d\omega_2\big) = \\ &= dd\omega_1 \wedge \omega_2 + (-1)^{\deg(d\omega_1)}d\omega_1 \wedge d\omega_2 + \\ &\quad + (-1)^{\deg \omega_1}d\omega_1 \wedge d\omega_2 + (-1)^{2\deg \omega_1}\omega_1 \wedge dd\omega_2. \end{aligned}$$

Первое и последнее слагаемые равны нулю по предположению, а два средних одинаковы с противоположными знаками. Таким образом, свойство 2) достаточно доказать для произвольной функции f и для формы dx^k. В последнем случае $d(dx^k) = 0$, поскольку x^k — тоже некоторая функция на многообразии. Итак, покажем, что $d(df) \equiv 0$. Имеем $(df)_i = \dfrac{\partial f}{\partial x^i}$, далее

$$\big(d(df)\big)_{ij} = \frac{\partial (df)_j}{\partial x^i} - \frac{\partial (df)_i}{\partial x^j} = \frac{\partial^2 f}{\partial x^j \partial x^i} - \frac{\partial^2 f}{\partial x^i \partial x^j} \equiv 0. \qquad \blacksquare$$

Отметим, что условие 2) теоремы 5.6 основывалось на замечательном свойстве функций многих переменных — независимости смешанных частных производных от порядка дифференцирования. Свойство 2) в некотором смысле является максимальным обобщением этого свойства на случай систем функций, образующих компоненты кососимметрического тензора.

Пример 1. В локальной системе координат всякая внешняя дифференциальная форма согласно определению имеет вид суммы $\omega = \sum\limits_{i_1 < \ldots < i_p} \omega_{i_1 \ldots i_p} dx^{i_1} \wedge \ldots \wedge dx^{i_p}$. Поэтому, по теореме 5.6 градиент формы ω вычисляется по формуле

$$d\omega = \sum_{i_1 < \ldots < i_p} d(\omega_{i_1 \ldots i_p}) \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}.$$

Далее, градиент гладкой функции df в локальной системе координат имеет компоненты $\left\{ \dfrac{\partial f}{\partial x^i} \right\}$. Следовательно, $df = \sum(\partial f/\partial x^i)dx^i$. Поэтому $d\omega = \sum\limits_{i_1 < \ldots < i_p} \sum\limits_{i=1}^{n} \dfrac{\partial \omega_{i_1 \ldots i_p}}{\partial x^i} dx^i \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}$. Последняя формула, очевидно, эквивалентна формуле (5.1). Рассмотрим задачу отыскания такой дифференциальной формы ω, чтобы выполнялось тождество

$$d\omega = \Omega, \tag{5.4}$$

где Ω — фиксированная дифференциальная форма. В каждой локальной системе координат уравнение (5.4) превращается в систему дифференциальных

уравнений в частных производных:

$$\sum_{s=1}^{k+1}(-1)^{s+1}\frac{\partial\omega_{i_1\ldots i_{s-1}i_{s+1}\ldots i_{k+1}}}{\partial x^{i_s}} = \Omega_{i_1\ldots i_{k+1}}.$$

По условию 2) теоремы 5.6 решения уравнения (5.4) существуют только тогда, когда $d\Omega = 0$. В самом деле, применив к левой и правой части (5.4) операцию d, получаем: $d\Omega = d(d\omega) = 0$. В частном случае, когда $\deg\Omega = 1$, задача сводится к отысканию такой гладкой функции f на многообразии M, градиент которой равен заданной форме Ω. В локальной системе координат $(x^1,\ldots,x^n)$ форма Ω имеет вид $\Omega = \sum\limits_{i=1}^{n}\Omega_i dx^i$, уравнение (5.4) эквивалентно системе

$$\frac{\partial f}{\partial x^i} = \Omega_i, \quad i = 1,\ldots,n. \tag{5.5}$$

Если решение системы (5.5) существует, то, продифференцировав (5.5) по переменным x^j, получаем: $\dfrac{\partial^2 f}{\partial x^i\partial x^j} = \dfrac{\partial\Omega_i}{\partial x^j}$. Поскольку смешанные частные производные гладкой функции f не зависят от порядка дифференцирования, то

$$\frac{\partial\Omega_i}{\partial x^j} = \frac{\partial\Omega_j}{\partial x^i}, \qquad 1 \leqslant i,j \leqslant n. \tag{5.6}$$

Условия (5.6) в бескоординатной записи эквивалентны условию $d\Omega = 0$. Покажем, что условия (5.6) достаточны для того, чтобы система (5.5) имела решение в достаточно малой окрестности любой точки P многообразия M. Для этого решим сначала одно из уравнений системы (5.5), скажем $\dfrac{\partial f}{\partial x^1} = \Omega_1$. Имеем

$$f(x^1,\ldots,x^n) = \int\limits_{x_0^1}^{x^1}\Omega_1(x^1,x^2,\ldots,x^n)\,dx^1 + \varphi_2(x^2,\ldots,x^n), \tag{5.7}$$

где $\varphi_2(x^2,\ldots,x^n)$ — произвольная гладкая функция от переменных $(x^2,\ldots,x^n)$. Подставим правую часть (5.7) во второе уравнение системы (5.5):

$$\frac{\partial f}{\partial x^2} = \int\limits_{x_0^1}^{x^1}\frac{\partial\Omega_1}{\partial x^2}(x^1,x^2,\ldots,x^n)\,dx^1 + \frac{\partial\varphi_2}{\partial x^2}(x^2,\ldots,x^n) = \Omega_2(x^1,\ldots,x^n).$$

Учитывая (5.6), получаем

$$\Omega_2(x^1,\ldots,x^n) = \int\limits_{x_0^1}^{x^1} \frac{\partial \Omega_2}{\partial x^1}(x^1,\ldots,x^n)\,dx^1 + \frac{\partial \varphi_2}{\partial x^2}(x^2,\ldots,x^n) =$$

$$= \Omega_2(x^1,\ldots,x^n) - \Omega_2(x_0^1,x^2,\ldots,x^n) + \frac{\partial \varphi_2}{\partial x^2}(x^2,\ldots,x^n),$$

т.е.

$$\frac{\partial \varphi_2}{\partial x^2}(x^2,\ldots,x^n) = \Omega_2(x_0^1,x^2,\ldots,x^n). \tag{5.8}$$

Общее решение уравнения (5.8) имеет вид

$$\varphi_2(x^2,\ldots,x^n) = \int\limits_{x_0^2}^{x^2} \Omega_2(x_0^1,x^2,\ldots,x^n)\,dx^2 + \varphi_3(x^3,\ldots,x^n). \tag{5.9}$$

Подставляя (5.7) и (5.9) в третье уравнение системы (5.5), получаем

$$\Omega_3(x^1,\ldots,x^n) = \int\limits_{x_0^1}^{x^1} \frac{\partial \Omega_1}{\partial x^3}(x^1,\ldots,x^n)\,dx^1 +$$

$$+ \int\limits_{x_0^2}^{x^2} \frac{\partial \Omega_2}{\partial x^3}(x_0^1,x^2,\ldots,x^n)\,dx^2 + \frac{\partial \varphi_3}{\partial x^3}(x^3,\ldots,x^n),$$

Снова, учитывая условие (5.6), получаем уравнение на функцию φ_3:

$$\Omega_3(x^1,\ldots,x^n) = \int\limits_{x_0^1}^{x^1} \frac{\partial \Omega_3}{\partial x^1}(x^1,\ldots,x^n)\,dx^1 +$$

$$+ \int\limits_{x_0^2}^{x^2} \frac{\partial \Omega_3}{\partial x^2}(x_0^1,x^2,\ldots,x^n)\,dx^2 + \frac{\partial \varphi_3}{\partial x^3}(x^3,\ldots,x^n) =$$

$$= \Omega_3(x^1,\ldots,x^n) - \Omega_3(x_0^1,x^2,\ldots,x^n) + \Omega_3(x_0^1,x^2,\ldots,x^n) -$$

$$- \Omega_3(x_0^1,x_0^2,x^3,\ldots,x^n) + \frac{\partial \varphi_3}{\partial x^3}(x^3,\ldots,x^n);$$

$$\frac{\partial \varphi_3}{\partial x^3}(x^3,\ldots,x^n) = \Omega_3(x_0^1,x_0^2,x^3,\ldots,x^n).$$

Продолжим этот процесс; получим последовательность функций $\varphi_k(x^k, \ldots, x^n)$, определяемых рекуррентно формулами

$$\varphi_k(x^k, \ldots, x^n) = \int\limits_{x_0^k}^{x^k} \Omega_k(x_0^1, \ldots, x_0^{k-1}, x^k, \ldots, x^n)\, dx^k +$$
$$+ \varphi_{k+1}(x^{k+1}, \ldots, x^n), \quad \varphi_{n+1} \equiv \mathrm{const},$$
$$f(x^1, \ldots, x^n) = \sum_{k=1}^{n} \int\limits_{x_0^k}^{x^k} \Omega_k(x_0^1, \ldots, x_0^{k-1}, x^k \ldots, x^n)\, dx^k + \varphi_{n+1}.$$

Последняя формула дает нам произвольное решение системы (5.5), зависящее от одного числового параметра $\varphi_{n+1} = \mathrm{const}$. Впрочем, неединственность решения системы (5.5) следует из общих алгебраических свойств внешних дифференциальных форм. В самом деле, если f_1 и f_2 — два решения системы (5.5), то $df_1 = \Omega$, $df_2 = \Omega$, т.е. $d(f_1 - f_2) = \Omega - \Omega = 0$. Функция $h = f_1 - f_2$ обладает тем свойством, что ее градиент тождественно равен нулю на многообразии M. Следовательно, h — локально постоянная функция. Если M — связное многообразие, то тогда h — постоянная функция. Таким образом, *любые два решения системы* (5.5) *отличаются на константу, а множество всех решений системы* (5.5) *или пусто или является одномерным многообразием в пространстве всех гладких функций на многообразии* M.

Переходя к общему случаю решений системы (5.4), мы можем сделать следующие заключения:

а) *необходимым условием существования решения системы* (5.4) *является условие* $d\Omega = 0$;

б) *если* ω *— решение системы* (5.4), *то всякая форма вида* $\omega + d\omega'$ *тоже является решением системы* (5.4).

5.5.2 Поведение форм при отображениях

Каждое гладкое отображение многообразий индуцирует отображение внешних дифференциальных форм, направленное в обратную сторону: $f\colon M_1 \to M_2$, $f^*\colon \Lambda^k(M_2) \to \Lambda^k(M_1)$.

Определение 5.16. Пусть $f\colon M_1 \to M_2$ — гладкое отображение многообразий, $\omega \in \Lambda^k(M_2)$ — внешняя дифференциальная форма степени k на многообразии M_2. *Обратным образом* $f^*(\omega)$ *формы* ω называется внешняя дифференциальная форма на многообразии M_1, задаваемая следующей формулой: $f^*(\omega)(\xi_1, \ldots, \xi_k) = \omega\big(df(\xi_1), \ldots, df(\xi_k)\big)$, где $\xi_1, \ldots, \xi_k \in T_P M_1$ — касательные векторы в точке P многообразия M_1, а $df(\xi_1), \ldots, df(\xi_k) \in T_{f(P)} M_2$ — их образы при дифференциале отображения f.

Так определенное отображение $f^*: \Lambda^k(M_2) \to \Lambda^k(M_1)$ пространств дифференциальных форм является, очевидно, линейным отображением линейных пространств.

Лемма 5.14. *Если в локальных координатах отображение f записывается системой функций $y^i = f^i(x^1, \ldots, x^n)$, а форма ω записывается в виде $\omega = \sum\limits_{i_1<\ldots<i_k} \omega_{i_1\ldots i_k}(y^1, \ldots, y^m) dy^{i_1} \wedge \ldots \wedge dy^{i_k}$, то*

$$f^*(\omega) = \sum_{i_1<\ldots<i_k} \omega_{i_1\ldots i_k}(f^1, \ldots, f^m) df^{i_1} \wedge \ldots \wedge df^{i_k}. \tag{5.10}$$

Соотношение (5.10) является обобщением так называемого свойства инвариантности дифференциала функции, гласящего, что вид дифференциала функции $y = f(x)$ не зависит от того, является переменная x независимой или, в свою очередь, функцией от другой переменной: $dy = \dfrac{df}{dx}(x)dx$. Доказательство леммы основано на непосредственной проверке определений.

Теорема 5.7. *Обратный образ дифференциальных форм удовлетворяет следующим соотношениям*:

1) если $f: M_1 \to M_2$ и $g: M_2 \to M_3$, то

$$(gf)^* = f^* g^*,$$

2) отображение f^* коммутирует с оператором d, т.е.

$$f^*\, d = d\, f^*.$$

Доказательство теоремы основано на проверке соотношений в произвольной локальной системе координат и оставляется читателю в качестве упражнения. ■

5.5.3 Интеграл дифференциальной формы по многообразию

Рассмотрим произвольное подмножество X в гладком многообразии M. Скажем, что множество $X \subset M$ имеет *меру нуль*, $\mu(X) = 0$, если для любой карты U_α и координатного гомеоморфизма $\varphi: U_\alpha \to V_\alpha \subset \mathbb{R}^n_+$ множество $\varphi(X \cap U_\alpha) \subset \mathbb{R}^n_+$ имеет меру нуль в евклидовом пространстве $\mathbb{R}^n$.

Определение множества меры нуль корректно и не зависит от выбора атласа карт.

Для этого достаточно показать, что если $f: V_1 \to V_2$ гладкий гомеоморфизм области $V_1 \subset \mathbb{R}^n$ на область $V_2 \subset \mathbb{R}^n$, то образ $f(X)$ любого множества $X \subset V_1$ меры нуль является множеством меры нуль.

Множества меры нуль на многообразии удовлетворяют стандартным свойствам, справедливым для множеств евклидового пространства:

а) $\mu\Big(\bigcup_{i=1}^{\infty} X_i\Big) \leqslant \sum\limits_{i=1}^{\infty} \mu(X_i) = 0$,

б) образ $f(Y)$ гладкого $(n-1)$-мерного многообразия Y при гладком отображении $f\colon Y \to M$ имеет меру нуль.

Определим теперь интеграл от внешней дифференциальной формы ω степени n на n-мерном ориентированном многообразии M в случае, когда форма ω имеет компактный носитель. Пусть носитель формы ω целиком лежит в одной карте карте U_α с координатами $(x_\alpha^1, \ldots, x_\alpha^n)$. Тогда форма ω в карте U_α имеет вид $\omega = f_\alpha(x_\alpha^1, \ldots, x_\alpha^n) dx_\alpha^1 \wedge \ldots \wedge dx_\alpha^n$. Положим

$$\int\limits_M \omega = \int\limits_{U_\alpha} \cdots \int f_\alpha(x_\alpha^1, \ldots, x_\alpha^n) dx_\alpha^1 \ldots dx_\alpha^n. \tag{5.11}$$

Правая часть формулы (5.11) существует и не зависит от выбора локальной системы координат $(x_\alpha^1, \ldots, x_\alpha^n)$ с сохранением ориентации. Для этого следует применить формулу замены переменных в кратном интеграле и учесть положительность якобиана замены переменных.

В общем случае произвольного компактного носителя формы ω интеграл от формы дается следующим образом,

Определение 5.17. Пусть M — n-мерное гладкое ориентированное многообразие с краем, ω — внешняя дифференциальная форма степени n с компактным носителем, $\{\varphi_\alpha\}$ — разбиение единицы, подчиненное атласу карт $\{U_\alpha\}$. Тогда полагаем

$$\int\limits_M \omega = \sum_\alpha \int\limits_{U_\alpha} \varphi_\alpha \omega.$$

Определение 5.17 корректно, т.е. не зависит от выбора разбиения единицы. Оставляем доказательство корректности читателю в качестве упражнения.

Предложение 5.7. *1)* $\int\limits_M (\lambda_1\omega_1 + \lambda_2\omega_2) = \lambda_1 \int\limits_M \omega_1 + \lambda_2 \int\limits_M \omega_2, \quad \lambda_1, \lambda_2 \in \mathbb{R}^1.$

2) Если M' — противоположная к M ориентация, то $\int\limits_{M'} \omega = -\int\limits_M \omega$.

Доказательство. Свойство 1) очевидно. Для доказательства свойства 2) заметим, противоположную ориентацию на многообразии M можно задать не меняя атласа карт, а изменив только знак у первой координаты в каждом атласе U_α. Тогда нетрудно убедиться, что в определении интеграла от формы тоже поменяется знак в каждом слагаемом. ■

5.5.4 Формула Стокса

Следующая теорема представляет формулу, которая обобщает многочисленные формулы интегрирования из математического анализа.

Рис. 5.24:

Теорема 5.8. *Пусть M — n-мерное ориентированное многообразие с краем ∂M, ω — внешняя дифференциальная форма степени $(n-1)$ с компактным носителем. Тогда*

$$(-1)^n \int\limits_M d\omega = \int\limits_{\partial M} \omega. \tag{5.12}$$

Формула (5.12) носит название *общей формулы Стокса*. Прежде чем доказывать теорему 5.8, рассмотрим некоторые частные случаи этой формулы.

Пример 2. Рассмотрим в плоскости $\mathbb{R}^2$ гладкую замкнутую кривую Γ без самопересечений, ограничивающую открытую область V в $\mathbb{R}^2$. Будем считать, что на кривой Γ введен параметр t, задающий направление обхода и, следовательно, ориентацию Γ как одномерного многообразия. Тогда замыкание $\overline{V}$ является ориентированным двумерным многообразием с краем $\partial V = \Gamma$. Если ориентация области $\overline{V}$ задается линейной системой координат (x^1, x^2), то ориентация на границе Γ окажется согласованной с ориентацией в области $\overline{V}$, если при обходе кривой Γ при увеличении параметра t область $\overline{V}$ остается слева (см. рис. 5.24). Пусть ω — произвольная форма степени 1 на плоскости $\mathbb{R}^2$. Форма ω в координатах (x^1, x^2) имеет вид $\omega = P(x^1, x^2)dx^1 + Q(x^1, x^2)dx^2$. Тогда интеграл от формы ω по кривой Γ совпадает с интегралом второго рода:

$$\int\limits_\Gamma \omega = \int\limits_\Gamma P dx^1 + Q dx^2 = \int\limits_{t_0}^{t_1} \Big(P\big(x^1(t), x^2(t)\big) \frac{dx^1}{dt} + Q\big(x^1(t), x^2(t)\big) \frac{dx^2}{dt} \Big) dt.$$

По формуле (5.12) имеем

$$\int\limits_\Gamma \omega = \int\limits_V d\omega = \int\limits_V (dP \wedge dx^1 + dQ \wedge dx^2) =$$

$$= \int\limits_V \left(\frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2} \right) dx^1 \wedge dx^2 = \iint\limits_V \left(\frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2} \right) dx^1 dx^2.$$

Таким образом, мы получили известную *формулу Грина* на плоскости:

$$\int\limits_{\Gamma}(Pdx^1+Qdx^2)=\iint\limits_{V}\left(\frac{\partial Q}{\partial x^1}-\frac{\partial P}{\partial x^2}\right)dx^1dx^2.$$

Пример 3. Аналогично, пусть Γ — гладкая замкнутая кривая без самопересечений в трехмерном пространстве $\mathbb{R}^3$, ограничивающая некоторую двумерную поверхность V. Тогда по формуле (5.12) получаем классическую *формулу Стокса*. Для этого рассмотрим форму ω степени 1 в трехмерном пространстве $\mathbb{R}^3$:

$$\omega=P(x^1,x^2,x^3)dx^1+Q(x^1,x^2,x^3)dx^2+R(x^1,x^2,x^3)dx^3.$$

Тогда интеграл второго рода по кривой Γ интерпретируется как интеграл от формы ω:

$$\int\limits_{\Gamma}\omega=\int\limits_{t_0}^{t_1}\Big(P(x^1,x^2,x^3)\frac{dx^1}{dt}+Q(x^1,x^2,x^3)\frac{dx^2}{dt}+R(x^1,x^2,x^3)\frac{dx^3}{dt}\Big)dt.$$

На основании формулы (5.12) имеем

$$\int\limits_{\Gamma}\omega=\int\limits_{V}d\omega;$$

$$d\omega=dP\wedge dx^1+dQ\wedge dx^2+dR\wedge dx^3=\Big(\frac{\partial Q}{\partial x^1}-\frac{\partial P}{\partial x^2}\Big)dx^1\wedge dx^2+$$
$$+\Big(\frac{\partial R}{\partial x^2}-\frac{\partial Q}{\partial x^3}\Big)dx^2\wedge dx^3+\Big(\frac{\partial P}{\partial x^3}-\frac{\partial R}{\partial x^1}\Big)dx^3\wedge dx^1.$$

Таким образом, интеграл по кривой Γ выражается через интеграл второго рода по поверхности V:

$$\int\limits_{\Gamma}(Pdx^1+Qdx^2+Rdx^3)=\iint\limits_{V}\Big(\Big(\frac{\partial Q}{\partial x^1}-\frac{\partial P}{\partial x^2}\Big)dx^1dx^2+$$
$$+\Big(\frac{\partial R}{\partial x^2}-\frac{\partial Q}{\partial x^3}\Big)dx^2dx^3+\Big(\frac{\partial P}{\partial x^3}-\frac{\partial R}{\partial x^1}\Big)dx^3dx^1\Big).$$

Пример 4. Последняя формула, *формула Гаусса–Остроградского*, тоже является частным случаем формулы (5.12). Пусть Γ — замкнутая поверхность в $\mathbb{R}^3$, ограничивающая область V. Пусть ω — внешняя дифференциальная форма степени 2. Форма ω в координатах (x^1,x^2,x^3) имеет вид

$$\omega=Pdx^1\wedge dx^2+Qdx^2\wedge dx^3+Rdx^3\wedge dx^1.$$

Тогда интеграл второго рода по поверхности Γ интерпретируется как интеграл формы ω. Применяя формулу (5.12), получаем

$$\int\limits_{\Gamma} \omega = -\int\limits_{V} d\omega;$$

$$d\omega = dP \wedge dx^1 \wedge dx^2 + dQ \wedge dx^2 \wedge dx^3 + dR \wedge dx^3 \wedge dx^1 = \\ = \left(\frac{\partial P}{\partial x^3} + \frac{\partial Q}{\partial x^1} + \frac{\partial R}{\partial x^2}\right) dx^1 \wedge dx^2 \wedge dx^3.$$

Следовательно,

$$\iint\limits_{\Gamma} (P dx^1 dx^2 + Q dx^2 dx^3 + R dx^3 dx^1) = \\ = -\iiint\limits_{V} \left(\frac{\partial P}{\partial x^3} + \frac{\partial Q}{\partial x^1} + \frac{\partial R}{\partial x^2}\right) dx^1 dx^2 dx^3.$$

Мы пришли к формуле Гаусса–Остроградского с обратным знаком. Знак минус связан с тем, что в классической формуле Гаусса–Остроградского ориентация поверхности Γ, задаваемая парой векторов $(\mathbf{e}_1, \mathbf{e}_2)$ в касательном пространстве к поверхности Γ, выбирается так, чтобы третий вектор $\mathbf{e}_3$ указывал направление вне области V.

Пример 5. Наконец, частным случаем формулы (5.12) можно считать и *формулу Ньютона–Лейбница*: $\int\limits_a^b (df/dx)dx = f(b) - f(a)$. Левую часть формулы Ньютона–Лейбница интерпретируем как интеграл формы $df = \dfrac{df}{dx}dx$ по отрезку $[a, b]$ вещественных чисел, который является одномерным ориентированным многообразием с краем. Краем отрезка $[a, b]$ является совокупность двух точек $\{a, b\}$, которые мы будем считать нульмерным многообразием. Однако для нульмерных многообразий мы не определяли понятие ориентации, поскольку у них отсутствуют касательные векторы. Поэтому можно поступить следующим образом. Будем считать, что точка края одномерного многообразия M имеет положительную ориентацию, если вектор, указывающий направление внутрь M, дает исходную ориентацию многообразия M, и, наоборот, имеет отрицательную, если вектор, указывающий направление внутрь M, дает противоположную ориентацию многообразия M. Нульмерная форма на многообразии является просто функцией. Поэтому интеграл функции по нульмерному многообразию будем считать суммой значений функции в точках, взятых со знаком, равным ориентации точек на многообразии. В нашем случае, правая часть формулы Ньютона–Лейбница будет истолкована как интеграл функции f по краю, состоящему из двух точек $\{a, b\}$. При этом точка a имеет положительную ориентацию, а точка

b — отрицательную. Поэтому $\int\limits_{\partial[a,b]} f = f(a) - f(b)$. Формула же (5.12) имеет вид: $\int\limits_{[a,b]} df = (-1)^n \int\limits_{\partial[a,b]} f$.

Доказательство. Левая и правая часть формулы (5.12) линейна относительно формы ω. Поэтому, разлагая форму ω в сумму $\omega = \omega_1 + \dots + \omega_N$, достаточно доказывать формулу (5.12) для случая, когда носитель формы ω компактен и лежит в одной карте. Более того, достаточно доказать формулу (5.12) для случая формы ω вида

$$\omega = f(x^1, \dots, x^n) dx^1 \wedge \dots \wedge dx^{k-1} \wedge dx^{k+1} \wedge \dots \wedge dx^n,$$

где f — функция с компактным носителем, определенная в $\mathbb{R}^n_+$. Тогда $d\omega = (-1)^{k-1}(\partial f/\partial x^k) dx^1 \wedge \dots \wedge dx^n$. Рассмотрим два случая. Пусть $k < n$. Тогда ограничение формы ω на край $\mathbb{R}^{n-1}_0 \subset \mathbb{R}^n_+$ равно нулю, поскольку $dx^n = 0$. Поэтому $\int\limits_{\mathbb{R}^{n-1}_0} \omega = 0$. С другой стороны, $\int\limits_{\mathbb{R}^n_+} d\omega = \int\limits_{\mathbb{R}^n_+} \dots \int (-1)^{k-1}(\partial f/\partial x^k) dx^1 \dots dx^n$. Переходя к повторному интегралу, сначала по переменной x^k, а потом по всем остальным переменным получаем

$$\int\limits_{\mathbb{R}^n_+} d\omega = \int\limits_{\mathbb{R}^{n-1}_+} \dots \int (-1)^{k-1} dx^1 \dots dx^{k-1} dx^{k+1} \dots dx^n \int\limits_{-\infty}^{\infty} \frac{\partial f}{\partial x^k} dx^k.$$

Внутренний интеграл равен нулю в силу следующих очевидных соотношений: $\int\limits_{-\infty}^{\infty} \frac{\partial f}{\partial x^k} dx^k = f(x^1, \dots, x^n)|_{x^k=-\infty}^{x^k=+\infty}$. Поэтому $\int\limits_{\mathbb{R}^n_+} d\omega = 0$. Пусть теперь $k = n$. Тогда $\omega = f(x^1, \dots, x^n) dx^1 \wedge \dots \wedge dx^{n-1}$,

$$d\omega = (-1)^{n-1} \frac{\partial f}{\partial x^n}(x^1, \dots, x^n) dx^1 \wedge \dots \wedge dx^{n-1} \wedge dx^n.$$

Имеем:

$$\int\limits_{\mathbb{R}^{n-1}_0} \omega = \int\limits_{\mathbb{R}^{n-1}_0} \dots \int f(x^1, \dots, x^{n-1}, 0) dx^1 \dots dx^{n-1};$$

$$\int\limits_{\mathbb{R}^n_+} d\omega = \int\limits_{\mathbb{R}^n_+} \cdots \int (-1)^{n-1} \frac{\partial f}{\partial x^n} dx^1 \dots dx^{n-1} dx^n =$$

$$= \int\limits_{\mathbb{R}^{n-1}_0} \cdots \int dx^1 \dots dx^{n-1} \int\limits_0^\infty (-1)^{n-1} \frac{\partial f}{\partial x^n} dx^n =$$

$$= \int\limits_{\mathbb{R}^{n-1}_0} \cdots \int dx^1 \dots dx^{n-1} \big((-1)^{n-1} f(x^1, \dots, x^n)|_{x^n=0}^{x^n=+\infty}\big) =$$

$$= (-1)^n \int\limits_{\mathbb{R}^{n-1}_0} \cdots \int f(x^1, \dots, x^{n-1}, 0) dx^1 \dots dx^{n-1}.$$

Следовательно, $\int\limits_{\mathbb{R}^{n-1}_0} = (-1)^n \int\limits_{\mathbb{R}^n_+} d\omega$. Таким образом, для завершения доказательства, достаточно разложить форму ω в сумму $\omega = \omega_1 + \dots + \omega_N$ так, чтобы носитель каждого слагаемого лежал в одной карте многообразия M. Рассмотрим атлас карт $\{U_\alpha\}$ и разбиение единицы $\{\varphi_\alpha\}$, подчиненное атласу $\{U_\alpha\}$. Тогда $1 \equiv \sum \varphi_\alpha$, $\omega = \sum \varphi_\alpha \omega$, $\operatorname{supp} \varphi_\alpha \omega \subset U_\alpha$. Теорема доказана. ■

Часть II

КОМПЬЮТЕРНАЯ ГЕОМЕТРИЯ

Глава 6

Гладкие кривые с вычислительной точки зрения

6.1 Приблизительная локальная форма кривой, определяемая кривизной и кручением

Рассмотрим произвольную бирегулярную гладкую кривую γ в $\mathbb{R}^3$, заданную радиус-вектором $\mathbf{r} = \mathbf{r}(s)$, где s — натуральный параметр. Будем, как обычно, обозначать дифференцирование по параметру s точками сверху. Пусть $\big(\mathbf{v}(s), \mathbf{n}(s), \mathbf{b}(s)\big)$ — репер Френе, где $\mathbf{v} = \dot{\mathbf{r}}$ — вектор скорости кривой γ, $\mathbf{n} = \ddot{\mathbf{r}}$ — вектор ускорения кривой γ и $\mathbf{b} = \mathbf{v} \times \mathbf{n}$ — бинормаль кривой γ. В разделе 2.1.2 были доказаны формулы Френе:

$$\begin{cases} \dot{\mathbf{v}} = k\mathbf{n}, \\ \dot{\mathbf{n}} = -k\mathbf{v} + \varkappa\mathbf{b}, \\ \dot{\mathbf{b}} = -\varkappa\mathbf{n}, \end{cases}$$

где $k = |\mathbf{v}|$ — кривизна кривой γ, а $\varkappa = -\langle\dot{\mathbf{b}}, \mathbf{n}\rangle$ — кручение. Заметим, что формулы Френе записаны здесь для натурально параметризованной кривой.

Имеем $\dot{\mathbf{r}} = \mathbf{v}$, $\ddot{\mathbf{r}} = k\mathbf{n}$, $\dddot{\mathbf{r}} = \dot{k}\mathbf{n} + k\dot{\mathbf{n}} = \dot{k}\mathbf{n} - k(k\mathbf{v} - \varkappa\mathbf{b})$. Вычисляя смешанное произведение $(\dot{\mathbf{r}}, \ddot{\mathbf{r}}, \dddot{\mathbf{r}})$ получаем

$$(\dot{\mathbf{r}}, \ddot{\mathbf{r}}, \dddot{\mathbf{r}}) = \big\langle \dot{\mathbf{r}} \times \ddot{\mathbf{r}}, \dddot{\mathbf{r}} \big\rangle = \langle k\mathbf{b}, \dot{k}\mathbf{n}\rangle - k^2\langle \mathbf{b}, k\mathbf{v} - \varkappa\mathbf{b}\rangle = k^2\varkappa.$$

Отсюда получаем формулу для кручения:

$$\varkappa = \frac{1}{k^2}(\dot{\mathbf{r}}, \ddot{\mathbf{r}}, \dddot{\mathbf{r}})$$

(в предположении, что параметр натуральный).

Выберем на кривой γ: $\mathbf{r} = \mathbf{r}(s)$ некоторую точку O и подберем натуральный параметр s так, чтобы $\mathbf{r}(0) = O$. Рассмотрим систему координат, определяемую репером Френе кривой γ в точке O (это всегда можно сделать). Разложим функцию $\mathbf{r} = \mathbf{r}(s)$ в окрестности $s = 0$ в ряд Тейлора по s:

$$\mathbf{r}(s) = \dot{\mathbf{r}}(0)s + \frac{1}{2}\ddot{\mathbf{r}}(0)s^2 + \frac{1}{6}\dddot{\mathbf{r}}(0)s^3 + o(s^3) =$$
$$= \Big(s - k^2\frac{s^3}{6}\Big)\mathbf{v} + \Big(k\frac{s^2}{2} + \dot{k}\frac{s^3}{6}\Big)\mathbf{n} + k\varkappa\frac{s^3}{6}\mathbf{b} + o(s^3).$$

Если обозначить компоненты координатной системы, определяемой репером Френе $(\mathbf{v}, \mathbf{n}, \mathbf{b})$ в точке O, через (x, y, z), то получим параметрическое уравнение кривой γ с точностью до $o(s^3)$ в виде

$$x = s - k^2\frac{s^3}{6}, \qquad y = k\frac{s^2}{2} + \dot{k}\frac{s^3}{6}, \qquad z = k\varkappa\frac{s^3}{6}.$$

Следовательно, уравнение кривой γ в окрестности точки O в выбранной системе координат (определяемой репером Френе) имеет вид

$$x = s\big(1 + o(1)\big), \qquad y = k\frac{s^2}{2}\big(1 + o(1)\big), \qquad z = k\varkappa\frac{s^3}{6}\big(1 + o(1)\big).$$

Приблизительно кривая γ в окрестности точки O может быть задана уравнениями

$$x = s, \qquad y = k\frac{s^2}{2}, \qquad z = k\varkappa\frac{s^3}{6}.$$

Исключая s в этих уравнениях, получаем уравнения упрощенной кривой $\widetilde{\gamma}$, приблизительно совпадающей с данной кривой γ в окрестности точки $s = 0$:

$$y = \frac{1}{2}kx^2, \qquad y^3 = \frac{9}{2}\frac{k}{\varkappa^2}z^2, \qquad z = \frac{1}{6}k\varkappa x^3.$$

Итак, для упрощенной кривой $\widetilde{\gamma}$ справедливы следующие утверждения:

1) проекция кривой $\widetilde{\gamma}$ на соприкасающуюся плоскость Oxy, натянутую на вектора $\mathbf{v}$ и $\mathbf{n}$, — это парабола $y = \frac{1}{2}kx^2$ с той же кривизной в вершине, что и у исходной кривой γ в данной точке, см. рис. 6.1,

2) проекция кривой $\widetilde{\gamma}$ на нормальную плоскость Oyz, натянутую на вектора $\mathbf{n}$ и $\mathbf{b}$, имеет вид $y = c_1 z^{2/3}$, $c_1 > 0$, см. рис. 6.1,

3) проекция кривой $\widetilde{\gamma}$ на спрямляющую плоскость Oxz, натянутую на вектора $\mathbf{v}$ и $\mathbf{b}$, (эта плоскость касается кривой в точке O) имеет вид $z = c_2 x^3$, см. рис. 6.1.

Замечание. Выше, упрощая кривую γ, мы отбросили слагаемое $\dot{k}\frac{s^3}{6}\mathbf{n}$ в разложение Тейлора функции $\mathbf{r} = \mathbf{r}(s)$. Если учесть это слагаемое, то в

Рис. 6.1:

Рис. 6.2:

случае, когда $\dot{k} \neq 0$ в точке O (т.е. когда кривизна не имеет экстремума или перегиба в точке O) упрощенная форма кривой с учетом всех слагаемых порядка до $O(s^3)$ включительно должна быть подправлена в направлении вектора $\mathbf{n}$ кососимметрично по s, см. рис. 6.2.

6.2 Формулы для кривизны и кручения кривой относительно произвольного параметра в координатах, задаваемых репером Френе

Пусть кривая γ задана параметрически уравнением

$$\mathbf{r} = \mathbf{r}(t),$$

где $t = t(s)$ — гладкая строго монотонная функция натурального параметра s. Тогда будет существовать гладкая обратная функция $s = s(t)$, причем $\dfrac{dt}{ds} = 1\Big/\dfrac{ds}{dt}$ $\left(\dfrac{dt}{ds} \neq 0,\ \dfrac{ds}{dt} \neq 0\right)$. Будем, как обычно, обозначать дифференцирование по t точкой сверху. Имеем

$$\dot{\mathbf{r}} = \frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}}{ds}\frac{ds}{dt} = \mathbf{v}\frac{ds}{dt}, \tag{6.1}$$

где $\mathbf{v}$ — вектор скорости кривой γ относительно натурального параметра s и, следовательно, $|\mathbf{v}| = 1$ (см. гл. 2). Отсюда следует, что

$$|\dot{\mathbf{r}}|^2 = |\mathbf{v}|^2\Big(\frac{ds}{dt}\Big)^2 = \Big(\frac{ds}{dt}\Big)^2. \tag{6.2}$$

Следовательно,

$$\frac{ds}{dt} = |\dot{\mathbf{r}}| \tag{6.3}$$

и

$$\mathbf{v} = \frac{\dot{\mathbf{r}}}{|\dot{\mathbf{r}}|}.$$

Дифференцируя (6.1) по t, получим

$$\ddot{\mathbf{r}} = \frac{d^2\mathbf{r}}{ds^2}\Big(\frac{ds}{dt}\Big)^2 + \frac{d\mathbf{r}}{ds}\frac{d^2s}{dt^2} = k\mathbf{n}\Big(\frac{ds}{dt}\Big)^2 + \mathbf{v}\frac{d^2s}{dt^2}, \tag{6.4}$$

где $\mathbf{n}$ — вектор главной нормали к кривой: $\mathbf{n} = \dfrac{d^2\mathbf{r}}{ds^2}\Big/\Big|\dfrac{d^2\mathbf{r}}{ds^2}\Big| = \dfrac{1}{k}\dfrac{d^2\mathbf{r}}{ds^2}$ (см. гл. 2). Умножим обе части векторного равенства (6.4) скалярно на $\mathbf{n}$:

$$\langle\ddot{\mathbf{r}}, \mathbf{n}\rangle = k\Big(\frac{ds}{dt}\Big)^2 = k|\dot{\mathbf{r}}|^2. \tag{6.5}$$

Последнее равенство следует из (6.2). Отсюда получается следующее выражение для кривизны k кривой γ:

$$k = \frac{\langle\ddot{\mathbf{r}}, \mathbf{n}\rangle}{|\dot{\mathbf{r}}|^2}.$$

Таким образом, кривизна — это компонента вектора ускорения кривой в направлении главной нормали, деленная на квадрат длины вектора скорости.

Продифференцируем (6.3) по t:

$$\frac{d^2s}{dt^2} = \frac{d}{dt}\langle\dot{\mathbf{r}}, \dot{\mathbf{r}}\rangle^{\frac{1}{2}} = \frac{1}{2\sqrt{\langle\dot{\mathbf{r}}, \dot{\mathbf{r}}\rangle}}\frac{d}{dt}\langle\dot{\mathbf{r}}, \dot{\mathbf{r}}\rangle = \frac{\langle\ddot{\mathbf{r}}, \dot{\mathbf{r}}\rangle}{|\dot{\mathbf{r}}|}.$$

Следовательно,

$$\frac{d^2s}{dt^2} = \langle\ddot{\mathbf{r}}, \mathbf{v}\rangle. \tag{6.6}$$

Из (6.4) и (6.6) получаем следующее выражение для вектора $k\mathbf{n} = \dfrac{d^2\mathbf{r}}{ds^2}$:

$$k\mathbf{n} = \frac{\ddot{\mathbf{r}} - \mathbf{v}\langle\ddot{\mathbf{r}}, \mathbf{v}\rangle}{|\dot{\mathbf{r}}|^2} = \frac{\ddot{\mathbf{r}} - \text{Пр}\ \ddot{\mathbf{r}}}{|\dot{\mathbf{r}}|^2} = \frac{1}{|\dot{\mathbf{r}}|^2}\,\text{Пр}\ \ddot{\mathbf{r}},$$

где через Пр обозначается проекция на направление вектора $\mathbf{a}$. (Из (6.4) следует, что $\ddot{\mathbf{r}}$ лежит в $O\mathbf{vn} = Oxy$.)

Найдем теперь выражение для кручения. Продифференцируем (6.4) по t еще раз. С учетом (6.5) имеем

$$\dddot{\mathbf{r}} = \frac{d^3\mathbf{r}}{dt^3} = \mathbf{n}\frac{d}{dt}\Big(k|\dot{\mathbf{r}}|^2\Big) + k\frac{d\mathbf{n}}{ds}\Big(\frac{ds}{dt}\Big)^3 + \frac{d\mathbf{v}}{ds}\frac{ds}{dt}\frac{d^2s}{dt^2} + \mathbf{v}\frac{d^3s}{dt^3}.$$

Подставляя сюда выражения для $\dfrac{d\mathbf{n}}{ds}$ и $\dfrac{d\mathbf{v}}{ds}$ из формул Френе и умножая скалярно на вектор $\mathbf{b}$, получим $\langle\dddot{\mathbf{r}}, \mathbf{b}\rangle = k\varkappa|\dot{\mathbf{r}}|^3$. Следовательно,

$$\varkappa = \frac{\langle\dddot{\mathbf{r}}, \mathbf{b}\rangle}{k|\dot{\mathbf{r}}|^3}.$$

Таким образом, кручение определяется величиной проекции вектора $\dddot{\mathbf{r}}$ на вектор бинормали $\mathbf{b}$.

6.3 Восстановление пространственной кривой по ее проекциям на координатные плоскости

Пусть кривая γ задана своими проекциями на плоскости Oxy и Oyz некоторой декартовой системы координат: $\mathbf{r}_{xy} = \mathbf{r}_{xy}(t)$, $\mathbf{r}_{yz} = \mathbf{r}_{yz}(t)$. Пусть s — натуральный параметр на кривой γ. Пренебрегая слагаемыми порядка $o(s^2)$, приблизительные параметрические уравнения кривой γ в окрестности точки $s = 0$ запишем в виде

$$x = sv^1 + \frac{1}{2}ks^2n^1, \quad y = sv^2 + \frac{1}{2}ks^2n^2, \quad z = sv^3 + \frac{1}{2}ks^2n^3, \tag{6.7}$$

где $\mathbf{v} = (v^1, v^2, v^3)$ — вектор скорости кривой γ, $\mathbf{n} = (n^1, n^2, n^3)$ — вектор главной нормали к кривой γ. Компоненты вектора скорости $\mathbf{v}$ легко определить по компонентам векторов скоростей проекций $\dot{\mathbf{r}}_{xy}$ и $\dot{\mathbf{r}}_{yz}$, пользуясь пропорцией:

$$dx : dy : dz = v^1 : v^2 : v^3$$

и теоремой об инвариантности первого дифференциала, согласно которой левая часть этой пропорции инвариантна относительно замены параметра на кривой и ее проекциях.

Кривизны проекций кривой γ на координатные плоскости Oxy и Oyz определяются по формулам:

$$\pm k_{xy} = \frac{y''x' - x''y'}{(x'^2 + y'^2)^{3/2}}, \qquad \pm k_{yz} = \frac{z''y' - y''z'}{(y'^2 + z'^2)^{3/2}}. \tag{6.8}$$

Подставляя (6.7) в (6.8), получаем (в точке $s = 0$)

$$\pm k_{xy} = \frac{k(n^2v^1 - n^1v^2)}{\big((v^1)^2 + (v^2)^2\big)^{3/2}}, \qquad \pm k_{yz} = \frac{k(n^3v^2 - n^2v^3)}{\big((v^2)^2 + (v^3)^2\big)^{3/2}}. \tag{6.9}$$

Обозначим

$$A = \pm k_{xy}\big((v^1)^2 + (v^2)^2\big)^{3/2}, \qquad B = \pm k_{yz}\big((v^2)^2 + (v^3)^2\big)^{3/2}. \tag{6.10}$$

Подберем знаки A и B так, чтобы (6.9) переписывалось в виде

$$k(n^2v^1 - n^1v^2) = A, \qquad k(n^3v^2 - n^2v^3) = B. \tag{6.11}$$

Умножим первое равенство в (6.11) на B/k, а второе — на A/k. Приравнивая левые части, получим тождество

$$Bn^2v^1 - (Bn^1 + An^3)v^2 + An^2v^3 = 0.$$

Прибавив к нему условие ортогональности векторов $\mathbf{v}$ и $\mathbf{n}$ и условие $|\mathbf{n}| = 1$, мы получим систему из трех уравнений для трех компонент вектора главной нормали $\mathbf{n}$. Решение этой системы будет иметь вид

$$\begin{aligned} n^1 &= C\big(-(Bv^1 + Av^3)v^3 - A(v^2)^2\big),\\ n^2 &= C\big(-Bv^2v^3 + Av^1v^2\big),\\ n^3 &= C\big((Bv^1 + Av^3)v^3 + B(v^2)^2\big), \end{aligned} \tag{6.12}$$

где постоянная C легко определяется из условия $|\mathbf{n}| = 1$:

$$C = \Big(\big((v^2)^2 + (v^1)^2\big)B^2 + \big((v^2)^2 + (v^3)^2\big)A^2\Big)^{-\frac{1}{2}}. \tag{6.13}$$

Кривизна k определяется из (6.11) и (6.12):

$$k = \left|\frac{1}{Cv^2}\right|. \tag{6.14}$$

Итак, формулы (6.10), (6.12), (6.13) определяют вектор $\mathbf{n}$ главной нормали к кривой в пространстве по ее двум проекциям, а формула (6.14) определяет кривизну этой кривой. Вектор бинормали определяется как $\mathbf{b} = \mathbf{v} \times \mathbf{n}$, поэтому репер Френе кривой определен нами полностью.

6.4 Приведение параметрического уравнения кривой к неявному виду

Предположим, что требуется найти точки пересечения двух кривых γ_1 и γ_2 на плоскости $\mathbb{R}^2$, заданных параметрическими уравнениями $\mathbf{r}_1 = \mathbf{r}_1(t)$ и $\mathbf{r}_2 = \mathbf{r}_2(q)$. Поскольку в точках пересечения кривых γ_1 и γ_2 значения параметров t и q, вообще говоря, не равны, то придется решать систему из

двух, вообще говоря, нелинейных уравнений с двумя неизвестными, которая в скалярном виде записывается так:

$$\begin{cases} r_1^1(t) = r_2^1(q), \\ r_1^2(t) = r_2^2(q). \end{cases}$$

В векторном виде эта система записывается как $\mathbf{r}_1(t) = \mathbf{r}_2(q)$. Получить решение такой системы численными методами может оказаться трудно. Однако, если одна из кривых (пусть для определенности это будет γ_2) задана неявным уравнением $f(x, y) = 0$, то для нахождения точек пересечения кривых γ_1 и γ_2 возникает одно уравнение с одним неизвестным t:

$$f\big(x(t), y(t)\big) = 0, \tag{6.15}$$

где $\mathbf{r}_1(t) = \big(x(t), y(t)\big)$. Для решения уравнений вида (6.15) есть ряд хорошо известных численных методов и стандартных алгоритмов.

В этой связи в компьютерной геометрии возникает задача преобразования уравнения кривой к неявному виду. Рассмотрим один из возможных способов решения этой задачи для частного (но важного для компьютерной геометрии) случая, когда рассматриваемая кривая γ задана параметрическим уравнением $\mathbf{r} = \mathbf{r}(t)$, где $\mathbf{r}(t)$ — рациональная функция от t вида

$$\mathbf{r}(t) = \frac{\mathbf{r}_0 + t\mathbf{r}_1 + \cdots + t^n\mathbf{r}_n}{\alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n},$$

где $\mathbf{r}_0, \ldots, \mathbf{r}_n \in \mathbb{R}^2$, $\alpha_0, \ldots, \alpha_n \in \mathbb{R}$. Принадлежность точки $(x, y) \in \mathbb{R}^2$ к кривой γ равносильна следующей системе уравнений:

$$\begin{cases} x(\alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n) = r_0^1 + tr_1^1 + \cdots + t^n r_n^1, \\ y(\alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n) = r_0^2 + tr_1^2 + \cdots + t^n r_n^2. \end{cases}$$

Приводя подобные члены с одинаковыми степенями по t получаем

$$\begin{cases} X(t) = a_0 + a_1 t + \cdots + a_n t^n = 0, \\ Y(t) = b_0 + b_1 t + \cdots + b_n t^n = 0, \end{cases} \tag{6.16}$$

где коэффициенты $a_0, \ldots, a_n$ и $b_0, \ldots, b_n$ — линейные функции от x и y соответственно. Например

$$\begin{cases} a_0 = r_n^1 - x\alpha_n, \\ b_0 = r_n^2 - y\alpha_n \end{cases}$$

и т.д.

Составим уравнение для кривой γ, не содержащее t^n:

$$b_n X(t) - a_n Y(t) = 0 \Longleftrightarrow c_{0(n-1)} t^{n-1} + \cdots + c_{01} t + c_{00} = 0.$$

Здесь коэффициенты c_{0i} — полиномы от x и y.

Аналогично, можно понизить степень t с n до $(n-1)$ еще $(n-1)$ способами:

$$
\begin{aligned}
(b_n t + b_{n-1})X(t) - (a_n t + a_{n-1})Y(t) = 0 \Longleftrightarrow \\
c_{1(n-1)}t^{n-1} + \dots + c_{11}t + c_{10} = 0, \\
\dots\dots\dots\dots\dots\dots\dots\dots \\
(b_n t^i + b_{n-1}t^{i-1} + \dots + b_{n-i})X(t) - (a_n t^i + a_{n-1}t^{i-1} + \dots + a_{n-i})Y(t) = 0 \Longleftrightarrow \\
c_{i(n-1)}t^{n-1} + \dots + c_{i1}t + c_{i0} = 0. \qquad (6.17)
\end{aligned}
$$

В итоге получаем следующую систему уравнений:

$$
\begin{pmatrix} c_{0(n-1)} & \dots & c_{00} \\ \vdots & \ddots & \vdots \\ c_{(n-1)(n-1)} & \dots & c_{(n-1)0} \end{pmatrix}
\begin{pmatrix} t^{n-1} \\ \vdots \\ t^0 \end{pmatrix} = 0.
$$

Эта система вида $C\tau = 0$, $\tau \in \mathbb{R}^n$, должна иметь ненулевое решение $(t^{n-1}, \dots, t^0)$, если точка (x, y) принадлежит кривой γ. Следовательно, если точка (x, y) принадлежит кривой γ, то указанная система имеет ненулевое решение, и следовательно определитель системы равен нулю: $\det C = 0$. Определитель $\det C$ — полином степени n от x и y, так как каждый элемент матрицы C представляет собой линейную функцию по x или по y. Следовательно, уравнение $\det C = 0$ — это алгебраическое уравнение относительно x и y. Берем это уравнение в качестве неявной формы записи кривой γ. При этом мы получаем, вообще говоря, некоторую кривую $\widetilde{\gamma}$, имеющую γ своей частью, $\widetilde{\gamma} \supset \gamma$, поскольку сделанные выше преобразования не были равносильными. Поэтому после нахождения точек пересечения указанным способом необходимо будет произвести проверку на принадлежность этих точек заданным кривым.

Пример 1. Рассмотрим рациональную кривую Безье второй степени (см. гл. 9):

$$
\mathbf{r}(t) = \frac{(1-t)^2\mathbf{p}_0 + 2t(1-t)\omega\mathbf{p}_1 + t^2\mathbf{p}_2}{(1-t)^2 + 2t(1-t)\omega + t^2}. \qquad (6.18)
$$

Сделаем замену $t = \dfrac{u}{1+u}$. Тогда (6.18) преобразуется к виду

$$
\mathbf{r}(u) = \frac{\mathbf{p}_0 + 2u\omega\mathbf{p}_1 + u^2\mathbf{p}_2}{1 + 2u\omega + u^2},
$$

где $\mathbf{r}(u) = \big(x(u), y(u)\big) = (x, y)$. В данном случае уравнения (6.16) примут вид

$$
\begin{cases} X(u) = u^2(x - x_2) + 2u\omega(x - x_1) + (x - x_0) = 0, \\ Y(u) = u^2(y - y_2) + 2u\omega(y - y_1) + (y - y_0) = 0, \end{cases} \qquad (6.19)
$$

а уравнения (6.17) запишутся в виде

$$\begin{cases} (y-y_2)X(u)-(x-x_2)Y(u)=0, \\ \big((y-y_2)u+2\omega(y-y_1)\big)X(u)-\big((x-x_2)u+2\omega(x-x_1)\big)Y(u)=0. \end{cases}$$

Пусть $O=(0,0)$, $\mathbf{p}_0=(-c,0)$, $\mathbf{p}_1=(b,h)$, $\mathbf{p}_2=(c,0)$, где c, b, $h>0$. Получаем общее уравнение кривой 2-го порядка

$$h^2x^2-2bhxy+\Big(\frac{c^2}{\omega^2}+\big(b^2-c^2\big)\Big)y^2+2c^2hy-c^2h^2=0. \tag{6.20}$$

В данном примере те же самые вычисления можно провести проще. Рассмотрим (6.19) как систему из двух уравнений с двумя неизвестными u^2 и u:

$$\begin{cases} u^2(c-x)+2u\omega(b-x)-(c+x)=0, \\ u^2(-y)+2u\omega(h-y)-y=0. \end{cases}$$

По формулам решения системы двух линейных уравнений с двумя неизвестными получаем

$$\begin{cases} u^2=\dfrac{2\omega\big((c+x)h-y(b+c)\big)}{D} \\ u=\dfrac{2cy}{D}, \end{cases} \tag{6.21}$$

где $D=2\omega\big((c-x)h+y(b-c)\big)$.

Подставляя второе уравнение из (6.21) в первое, получаем (6.20).

Таким образом, мы получили способ преобразования рациональной кривой Безье второй степени к виду кривой второго порядка.

Глава 7

Сплайны и кривые Безье

7.1 Сплайны

Определение 7.1. *Сплайном* называется кривая, построенная по совокупности точек $\mathbf{p}_0, \ldots, \mathbf{p}_n$, которые называются *характеристическими или опорными точками сплайна.* Требуется, чтобы сплайн $\mathbf{r} = \mathbf{r}(t)$ проходил через заданные точки $\mathbf{p}_0, \ldots, \mathbf{p}_n$ при заданных значениях параметра $t_0 \leqslant \ldots \leqslant t_n$: $\mathbf{r}(t_i) = \mathbf{p}_i$, $0 \leqslant i \leqslant n$. Значения параметра t_i $(0 \leqslant i \leqslant n)$ называются *узловыми*, а точки $t_0, \ldots, t_n$ на оси $\mathbb{R}$ — *узлами сплайна.*

7.1.1 Примеры сплайнов

1) *Ломаная линия* с вершинами в точках $\mathbf{p}_i$:

$$\mathbf{r}(t) = \mathbf{p}_i(1 - \omega) + \mathbf{p}_{i+1}\omega,$$

где $\omega = \dfrac{t - t_i}{t_{i+1} - t_i}$ и $t_i \leqslant t \leqslant t_{i+1}$. Параметр ω называется *местным* параметром на участке $t_i \leqslant t \leqslant t_{i+1}$.

Если $t_i = i$, то параметризация ломаной называется *равномерной.* В этом случае *параметрическая длина* $T = t_n - t_0$ ломаной равна n, т.е. — числу опорных точек минус 1, если все опорные точки $\mathbf{p}_0, \ldots, \mathbf{p}_n$ различны.

Если $\mathbf{p}_0 = \mathbf{p}_n$, т.е. ломаная замкнута, ее параметрическая длина равна числу опорных точек (в предположении, что точки $\mathbf{p}_0, \ldots, \mathbf{p}_{n-1}$ различны).

2) *Сплайн Эрмита.*

Пусть имеется $(m + 1)$ точка на плоскости, и в каждой из них задана производная радиус-вектора $\mathbf{r}(t)$ кривой, т.е. заданы $\mathbf{r}'(t_0), \ldots, \mathbf{r}'(t_m)$. Тогда существует так называемый *сплайн Эрмита*, задаваемый полиномом степени $(2m + 1)$ от t, проходящий через заданные точки и имеющий в них заданные производные $\mathbf{r}'(t_0), \ldots, \mathbf{r}'(t_m)$. Этот полином имеет вид $\mathbf{r}(t) = \alpha_{2m+1}t^{2m+1} + \cdots + \alpha_1 t + \alpha_0$, $\alpha_i \in \mathbb{R}^2$.

Если $m = 1$, т.е. заданы две точки и производные в них, то существует сплайн Эрмита порядка три, идущий из точки $\mathbf{r}_0$ в точку $\mathbf{r}_1$ и имеющий заданные производные на концах.

В компьютерной геометрии используются составные сплайны Эрмита 3-го порядка, проходящие через заданные точки $\mathbf{p}_0, \ldots, \mathbf{p}_n$ и имеющие в этих точках заданные производные $\mathbf{q}_0, \ldots, \mathbf{q}_n$. На участке (t_i, t_{i+1}) он задается формулой

$$\mathbf{r}_i(\omega) = \mathbf{a}_0 + \mathbf{a}_1\omega + \mathbf{a}_2\omega^2 + \mathbf{a}_3\omega^3,$$

где $\omega = \dfrac{t - t_i}{t_{i+1} - t_i}$ — местный параметр, $0 \leqslant \omega \leqslant 1$. Вектора $\mathbf{a}_j$, $0 \leqslant j \leqslant 3$, определяются из граничных условий: $\mathbf{r}_i(0) = \mathbf{p}_i$.

7.1.2 Построение сплайнов Эрмита

Постановка задачи. Пусть даны две точки $\mathbf{r}(0) = \mathbf{r}_0$, $\mathbf{r}(1) = \mathbf{r}_1$. Требуется построить кривую $\mathbf{r} = \mathbf{r}(t)$, $0 \leqslant t \leqslant 1$, такую, что ее производные до порядка m включительно принимают на концах отрезка $0 \leqslant t \leqslant 1$ заданные значения:

$$\begin{cases} \left.\dfrac{d^k}{dt^k}\right|_{t=0} \mathbf{r}(t) = \mathbf{r}_0^{(k)}, \\ \left.\dfrac{d^k}{dt^k}\right|_{t=1} \mathbf{r}(t) = \mathbf{r}_1^{(k)}, \ 0 \leqslant k \leqslant m. \end{cases} \tag{7.1}$$

Здесь $\mathbf{r}_0^{(k)}$, $\mathbf{r}_1^{(k)}$, $0 \leqslant k \leqslant m$, — заданные вектора, причем $\mathbf{r}_0^{(0)} = \mathbf{r}_0$, $\mathbf{r}_1^{(0)} = \mathbf{r}_1$. Требуется, чтобы зависимость каждой компоненты $\mathbf{r}(t)$ имела вид полинома порядка $(2m + 1)$. Такой полином имеет $(2m + 2)$ коэффициентов, которые можно определить из системы (7.1), состоящей из $(2m + 2)$ уравнений для каждой компоненты $\mathbf{r}(t)$.

Утверждение 7.1. *Решение поставленной задачи существует и имеет вид полинома степени* $(2m + 1)$ *по* t *с векторными коэффициентами*:

$$\mathbf{r}(t) = \sum_{i=0}^{m} \big(g_i(t, m)\mathbf{r}_0^i + h_i(t, m)\mathbf{r}_1^i\big), \tag{7.2}$$

где $g_i(t, m)$, $h_i(t, m)$, $0 \leqslant i \leqslant m$, *— полиномы по* t *степени* $(2m + 1)$, *коэффициенты которых определяются из системы уравнений* (7.1) *и имеют вид*:

$$\begin{cases} g_i(t, m) = \dfrac{1}{i!} \displaystyle\sum_{j=0}^{m-i} C_{2m+1-i}^{j} t^{i+j} (1 - t)^{2m+1-(i+j)} \\ h_i(t, m) = (-1)^i \dfrac{1}{i!} \displaystyle\sum_{j=0}^{m-i} C_{2m+1-i}^{j} t^{2m+1-(i+j)} (1 - t)^{i+j}. \end{cases} \tag{7.3}$$

Доказательство. Справедливость утверждения проверяется прямой подстановкой (7.3) в (7.2), см. [53, 54]. ■

Рис. 7.1:

Из (7.3) следует, что функции $g_i(t,m)$ и $h_i(t,m)$ связаны следующими соотношениями:

$$g_0(t,m) + h_0(t,m) = (t+1-t)^{2m+1} = 1, \qquad h_i(t,m) = (-1)^i g_i(1-t,m).$$

Однако на практике применение формул (7.3) ограничивается, как правило, лишь теми случаями, когда необходимо заполнить пробел уже существующей кривой с заданной степенью гладкости (т.е. с заданными производными на концах, которые требуется соединить). Если же требуется построить новую кривую по n точкам, то значения производных в этих точках, как правило, не заданы. Поэтому их приходится доопределять из каких-то дополнительных соображений. Например, если мы ищем составной сплайн Эрмита порядка $2m+1=3$ $(m=1)$ для последовательности опорных точек $\mathbf{p}_0,\ldots,\mathbf{p}_n \in \mathbb{R}^2$ (или, более общо, $\mathbf{p}_0,\ldots,\mathbf{p}_n \in \mathbb{R}^l$), то неизвестные производные $\mathbf{q}_i$ во внутренних точках $\mathbf{p}_1,\ldots,\mathbf{p}_{n-1}$ можно определить по формулам:

$$\mathbf{q}_i = \frac{\mathbf{p}_{i+1} - \mathbf{p}_{i-1}}{t_{i+1} - t_{i-1}} \qquad (1 \leqslant i \leqslant n-1). \tag{7.4}$$

При неравномерном расположении опорных точек такой способ определения $\mathbf{q}_i$ может привести к появлению нежелательных петель у сплайна Эрмита.

Поэтому пользуются другими схемами определения $\mathbf{q}_i$. Например, можно положить:

$$\mathbf{q}_i = s_{i+1}\frac{\mathbf{p}_i - \mathbf{p}_{i-1}}{s_i + s_{i+1}} + s_i\frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{s_i + s_{i+1}} \tag{7.5}$$

или

$$\mathbf{q}_i = s_i\frac{\mathbf{p}_i - \mathbf{p}_{i-1}}{s_i + s_{i+1}} + s_{i+1}\frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{s_i + s_{i+1}} \qquad (1 \leqslant i \leqslant n-1), \tag{7.6}$$

где $s_i = |\mathbf{p}_i - \mathbf{p}_{i-1}|$, $s_{i+1} = |\mathbf{p}_{i+1} - \mathbf{p}_i|$.

На рис. 7.1 приведены ломаная линия и сплайн Эрмита, построенный данным способом.

Производные в крайних точках $\mathbf{p}_0$ и $\mathbf{p}_n$, если они не заданы исходя из каких-то других соображений, можно определить из условия, что третьи производные $\mathbf{r}(t)$ обращаются в нуль на концах:

$$\mathbf{r}^{(3)}(t_0) = 0, \qquad \mathbf{r}^{(3)}(t_n) = 0. \tag{7.7}$$

Задача. Написать программу построения составного сплайна Эрмита порядка 3 для опорных точек $\mathbf{p}_0, \ldots, \mathbf{p}_n$. Проследить появление петель при неравномерном расположение опорных точек. Сравнить метод (7.4) с методами (7.5) и (7.6).

Пример 2. Сплайн Эрмита порядка 3 для последовательности опорных точек $\mathbf{p}_0, \ldots, \mathbf{p}_n$ имеет вид

$$\begin{aligned}\mathbf{r}(t) = g_0(w,1)\mathbf{p}_i + g_1(w,1)\mathbf{q}_i + h_0(w,1)\mathbf{p}_{i+1} + h_1(w,1)\mathbf{q}_{i+1} = \\ = g_0(w)\mathbf{p}_i + g_1(w)\mathbf{q}_i + h_0(w)\mathbf{p}_{i+1} + h_1(w)\mathbf{q}_{i+1},\end{aligned} \tag{7.8}$$

где $w = \dfrac{t - t_i}{t_{i+1} - t_i}$, $0 \leqslant w \leqslant 1$, — местный параметр на участке $t_i \leqslant t \leqslant t_{i+1}$. Здесь

$$\begin{aligned} g_0(w) &= 1 - 3w^2 + 2w^3, & g_1(w) &= w - 2w^2 + w^3, \\ h_0(w) &= 3w^2 - 2w^3, & h_1(w) &= -w^2 + w^3. \end{aligned} \tag{7.9}$$

Значения $\mathbf{q}_i$ при $1 \leqslant i \leqslant n-1$ могут быть определены, например, по одной из формул (7.4), (7.5) или (7.6). Значения $\mathbf{q}_0$ и $\mathbf{q}_n$ часто определяют из условий (7.7), что дает

$$\begin{aligned} \mathbf{q}_0 &= 2(\mathbf{p}_1 - \mathbf{p}_0) - \mathbf{q}_1, \\ \mathbf{q}_n &= 2(\mathbf{p}_n - \mathbf{p}_{n-1}) - \mathbf{q}_{n-1}. \end{aligned} \tag{7.10}$$

Равенства (7.10) получаются постановкой (7.8) в (7.7).

7.1.3 Псевдоупругие сплайны Эрмита

При построении составного сплайна Эрмита третьего порядка неизвестные производные $\mathbf{q}_0, \ldots, \mathbf{q}_n$ в точках $\mathbf{p}_0, \ldots, \mathbf{p}_n$ могут быть определены также из следующих соображений.

Представим себе, что через точки $\mathbf{p}_0, \ldots, \mathbf{p}_n$ должна пройти упругая нить, которая стремится занять положение, соответствующее минимуму запасенной энергии ее внутренних сил упругости. Эта энергия, как известно из физики, выражается интегралом от квадрата кривизны $k(s)$ кривой, по которой пролегает нить:

$$E = \frac{1}{2}\int\limits_{t_0}^{t_n} c(s)k^2(s)\,ds, \tag{7.11}$$

где s — натуральный параметр на кривой, а $c(s)$ — коэффициент упругости нити. Мы будем считать, что $c(s) = c = \mathrm{const}$.

1) Предположим, что искомый сплайн задается графиком функции $y = y(x)$. Поскольку минимизация энергии упругих сил (7.11) является, вообще говоря, достаточно трудоемкой вычислительной задачей, мы заменим ее на

более простую, но достаточную для наших целей задачу. А именно, заменим энергию E, задаваемую формулой (7.11), на *псевдоупругую* энергию, заменив кривизну k на модуль второй производной $\left|\dfrac{d^2y}{dx^2}\right|$. Это приводит к задаче минимизации интеграла

$$E = \frac{1}{2}\mathrm{c}\sum_{i=0}^{n-1}\int\limits_{x_i}^{x_{i+1}} \left(y''(x)\right)^2 dx \tag{7.12}$$

за счет выбора величин $y'_0, \ldots, y'_n$, где $y'_i = y'(x_i)$. Условие стационарной точки целевой функции (7.12), получаемое приравниванием к нулю ее частных производных, имеет вид

$$\frac{\partial E}{\partial y'_j} = \mathrm{c}\sum_{i=0}^{n-1}\int\limits_{x_i}^{x_{i+1}} y''(x)\frac{\partial y''(x)}{\partial y'_j}\,dx = 0, \qquad j = 0, \ldots, n. \tag{7.13}$$

Пусть точки $\mathbf{p}_0, \ldots, \mathbf{p}_n$ на $\mathbb{R}^2$ имеют координаты $(x_0, y_0), \ldots, (x_n, y_n)$ соответственно. Будем считать, что $x_i < x_{i+1}$ для любого i, $0 \leqslant i \leqslant n-1$. Введем обозначения: $\Delta_i = x_{i+1} - x_i$. Тогда местный параметр w на участке (x_i, x_{i+1}) будет равен $w = \dfrac{x - x_i}{\Delta_i}$. На отрезке $[x_i, x_{i+1}]$ обозначим $y(x)$ через $y_{i,i+1}(x)$.

Согласно (7.8) на участке (x_i, x_{i+1}), $i = 0, \ldots, n-1$, зависимость $\widetilde{y}_{i,i+1}(w) = y_{i,i+1}\big(x(w)\big)$ имеет вид

$$\widetilde{y}_{i,i+1}(w) = y_i g_0(w) + y_{i+1}h_0(w) + y'_i\Delta_i g_1(w) + y'_{i+1}\Delta_i h_1(w), \tag{7.14}$$

поскольку $y'_i = \left.\dfrac{dy_{i,i+1}}{dx}\right|_{x=x_i} = \dfrac{1}{\Delta_i}\left.\dfrac{d\widetilde{y}_{i,i+1}}{dw}\right|_{w=0}$ и $y'_{i+1} = \left.\dfrac{dy_{i,i+1}}{dx}\right|_{x=x_{i+1}} = \dfrac{1}{\Delta_i}\left.\dfrac{d\widetilde{y}_{i,i+1}}{dw}\right|_{w=1}$.

Подставляя (7.14) в (7.13), интегрируя по частям, пользуясь явным видом функций $g_0(w)$, $h_0(w)$, $g_1(w)$, $h_1(w)$ из (7.9), а также тем, что $\dfrac{d^4y}{dx^4} = 0$, приходим к следующей системе уравнений:

$$\begin{cases} y''_{i-1,i}(x_i) = y''_{i,i+1}(x_i), \ i = 1, \ldots, n-1 \\ y''_{0,1}(x_0) = y''_{n-1,n}(x_n) = 0. \end{cases} \tag{7.15}$$

Которая обозначает, что вторые производные $y''(x)$, а следовательно и кривизна кривой, должны быть непрерывны в точках стыковки $\mathbf{p}_1, \ldots, \mathbf{p}_{n-1}$, а на концах y'' обращается в нуль.

Подставим теперь (7.14) в (7.15) и снова воспользуемся явным видом функций $g_0(w)$, $h_0(w)$, $g_1(w)$, $h_1(w)$ из (7.9). Используя равенства $y''_{i,i+1}(x_i) = \dfrac{1}{\Delta_i^2}\left.\dfrac{d\widetilde{y}_{i,i+1}}{dw}\right|_{w=0}$ и $y''_{i,i+1}(x_{i+1}) = \dfrac{1}{\Delta_i^2}\left.\dfrac{d\widetilde{y}_{i,i+1}}{dw}\right|_{w=1}$, получим систему уравнений

на y'_i с ленточной 3-шаговой матрицей:

$$\begin{cases} y'_{i-1}\dfrac{1}{\Delta_{i-1}} + 2y'_i\Big(\dfrac{1}{\Delta_{i-1}} + \dfrac{1}{\Delta_i}\Big) + y'_{i+1}\dfrac{1}{\Delta_i} = 3\dfrac{y_i - y_{i-1}}{\Delta_{i-1}^2} + 3\dfrac{y_{i+1} - y_i}{\Delta_i^2}, \\ 1 \leqslant i \leqslant n-1; \\ 2y'_0\Delta_0 + y'_1\Delta_0 = 3(y_1 - y_0) \text{ (граничное условие в точке } \mathbf{p}_0); \\ y'_{n-1}\Delta_{n-1} + 2y'_n\Delta_{n-1} = 3(y_n - y_{n-1}) \text{ (граничное условие в } \mathbf{p}_n). \end{cases}$$

Это — система из $(n+1)$-го уравнения с $(n+1)$ неизвестным $y'_0, \ldots, y'_n$. Поскольку матрица системы имеет ленточный вид, то она допускает численное решение за линейное по n время.

2) Теперь рассмотрим более общий случай параметрически заданного сплайна $\mathbf{r} = \mathbf{r}(t)$ (т.е. не обязательно в виде графика функции $y = y(x)$).

На участке между точками $\mathbf{p}_i$ и $\mathbf{p}_{i+1}$ обозначим искомый сплайн Эрмита порядка 3 через $\mathbf{r}_i(w)$ $(0 \leqslant i \leqslant n-1)$, где $0 \leqslant w \leqslant 1$ — местный параметр. Имеем, согласно (7.8):

$$\mathbf{r}_i(w) = \mathbf{p}_i g_0(w) + \mathbf{p}_{i+1} h_0(w) + \dot{\mathbf{r}}_i(0) g_1(w) + \dot{\mathbf{r}}_i(1) h_1(w), \qquad i = 0, \ldots, n-1. \tag{7.16}$$

К этому надо добавить условия непрерывности первых производных в точках стыковки:

$$\dot{\mathbf{r}}_{i-1}(1) = k_i \dot{\mathbf{r}}_i(0), \qquad i = 1, \ldots, n-1, \tag{7.17}$$

где $k_i = \dfrac{t_i - t_{i-1}}{t_{i+1} - t_i}$. Будем считать, что $\dfrac{|t_i - t_{i-1}|}{|t_{i+1} - t_i|} = \dfrac{|\mathbf{p}_i - \mathbf{p}_{i-1}|}{|\mathbf{p}_{i+1} - \mathbf{p}_i|}$, т.е. параметр t пропорционален натуральному. Тогда и местный параметр w будет также пропорционален натуральному. Обозначим $\mathbf{a}_i = \mathbf{p}_{i+1} - \mathbf{p}_i$, $|\mathbf{a}_i| = a_i$. Из пропорциональности параметра w натуральному следует, что $ds = a_i dw$ на участке (t_i, t_{i+1}). Следовательно, кривизна k пропорциональна $\dfrac{\ddot{\mathbf{r}}_i}{a_i^2}$, а потому псевдоупругая энергия сплайна Эрмита может быть определена как:

$$E = \sum_{i=0}^{n-1} \frac{1}{a_i^3} \int_0^1 \big(\ddot{\mathbf{r}}_i(w)\big)^2 \, dw. \tag{7.18}$$

При этом, согласно (7.16) и (7.17) $\ddot{\mathbf{r}}_i(w)$ выражается через $\dot{\mathbf{r}}_i(0)$ и $\dot{\mathbf{r}}_{i+1}(1)$. Условие стационарности (необходимое условие экстремума функции (7.18)) имеет вид

$$\begin{aligned} &\frac{\partial E}{\partial \dot{\mathbf{r}}_i(0)} = 0, \ i = 0, \ldots, n-1, \\ &\frac{\partial E}{\partial \dot{\mathbf{r}}_{n-1}(1)} = 0. \end{aligned} \tag{7.19}$$

Подставляя (7.16) и (7.17) в (7.19), приведем (7.19) к виду

$$\begin{cases} \ddot{\mathbf{r}}_{i-1}(1) = k_i^2 \ddot{\mathbf{r}}_i(0), \ i = 1, \ldots, n-1 \\ \ddot{\mathbf{r}}_0(0) = \ddot{\mathbf{r}}_{n-1}(1) = 0, \end{cases}$$

где $k_i = \dfrac{a_{i-1}}{a_i}$. Подставим сюда (7.16). Получим после алгебраических преобразований следующее выражение:

$$\begin{cases} \dot{\mathbf{r}}_{i-1} + 2\dot{\mathbf{r}}_i(k_i + k_i^2) + \dot{\mathbf{r}}_{i+1} k_i^2 k_{i+1} = \\ = 3(\mathbf{p}_i - \mathbf{p}_{i-1}) + 3(\mathbf{p}_{i+1} - \mathbf{p}_i)k_i^2, \ 1 \leqslant i \leqslant n-1, \\ 2\dot{\mathbf{r}}_0 + \dot{\mathbf{r}}_1 k_1 = 3(\mathbf{p}_1 - \mathbf{p}_0), \\ \dot{\mathbf{r}}_{n-1} + 2\dot{\mathbf{r}}_n = 3(\mathbf{p}_n - \mathbf{p}_{n-1}), \end{cases} \tag{7.20}$$

где $\dot{\mathbf{r}}_i = \dot{\mathbf{r}}_i(0)$, $i = 1, \ldots, n-1$, $\dot{\mathbf{r}}_n = \dot{\mathbf{r}}_{n-1}(1)$. Эта система состоит из $(n+1)$ уравнений с $(n+1)$ неизвестными $\dot{\mathbf{r}}_0, \ldots, \dot{\mathbf{r}}_n$. Как и выше, матрица системы является ленточной, т.е. ненулевые элементы в ней расположены лишь в узкой окрестности главной диагонали. Такие системы уравнений допускают численное решение за линейное время с помощью хорошо разработанных алгоритмов.

7.1.4 Случай, когда на концах кривой заданы направления касательных векторов

Иногда направления касательных векторов на концах кривой заданы заранее. Например — в тех случаях, когда искомая кривая (составной сплайн Эрмита) должна гладко стыковаться на концах с другими, уже построенными кривыми. В этом случае вместо уравнений $\dfrac{\partial E}{\partial \dot{\mathbf{r}}_0} = 0$, $\dfrac{\partial E}{\partial \dot{\mathbf{r}}_n} = 0$ необходимо рассматривать условия

$$\left\langle \dot{\mathbf{r}}_0, \frac{\partial E}{\partial \dot{\mathbf{r}}_0} \right\rangle = 0, \qquad \left\langle \dot{\mathbf{r}}_n, \frac{\partial E}{\partial \dot{\mathbf{r}}_n} \right\rangle = 0, \tag{7.21}$$

поскольку теперь $\dot{\mathbf{r}}_0$ и $\dot{\mathbf{r}}_n$ уже не являются свободными параметрами. Их направления заданы, поэтому производные надо брать по заданным направлениям. Из (7.20) и (7.21) следует, что

$$\begin{cases} 2\langle \dot{\mathbf{r}}_0, \dot{\mathbf{r}}_0 \rangle + \langle \dot{\mathbf{r}}_0, \dot{\mathbf{r}}_1 \rangle k_1 = 3\langle \dot{\mathbf{r}}_0, \mathbf{p}_1 - \mathbf{p}_0 \rangle, \\ \langle \dot{\mathbf{r}}_n, \dot{\mathbf{r}}_{n-1} \rangle + 2\langle \dot{\mathbf{r}}_n, \dot{\mathbf{r}}_n \rangle = 3\langle \dot{\mathbf{r}}_n, \mathbf{p}_n - \mathbf{p}_{n-1} \rangle. \end{cases}$$

Эти уравнения надо поставить в (7.20) вместо последних двух уравнений, соответствующих крайним точкам $i = 0$ и $i = n$.

Замечание. В случае, когда опорные точки $\mathbf{p}_0, \ldots, \mathbf{p}_n$ равноудалены друг от друга, т.е. $|\mathbf{p}_{i+1} - \mathbf{p}_i| = |\mathbf{p}_i - \mathbf{p}_{i-1}|$ для любого $i = 1, \ldots, n-1$, можно положить $k_i = 1$ для каждого $i = 1, \ldots, n-1$. Однако, если точки расположены неравномерно, расчет значений k_i обязателен, так как кривая с одинаковыми k_i будет, вообще говоря, плохой. Описанный прием построения псевдоупругих сплайнов может быть применен и для кривых Безье.

7.1.5 Кубические сплайны. Построение кубического сплайна

Пусть дана $(n+1)$ точка $\mathbf{p}_0, \ldots, \mathbf{p}_n \in \mathbb{R}^2$. Требуется провести через эти точки кривую класса C^2, т.е. имеющую непрерывные первые и вторые производные радиус-вектора по параметру: $\mathbf{r} = \mathbf{r}(t)$, $\mathbf{r}(t_i) = \mathbf{p}_i$, $i = 0, \ldots, n$, и $\dot{\mathbf{r}}(t)$, $\ddot{\mathbf{r}}(t)$ — непрерывны по t. Решение этой задачи можно дать в виде кубического сплайна. На каждом участке $[t_i, t_{i+1}]$ такой сплайн описывается полиномом третьей степени по t. При полиномиальной интерполяции степень полинома в данной задаче не может быть, вообще говоря, сделана меньше 3. Поэтому кубический сплайн имеет наименьшую степень в данной постановки задачи и определяется однозначно.

Пусть искомый сплайн задается параметрическим уравнением $\mathbf{r} = \mathbf{r}(t)$. Обозначим $\mathbf{s}_i = \ddot{\mathbf{r}}(t_i)$, $i = 0, \ldots, n$. Поскольку сплайн является кубическим, вторая производная радиус-вектора будет линейной функцией на каждом отрезке $[t_i, t_{i+1}]$:

$$\frac{d^2\mathbf{r}}{dt^2} = \mathbf{s}_i \frac{t_{i+1} - t}{t_{i+1} - t_i} + \mathbf{s}_{i+1} \frac{t - t_i}{t_{i+1} - t_i}. \tag{7.22}$$

Проинтегрируем (7.22) два раза. Получим

$$\mathbf{r}(t) = \mathbf{s}_i \frac{(t_{i+1} - t)^3}{6(t_{i+1} - t_i)} + \mathbf{s}_{i+1} \frac{(t - t_i)^3}{6(t_{i+1} - t_i)} + \mathbf{c}_1 t + \mathbf{c}_2. \tag{7.23}$$

Постоянные $\mathbf{c}_1$ и $\mathbf{c}_2$ определим из условий, заданных на концах данного участка кривой:

$$\begin{cases} \mathbf{r}(t_i) = \mathbf{p}_i, \\ \mathbf{r}(t_{i+1}) = \mathbf{p}_{i+1}. \end{cases} \tag{7.24}$$

Подстановка (7.23) в (7.24) и вычисление $\mathbf{c}_1$, $\mathbf{c}_2$ приводит к следующему ответу:

$$\begin{aligned} \mathbf{r}(t) = \mathbf{s}_i \frac{(t_{i+1} - t)^3}{6(t_{i+1} - t_i)} + \mathbf{s}_{i+1} \frac{(t - t_i)^3}{6(t_{i+1} - t_i)} + \\ + \Big(\frac{\mathbf{p}_i}{t_{i+1} - t_i} - \mathbf{s}_i \frac{t_{i+1} - t_i}{6} \Big)(t_{i+1} - t) + \\ + \Big(\frac{\mathbf{p}_{i+1}}{t_{i+1} - t_i} - \mathbf{s}_{i+1} \frac{t_{i+1} - t_i}{6} \Big)(t - t_i). \end{aligned} \tag{7.25}$$

Правая часть (7.25) является кубическим полином по t на отрезке $[t_i, t_{i+1}]$. Она содержит два неизвестных векторных параметра $\mathbf{s}_i$ и $\mathbf{s}_{i+1}$. Их можно определить из условия согласованности первых производных на концах отрезков $[t_i, t_{i+1}]$. Пока мы еще не использовали этого условия (вторые производные должны быть согласованы в силу определения $\mathbf{s}_i$). На левом конце отрезка $[t_i, t_{i+1}]$ мы имеем

$$\frac{d\mathbf{r}}{dt}\Big|_{t=t_i} = -\frac{(2\mathbf{s}_i + \mathbf{s}_{i+1})(t_{i+1} - t_i)}{6} + \frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{t_{i+1} - t_i}. \tag{7.26}$$

Аналогично, на правом конце отрезка $[t_{i-1}, t_i]$, т.е. в той же точке t_i, имеем

$$\left.\frac{d\mathbf{r}}{dt}\right|_{t=t_i} = \frac{(2\mathbf{s}_i + \mathbf{s}_{i-1})(t_i - t_{i-1})}{6} + \frac{\mathbf{p}_i - \mathbf{p}_{i-1}}{t_i - t_{i-1}}. \tag{7.27}$$

Приравнивая правые части (7.26) и (7.27) получим $(n-1)$ уравнение для внутренних точек $i = 1, \ldots, n-1$:

$$\begin{aligned}\mathbf{s}_{i-1}(t_i - t_{i-1}) + 2\mathbf{s}_i(t_{i+1} - t_{i-1}) + \mathbf{s}_{i+1}(t_{i+1} - t_i) = \\ = 6\frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{t_{i+1} - t_i} - 6\frac{\mathbf{p}_i - \mathbf{p}_{i-1}}{t_i - t_{i-1}}, \qquad i = 1, \ldots, n-1.\end{aligned} \tag{7.28}$$

Система (7.28) представляет собой систему из $(n-1)$ векторного уравнения (или, что то же самое, $2(n-1)$ скалярного уравнения) с ленточной трех-диагональной матрицей и $(n+1)$ векторным неизвестным $\mathbf{s}_0, \ldots, \mathbf{s}_n$ (что соответствует $2(n+1)$ скалярных неизвестных). Если кривая замкнута, то $\mathbf{s}_0 = \mathbf{s}_n$ и в систему (7.28) добавляется еще одно уравнение для $i = n$:

$$\mathbf{s}_{n+1} = \mathbf{s}_1.$$

Таким образом, для замкнутой кривой число неизвестных равно числу уравнений и система имеет единственное решение.

Если кривая незамкнута, то необходимо задать граничные условия. Например, если рассматривать сплайны как гибкую нить, концы которой не имеют нагрузки на изгиб, получим такое условие:

$$\mathbf{s}_0 = \mathbf{s}_n = 0. \tag{7.29}$$

Еще один возможный вариант граничных условий:

$$\mathbf{s}_0 = \mathbf{s}_1, \ \mathbf{s}_n = \mathbf{s}_{n-1}. \tag{7.30}$$

При этом концевые участки кривой будут иметь постоянную кривизну.

Определив из системы (7.28) и граничных условий вида (7.29) или (7.30) (или каких-либо других, следующих из постановки задачи), получим формулу для искомого кубического сплайна на участке $[t_i, t_{i+1}]$:

$$\begin{aligned}\mathbf{r}(t) = (1-w)\mathbf{p}_i + w\mathbf{p}_{i+1} + \\ + \left((-2w + 3w^2 - w^3)\mathbf{s}_i + (-w + w^3)\mathbf{s}_{i+1}\right)\frac{(t_{i+1} - t_i)^2}{6},\end{aligned}$$

где $w = \dfrac{t - t_i}{t_{i+1} - t_i}$ — местный параметр на участке $[t_i, t_{i+1}]$.

Замечания. 1) В отличии от составного сплайна Эрмита, для кубического сплайна изменение положения одной из опорных точек всегда приводит к перевычислению всего сплайна.

2) Если взять равномерную параметризацию, т.е. положить $t_i = i$, $i = 0, \ldots, n$, то уравнения (7.28) примут совсем простой вид

$$\mathbf{s}_{i-1} + 4\mathbf{s}_i + \mathbf{s}_{i+1} = 6(\mathbf{p}_{i+1} - 2\mathbf{p}_i + \mathbf{p}_{i-1}), \quad i = 1, \ldots, n-1.$$

Однако равномерная параметризация в случае неравномерного расположения опорных точек может привести к появлению необоснованных изгибов сплайна. Для того, чтобы сплайн не имел подобных изгибов, параметризацию следует выбирать приблизительно натуральной. Для этого достаточно потребовать выполнения приблизительного равенства:

$$\frac{|\mathbf{p}_{i+1} - \mathbf{p}_i|}{t_{i+1} - t_i} \approx \frac{|\mathbf{p}_i - \mathbf{p}_{i-1}|}{t_i - t_{i-1}}.$$

7.1.6 Сплайн Лагранжа

Пусть дана $(n+1)$ опорная точка $\mathbf{p}_i = (x_i, y_i) \in \mathbb{R}^2$, $0 \leqslant i \leqslant n$. Требуется найти полиномиальную по t кривую $\mathbf{r} = \mathbf{r}(t) = P^n(t)$ такую, что $\mathbf{r}(t_i) = \mathbf{p}_i$, $0 \leqslant i \leqslant n$, при заданных значениях параметра $t_0 < \ldots < t_n$. Будем искать решение этой задачи в виде

$$\mathbf{r}(t) = \sum_{i=0}^{n} L(t_i)\mathbf{p}_i,$$

где

$$L_i(t_j) = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \qquad 0 \leqslant i, j \leqslant n.$$

Таким свойствами обладают многочлены

$$L_i(t) = \frac{(t-t_0)\ldots(t-t_{i-1})(t-t_{i+1})\ldots(t-t_n)}{(t_i-t_0)\ldots(t_i-t_{i-1})(t_i-t_{i+1})\ldots(t_i-t_n)} =$$

$$= \frac{\prod\limits_{j=0, j\neq i}^{n}(t-t_j)}{\prod\limits_{j=0, j\neq i}^{n}(t_i-t_j)}.$$

С их помощью получаем *сплайн Лагранжа*

$$\mathbf{r}(t) = \sum_{i=0}^{n} \frac{\prod\limits_{j=0, j\neq i}^{n}(t-t_j)}{\prod\limits_{j=0, j\neq i}^{n}(t_i-t_j)}\mathbf{p}_i$$

или

$$\mathbf{r}(t) = \sum_{i=0}^{n} \frac{w_{0,n}(t)}{(t-t_i)w'_{0,n}(t_i)}\mathbf{p}_i,$$

Рис. 7.2:

где

$$w_{0,n}(t) = (t - t_0)\dots(t - t_n),$$

$$w'_{0,n}(t_i) = \frac{dw_{0,n}}{dt}\Big|_{t=t_i} = (t_i - t_0)\dots(t_i - t_{i-1})(t_i - t_{i+1})\dots(t_i - t_n).$$

На рис. 7.2 приведены кубический сплайн (1), сплайн Эрмита (2) и сплайн Лагранжа (3), построенные по одним и тем же восьми точкам.

7.1.7 Сплайн Ньютона

Сплайн Ньютона представляет собой тот же самый сплайн Лагранжа, но в другой форме, удобной для перевычисления сплайна в случае добавления новых опорных точек.

А именно, сплайн Ньютона записывается в виде

$$\mathbf{r}(t) = \mathbf{a}_0 + \mathbf{a}_1(t - t_0) + \mathbf{a}_2(t - t_0)(t - t_1) + \dots + \mathbf{a}_n(t - t_0)\dots(t - t_{n-1}),$$

где векторные коэффициенты $\mathbf{a}_0, \dots, \mathbf{a}_n$ ищутся из условий прохождения сплайна через точки $\mathbf{p}_0, \dots, \mathbf{p}_n$. При этом точки $\mathbf{p}_0, \dots, \mathbf{p}_n$ полностью определяют коэффициенты $\mathbf{a}_0, \dots, \mathbf{a}_k$, $0 \leqslant k \leqslant n$, с помощью системы уравнений

$$\begin{cases} \mathbf{p}_0 = \mathbf{a}_0, \\ \mathbf{p}_1 = \mathbf{a}_0 + \mathbf{a}_1(t_1 - t_0), \\ \mathbf{p}_2 = \mathbf{a}_0 + \mathbf{a}_1(t_2 - t_0) + \mathbf{a}_2(t_2 - t_0)(t_2 - t_1), \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \mathbf{p}_n = \mathbf{a}_0 + \mathbf{a}_1(t_n - t_0) + \mathbf{a}_2(t_n - t_0)(t_n - t_1) + \dots \\ \dots + \mathbf{a}_n(t_n - t_0)\dots(t_n - t_{n-1}). \end{cases}$$

Последняя система решается за один проход, так как матрица системы является нижнетреугольной. Из первых k уравнений определяются первые k неизвестные $\mathbf{a}_0, \dots, \mathbf{a}_{k-1}$. Поэтому, если мы добавим еще одну точку $\mathbf{p}_{n+1}$, нам будет нужно лишь определить еще один коэффициент $\mathbf{a}_{n+1}$ из еще одного, добавленного уравнения. Коэффициенты $\mathbf{a}_0, \dots, \mathbf{a}_n$ при этом останутся прежними.

Утверждение 7.2. *Пусть дана $(n+1)$ различная опорная точка $\mathbf{p}_0, \ldots, \mathbf{p}_n$. Тогда для любого $k \geqslant 3$ существует единственный кусочно-полиномиальный сплайн порядка k, принадлежащий классу C^{k-1} и удовлетворяющий $(k-1)$ векторным граничным условиям общего положения (т.е. независимым от остальных условий задачи).*

Доказательство. На отрезке $[t_i, t_{i+1}]$ функция $\mathbf{r} = \mathbf{r}(t)$ имеет вид полинома степени k с векторными коэффициентами. Следовательно, ее производная $(k-1)$ порядка $\mathbf{r}^{(k-1)}(t)$ является линейной функцией на этом отрезке, и поэтому $\mathbf{r}^{(k-1)}(t)$ полностью определяется своими значениями на концах отрезка.

Обозначим $\mathbf{s}_i = \mathbf{r}^{(k-1)}(t_i)$, $i = 0, \ldots, n$. Тогда для любого отрезка $[t_i, t_{i+1}]$ производная $\mathbf{r}^{(k-1)}(t)$ выражается через $\mathbf{s}_i$ и $\mathbf{s}_{i+1}$. Интегрируя это выражение $(k-1)$ раз, получим на отрезке $[t_i, t_{i+1}]$ зависимость $\mathbf{r}(t) = f(\mathbf{s}_i, \mathbf{s}_{i+1}, \mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{k-1}, t)$, где $\mathbf{c}_1, \ldots, \mathbf{c}_{k-1}$ — неопределенные постоянные, возникшие при интегрировании. Из условий

$$\begin{cases} \mathbf{r}(t_i) = \mathbf{p}_i, \\ \mathbf{r}(t_{i+1}) = \mathbf{p}_{i+1} \end{cases}$$

мы можем определить две постоянные из множества $\{\mathbf{c}_1, \ldots, \mathbf{c}_{k-1}\}$. Остается $(k-3)$ постоянных для каждого из n отрезков $[t_0, t_1]$, ..., $[t_{n-1}, t_n]$. Из условий согласования первых $(k-2)$ производных в точках $t_1, \ldots, t_{n-1}$ (т.е. во внутренних точках) получаем $(n-1)$ векторное уравнение для каждой из упомянутых $(k-2)$ производных. Итого, получается $(n-1)(k-2)$ векторных уравнений. Число неизвестных равно $(n+1) + n(k-3)$, поскольку имеется $(n+1)$ неизвестное $\mathbf{s}_i$, $i = 0, \ldots, n$, и $n(k-3)$ неопределенных постоянных. Получаем, что число неизвестных минус число уравнений равно $(k-1)$. Следовательно, система будет иметь единственное решение, если к ней добавить $(k-1)$ независимых граничных условий. Утверждение доказано. ■

Утверждение 7.3. *Пусть дана $(n+1)$ опорная точка $\mathbf{p}_0, \ldots, \mathbf{p}_n$. Тогда кусочно-полиномиальный сплайн степени k класса гладкости C^k существует, вообще говоря, лишь при $k \geqslant n$. При этом, он будет полиномом на всем отрезке $t_0 \leqslant t \leqslant t_n$ и, следовательно, будет принадлежать классу C^∞. При $k = n$ такой сплайн единственен.*

Доказательство. Кусочно-полиномиальная функция класса гладкости C^k, где k — степень составляющих ее полиномов, является полиномом на всей оси, так как ряд Тейлора полинома степени k совпадает с ним самим и однозначно определяется его первыми k производными в любой точке. Если $k = n$, то для определения $(n+1)$ векторного коэффициента такого полинома $\mathbf{r}(t) = \mathbf{a}_0 + \mathbf{a}_1 t + \cdots + \mathbf{a}_n t^n$ мы имеем $(n+1)$ векторное уравнение: $\mathbf{r}(t_i) = \mathbf{p}_i$, $i = 0, \ldots, n$. Эта система имеет единственное решение — полином Лагранжа. При $k < n$ число уравнений больше числа неизвестных, поэтому система $\mathbf{r}(t_i) = \mathbf{p}_i$, $i = 0, \ldots, n$, вообще говоря, не будет иметь решений. Утверждение доказано. ■

7.2 Кривые Безье

Поставим следующую задачу. Для любого набора из $(n+1)$ опорных точек $\mathbf{p}_0, \ldots, \mathbf{p}_n \in \mathbb{R}^3$, требуется найти такие функции $f_0(t), \ldots, f_n(t)$, не зависящие от $\mathbf{p}_0, \ldots, \mathbf{p}_n$ и имеющие вид полиномов по t, что кривая

$$\mathbf{r}(t) = \mathbf{p}_0 f_0(t) + \cdots + \mathbf{p}_n f_n(t), \qquad t \in [0,1], \tag{7.31}$$

обладает следующими свойствами:

1) $\mathbf{r}(0) = \mathbf{p}_0$, $\mathbf{r}(1) = \mathbf{p}_n$;

2) кривая (7.31) не зависит от выбора декартовой системы координат в $\mathbb{R}^3$ или, другим словами, формула (7.31) инвариантна относительно движений пространства;

3) если все опорные точки $\mathbf{p}_0, \ldots, \mathbf{p}_n$ лежат в одной плоскости, то кривая (7.31) плоская и лежит внутри выпуклой оболочки опорных точек $\mathbf{p}_0, \ldots, \mathbf{p}_n$.

Утверждение 7.4. *Условия* 1) – 3) *равносильны условиям*:

a) $\sum\limits_{i=0}^{n} f_i(t) = 1$ *для любой точке* $t \in [0,1]$;

b) $f_i(t) \geqslant 0$ *для любого* $t \in [0,1]$;

c) $f_0(0) = 1$, $f_i(0) = 0$ *для любого* $1 \leqslant i \leqslant n$ *и* $f_n(1) = 1$, $f_i(1) = 0$ *для любого* $0 \leqslant i \leqslant n-1$.

Доказательство. Пусть выполнены условия 1) – 3).

a) Если $\mathbf{p}_0 = \mathbf{p}_1 = \cdots = \mathbf{p}_n$, то из 3) следует, что $\mathbf{r}(t) = \mathbf{p}_0$, $t \in [0,1]$. Поэтому $\sum\limits_{i=0}^{n} f_i(t) \equiv 1$. (Здесь мы воспользовались свойством независимости функций $f_0(t), \ldots, f_n(t)$ от $\mathbf{p}_0, \ldots, \mathbf{p}_n$.)

b) Докажем от противного. Пусть $f_j(t^*) < 0$ для некоторого $t^* \in [0,1]$. Выберем опорные точки $\mathbf{p}_0, \ldots, \mathbf{p}_{j-1}, \mathbf{p}_j, \ldots, \mathbf{p}_n$ на оси Ox, а $\mathbf{p}_j$ — в верхней полуплоскости. Имеем: $\mathbf{r}(t) = \big(x(t), y(t)\big)$, $\mathbf{p}_i = (x_i, y_i)$. Следовательно, $y_i = 0$, если $i \neq j$, $y_j > 0$. Для второй компоненты радиус-вектора кривой в точке t^* имеем: $r^2(t^*) = y(t^*) = \sum\limits_{i=0}^{n} f_i(t^*) y_i = f_j(t^*) y_j < 0$, поскольку $f_j(t^*) < 0$ по предположению. Следовательно, $y(t^*) < 0$, поэтому кривая $\mathbf{r}(t)$ заходит в нижнюю полуплоскость $y < 0$ и не лежит в выпуклой оболочке точек $\mathbf{p}_0, \ldots, \mathbf{p}_n$, которая находится в верхней полуплоскости (здесь мы воспользовались свойством 2) независимости кривой от выбора системы координат).

c) Следует из 1), если взять точки $\mathbf{p}_0$ и $\mathbf{p}_n$ различными, а точки $\mathbf{p}_1, \ldots, \mathbf{p}_{n-1}$ — совпадающими с началом координат.

Пусть теперь выполнены условия *a)* – *c)*.

1) следует из *c)*.

2) Пусть дана кривая вида $\mathbf{r}(t) = \sum\limits_{i=0}^{n} f_i(t)\mathbf{p}_i$. Если A — движение пространства, то оно по теореме о группе движений пространства представимо в виде композиции: $A = O \circ L$, где O — вращение, L — сдвиг. Имеем $O\big(\mathbf{r}(t)\big) = O\big(\sum\limits_{i=0}^{n} f_i(t)\mathbf{p}_i\big) = \sum\limits_{i=0}^{n} f_i(t)O(\mathbf{p}_i)$, поскольку вращение — линейное преобразование. Далее, $L\big(\mathbf{r}(t)\big) = \mathbf{r}(t) + \mathbf{b}$ для некоторого $\mathbf{b} \in \mathbb{R}^3$, $\mathbf{b} = \mathrm{const}$. Следовательно, из *a*) получаем: $L\big(\mathbf{r}(t)\big) = \sum\limits_{i=0}^{n} f_i(t)\mathbf{p}_i + \mathbf{b} = \sum\limits_{i=0}^{n} f_i(t)\mathbf{p}_i + \sum\limits_{i=0}^{n} f_i(t)\mathbf{b} = \sum\limits_{i=0}^{n} f_i(t)(\mathbf{p}_i + \mathbf{b}) = \sum\limits_{i=0}^{n} f_i(t)L(\mathbf{p}_i)$ (поскольку $\sum\limits_{i=0}^{n} f_i(t) = 1$). Следовательно, кривая $\mathbf{r}(t) = \sum\limits_{i=0}^{n} f_i(t)\mathbf{p}_i$ не изменится, если при ее вычислении воспользоваться другой декартовой системой координат в пространстве.

3) Выпуклая оболочка точек $\mathbf{p}_0, \ldots, \mathbf{p}_n$ — это выпуклый многоугольник, лежащий по одну сторону от прямой, продолжающей любую его сторону. Обратно, область пересечения соответствующих полуплоскостей, образованных этими прямыми, совпадает с выпуклой оболочкой $\mathbf{p}_0, \ldots, \mathbf{p}_n$.

Возьмем одну из сторон этого многоугольника и выберем ось Ox, совпадающей с ней, а ось Oy — направленной в ту полуплоскость, где лежит многоугольник. Тогда если $\mathbf{p}_i = (x_i, y_i)$, то $y_i \geqslant 0$ для любого $0 \leqslant i \leqslant n$. Обозначим $\mathbf{r}(t) = \big(x(t), y(t)\big)$. Имеем

$$y(t) = \sum_{i=0}^{n} f_i(t)y_i \geqslant 0,$$

поскольку $f_i(t) \geqslant 0$ согласно *c*), а $y_i \geqslant 0$ по построению ($0 \leqslant i \leqslant n$). Следовательно, кривая $\mathbf{r}(t)$ целиком лежит в соответствующей полуплоскости $y \geqslant 0$. Так как это верно для любой стороны многоугольника — выпуклой оболочки точек $\mathbf{p}_0, \ldots, \mathbf{p}_n$, — то $\mathbf{r}(t)$ целиком принадлежит этой выпуклой оболочке. Утверждение доказано. ■

Определение 7.2. Фиксируем целое $n > 0$. *Базисом Бернштейна* на отрезке $[0, 1]$ называется система из $(n + 1)$ полинома степени n, задаваемых формулами

$$B_i^n(t) = \frac{n!}{i!(n-i)!}t^i(1-t)^{n-i} = C_n^i t^i (1-t)^{n-i}, \qquad 0 \leqslant i \leqslant n.$$

На рис. 7.3 приведены полиномы Бернштейна третьей и четвертой степени.

Утверждение 7.5. *Базис Бернштейна удовлетворяет свойствам a) – c), сформулированным в утверждении* 7.4.

Доказательство. *a*) Согласно биному Ньютона имеем $(a+b)^n = \sum\limits_{k=0}^{n} C_n^k a^k b^{n-k}$.

Рис. 7.3:

Возьмем здесь $a = t$, $b = 1 - t$. Получим

$$1 = 1^n = (t + 1 - t)^n = \sum_{k=0}^{n} C_n^k t^k (1-t)^{n-k} = \sum_{k=0}^{n} B_k^n(t).$$

Следовательно, $\sum\limits_{k=0}^{n} B_k^n(t) = 1$.

b) Условие $B_i^n(t) \geqslant 0$ на отрезке $[0, 1]$ очевидно.

c) $B_0^n(0) = C_n^0 0^0 (1-0)^0 = 1$, $B_i^n(0) = C_n^i 0^i (1-0)^{n-i} = 0$, если $i > 0$. Аналогично доказывается, что $B_n^n(1) = 1$, $B_i^n(1) = 0$, если $i < n$. Утверждение доказано. ■

Определение 7.3. Пусть дана $(n+1)$ опорная точка $\mathbf{p}_0, \ldots, \mathbf{p}_n$. Тогда *кривой Безье, определяемой точками* $\mathbf{p}_0, \ldots, \mathbf{p}_n$, называется кривая

$$\mathbf{r}_B(t) = \sum_{i=0}^{n} B_i^n(t)\mathbf{p}_i, \qquad t \in [0, 1].$$

Из утверждений 7.4 и 7.5 следует, что кривая Безье является решением поставленной выше задачи (разумеется — не единственным).

Лемма 7.1. *Полиномы Бернштейна $B_i^n(t)$ удовлетворяют соотношению*

$$B_i^n(t) = tB_{i-1}^{n-1}(t) + (1-t)B_i^{n-1}(t)$$

для каждого $t \in [0, 1]$ и $n \geqslant 1$, или, опуская аргумент t,

$$B_i^n = tB_{i-1}^{n-1} + (1-t)B_i^{n-1}. \tag{7.32}$$

Доказательство. Имеем

$$tB_{i-1}^{n-1} + (1-t)B_i^{n-1} = \frac{i}{n}\frac{n!}{i!(n-i)!}t^i(1-t)^{n-i} + \\ + \frac{n-i}{n}\frac{n!}{i!(n-i)!}t^i(1-t)^{n-i} = B_i^n.$$

Лемма доказана. ■

Рис. 7.4:

Следствие 7.1. *Полиномы Бернштейна для любых n, i могут быть вычислены рекуррентно с помощью соотношения* (7.32) *и формул, принимаемых по определению*: $B_0^0(t) \equiv 1$; $B_i^n(t) = 0$, *если* $i < 0$ *или* $i > n$.

Утверждение 7.6. *Касательные к кривой Безье в точках* $\mathbf{p}_0$ *и* $\mathbf{p}_n$ *совпадают с направлением векторов* $\mathbf{p}_1 - \mathbf{p}_0$ *и* $\mathbf{p}_n - \mathbf{p}_{n-1}$ *соответственно.*

Доказательство. Имеем $\mathbf{r}'(t) = \sum\limits_{i=0}^{n} \big(B_i^n(t)\big)'\mathbf{p}_i$. Вычислим $\big(B_0^n(t)\big)'|_{t=0}$ и $\big(B_1^1(t)\big)'|_{t=0}$:

$$\big(B_0^n(t)\big)'|_{t=0} = \big(C_n^0 t^0 (1-t)^n\big)'\big|_{t=0} = \big((1-t)^n\big)'\big|_{t=0} = -n,$$
$$\big(B_1^1(t)\big)'|_{t=0} = \big(nt(1-t)^{n-1}\big)'\big|_{t=0} = n.$$

Следовательно, $\mathbf{r}'(0) = n(\mathbf{p}_1 - \mathbf{p}_0)$, т.е. параллельно $\mathbf{p}_1 - \mathbf{p}_0$.

Аналогично, $\mathbf{r}'(1) = n(\mathbf{p}_n - \mathbf{p}_{n-1})$, т.е. параллельно $\mathbf{p}_n - \mathbf{p}_{n-1}$. Утверждение доказано. ■

Следствие 7.2. *Кривая Безье касается своей опорной ломаной* (*характеристической ломаной*) *в крайних точках.*

Пример 3 (Кубическая кривая Безье). Если заданы четыре опорные точки $\mathbf{p}_0$, $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, то кривая Безье, построенная по этим четырем точкам, имеет вид: $\mathbf{r}(t) = (1-t)^3\mathbf{p}_0 + 3t(1-t)^2\mathbf{p}_1 + 3t^2(1-t)\mathbf{p}_2 + t^3\mathbf{p}_3$, см. рис. 7.4.

Пример 4 (Квадратичная кривая Безье). Если заданы три опорные точки $\mathbf{p}_0$, $\mathbf{p}_1$, $\mathbf{p}_2$, то кривая Безье имеет вид: $\mathbf{r}(t) = (1-t)^2\mathbf{p}_0 + 2t(1-t)\mathbf{p}_1 + t^2\mathbf{p}_2$. Другое представление той же кривой: $\mathbf{r}(t) = (1-t)\big(\mathbf{p}_1 + (\mathbf{p}_0 - \mathbf{p}_1)(1-t)\big) + t\big(\mathbf{p}_1 + (\mathbf{p}_2 - \mathbf{p}_1)t\big)$, см. рис. 7.5.

7.2.1 Алгоритм де Кастелье

Алгоритм де Кастелье позволяет построить кривую Безье с помощью реккурентного соотношения (7.32).

Рис. 7.5:

Пусть даны опорные точки $\mathbf{p}_0, \ldots, \mathbf{p}_n$. Обозначим через $\mathbf{r}_i(t;k)$ кривую Безье, построенную по точкам $\mathbf{p}_i, \ldots, \mathbf{p}_{i+k}$ $(i+k \leqslant n)$.

Тогда искомая кривая Безье — это кривая $\mathbf{r}_0(t;n)$. Имеем $\mathbf{r}_i(t;0) \equiv \mathbf{p}_i$, $0 \leqslant i \leqslant n$. Выведем реккурентные формулы для определения $\mathbf{r}_0(t;n) = \mathbf{r}(t)$ исходя из начальных данных: $\mathbf{r}_i(t;0) \equiv \mathbf{p}_i$ $(0 \leqslant i \leqslant n)$. Имеем

$$\mathbf{r}(t) = \mathbf{p}_0(1-t)^n + \mathbf{p}_n t^n + \sum_{i=1}^{n-1} \big(tB_{i-1}^{n-1} + (1-t)B_i^{n-1}\big)\mathbf{p}_i =$$

$$= (1-t)\big(\mathbf{p}_0(1-t)^{n-1} + \sum_{i=1}^{n-1} B_i^{n-1}\mathbf{p}_i\big) + t\big(\mathbf{p}_n t^{n-1} + \sum_{i=1}^{n-1} B_{i-1}^{n-1}\mathbf{p}_i\big) =$$

$$= (1-t)\sum_{i=0}^{n-1} B_i^{n-1}\mathbf{p}_i + t\sum_{i=0}^{n-1} B_i^{n-1}\mathbf{p}_{i+1} = (1-t)\mathbf{r}_0(t;n-1) + t\mathbf{r}_1(t;n-1).$$

Отсюда получаем следующую реккурентную формулу де Кастелье для вычисления кривой Безье:

$$\mathbf{r}_i(t;k) = (1-t)\mathbf{r}_i(t;k-1) + t\mathbf{r}_{i+1}(t;k-1), \tag{7.33}$$

где $i+k \leqslant n$.

По формуле (7.33), начиная от значений $\mathbf{r}_i(t;0) = \mathbf{p}_i$ получаем последовательно значения $\mathbf{r}_i(t;k)$, $1 \leqslant k \leqslant n$. На последнем шаге вычисляем $\mathbf{r}_0(t;n) = \mathbf{r}(t)$. Это и есть алгоритм де Кастелье.

7.2.2 Операторная форма кривой Безье

Пусть даны опорные точки $\mathbf{p}_0, \ldots, \mathbf{p}_n$ и параметр $t \in [0,1]$. Определим оператор *правого сдвига* E, действующий на кривых вида $\mathbf{r}(t) = \sum\limits_{i=0}^{n-1} f_i(t)\mathbf{p}_i$, где $f_i(t)$, $i = 0, \ldots, n-1$, — гладкие функции от $t \in [0,1]$, по формуле:

$$E\mathbf{r}(t) = \sum_{i=0}^{n-1} f_i(t)\mathbf{p}_{i+1}.$$

Аналогично, на кривых вида $\mathbf{r}(t) = \sum\limits_{i=1}^{n} g_i(t)\mathbf{p}_i$, $t \in [0,1]$, определен оператор *левого сдвига* L:

$$L\mathbf{r}(t) = \sum_{i=1}^{n} g_i(t)\mathbf{p}_{i-1}.$$

Тогда на множестве кривых для которых определен оператор $L \circ E$ или $E \circ L$ имеем: $L \circ E = \mathrm{Id}$, $E \circ L = \mathrm{Id}$. Другими словами $L = E^{-1}$, $E = L^{-1}$. Мы будем пользоваться также степенями операторов L и E: $L^k = L \circ \ldots \circ L$, $E^k = E \circ \ldots \circ E$ (где L и E повторяются k раз), применяя их к тем кривым, для которых эти степени определены.

Кривую $\mathbf{r}(t) \equiv \mathbf{p}_i$ мы будем обозначать просто $\mathbf{p}_i$.

Лемма 7.2. *Кривая Безье, построенная по точкам* $\mathbf{p}_0, \ldots, \mathbf{p}_n$, *может быть задана в виде*:

$$\mathbf{r}(t) = \big((1-t)L + t\big)^n \mathbf{p}_n, \tag{7.34}$$

а также в виде:

$$\mathbf{r}(t) = \big((1-t) + tE\big)^n \mathbf{p}_0. \tag{7.35}$$

Доказательство. Имеем

$$\big((1-t)L + t\big)^n \mathbf{p}_n = \sum_{i=0}^{n} C_n^i t^i (1-t)^{n-i} L^{n-i} \mathbf{p}_n = \sum_{i=0}^{n} C_n^i t^i (1-t)^{n-i} \mathbf{p}_i,$$

поскольку $L^{n-i}\mathbf{p}_n = \mathbf{p}_i$.

Аналогично,

$$\big((1-t) + tE\big)^n \mathbf{p}_0 = \sum_{i=0}^{n} C_n^i t^i (1-t)^{n-i} E^i \mathbf{p}_0 = \sum_{i=0}^{n} C_n^i t^i (1-t)^{n-i} \mathbf{p}_i,$$

так как $E^i\mathbf{p}_0 = \mathbf{p}_i$. Лемма доказана. ■

В операторных обозначениях алгоритм де Кастелье имеет вид

$$\mathbf{r}_i(t;k) = \big((1-t) + tE\big)\mathbf{r}_i(t;k-1) \qquad (i + k \leqslant n),$$

поскольку $E\mathbf{r}_i(t;k-1) = \mathbf{r}_{i+1}(t;k-1)$.

В частности

$$\begin{aligned}
&\mathbf{r}_0(t;n) = \big((1-t) + tE\big)\mathbf{r}_0(t;n-1),\\
&\mathbf{r}_0(t;n-1) = \big((1-t) + tE\big)\mathbf{r}_0(t;n-2),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\mathbf{r}_0(t;1) = \mathbf{p}_0,
\end{aligned}$$

что опять дает нам формулу (7.35):

$$\mathbf{r}(t) = \mathbf{r}_0(t;n) = \big((1-t) + tE\big)^n \mathbf{p}_0.$$

Рис. 7.6:

Лемма 7.3. *Для любого значения $t \in [0, 1]$ соответствующая точка $\mathbf{r}(t)$ на кривой Безье $\mathbf{r}(t) = \mathbf{r}_0(t; n)$ делит отрезок, соединяющий точки $\mathbf{r}_0(t; n-1)$ и $\mathbf{r}_1(t; n-1)$, в отношении $t : (1-t)$. Сама кривая $\mathbf{r}(t) = \mathbf{r}_0(t; n)$ касается этого отрезка в точке $\mathbf{r}(t)$, см. рис.* 7.6.

Доказательство. Первое утверждение сразу следует из формулы (7.33):

$$\mathbf{r}_0(t; n) = (1-t)\mathbf{r}_0(t; n-1) + t\mathbf{r}_1(t; n-1).$$

Для доказательства второго утверждения продифференцируем $\mathbf{r}(t)$ воспользовавшись формулой (7.35):

$$\begin{aligned}\frac{d\mathbf{r}(t)}{dt} = \frac{d\mathbf{r}_0(t; n)}{dt} = \frac{d}{dt}\Big(\big((1-t)+tE\big)^n \mathbf{p}_0\Big) = \\ = n\big((1-t)+tE\big)^{n-1}(E-1)\mathbf{p}_0 = n\big((1-t)+tE\big)^{n-1}(\mathbf{p}_1 - \mathbf{p}_0) = \\ = n\big(\mathbf{r}_1(t; n-1) - \mathbf{r}_0(t; n-1)\big). \quad (7.36)\end{aligned}$$

Следовательно, $\dfrac{d\mathbf{r}(t)}{dt}$ параллельно $\mathbf{r}_1(t; n-1) - \mathbf{r}_0(t; n-1)$. Лемма доказана. ■

7.2.3 Годографы кривых Безье

Пусть дана $(n+1)$ опорная точка $\mathbf{p}_0, \ldots, \mathbf{p}_n$. Построим по ним кривую Безье $\mathbf{r}(t) = \big((1-t)+tE\big)^n \mathbf{p}_0$. Обозначим через $\mathbf{a}_0, \ldots, \mathbf{a}_{n-1}$ векторы, образованные сторонами опорного многоугольника с вершинами в точках $\mathbf{p}_0, \ldots, \mathbf{p}_n$: $\mathbf{a}_i = \mathbf{p}_{i+1} - \mathbf{p}_i$.

Тогда согласно (7.36) имеем

$$\frac{d\mathbf{r}(t)}{dt} = \dot{\mathbf{r}}(t) = n\big((1-t)+tE\big)^{n-1}\mathbf{a}_0.$$

Построим теперь новую кривую Безье, взяв в качестве ее опорных точек, концы векторов $\mathbf{a}_0, \ldots, \mathbf{a}_{n-1}$, приложенных к началу координат. Эта кривая называется *годографом* исходной кривой Безье $\mathbf{r}(t)$.

Лемма 7.4. *Для любой точки $t \in [0,1]$ радиус-вектор точки годографа кривой Безье с параметром t пропорционален касательной к самой этой кривой Безье в точке $\mathbf{r}(t)$.*

Доказательство. Уравнение годографа, согласно (7.35), имеет вид

$$\mathbf{r}_{hod}(t) = \big((1-t)+tE\big)^{n-1}\mathbf{a}_0.$$

Но согласно (7.36) для самой кривой Безье имеем

$$\frac{d\mathbf{r}(t)}{dt} = n\big((1-t)+tE\big)^{n-1}\mathbf{a}_0. \tag{7.37}$$

Отсюда следует утверждение леммы. ■

Можно определить также годографы высших порядков:

$$\mathbf{r}_{hod(j)}(t) = \big((1-t)+tE\big)^{n-j}(E-1)^j\mathbf{p}_0. \tag{7.38}$$

Их радиус-векторы будут пропорциональны высшим производным радиус-вектора исходной кривой Безье и связаны с этими производными формулой, аналогичной формуле (7.37):

$$\frac{d^j\mathbf{r}(t)}{dt^j} = \frac{n!}{(n-j)!}\mathbf{r}_{hod(j)}(t) = n(n-1)\ldots(n-j+1)\mathbf{r}_{hod(j)}(t). \tag{7.39}$$

Следствие 7.3. *Производная порядка j на концах кривой Безье полностью определяется $(j+1)$ крайней опорной точкой. А именно: опорными точками $\mathbf{p}_0,\ldots,\mathbf{p}_j$ на конце $\mathbf{p}_0$ и опорными точками $\mathbf{p}_{n-j},\ldots,\mathbf{p}_n$ на конце $\mathbf{p}_n$*

Доказательство. По формулам (7.38) и (7.39) при $t=0$ имеем

$$\frac{d^j\mathbf{r}(t)}{dt^j}\Big|_{t=0} = \frac{n!}{(n-j)!}(E-1)^j\mathbf{p}_0 =$$
$$= \frac{n!}{(n-j)!}\sum_{i=0}^{j}(-1)^{j-i}C_j^iE^i\mathbf{p}_0 = \frac{n!}{(n-j)!}\sum_{i=0}^{j}(-1)^{j-i}C_j^i\mathbf{p}_i.$$

Аналогично для производных на конце $\mathbf{p}_n$. Следствие доказано. ■

Следствие 7.4. *Репер Френе на конце кривой Безье может быть выражен через первые три (или соответственно последние три) опорные точки.*

Доказательство. Репер Френе определяется первой и второй производной кривой в точке, откуда согласно предыдущему следствию вытекает искомое утверждение. Следствие доказано. ■

Лемма 7.5. *В начальной точке кривой Безье репер Френе* $(\mathbf{v}, \mathbf{n}, \mathbf{b})$, *кривизна и кручение задаются следующими формулами*:

вектор скорости:

$$\mathbf{v}_0 = \frac{\mathbf{a}_0}{|\mathbf{a}_0|} = \frac{\mathbf{p}_1 - \mathbf{p}_0}{|\mathbf{p}_1 - \mathbf{p}_0|};$$

вектор главной нормали:

$$\mathbf{n}_0 = \frac{\mathbf{a}_1 - \mathbf{v}_0\langle \mathbf{a}_1, \mathbf{v}_0\rangle}{|\mathbf{a}_1 - \mathbf{v}_0\langle \mathbf{a}_1, \mathbf{v}_0\rangle|};$$

вектор бинормали:

$$\mathbf{b}_0 = \mathbf{v}_0 \times \mathbf{n}_0;$$

кривизна:

$$k_0 = \frac{n-1}{n}\frac{\langle \mathbf{a}_1, \mathbf{n}_0\rangle}{|\mathbf{a}_0|^2};$$

кручение:

$$\varkappa_0 = \frac{n-2}{n}\frac{\langle \mathbf{a}_2, \mathbf{b}_0\rangle}{\langle \mathbf{a}_1, \mathbf{n}_0\rangle|\mathbf{a}_0|},$$

где $\mathbf{a}_l = \mathbf{p}_{l+1} - \mathbf{p}_l$, $l = 0, 1, 2$.

Доказательство. Докажем например формулу для кривизны k_0.

Пусть s — натуральный параметр на кривой $\mathbf{r} = \mathbf{r}(t)$. Имеем

$$\ddot{\mathbf{r}}(t) = \frac{d^2\mathbf{r}}{ds^2}\left(\frac{ds}{dt}\right)^2 + \frac{d\mathbf{r}}{ds}\frac{d^2s}{dt^2}.$$

Умножая на вектор главной нормали, с учетом равенств $\left\langle \frac{d\mathbf{r}}{ds}, \mathbf{n}\right\rangle = 0$, $\frac{ds}{dt} = |\dot{\mathbf{r}}|$, $\frac{d^2\mathbf{r}}{ds^2} = k\mathbf{n}$ получаем

$$\langle \ddot{\mathbf{r}}, \mathbf{n}\rangle = k|\dot{\mathbf{r}}|^2.$$

Следовательно, $k = \frac{\langle \ddot{\mathbf{r}}, \mathbf{n}\rangle}{|\dot{\mathbf{r}}|^2}$.

С другой стороны, согласно (7.38) и (7.39) имеем

$$\dot{\mathbf{r}}(0) = n(\mathbf{p}_1 - \mathbf{p}_0) = n\mathbf{a}_0, \tag{7.40}$$

$$\begin{aligned}\ddot{\mathbf{r}}(0) = n(n-1)(E-1)^2\mathbf{p}_0 = \\ = n(n-1)(\mathbf{p}_2 - 2\mathbf{p}_1 + \mathbf{p}_0) = n(n-1)(\mathbf{a}_1 - \mathbf{a}_0).\end{aligned} \tag{7.41}$$

Отсюда

$$k_0 = \frac{\langle \ddot{\mathbf{r}}(0), \mathbf{n}_0\rangle}{|\dot{\mathbf{r}}(0)|^2} = \frac{n-1}{n}\frac{\langle \mathbf{a}_1, \mathbf{n}_0\rangle}{|\mathbf{a}_0|^2},$$

поскольку $\langle \mathbf{a}_0, \mathbf{n}_0\rangle = 0$, так как $\mathbf{n}_0 \perp \dot{\mathbf{r}}(0) = n\mathbf{a}_0$.

Остальные формулы доказываются аналогично. Формула для вектора главной нормали $\mathbf{n}_0$ следует из того, что главная нормаль ортогональна вектору $\mathbf{a}_0$, сонаправленному с $\dot{\mathbf{r}}(0)$, лежит в плоскости $\mathrm{Lin}\big(\dot{\mathbf{r}}(0), \ddot{\mathbf{r}}(0)\big)$, которая согласно (7.40) и (7.41) совпадает с плоскостью $\mathrm{Lin}(\mathbf{a}_0, \mathbf{a}_1)$, и имеет единичную длину, откуда:

$$\mathbf{n}_0 = \frac{\mathbf{a}_1 - \mathrm{Пр}_{\mathbf{a}_0}\ \mathbf{a}_1}{|\mathbf{a}_1 - \mathrm{Пр}_{\mathbf{a}_0}\ \mathbf{a}_1|}.$$

Что и дает формулу для $\mathbf{n}_0$. Лемма доказана. ■

Задача. Выписать аналогичные формулы для правого конца кривой Безье.

7.2.4 Деление кривой Безье на две кривые Безье того же порядка в отношении $t^* : (1 - t^*)$

Пусть даны опорные точки $\mathbf{p}_0, \ldots, \mathbf{p}_n$ и кривая Безье $\mathbf{r} = \mathbf{r}(t)$, построенная по ним. Требуется для некоторого заданного $t^* \in [0, 1]$ построить опорные точки $\mathbf{p}_0^a, \ldots, \mathbf{p}_n^a$ и $\mathbf{p}_0^b, \ldots, \mathbf{p}_n^b$ так, чтобы кривые Безье для этих новых точек совпали бы с двумя дугами исходной кривой Безье, на которые она разбивается точкой $\mathbf{r}(t^*)$.

Данная задача имеет важное значение в теории кривых Безье. Необходимость деления кривых Безье возникает в тех случаях, когда требуется изменить какую-то часть кривой, оставляя некоторую другую ее часть неизменной. Поскольку смещение любой опорной точки меняет всю кривую Безье в целом, в таких случаях приходится разбивать кривую Безье на несколько частей. Это дает возможность менять только одну из получившихся частей исходной кривой Безье (ставших уже самостоятельными кривыми Безье), не трогая остальные части исходной кривой.

Для решения поставленной задачи введем новый параметр $\tau = \dfrac{t}{t^*}$. Имеем: $t = \tau t^*$. Запишем формулу (7.35) в виде

$$\mathbf{r}(t) = \mathbf{r}_0(t; n) = (1 - \tau t^* + \tau t^* E)^n \mathbf{p}_0 = \\ = \big((1 - \tau) + (1 - t^* + t^* E)\tau\big)^n \mathbf{p}_0. \quad (7.42)$$

Введем новый оператор сдвига $E_a = (1 - t^* + t^* E)$ и перепишем (7.42) в виде (7.35), положив $\mathbf{p}_0^a = \mathbf{p}_0$, $\mathbf{p}_i^a = E_a^i \mathbf{p}_0^a$:

$$\mathbf{r}_0(t; n) = \mathbf{r}_0^a(t; n) = \big((1 - \tau) + \tau E_a\big)^n \mathbf{p}_0^a,$$

где $\tau \in [0, 1]$, $t = \tau t^*$.

Таким образом, если мы построим кривую Безье по $(n + 1)$ точкам $\mathbf{p}_0^a, \ldots, \mathbf{p}_n^a$, определенным по формуле $\mathbf{p}_i^a = (1 - t^* + t^* E)^i \mathbf{p}_0$, эта новая кривая Безье совпадает с дугой исходной кривой Безье $\mathbf{r}(t)$, $0 \leqslant t \leqslant t^*$.

Аналогично, с помощью нового параметра $\xi = \dfrac{t - t^*}{1 - t^*}$ и нового оператора левого сдвига $L_b = (1 - t^*)L + t^*$ $(L = E^{-1})$ определим $(n+1)$ новую опорную

точку $\mathbf{p}_0^b, \ldots, \mathbf{p}_n^b$ по формуле $\mathbf{p}_{n-i}^b = L_b^i \mathbf{p}_n^b$, $i = 0, \ldots, n$, $\mathbf{p}_n^b = \mathbf{p}_n$. Тогда дуга исходной кривой Безье на участке $[t^*, 1]$ может быть задана формулой (7.34), которую можно записать в виде

$$\begin{aligned}\mathbf{r}(t) = \big((1-t)L + t\big)^n \mathbf{p}_n = \\ = \Big(\big(1-(1-t^*)\xi - t^*\big)L + (1-t^*)\xi + t^*\Big)^n \mathbf{p}_n = \\ = \Big(\big((1-t^*)L + t^*\big)(1-\xi) + \xi\Big)^n \mathbf{p}_n = \big((1-\xi)L_b + \xi\big)^n \mathbf{p}_n^b.\end{aligned}$$

Следовательно, эта дуга будет совпадать с кривой Безье, построенной по точкам $\mathbf{p}_0^b, \ldots, \mathbf{p}_n^b$.

Процедура построения точек $\mathbf{p}_0^a, \ldots, \mathbf{p}_n^a$ и $\mathbf{p}_0^b, \ldots, \mathbf{p}_n^b$ может быть описана более наглядно следующим образом.

1) Исходный многоугольник с вершинами в опорных точках назовем *многоугольником нулевого уровня* $\mathbf{p}_0^{(0)} = \mathbf{p}_0, \ldots, \mathbf{p}_n^{(0)} = \mathbf{p}_n$.

2) Строим многоугольники уровней $1, 2, \ldots, n$ так, что вершины многоугольника следующего уровня делят стороны многоугольника предыдущего уровня в отношении $t^* : (1-t^*)$:

$$\mathbf{p}_i^{(k)} = (1-t^*)\mathbf{p}_i^{(k-1)} + t^*\mathbf{p}_{i+1}^{(k-1)}, \ i = 0, \ldots, n-k.$$

Многоугольник k-того уровня будет иметь $(n-k+1)$ вершин: $\mathbf{p}_0^{(k)}, \ldots, \mathbf{p}_{n-k}^{(k)}$. При этом, согласно построению, сдвигая все исходные вершины вперед, мы сдвигаем вперед и все вершины многоугольника любого уровня, т.е. $E\mathbf{p}_i^{(k)} = \mathbf{p}_{i+1}^{(k)}$. Тогда

$$\mathbf{p}_i^a = \mathbf{p}_0^{(i)}.$$

В самом деле, имеем по построению:

$$\begin{aligned}&\mathbf{p}_0^{(0)} = \mathbf{p}_0, \\ &\mathbf{p}_0^{(1)} = (1-t^*)\mathbf{p}_0 + t^*\mathbf{p}_1 = (1-t^*+t^*E)\mathbf{p}_0 = E_a\mathbf{p}_0^{(0)}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\mathbf{p}_0^{(k)} = (1-t^*)\mathbf{p}_0^{(k-1)} + t^*\mathbf{p}_1^{(k-1)} = \\ &= (1-t^*+t^*E)\mathbf{p}_0^{(k-1)} = E_a\mathbf{p}_0^{(k-1)},\end{aligned} \tag{7.43}$$

так как $E\mathbf{p}_0^{(k-1)} = \mathbf{p}_1^{(k-1)}$. Следовательно, $\mathbf{p}_0^{(k)} = E_a^k\mathbf{p}_0^{(0)} = E_a^k\mathbf{p}_0 = \mathbf{p}_k^a$, где $E_a = (1-t^*) + t^*E$.

Аналогично строятся точки $\mathbf{p}_0^b, \ldots, \mathbf{p}_n^b$, только надо брать последовательность последних, а не первых вершин многоугольников k-го уровня.

7.2.5 Условия сохранения гладкости сопряжения при делении кривой Безье

Деление кривой Безье, описанное выше, применяется, в частности, тогда, когда надо управлять формой этой кривой, зафиксировав некоторую точку на ней (в этой точке мы и делим кривую). Однако при последующем

управлении двумя частями кривой гладкость сопряжения в точке деления может падать. Если мы хотим, чтобы эта гладкость была не менее C^k, то мы можем добиться этого, оставив на месте точки

$$\begin{array}{l} \mathbf{p}_n^a, \ldots, \mathbf{p}_{n-k}^a, \\ \mathbf{p}_0^b, \ldots, \mathbf{p}_k^b. \end{array} \tag{7.44}$$

Данное утверждение следует из доказанного выше утверждения о том, что первые k производных на конце кривой Безье определяются соответствующим крайними $(k+1)$ опорными точками. Поскольку исходная кривая была бесконечно гладкой, то оставляя на месте точки (7.44) мы обеспечим сопряжение производных до порядка k включительно в точке деления.

7.2.6 Увеличение числа опорных точек без изменения формы кривой Безье

Поскольку опорные точки кривой Безье используются для управления этой кривой, причем некоторые из опорных точек, как мы видели, по тем или иным соображениям должны оставаться на месте и не могут участвовать в управлении кривой, то иногда возникает необходимость добавить число опорных точек данной кривой Безье.

Для кривой Безье с опорными точками $\mathbf{p}_0, \ldots, \mathbf{p}_n$ имеем

$$\mathbf{r}(t) = \mathbf{r}_0(t; n) = (1 - t + Et)^n \mathbf{p}_0$$

или

$$\mathbf{r}(t) = \sum_{i=0}^{n} C_n^i (1-t)^{n-i} t^i \mathbf{p}_i.$$

Произведем следующие преобразования параметрического представления кривой Безье с целью увеличения числа опорных точек на единицу без изменения формы исходной кривой:

$$\mathbf{r}(t) = (1 - t + t) \sum_{i=0}^{n} C_n^i (1-t)^{n-i} t^i \mathbf{p}_i =$$

$$= \sum_{i=0}^{n} C_n^i (1-t)^{n-i+1} t^i \mathbf{p}_i + \sum_{i=0}^{n} C_n^i (1-t)^{n-i} t^{i+1} \mathbf{p}_i =$$

$$= \sum_{i=0}^{n+1} C_{n+1}^i (1-t)^{n+1-i} t^i \Big(\frac{C_n^i}{C_{n+1}^i} (1 - \delta_{(n+1)i}) \mathbf{p}_i + \frac{C_n^{i-1}}{C_{n+1}^i} (1 - \delta_{0i}) \mathbf{p}_{i-1} \Big),$$

где δ_{ij} — символ Кронекера-Коппелли:

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

Рис. 7.7:

Положим теперь

$$\mathbf{q}_i = \frac{n+1-i}{n+1}\mathbf{p}_i(1-\delta_{(n+1)i}) + \frac{i}{n+1}\mathbf{p}_{i-1}(1-\delta_{0i}), \quad i = 0, \ldots, n+1.$$

Отметим, что $\mathbf{p}_{n+1}$ и $\mathbf{p}_{-1}$ входят в последнее равенство с нулевыми коэффициентами, поэтому доопределять эти (не определенные пока) точки нет необходимости.

Пример 5. На рис. 7.7 показана процедура увеличения числа опорных точек кривой Безье с 4 до 5.

Глава 8

Поверхности Безье

Определение 8.1. Пусть дано $(m+1)\times(n+1)$ точек в пространстве $\mathbb{R}^3$, образующих прямоугольную матрицу (сетку Безье)

$$\mathbf{P}_{mn} = \begin{pmatrix} \mathbf{p}_{00} & \cdots & \mathbf{p}_{0n} \\ \mathbf{p}_{10} & \cdots & \mathbf{p}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{p}_{m0} & \cdots & \mathbf{p}_{mn} \end{pmatrix}.$$

Поверхностью Безье порядка $m\times n$, соответствующей сетке $\mathbf{P}_{mn}$, называется поверхность

$$\begin{aligned} S_{00}(u,v;m,n) &= (1-v+Fv)^n(1-u+Eu)^m\mathbf{p}_{00} = \\ &= (1-u+Eu)^m(1-v+Fv)^n\mathbf{p}_{00}, \qquad 0\leqslant u,v\leqslant 1, \end{aligned} \tag{8.1}$$

где E — оператор сдвига вперед по первому индексу, F — оператор сдвига вперед по второму индексу:

$$E\mathbf{p}_{ij} = \mathbf{p}_{(i+1)j}, \qquad F\mathbf{p}_{ij} = \mathbf{p}_{i(j+1)}.$$

Операторы E и F очевидно коммутируют друг с другом: $EF\mathbf{p}_{ij} = FE\mathbf{p}_{ij} = \mathbf{p}_{(i+1)(j+1)}$. Поэтому $(1-v+Fv)^n(1-u+Eu)^m\mathbf{p}_{00} = (1-u+Eu)^m(1-v+Fv)^n\mathbf{p}_{00}$, т.е. формула (8.1) непротиворечива.

8.1 Геометрический смысл поверхности Безье

Поверхность Безье можно получить следующим образом:

1) Строим $(n+1)$ кривую Безье по столбцам матрицы $\mathbf{P}_{mn}$:

$$\mathbf{r}_j(u) = (1-u+Eu)^m\mathbf{p}_{0j}, \qquad j=0,\ldots,n.$$

2) Далее, начиная с каждой точки кривой Безье $\mathbf{r}_0(u)$, строим кривые Безье, имеющие опорные точки на кривых $\mathbf{r}_j(u)$ $(0\leqslant j\leqslant n)$, соответствующие одному и тому же значению параметра u.

При этом, соответствующий оператор перехода от опорной точки, взятой на кривой $\mathbf{r}_j(u)$, к следующей опорной точке, взятой на кривой $\mathbf{r}_{j+1}(u)$, в терминах узлов сетки $\mathbf{p}_{ij}$ описывается оператором сдвига F по второму индексу.

Следовательно,

$$S(u,v) = (1-v+Fv)^n \mathbf{r}_0(u) = (1-v+Fv)^n (1-u+Eu)^m \mathbf{p}_{00} = \\ = S_{00}(u,v;m,n),$$

где m и n — количество шагов по первому и второму индексу соответственно, начиная от угловой точки сетки Безье, 00 — индекс этой угловой точки, с которой мы начинаем образовывать все остальные узлы сетки операторами сдвига E и F.

Имеем

$$\begin{aligned} &S_{00}(u,v;0,0) = \mathbf{p}_{00}, \\ &S_{00}(u,v;1,1) = (1-u+Eu)(1-v+Fv)\mathbf{p}_{00} = \\ &\qquad = (1-u)(1-v)\mathbf{p}_{00} + (1-u)v\mathbf{p}_{01} + u(1-v)\mathbf{p}_{10} + uv\mathbf{p}_{11} \end{aligned} \tag{8.2}$$

(*четырехугольный лист Безье*);

$$\begin{aligned} S_{00}(u,v;2,2) &= (1-u+Eu)^2(1-v+Fv)^2\mathbf{p}_{00} = \\ &= (1-u+Eu)(1-v+Fv)S_{00}(u,v;1,1) = \\ &= (1-u)(1-v)S_{00}(u,v;1,1) + (1-u)vS_{01}(u,v;1,1) + \\ &\quad + u(1-v)S_{10}(u,v;1,1) + uvS_{11}(u,v;1,1). \end{aligned} \tag{8.3}$$

Здесь через S_{ij} обозначены четырехугольные листы Безье, построенные из точек $\mathbf{p}_{ij}$, взятых в качестве угловой точки, аналогично формуле 8.2.

Аналогично формуле (8.3), любая точка на поверхности Безье степени $m \times n$ может быть представлена в виде точки четырехугольного листа Безье с углами в точках соответственно подобранных поверхностей Безье порядка $(m-1) \times (n-1)$:

$$\begin{aligned} S_{00}(u,v;m,n) &= (1-u)(1-v)S_{00}(u,v;m-1,n-1) + \\ &+ (1-u)vS_{01}(u,v;m-1,n-1) + u(1-v)S_{10}(u,v;m-1,n-1) + \\ &\quad + uvS_{11}(u,v;m-1,n-1). \end{aligned}$$

8.2 Формулы вычисления координат точек на поверхности Безье

Поверхность Безье можно представить в следующим виде:

$$S_{00}(u,v;m,n) = F(u,m)\mathbf{P}F^{\top}(v,n),$$

где $\mathbf{P} = (\mathbf{p}_{ij})$ и

$$F(u,m) = \big(f_0(u,m), \ldots, f_m(u,m)\big),$$
$$F(v,n) = \big(f_0(v,n), \ldots, f_n(v,n)\big),$$
$$f_i(\xi,k) = C_k^i (1-\xi)^{k-i}\xi^i.$$

Далее нам будет удобно изучать геометрические характеристики поверхности Безье в ее угловой точке $S_{00}(0,0;m,n)$. Для того, чтобы представить произвольную точку поверхности Безье в таком виде, определим операцию деления поверхности Безье в ее произвольной точке $S_{00}(u^*,v^*;m,n)$, аналогичную делению кривой Безье (т.е. без изменения самой поверхности).

8.3 Деление поверхности Безье

Выше в разделе 7.2.4 было показано, что каждая из кривых, построенных по столбцам сетки $\mathbf{P}_{mn}$ может быть разделена в точке $u = u^*$ без изменения ее формы. При делении в точке $u = u^*$ возникает две матрицы узлов Безье того же порядка, что и исходная матрица узлов: $\mathbf{P}^a_{mn}$ и $\mathbf{P}^b_{mn}$. Аналогично, каждую из них можно поделить в точке $v = v^*$ (т.е. — поделить кривые Безье, построенные по строкам этих матриц). Получим четыре матрицы сеток Безье: $\mathbf{P}^{ac}_{mn}$, $\mathbf{P}^{ad}_{mn}$, $\mathbf{P}^{bc}_{mn}$, $\mathbf{P}^{bd}_{mn}$. Соответствующие им поверхности Безье будут лежать на исходной поверхности Безье и делить ее на четыре части.

Новые сети опорных точек $\mathbf{P}^{ac}_{mn}$, $\mathbf{P}^{ad}_{mn}$, $\mathbf{P}^{bc}_{mn}$ и $\mathbf{P}^{bd}_{mn}$ получаются из исходной сети $\mathbf{P}_{mn}$ по следующим формулам, аналогичным формулам (7.43):

$$\mathbf{p}^{ac}_{ij} = (1-u^*+Eu^*)^i(1-v^*+Fv^*)^j\mathbf{p}_{00},$$
$$\mathbf{p}^{bc}_{(m-i)j} = \big((1-u^*)E^{-1}+u^*\big)^i(1-v^*+Fv^*)^j\mathbf{p}_{m0},$$
$$\mathbf{p}^{ad}_{i(n-j)} = (1-u^*+Eu^*)^i\big((1-v^*)F^{-1}+v^*\big)^j\mathbf{p}_{0n},$$
$$\mathbf{p}^{bd}_{(m-i)(n-j)} = \big((1-u^*)E^{-1}+u^*\big)^i\big((1-v^*)F^{-1}+v^*\big)^j\mathbf{p}_{mn}.$$

8.4 Геометрические свойства поверхности Безье в угловой точке

Выше было показано, что любая точка поверхности Безье может быть представлена как угловая точка новой поверхности Безье, целиком лежащей на предыдущей. Следовательно, при изучении локальных геометрических свойств поверхности Безье мы можем ограничиться рассмотрением лишь ее угловой точки.

Имеем

$$\frac{\partial^i}{\partial u^i}\Big|_{u=0} S_{00}(u,v;m,n) = \frac{m!}{(m-i)!}(E-1)^i S_{00}(0,v;m,n) =$$
$$= \frac{m!}{(m-i)!}(1-v+Fv)^n(E-1)^i\mathbf{p}_{00}.$$

В частности

$$T_u(v) = \frac{\partial}{\partial u}\Big|_{u=0} S_{00}(u,v;m,n) = m(1-v+Fv)^n(\mathbf{p}_{10}-\mathbf{p}_{00})$$

и

$$T_v(u) = \frac{\partial}{\partial v}\Big|_{v=0} S_{00}(u,v;m,n) = n(1-u+Eu)^m(\mathbf{p}_{01}-\mathbf{p}_{00}).$$

Следовательно касательные вектора в угловой точке $\mathbf{p}_{00}$ имеют вид: $T_u(0) = m(\mathbf{p}_{10}-\mathbf{p}_{00})$ и $T_v(0) = n(\mathbf{p}_{01}-\mathbf{p}_{00})$. Нормаль к поверхности в угловой точке дается формулой: $\mathbf{n} = \dfrac{T_u(0)\times T_v(0)}{|T_u(0)\times T_v(0)|}$.

Далее, возьмем вторые производные:

$$\frac{\partial^2}{\partial u^2}\Big|_{u=0,v=0} S_{00}(u,v;m,n) = \frac{\partial^2}{\partial u^2}\Big|_{u=0,v=0}(1-u+Eu)^m(1-v+Fv)^n\mathbf{p}_{00} =$$
$$= m(m-1)(E-1)^2\mathbf{p}_{00},$$
$$\frac{\partial^2}{\partial v^2}\Big|_{u=0,v=0} S_{00}(u,v;m,n) = n(n-1)(F-1)^2\mathbf{p}_{00},$$
$$\frac{\partial^2}{\partial v\partial v}\Big|_{u=0,v=0} S_{00}(u,v;m,n) = mn(E-1)(F-1)\mathbf{p}_{00}.$$

Имеем

$$(E-1)^2\mathbf{p}_{00} = \mathbf{p}_{20} - 2\mathbf{p}_{10} + \mathbf{p}_{00},$$
$$(F-1)^2\mathbf{p}_{00} = \mathbf{p}_{02} - 2\mathbf{p}_{01} + \mathbf{p}_{00},$$
$$(F-1)(E-1)\mathbf{p}_{00} = (\mathbf{p}_{11}-\mathbf{p}_{01}) - (\mathbf{p}_{10}-\mathbf{p}_{00}) =$$
$$= (E-1)(F-1)\mathbf{p}_{00} = (\mathbf{p}_{11}-\mathbf{p}_{10}) - (\mathbf{p}_{01}-\mathbf{p}_{00}).$$

Вычислив, согласно приведенным формулам, первые и вторые производные в угловой точке, а также нормаль к поверхности в этой точке, мы сможем вычислить гауссову и среднюю кривизну, главные направления и главные кривизны в угловой точке. Все эти величины будут выражаться через узлы сетки: $\mathbf{p}_{00}$, $\mathbf{p}_{01}$, $\mathbf{p}_{02}$, $\mathbf{p}_{10}$, $\mathbf{p}_{20}$.

Для произвольной точки $S_{00}(u^*,v^*;m,n)$ поступаем так:

1) делим поверхность в этой точке и вычисляем узлы $\mathbf{p}^{bd}_{00}$, $\mathbf{p}^{bd}_{01}$, $\mathbf{p}^{bd}_{02}$, $\mathbf{p}^{bd}_{10}$, $\mathbf{p}^{bd}_{20}$;

2) вычисляем геометрические характеристики в данной точке через эти узлы.

8.5 Измельчение сетки при сохранение поверхности Безье

Добавление новых узлов в сетку Безье без изменения самой поверхности происходит по формулам, полностью аналогичным выше, в разделе 7.2.6, добавления опорных точек в кривую Безье. Если дана сетка Безье

$\mathbf{P}_{mn} = (\mathbf{p}_{ij})_{i,j=0}^{m,n}$, то увеличенная на одну строку и на один столбец сетка $\mathbf{Q}_{(m+1)(n+1)}$ для той же самой поверхности Безье строится по формулам:

$$\mathbf{q}_{ij} = \frac{1}{m+1}\frac{1}{n+1}\big((m+1-i)(1-\delta_{(m+1)i}) + i(1-\delta_{0i})E^{-1}\big) \times \\ \times\big((n+1-j)(1-\delta_{(n+1)j}) + j(1-\delta_{0j})F^{-1}\big)\mathbf{p}_{ij}.$$

При этом углы новой сетки совпадут с углами старой: $\mathbf{q}_{00} = \mathbf{p}_{00}$, $\mathbf{q}_{(m+1)0} = \mathbf{p}_{m0}$, $\mathbf{q}_{0(n+1)} = \mathbf{p}_{0n}$, $\mathbf{q}_{(m+1)(n+1)} = \mathbf{p}_{mn}$.

Глава 9

Проективные (рациональные) кривые Безье

Обычные кривые Безье получаются с помощью итерирования операции деления отрезка в отношение $t : (1-t)$. Так, например: $\mathbf{p}_t = (1-t)\mathbf{p}_0 + t\mathbf{p}_1 = (1-t+tE)\mathbf{p}_0 = \mathbf{r}_0(t;1)$ — кривая Безье первого порядка (здесь отрезок $\mathbf{p}_0\mathbf{p}_1$ разделен в отношении $t : (1-t)$). Аналогично, $\mathbf{r}_0(t;n) = (1-t+tE)^n\mathbf{p}_0$ — кривая Безье порядка n. Здесь уже операция деления отрезка в отношении $t : (1-t)$ проитерирована n раз.

Рациональные кривые Безье получаются таким же образом с помощью операции *рационального деления отрезка с весами* ω_0 *и* ω_1.

9.1 Операция рационального деления отрезка

Пусть даны точки $\mathbf{p}_0, \mathbf{p}_1 \in \mathbb{R}^n$ и их веса $\omega_0, \omega_1 \in \mathbb{R}$. Тогда *операция рационального деления отрезка* $\mathbf{p}_0$ *и* $\mathbf{p}_1$ *с весами* ω_0 *и* ω_1 определяется следующим образом:

1) Выбираем точку O не равную $\mathbf{p}_0$ и $\mathbf{p}_1$ и помещаем в нее начало координат.

2) Строим вектора $\widehat{\mathbf{p}}_0 = \omega_0\mathbf{p}_0$, $\widehat{\mathbf{p}}_1 = \omega_1\mathbf{p}_1$.

3) Делим отрезок $\widehat{\mathbf{p}}_0\widehat{\mathbf{p}}_1$ в отношение $t : (1-t)$. Пусть это будет точка $\widehat{\mathbf{p}}_t$.

4) Проецируем точку деления $\widehat{\mathbf{p}}_t$ из центра O на прямую $\mathbf{p}_0\mathbf{p}_1$. Получаем искомую точку $\mathbf{p}_t$ рационального деления отрезка $\mathbf{p}_0\mathbf{p}_1$.

Лемма 9.1. *1) Операция рационального деления отрезка не зависит от выбора точки O, не совпадающей с $\mathbf{p}_0$ и $\mathbf{p}_1$.*

Рис. 9.1:

2)

$$\frac{\mathbf{p}_0\mathbf{p}_t}{\mathbf{p}_t\mathbf{p}_1} = \frac{t\omega_1}{(1-t)\omega_0}.$$

3)

$$\frac{O\mathbf{p}_t}{O\widehat{\mathbf{p}}_t} = \frac{1}{(1-t)\omega_0 + t\omega_1}.$$

4)

$$\mathbf{p}_t = \frac{(1-t+Et)(\omega_0\mathbf{p}_0)}{(1-t+Et)\omega_0},$$

где оператор сдвига E действует одновременно и на точки $\mathbf{p}_0$, $\mathbf{p}_1$, и на их веса ω_0, ω_1.

Доказательство. Пункт 1) следует из пункта 2), так как правая часть 2) не зависит от выбора точки O.

Пункт 4) следует из пункта 2). В самом деле, согласно 2)

$$\mathbf{p}_t = \frac{(1-t)\omega_0}{t\omega_1 + (1-t)\omega_0}\mathbf{p}_0 + \frac{t\omega_1}{t\omega_1 + (1-t)\omega_0}\mathbf{p}_1$$

(поскольку $\mathbf{p}_t$ делит отрезок $\mathbf{p}_0\mathbf{p}_1$ в отношении $t\omega_1 : (1-t)\omega_0$).

Последнюю формулу можно переписать в виде

$$\mathbf{p}_t = \frac{(1-t)\omega_0\mathbf{p}_0 + t\omega_1\mathbf{p}_1}{(1-t+tE)\omega_0} = \frac{(1-t+tE)\omega_0\mathbf{p}_0}{(1-t+tE)\omega_0},$$

что и дает пункт 4).

Следовательно, достаточно доказать пункты 2) и 3). Докажем их:

2) Введем следующие обозначения: $A = \mathbf{p}_0$, $B = \mathbf{p}_1$, $C = \widehat{\mathbf{p}}_0 = \omega_0\mathbf{p}_0$, $D = \widehat{\mathbf{p}}_1 = \omega_1\mathbf{p}_1$, $E = \widehat{\mathbf{p}}_t$, $E' = \mathbf{p}_t$, см. рис. 9.1.

Построим линии CC' и DD' параллельные OEE': $OEE' \parallel CC' \parallel DD'$ (точки C' и D' лежат на отрезке AB).

Тогда $\dfrac{t}{1-t} = \dfrac{CE}{ED} = \dfrac{C'E'}{E'D'}$.

Но $\dfrac{AE'}{C'E'} = \dfrac{OA}{OC} = \dfrac{1}{\omega_0}$ из треугольника $\triangle OAE'$ и $\dfrac{E'B}{E'D'} = \dfrac{OB}{OD} = \dfrac{1}{\omega_1}$ из треугольника $\triangle OBE'$.

Следовательно,

$$\frac{\mathbf{p}_0\mathbf{p}_t}{\mathbf{p}_t\mathbf{p}_1} = \frac{AE'}{E'B'} = \frac{C'E'}{E'D'}\frac{\omega_1}{\omega_0} = \frac{t}{1-t}\frac{\omega_1}{\omega_0} = \frac{t\omega_1}{(1-t)\omega_0}.$$

Утверждение 2) доказано.

Докажем пункт 3). Имеем, согласно 2)

$$\mathbf{p}_t = \frac{(1-t)\omega_0}{t\omega_1 + (1-t)\omega_0}\mathbf{p}_0 + \frac{t\omega_1}{t\omega_1 + (1-t)\omega_0}\mathbf{p}_1.$$

Кроме того, $\widehat{\mathbf{p}}_t = (1-t)(\omega_0\mathbf{p}_0) + t(\omega_1\mathbf{p}_1) = (1-t)\widehat{\mathbf{p}}_0 + t\widehat{\mathbf{p}}_1$. Отсюда получаем равенство для $\dfrac{O\mathbf{p}_t}{O\widehat{\mathbf{p}}_t}$. Лемма доказана. ■

Определение 9.1. *Рациональной (или проективной) кривой Безье порядка* n, построенной по заданным опорным точкам $\mathbf{p}_0, \ldots, \mathbf{p}_n$ с заданными весами $\omega_0, \ldots, \omega_n$ $(\omega_i > 0)$, называется кривая

$$\mathbf{r}_0(t;n) = \frac{(1-t+tE)^n(\omega_0\mathbf{p}_0)}{(1-t+tE)^n\omega_0} = \frac{(1-t+tE)^n(\omega_0\mathbf{p}_0)}{\omega_0(t;n)}, \tag{9.1}$$

где $t \in [0,1]$, $\omega_0(t;n) = (1-t+tE)^n\omega_0$. Здесь, как и выше, оператор E действует одновременно и на последовательности опорных точек $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_n$ и на последовательности весов $\omega_0, \omega_1, \ldots, \omega_n$.

На рис. 9.2 приведены рациональные кривые Безье, построенные по одним и тем же заданным опорным точкам $\mathbf{p}_i$, $i = 0, \ldots, 5$. Веса всех точек, кроме $\mathbf{p}_3$, равны единице, а вес ω_3 точки $\mathbf{p}_3$ приведен на рисунке.

Согласно биному Ньютона, рациональная кривая Безье $\mathbf{r}_0(t;n)$ может быть представлена в виде

$$\mathbf{r}_0(t;n) = \sum_{i=0}^{n} C_n^i t^i (1-t)^{n-i} \frac{\omega_i}{\omega_0(t;n)}\mathbf{p}_i. \tag{9.2}$$

В случае равных весов $\omega_0 = \cdots = \omega_n$ знаменатель $\omega_0(t;n)$ превращается в постоянную, а рациональная кривая Безье — в обычную кривую Безье.

9.2 Свойства рациональных кривых Безье

Лемма 9.2. *Рациональная кривая Безье находится в выпуклой оболочке своих опорных точек и проходит через первую и последнюю опорные точки.*

Доказательство. Лемма доказывается полностью аналогично случаю обычной кривой Безье. Для доказательства принадлежности кривой выпуклой оболочке опорных точек достаточно воспользоваться тем, что все веса $\omega_0, \ldots, \omega_n$ по определению положительны. Лемма доказана. ■

Рис. 9.2:

Лемма 9.3. *При увеличение веса ω_i рациональная кривая Безье приближается к опорной точке $\mathbf{p}_i$. Вектор смещения при этом всегда направлен от точки исходной кривой к опорной точке $\mathbf{p}_i$ (его величина различна для разных точек кривой).*

Доказательство. Заменим ω_i на $\omega_i + \Delta$, где $\Delta > 0$. Тогда кривая $\mathbf{r}_0(t;n)$ заменится на кривую $\mathbf{r}_0^{\Delta}(t;n)$, определяемую по формуле

$$\mathbf{r}_0^{\Delta}(t;n) = \frac{\mathbf{r}_0(t;n) + \gamma_i \Delta \mathbf{p}_i}{1 + \gamma_i \Delta},$$

где $\gamma_i = \dfrac{C_n^i t^i (1-t)^{n-i}}{\omega_0(t;n)}$. Следовательно, для любого $\Delta > 0$ $\mathbf{r}_0^{\Delta}(t;n) \in [\mathbf{r}_0(t;n), \mathbf{p}_i]$ и делит этот отрезок в отношении $\gamma_i \Delta : 1$. Лемма доказана. ■

Лемма 9.4. *При смещении одной из опорных точек $\mathbf{p}_i$ на вектор ξ все точки рациональной кривой Безье смещаются в направлении вектора ξ (но на разные величины).*

Доказательство сразу следует из формулы (9.2). ■

Если $\omega_i \to 0$, то влияние опорной точки $\mathbf{p}_i$ на рациональную кривую Безье стремится к нулю. Если положить $\omega_i = 0$, то рациональная кривая Безье будет лежать в выпуклой оболочке остальных опорных точек $\mathbf{p}_0, \ldots, \mathbf{p}_{i-1}, \mathbf{p}_{i+1}, \ldots, \mathbf{p}_n$.

9.3 Деление рациональной кривой Безье

Деление рациональной кривой Безье сводится к делению обычной кривой Безье, которая порождает данную рациональную кривую.

Рассмотрим рациональную кривую Безье, построенную по опорным точкам $\mathbf{p}_0, \ldots, \mathbf{p}_n$ с весами $\omega_0 > 0, \ldots, \omega_n > 0$. Обозначим ее через $\mathbf{r}_0(t;n)$, $t \in [0,1]$. Как мы знаем, эта кривая порождается обычной кривой Безье, построенной по вспомогательным опорным точкам $\{\widehat{\mathbf{p}}_0, \ldots, \widehat{\mathbf{p}}_n\} = \{\omega_0\mathbf{p}_0, \ldots, \omega_n\mathbf{p}_n\}$, которые зависят от выбора начала координат. При этом, начало координат необходимо выбирать так, чтобы оно не совпадало ни с одной из опорных точек $\mathbf{p}_0, \ldots, \mathbf{p}_n$. Обозначим построенную таким образом вспомогательную обычную кривую Безье через $\widehat{\mathbf{r}}_0(t;n)$, $t \in [0,1]$. Разделим вспомогательную кривую $\widehat{\mathbf{r}}_0(t;n)$ в точке $t = t^*$ в отношении $t^* : (1-t^*)$, как это делалось выше. Таким образом, возникает три вспомогательные (обычные) кривые Безье:
$\widehat{\mathbf{r}}_0(t;n)$ — кривая Безье, построенная по опорным точкам $\widehat{\mathbf{p}}_0, \ldots, \widehat{\mathbf{p}}_n$, $t \in [0,1]$,
$\widehat{\mathbf{r}}_0^a(s;n)$ — кривая Безье, построенная по опорным точкам $\widehat{\mathbf{p}}_0^a, \ldots, \widehat{\mathbf{p}}_n^a$, $s \in [0,1]$,
$\widehat{\mathbf{r}}_0^b(q;n)$ — кривая Безье, построенная по опорным точкам $\widehat{\mathbf{p}}_0^b, \ldots, \widehat{\mathbf{p}}_n^b$, $q \in [0,1]$. Здесь

$$\begin{aligned} &\widehat{\mathbf{p}}_i^a = \widehat{\mathbf{r}}_0(t^*;i) = (1 - t^* + Et^*)^i \widehat{\mathbf{p}}_0, \\ &\widehat{\mathbf{p}}_{n-i}^b = \widehat{\mathbf{r}}_{n-i}(t^*;i) = \big((1-t^*)L + t^*\big)^i \widehat{\mathbf{p}}_n, \qquad 0 \leqslant i \leqslant n. \end{aligned} \tag{9.3}$$

Кривые Безье $\widehat{\mathbf{r}}_0^a(s;n)$ и $\widehat{\mathbf{r}}_0^b(q;n)$ порождают две рациональные кривые Безье $\mathbf{r}_0^a(s;n)$ и $\mathbf{r}_0^b(q;n)$ по правилу:
$\mathbf{r}_0^a(s;n)$ — рациональная кривая Безье, построенная по опорным точкам $\mathbf{p}_0^a, \ldots, \mathbf{p}_n^a$ с весами $\omega_0^a, \ldots, \omega_n^a$,
$\mathbf{r}_0^b(q;n)$ — рациональная кривая Безье, построенная по опорным точкам $\mathbf{p}_0^b, \ldots, \mathbf{p}_n^b$ с весами $\omega_0^b, \ldots, \omega_n^b$, где

$$\begin{aligned} &\mathbf{p}_i^a = \mathbf{r}_0(t^*;i) = \frac{(1 - t^* + Et^*)^i \omega_0 \mathbf{p}_0}{(1 - t^* + Et^*)^i \omega_0}, \\ &\mathbf{p}_{n-i}^b = \mathbf{r}_{n-i}(t^*;i) = \frac{\big((1-t^*)L + t^*\big)^i \omega_n \mathbf{p}_n}{\big((1-t^*)L + t^*\big)^i \omega_n}, \end{aligned} \tag{9.4}$$

$$\begin{aligned} &\omega_i^a = (1 - t^* + Et^*)^i \omega_0, \\ &\omega_{n-i}^b = \big((1-t^*)L + t^*\big)^i \omega_n. \end{aligned} \tag{9.5}$$

Из формул (9.3), (9.4) и (9.5) следует, что

$$\begin{aligned} &\widehat{\mathbf{p}}_i^a = \omega_i^a \mathbf{p}_i^a, \\ &\widehat{\mathbf{p}}_i^b = \omega_i^b \mathbf{p}_i^b, \quad i = 0, \ldots, n. \end{aligned} \tag{9.6}$$

Равенства (9.6) означают, что кривые Безье $\widehat{\mathbf{r}}_0^a$ и $\widehat{\mathbf{r}}_0^b$ соответствуют рациональным кривым Безье $\mathbf{r}_0^a$ и $\mathbf{r}_0^b$ с весами $\{\omega_i^a\}$ и $\{\omega_i^b\}$ соответственно.

Лемма 9.5. *Кривые $\mathbf{r}_0^a(s;n)$ и $\mathbf{r}_0^b(q;n)$ делят кривую $\mathbf{r}_0(t;n)$ в точке $t = t^*$ в отношении $t^* : (1-t^*)$.*

Доказательство. Поставим в соответствии рациональной кривой Безье $\mathbf{r}_0(t;n)$, построенной по исходным данным $\big\{(\mathbf{p}_0,\omega_0),\dots,(\mathbf{p}_n,\omega_n)\big\}$, обычную кривую Безье $\widetilde{\mathbf{r}}_0(t;n)$ в пространстве $\mathbb{R}^{k+1}$ с координатами $(x^1,\dots,x^k,\omega)$, если $\mathbf{p}_i\in\mathbb{R}^k$, определяемую опорными точками $\big\{(\omega_0\mathbf{p}_0,\omega_0),\dots,(\omega_n\mathbf{p}_n,\omega_n)\big\}$. Обозначим $\widetilde{\mathbf{p}}_i=(\omega_i\mathbf{p}_i,\omega_i)\in\mathbb{R}^{k+1}$.

Определим отображение $\varphi\colon\mathbb{R}^{k+1}_+\to\mathbb{R}^k$, где $\mathbb{R}^{k+1}_+=\{(x,\omega):\omega>0\}$, следующим образом: $\varphi(x^1,\dots,x^k,\omega)=\Big(\dfrac{x^1}{\omega},\dots,\dfrac{x^k}{\omega}\Big)$. Таким образом, φ — это центральная проекция из начала координат на плоскость $\omega=1$. Покажем, что φ отображает кривую $\widetilde{\mathbf{r}}_0(t;n)$ в кривую $\mathbf{r}_0(t;n)$.

Имеем

$$\widetilde{\mathbf{r}}_0(t;n)=(1-t+tE)^n\widetilde{\mathbf{p}}_0=\big((1-t+tE)^n\omega_0\mathbf{p}_0,(1-t+tE)^n\omega_0\big).$$

Следовательно, согласно (9.1), получаем

$$\varphi\big(\widetilde{\mathbf{r}}_0(t;n)\big)=\frac{(1-t+Et)^n\omega_0\mathbf{p}_0}{(1-t+Et)^n\omega_0}=\mathbf{r}_0(t;n).$$

Следовательно, деление кривой $\widetilde{\mathbf{r}}_0(t;n)$ переходит в деление кривой $\mathbf{r}_0(t;n)$. Но согласно формулам (9.4) и (9.5) кривые Безье, построенные по точкам $\{(\omega_i^a\mathbf{p}_i^a,\omega_i^a)\}_{i=0}^n$ и $\{(\omega_i^b\mathbf{p}_i^b,\omega_i^b)\}_{i=0}^n$, делят кривую Безье $\widetilde{\mathbf{r}}_0(t;n)$ в отношении $t^*:(1-t^*)$. В самом деле, из того, что

$$\omega_i^a\mathbf{p}_i^a=(1-t^*+Et^*)^i\omega_0\mathbf{p}_0,\ \omega_i^a=(1-t^*+Et^*)^i\omega_0$$

следует, что $(\omega_i^a\mathbf{p}_i^a,\omega_i^a)=(1-t^*+Et^*)^i\widetilde{\mathbf{p}}_0=\widetilde{\mathbf{p}}_i^a$.

Аналогично: $(\omega_{n-i}^b\mathbf{p}_{n-i}^b,\omega_{n-i}^b)=\big((1-t^*)L+t^*\big)^i\widetilde{\mathbf{p}}_n=\widetilde{\mathbf{p}}_i^b$.

Это означает, что операция деления обычной кривой Безье $\widetilde{\mathbf{r}}_0(t;n)$ в точке t^* переходит при проектировании $\mathbb{R}^{k+1}\to\mathbb{R}^k$ (отбрасыванием последней компоненты) в операцию деления обычной кривой Безье $\widehat{\mathbf{r}}_0(t;n)$ в той же точке t^*. Но две части разделенной в точке t^* кривой Безье $\widehat{\mathbf{r}}_0(t;n)$ (которые мы обозначили выше через $\widehat{\mathbf{r}}_0^a(t;n)$ и $\widehat{\mathbf{r}}_0^b(t;n)$) порождают, по определению, те две рациональные кривые Безье, которые являются результатом деления рациональной кривой Безье $\mathbf{r}_0(t;n)$ в точке t^*. Выше мы обозначили их через $\mathbf{r}_0^a(t;n)$ и $\mathbf{r}_0^b(t;n)$. В итоге получаем, что кривые $\mathbf{r}_0^a(t;n)$ и $\mathbf{r}_0^b(t;n)$, на которые делится рациональная кривая Безье $\mathbf{r}_0(t;n)$, являются образами при отображении φ кривых $\widetilde{\mathbf{r}}_0^a(t;n)$ и $\widetilde{\mathbf{r}}_0^b(t;n)$, на которые делится в точке t^* обычная кривая Безье $\widetilde{\mathbf{r}}_0(t;n)$. Последнее утверждение следует из того, что отображение φ раскладывается в композицию проектирования $\mathbb{R}^{k+1}\to\mathbb{R}^k$ (отбрасыванием последней компоненты) и операции порождения рациональной кривой Безье из обычной вспомогательной кривой Безье с помощью весов $\omega_0,\dots,\omega_n$. Изобразим это в виде диаграмм:

$$\begin{array}{ccc}\widetilde{\mathbf{r}}_0(t;n) & \xrightarrow{\varphi} & \mathbf{r}_0(t;n)\\ {\scriptstyle\text{Пр}}\big\downarrow & & \big\|\\ \widehat{\mathbf{r}}_0(t;n) & \xrightarrow{\omega} & \mathbf{r}_0(t;n),\end{array}\qquad\begin{array}{ccc}\widetilde{\mathbf{r}}_0^a(t;n) & \xrightarrow{\varphi} & \mathbf{r}_0^a(t;n)\\ {\scriptstyle\text{Пр}}\big\downarrow & & \big\|\\ \widehat{\mathbf{r}}_0^a(t;n) & \xrightarrow{\omega^a} & \mathbf{r}_0^a(t;n),\end{array}$$

$$\begin{array}{ccc} \widetilde{\mathbf{r}}_0^b(t;n) & \xrightarrow{\varphi} & \mathbf{r}_0^b(t;n) \\ {\scriptstyle \text{Пр}}\Big\downarrow & & \Big\| \\ \widehat{\mathbf{r}}_0^b(t;n) & \xrightarrow{\omega^b} & \mathbf{r}_0^b(t;n). \end{array}$$

Здесь через Пр обозначена проекция $\mathbb{R}^{k+1} \to \mathbb{R}^k$, получаемая отбрасыванием последней компоненты, а через ω, ω^a и ω^b — операции порождения рациональной кривой Безье из вспомогательной обычной кривой Безье с помощью весов $\{\omega_0, \ldots, \omega_n\}$, $\{\omega_0^a, \ldots, \omega_n^a\}$ и $\{\omega_0^b, \ldots, \omega_n^b\}$ соответственно.

Отсюда следует, что кривые $\mathbf{r}_0^a(t;n)$ и $\mathbf{r}_0^b(t;n)$ действительно делят кривую $\mathbf{r}_0(t;n)$ — т.е. действительно представляют собой две части кривой $\mathbf{r}_0(t;n)$, стыкующиеся в точке $\mathbf{r}_0(t^*;n)$ и составляющие в совокупности всю исходную кривую $\mathbf{r}_0(t;n)$. (Заметим, что из определения это непосредственно не следует).

В самом деле, аналогичное утверждение верно по определению для кривых $\widetilde{\mathbf{r}}_0(t;n)$, $\widetilde{\mathbf{r}}_0^a(t;n)$ и $\widetilde{\mathbf{r}}_0^b(t;n)$, которые при взаимно-однозначном отображении φ переходят в кривые $\mathbf{r}_0(t;n)$, $\mathbf{r}_0^a(t;n)$ и $\mathbf{r}_0^b(t;n)$ соответственно. Значит, оно верно и для кривых $\mathbf{r}_0(t;n)$, $\mathbf{r}_0^a(t;n)$ и $\mathbf{r}_0^b(t;n)$. Лемма доказана. ■

9.4 Увеличение числа опорных точек рациональной кривой Безье

Операция увеличения числа опорных точек рациональной кривой Безье сводится к уже известной нам операции увеличения числа опорных точек обычной кривой Безье в пространстве $\mathbb{R}_+^{k+1}$ и применению введенного в доказательстве леммы 9.5 отображения $\varphi\colon \mathbb{R}_+^{k+1} \to \mathbb{R}^k$, $\varphi(\mathbf{x}, \omega) = \dfrac{\mathbf{x}}{\omega}$. Более подробно, алгоритм состоит в следующем.

1) По данной рациональной кривой Безье $\mathbf{r}_0(t;n)$, определяемой опорными точками $\mathbf{p}_0, \ldots, \mathbf{p}_n$ и весами $\omega_0, \ldots, \omega_n$, строим обычную кривую Безье $\widetilde{\mathbf{r}}_0(t;n)$ в пространстве $\mathbb{R}_+^{k+1}$, определяемую опорными точками $\big\{(\omega_0\mathbf{p}_0, \omega_0), \ldots, (\omega_n\mathbf{p}_n, \omega_n)\big\}$, $\widetilde{\mathbf{p}}_i = (\omega_i\mathbf{p}_i, \omega_i) \in \mathbb{R}_+^{k+1}$. Тогда $\varphi : \widetilde{\mathbf{r}}_0(t;n) \longmapsto \mathbf{r}_0(t;n)$.

2) Увеличиваем число опорных точек обычной кривой Безье $\widetilde{\mathbf{r}}_0(t;n)$ на единицу способом, описанным в разделе 7.2.6. А именно, строим новые опорные точки $\widetilde{\mathbf{q}}_i \in \mathbb{R}^{k+1}$, $i = 0, \ldots, n+1$, по формуле

$$\widetilde{\mathbf{q}}_i = \frac{n+1-i}{n+1}\widetilde{\mathbf{p}}_i(1-\delta_{(n+1)i}) + \frac{i}{n+1}\widetilde{\mathbf{p}}_{i-1}(1-\delta_{0i}), \qquad i = 0, \ldots, n+1.$$

Точки $\widetilde{\mathbf{p}}_{-1}$ и $\widetilde{\mathbf{p}}_{n+1}$ можно определить произвольно, т.к. они входят в правую часть последней формулы с нулевыми коэффициентами.

3) Применяем к новым опорным точкам $\widetilde{\mathbf{q}}_i$ преобразование φ: $\mathbf{q}_i = \varphi(\widetilde{\mathbf{q}}_i)$. Поскольку кривая Безье $\widetilde{\mathbf{r}}_0(t;n+1)$, определяемая опорными точками $\{\widetilde{\mathbf{q}}_0, \ldots, \widetilde{\mathbf{q}}_{n+1}\}$, — это та же самая кривая, что и $\widetilde{\mathbf{r}}_0(t;n)$, то $\varphi\big(\widetilde{\mathbf{r}}_0(t;n+1)\big) =$

$\mathbf{r}_0(t;n)$. Но одновременно, по определению рациональной кривой Безье имеем: $\varphi\big(\widetilde{\mathbf{r}}_0(t;n+1)\big) = \mathbf{r}_0(t;n+1)$, где $\mathbf{r}_0(t;n+1)$ — рациональная кривая Безье, определяемая опорными точками и весами $\big\{(\mathbf{q}_0,\omega_0'),\ldots,(\mathbf{q}_{n+1},\omega_{n+1}')\big\}$, где $\omega_i' = \widetilde{q}_i^{k+1}$ — $(k+1)$-ые компоненты точек $\widetilde{\mathbf{q}}_i$ в $\mathbb{R}^{k+1}$.

Окончательные формулы имеют вид

$$\begin{cases} \mathbf{q}_i = \dfrac{i\omega_{i-1}\mathbf{p}_{i-1} + (n+1-i)\omega_i\mathbf{p}_i}{i\omega_{i-1} + (n+1-i)\omega_i}, \\ \omega_i' = \dfrac{i\omega_{i-1} + (n+1-i)\omega_i}{n+1}, \end{cases} \qquad i = 0,\ldots,n+1,$$

где $\big\{(\mathbf{p}_i,\omega_i)\big\}_{i=0}^{n}$ — исходные опорные точки и веса данной рациональной кривой Безье, а $\big\{(\mathbf{q}_i,\omega_i')\big\}_{i=0}^{n+1}$ — новые опорные точки и веса той же кривой, но — с увеличенным на единицу количеством опорных точек (и, соответственно, с большим на единицу числом весов).

9.5 Производные на концах рациональной кривой Безье

Так же, как и в случае обычных кривых Безье, i-тая производная на конце рациональной кривой Безье определяется $(i+1)$ крайней опорной точкой, а также весами этих точек.

Пусть $\mathbf{r}_0(t;n)$ — рациональная кривая Безье, определяемая опорными точками и весами: $\big\{(\mathbf{p}_0,\omega_0),\ldots,(\mathbf{p}_n,\omega_n)\big\}$. Имеем

$$\mathbf{r}_0(t;n) = \frac{(1-t+Et)^n\omega_0\mathbf{p}_0}{(1-t+Et)^n\omega_0}, \qquad t\in[0,1].$$

Перепишем это равенство в виде

$$\mathbf{r}_0(t;n)(1-t+Et)^n\omega_0 = (1-t+Et)^n\omega_0\mathbf{p}_0.$$

Продифференцируем по t обе части последнего равенства i раз в точке $t=0$. Получим

$$\begin{aligned} \omega_0\mathbf{r}_0^{(i)}(0;n) + \sum_{j=1}^{i} C_i^j\mathbf{r}_0^{(i-j)}(0;n)\big((1-t+Et)^n\big)^{(j)}\big|_{t=0}\omega_0 = \\ = \omega_0\mathbf{r}_0^{(i)}(0;n) + \sum_{j=1}^{i} C_i^j\frac{n!}{(n-j)!}\mathbf{r}_0^{(i-j)}(0;n)(E-1)^j\omega_0 = \\ = \frac{n!}{(n-i)!}(E-1)^i\omega_0\mathbf{p}_0. \end{aligned} \tag{9.7}$$

Здесь верхний индекс в скобках обозначает порядок производной, взятой по t. Из соотношения (9.7) следует, что:

$$\mathbf{r}_0(0;n) = \mathbf{p}_0, \qquad \omega_0\mathbf{r}_0^{(1)}(0;n) + n(\omega_1 - \omega_0)\mathbf{p}_0 = n(\omega_1\mathbf{p}_1 - \omega_0\mathbf{p}_0),$$
$$\omega_0\mathbf{r}_0^{(2)}(0;n) + 2n(\omega_1 - \omega_0)\mathbf{r}_0^{(1)}(0;n) + n(n-1)(\omega_2 - 2\omega_1 + \omega_0)\mathbf{p}_0 =$$
$$= n(n-1)(\omega_2\mathbf{p}_2 - 2\omega_1\mathbf{p}_1 + \omega_0\mathbf{p}_0).$$

Откуда получаем искомые формулы для производных на концах рациональной кривой Безье:

$$\mathbf{r}_0(0;n) = \mathbf{p}_0, \qquad \mathbf{r}_0^{(1)}(0;n) = n\frac{\omega_1}{\omega_0}(\mathbf{p}_1 - \mathbf{p}_0),$$
$$\mathbf{r}_0^{(2)}(0;n) = n(n-1)\frac{\omega_2}{\omega_0}(\mathbf{p}_2 - \mathbf{p}_1) +$$
$$+ n\frac{2\omega_1\omega_0 + (n-1)\omega_0\omega_2 - 2n\omega_1^2}{\omega_0^2}(\mathbf{p}_1 - \mathbf{p}_0).$$

Аналогично, на другом конце кривой при $t = 1$ имеем:

$$\mathbf{r}_0(1;n) = \mathbf{p}_n, \qquad \mathbf{r}_0^{(1)}(1;n) = n\frac{\omega_{n-1}}{\omega_n}(\mathbf{p}_n - \mathbf{p}_{n-1}),$$
$$\mathbf{r}_0^{(2)}(1;n) = -n(n-1)\frac{\omega_{n-2}}{\omega_n}(\mathbf{p}_{n-1} - \mathbf{p}_{n-2}) +$$
$$+ n\frac{2n\omega_{n-1}^2 - (n-1)\omega_n\omega_{n-2} - 2\omega_{n-1}\omega_n}{\omega_n^2}(\mathbf{p}_n - \mathbf{p}_{n-1}).$$

Задача. Доказать, что кривизна рациональной кривой Безье в точке $t = 0$ отличается от кривизны обычной кривой Безье, построенной по тем же опорным точкам без учета весов, на множитель $\dfrac{\omega_0\omega_2}{\omega_1^2}$:

$$k_{\text{рац}}(0) = \frac{\omega_0\omega_2}{\omega_1^2}k_{\text{обыч}}(0).$$

9.6 Рациональные поверхности Безье

Напомним, что поверхность Безье порядка $(1,1)$ определяется матрицей опорных точек $(\widehat{\mathbf{p}}_{ij})_{i,j=0}^1$ по формуле:

$$\widehat{S}_{00}(u,v;1,1) = (1 - u + Eu)(1 - v + Fv)\widehat{\mathbf{p}}_{00}.$$

Пусть теперь каждая точка $\mathbf{p}_{ij}$ в матрице опорных точек имеет вес $\omega_{ij} > 0$.

Определение 9.2. *Рациональная поверхность Безье, построенная по точкам* $\mathbf{p}_{ij}$ *с весами* ω_{ij} определяется следующим образом:

$$S_{00}(u,v;1,1) = \frac{(1 - u + Eu)(1 - v + Fv)\omega_{00}\mathbf{p}_{00}}{(1 - u + Eu)(1 - v + Fv)\omega_{00}}. \tag{9.8}$$

Формула (9.8) означает, что мы строим поверхность Безье в $\mathbb{R}_+^{k+1}$ по точкам $(\omega_{ij}\mathbf{p}_{ij}, \omega_{ij}) \in \mathbb{R}_+^{k+1}$, а затем применяем преобразование $\varphi\colon \mathbb{R}_+^{k+1} \to \mathbb{R}^k$, $\varphi(\mathbf{x}, \omega) = \dfrac{\mathbf{x}}{\omega}$.

Из этого определения следует, что рациональная поверхность Безье $S_{00}(u, u; 1, 1)$ может быть получена (аналогично обычной поверхности Безье) с помощью следующего алгоритма:

1) соединяем $\mathbf{p}_{00}$ и $\mathbf{p}_{01}$ рациональной кривой Безье $S_{00}(0, v)$ с весами ω_{00} и ω_{01};

2) соединяем $\mathbf{p}_{10}$ и $\mathbf{p}_{11}$ рациональной кривой Безье $S_{00}(1, v)$ с весами ω_{10} и ω_{11};

3) фиксируем $v = v_0$, соединяем точки $S_{00}(0, v_0)$ и $S_{00}(1, v_0)$ с весами $(1 - v_0 + Fv_0)\omega_{00}$ и $(1 - v_0 + Fv_0)\omega_{10}$ рациональной кривой Безье и берем на ней точку с параметром $u = u_0 \in [0, 1]$.

Полученная точка и будет точкой с параметрами (u_0, v_0) на рациональной поверхности Безье, построенной по $(\mathbf{p}_{ij}, \omega_{ij})_{i,i=0}^1$.

9.7 Представление кривых второго порядка рациональными кривыми Безье порядка 2

Рассмотрим рациональную кривую Безье порядка 2, построенную по трем опорным точкам $\mathbf{p}_0$, $\mathbf{p}_1$, $\mathbf{p}_2$ с весами ω_0, ω_1, ω_2 соответственно:

$$\mathbf{r}(t; 2) = \frac{(1-t)^2\omega_0\mathbf{p}_0 + 2t(1-t)\omega_1\mathbf{p}_1 + t^2\omega_2\mathbf{p}_2}{(1-t)^2\omega_0 + 2t(1-t)\omega_1 + t^2\omega_2}. \tag{9.9}$$

Покажем, что при подходящем выборе параметров $(\mathbf{p}_i, \omega_i)_{i=0}^2$ кривая (9.9) задает на плоскости дугу кривой второго порядка общего вида.

Введем новую переменную τ по формуле $t = \dfrac{\tau}{\xi(1-\tau)+\tau}$, $1 - t = \dfrac{\xi(1-\tau)}{\xi(1-\tau)+\tau}$, где величину ξ подберем позже. Тогда (9.9) преобразуется к виду

$$\mathbf{r}(t; 2) = \mathbf{r}(\tau; \xi) = \frac{\xi^2\omega_0(1-\tau)^2\mathbf{p}_0 + 2\xi\omega_1\tau(1-\tau)\mathbf{p}_1 + \omega_2\tau^2\mathbf{p}_2}{\xi^2\omega_0(1-\tau)^2 + 2\xi\omega_1\tau(1-\tau) + \omega_2\tau^2}. \tag{9.10}$$

Подберем теперь ξ так, чтобы в числителе правой части последнего равенства коэффициенты при $\mathbf{p}_0(1-\tau)^2$ и $\mathbf{p}_2\tau^2$ совпали: $\xi^2\omega_0 = \omega_2$. Имеем: $\xi = \sqrt{\dfrac{\omega_2}{\omega_0}}$. Далее, разделим числитель и знаменатель (9.10) на ω_2. Тогда коэффициенты при $\mathbf{p}_0(1-\tau)^2$ и $\mathbf{p}_2\tau^2$ стали равными единице. При этом, коэффициент при $2\mathbf{p}_1\tau(1-\tau)$ станет равным $\omega = \dfrac{\xi\omega_1}{\omega_2} = \sqrt{\dfrac{\omega_2}{\omega_0}}\dfrac{\omega_1}{\omega_2} = \dfrac{\omega_1}{\sqrt{\omega_0\omega_2}}$.

Рис. 9.3:

Выберем ω в качестве нового параметра, вместо ξ. Тогда (9.10) преобразуется к виду

$$\mathbf{r}(\tau;\xi) = \mathbf{r}(\tau;\omega) = \frac{(1-\tau)^2\mathbf{p}_0 + 2\omega\tau(1-\tau)\mathbf{p}_1 + \tau^2\mathbf{p}_2}{(1-\tau)^2 + 2\omega\tau(1-\tau) + \tau^2}. \tag{9.11}$$

Далее, сделаем еще одну замену переменной: $u = \dfrac{\tau}{1-\tau}$. Тогда (9.11) преобразуется к виду

$$\mathbf{r}(u;\omega) = \frac{\mathbf{p}_0 + 2u\omega\mathbf{p}_1 + u^2\mathbf{p}_2}{1 + 2u\omega + u^2}. \tag{9.12}$$

Подберем декартову систему координат так, чтобы выполнялись соотношения: $\mathbf{p}_0 = (-c, 0)$, $\mathbf{p}_1 = (b, h)$, $\mathbf{p}_2 = (c, 0)$, $\mathbf{r}(u;\omega) = (x, y)$. Это всегда можно сделать, направив ось Ox по прямой $\mathbf{p}_0\mathbf{p}_2$, выбрав соответствующий масштаб и поместив начало координат посередине между точками $\mathbf{p}_0$ и $\mathbf{p}_2$, см. рис. 9.3. В компонентах выбранной декартовой системы (9.12) переписывается в виде

$$\mathbf{r}(u;\omega) = (x, y), \qquad \begin{cases} u^2(c-x) + 2\omega u(b-x) - (c+x) = 0, \\ u^2(-y) + 2\omega u(h-y) - y = 0. \end{cases} \tag{9.13}$$

Выразим из последних уравнений значения u^2 и u, считая их независимыми переменными:

$$\begin{aligned} u^2 &= \frac{2\omega\big((c+x)h - y(b+c)\big)}{\Delta}, \\ u &= \frac{2cy}{\Delta}, \end{aligned} \tag{9.14}$$

где Δ — определитель матрицы системы (9.13) относительно неизвестных u^2 и u, который равен: $\Delta = 2\omega\big((c-x)h + y(b-c)\big)$. Исключая u из (9.14), получим неявную форму задания кривой $\mathbf{r}(u;\omega)$ в координатах (x, y):

$$h^2x^2 - 2hbxy + \Big(\frac{c^2}{\omega^2} + (b^2 - c^2)\Big)y^2 + 2c^2hy - c^2h^2 = 0. \tag{9.15}$$

Рис. 9.4:

Отметим, что последнее преобразование не было равносильным, поэтому уравнение (9.15) является следствием уравнения (9.14) и, вообще говоря, не равносильно ему. Следовательно, уравнение (9.15) задает, вообще говоря, большую кривую, частью которой является исходная кривая $\mathbf{r}(u;\omega)$.

Перепишем уравнение (9.15) в виде

$$Ax^2 + 2Bxy + Dy^2 + 2Gy + F = 0, \tag{9.16}$$

где $A = h^2$, $B = -hb$, $D = \dfrac{c^2}{\omega^2} + (b^2 - c^2)$, $G = c^2 h$, $F = -c^2 h^2$. Известно, что уравнение (9.16) задает кривую второго порядка. Тип кривой зависит от величин инвариантов:

$$\begin{aligned}
I_1 &= A + D,\\
I_2 &= \begin{vmatrix} A & B \\ B & D \end{vmatrix} = h^2\Big(\frac{c^2}{\omega^2} + (b^2 - c^2)\Big) - (hb)^2 = h^2c^2\Big(\frac{1}{\omega^2} - 1\Big),\\
I_3 &= \begin{vmatrix} A & B & 0 \\ B & D & G \\ 0 & G & F \end{vmatrix} = FI_2 - AG^2.
\end{aligned}$$

Согласно теореме о классификации кривых второго порядка, тип кривой зависит от знака определителя I_2, см. рис. 9.4.

1) Если $I_2 > 0$ (т.е. если $\omega < 1$), то кривая эллиптического типа:
   - (а) если $I_1 I_3 < 0$, то это — эллипс;
   - (б) если $I_1 I_3 = 0$, то это — вырожденный эллипс (одна точка);
   - (в) если $I_1 I_3 > 0$, то это — мнимый эллипс.
2) Если $I_2 = 0$ (т.е. если $\omega = 1$), то кривая параболического типа:
   - (а) если $I_3 \neq 0$, то это — парабола;
   - (б) если $I_3 = 0$, то это — пара параллельных прямых (действительных или мнимых).
3) Если $I_2 < 0$ (т.е. если $\omega > 1$), то кривая гиперболического типа:

(а) если $I_3 \neq 0$, то это — гипербола;

(б) если $I_3 = 0$, то это — пара пересекающихся прямых.

Далее, дифференцируем (9.16) по x и приравнивая производную нулю, получим: $2Ax + 2By = 0$, т.е. $hx = by$. Следовательно, максимальное значение y достигается на прямой $O\mathbf{p}_1$ и, согласно (9.14) и (9.16), равно $y_0 = \dfrac{\omega h}{1+\omega}$. Соответствующее y_0 значение переменной t равно $t_0 = \dfrac{1}{2}$.

Задача. Доказать, что если $\omega = \dfrac{c}{|\mathbf{p}_0\mathbf{p}_1|}$, то рациональная кривая Безье является дугой окружности.

Глава 10

B-сплайны (бета-сплайны), B-кривые (бета-кривые) и B-поверхности (бета-поверхности)

10.1 Постановка задачи

Основными недостатками кривых Безье являются следующие:

1) смещение любой опорной точки глобально влияет на всю кривую;

2) сложно обеспечить точное прохождение кривой через заданные точки (кроме крайних опорных точек, являющихся концами кривой).

Поэтому при работе с кривыми Безье часто приходиться прибегать к делению кривой Безье, т.е. по сути дела — к замене одной кривой другими. Поэтому естественным образом возникает задача построения такого класса кривых, который бы позволял избежать недостатков 1), 2) и решать соответствующие задачи управления кривой путем изменения управляющих параметров кривой, а не путем ее замены на другие кривые.

Вспомним, что рациональная кривая Безье, построенная по точкам $\mathbf{p}_0, \ldots, \mathbf{p}_n$ с весами $\omega_0, \ldots, \omega_n$, определялась формулой

$$\mathbf{r}_0(t;n) = \frac{(1-t+tE)^n \omega_0 \mathbf{p}_0}{(1-t+tE)^n \omega_0} = \\ = \frac{\sum\limits_{i=0}^{n} C_n^i t^i (1-t)^{n-i} \omega_i \mathbf{p}_i}{\sum\limits_{i=0}^{n} C_n^i t^i (1-t)^{n-i} \omega_i} = \frac{\sum\limits_{i=0}^{n} B_i^n(t) \omega_i \mathbf{p}_i}{\sum\limits_{i=0}^{n} B_i^n(t) \omega_i}, \quad (10.1)$$

где $\left\{B_i^n(t) = C_n^i t^i (1-t)^{n-i}\right\}$ — базис Бернштейна на $[0,1]$.

Из формулы (10.1) видно, что глобальное влияние каждой опорной точки $\mathbf{p}_i$ на всю кривую происходит из-за того, что каждая из функций B_i^n не равна нулю на всем интервале $(0,1)$. Если бы вместо функций Бернштейна в формуле (10.1) выступали функции с локальными носителями, существенно меньшими, чем область определения параметра кривой t, то удалось бы как локализовать влияние отдельной точки на всю кривую, так и добиться (за счет выбора таких функций) прохождения кривой через некоторую заданную точку.

Поэтому поставим задачу следующим образом. Пусть даны опорные точки $\mathbf{p}_1, \ldots, \mathbf{p}_n$ с весами $\omega_1, \ldots, \omega_n$. Определим кривую $\mathbf{r}(t)$ по формуле

$$\mathbf{r}(t) = \frac{\sum\limits_{i=1}^{n} N_i(t)\omega_i \mathbf{p}_i}{\sum\limits_{i=1}^{n} N_i(t)\omega_i}, \qquad t_{min} \leqslant t \leqslant t_{max}, \tag{10.2}$$

где $N_i(t)$ — набор кусочно-полиномиальных функций таких, что

1) $N_i(t) = 0$ при $t \notin [a_i, b_i] \subset [t_{min}, t_{max}]$;

2) для любой гладкой функции $f(t)$ существует линейная комбинация $\sum c_i N_i(t)$, интерполирующая функцию $f(t)$ в заданных узлах $\{t_i\}_{i=1}^n$:

$$f(t_i) = \sum_{j=1}^{n} c_j N_j(t_i), \ i = 1, \ldots, n;$$

3) $\sum\limits_{i=1}^{n} N_i(t) = 1$ для каждого $t \in [t_{min}, t_{max}]$.

Последнее условие вводится для того, чтобы в случае совпадения всех опорных точек кривая превращалась бы в ту же точку.

Решение поставленной задачи основывается на понятии разделенных разностей и называется *B-кривыми*. B-кривые обобщают кривые Безье.

10.2 Разделенные разности

Пусть имеется гладкая функция $f(t)$ и фиксированы некоторые значения $t_1, \ldots, t_{m+1}$ переменной t (не обязательно различные и не обязательно упорядоченные по возрастанию или убыванию t). Мы также будем рассматривать переменное значение t, которое будем считать первым в последовательности. Поскольку оно предшествует значению t_1, припишем ему индекс 0 (хотя в большинстве случаев будем обозначать его просто t без индекса): $t_0 = t, t_1, \ldots, t_{m+1}$. Обозначим $f_i = f(t_i)$, $f_t = f(t) = f(t_0) = f_0$. Определим оператор правого сдвига E:

$$Et_i = t_{i+1}, \qquad Ef_i = f_{i+1} \qquad (0 \leqslant i \leqslant m).$$

Определение 10.1. *Разделенной разностью* функции $f(t)$ в узле t_i порядка $k \leqslant m - i + 1$ называется величина

$$R^{(k)} f_i = \frac{(E-1)R^{(k-1)} f_i}{(E^k - 1)t_i} = \frac{R^{(k-1)} f_{i+1} - R^{(k-1)} f_i}{t_{i+k} - t_i}, \qquad (10.3)$$
$$R^{(0)} f_i = f_i.$$

При этом, если $t_i = t_{i+1} = \cdots = t_{i+k}$, то в правой части (10.3) появляется отношение вида $\dfrac{0}{0}$, которое по определению полагается равным

$$R^{(k)} f_i = \lim_{\delta_1, \ldots, \delta_k \to 0,\ \delta_i > 0} R^{(k)}_{\delta_1, \ldots, \delta_k} f_i,$$

где $R^{(k)}_{\delta_1, \ldots, \delta_k} f_i$ определяется по возмущенным узлам $\widetilde{t}_{i+j} = t_i + \delta_1 + \cdots + \delta_j$, где $1 \leqslant j \leqslant k$, $\delta_j > 0$.

Теорема 10.1. *Если $t_i = t_{i+1} = \cdots = t_{i+k}$, то*

$$R^{(k)} f_i = \frac{1}{k!} \frac{d^k}{d\,t^k} f(t_i).$$

Доказательство будет дано для $k = 1$, $k = 2$, а затем, на основе этих доказательств — для произвольного k. Рассмотрим сначала случай $k = 1$. Имеем

$$R^{(1)} f_i = \lim_{\delta \to 0} \frac{f(t_i + \delta) - f(t_i)}{(t_i + \delta) - t_i} = \lim_{\delta \to 0} \frac{f(t_i + \delta) - f(t_i)}{\delta} = f'(t_i).$$

Пусть теперь $k = 2$. Тогда

$$R^{(2)} f_i = \lim_{0 < \delta_1, \delta_2 \to 0} \frac{\widetilde{R}^{(1)} f_{i+1} - \widetilde{R}^{(1)} f_i}{(t_i + \delta_1 + \delta_2) - t_i} = \lim_{0 < \delta_1, \delta_2 \to 0} \frac{\widetilde{R}^{(1)} f_{i+1} - \widetilde{R}^{(1)} f_i}{\delta_1 + \delta_2}, \qquad (10.4)$$

где через $\widetilde{R}^{(1)}$ обозначены разделенные разности первого порядка, построенные по возмущенным точкам: $\widetilde{t}_{i+1} = t_i + \delta_1$, $\widetilde{t}_{i+2} = t_i + \delta_1 + \delta_2$.

Поскольку $f(t)$ по условию — гладкая функция, то по формуле Тейлора имеем

$$\widetilde{R}^{(1)} f_i = \frac{f(t_i + \delta_1) - f(t_i)}{\delta_1} = \frac{f'(t_i)\delta_1 + \frac{1}{2} f''(t_i)\delta_1^2 + o(\delta_1^2)}{\delta_1} =$$
$$= f'(t_i) + \frac{1}{2} f''(t_i)\delta_1 + o(\delta_1). \qquad (10.5)$$

Аналогично:

$$\widetilde{R}^{(1)} f_{i+1} = f'(\widetilde{t}_{i+1}) + \frac{1}{2} f''(\widetilde{t}_{i+1})\delta_2 + o(\delta_2). \qquad (10.6)$$

Кроме того, по формуле Тейлора, примененной к $f'(t)$, имеем

$$f'(\widetilde{t}_{i+1}) = f'(t_i + \delta_1) = f'(t_i) + f''(t_i)\delta_1 + o(\delta_1).$$

Подставляя это равенство в (10.6), получим

$$\widetilde{R}^{(1)}f_{i+1} \;=\; f'(t_i) \;+\; f''(t_i)\delta_1 \;+\; \frac{1}{2}f''(t_i \;+\; \delta_1)\delta_2 \;+\; o(\delta_1) \;+\; o(\delta_2). \quad (10.7)$$

Подставим теперь (10.5) и (10.7) в (10.4). Получаем

$$R^{(2)}f_i = \lim_{0<\delta_1,\delta_2\to 0} \frac{f''(t_i)\delta_1 + \frac{1}{2}f''(t_i+\delta_1)\delta_2 - \frac{1}{2}f''(t_i)\delta_1 + o(\delta_1) + o(\delta_2)}{\delta_1+\delta_2} =$$

$$= \lim_{0<\delta_1,\delta_2\to 0}\left\{\frac{1}{2}\Big(\frac{\delta_1}{\delta_1+\delta_2}f''(t_i) + \frac{\delta_2}{\delta_1+\delta_2}f''(t_i+\delta_1)\Big) + \frac{o(\delta_1)}{\delta_1+\delta_2} + \frac{o(\delta_2)}{\delta_1+\delta_2}\right\} =$$

$$= \lim_{\delta_1\to 0}\frac{1}{2}\big(\alpha f''(t_i) + (1-\alpha)f''(t_i+\delta_1)\big) = \lim_{\delta_1\to 0}\frac{1}{2}f''(\xi) = \frac{1}{2}f''(t_i),$$

где $\alpha = \dfrac{\delta_1}{\delta_1+\delta_2}$, $\xi \in [t_i, t_i+\delta_1]$.

Случай произвольного k. Из приведенного доказательства для случая $k = 2$ видно, что утверждение достаточно доказать для случая, когда $f(t)$ — полином. В самом деле, если заменить функцию $f(t)$ ее рядом Тейлора в точке $t = t_0$, то остаточный член ряда Тейлора достаточно высокого порядка даст ноль при переходе к пределу $\delta_1 \to 0, \ldots, \delta_k \to 0$. (Достаточно взять остаточный член порядка $(k+1)$).

Пусть $p(t)$ — полином степени m. Обозначим через $p[t_1, \ldots, t_k]$ его разделенную разность порядка $(k-1)$, построенную по узлам $t_1, \ldots, t_k$.

Лемма 10.1. *Если $p(t)$ — полином степени m, то его разделенная разность порядка $q \leqslant m$ вида $p[t, t_1, \ldots, t_q]$ как функция первого узла t, где узлы $t_1, \ldots, t_q$ — фиксированы и различны, является полиномом степени $(m-q)$.*

Доказательство. Разделенная разность первого порядка полинома $p(t)$ имеет вид

$$p[t, t_1] = \frac{p(t_1) - p(t)}{t_1 - t}.$$

Но полином $p(t_1) - p(t)$ имеет корень $t = t_1$. Следовательно, по теореме Безу он делится без остатка на $(t_1 - t)$, причем частное от деления — полином степени $(m-1)$.

Далее доказываем по индукции. Предположим, что $p[t, t_1, \ldots, t_{q-1}]$ — полином степени $(m-q+1)$. Тогда его разделенная разность порядка q равна

$$p[t, t_1, \ldots, t_q] = \frac{p[t_1, \ldots, t_q] - p[t, t_1, \ldots, t_{q-1}]}{t_q - t}. \quad (10.8)$$

В числители правой части формулы (10.8) стоит полином $p[t_1, \ldots, t_q] - p[t, t_1, \ldots, t_{q-1}]$ степени $(m-q+1)$ (как функция t). Достаточно показать, что он имеет корень $t = t_q$. Отсюда по теореме Безу будет следовать, что $p[t, t_1, \ldots, t_q]$ является полиномом степени $(m-q)$, что и докажет лемму.

Подставим $t = t_q$ в числитель правой части формулы (10.8). Получим $p[t_1, \ldots, t_q] - p[t_q, t_1, \ldots, t_{q-1}]$. Это выражение равно нулю согласно следующей лемме.

Лемма 10.2. *Пусть имеется q попарно несовпадающих узлов $t_1, \ldots, t_q$. Тогда разделенная разность функции f в узлах $t_1, \ldots, t_q$ выражается формулой*

$$f[t_1, \ldots, t_q] = \sum_{j=1}^{q} \frac{f(t_j)}{(t_j - t_1)\ldots(t_j - t_{j-1})(t_j - t_{j+1})\ldots(t_j - t_q)} = \\ = \sum_{j=1}^{q} \frac{f(t_j)}{w'_{1,q}(t_j)}, \quad (10.9)$$

где $w_{1,q}(t) = (t - t_1)\ldots(t - t_q)$, а $w'_{1,q}(t)$ — производная функции $w_{1,q}(t)$ по t. В частности, $f[t_1, \ldots, t_q]$ не меняется при произвольной перестановке узлов $t_1, \ldots, t_q$.

Доказательство. Докажем (10.9) по индукции. Если $q = 2$, то (10.9) принимает вид

$$f[t_1, t_2] = \frac{f(t_1)}{t_1 - t_2} + \frac{f(t_2)}{t_2 - t_1} = \frac{f(t_2) - f(t_1)}{t_2 - t_1}.$$

Это совпадает с определением разделенной разности первого порядка.

Предположим теперь, что формула (10.9) справедлива для разделенных разностей порядка $(k-2)$ и докажем ее для разделенных разностей порядка $(k-1)$.

Имеем по предположению

$$f[t_2, \ldots, t_k] = \sum_{j=2}^{k} \frac{f(t_j)}{w'_{2,k-1}(t_j)} = \sum_{j=2}^{k} \frac{(t_j - t_1)f(t_j)}{(t_j - t_1)w'_{2,k-1}(t_j)}$$

и

$$f[t_1, \ldots, t_{k-1}] = \sum_{j=1}^{k-1} \frac{(t_j - t_k)f(t_j)}{(t_j - t_k)w'_{1,k-1}(t_j)}.$$

Правые части обеих формул приведены к общему знаменателю, который имеет вид

$$(t_j - t_1)\ldots(t_j - t_{j-1})(t_j - t_{j+1})\ldots(t_j - t_k) = w'_{1,k}(t_j).$$

Вычитая второе равенство из первого и деля его на $(t_k - t_1)$, получим равенство (10.9)

$$f[t_1, \ldots, t_k] = \frac{f[t_2, \ldots, t_k] - f[t_1, \ldots, t_{k-1}]}{t_k - t_1} = \\ = \frac{(t_k - t_1)f(t_1)}{(t_k - t_1)w'_{1,k}(t_1)} + \sum_{j=2}^{k-1} \frac{(t_j - t_1)f(t_j) - (t_j - t_k)f(t_j)}{(t_k - t_1)w'_{1,k}(t_j)} + \\ + \frac{(t_k - t_1)f(t_k)}{(t_k - t_1)w'_{1,k}(t_k)} = \sum_{j=1}^{k} \frac{f(t_j)}{w'_{1,q}(t_j)}.$$

Лемма 10.2 доказана. ■

Лемма 10.1 доказана. ■

Следствие 10.1. *Разделенная разность $(m+1)$-го порядка от полинома степени m равна нулю, а разделенные разности m-го порядка равна константе.*

Лемма 10.3. *Пусть $p(t)$ является интерполяционным полиномом степени m для функции $f(t)$, совпадающим с ней в $(m+1)$ узлах $t_1, \ldots, t_{m+1}$. Тогда*

$$p(t) = f(t_1) + f[t_1, t_2](t - t_1) + f[t_1, t_2, t_3](t - t_1)(t - t_2) + \ldots \\ \ldots + f[t_1, t_2, \ldots, t_{m+1}](t - t_1)(t - t_2)\ldots(t - t_m) \quad (10.10)$$

(*интерполяционная формула Ньютона*).

Доказательство. По определению разделенных разностей имеем для полинома $p(t)$ степени m, интерполирующего $f(t)$ в узлах $t_1, \ldots, t_{m+1}$, следующее выражение:

$$p[t, t_1, \ldots, t_m] = f[t_1, \ldots, t_{m+1}] + (t - t_{m+1})p[t, t_1, \ldots, t_{m+1}] = f[t_1, \ldots, t_{m+1}].$$

В последнем равенстве использовано, что $p[t, t_1, \ldots, t_{m+1}] = 0$ как разделенная разность $(m+1)$-го порядка от полинома степени m. Аналогично:

$$p[t, t_1, \ldots, t_{m-1}] = f[t_1, \ldots, t_m] + (t - t_m)p[t, t_1, \ldots, t_m],$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$p[t, t_1] = f[t_1, t_2] + (t - t_2)p[t, t_1, t_2],$$
$$p(t) = p[t] = f[t_1] + (t - t_1)p[t, t_1].$$

Здесь мы несколько раз воспользовались тем, что полином p и функция f совпадают в узлах $t_1, \ldots, t_{m+1}$, а, следовательно, все их разделенные разности, построенные по этим узлам, также совпадают.

Подставляя полученные равенства последовательно одно в другое, начиная с последнего, получим искомую формулу (10.10). Лемма доказана. ■

Следствие 10.2. *Для любого полинома степени m имеет место*

$$p(t) = p(t_1) + p[t_1, t_2](t - t_1) + p[t_1, t_2, t_3](t - t_1)(t - t_2) + \ldots \\ \ldots + p[t_1, t_2, \ldots, t_{m+1}](t - t_1)(t - t_2)\ldots(t - t_m). \quad (10.11)$$

Доказательство. Формула (10.11) следует из леммы 10.3 поскольку все разделенные разности, построенные по узлам $t_1, \ldots, t_{m+1}$ для $p(t)$ и $f(t)$ совпадают. Следствие доказано. ■

Для доказательства теоремы 10.1 устремим в формуле (10.11) $t_2 \to t_1, \ldots, t_{m+1} \to t_1$. Тогда $p[t_1, \ldots, t_k] \to R^{(k-1)}p(\xi)$ — разделенная разность порядка $(k-1)$ полинома $p(t)$, построенная по k совпадающим точкам $t = t_1 = \cdots = t_k = \xi$. Следовательно, в пределе получим

$$p(t) = p(\xi) + R^{(1)}p(\xi)(t - \xi) + R^{(2)}p(\xi)(t - \xi)^2 + \cdots + R^{(m)}p(\xi)(t - \xi)^m$$

для любого t. Сравнивая эту формулу с формулой Тейлора и пользуясь произвольностью t, получаем

$$R^{(i)}p(\xi) = \frac{p^i(\xi)}{i!}.$$

Теорема доказана 10.1. ■

Пример 6. Пусть дан $(m+1)$ узел $t_1, \ldots, t_{m+1}$, причем узлы $t_1, \ldots, t_m$ совпадают: $t_i = t_1$, где $2 \leqslant i \leqslant m$. Вычислим разделенную разность $f[t_1, \ldots, t_{m+1}]$ для гладкой функции $f(t)$. Согласно (10.10) имеем при $t = t_{m+1}$:

$$\begin{aligned} f(t_{m+1}) = f[t_1] + f[t_1, t_1](t_{m+1} - t_1) + \ldots \\ \ldots + f[t_1, \ldots, t_m = t_1](t_{m+1} - t_1)^{m-1} + \\ + f[t_1, \ldots, t_m = t_1, t_{m+1}](t_{m+1} - t_1)^m = \\ = \sum_{j=0}^{m-1} \frac{f^{(j)}(t_1)}{j!}(t_{m+1} - t_1)^j + f[t_1, \ldots, t_m = t_1, t_{m+1}](t_{m+1} - t_1)^m. \end{aligned}$$

Отсюда получаем искомый ответ:

$$f[t_1, \ldots, t_1, t_{m+1}] = \frac{f(t_{m+1})}{(t_{m+1} - t_1)^m} - \sum_{j=0}^{m-1} \frac{f^{(j)}(t_1)}{j!(t_{m+1} - t_1)^{m-j}}.$$

10.3 Свойства разделенных разностей

1) Если $f(t) = a_0 + a_1 t + \cdots + a_m t^m$, то $f[t_1, \ldots, t_{m+1}] = a_m$, $f[t_1, \ldots, t_{m+2}] = 0$ для любого набора узлов $t_1, \ldots, t_{m+2}$. Другими словами, разделенная разность порядка m для полинома степени m равна его старшему коэффициенту.

Следствие 10.3. *Разделенная разность порядка m, построенная по узлам $t_1, \ldots, t_{m+1}$ для функции $f(t)$ равна старшему коэффициенту его интерполяционного полинома степени m, построенного по узлам $t_1, \ldots, t_{m+1}$ (это следует из совпадения разделенных разностей, построенных по узлам $t_1, \ldots, t_{m+1}$, для $f(t)$ и его интерполяционного полинома).*

2) Любая разделенная разность является симметричной функцией своих аргументов. Это следует из симметричности разделенных разностей как функций различных узлов и из определения разделенных разностей при совпадающих узлах, а также — из единственности интерполяционного полинома.

3) Линейность: если $f(t) = \alpha g(t) + \beta h(t)$, то $f[t_1, \ldots, t_{m+1}] = \alpha g[t_1, \ldots, t_{m+1}] + \beta h[t_1, \ldots, t_{m+1}]$. Вытекает из определения разделенных разностей.

4) Если $f(t) = g(t)h(t)$, то

$$f[t_1, \ldots, t_{m+1}] = \sum_{k=1}^{m+1} g[t_1, \ldots, t_k] h[t_k, \ldots, t_{m+1}]$$

(формула Лейбница). Доказательство несложно провести по индукции. Оставляем его в качестве упражнения.

5) Если $p_m(t)$ — интерполяционный полином для функции $f(t)$, построенный по узлам $t_1, \ldots, t_{m+1}$, то для любого t

$$f(t) = p_m(t) + f[t_1, \ldots, t_{m+1}, t](t - t_1) \ldots (t - t_{m+1}).$$

Доказательство оставляем в качестве упражнения.

10.4 Усеченная степенная функция

Пусть $m \geqslant 0$ — целое число. Рассмотрим функцию $\sigma_m(z,t) = (z-t)_+^m$, где индексом "+" мы обозначаем положительную часть функции: $f_+(t) = \max\{f(t), 0\}$. Таким образом, $\sigma_m(z,t) = \big(\max(0, z-t)\big)^m$. При фиксированном t функция $\sigma_m(z,t)$, рассматриваемая как функция z, называется *усеченной степенной функцией.*

Имеет место

Лемма 10.4. *Для производных усеченной степенной функции справедливы формулы*:

$$\begin{aligned} &\frac{\partial \sigma_m}{\partial z} = m\sigma_{m-1}, \ \frac{\partial \sigma_m}{\partial t} = -m\sigma_{m-1}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\frac{\partial^k \sigma_m}{\partial z^k} = \frac{m!}{(m-k)!}\sigma_{m-k}, \ \frac{\partial^k \sigma_m}{\partial t^k} = \frac{(-1)^k m!}{(m-k)!}\sigma_{m-k}, \qquad k = 0, \ldots, m \end{aligned} \tag{10.12}$$

Доказательство получается простым дифференцированием. ■

Фиксируем набор *узлов* $t_0 < t_1 < \ldots < t_m$ и рассмотрим функцию $M(t) = \sigma_{m-1}[t_0, \ldots, t_m] = \sigma_{m-1}[t_0, \ldots, t_m](t)$, где разделенная разность $\sigma_{m-1}[t_0, \ldots, t_m]$ вычисляется от $\sigma_{m-1}(z,t)$, рассматриваемой как функция z при фиксированном t.

Лемма 10.5. *Пусть $m > 0$. Тогда $M(t) = 0$, если $t \leqslant t_0$ или $t \geqslant t_m$.*

Доказательство. Если $t \leqslant t_0$, то $\sigma_{m-1}[t_0, \ldots, t_m](t)$ — это разделенная разность порядка m от полинома степени $(m-1)$. Следовательно, она равна нулю.

Если $t \geqslant t_m$, то во всех узлах $t_0, \ldots, t_m$ справедливо равенство $\sigma_{m-1}(z,t)\big|_{z=t_i} = 0$. Следовательно, соответствующая разделенная разность равна нулю. Лемма доказана. ■

Из доказанной леммы вытекает, что разделенная разность $\sigma_{m-1}[t_0,\ldots,t_m](t)$ является непрерывной функцией параметра t, отличной от нуля лишь на интервале (t_0,t_m). Согласно формуле (10.9) имеем

$$\sigma_{m-1}[t_0,\ldots,t_m](t)=\sum_{j=0}^{m}\frac{(t_j-t)_+^{m-1}}{(t_j-t_0)\ldots(t_j-t_{j-1})(t_j-t_{j+1})\ldots(t_j-t_m)}=$$
$$=\sum_{j=0}^{m}\frac{(t_j-t)_+^{m-1}}{w'_{0,m}(t_j)},$$

где $w_{0,m}(t)=(t-t_0)\ldots(t-t_m)$, а $w'_{0,m}(t)$ — производная функции $w_{0,m}(t)$.

Согласно выведенным выше формулам (10.12) для производных усеченной разности имеем

$$\frac{\partial^k\sigma_{m-1}[t_0,\ldots,t_m]}{\partial t^k}=\frac{(-1)^k(m-1)!}{(m-1-k)!}\sigma_{m-1-k}[t_0,\ldots,t_m]=$$
$$=\frac{(-1)^k(m-1)!}{(m-1-k)!}\sum_{j=0}^{m}\frac{(t_j-t)_+^{m-1-k}}{w'_{0,m}(t_j)},\qquad k=1,\ldots,m-2.\quad(10.13)$$

В правой части (10.13) стоят непрерывные функции от t.

На основе разделенных разностей мы будем строить весовые функции $N_i(t)$ в формуле (10.2) с целью обобщения рациональных кривых Безье:

$$\mathbf{r}(t)=\frac{\sum\limits_{i=1}^{n}N_i(t)\omega_i\mathbf{p}_i}{\sum\limits_{i=1}^{n}N_i(t)\omega_i},\qquad t_0\leqslant t\leqslant t_m.$$

Определение 10.2. *Нормированным B-сплайном m-го порядка* $N_{i,m}(t)$ для неубывающей последовательности узлов $t_i\leqslant t_{i+1}\leqslant\ldots\leqslant t_{i+m}$, отсчитываемой от первого узла t_i, называется функция

$$N_{i,m}(t)=(t_{i+m}-t_i)\sigma_{m-1}[t_i,\ldots,t_{i+m}](t).$$

Соответственно, *ненормированным B-сплайном* $M_{i,m}(t)$ называется функция

$$M_{i,m}(t)=\sigma_{m-1}[t_i,\ldots,t_{i+m}](t).$$

Определение 10.3. *Нормированным B-сплайном m-го порядка* $N_{m,i}(t)$ для неубывающей последовательности узлов $t_{i-m}\leqslant t_{i-m+1}\leqslant\ldots\leqslant t_{i-1}\leqslant t_i$, отсчитываемой от последнего узла t_i, называется функция

$$N_{m,i}(t)=(t_i-t_{i-m})\sigma_{m-1}[t_{i-m},\ldots,t_i](t).$$

Соответственно определяется и *ненормированный B-сплайн* $M_{m,i}(t)$:

$$M_{m,i}(t)=\sigma_{m-1}[t_{i-m},\ldots,t_i](t).$$

Замечание. Если порядок B-сплайна определять не числом узлов минус единица, а степенью усеченной функции (как многочлена t), то введенные выше B-сплайны будут иметь (по определению) порядок не m, а $(m-1)$. В этом случае для них применяются другие обозначения: N_{i-}^{m-1}, N_{-i}^{m-1}, M_{i-}^{m-1}, M_{-i}^{m-1}. При этом:

$$N_{i-}^{m-1}(t) = N_{i,m}(t), \qquad N_{-i}^{m-1}(t) = N_{m,i}(t),$$
$$M_{i-}^{m-1}(t) = M_{i,m}(t), \qquad M_{-i}^{m-1}(t) = M_{m,i}(t).$$

И те и другие обозначения присутствуют в литературе по B-кривым.

Рассмотрим (для простоты) бесконечную в обе стороны последовательность неубывающих узлов $\{t_i\}$, где $t_i \in \mathbb{R}$, $i \in \mathbb{Z}$, $t_i \leqslant t_{i+1}$ для любого i. Фиксируем $m > 0$. Тогда определены множества B-сплайнов: сплайны $N_{i,m}(t)$ привязаны к узлу i как к первому узлу, сплайны $N_{m,i}(t)$ привязаны к узлу i как к последнему узлу ($i \in \mathbb{Z}$).

Лемма 10.6. *Если $t \in [t_j, t_{j+1}]$, то среди функций $N_{i,m}(t)$ и $N_{m,i}(t)$ только m функций в каждом из этих двух классов могут быть отличны от нуля в точке t. А именно, отличны от нуля в точке t могут быть лишь следующие функции*:

$$N_{j-m+1,m}(t), \ldots, N_{j,m}(t), \qquad \text{и} \qquad N_{m,j+1}(t), \ldots, N_{m,j+m}(t).$$

Доказательство. Имеем $N_{i,m}(t) = (t_{i+m} - t_i)\sigma_{m-1}[t_i, \ldots, t_{i+m}](t)$. Следовательно, $N_{i,m}(t) = 0$ при $t \notin (t_i, t_{i+m})$. Но по условию $t \in [t_j, t_{j+1}]$. Значит $N_{i,m}(t) = 0$, если пересечение отрезка $[t_j, t_{j+1}]$ и интервала (t_i, t_{i+m}) пусто. Получаем, что $N_{i,m}(t) = 0$, если $i > j$ или $i < j - m + 1$.

Для B-сплайнов $N_{m,i}(t)$, привязанных к правой точке, доказательство аналогично. Лемма доказана. ■

Лемма 10.7. *Для любого $t \in \mathbb{R}$ имеет место*

$$\sum_i N_{i,m}(t) = \sum_i N_{m,i}(t) \equiv 1.$$

Доказательство. Имеем по определению разделенной разности и по определению B-сплайнов:

$$N_{\alpha,m}(t) = \sigma_{m-1}[t_{\alpha+1}, \ldots, t_{\alpha+m}] - \sigma_{m-1}[t_\alpha, \ldots, t_{\alpha+m-1}],$$
$$N_{m,\beta}(t) = \sigma_{m-1}[t_{\beta-m+1}, \ldots, t_\beta] - \sigma_{m-1}[t_{\beta-m}, \ldots, t_{\beta-1}].$$

Следовательно, согласно лемме 10.6, при $t \in [t_j, t_{j+1}]$ имеем

$$\sum_{i=-\infty}^{\infty} N_{i,m}(t) = \sum_{i=j-m+1}^{j} N_{i,m}(t) =$$
$$= \sum_{i=j-m+1}^{j} \sigma_{m-1}[t_{i+1}, \ldots, t_{i+m}] - \sum_{i=j-m+1}^{j} \sigma_{m-1}[t_i, \ldots, t_{i+m-1}] =$$
$$= \sigma_{m-1}[t_{j+1}, \ldots, t_{j+m}] - \sigma_{m-1}[t_{j-m+1}, \ldots, t_j].$$

Но $\sigma_{m-1}[t_{j+1},\dots,t_{j+m}]$ есть разделенная разность порядка $(m-1)$ от полинома степени $(m-1)$, равного $(z-t)^{m-1}$ — поскольку усеченная степенная функция $(z-t)_+^{m+1}$ при $t\in[t_j,t_{j+1}]$ на узлах $t_{j+1},\dots,t_{j+m}$ совпадает с неусеченной степенной функцией $(z-t)^{m-1}$, т.е. с полиномом.

Разделенная разность порядка $(m-1)$ от полинома степени $(m-1)$ равна его старшему коэффициенту. В данном случае она равна 1, так как старший коэффициент полинома $(z-t)^{m-1}$ равен 1. Следовательно, $\sigma_{m-1}[t_{j+1},\dots,t_{j+m}]\equiv 1$. Далее, при $t\in[t_j,t_{j+1}]$ имеем: $\sigma_{m-1}[t_{j-m+1},\dots,t_j]=0$. Следовательно, $\sum\limits_i N_{i,m}(t)\equiv 1-0=1$.

Аналогично — для функции $N_{m,i}(t)$. Лемма доказана. ■

10.5 Рекуррентные соотношения для B-сплайнов

Теорема 10.2 (формула Кокса-Де Бура). *Справедливы следующие формулы для ненормированных B-сплайнов при любых t, $m\geqslant 1$, $\alpha\in\mathbb{Z}$ и $\beta=\alpha+m$:*

$$\begin{aligned}
M_{\alpha,m}(t)&=\frac{(t_\beta-t)M_{\alpha+1,m-1}(t)+(t-t_\alpha)M_{\alpha,m-1}(t)}{t_\beta-t_\alpha},\\
M_{m,\beta}(t)&=\frac{(t_\beta-t)M_{m-1,\beta}(t)+(t-t_\alpha)M_{m-1,\beta-1}(t)}{t_\beta-t_\alpha},\\
M_{\alpha-}^{m-1}(t)&=\frac{(t_\beta-t)M_{\alpha+1-}^{m-2}(t)+(t-t_\alpha)M_{\alpha-}^{m-2}(t)}{t_\beta-t_\alpha},\\
M_{-\beta}^{m-1}(t)&=\frac{(t_\beta-t)M_{-\beta}^{m-2}(t)+(t-t_\alpha)M_{-\beta-1}^{m-2}(t)}{t_\beta-t_\alpha}.
\end{aligned}\tag{10.14}$$

Здесь ненормированные B-сплайны первого порядка равны по определению разделенным разностям функции $\sigma_0(z)=(z-t)_+^0$:

$$\begin{aligned}
M_{i,1}(t)=M_{i-}^0&=\begin{cases}\dfrac{1}{t_{i+1}-t_i}, & \text{если } t\in[t_i,t_{i+1}),\\ 0, & \text{иначе};\end{cases}\\
M_{1,i}(t)=M_{-i}^0&=\begin{cases}\dfrac{1}{t_i-t_{i-1}}, & \text{если } t\in[t_{i-1},t_i),\\ 0, & \text{иначе}.\end{cases}
\end{aligned}\tag{10.15}$$

Доказательство. Имеем

$$\sigma_{m-1}(z)=(z-t)_+^{m-1}=(z-t)(z-t)_+^{m-2}=g(z)\sigma_{m-2}(z),$$

где $g(z)=(z-t)$.

Предположим, что $t\in[t_\alpha,t_\beta]$, где $\beta=\alpha+m$. Тогда по формуле Лейбница для разделенных разностей:

$$\begin{aligned}
\sigma_{m-1}[t_\alpha,\dots,t_\beta]&=g[t_\alpha]\sigma_{m-2}[t_\alpha,\dots,t_\beta]+\\
&\quad+g[t_\alpha,t_{\alpha+1}]\sigma_{m-2}[t_{\alpha+1},\dots,t_\beta]+0+\dots+0=\\
&\qquad=(t_\alpha-t)\sigma_{m-2}[t_\alpha,\dots,t_\beta]+\sigma_{m-2}[t_{\alpha+1},\dots,t_\beta],
\end{aligned}$$

так как $g[t_\alpha, t_{\alpha+1}, t_{\alpha+2}, \ldots, t_{\alpha+k}] = 0$ при $k \geqslant 2$, а $g[t_\alpha, t_{\alpha+1}] = 1$ — это коэффициент при старшей степени многочлена $g(z) = z - t$.

Подставим в последнее равенство определение разделенной разности:

$$\sigma_{m-2}[t_\alpha, \ldots, t_\beta] = \frac{\sigma_{m-2}[t_{\alpha+1}, \ldots, t_\beta] - \sigma_{m-2}[t_\alpha, \ldots, t_{\beta-1}]}{t_\beta - t_\alpha}.$$

Получим

$$\begin{aligned}\sigma_{m-1}[t_\alpha, \ldots, t_\beta] &= \sigma_{m-2}[t_{\alpha+1}, \ldots, t_\beta] + \\ &+ \frac{t_\alpha - t}{t_\beta - t_\alpha}\big(\sigma_{m-2}[t_{\alpha+1}, \ldots, t_\beta] - \sigma_{m-2}[t_\alpha, \ldots, t_{\beta-1}]\big) = \\ &= \frac{(t_\beta - t)\sigma_{m-2}[t_{\alpha+1}, \ldots, t_\beta] + (t - t_\alpha)\sigma_{m-2}[t_\alpha, \ldots, t_{\beta-1}]}{t_\beta - t_\alpha}.\end{aligned}$$

Если переписать последнее равенство, подставив в него определение ненормированного B-сплайна, получим утверждение теоремы. Теорема доказана. ■

Теорема Кокса-Де Бура позволяет вычислять B-сплайны рекуррентно с использованием формул (10.14) и (10.15).

Следствие 10.4. *Пусть имеется произвольная последовательность (не обязательно неубывающих) узлов $T = \{t_0, \ldots, t_m\}$. Пусть среди них есть 2 неравных друг другу узла t_a и t_b: $t_a \neq t_b$, где $0 \leqslant a, b \leqslant m$. Обозначим*

$$\begin{gathered}M^T(t) = \sigma_{m-1}[T] = \sigma_{m-1}[t_0, \ldots, t_m], \\ M^{T\setminus a}(t) = \sigma_{m-2}[T \setminus a] = \sigma_{m-2}[t_0, \ldots, \widehat{t}_a, \ldots, t_m], \\ M^{T\setminus b}(t) = \sigma_{m-2}[T \setminus b] = \sigma_{m-2}[t_0, \ldots, \widehat{t}_b, \ldots, t_m], \\ p_{ab}(t) = \frac{t_b - t}{t_b - t_a}, \qquad q_{ab}(t) = \frac{t - t_a}{t_b - t_a}, \qquad (p_{ab} + q_{ab} \equiv 1),\end{gathered}$$

где "домиком" отмечены отброшенные узлы. Тогда

$$\begin{gathered}M^T(t) = p_{ab}(t)M^{T\setminus a}(t) + q_{ab}(t)M^{T\setminus b}(t), \\ M^{\{t_a, t_b\}}(t) = \begin{cases} \dfrac{1}{|t_b - t_a|}, & \text{если } t \in [t_a \wedge t_b, t_a \vee t_b), \\ 0, & \text{иначе,}\end{cases}\end{gathered}$$

где $\alpha \vee \beta = \max\{\alpha, \beta\}$, $\alpha \wedge \beta = \min\{\alpha, \beta\}$.

Доказательство следует из теоремы Кокса-Де Бура и из того, что разделенная разность не зависит от порядка узлов. Следствие доказано. ■

Следствие 10.5. *B-сплайны — неотрицательные функции*:

$$\begin{gathered}M_{i,m}(t) \geqslant 0, \qquad M_{i-}^{m-1}(t) \geqslant 0, \\ M_{m,i}(t) \geqslant 0, \qquad M_{-i}^{m-1}(t) \geqslant 0.\end{gathered}$$

Доказательство следует из неотрицательности B-сплайнов первого порядка (формула (10.15)) и из того, что рекуррентная формула (10.14) переводит совокупность неотрицательных функций порядка $(m-1)$ в совокупность неотрицательных функций порядка m. Следствие доказано. ■

Следствие 10.6. *Для нормированных B-сплайнов формулы Кокса-Де Бура принимают вид* ($\beta = \alpha + m$):

$$N_{\alpha,m}(t) = \frac{t_\beta - t}{t_\beta - t_{\alpha+1}} N_{\alpha+1,m-1}(t) + \frac{t - t_\alpha}{t_{\beta-1} - t_\alpha} N_{\alpha,m-1}(t),$$

$$N_{m,\beta}(t) = \frac{t_\beta - t}{t_\beta - t_{\alpha+1}} N_{m-1,\beta}(t) + \frac{t - t_\alpha}{t_{\beta-1} - t_\alpha} N_{m-1,\beta-1}(t),$$

$$N^0_{i-}(t) = N_{i,1}(t) = \begin{cases} 1, & \text{если } t \in [t_i, t_{i+1}), \\ 0, & \text{иначе}; \end{cases}$$

$$N^0_{-i}(t) = N_{1,i}(t) = \begin{cases} 1, & \text{если } t \in [t_{i-1}, t_i), \\ 0, & \text{иначе}. \end{cases}$$

Доказательство следует из соотношений:

$$N_{\alpha,m}(t) = (t_{\alpha+m} - t_\alpha) M_{\alpha,m}(t), \qquad N_{m,\beta}(t) = (t_\beta - t_{\beta-m}) M_{m,\beta}(t).$$

Следствие доказано. ■

Для сокращения объема вычислений следует сначала вычислить ненормированные B-сплайны нужного порядка по формулам Кокса-Де Бура, а затем нормировать их. Если требуется вычислить все ненулевые B-сплайны порядка m для $t \in [t_i, t_{i+1}]$ в неубывающей последовательности узлов $\{t_i\}$, то вычисления производятся по схеме:

$$\begin{array}{cccc}
 & & & M_{i,1}(t) \\
 & & M_{i-1,2}(t) & M_{i,2}(t) \\
 & & \vdots & \vdots \\
M_{i-m+1,m}(t) & \dots & M_{i-1,m}(t) & M_{i,m}(t) \\
\downarrow & \downarrow & \downarrow & \downarrow \\
N_{i-m+1,m}(t) & \dots & N_{i-1,m}(t) & N_{i,m}(t).
\end{array}$$

Аналогичная схема вычислений применяется для B-сплайнов остальных трех типов.

Следствие 10.7. *Для любых $m \geqslant 2$ и $\alpha \in \mathbb{Z}$ функции $N_{\alpha,m}(t)$ и $N_{m,\beta}(t)$, где $\beta = \alpha + m$, построенные по неубывающей последовательности узлов $\{t_i\}_{i\in\mathbb{Z}}$, являются непрерывными $(m-2)$ раза дифференцируемыми функциями, обращающимися в ноль вне отрезка $[t_\alpha, t_\beta]$. На отрезке $[t_\alpha, t_\beta]$ они имеют производные до порядка $(m-1)$ включительно. Функции $N_{\alpha,m}(t)$ и $N_{m,\beta}(t)$ зависят (как от параметров) лишь от значений узлов $t_\alpha, \dots, t_\beta$.*

Определение 10.4. *B-кривой* порядка m, построенной по вершинам $\mathbf{p}_1, \ldots, \mathbf{p}_n$, $n \geqslant m$, с весами $\omega_1, \ldots, \omega_n$, называется кривая

$$\mathbf{r}(t) = \frac{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i \mathbf{p}_i}{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i}, \tag{10.16}$$

определяемая опорными точками $\{\mathbf{p}_i\}_{i=1}^n$, весами $\{\omega_i\}_{i=1}^n$ и последовательностью узлов $t_1, \ldots, t_{n+m}$.

Пример 7 (Частный случай). Если $\omega_1 = \cdots = \omega_n$, то формула (10.16) приобретает вид

$$\mathbf{r}(t) = \sum_{i=1}^{n} N_{i,m}(t)\mathbf{p}_i.$$

Определение 10.5. B-кривая, построенная по множеству неравностоящих узлов, называется *неоднородной*.

Определение 10.6. Множество узлов $\{t_1, \ldots, t_{n+m}\}$ называется *расширенным* множеством узлов B-кривой.

10.6 Алгоритм вычисления радиус-вектора B-кривой

Пусть даны опорные точки $\mathbf{p}_1, \ldots, \mathbf{p}_n$ и их веса $\omega_1, \ldots, \omega_n$, а также расширенное множество неубывающих узлов $\{t_1, \ldots, t_{n+m} : t_1 \leqslant \ldots \leqslant t_{n+m}\}$. Опишем алгоритм вычисления радиус-вектора B-кривой

$$\mathbf{r}(t) = \frac{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i \mathbf{p}_i}{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i}.$$

1) Фиксируем t. Положим $t_0 = -\infty$, $t_{n+m+1} = +\infty$. Тогда существует единственное i_0 такое, что $t \in [t_{i_0}, t_{i_0+1})$, где $0 \leqslant i_0 \leqslant n+m$.

2) Для определенного выше индекса i_0, если $1 \leqslant i_0 \leqslant n+m$, вычисляем единственное ненулевое значение ненормированного B-сплайна первого уровня ($m = 1$):

$$M_{i_0,1}(t) = \frac{1}{t_{i_0+1} - t_{i_0}} \qquad (1 \leqslant i_0 \leqslant n+m).$$

Напомним, что i_0 было выбрано на первом шаге, так чтобы $t \in [t_{i_0}, t_{i_0+1})$. При этом, в силу условия $1 \leqslant i_0 \leqslant n+m$, знаменатель $t_{i_0+1} - t_{i_0} < \infty$, а в силу условия $t \in [t_{i_0}, t_{i_0+1})$ имеем $t_{i_0+1} - t_{i_0} > 0$. Следовательно, $0 < M_{i_0,1}(t) < \infty$.

Рис. 10.1:

Для индекса i_0, находящегося вне множества $\{1, \ldots, n+m\}$, значение $M_{i_0,1}$ не вычисляется.

3) С помощью соотношений Кокса-Де Бура вычисляем все отличные от нуля в точке t ненормированные сплайны m-го порядка $M_{j,m}(t)$ при $1 \leqslant j \leqslant n$, см. рис. 10.1:

$$M_{j,m}(t) = \frac{(t_{j+m} - t)M_{j+1,m-1}(t) + (t - t_j)M_{j,m-1}(t)}{t_{j+m} - t_j}.$$

В частности:

$$\begin{aligned} M_{i_0,s}(t) &= \frac{(t - t_{i_0})^{s-1}}{(t_{i_0+1} - t_{i_0}) \ldots (t_{i_0+s} - t_{i_0})}, \\ N_{i_0,s}(t) &= \frac{(t - t_{i_0})^{s-1}}{(t_{i_0+1} - t_{i_0}) \ldots (t_{i_0+s-1} - t_{i_0})}, \end{aligned} \tag{10.17}$$

где $1 \leqslant s \leqslant m$, $t \in [t_{i_0}, t_{i_0+1})$. Положив в (10.17) $i_0 = 1$, получим

$$N_{1,s}(t) = \frac{(t - t_1)^{s-1}}{(t_2 - t_1) \ldots (t_s - t_1)}, \qquad 1 \leqslant s \leqslant m, \qquad t \in [t_1, t_2).$$

Лемма 10.8. *Имеет место*:

$$N_{1,m}(t_1) = \begin{cases} 1, & \textit{если} \quad t_1 = \cdots = t_m, \\ 0, & \textit{иначе}. \end{cases}$$

Доказательство. Пусть $t_1 = \cdots = t_k < t_{k+1}$ для некоторого $k \leqslant m$. Требуется доказать, что в этом случае

$$N_{1,m}(t_1) = \begin{cases} 1, & \text{если} \quad k = m, \\ 0, & \text{если} \quad k < m. \end{cases} \tag{10.18}$$

Выберем $t \in [t_k, t_{k+1})$. В частности, t может принимать значение $t = t_k = t_1$.

Согласно (10.17),

$$M_{k,1}(t_k) = \frac{1}{t_{k+1} - t_k},$$

$$M_{k,s}(t_k) = \frac{(t_k - t_k)^{s-1}}{(t_{k+1} - t_k)\dots(t_{k+s} - t_k)} = 0 \qquad (2 \leqslant s \leqslant m).$$

Знаменатель в последней формуле отличен от нуля, поскольку $t_{k+1}, \dots, t_{k+s} > t_k$. Следовательно, поскольку $t_1 = t_k$, имеем

$$M_{k,1}(t_1) = \frac{1}{t_{k+1} - t_k}, \qquad M_{k,s}(t_1) = 0 \qquad (2 \leqslant s \leqslant m).$$

Отсюда следует, что

$$N_{k,1}(t_1) = 1, \qquad N_{k,s}(t_1) = 0 \qquad (2 \leqslant s \leqslant m). \tag{10.19}$$

Покажем, что для любого $2 \leqslant i \leqslant m$ выполнены равенства:

$$N_{k-j,i}(t_1) = N_{k-j,i}(t_k) = \begin{cases} 1, & \text{если} \quad j = i-1, \\ 0, & \text{если} \quad 0 \leqslant j \leqslant i-2. \end{cases} \tag{10.20}$$

Будем доказывать (10.20) по индукции.

Для любого $t \in [t_k, t_{k+1})$ при $i = 2$ имеем по формуле Кокса-Де Бура:

$$M_{k-1,2}(t) = \frac{(t_{k+1} - t)M_{k,1} + (t - t_{k-1})M_{k-1,1}}{(t_{k+1} - t_{k-1})} = \frac{\dfrac{t_{k+1} - t}{t_{k+1} - t_k} + 0}{t_{k+1} - t_k},$$

так как $t_{k-1} = t_k$, $M_{k-1,1} = 0$.

Положим $t = t_k = t_1$, тогда:

$$M_{k-1,2}(t_1) = M_{k-1,2}(t_k) = \frac{1}{t_{k+1} - t_k} = \frac{1}{t_{k+1} - t_1}.$$

Следовательно, поскольку по предположению $t_{k-1} = t$, имеем

$$N_{k-1,2}(t_1) = (t_{k+1} - t_{k-1})M_{k-1,2}(t_1) = \frac{t_{k+1} - t_1}{t_{k+1} - t_1} = 1.$$

Далее, согласно (10.19) справедливо равенство $N_{k,2}(t_1) = 0$. Таким образом, (10.20) доказано при $i = 2$. Предположим теперь, что (10.20) верно для некоторого $2 \leqslant i < m$ и докажем, что тогда оно верно и для $(i+1)$. Согласно предположению индукции для значений $M_{k-j,i}(t_1)$ выполнены равенства:

$$M_{k-i+1,i}(t_1) = \frac{1}{t_{k+1} - t_{k-i+1}}$$

и $M_{k-j,i}(t_1) = 0$ при $0 \leqslant j \leqslant i-2$. По формулам Кокса-Де Бура величины $M_{k-j,i+1}$ вычисляются по следующему правилу, изображенному на диаграмме 10.2.

Рис. 10.2:

Из приведенной диаграммы видно, что $M_{k-j,i+1}(t_1) = 0$ при $j \leqslant (i+1) - 3 = i-2$, поскольку согласно формуле Кокса-Де Бура эти величины являются линейной комбинацией нулевых (по предположению индукции) элементов предыдущего уровня. Для завершения доказательства (10.20) осталось проверить, что

$$M_{k-i,i+1}(t_1) = \frac{1}{t_{k+1} - t_{k-i}} \qquad \text{(что равносильно } N_{k-i,i+1}(t_1) = 1\text{)},$$
$$M_{k-i+1,i+1}(t_1) = 0.$$

По формуле Кокса-Де Бура имеем

$$M_{k-i,i+1}(t_1) = \frac{(t_{k+1} - t_1)M_{k-i+1,i}(t_1) + (t_1 - t_{k-i})M_{k-i,i}(t_1)}{t_{k+1} - t_{k-i}} =$$
$$= \frac{(t_{k+1} - t_1)M_{k-i+1,i}(t_1)}{t_{k+1} - t_1} = M_{k-i+1,i}(t_1) = \frac{1}{t_{k+1} - t_{k-i+1}} = \frac{1}{t_{k+1} - t_{k-i}},$$
$$M_{k-i+1,i+1}(t_1) = \frac{0 + (t_1 - t_{k-i+1})M_{k-i+1,i}(t_1)}{t_{k+2} - t_{k-i+1}} = 0,$$

так как $t_{k-i} = t_1$, $t_{k-i+1} = t_{k-i}$.

Формула (10.20) доказана. Но из (10.20) следует (10.18), поскольку

$$N_{1,m}(t_1) = N_{k-(k-1),m}(t_1) =$$
$$= N_{k-j,i}(t_1) = \begin{cases} 1, & \text{если} \quad j = k-1 = i-1 = m-1, \\ 0, & \text{если} \quad 0 \leqslant j = k-1 \leqslant i-2 = m-2. \end{cases}$$

Здесь $j = k-1$, $i = m$.

Следовательно,

$$N_{1,m}(t_1) = \begin{cases} 1, & \text{если} \quad k = m, \\ 0, & \text{если} \quad k < m. \end{cases}$$

Лемма доказана. ■

Следствие 10.8. *Если $t_1 = \cdots = t_m$, то B-кривая начинается в первой опорной точке $\mathbf{p}_1$.*

Доказательство. Имеем

$$\mathbf{r}(t_1) = \frac{\sum\limits_{i=1}^{n} N_{i,m}(t_1)\omega_i\mathbf{p}_i}{\sum\limits_{i=1}^{n} N_{i,m}(t_1)\omega_i} = \frac{\omega_1\mathbf{p}_1 N_{1,m}(t_1) + \sum\limits_{i=2}^{n} N_{i,m}(t_1)\omega_i\mathbf{p}_i}{\omega_1 N_{1,m}(t_1) + \sum\limits_{i=2}^{n} N_{i,m}(t_1)\omega_i} = \frac{\omega_1\mathbf{p}_1}{\omega_1} = \mathbf{p}_1,$$

поскольку $N_{1,m}(t_1) = 1$ в силу доказанной леммы 10.8, а $N_{i,m}(t_1) = 0$ при $i \geqslant 2$ по свойству локальности B-сплайнов. Следствие доказано. ■

Аналогично доказывается, что если $t_{n+1} = \cdots = t_{n+m}$, то B-кривая имеет правый конец в последней опорной точке $\mathbf{p}_n$.

Таким образом, для того, чтобы B-кривая начиналась в точке $\mathbf{p}_1$ и заканчивалась в точке $\mathbf{p}_n$ достаточно взять равными между собой первые m и последние m узлов в расширенном множестве узлов $t_1, \ldots, t_{n+m}$.

Вернемся к описанию алгоритма вычисления радиус-вектора точки B-кривой. На предыдущем шаге были определены значения ненормированных сплайнов $M_{j,m}(t)$ в заданной точке t. Затем производятся следующие шаги.

4) Вычисляем нормированные сплайны $N_{j,m}(t)$ для каждого $1 \leqslant j \leqslant n$ по формуле: $N_{j,m}(t) = (t_{j+m} - t_j)M_{j,m}(t)$ (при этом $1 \leqslant j, j+m \leqslant n+m$).

5) Окончательно вычисляем $\mathbf{r}(t)$ по формуле

$$\mathbf{r}(t) = \frac{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i\mathbf{p}_i}{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i}.$$

10.7 Алгоритм вычисления производных B-кривой при условии, что $t_1 = \cdots = t_m$, $t_{n+1} = \cdots = t_{n+m}$

Пусть даны опорные точки, веса и расширенное множество узлов: $\mathbf{p}_1, \ldots, \mathbf{p}_n$; $\omega_1, \ldots, \omega_n$; $t_1, \ldots, t_{n+m}$, причем первые m и последние m узлов совпадают (а, следовательно, соответствующая B-кривая проходит через крайние опорные точки $\mathbf{p}_1$ и $\mathbf{p}_n$): $t_1 = \cdots = t_m$; $t_{n+1} = \cdots = t_{n+m}$. Радиус-вектор соответствующей B-кривой задается формулой:

$$\mathbf{r}(t) = \frac{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i\mathbf{p}_i}{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i} = \frac{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i\mathbf{p}_i}{\omega(t)},$$

где $\omega(t) = \sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i$. Перепишем последнюю формулу в виде

$$\mathbf{r}(t) = \frac{\omega\mathbf{r}}{\omega} = \frac{\omega(t)\mathbf{r}(t)}{\omega(t)}.$$

Дифференцируя обе части этого равенства k раз по t, получим определенные выражения для производных $\dfrac{d^k}{d\,t^k}\mathbf{r}(t)$ через производные $\dfrac{d^s}{d\,t^s}(\omega\mathbf{r})$ и $\dfrac{d^s}{d\,t^s}\omega$ при $1 \leqslant s \leqslant k$, значения которых, в свою очередь, даются в нижеследующих двух леммах (лемма 10.9 и лемма 10.10).

Ниже мы будем считать, что $2 \leqslant m < n$, и пользоваться сделанным выше предположением о том, что $t_1 = \cdots = t_m$, $t_{n+1} = \cdots = t_{n+m}$.

Лемма 10.9. *В сделанных предположениях для любого* $1 \leqslant k \leqslant m-1$ *и любого* $t \in \mathbb{R}$ *выполнены равенства*:

$$\frac{d^k}{d\,t^k}\big(\omega(t)\mathbf{r}(t)\big) = \frac{d^k}{d\,t^k}\Big(\sum_{i=1}^{n} N_{i,m}(t)\omega_i\mathbf{p}_i\Big) = \sum_{i=k+1}^{n} N_{i,m-k}(t)\mathbf{p}_i^{(k)},$$

где $\mathbf{p}_i^{(0)} = \omega_i\mathbf{p}_i$ $(1 \leqslant i \leqslant n)$ *и*

$$\mathbf{p}_i^{(k)} = (m-k)\frac{\mathbf{p}_i^{(k-1)} - \mathbf{p}_{i-1}^{(k-1)}}{t_{i+m-k} - t_i} \qquad (1+k \leqslant i \leqslant n).$$

Доказательство. Для любого $t \in \mathbb{R}$ имеем

$$\begin{aligned}
\frac{d\,\sigma_{m-1}}{d\,t} &= -(m-1)(z-t)_+^{m-2} = -(m-1)\sigma_{m-2},\\
\frac{d\,N_{i,m}(t)}{d\,t} &= \frac{d}{d\,t}\big(\sigma_{m-1}[t_{i+1},\ldots,t_{i+m}] - \sigma_{m-1}[t_i,\ldots,t_{i+m-1}]\big) =\\
&= -\frac{m-1}{t_{i+m}-t_{i+1}}N_{i+1,m-1} + \frac{m-1}{t_{i+m-1}-t_i}N_{i,m-1}.
\end{aligned}$$

Пусть $t \in [t_{i_0}, t_{i_0+1})$. Поскольку нас интересуют лишь значения $t \in [t_1, t_{n+m})$, для которых рассматриваемые разделенные разности могут быть отличны от нуля, а по предположению выполнены равенства $t_1 = \cdots = t_m$, $t_{n+1} = \cdots = t_{n+m}$, то достаточно рассмотреть случай $m \leqslant i_0 \leqslant n+1$.

Имеем, учитывая, что для $t \in [t_{i_0}, t_{i_0+1})$ $N_{i,m}(t) = 0$ при $i < i_0 - m + 1$ и

при $i > i_0$:

$$\frac{d}{dt}(\omega\mathbf{r}) = \frac{d}{dt}\Big(\sum_{i=1}^{n} N_{i,m}(t)\omega_i\mathbf{p}_i\Big) = \frac{d}{dt}\Big(\sum_{i=i_0-m+1}^{i_0} N_{i,m}(t)\omega_i\mathbf{p}_i\Big) =$$

$$= \sum_{i=i_0-m+1}^{i_0} \frac{d\,N_{i,m}(t)}{dt}\omega_i\mathbf{p}_i = (m-1)\sum_{i=i_0-m+1}^{i_0} N_{i,m-1}\frac{\omega_i\mathbf{p}_i}{t_{i+m-1}-t_i} -$$

$$-(m-1)\sum_{i=i_0-m+1}^{i_0} N_{i+1,m-1}\frac{\omega_i\mathbf{p}_i}{t_{i+m}-t_{i+1}} = (m-1)N_{i_0-m+1,m-1}\frac{\omega_{i_0-m+1}\mathbf{p}_{i_0-m+1}}{t_{i_0}-t_{i_0-m+1}} +$$

$$+(m-1)\sum_{i=i_0-m+2}^{i_0} N_{i,m-1}\frac{\omega_i\mathbf{p}_i}{t_{i+m-1}-t_i} -$$

$$-(m-1)\sum_{i=i_0-m+2}^{i_0} N_{i,m-1}\frac{\omega_{i-1}\mathbf{p}_{i-1}}{t_{i+m-1}-t_i} - (m-1)N_{i_0+1,m-1}\frac{\omega_{i_0}\mathbf{p}_{i_0}}{t_{i_0+m}-t_{i_0+1}}.$$

Но $N_{i_0-m+1,m-1}(t) = 0$ и $N_{i_0+1,m-1}(t) = 0$ при $t \in [t_{i_0}, t_{i_0+1})$. Следовательно,

$$\frac{d\,\omega\mathbf{r}}{dt} = (m-1)\sum_{i=i_0-m+2}^{i_0} N_{i,m-1}\frac{\omega_i\mathbf{p}_i - \omega_{i-1}\mathbf{p}_{i-1}}{t_{i+m-1}-t_i} = \sum_{i=1+1}^{n} N_{i,m-1}(t)\mathbf{p}_i^{(1)}.$$

Это доказывает лемму при $k = 1$.

Для произвольного k лемма доказывается по индукции. Предположим, что утверждение леммы верно при некотором k и докажем его при $(k+1)$. Имеем при $t \in [t_{i_0}, t_{i_0+1})$, $m \leqslant i_0 \leqslant n+1$:

$$\frac{d^{k+1}}{dt^{k+1}}(\omega\mathbf{r}) = \frac{d}{dt}\Big(\frac{d^k}{dt^k}(\omega\mathbf{r})\Big) =$$

$$= \frac{d}{dt}\Big(\sum_{i=k+1}^{n} N_{i,m-k}(t)\mathbf{p}_i^{(k)}\Big) = \sum_{i=k+1}^{n} \frac{d\,N_{i,m-k}(t)}{dt}\mathbf{p}_i^{(k)} =$$

$$= \sum_{i=k+1}^{n} \mathbf{p}_i^{(k)}\Big(-\frac{m-k-1}{t_{i+m-k}-t_{i+1}}N_{i+1,m-k-1} + \frac{m-k-1}{t_{i+m-k-1}-t_i}N_{i,m-k-1}\Big) =$$

$$= -\sum_{i=i_0-m+k+2}^{i_0} \frac{m-k-1}{t_{i+m-k-1}-t_i}N_{i,m-k-1}\mathbf{p}_{i-1}^{(k)} +$$

$$+\sum_{i=i_0-m+k+2}^{i_0} \frac{m-k-1}{t_{i+m-k-1}-t_i}N_{i,m-k-1}\mathbf{p}_i^{(k)}.$$

В первой сумме мы заменили в индексе суммирования $(i+1)$ на i. Использовано также, что $N_{j,s}(t) = 0$ при $j < i_0 - s + 1$ и при $j > i_0$. Из последнего

равенства вытекает, что

$$\frac{d^{k+1}}{d\,t^{k+1}}(\omega\mathbf{r}) = \sum_{i=i_0-m+k+2}^{i_0} \frac{(m-k-1)\mathbf{p}_i^{(k)} - (m-k-1)\mathbf{p}_{i-1}^{(k)}}{t_{i+m-k-1} - t_i} N_{i,m-k-1} =$$

$$= \sum_{i=k+2}^{n} \mathbf{p}_i^{(k+1)} N_{i,m-(k+1)}.$$

Лемма доказана. ■

Лемма 10.10. *В сделанных выше предположениях для любого* $1 \leqslant k \leqslant m-1$ *и любого* $t \in \mathbb{R}$ *выполнены равенства*:

$$\frac{d^k}{d\,t^k}\big(\omega(t)\big) = \frac{d^k}{d\,t^k}\Big(\sum_{i=1}^{n} N_{i,m}(t)\omega_i\Big) = \sum_{i=k+1}^{n} N_{i,m-k}(t)\omega_i^{(k)},$$

где $\omega_i^{(0)} = \omega_i$ $(1 \leqslant i \leqslant n)$ *и*

$$\omega_i^{(k)} = (m-k)\frac{\omega_i^{(k-1)} - \omega_{i-1}^{(k-1)}}{t_{i+m-k} - t_i} \qquad (1+k \leqslant i \leqslant n).$$

Доказательство полностью аналогично доказательству леммы 10.9 с заменой $\omega_i\mathbf{p}_i$ на ω_i. ■

С помощью леммы 10.9 и леммы 10.10 могут быть вычислены производные любого порядка радиус-вектора B-кривой $\mathbf{r}(t)$ такой, что в ее расширенном множестве узлов первые m узлов и последние m узлов совпадают (а, следовательно, кривая проходит через свои опорные точки).

10.8 Алгоритм Де Бура вычисления радиус-вектора B-кривой

Теорема 10.3. *Радиус-вектор* $\mathbf{r}(t)$ *B-кривой*:

$$\mathbf{r}(t) = \frac{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i\mathbf{p}_i}{\sum\limits_{i=1}^{n} N_{i,m}(t)\omega_i} = \frac{\omega\mathbf{r}}{\omega} \tag{10.21}$$

может быть вычислен с помощью следующего алгоритма:

1) *Для заданного* $t \in [t_{i_0}, t_{i_0+1})$ $(m \leqslant i_0 \leqslant n+1)$ *вычисляем величины* $\mathbf{r}_i^{(k)}$, $\omega_i^{(k)}$ *при* $k = 1, \ldots, m-1$, $i = i_0 - m + k + 1, \ldots, i_0$ *по рекуррентным*

формулам:

$$\mathbf{r}_i^{(k)} = \frac{t_{i+m-k} - t}{t_{i+m-k} - t_i}\mathbf{r}_{i-1}^{(k-1)} + \frac{t - t_i}{t_{i+m-k} - t_i}\mathbf{r}_i^{(k-1)},$$
$$\omega_i^{(k)} = \frac{t_{i+m-k} - t}{t_{i+m-k} - t_i}\omega_{i-1}^{(k-1)} + \frac{t - t_i}{t_{i+m-k} - t_i}\omega_i^{(k-1)},$$
$$\mathbf{r}_i^{(0)} = \omega_i\mathbf{p}_i,\ \omega_i^{(0)} = \omega_i,\ i = i_0 - m + 1, \ldots, i_0.$$

2) Вычисляем $\mathbf{r}(t)$ *по формуле*

$$\mathbf{r}(t) = \frac{\mathbf{r}_{i_0}^{(m-1)}}{\omega_{i_0}^{(m-1)}}.$$

Данный алгоритм иллюстрируется следующей диаграммой ($t \in [t_{i_0}, t_{i_0+1})$):

$$\begin{array}{ccccc}
\mathbf{r}_{i_0-m+1}^{(0)} & \mathbf{r}_{i_0-m+2}^{(0)} & \cdots & \mathbf{r}_{i_0-1}^{(0)} & \mathbf{r}_{i_0}^{(0)} \\
 & \mathbf{r}_{i_0-m+2}^{(1)} & \cdots & \mathbf{r}_{i_0-1}^{(1)} & \mathbf{r}_{i_0}^{(1)} \\
 & & \ddots & \vdots & \vdots \\
 & & & \mathbf{r}_{i_0-1}^{(m-2)} & \mathbf{r}_{i_0}^{(m-2)} \\
 & & & & \mathbf{r}_{i_0}^{(m-1)}
\end{array}$$

Доказательство. Пусть $t \in [t_{i_0}, t_{i_0+1})$, $m \leqslant i_0 \leqslant n+1$.

Согласно формулам Кокса-Де Бура имеем

$$N_{a,m}(t) = \frac{t_{a+m} - t}{t_{a+m} - t_{a+1}}N_{a+1,m-1}(t) + \frac{t - t_a}{t_{a+m-1} - t_a}N_{a,m-1}(t).$$

Следовательно, числитель в правой части (10.21) может быть записан в виде:

$$\omega\mathbf{r} = \sum_{i=1}^{n} N_{i,m}(t)\omega_i\mathbf{p}_i = \sum_{i=i_0-m+1}^{i_0} N_{i,m}(t)\omega_i\mathbf{p}_i =$$
$$= \sum_{i=i_0-m+1}^{i_0}\Big(\frac{t_{i+m} - t}{t_{i+m} - t_{i+1}}N_{i+1,m-1}(t) + \frac{t - t_i}{t_{i+m-1} - t_i}N_{i,m-1}(t)\Big)\omega_i\mathbf{p}_i =$$
$$= \sum_{i=i_0-m+2}^{i_0}\frac{t_{i+m-1} - t}{t_{i+m-1} - t_i}N_{i,m-1}(t)\omega_{i-1}\mathbf{p}_{i-1} + \sum_{i=i_0-m+2}^{i_0}\frac{t - t_i}{t_{i+m-1} - t_i}N_{i,m-1}(t)\omega_i\mathbf{p}_i.$$

Остальные слагаемые, опущенные здесь, равны нулю в силу равенства $N_{s,m}(t) = 0$ при $s < i_0 - m + 1$ и $s > i_0$. Следовательно,

$$\omega\mathbf{r} = \sum_{i=i_0-m+2}^{i_0} N_{i,m-1}(t)\Big(\frac{t_{i+m-1} - t}{t_{i+m-1} - t_i}\omega_{i-1}\mathbf{p}_{i-1} + \frac{t - t_i}{t_{i+m-1} - t_i}\omega_i\mathbf{p}_i\Big) =$$
$$= \sum_{i=i_0-m+2}^{i_0} N_{i,m-1}(t)\mathbf{r}_i^{(1)}.$$

Продолжая аналогичные вычисления, получим цепочку равенств:

$$\omega\mathbf{r} = \sum_{i=i_0-m+1}^{i_0} N_{i,m}(t)\mathbf{r}_i^{(0)} = \sum_{i=i_0-m+2}^{i_0} N_{i,m-1}(t)\mathbf{r}_i^{(1)} = \ldots$$
$$\ldots = \sum_{i=i_0-1}^{i_0} N_{i,2}(t)\mathbf{r}_i^{(m-2)} = \sum_{i=i_0}^{i_0} N_{i,1}(t)\mathbf{r}_i^{(m-1)} = \mathbf{r}_i^{(m-1)}.$$

В последнем равенстве использовано, что $N_{i_0,1}(t) = 1$ при $t \in [t_{i_0}, t_{i_0+1})$. Аналогично, для знаменателя правой части (10.21) получаем при $t \in [t_{i_0}, t_{i_0+1})$:

$$\omega = \sum_{i=1}^{n} N_{i,m}(t)\omega_i = \sum_{i=i_0-m+1}^{i_0} N_{i,m}(t)\omega_i^{(0)} = \sum_{i=i_0-m+2}^{i_0} N_{i,m-1}(t)\omega_i^{(1)} = \ldots$$
$$\ldots = \sum_{i=i_0-1}^{i_0} N_{i,2}(t)\omega_i^{(m-2)} = N_{i_0,1}(t)\omega_{i_0}^{(m-1)} = \omega_{i_0}^{(m-1)}.$$

Теорема доказана. ■

Пример 8 (B-представление ломаной с вершинами в опорных точках). Пусть даны n вершин $\mathbf{p}_1, \ldots, \mathbf{p}_n$ с весами равными единице: $\omega_1 = \cdots = \omega_n = 1$. Рассмотрим расширенное множество узлов вида $t_1 = t_2 = 0$, $t_3 = 1$, $t_4 = 2, \ldots, t_i = i-2, \ldots, t_n = n-2$, $t_{n+1} = t_{n+2} = n-1$. Тогда соответствующая B-кривая второго порядка будет иметь вид:

$$\mathbf{r}(t) = \frac{\sum\limits_{i=1}^{n} N_{i,2}(t)\omega_i\mathbf{p}_i}{\sum\limits_{i=1}^{n} N_{i,2}(t)\omega_i} = \sum_{i=1}^{n} N_{i,2}\mathbf{p}_i.$$

Имеем

$$N_{i,2}(t) = \begin{cases} \dfrac{t-t_i}{t_{i+1}-t_i}, & \text{если} \quad t \in [t_i, t_{i+1}), \\[2ex] \dfrac{t_{i+2}-t}{t_{i+2}-t_{i+1}}, & \text{если} \quad t \in [t_{i+1}, t_{i+2}), \\[2ex] 0, & \text{иначе.} \end{cases}$$

Отсюда следует, что при $2 \leqslant j \leqslant n+1$:

$$N_{i,2}(t_j) = \begin{cases} 1, & \text{если} \quad i = j-1, \\ 0, & \text{если} \quad i \neq j-1. \end{cases}$$

Поэтому $\mathbf{r}(t_j) = N_{j-1,2}(t_j)\mathbf{p}_{j-1} = \mathbf{p}_{j-1}$. Кроме того, каждая из функций $N_{i,2}(t)$ линейна на каждом из интервалов $[t_i, t_{i+1}]$. Следовательно, $\mathbf{r}(t)$ представляет собой ломаную с вершинами в опорных точках $\mathbf{p}_i$.

Пример 9 (B-представление кривой Безье). Пусть дана $(n+1)$ опорная точка $\mathbf{p}_0, \ldots, \mathbf{p}_n$ с весами $\omega_0, \ldots, \omega_n$. Возьмем расширенное множество узлов вида: $t_{-n-1}, \ldots, t_n$, где $t_{-1-n} = t_{-n} = \cdots = t_{-1} = 0$, $t_0 = t_1 = \cdots = t_n = 1$. Построим B-кривую на основе B-сплайнов типа $N_{-i}^m(t)$:

$$\mathbf{r}(t) = \frac{\sum\limits_{i=0}^{n} N_{-i}^n(t)\omega_i \mathbf{p}_i}{\sum\limits_{i=0}^{n} N_{-i}^n(t)\omega_i}, \qquad 0 \leqslant t \leqslant 1.$$

Покажем, что эта кривая совпадает с рациональной кривой Безье, построенной по тем же точкам $\mathbf{p}_0, \ldots, \mathbf{p}_n$ с теми же весами $\omega_0, \ldots, \omega_n$, т.е. с кривой:

$$\mathbf{r}_B(t) = \frac{\sum\limits_{i=0}^{n} B_i^n(t)\omega_i \mathbf{p}_i}{\sum\limits_{i=0}^{n} B_i^n(t)\omega_i}, \qquad 0 \leqslant t \leqslant 1,$$

где $B_i^n(t) = C_n^i t^i (1-t)^{n-i}$ — полиномы Бернштейна. Другими словами, покажем, что при указанном выборе расширенного множества узлов $B_i^n(t) = N_{-i}^n(t)$ для любых $t \in [0,1]$, $0 \leqslant i \leqslant n$. Имеем

$$B_0^0(t) = 1, \qquad N_{-0}^0(t) = 1.$$

Далее, для любого $1 \leqslant i \leqslant n-1$ имеем

$$B_i^n(t) = tB_{i-1}^{n-1}(t) + (1-t)B_i^{n-1}(t),$$
$$N_{-i}^n = \frac{t_i - t}{t_i - t_{i-n}} N_{-i}^{n-1}(t) + \frac{t - t_{i-n-1}}{t_i - t_{i-n-1}} N_{-(i-1)}^{n-1}(t) =$$
$$= (1-t)N_{-i}^{n-1}(t) + tN_{-(i-1)}^{n-1}(t).$$

Таким образом, рекуррентные формулы, определяющие B_i^n и N_{-i}^n, в данном случае совпадают и начальные значения для них тоже совпадают. Следовательно, $N_{-i}^n(t) = B_i^n(t)$ для любого $t \in [0,1]$.

С помощью B-кривых могут быть представлены кривые второго порядка, кубические сплайны, а также интерполяционные кривые, проходящие через заданные опорные точки.

10.9 Интерполяция с помощью B-кривых

Пусть требуется построить кривую класса C^{m-2}, проходящую через точки $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Можно воспользоваться таким методом.

1) Строим B-кривую порядка m по опорным точкам $\mathbf{p}_1 = \mathbf{a}_1, \ldots, \mathbf{p}_n = \mathbf{a}_n$ с весами $\omega_i = 1$ по некоторой (произвольной) последовательности неубывающих узлов $\{t_1, \ldots, t_{n+m}\}$.

2) Определяем ближайшие к $\mathbf{a}_i$ точки на построенной кривой и определяем значения параметров x_i, соответствующих полученным точкам.

3) Определяем новые опорные точки $\widetilde{\mathbf{p}}_1, \ldots, \widetilde{\mathbf{p}}_n$ из условия:

$$\mathbf{r}(x_i) = \mathbf{a}_i, \qquad i = 1, \ldots, n.$$

Эти условия записываются в виде системы линейных уравнений на $\widetilde{\mathbf{p}}_i$:

$$\begin{cases} N_{1,m}(x_1)\widetilde{\mathbf{p}}_1 + \cdots + N_{n,m}(x_1)\widetilde{\mathbf{p}}_n = \mathbf{a}_1, \\ N_{1,m}(x_2)\widetilde{\mathbf{p}}_1 + \cdots + N_{n,m}(x_2)\widetilde{\mathbf{p}}_n = \mathbf{a}_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ N_{1,m}(x_n)\widetilde{\mathbf{p}}_1 + \cdots + N_{n,m}(x_n)\widetilde{\mathbf{p}}_n = \mathbf{a}_n. \end{cases} \tag{10.22}$$

Данный способ не очень хорош, так как у кривой могут появиться необоснованные изгибы на концах. Для интерполяции лучше применять составные кубические сплайны или сплайны Эрмита.

10.10 Представление кубического сплайна в виде B-кривой

Представим в B-виде кубический сплайн:

$$\mathbf{a}(x) = (1-\omega)\mathbf{a}_i + \omega\mathbf{a}_{i+1} + \big((-2\omega + 3\omega^2 - \omega^3)\mathbf{s}_i + (-\omega + \omega^3)\mathbf{s}_{i+1}\big)\frac{(x_{i+1} - x_i)^2}{6},$$

где $\omega = \dfrac{x - x_i}{x_{i+1} - x_i}$ — местный параметр при $x \in [x_i, x_{i+1}]$;
$\mathbf{a}_1, \ldots, \mathbf{a}_n$ — интерполируемые точки;
$\mathbf{s}_1, \ldots, \mathbf{s}_n$ — вторые производные в точках x_i: $\mathbf{r}''(x_i) = \mathbf{s}_i$ $(i = 1, \ldots, n)$, определяемые из системы уравнений

$$\mathbf{s}_{i-1}(x_i - x_{i-1}) + 2\mathbf{s}_i(x_{i+1} - x_{i-1}) + \mathbf{s}_{i+1}(x_{i+1} - x_i) = 6\frac{\mathbf{a}_{i+1} - \mathbf{a}_i}{x_{i+1} - x_i} - 6\frac{\mathbf{a}_i - \mathbf{a}_{i-1}}{x_i - x_{i-1}}$$

и граничных условий на $\mathbf{s}_1$, $\mathbf{s}_n$ (например, $\mathbf{s}_1 = \mathbf{s}_n = 0$);
$x_1, \ldots, x_n$ — значения параметра x, при которых кривая должна пройти через точки $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

Будем искать B-кривую 4-го порядка:

$$\mathbf{r}(t) = \sum_{i=1}^{n+2} N_{i,4}(t)\mathbf{p}_i,$$

определенную на расширенном множестве узлов:

$$t_1 = t_2 = t_3 = t_4 = x_1, \qquad t_5 = x_2, \qquad t_6 = x_3,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$t_{n+3} = t_{n+4} = t_{n+5} = t_{n+6} = x_n,$$

совпадающую с заданным кубическим сплайном. Для этого определим опорные точки $\mathbf{p}_1, \ldots, \mathbf{p}_{n+2}$ из условий:

$$\begin{aligned} &\mathbf{r}(x_1) = \mathbf{a}_1, \quad \left.\frac{d\,\mathbf{r}}{d\,x}\right|_{x=x_1} = \mathbf{a}'(x_1), \\ &\mathbf{r}(x_j) = \mathbf{a}_j, \quad j = 2, \ldots, n-1, \\ &\mathbf{r}(x_n) = \mathbf{a}_n, \quad \left.\frac{d\,\mathbf{r}}{d\,x}\right|_{x=x_n} = \mathbf{a}'(x_n). \end{aligned} \tag{10.23}$$

Положим все веса опорных точек B-кривой равными 1. Тогда такая B-кривая и кубический сплайн $\mathbf{a}(x)$ будут совпадать, так как оба они состоят из дуг кубических кривых, имеющих в одних и тех же точках $\mathbf{a}_1, \ldots, \mathbf{a}_n$ непрерывные первые и вторые производные (что, вместе с граничными условиями

$$\left.\frac{d\,\mathbf{r}}{d\,x}\right|_{x=x_1} = \mathbf{a}'(x_1), \qquad \left.\frac{d\,\mathbf{r}}{d\,x}\right|_{x=x_n} = \mathbf{a}'(x_n).$$

полностью определяет обе кривые).

Приведенные выше уравнения (10.23) для опорных точек $\mathbf{p}_1, \ldots, \mathbf{p}_{n+2}$ могут быть записаны в виде

$$\begin{cases} N_{1,4}(x_1)\mathbf{p}_1 + N_{2,4}(x_1)\mathbf{p}_2 + \cdots + N_{n+2,4}(x_1)\mathbf{p}_{n+2} = \mathbf{a}_1, \\ 3N_{2,3}(x_1)\dfrac{\mathbf{p}_2 - \mathbf{p}_1}{t_5 - t_2} + 3N_{3,3}(x_1)\dfrac{\mathbf{p}_3 - \mathbf{p}_2}{t_6 - t_3} + \ldots \\ \ldots + 3N_{n+2,3}(x_1)\dfrac{\mathbf{p}_{n+2} - \mathbf{p}_{n+1}}{t_{n+5} - t_{n+2}} = \mathbf{a}'(x_1), \\ N_{1,4}(x_2)\mathbf{p}_1 + N_{2,4}(x_2)\mathbf{p}_2 + \cdots + N_{n+2,4}(x_2)\mathbf{p}_{n+2} = \mathbf{a}_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ N_{1,4}(x_{n-1})\mathbf{p}_1 + N_{2,4}(x_{n-1})\mathbf{p}_2 + \cdots + N_{n+2,4}(x_{n-1})\mathbf{p}_{n+2} = \mathbf{a}_{n-1}, \\ 3N_{2,3}(x_n)\dfrac{\mathbf{p}_2 - \mathbf{p}_1}{t_5 - t_2} + 3N_{3,3}(x_n)\dfrac{\mathbf{p}_3 - \mathbf{p}_2}{t_6 - t_3} + \ldots \\ \ldots + 3N_{n+2,3}(x_n)\dfrac{\mathbf{p}_{n+2} - \mathbf{p}_{n+1}}{t_{n+5} - t_{n+2}} = \mathbf{a}'(x_n), \\ N_{1,4}(x_n)\mathbf{p}_1 + N_{2,4}(x_n)\mathbf{p}_2 + \cdots + N_{n+2,4}(x_n)\mathbf{p}_{n+2} = \mathbf{a}_n. \end{cases} \tag{10.24}$$

Приведем эту систему к более простому виду. Заметим, что из соотношений

$$M_{i,1}(t) = \begin{cases} \dfrac{1}{t_{i+1} - t_i} & \text{если} \quad t \in [t_i, t_{i+1}), \\ 0, & \text{иначе}, \end{cases}$$

$$M_{i,m}(t) = \frac{(t_{i+m} - t)M_{i+1,m-1}(t) + (t - t_i)M_{i,m-1}(t)}{t_{i+m} - t_i}$$

следует, что для каждого t_i, где $i = 4, 5, \ldots, n+3$, при $t = t_i$ отличны от

нуля только три B-сплайна четвертого порядка:

$$N_{i-3,4}(t_i) = \frac{t_{i+1}-t_i}{t_{i+1}-t_{i-2}} \cdot \frac{t_{i+1}-t_i}{t_{i+1}-t_{i-1}},$$
$$N_{i-2,4}(t_i) = \frac{t_i-t_{i-1}}{t_{i+2}-t_{i-1}} \cdot \frac{t_{i+2}-t_i}{t_{i+1}-t_{i-1}} + \frac{t_{i+1}-t_i}{t_{i+1}-t_{i-2}} \cdot \frac{t_i-t_{i-2}}{t_{i+1}-t_{i-1}},$$
$$N_{i-1,4}(t_i) = \frac{t_i-t_{i-1}}{t_{i+2}-t_{i-1}} \cdot \frac{t_i-t_{i-1}}{t_{i+1}-t_{i-1}}$$

и два B-сплайна третьего порядка:

$$N_{i-2,3}(t_i) = \frac{t_{i+1}-t_i}{t_{i+1}-t_{i-1}}, \qquad N_{i-1,3}(t_i) = \frac{t_i-t_{i-1}}{t_{i+1}-t_{i-1}}.$$

После подстановки последних формул в систему уравнений (10.24), получим

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ -b & b & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & a_{32} & a_{33} & a_{34} & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{43} & a_{44} & a_{45} & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & a_{n\,n-1} & a_{nn} & a_{n\,n+1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & -c & c \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \\ \mathbf{p}_4 \\ \vdots \\ \mathbf{p}_n \\ \mathbf{p}_{n+1} \\ \mathbf{p}_{n+2} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}'_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \vdots \\ \mathbf{a}_{n-1} \\ \mathbf{a}'_n \\ \mathbf{a}_n \end{pmatrix},$$

где

$$\mathbf{a}'_1 = \mathbf{a}'(x_1), \qquad \mathbf{a}'_n = \mathbf{a}'(x_n), \qquad b = \frac{3}{x_2-x_1}, \qquad c = \frac{3}{x_n-x_{n-1}},$$
$$a_{k\,k-1} = \frac{x_k-x_{k-1}}{x_k-x_{k-3}} \cdot \frac{x_k-x_{k-1}}{x_k-x_{k-2}},$$
$$a_{kk} = \frac{x_{k-1}-x_{k-2}}{x_{k+1}-x_{k-2}} \cdot \frac{x_{k+1}-x_{k-1}}{x_k-x_{k-2}} + \frac{x_k-x_{k-1}}{x_k-x_{k-3}} \cdot \frac{x_{k-1}-x_{k-3}}{x_k-x_{k-2}},$$
$$a_{k\,k+1} = \frac{x_{k-1}-x_{k-2}}{x_{k+1}-x_{k-2}} \cdot \frac{x_{k-1}-x_{k-2}}{x_k-x_{k-2}}.$$

Получается система уравнений с трех-диагональной матрицей системы.

В случае замкнутого кубического сплайна число опорных точек будет равно числу узлов $\mathbf{a}_i$: $\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$, второе и предпоследние уравнения, описывающие производные на концах, пропадут, а расширенное множество узлов будет иметь вид: $t_1 = x_{n-1}$, $t_2 = x_n$, $t_3 = x_1$, $t_i = x_{i-2}$, где $i = 4, 5, \dots, n$.

10.11 B-поверхности

Пусть задана матрица опорных точек $(\mathbf{p}_{ij})$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$, с весами (ω_{ij}). Фиксируем числа p и q — порядки B-сплайнов в двух координатных

Рис. 10.3:

направлениях u и v соответственно. Тогда B-поверхность порядка (p, q), построенная по матрице опорных точек $(\mathbf{p}_{ij})$ с весами (ω_{ij}) и по двум расширенным множествам узлов в направлениях u и v соответственно, определяется формулой

$$\mathbf{r}(u,v) = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} N_{i,q}(v)N_{j,p}(u)\omega_{ij}\mathbf{p}_{ij}}{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} N_{i,q}(v)N_{j,p}(u)\omega_{ij}}. \tag{10.25}$$

Если ввести новые обозначения для входящих в эту формулу произведений B-сплайнов:

$$N_{ij,qp}(v,u) = N_{i,q}(v)N_{j,p}(u),$$

то формула (10.25) перепишется в виде

$$\mathbf{r}(u,v) = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} N_{ij,qp}(v,u)\omega_{ij}\mathbf{p}_{ij}}{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} N_{ij,qp}(v,u)\omega_{ij}}. \tag{10.26}$$

Если расширенные множества узлов в u и v-направлениях имеют вид

$$\begin{gathered} u_1 = \dots = u_p, \quad u_{p+1}, \dots, u_n, \quad u_{n+1} = \dots = u_{n+k}, \\ v_1 = \dots = v_q, \quad v_{q+1}, \dots, v_n, \quad v_{n+1} = \dots = v_{n+q}, \end{gathered}$$

то область изменения параметров (u, v) на B-поверхности (10.26) будет следующей:

$$u_p \leqslant u \leqslant u_{n+1}, \quad v_q \leqslant v \leqslant v_{n+1}.$$

На рис. 10.3 показана B-поверхность порядка $(4, 4)$, построенная по матрице 7×7 опорных точек. Все вершины имеют вес равный единице и, кроме центральной, лежат в одной плоскости.

Глава 11

Другие способы представления поверхностей в компьютерной геометрии

11.1 Линейчатые поверхности

Линейчатая поверхность — геометрическое место точек отрезков прямых, соединяющих соответствующие друг другу точки двух заданных кривых. Пусть $\rho_1(t)$, $t_0 \leqslant t \leqslant t_1$, и $\rho_2(w)$, $w_0 \leqslant w \leqslant w_1$, — две заданные кривые. Тогда определяемая ими линейчатая поверхность имеет вид

$$\mathbf{r}(u,v) = (1-v)\rho_1(t) + v\rho_2(w),$$

где $t = t_0(1-u) + t_1 u$, $w = w_0(1-u) + w_1 u$, $0 \leqslant u, v \leqslant 1$, см. рис. 11.1.

Если обе кривые ρ_1 и ρ_2 замкнуты, то линейчатая поверхность замкнута по u. По параметру v линейчатая поверхность всегда незамкнута.

Если обе кривые ρ_1 и ρ_2 — отрезки прямых, то линейчатая поверхность называется *билинейной* и определяется только своими четырьмя угловыми точками $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{p}_4$:

$$\begin{gathered}\mathbf{r}(u,v) = (1-v)\big((1-u)\mathbf{p}_1 + u\mathbf{p}_2\big) + v\big((1-u)\mathbf{p}_3 + u\mathbf{p}_4\big) = \\ = (1-u)(1-v)\mathbf{p}_1 + u(1-v)\mathbf{p}_2 + (1-u)v\mathbf{p}_3 + uv\mathbf{p}_4, \qquad 0 \leqslant u, v \leqslant 1.\end{gathered}$$

11.2 Секториальные поверхности

Секториальная поверхность — частный случай линейчатой поверхности,

Рис. 11.1:

когда одна из кривых ρ_1 и ρ_2 вырождается в точку:

$$\mathbf{r}(u,v) = (1-v)\rho(u) + v\mathbf{p},$$

где $\rho(u)$ — кривая, $\mathbf{p}$ — точка, $u_1 \leqslant u \leqslant u_2$, $0 \leqslant v \leqslant 1$.

Секториальная билинейная поверхность называется *треугольной поверхностью.*

11.3 Поверхности Кунса

11.3.1 Линейные поверхности Кунса

Вычтем и добавим к уравнению линейчатой поверхности уравнение билинейной поверхности, построенной по четырем угловым точкам (см. выше). Уравнение линейчатой поверхности примет вид

$$\begin{gathered}\mathbf{r}(u,v) = (1-v)\rho_1(u)+v\rho_2(u) = (1-v)\rho_1(u)+v\rho_2(u)+(1-u)\big((1-v)\mathbf{p}_1+v\mathbf{p}_3\big)+\\ + u\big((1-v)\mathbf{p}_2 + v\mathbf{p}_4\big) - (1-u)(1-v)\mathbf{p}_1 - u(1-v)\mathbf{p}_2 - (1-u)v\mathbf{p}_3 - uv\mathbf{p}_4 =\\ = (1-v)\rho_1(u) + v\rho_2(u) + (1-u)\rho_3(v) + u\rho_4(v) - (1-u)(1-v)\mathbf{p}_1 -\\ - u(1-v)\mathbf{p}_2 - (1-u)v\mathbf{p}_3 - uv\mathbf{p}_4, \qquad 0 \leqslant u, v \leqslant 1, \qquad (11.1)\end{gathered}$$

где $\rho_1(u)$ и $\rho_2(u)$ — кривые, по которым строится данная линейчатая поверхность, причем считается, что на них обеих задан один и тот же параметр $0 \leqslant u \leqslant 1$, и

$$\rho_3(v) = (1-v)\mathbf{p}_1 + v\mathbf{p}_3, \qquad \rho_4(v) = (1-v)\mathbf{p}_2 + v\mathbf{p}_4 \qquad (11.2)$$

— отрезки прямых.

Рис. 11.2:

Подставим теперь в уравнение (11.1) вместо отрезков прямых (11.2) уравнения двух произвольных кривых $\rho_3(v)$ и $\rho_4(v)$, соединяющих соответственно пары точек $(\mathbf{p}_1, \mathbf{p}_3)$ и $(\mathbf{p}_2, \mathbf{p}_4)$. Тогда новая поверхность, задаваемая уравнением (11.1), будет определяться четырьмя кривыми ρ_1, ρ_2, ρ_3, ρ_4, образующими криволинейный четырехугольник, натянутый на угловые точки $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4$, см. рис. 11.2. При этом, все четыре кривые $\rho_1, \rho_2, \rho_3, \rho_4$ войдут в уравнение (11.1) симметрично. Получится *линейная поверхность Кунса.*

Определение 11.1. Пусть заданы четыре кривые, образующие криволинейный четырехугольник с углам в точках $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{p}_4$:

$$\begin{aligned}
&\gamma_a = \mathbf{r}(u,0), \quad 0 \leqslant u \leqslant 1, \qquad \mathbf{r}(0,0) = \mathbf{p}_1, \quad \mathbf{r}(1,0) = \mathbf{p}_2,\\
&\gamma_b = \mathbf{r}(0,v), \quad 0 \leqslant v \leqslant 1, \qquad \mathbf{r}(0,0) = \mathbf{p}_1, \quad \mathbf{r}(0,1) = \mathbf{p}_3,\\
&\gamma_c = \mathbf{r}(1,v), \quad 0 \leqslant v \leqslant 1, \qquad \mathbf{r}(1,0) = \mathbf{p}_2, \quad \mathbf{r}(1,1) = \mathbf{p}_4,\\
&\gamma_d = \mathbf{r}(u,1), \quad 0 \leqslant u \leqslant 1, \qquad \mathbf{r}(0,1) = \mathbf{p}_3, \quad \mathbf{r}(1,1) = \mathbf{p}_4.
\end{aligned}$$

Обозначим:

$$\alpha_0(t) = 1 - t, \qquad \alpha_1(t) = t.$$

Функции $\alpha_0(t)$, $\alpha_1(t)$ будем называть *функциями смещения линейной поверхности Кунса.* Сама линейная поверхность Кунса определяется следующей формулой:

$$\begin{aligned}
\mathbf{r}(u,v) = \alpha_0(v)\mathbf{r}(u,0) + \alpha_0(u)\mathbf{r}(0,v) + \alpha_1(v)\mathbf{r}(u,1) +\\
+ \alpha_1(u)\mathbf{r}(1,v) - \alpha_0(u)\alpha_0(v)\mathbf{r}(0,0) - \alpha_1(u)\alpha_0(v)\mathbf{r}(1,0) -\\
- \alpha_0(u)\alpha_1(v)\mathbf{r}(0,1) - \alpha_1(u)\alpha_1(v)\mathbf{r}(1,1). \qquad (11.3)
\end{aligned}$$

Замечание. Линейная поверхность Кунса не является линейчатой поверхностью (вообще говоря). Название “линейная” обусловлено тем, что в ее определении используются линейные функции смещения α_0 и α_1.

Замечание. Параметрическая область поверхности Кунса может иметь вид прямоугольника, не обязательно совпадающего с квадратом $[0,1]\times[0,1]$.

11.3.2 Матричный вид уравнения поверхности Кунса

Преобразуем формулу (11.3) к другому виду. Имеем в матричных обозначениях:

$$\mathbf{r}(u,v) = \begin{pmatrix} \alpha_0(u) & \alpha_1(u) \end{pmatrix} \begin{pmatrix} \mathbf{r}(0,v) \\ \mathbf{r}(1,v) \end{pmatrix} + \begin{pmatrix} \alpha_0(v) & \alpha_1(v) \end{pmatrix} \begin{pmatrix} \mathbf{r}(u,0) \\ \mathbf{r}(u,1) \end{pmatrix} -$$

$$- \begin{pmatrix} \alpha_0(u) & \alpha_1(u) \end{pmatrix} \begin{pmatrix} \mathbf{r}(0,0) & \mathbf{r}(0,1) \\ \mathbf{r}(1,0) & \mathbf{r}(1,1) \end{pmatrix} \begin{pmatrix} \alpha_0(v) \\ \alpha_1(v) \end{pmatrix} =$$

$$= \sum_{i=0}^{1} \alpha_i(u)\mathbf{r}(i,v) + \sum_{j=0}^{1} \alpha_j(v)\mathbf{r}(u,j) - \sum_{i=0}^{1}\sum_{j=0}^{1} \alpha_i(u)\alpha_j(v)\mathbf{r}(i,j).$$

Пример 10. Зададим с помощью поверхности Кунса область D плоскости, ограниченную эллипсом $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$ или $\begin{cases} x = a\cos\varphi \\ y = b\cos\varphi, \end{cases}$ где $0 \leqslant \varphi < 2\pi$, с полуосями $a > 0$, $b > 0$. Представим границу области D в виде четырех кривых — дуг эллипса, стыкующихся в точках $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{p}_4$. Соответствующая линейная поверхность Кунса имеет вид

$$\mathbf{r}(u,v) = a\Big(\cos u + \cos v - 1 + \frac{2u(1-\cos v - \sin v)}{\pi} + \frac{2v(1-\cos u - \sin u)}{\pi}\Big)\mathbf{e}_1 +$$
$$+b\Big(\sin u - \sin v - \frac{2u(1-\cos v - \sin v)}{\pi} + \frac{2v(1-\cos u - \sin u)}{\pi}\Big)\mathbf{e}_2,$$
$$0 \leqslant u, v \leqslant \frac{\pi}{2}.$$

Преимущество задания плоских областей в виде линейных поверхностей Кунса состоит в том, что у поверхности Кунса параметрическая область — всегда прямоугольник.

11.3.3 Обобщенные поверхности Кунса

При стыковке линейных поверхностей Кунса по граничным кривым производные в направлении, ортогональном границе, претерпевают разрыв. Чтобы гладко склеивать поверхности Кунса необходимо строить их таким образом, чтобы они имели на границе заданные (согласованные) производные до k-го порядка включительно. Производные в касательном направлении определяются уравнениями самих граничных кривых, а производные в трансверсальном к границе направлении необходимо задавать дополнительно.

Определение 11.2. *Обобщенной поверхностью Кунса n-го порядка* называется поверхность

$$\mathbf{r}(u,v) = \sum_{i=0}^{1}\sum_{k=0}^{n} \alpha_{k,i}(u)\mathbf{r}_u^{(k)}(i,v) + \sum_{j=0}^{1}\sum_{k=0}^{n} \alpha_{k,j}(v)\mathbf{r}_v^{(k)}(u,j) -$$
$$- \sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{n}\sum_{l=0}^{n} \alpha_{k,i}(u)\alpha_{l,j}(v)\mathbf{r}_{uv}^{(k,l)}(i,j), \qquad 0 \leqslant u,v \leqslant 1, \quad (11.4)$$

задаваемая четырьмя граничными кривыми $\mathbf{r}(u,0)$, $\mathbf{r}(0,v)$, $\mathbf{r}(1,v)$, $\mathbf{r}(u,1)$, стыкующимися в точках $\mathbf{p}_1 = \mathbf{r}(0,0)$, $\mathbf{p}_2 = \mathbf{r}(1,0)$, $\mathbf{p}_3 = \mathbf{r}(0,1)$ и $\mathbf{p}_4 = \mathbf{r}(1,1)$, а также чистыми частными производными в трансверсальном направлении на граничных кривых до n-го порядка включительно и смешанными частными производными в угловых точках $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{p}_4$ до $2n$-го порядка включительно:

$$\mathbf{r}_u^{(k)}(i,v) = \frac{\partial^k \mathbf{r}(u,v)}{\partial\, u^k}\Big|_{u=i}, \qquad i = 0,1,$$
$$\mathbf{r}_v^{(k)}(u,j) = \frac{\partial^k \mathbf{r}(u,v)}{\partial\, v^k}\Big|_{v=j}, \qquad j = 0,1, \qquad (11.5)$$
$$\mathbf{r}_{uv}^{(k,l)}(i,j) = \frac{\partial^{kl} \mathbf{r}(u,v)}{\partial\, u^k \partial\, v^l}\Big|_{u=i,v=j}, \qquad i,j = 0,1,$$

а также — *обобщенными функциями смещения* $\alpha_{k,i}(t)$, где $0 \leqslant k \leqslant n$, $i = 0$, 1, в качестве которых могут быть взяты произвольные гладкие функции, удовлетворяющие соотношениям:

$$\frac{d^m \alpha_{k,i}(t)}{d\, t^m}\Big|_{t=j} = \delta_i^j \delta_k^m = \begin{cases} 1, & \text{если} \quad i=j, k=m, \\ 0, & \text{иначе.} \end{cases} \qquad (11.6)$$

Здесь $0 \leqslant k, m \leqslant n$.

Задача. Проверить, что поверхность (11.4), определенная с помощью (11.5) и (11.6), имеет заданные в (11.5) производные на границе.

Замечание. В качестве обобщенных функций смещения всегда можно выбрать многочлены.

Пример 11 (Кубическая поверхность Кунса). Определим функции смещения формулами:

$$\alpha_0(t) = 1 - 3\,t^2 + 2\,t^3, \qquad \alpha_1(t) = 3\,t^2 - 2\,t^3,$$
$$\beta_0(t) = t - 2\,t^2 + t^3, \qquad \beta_1(t) = -\,t^2 + t^3$$

(здесь для простоты использованы обозначения: $\alpha_{0,0}(t) = \alpha_0(t)$, $\alpha_{0,1}(t) = \alpha_1(t)$, $\alpha_{1,0}(t) = \beta_0(t)$, $\alpha_{1,1}(t) = \beta_1(t)$). Пусть заданы четыре кривые $\mathbf{r}(u,i)$,

$\mathbf{r}(i,v)$, где $i = 0,\ 1$, стыкующиеся в точках $\mathbf{r}(0,0)$, $\mathbf{r}(0,1)$, $\mathbf{r}(1,0)$, $\mathbf{r}(1,1)$. Пусть заданы первые производные на границе в трансверсальном направлении: $\mathbf{r}_u(i,v)$, $\mathbf{r}_v(u,i)$, $i = 0,\ 1$, и вторые смешанные производные в угловых точках: $\mathbf{r}_{uv}(i,j)$, $i,j = 0,\ 1$.

Кубическая поверхность Кунса определяется формулой (в матричных обозначениях):

$$\mathbf{r}(u,v) = \begin{pmatrix} \alpha_0(u) & \alpha_1(u) & \beta_0(u) & \beta_1(u) \end{pmatrix} \begin{pmatrix} \mathbf{r}(0,v) \\ \mathbf{r}(1,v) \\ \mathbf{r}_u(0,v) \\ \mathbf{r}_u(1,v) \end{pmatrix} +$$

$$+ \begin{pmatrix} \alpha_0(v) & \alpha_1(v) & \beta_0(v) & \beta_1(v) \end{pmatrix} \begin{pmatrix} \mathbf{r}(u,0) \\ \mathbf{r}(u,1) \\ \mathbf{r}_v(u,0) \\ \mathbf{r}_v(u,1) \end{pmatrix} -$$

$$- \begin{pmatrix} \alpha_0(u) \\ \alpha_1(u) \\ \beta_0(u) \\ \beta_1(u) \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} \mathbf{r}(0,0) & \mathbf{r}(0,1) & \mathbf{r}_v(0,0) & \mathbf{r}_v(0,1) \\ \mathbf{r}(1,0) & \mathbf{r}(1,1) & \mathbf{r}_v(1,0) & \mathbf{r}_v(1,1) \\ \mathbf{r}_u(0,0) & \mathbf{r}_u(0,1) & \mathbf{r}_{uv}(0,0) & \mathbf{r}_{uv}(0,1) \\ \mathbf{r}_u(1,0) & \mathbf{r}_u(1,1) & \mathbf{r}_{uv}(1,0) & \mathbf{r}_{uv}(1,1) \end{pmatrix} \begin{pmatrix} \alpha_0(v) \\ \alpha_1(v) \\ \beta_0(v) \\ \beta_1(v) \end{pmatrix},$$

$0 \leqslant u, v \leqslant 1$.

11.4 Поверхности, построенные по остову из кривых

В проектировании поверхностей иногда применяется метод построения поверхности по дискретной совокупности ее сечений. Для затягивания остова, образованного сечениями, гладкой поверхностью, применяются методы, аналогичные построению интерполяционных кривых. Поэтому получаемые таким способом поверхности имеют такие же названия, как и соответствующие интерполяционные сплайны.

11.4.1 Поверхности Эрмита

Напомним, что сплайном Эрмита называется кусочно-кубическая кривая, проходящая через заданные точки $\mathbf{p}_0, \ldots, \mathbf{p}_n \in \mathbb{R}^N$ и имеющая в этих точках заданные производные $\mathbf{q}_0, \ldots, \mathbf{q}_n \in \mathbb{R}^N$. Если производные $\mathbf{q}_i$ не заданы, то в точках $\mathbf{p}_1, \ldots, \mathbf{p}_{n-1}$ их можно определить из некоторых “естественных” соображений. Одним из примеров такого рода является псевдоупругий сплайн Эрмита (см. раздел 7.1.3). Еще один способ — положить по определению:

$$\mathbf{q}_i = \frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{t_{i+1} - t_i}$$

(есть и другие способы).

Задание производных на концах $\mathbf{p}_0$ и $\mathbf{p}_n$ иногда заменяют на условие обращения в ноль третей производной на концах кривой. Это условие равносильно следующим формулам:

$$\begin{cases} \mathbf{q}_0 = 2(\mathbf{p}_1 - \mathbf{p}_0) - \mathbf{q}_1, \\ \mathbf{q}_n = 2(\mathbf{p}_n - \mathbf{p}_{n-1}) - \mathbf{q}_{n-1}. \end{cases}$$

Напомним, что вторые производные в опорных точках сплайна Эрмита (где происходит стык кубических кривых) претерпевают разрыв, т.е. сплайны Эрмита принадлежат классу C^1. Аналогично сплайнам Эрмита строятся *поверхности Эрмита.*

Пусть имеется семейство кривых:

$$\begin{cases} \mathbf{r} = \mathbf{r}_0(u) = \mathbf{r}(u, 0), \\ \dots\dots\dots\dots \\ \mathbf{r} = \mathbf{r}_i(u) = \mathbf{r}(u, i), \\ \dots\dots\dots\dots \\ \mathbf{r} = \mathbf{r}_n(u) = \mathbf{r}(u, n). \end{cases} \tag{11.7}$$

Без ограничения общности будем считать, что параметр u для всех кривых пробегает один и тот же отрезок: $u_0 \leqslant u \leqslant u_1$. Тогда поверхность Эрмита, затягивающая остов кривых $\mathbf{r}(u, i)$, $0 \leqslant i \leqslant n$, определяется формулой аналогичной формуле для сплайна Эрмита:

$$\mathbf{r}(u, v) = \mathbf{r}(u, j)\alpha_0(w) + \mathbf{r}(u, j+1)\alpha_1(w) + \\ + \mathbf{r}_v(u, j)\beta_0(w) + \mathbf{r}_v(u, j+1)\beta_1(w),$$

если $j \leqslant v \leqslant j+1$, см. рис. 11.3.. Здесь $w = v - j$ — местный параметр на данном участке поверхности ($0 \leqslant w \leqslant 1$), $\mathbf{r}_v(u, j)$ и $\mathbf{r}_v(u, j+1)$ — производные в трансверсальном направлении на заданном остове кривых (11.7), $\alpha_0(w) = 1 - 3w^2 + 2w^3$, $\alpha_1(w) = 3w^2 - 2w^3$, $\beta_0(w) = w - 2w^2 + w^3$, $\beta_1(w) = -w^2 + w^3$. Параметрическая область поверхности Эрмита имеет вид

$$\begin{cases} u_0 \leqslant u \leqslant u_1, \\ 0 \leqslant v \leqslant n. \end{cases}$$

Если трансверсальные производные $\mathbf{r}_v(u, i)$ на линиях остова не заданы, то их, аналогично случаю сплайна Эрмита, можно определить, например так:

$$\begin{cases} \mathbf{r}_v(u, j) = \dfrac{\mathbf{r}(u, j+1) - \mathbf{r}(u, j-1)}{2}, \qquad 1 \leqslant j \leqslant n-1, \\ \mathbf{r}_v(u, 0) = 2\mathbf{r}(u, 1) - 2\mathbf{r}(u, 0) - \mathbf{r}_v(u, 1), \\ \mathbf{r}_v(u, n) = 2\mathbf{r}(u, n) - 2\mathbf{r}(u, n-1) - \mathbf{r}_v(u, n-1). \end{cases}$$

При таком определении третьи производные на кривых $\mathbf{r}_0$ и $\mathbf{r}_n$ в трансверсальном направлении обратятся в ноль:

$$\left.\frac{\partial^3 \mathbf{r}(u, v)}{\partial v^3}\right|_{v=0} = \left.\frac{\partial^3 \mathbf{r}(u, v)}{\partial v^3}\right|_{v=n} = 0.$$

Рис. 11.3:

Сами исходные кривые $\mathbf{r}_0, \ldots, \mathbf{r}_n$, образующие остов поверхности Эрмита, могут быть построены как кривые Безье или B-кривые. При этом получаются поверхности класса C^1, имеющие на двух участках границы управляемые производные в трансверсальном направлении (а именно, на участках $\mathbf{r}(u,0)$ и $\mathbf{r}(u,n)$). На участках границы $\mathbf{r}(u_0,v)$ и $\mathbf{r}(u_1,v)$ производные, вообще говоря, управляются лишь приблизительно — с помощью выбора кривых остова $\mathbf{r}_i(u)$.

11.4.2 Применение поверхностей Эрмита: поверхность перехода

Примером применения поверхностей Эрмита может служить задача объединения двух поверхностей в одну общую поверхность путем соединения их краев с помощью *поверхности перехода*.

Пусть даны две поверхности M и N, которые надо сопрячь, соединив поверхностью перехода край $\mathbf{r} = \mathbf{r}_M(t, u_1^M)$ первой поверхности и край $\mathbf{r} = \mathbf{r}_N(t, u_0^N)$ второй поверхности. Будем считать, что оба сопрягаемых края отнесены к одному и тому же параметру t, изменяющемуся в пределах $t_0 \leqslant t \leqslant t_1$. Второй параметр на поверхностях M и N обозначим (в обоих случаях) через u. Пусть параметрические пространства поверхностей имеют вид

$$\begin{cases} t_0 \leqslant t \leqslant t_1, \\ u_0^M \leqslant u \leqslant u_1^M \end{cases}$$

для поверхности $M : \mathbf{r} = \mathbf{r}_M(t,u)$ и

$$\begin{cases} t_0 \leqslant t \leqslant t_1, \\ u_0^N \leqslant u \leqslant u_1^N \end{cases}$$

для поверхности $N : \mathbf{r} = \mathbf{r}_N(t,u)$. Без ограничения общности будем считать, что $u_1^M < u_0^N$. Тогда поверхность перехода от M к N может быть описана

формулой поверхности Эрмита:

$$\mathbf{r}(t,w) = \mathbf{r}_M(t,u_1^M)(1-3w^2+2w^3) + \mathbf{r}_N(t,u_0^N)(3w^2-2w^3) + \\ + \frac{\partial}{\partial w}\Big|_{u=u_1^M}\mathbf{r}_M(t,u)(w-2w^2+w^3) + \frac{\partial}{\partial w}\Big|_{u=u_0^N}\mathbf{r}_N(t,u)(-w^2+w^3),$$

где $w = \dfrac{u-u_1^M}{u_0^N-u_1^M}$ — местный параметр по направлению u, $t_0 \leqslant t \leqslant t_1$, $0 \leqslant w \leqslant 1$. Данная поверхность Эрмита гладко сопрягает поверхности M и N по заданным краям. Класс гладкости сопряжения — C^1.

11.4.3 Поверхности Лагранжа

Аналогично интерполяционной кривой Лагранжа, можно построить *поверхность Лагранжа*, затягивающую заданный остов кривых (11.7). Эта поверхность описывается радиус-вектором:

$$\mathbf{r}(u,v) = \sum_{i=0}^{n} L_i(v)\mathbf{r}(u,i), \qquad u_0 \leqslant u \leqslant u_1, \qquad 0 \leqslant v \leqslant n,$$

где

$$L_i(v) = \frac{\prod\limits_{j=0,j\neq i}^{n}(v-j)}{\prod\limits_{j=0,j\neq i}^{n}(i-j)}.$$

Последняя формула получается из формулы для коэффициентов интерполяционного полинома Лагранжа подстановкой $t_i = i$ $(0 \leqslant i \leqslant n)$.

Поверхность Лагранжа имеет те же недостатки, что и интерполяционный полином Лагранжа — необоснованные изгибы при неравномерном расположении кривых остова в пространстве. По параметру v поверхность будет бесконечно гладкой и представляет собой многочлен. Гладкость по параметру u зависит от гладкости кривых остова.

11.4.4 Поверхности Гордона

Пусть задан остов поверхности, состоящий из двух семейств кривых в обоих параметрических направлениях, которые образуют *сетку* — т.е. каждая кривая одного семейства пресекает в одной и только одной точке каждую кривую другого семейства. Различные кривые одного и того же семейства не пересекаются. Обозначим заданную таким образом сетку кривых через

$$\begin{array}{ll} \mathbf{r} = \mathbf{r}(u,0),\ldots,\mathbf{r} = \mathbf{r}(u,n), & \text{(первое семейство)} \\ \mathbf{r} = \mathbf{r}(0,v),\ldots,\mathbf{r} = \mathbf{r}(m,v), & \text{(второе семейство)}. \end{array} \tag{11.8}$$

Здесь $0 \leqslant u \leqslant m$, $0 \leqslant v \leqslant n$. Тогда радиус-вектор точки пересечения i-ой кривой первого семейства с j-ой кривой второго семейства будет равен $\mathbf{r}(i,j)$, $0 \leqslant i \leqslant m$, $0 \leqslant j \leqslant n$.

Рис. 11.4:

Определение 11.3. *Поверхностью Гордона*, затягивающей сетку кривых (11.8), называется поверхность:

$$\mathbf{r}(u,v) = \sum_{j=0}^{n} L_j^n(v)\mathbf{r}(u,j) + \sum_{i=0}^{m} L_i^m(u)\mathbf{r}(i,v) - \\ - \sum_{i=0}^{m}\sum_{j=0}^{n} L_i^m(u)L_j^n(v)\mathbf{r}(i,j), \qquad 0 \leqslant u \leqslant m, \qquad 0 \leqslant v \leqslant n, \quad (11.9)$$

где

$$L_i^m(u) = \frac{\prod\limits_{r=0, r\neq i}^{m}(u-r)}{\prod\limits_{r=0, r\neq i}^{m}(i-r)}, \qquad L_j^n(v) = \frac{\prod\limits_{s=0, s\neq j}^{n}(v-s)}{\prod\limits_{s=0, s\neq j}^{n}(j-s)} \qquad (11.10)$$

— коэффициенты Лагранжа (рис. 11.4).

Обобщенная поверхность Гордона задается формулой (11.9), в которой вместо коэффициентов Лагранжа (11.10) используются произвольные C^1–функции $\alpha_k(w)$, удовлетворяющие условию

$$\alpha_k(w) = \begin{cases} 1, & \text{если} \quad w = k, \\ 0, & \text{если} \quad w \neq k. \end{cases}$$

Например, в качестве $\alpha_k(w)$ можно взять кубические сплайны. Гладкость обобщенной поверхности Гордона зависит от гладкости используемых функций $\alpha_k(w)$. Поверхности Гордона, построенные по коэффициентам Лагранжа, имеют гладкость C^∞, но могут образовывать необоснованные изгибы.

11.4.5 Поверхности, затягивающие сетку кривых заплатами Кунса

Остов, состоящий из сетки кривых (11.8) можно затянуть поверхностью, состоящей из заплат, затягивающих каждую отдельную ячейку сетки и гладко сопрягающихся между собой. В качестве таких заплат могут выступать поверхности Кунса. Получается следующая поверхность

$$\mathbf{r}(u,v) = \begin{pmatrix} \alpha_0(t) & \alpha_1(t) & \beta_0(t) & \beta_1(t) \end{pmatrix} \begin{pmatrix} \mathbf{r}(i,v) \\ \mathbf{r}(i+1,v) \\ \mathbf{r}_u(i,v) \\ \mathbf{r}_u(i+1,v) \end{pmatrix} +$$

$$+ \begin{pmatrix} \alpha_0(w) & \alpha_1(w) & \beta_0(w) & \beta_1(w) \end{pmatrix} \begin{pmatrix} \mathbf{r}(u,j) \\ \mathbf{r}(u,j+1) \\ \mathbf{r}_v(u,j) \\ \mathbf{r}_v(u,j+1) \end{pmatrix} -$$

$$- \begin{pmatrix} \alpha_0(t) & \alpha_1(t) & \beta_0(t) & \beta_1(t) \end{pmatrix} \times$$

$$\times \begin{pmatrix} \mathbf{r}(i,j) & \mathbf{r}(i,j+1) & \mathbf{r}_v(i,j) & \mathbf{r}_v(i,j+1) \\ \mathbf{r}(i+1,j) & \mathbf{r}(i+1,j+1) & \mathbf{r}_v(i+1,j) & \mathbf{r}_v(i+1,j+1) \\ \mathbf{r}_u(i,j) & \mathbf{r}_u(i,j+1) & \mathbf{r}_{uv}(i,j) & \mathbf{r}_{uv}(i,j+1) \\ \mathbf{r}_u(i+1,j) & \mathbf{r}_u(i+1,j+1) & \mathbf{r}_{uv}(i+1,j) & \mathbf{r}_{uv}(i+1,j+1) \end{pmatrix} \times$$

$$\times \begin{pmatrix} \alpha_0(w) \\ \alpha_1(w) \\ \beta_0(w) \\ \beta_1(w) \end{pmatrix}, \tag{11.11}$$

$t = u - i$, $w = v - j$, $0 \leqslant u \leqslant m$, $0 \leqslant v \leqslant n$. Здесь $i = [u]$ — целая часть u, $j = [v]$ — целая часть v, $t = u - i = \{u\}$ — дробная часть u, $w = v - j = \{v\}$ — дробная часть v, $t, w \in [0,1)$ — местные параметры в данной ячейке сетки (11.8). Согласно (11.8) области изменения параметров u и v таковы: $0 \leqslant u \leqslant m$, $0 \leqslant v \leqslant n$.

Для вычисления радиус-вектора поверхности с помощью формулы (11.11), необходимо задать производные на кривых сетки в трансверсальном направлении к этим кривым. По аналогии с кубическими сплайнами Эрмита, это можно сделать, например, по следующим формулам:

$$\begin{aligned} \mathbf{r}_u(i,v) &= \frac{\mathbf{r}(i+1,v) - \mathbf{r}(i-1,v)}{2}, \\ \mathbf{r}_v(u,j) &= \frac{\mathbf{r}(u,j+1) - \mathbf{r}(u,j-1)}{2} \end{aligned} \tag{11.12}$$

при $1 \leqslant u \leqslant m-1$, $1 \leqslant v \leqslant n-1$ (т.е. для внутренних линий сетки) и по формулам

$$\begin{aligned}
&\mathbf{r}_u(0,v) = 2\mathbf{r}(1,v) - 2\mathbf{r}(0,v) - \mathbf{r}_u(1,v),\\
&\mathbf{r}_u(m,v) = 2\mathbf{r}(m,v) - 2\mathbf{r}(m-1,v) - \mathbf{r}_u(m-1,v),\\
&\mathbf{r}_v(u,0) = 2\mathbf{r}(u,1) - 2\mathbf{r}(u,0) - \mathbf{r}_v(u,1),\\
&\mathbf{r}_v(u,n) = 2\mathbf{r}(u,n) - 2\mathbf{r}(u,n-1) - \mathbf{r}_v(u,n-1)
\end{aligned} \tag{11.13}$$

для крайних линий сетки.

Геометрически это означает, что производная $\mathbf{r}_u(i,v)$ в произвольной точке $\mathbf{r}(i,v)$ внутренней линии сетки полагается пропорциональной вектору, соединяющему точки $\mathbf{r}(i-1,v)$ и $\mathbf{r}(i+1,v)$ на соседних с ней линиях сетки, отвечающие тому же значению параметра v, что и данная точка. Аналогичное утверждение справедливо и для производной $\mathbf{r}_v(u,j)$.

Так же, как и в случае кубических сплайнов Эрмита, формулы (11.13) обеспечивают равенство нулю третьих производных поверхности на границе в трансверсальном к граничной кривой направлении.

11.4.6 Поверхности тензорного произведения

Рассмотрим поверхность Гордона (11.9), построенную по сетке кривых, которые являются кубическими сплайнами Эрмита. Ясно, что такая поверхность целиком определяется узлами сетки и первыми производными в узлах, заданными в обоих параметрических направлениях (поскольку этими величинами целиком определяются кубические сплайны Эрмита, являющиеся кривыми сетки). Таким образом, в данном случае поверхность строится фактически не по сетке кривых, а по набору точек (узлов) с заданными в них производными. Такие поверхности называются *поверхностями тензорного произведения*.

Для того, чтобы получить формулу поверхности тензорного произведения, поставим в формулу (11.9) выражения, определяющие кубические сплайны Эрмита (линии сетки) и их производные по u и v:

$$\begin{gathered}
\mathbf{r}(i,v) = \alpha_0(w)\mathbf{r}(i,j) + \alpha_1(w)\mathbf{r}(i,j+1) + \beta_0(w)\mathbf{r}_v(i,j) + \\
+\beta_1(w)\mathbf{r}_v(i,j+1),\\[2ex]
\mathbf{r}(u,j) = \alpha_0(t)\mathbf{r}(i,j) + \alpha_1(t)\mathbf{r}(i+1,j) + \beta_0(t)\mathbf{r}_u(i,j) + \\
+\beta_1(t)\mathbf{r}_u(i+1,j),\\[2ex]
\mathbf{r}_u(i,v) = \alpha_0(w)\mathbf{r}_u(i,j) + \alpha_1(w)\mathbf{r}_u(i,j+1) + \beta_0(w)\mathbf{r}_{vu}(i,j) + \\
+\beta_1(w)\mathbf{r}_{vu}(i,j+1),\\[2ex]
\mathbf{r}_v(u,j) = \alpha_0(t)\mathbf{r}_v(i,j) + \alpha_1(t)\mathbf{r}_v(i+1,j) + \beta_0(t)\mathbf{r}_{uv}(i,j) + \\
+\beta_1(t)\mathbf{r}_{uv}(i+1,j),
\end{gathered} \tag{11.14}$$

где, напомним:

$$\alpha_0(x) = 1 - 3x^2 + 2x^3, \qquad \alpha_1(x) = 3x^2 - 2x^3,$$
$$\beta_0(x) = x - 2x^2 + x^3, \qquad \alpha_1(x) = -x^2 + x^3,$$

где $x \in [0,1]$ — местный параметр. В формулах (11.14) через i и j обозначены целые части параметров u и v соответственно. Местные параметры t и w определяются как $t = u - i$, $w = v - j$. Вторые две формулы в (11.14) получаются дифференцированием первых двух. После подстановки формул (11.14) в (11.9) и несложных алгебраических преобразований получим следующее выражение для радиус-вектора поверхности тензорного произведения (в матричных обозначениях):

$$\mathbf{r}(u,v) = \begin{pmatrix} \alpha_0(t) & \alpha_1(t) & \beta_0(t) & \beta_1(t) \end{pmatrix} \times$$
$$\times \begin{pmatrix} \mathbf{r}(i,j) & \mathbf{r}(i,j+1) & \mathbf{r}_v(i,j) & \mathbf{r}_v(i,j+1) \\ \mathbf{r}(i+1,j) & \mathbf{r}(i+1,j+1) & \mathbf{r}_v(i+1,j) & \mathbf{r}_v(i+1,j+1) \\ \mathbf{r}_u(i,j) & \mathbf{r}_u(i,j+1) & \mathbf{r}_{uv}(i,j) & \mathbf{r}_{uv}(i,j+1) \\ \mathbf{r}_u(i+1,j) & \mathbf{r}_u(i+1,j+1) & \mathbf{r}_{uv}(i+1,j) & \mathbf{r}_{uv}(i+1,j+1) \end{pmatrix} \times$$
$$\times \begin{pmatrix} \alpha_0(w) \\ \alpha_1(w) \\ \beta_0(w) \\ \beta_1(w) \end{pmatrix}, \qquad t = u - i, \quad w = v - j, \quad 0 \leqslant u \leqslant m, \quad 0 \leqslant v \leqslant n. \tag{11.15}$$

Из формулы (11.15) хорошо видно, что поверхность тензорного произведения определяется величинами, заданными лишь в узлах сетки.

11.5 Поверхности с треугольной параметрической областью

При триангуляции поверхностей иногда бывает удобно рассматривать криволинейные поверхности треугольной формы, имеющие в качестве параметрической области треугольник на плоскости. При этом удобно пользоваться не декартовыми, а барицентрическими параметрическими координатами.

11.5.1 Барицентрические координаты

Лемма 11.1. *Пусть даны три точки A, B, $C \in \mathbb{R}^2$, не лежащие на одной прямой. Тогда для любой точки $P \in \mathbb{R}^2$ существует единственная тройка чисел $\{a, b, c\}$ таких, что*

$$\begin{cases} \mathbf{r}_P = a\mathbf{r}_A + b\mathbf{r}_B + c\mathbf{r}_C, \\ a + b + c = 1. \end{cases} \tag{11.16}$$

Здесь через $\mathbf{r}_P$, $\mathbf{r}_A$, $\mathbf{r}_B$, $\mathbf{r}_C$ *обозначены радиус-вектора точек* P, A, B, C *соответственно.*

Числа a, b, c называются *барицентрическим координатами* точки P относительно треугольника $\triangle ABC$.

Доказательство. Перепишем (11.16) в виде системы уравнений с неизвестными a, b, c. Пусть $A = (x_a, y_a)$, $B = (x_b, y_b)$, $C = (x_c, y_c)$, $P = (x, y)$. Имеем

$$\begin{cases} ax_a + bx_b + cx_c = x, \\ ay_a + by_b + cy_c = y, \\ a + b + c = 1. \end{cases} \tag{11.17}$$

Это — система из трех уравнений с тремя неизвестными. Определитель системы (11.17) равен

$$\Delta = \begin{vmatrix} 1 & 1 & 1 \\ x_a & x_b & x_c \\ y_a & y_b & y_c \end{vmatrix} = 2S_{ABC} \neq 0.$$

Следовательно, система (11.17) имеет единственное решение для любой правой части. Поэтому числа a, b, c определены однозначно. По правилу Крамера они равны

$$a = \frac{1}{\Delta} \begin{vmatrix} 1 & 1 & 1 \\ x & x_b & x_c \\ y & y_b & y_c \end{vmatrix} = \frac{xy_b + x_by_c + x_cy - yx_b - y_bx_c - y_cx}{2S_{ABC}},$$

$$b = \frac{1}{\Delta} \begin{vmatrix} 1 & 1 & 1 \\ x_a & x & x_c \\ y_a & y & y_c \end{vmatrix} = \frac{x_ay + xy_c + x_cy_a - y_ax - yx_c - y_cx_a}{2S_{ABC}},$$

$$c = \frac{1}{\Delta} \begin{vmatrix} 1 & 1 & 1 \\ x_a & x_b & x \\ y_a & y_b & y \end{vmatrix} = \frac{x_ay_b + x_by + xy_a - y_ax_b - y_bx - yx_a}{2S_{ABC}}.$$

Лемма доказана. ■

Следствие 11.1. *Имеют место формулы*

$$|a| = \frac{S_{PBC}}{S_{ABC}}, \qquad |b| = \frac{S_{APC}}{S_{ABC}}, \qquad |c| = \frac{S_{ABP}}{S_{ABC}},$$

т.е. коэффициенты a, b, c *по модулю равны отношению площади треугольника* $\triangle ABC$, *в котором соответствующая вершина заменена на* P, *к площади исходного треугольника* $\triangle ABC$.

Следствие 11.2. *Все три барицентрические координаты точки положительны тогда и только тогда, когда точка принадлежит внутренности треугольника* $\triangle ABC$. *Все три барицентрические координаты точки неотрицательны тогда и только тогда, когда точка принадлежит треугольнику* $\triangle ABC$.

Рис. 11.5:

11.5.2 Билинейная треугольная поверхность

Пусть даны три точки P_1, P_2, P_3 в $\mathbb{R}^3$. В качестве области параметров возьмем треугольник $\triangle ABC \subset \mathbb{R}^2$. Точка параметрического пространства описывается координатами $0 \leqslant a, b, c \leqslant 1$, $a + b + c = 1$. Радиус-вектор треугольной поверхности с вершинами в точках P_1, P_2, P_3 задается формулой $\mathbf{r}(a, b, c) = a\mathbf{r}_1 + b\mathbf{r}_2 + c\mathbf{r}_3$, где $\mathbf{r}_i$ — радиус-вектор точки P_i.

11.5.3 Треугольная поверхность на трех кривых

Пусть даны три попарно пересекающиеся кривые $\mathbf{r} = \mathbf{r}_1(u)$, $\mathbf{r} = \mathbf{r}_2(v)$, $\mathbf{r} = \mathbf{r}_3(w)$, $0 \leqslant u, v, w \leqslant 1$, пересекающиеся в точках $P_1 = \mathbf{r}_2(0) = \mathbf{r}_3(1)$, $P_2 = \mathbf{r}_1(1) = \mathbf{r}_3(0)$, $P_3 = \mathbf{r}_1(0) = \mathbf{r}_2(1)$. Определим треугольную поверхность формулой:

$$\mathbf{r}(a, b, c) = a\big(\mathbf{r}_3(1 - b) + \mathbf{r}_2(c) - P_1\big) + b\big(\mathbf{r}_1(1 - c) + \mathbf{r}_3(a) - P_2\big) + \\ + c\big(\mathbf{r}_2(1 - a) + \mathbf{r}_1(b) - P_3\big), \quad (11.18)$$

где $0 \leqslant a, b, c \leqslant 1$, $a + b + c = 1$ (рис. 11.5). Легко проверить, что при $c = 0$ формула (11.18) описывает кривую $\mathbf{r}_3(a)$ $(0 \leqslant a \leqslant 1)$, при $a = 0$ — кривую $\mathbf{r}_1(b)$ $(0 \leqslant b \leqslant 1)$, при $b = 0$ — кривую $\mathbf{r}_2(c)$ $(0 \leqslant c \leqslant 1)$. Например, если $a = 0$, то $c = 1 - b$ и, согласно (11.18), имеем

$$\mathbf{r}(0, b, 1 - b) = b\big(\mathbf{r}_1(b) + \mathbf{r}_3(0) - P_2\big) + (1 - b)\big(\mathbf{r}_2(1) + \mathbf{r}_1(b) - P_3\big) = \\ = b\big(\mathbf{r}_1(b) + P_2 - P_2\big) + (1 - b)\big(P_3 + \mathbf{r}_1(b) - P_3\big) = \mathbf{r}_1(b).$$

Упражнение. Если кривые $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ являются дугами окружностей, расположенных в трех взаимно-перпендикулярных координатных плоскостях,

пересекающихся в общем центре этих дуг, то треугольная поверхность (11.18) — область на сфере, высекаемая соответствующим координатным квадрантом.

Глава 12

Компьютерная геометрия проективно преобразованных изображений (Г. В. Носовский, Е. С. Скрипка, А. А. Толченников)

12.1 Введение

Одной из важных задач современной компьютерной геометрии является построение эффективных алгоритмов склейки плоских изображений одного и того же объекта, полученных путем центрального проектирования из различных точек трехмерного пространства. Эта задача возникает, например, в стереофотограмметрии [42], в робототехнике ("зрение роботов") [46], а также при построении объемного изображения объекта на основе его плоских изображений, полученных с помощью нескольких видеокамер [51].

С чисто геометрической точки зрения рассматриваемая задача сводится к вычислению проективного преобразования F, связывающего две (вообще говоря, неизвестные заранее) области D_1 и D_2, лежащие в одной аффинной карте проективного пространства $\mathbb{R}P^2$, см. рис. 12.1. Для ее решения необходимо уметь распознавать соответствующие друг другу при искомом преобразовании точки. Следуя терминологии, принятой в приложениях, мы

Рис. 12.1: Задача склейки двух изображений одного объекта, полученных центральным проектированием из разных точек пространства

будем называть такие точки *сопряженными относительно* F или просто *сопряженными*.

Известные к настоящему времени алгоритмы распознавания сопряженных точек в проективно преобразованных изображениях либо недостаточно эффективны, либо требуют привлечения существенного объема дополнительной информации, которая далеко не всегда бывает доступна.

Из проективной геометрии известно, что для того, чтобы полностью определить проективное преобразование F, действующее в $\mathbb{R}P^2$, достаточно знать с абсолютной точностью две четверки сопряженных точек общего положения в $\mathbb{R}P^2$. В практических задачах, однако, сопряженные точки определяются не абсолютно точно, а лишь с некоторой ошибкой. В то же время, их количество может быть достаточно велико, что позволяет применять методы математической статистики для вычисления оптимального в той или иной ситуации проективного преобразования. Однако при этом необходимо уметь оценивать устойчивость вычисляемого проективного преобразования к возмущению начальных данных. Подобные оценки зависит как от применяемого метода определения проективного преобразования, так и от конфигурации (и, естественно, от точности задания) сопряженных точек. В настоящей главе получен ряд таких оценок — как для общей ситуации, так и для частного случая часто применяемого на практике — как правило, на основе лишь эвристических оценок точности, — “нормированного линейного алгоритма” (НЛА). Полученные оценки позволяют, в частности, сформулировать некоторые полезные общие правила выбора оптимальной конфигурации сопряженных точек.

12.2 Основные понятия проективной геометрии

На фотографическом снимке или реалистическом рисунке плоской местности изображения параллельных прямых, вообще говоря, пересекаются, а изображения равных отрезков одной прямой, вообще говоря, не равны.

Это говорит о том, что отображение местности на плоскости снимка или рисунка не является аффинным. Более того, на плоском рисунке местности изображения параллельных прямых пересекаются в точке, которая не имеет прообраза на местности (иначе прямые не были бы параллельными). Таким образом мы имеем дело с центральным проектированием. Это отображение не является аффинным, не взаимно однозначным и и не всюду определено.

Для изучения центрального проектирования удобно рассматривать множество, называемое *проективной плоскостью*, "точками" которого являются прямые, проходящие через центр проектирования, а пересечения этих прямых с плоскостью проектирования считать изображением соответствующих "точек". При этом "точки", соответствующие прямым, параллельным выбранной плоскости проектирования, не получат никакого изображения (но они будут иметь изображение при другом выборе плоскости проектирования). Такие точки называются "бесконечно удаленными точками" по отношению к данной плоскости проектирования.

Далее, множество "точек", соответствующих прямым, лежащим в фиксированной плоскости, проходящей через центр проектирования, естественно назвать "прямой" проективной плоскости. На плоскости проектирования такая "прямая", за вычетом ее "бесконечно удаленной точки", изображается в виде обычной прямой. Единственным исключением является "прямая", соответствующая плоскости, параллельной плоскости проектирования. Она целиком состоит из "бесконечно удаленных точек" и не получает никакого изображения на данной плоскости проектирования (или, как говорят в геометрии — на данной аффинной карте). Эта "прямая" называется "бесконечно удаленной прямой" по отношению к данной плоскости проектирования.

Определение 12.1. Множество одномерных подпространств $(n+1)$-мерного евклидова пространства $\mathbb{R}^{n+1}$ называется *n-мерным проективным пространством* и обозначается $\mathbb{R}P^n$.

Для любого ненулевого вектора $\mathbf{x} \in \mathbb{R}^{n+1}$ будем обозначать через $\widehat{\mathbf{x}}$ одномерное подпространство $\langle \mathbf{x} \rangle$, рассматриваемое как точка проективного пространства $\mathbb{R}P^n$.

Определение 12.2. *Однородными координатами* точки $\widehat{\mathbf{x}} \in \mathbb{R}P^n$ называются координаты вектора $\mathbf{x}$ в каком-либо базисе пространства $\mathbb{R}^{n+1}$. Однородные координаты определены лишь с точностью до общего множителя $\lambda \neq 0$. Точку с однородными координатами $x_0, x_1, \ldots, x_n$ будем обозначать $(x_0 : \ldots : x_n)$.

Определение 12.3. *Неоднородными координатами* точки $\widehat{\mathbf{x}} \in \mathbb{R}P^n$ называются аффинные координаты ее пересечения (как прямой в $\mathbb{R}^{n+1}$) с какой-либо фиксированной плоскостью в $\mathbb{R}^{n+1}$, не проходящей через начало координат. При указанном соответствии точки выбранной плоскости взаимно-однозначно соответствуют точкам некоторой области в $\mathbb{R}P^n$. Любая такая плоскость (с заданной на ней системой аффинных координат), изображающая описанным способом область в $\mathbb{R}P^n$, называется *аффинной картой* в $\mathbb{R}P^n$.

Проиллюстрируем на примере связь между однородными и неоднородными координатами.

Пусть $\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_n$ — базис пространства $\mathbb{R}^{n+1}$. Рассмотрим аффинную карту $S_0 = \mathbf{e}_0 + \langle \mathbf{e}_1, \ldots, \mathbf{e}_n \rangle$. Изображением точки $\widehat{\mathbf{x}} = (x_0 : \ldots : x_n)$ на S_0 служит точка $\mathbf{e}_0 + (x_1/x_0)\mathbf{e}_1 + \cdots + (x_n/x_0)\mathbf{e}_n$, аффинные координаты которой относительно репера $\{\mathbf{e}_0 : \ldots : \mathbf{e}_n\}$ суть $(x_1/x_0, \ldots, x_n/x_0)$. Таким образом, при указанном выборе аффинной карты неоднородными координатами точки $(x_0 : \ldots : x_n)$ служат отношения $x_1/x_0, \ldots, x_n/x_0$. Точки с $x_0 = 0$ являются бесконечно удаленными по отношению к S_0 (т.е. — не входят в нее).

Всякий невырожденный линейный оператор $\mathcal{A}$, действующий в $\mathbb{R}^{n+1}$, переводит одномерные подпространства в одномерные подпространства и тем самым определяет некоторое взаимно однозначное преобразование $\widehat{\mathcal{A}}$ пространства $\mathbb{R}P^n$ в себя.

Определение 12.4. Преобразования вида $\widehat{\mathcal{A}}$, где $\mathcal{A}$ — некоторый линейный оператор в $\mathbb{R}^{n+1}$, называются *проективными преобразованиями* пространства $\mathbb{R}P^n$.

Определение 12.5. Будем говорить, что $(n+2)$ точки n-мерного проективного пространства находятся в *общем положении*, если никакая $(n+1)$ из них не лежит в одной и той же гиперплоскости.

В курсах проективной геометрии доказывается следующая теорема

Теорема 12.1. *Пусть $\{\mathbf{p}_0, \ldots, \mathbf{p}_{n+1}\}$ и $\{\mathbf{q}_0, \ldots, \mathbf{q}_{n+1}\}$ — два множества точек, каждое из которых состоит из $(n+2)$ точек общего положения n-мерного проективного пространства $\mathbb{R}P^n$. Тогда существует единственное проективное преобразование, переводящее каждую точку $\mathbf{p}_i$ первого множества в соответствующую ей точку $\mathbf{q}_i$ второго множества, $i = 0, \ldots, n+1$.*

12.3 Метод распознавания сопряженных точек по разнесениям связанных признаков (Г. В. Носовский)

Пусть имеется два фото (или видео) снимка E_1 и E_2, полученные с помощью двух различных камер, в поле зрения которых попал один и тот же объект X, который без ограничения общности можно считать двумерной поверхностью. Таким образом, с математической точки зрения можно считать, что мы имеем дело с двумя изображениями одной и той же поверхности X, полученными центральным проектированием из разных точек пространства в двух различных направлениях, рис. 12.1. Взаимное расположение центров проектирования и направления осей проектирования в пространстве нам, вообще говоря, неизвестно. Кроме того, возможен случай, когда изображение поверхности X занимает лишь часть поля снимков E_1 и E_2.

Будем называть точки (в дискретном случае — ячейки) снимков E_1 и E_2 *сопряженными*, если в них спроектировались лучи, исходящие из одной и той же точки поверхности X.

Будем считать, что каждый из снимков E_k $(k = 1, 2)$ уже оцифрован, т.е. представляет собой матрицу размера $n \times n$, в каждой ячейке которой стоит некоторое, вообще говоря, векторное значение e_k^{ij}, характеризующее соотношение цветов и интенсивность изображения в данной точке.

Будем считать, что величина e_k^{ij} может принимать достаточно много существенно различных значений. В данном случае существенно различными для нас будут такие значения, которые, за исключением случаев случайных помех или возмущений, не появляются в сопряженных точках снимков E_1 и E_2.

Ниже мы будем предполагать, что область возможных значений e_k^{ij} заранее разбита на конечное число подмножеств (групп) близких значений. Другими словами, мы будем рассматривать группированную выборку. Будем считать, что имеется число m таких групп, называть их *признаками* или *именами* и обозначать через $a_1, \ldots, a_m$. Множество всех возможных имен обозначим через

$$I = \{a_1, \ldots, a_m\}.$$

Таким образом, каждый из снимков $E_k (k = 1, 2)$ может быть представлен в виде матрицы размера $n \times n$ в ячейках которой стоят целые числа $1 \leqslant e_k^{ij} \leqslant m$ — номера признаков (имен).

Ниже мы будем предполагать, что число m достаточно велико. Так будет, например, если снимки E_1 и E_2 являются цветными. В случае черно-белых или полутоновых снимков для увеличения числа m можно в качестве элементарной ячейки брать целый блок оцифрованных точек размером $d \times d$, где $d > 1$.

Нашей задачей будет построение такого алгоритма распознавания сопряженных ячеек снимков E_1 и E_2, который бы работал в общей ситуации, когда не имеется никакой априорной информации об объекте X, изображения которого присутствуют на обоих снимках. Это значит, что вся информация о сопряженных точках и их окружениях должна черпаться только из самих снимков E_1 и E_2. Предполагается, что это — обычные фотографии без каких-либо добавлений, помогающих распознаванию.

Мы будем рассматривать случай, когда вертикальные оси снимков E_1 и E_2 примерно соответствуют друг другу. Достаточно предположить, что столбцы и строки матриц E_1 и E_2 изображают вертикальные и горизонтальные плоскости трехмерного пространства, в котором производилась съемка. Для большинства приложений этого достаточно, поскольку в реальности вертикаль фотографии обычно соответствует вертикальным плоскостям в пространстве.

Воспользуемся непараметрическим подходом к поиску дубликатов в перемешанных последовательностях, предложенным в работах А. Т. Фоменко и Г. В. Носовского [25, 26].

Пусть число n (размер матриц E_1, E_2) представлено в виде произведения

двух своих делителей $n = Np$. Фиксируем разбиение каждой из матриц E_1 и E_2 на квадратные блоки размера $p \times p$, состоящие из p^2 элементарных ячеек. Всего в каждой из матриц будет N^2 таких блоков, которые будем обозначать $E_k^{\alpha\beta}$, $1 \leqslant k \leqslant 2$, $1 \leqslant \alpha, \beta \leqslant N$:

$$E_k^{\alpha\beta} = \{e_k^{ij} \mid (\alpha - 1)p < i \leqslant \alpha p, \quad (\beta - 1)p < i \leqslant \beta p\}.$$

Блоки $E_k^{\alpha\beta}$ в нашем алгоритме будут выступать как участки изображения, состоящие из достаточно близких друг к другу ячеек. Число p (размер блока) является параметром, который выбирается в зависимости от конкретной прикладной задачи. Ниже мы будем считать, что $p \geqslant 2$.

Определение 12.6. *Разнесением* двух ячеек $e_1^{ij} \in E_1$ и $e_2^{i'j'} \in E_2$ назовем двумерный вектор ρ с целочисленными компонентами:

$$\rho(e_1^{ij}, e_2^{i'j'}) = (i' - i, j' - j). \tag{12.1}$$

Определение 12.7. Два имени $n_i, n_j \in I$ назовем *связанными* (относительно данной пары матриц E_1 и E_2 с данным разбиением на блоки), если они появились в ячейках одного и того же блока хотя бы одной из матриц E_1 и E_2:

$$\exists\, k \in \{1, 2\}, \quad \exists\, \alpha, \beta \in \{1, \ldots, N\} : \quad n_i, n_j \in E_k^{\alpha\beta}.$$

Связанность двух имен $n_i, n_j \in I$ $(1 \leqslant i, j \leqslant m)$ мы будем обозначать $n_i \sim n_j$.

Если две ячейки $e_1^{ij} \in E_1$ и $e_2^{i'j'} \in E_2$ содержат связанные между собой имена, то эти ячейки мы также будем называть *связанными* и обозначать: $e_1^{ij} \sim e_2^{i'j'}$.

В частности, любое имя, хотя бы раз встретившееся на одном из на снимков E_1 и E_2, связано с самим собой относительно этих снимков.

Во избежание путаницы отметим, что мы используем один и тот же знак включения "$\in$" как для ячеек матриц E_1 и E_2, так и для имен, которые появляются в этих ячейках. В первом случае имеется в виду, что ячейка является элементом матрицы (или ее блока), а во втором — что данное имя присутствует в какой-то из ячеек этой матрицы (или ее блока).

С помощью матриц E_1 и E_2 построим конечную вероятностную схему равновероятного выбора одной ячейки из E_1 и одной ячейки из E_2. Для этого возьмем в качестве пространства элементарных событий декартово произведение $\Omega = E_1 \times E_2$, в качестве алгебры событий Σ — совокупность всех подмножеств Ω, в качестве вероятностной меры $\mathbf{P}$ — равномерное распределение на Ω: $\mathbf{P}\{\omega\} = 1/n^2$ для любого $\omega \in \Omega$. Ячейку, случайно выбранную в матрице E_1, обозначим через $e_1^{ij}(\omega)$, а ячейку, случайно выбранную в матрице E_2 — $e_2^{i'j'}(\omega)$. Здесь индексы $i = i(\omega)$, $j = j(\omega)$, $i' = i'(\omega)$, $j' = j'(\omega)$ — это случайные величины, определенные на вероятностном пространстве $(\Omega, \Sigma, \mathbf{P})$, но для упрощения записи аргумент ω у индексов мы не пишем.

Обозначим $b_1 = b_1(\omega)$ значение (имя), содержащееся в первой выбранной ячейке $e_1^{ij}(\omega) \in E_1$, а через $b_2(\omega)$ — имя из второй выбранной ячейки $e_2^{i'j'}(\omega) \in E_2$.

Определим в построенной вероятностной схеме случайную величину $\xi = \xi(\omega)$ — разнесение выбранной пары ячеек:

$$\xi(\omega) = \rho\big(e_1^{ij}(\omega), e_2^{i'j'}(\omega)\big). \tag{12.2}$$

Случайная величина ξ с вероятностью единица принимает значения на участке целочисленной двумерной решетки: $\{(i_1, i_2) : 1 - n \leqslant i, j \leqslant n - 1\}$.

Предположим, в матрицах E_1 и E_2 есть хотя бы одна пара связанных между собой ячеек. Это предположение, как правило, бывает выполнено на практике, поскольку по предположению снимки E_1 и E_2 содержат изображения одного и того же объекта X. Следовательно, даже при наличии некоторой зашумленности, должны найтись такие пары близких (попавших в один блок) имен на одном снимке, что оба эти имени или хотя бы одно из них появится и на другом снимке. Легко видеть, что это означает существование связанных пар ячеек в матрицах E_1 и E_2.

В этом предположении будет отлична от нуля вероятность $\mathbf{P}(A)$ события

$$A = \big\{\omega : b_1(\omega) \sim b_2(\omega)\big\}, \tag{12.3}$$

наступающего тогда, когда выбранная пара ячеек связана между собой (см. определение 12.7).

Предлагаемый алгоритм поиска сопряженных точек включает в себя предварительный отсев сравниваемых участков двух изображений на основе сравнении двух распределений:

$$\begin{aligned} f_0(i,j) &= \mathbf{P}\big\{\xi(\omega) = (i,j)\big\}, \\ f_1(i,j) &= \mathbf{P}\big\{\xi(\omega) = (i,j) | A\big\}, \end{aligned} \qquad 1 - n \leqslant i, j \leqslant n - 1. \tag{12.4}$$

Здесь f_0 — это безусловное распределение случайной величины ξ на целочисленной решетке, а f_1 — ее условное распределение при условии, что наступило событие A, определенное в (12.3).

Ниже нам понадобится явный вид распределения f_0. Он дается в следующих двух леммах.

Лемма 12.1. *Для любых* $1 - n \leqslant i, j \leqslant n - 1$ *имеют место равенства*:

$$f_0(i,j) = f_0(-i,j) = f_0(i,-j) = f_0(-i,-j).$$

Доказательство. Обозначим через V вращение квадратной матрицы на $\pi/2$:

$$V(e_k^{ij}) = e_k^{j,n+1-i} \quad (1 \leqslant k \leqslant 2, \quad 1 \leqslant i, j \leqslant n).$$

Преобразование V и его итерации являются перестановками множества ячеек матрицы E_k. Кроме того, V, очевидно, является автоморфизмом пространства элементарных событий $\Omega = E_1 \times E_2$ и сохраняет на нем равномерное распределение: $\mathbf{P}\{V\omega\} = \mathbf{P}\{\omega\}$. Далее, поскольку при вращении

обоих матриц на один и тот же угол, вектор разнесения между любой парой ячеек вращается на тот же угол, имеем

$$f_0(i,j) = \mathbf{P}\{\xi(\omega) = (i,j)\} = \mathbf{P}\{\xi(V\omega) = (i,j)\} = \\ = \mathbf{P}\{\xi(\omega) = (-i,j)\} = f_0(-i,j).$$

Аналогично:

$$f_0(i,j) = \mathbf{P}\{\xi(V^2\omega) = (i,j)\} = \mathbf{P}\{\xi(\omega) = (-i,-j)\} = f_0(-i,-j),$$
$$f_0(i,j) = \mathbf{P}\{\xi(V^3\omega) = (i,j)\} = \mathbf{P}\{\xi(\omega) = (i,-j)\} = f_0(i,-j).$$

Лемма доказана. ■

В силу доказанной леммы, для нахождения распределения f_0 достаточно вычислить значения $f_0(i,j)$ при $i,j \geqslant 0$. Они даются следующей леммой.

Лемма 12.2. *Пусть* $0 < i,j \leqslant n-1$. *Тогда*

$$\begin{aligned} &f_0(i,j) = \mathbf{P}\{\xi(\omega) = (i,j)\} = \frac{1}{n^2} - \frac{i+j}{n^3} + \frac{ij}{n^4}, \\ &f_0(i,0) = \frac{1}{n^2} - \frac{i}{n^3}, \quad f_0(0,j) = \frac{1}{n^2} - \frac{j}{n^3}, \quad f_0(0,0) = \frac{1}{n^2}. \end{aligned} \tag{12.5}$$

Доказательство. Докажем первое равенство из (12.5). Вероятность получить разнесение (i,j) между парой ячеек $e_1^{ij}(\omega)$ и $e_2^{i'j'}(\omega)$ при равномерном распределении $\mathbf{P}$ равна отношению числа благоприятствующих исходов к общему числу исходов. Общее число исходов равно n^4 – число различных способов выбрать пару ячеек из матриц E_1 и E_2 соответственно. Число благоприятствующих исходов в данном случае равно числу способов выбрать такую ячейку в E_1, для которой существует разнесенная с ней на (i,j) пара. В этом случае парная ячейка в E_2 определяется однозначно. Но количество ячеек в матрице размера $n \times n$ таких, что существует ячейка, сдвинутая относительно нее на вектор (i,j), равно, очевидно, $n^2 - ni - nj + ij$. Следовательно,

$$f_0(i,j) = \frac{n^2 - ni - nj + ij}{n^4} = \frac{1}{n^2} - \frac{i+j}{n^3} + \frac{ij}{n^4}.$$

Остальные равенства доказываются аналогично. Лемма доказана. ■

Замечание. Нетрудно показать, что распределение f_0 не изменится, если разнесение ячеек $e_1^{ij} \in E_1^{\alpha\beta}$ и $e_2^{i'j'} \in E_1^{\alpha'\beta'}$ определить не по формуле (12.1), а по формуле:

$$\widetilde{\rho}(e_1^{ij}, e_2^{i'j'}) = (\alpha' - \alpha, \beta' - \beta').$$

Следствие 12.1. *Распределение f_0 не зависит от того, какие значения находятся в ячейках матриц E_1 и E_2 и определяется только размером матриц n.*

Разность распределений f_1 и f_0 будет иметь всплеск на величинах типичных сдвигов между сопряженными ячейками матриц E_1 и E_2 [25, 26]. Этот всплеск можно статистически распознать и определить по нему значения типичных сдвигов между сопряженными ячейками.

Предлагаемый алгоритм распознавания сопряженных точек на изображениях E_1 и E_2 состоит из двух этапов вычислений. На первом этапе вычисляются и сравниваются между собой распределения f_0 и f_1. На основе их статистического сравнения определяются типичные сдвиги между именами, связанными относительно матриц E_1 и E_2. Они интерпретируются как сдвиги между сопряженными ячейками [25, 26]. Алгоритмически этот этап может быть осуществлен за один проход по ячейкам матриц E_1 и E_2. Сложность соответствующих вычислений линейна по числу ячеек $2n^2$ при фиксированных N и m.

Затем, на втором этапе, при окончательном определении сопряженных точек сравниваются между собой уже не произвольные фрагменты матриц E_1 и E_2, а только такие пары блоков $(E_1^{\alpha\beta}, E_2^{\alpha'\beta'})$, которые содержат достаточно много пар связанных между собой ячеек, разнесенных на величины определенных на первом этапе типичных сдвигов. Это резко сокращает количество пар фрагментов E_1 и E_2, которые надо сравнивать. На этом этапе можно использовать стандартные методы корреляционного анализа, обычно применяемые в этих случаях [51], но часто бывает достаточно простое совмещение пиков интенсивности (яркости или цвета).

Подчеркнем, что мы не предполагаем, что изображения объекта X на снимках E_1 и E_2 могут быть совмещены с помощью сдвига или с помощью некоторого аффинного преобразования. Искомое преобразование предполагается проективным. Однако распределение сдвигов между сопряженными точками при проективном преобразовании в рассматриваемом случае, как правило, будет иметь всплеск, соответствующий типичным значением сдвигов между сопряженными точками, см. рис. 12.2. Напомним, что мы считаем, что оси снимков E_1 и E_2 приблизительно соответствуют друг другу (т.е. снимки E_1 и E_2 не сильно повернуты друг относительно друга). В противном случае, указанный метод будет эффективен после соответствующей модификации, компенсирующей поворот.

В заключение этого раздела обсудим вкратце вопрос об отсечении случайных всплесков разности $f_1 - f_0$, мешающих распознаванию "полезного" всплеска, указывающего на величины типичных сдвигов между сопряженными точками.

Предположим на мгновение, что на снимках E_1 и E_2 сопряженных точек нет. Тогда естественно считать, что матрицы E_1 и E_2 являются независимыми друг от друга случайными элементами, определенными на некотором вероятностном пространстве $(\Omega_1, \Sigma_1, \mathbf{P}_1)$. В этом случае случайная величина ξ, заданная соотношением (12.2), и событие A, введенное в (12.3), будут определены на произведении вероятностных пространств $(\Omega_1, \Sigma_1, \mathbf{P}_1) \times (\Omega, \Sigma, \mathbf{P})$. Если E_1 и E_2 независимы, то случайная величина ξ будет независима от события A. Следовательно, в этом случае распре-

Рис. 12.2: Возникновение области типичных значений сдвига между двумя проективными изображениями объекта X при наложении снимков E_1 и E_2

деления f_0 и f_1 как распределения на $\Omega_1 \times \Omega$ будут совпадать. При этом из леммы 12.2 следует, что f_0 будет задаваться теми же соотношениями (12.5). А следовательно, и распределение f_1, рассматриваемое как распределение на произведении $\Omega_1 \times \Omega$, также будет определяться формулами (12.5). Для фиксированной реализации ω_1, (т.е. на конкретных матрицах E_1 и E_2) распределение f_1 может отличаться от f_0. Но для больших n возникновение значительных отличий между ними в силу центральной предельной теоремы будет маловероятно в широком классе вероятностных законов $\mathbf{P}_1$. Поэтому при больших n (порядка нескольких сотен) случайные всплески разности $f_1 - f_0$, как правило, невелики и типичный сдвиг между сопряженными точками обычно легко распознается.

Для оценки значимости всплесков $f_1 - f_0$ можно пользоваться, например, следующей леммой 12.3. Перед ее формулировкой введем необходимые обозначения и предположения.

Предположим, что данная пара квадратных матриц (E_1, E_2) размера $n \times n$ является реализацией некоторого случайного эксперимента $(\Omega_1, \Sigma_1, \mathbf{P}_1)$. Без ограничения общности будем считать, что $(E_1, E_2) = \omega_1^0 \in \Omega_1$ — некоторое элементарное событие пространства Ω_1 и что произвольный элемент $\omega_1 \in \Omega_1$ имеет вид пары квадратных матриц размера $n \times n$, в ячейках которых находятся целые числа от 1 до m.

Рассмотрим описанную выше конечную схему $(\Omega, \Sigma, \mathbf{P})$ равновероятного выбора двух ячеек из пары матриц $\omega_1 = (E_1, E_2)$. Определим на произведении вероятностных пространств $(\Omega_1, \Sigma_1, \mathbf{P}_1) \times (\Omega, \Sigma, \mathbf{P})$ случайную величину ξ соотношением (12.2) и событие A соотношением (12.3).

Определим функции f_0, f_1 (плотности безусловного и условного распределения ξ) соотношениями, аналогичными (12.4), где вместо меры $\mathbf{P}$ возьмем произведение мер $\mathbf{P}_1 \times \mathbf{P}$. При этом, из леммы 12.2 следует, что функция f_0 будет по-прежнему определяться формулами (12.5).

Ниже через ω_1 и ω мы, как обычно, будем обозначать не только элементы множеств Ω_1 и Ω, но и подмножества в $\Omega_1 \times \Omega$ вида $\omega_1 \times \Omega$ и $\Omega_1 \times \omega$.

Фиксируем некоторые $\varepsilon > 0$, $1 - n \leqslant i, j \leqslant n - 1$. Определим событие:

$$A_{ij}^{\varepsilon} = \big\{(\omega_1, \omega) : \mathbf{P}\{\xi = (i,j)|A, \omega_1\} - \mathbf{P}\{\xi = (i,j)|A\} \geqslant \varepsilon\big\}. \tag{12.6}$$

При условии независимости ξ и A событие A_{ij}^{ε} будет состоять из таких $(E_1, E_2) \times \omega$, что вычисленная для (E_1, E_2) в точке (i,j) разность $\big(f_1(i,j) - f_0(i,j)\big)$ функций, определенных в (12.4), будет не меньше, чем ε.

В указанных обозначениях справедлива следующая

Лемма 12.3. *Предположим, что ξ и A независимы. Тогда для любых $\varepsilon > 0$ и $1 - n \leqslant i, j \leqslant n - 1$:*

$$\mathbf{P}\big\{A_{ij}^{\varepsilon}|A\big\} \leqslant \frac{f_0(i,j)}{f_0(i,j) + \varepsilon}.$$

Доказательство. Фиксируем некоторые $\varepsilon > 0$, $1 - n \leqslant i, j \leqslant n - 1$. Обозначим через B событие $\{\xi = (i,j)\}$. По предположению, B и A независимы. Имеем по формуле полной вероятности и в силу (12.6):

$$\begin{aligned}\mathbf{P}\{B|A\} &= \frac{\mathbf{P}\{BA\}}{\mathbf{P}\{A\}} = \frac{\sum\limits_{\omega_1} \mathbf{P}\{BA\omega_1\}}{\mathbf{P}\{A\}} = \frac{\sum\limits_{\omega_1} \mathbf{P}\{B|A\omega_1\}\mathbf{P}\{A\omega_1\}}{\mathbf{P}\{A\}} = \\ &= \sum_{\omega_1} \mathbf{P}\{B|A\omega_1\}\mathbf{P}\{\omega_1|A\} \geqslant \sum_{\omega_1 \subset A_{ij}^{\varepsilon}} \mathbf{P}\{B|A\omega_1\}\mathbf{P}\{\omega_1|A\} \geqslant \\ &\geqslant \sum_{\omega_1 \subset A_{ij}^{\varepsilon}} \big(\mathbf{P}\{B|A\} + \varepsilon\big)\mathbf{P}\{\omega_1|A\} = \big(\mathbf{P}\{B|A\} + \varepsilon\big)\mathbf{P}\{A_{ij}^{\varepsilon}|A\}.\end{aligned}$$

Следовательно, в силу независимости B и A:

$$\mathbf{P}\{A_{ij}^{\varepsilon}|A\} \leqslant \frac{\mathbf{P}\{B|A\}}{\mathbf{P}\{B|A\} + \varepsilon} = \frac{\mathbf{P}\{B\}}{\mathbf{P}\{B\} + \varepsilon} = \frac{f_0(i,j)}{f_0(i,j) + \varepsilon}.$$

Лемма доказана. ■

12.4 Устойчивость проективного преобразования к возмущениям конфигурации сопряженных точек (Г. В. Носовский, Е. С. Скрипка)

Для любого ненулевого вектора $\mathbf{x} \in \mathbb{R}^3$ будем обозначать $\widehat{\mathbf{x}}$ прямую, проходящую через 0 и содержащую вектор $\mathbf{x}$. Каждая такая прямая единственным образом определяет точку проективной плоскости $\mathbb{R}P^2$, поэтому мы можем считать, что $\widehat{\mathbf{x}} \in \mathbb{R}P^2$. Напомним, что *однородными координатами* точки $\widehat{\mathbf{x}} \in \mathbb{R}P^2$ называются координаты вектора $\mathbf{x}$ в $\mathbb{R}^3$. Однородные координаты определены с точностью до общего множителя $\lambda \neq 0$. Точку из $\mathbb{R}P^2$

с однородными координатами x_1, x_2, x_3 будем обозначать $(x_1 : x_2 : x_3)$. *Неоднородными или аффинными* координатами точки $\widehat{\mathbf{x}} \in \mathbb{R}P^2$ будем называть ее координаты в некоторой содержащей эту точку аффинной карте. Напомним, что аффинная карта в $\mathbb{R}P^2$ может быть задана с помощью произвольной плоскости π в $\mathbb{R}^3$, не проходящей через 0, с заданным на ней двумерным базисом. Точка $\widehat{\mathbf{x}} \in \mathbb{R}P^2$ принадлежит этой карте, если прямая $\widehat{\mathbf{x}}$ пересекает плоскость π. Соответствующие аффинные координаты точки $\widehat{\mathbf{x}} \in \mathbb{R}P^2$ определяются как координаты точки пересечения прямой $\widehat{\mathbf{x}}$ с плоскостью π в заданном на π базисе.

Пусть фиксирован ортонормированный базис $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ пространства $\mathbb{R}^3$. Фиксируем аффинную карту S_3, взяв в качестве плоскости π плоскость $x_3 = 1$ с базисом $\{\mathbf{e}_1, \mathbf{e}_2\}$. Без ограничения общности будем считать, что связанные искомым проективным преобразованием F области лежат в карте S_3. Это соответствует тому, что вектор $\mathbf{e}_3$ направляется каждый раз по оси объектива камеры. Аффинными координатами точки $\widehat{\mathbf{x}} = (x_1 : x_2 : x_3)$ в карте S_3 будет служить пара чисел $(x_1/x_3, x_2/x_3)$.

Приведенное выше определение точек общего положения в $\mathbb{R}P^n$ для случая $n = 2$ приобретает следующий вид: набор из 4-х точек проективной плоскости $\mathbb{R}P^2$ называется *системой точек общего положения*, если никакие 3 из них, рассматриваемые как прямые в $\mathbb{R}^3$, проходящие через 0, не лежат в одной плоскости. Соответственно формулируется в частном случае $\mathbb{R}P^2$ и приведенная выше теорема о восстановлении проективного преобразования по множествам сопряженных точек общего положения:

Теорема 12.2 (см., например, [43]). *Пусть $\{P_1, P_2, P_3, P_4\}$ и $\{Q_1, Q_2, Q_3, Q_4\}$ — две четверки точек общего положения проективного пространства $\mathbb{R}P^2$. Тогда существует единственное проективное преобразование, переводящее P_i в Q_i для любого $i = 1, 2, 3, 4$.*

Ниже нам понадобится явный вид системы линейных уравнений, определяющих это проективное преобразование.

Будем обозначать точки из $\mathbb{R}P^2$ и соответствующие им вектора из $\mathbb{R}^3$, (определенные неоднозначно) одинаковыми буквами. Пусть даны две четверки точек P, Q, R, T и P', Q', R', T' общего положения, лежащие в аффинной карте S_3 пространства $\mathbb{R}P^2$. Обозначим через $F = (f_{ij})$ $(1 \leqslant i, j \leqslant 3)$ матрицу линейного преобразования пространства $\mathbb{R}^3$, соответствующего проективному преобразованию, переводящему четверку P, Q, R, T в четверку P', Q', R', T'. Эта матрица определена с точностью до ненулевого скалярного множителя. Для того, чтобы определить ее однозначно, будем считать, что:

$$\begin{aligned} &P = (p_1, p_2, 1), \quad Q = (q_1, q_2, 1), \quad R = (r_1, r_2, 1), \quad T = (t_1, t_2, 1), \\ &P' = (p'_1, p'_2, 1), \quad Q' = (q'_1, q'_2, q'_3), \quad R' = (r'_1, r'_2, r'_3), \quad T' = (t'_1, t'_2, t'_3). \end{aligned} \tag{12.7}$$

И при этом:

$$F(P) = P', \quad F(Q) = Q', \quad F(R) = R', \quad F(T) = T'. \tag{12.8}$$

Другими словами, будем считать, что точки P, Q, R, T, рассматриваемые как вектора в $\mathbb{R}^3$, имеют третью координату, равную единице. Это возможно, поскольку по предположению все эти точки принадлежат аффинной карте S_3. Кроме того, умножением матрицы F на подходящий ненулевой множитель, в силу условия $P' \in S_3$, всегда можно добиться, чтобы третья координата вектора $P' = F(P) \in \mathbb{R}^3$ равнялась единице.

Обозначим:

$$a_p = p'_1, \quad b_p = p'_2, \qquad a_q = \frac{q'_1}{q'_3}, \quad b_q = \frac{q'_2}{q'_3},$$
$$a_r = \frac{r'_1}{r'_3}, \quad b_r = \frac{r'_2}{r'_3}, \qquad a_t = \frac{t'_1}{t'_3}, \quad b_t = \frac{t'_2}{t'_3}.$$

Четыре векторных равенства (12.8) задают систему из 12 скалярных линейных уравнений на 9 неизвестных коэффициентов f_{ij} матрицы F и на три неизвестные нам координаты q'_3, r'_3, t'_3. Запишем ее в виде

$$A\mathbf{x} = \mathbf{y}, \tag{12.9}$$

где

$$\mathbf{x} = (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33}, q'_3, r'_3, t'_3)^\top,$$
$$\mathbf{y} = (a_p, 0, 0, 0, b_p, 0, 0, 0, 1, 0, 0, 0)^\top,$$
$$A = \begin{pmatrix}
p_1 & p_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
q_1 & q_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -a_q & 0 & 0 \\
r_1 & r_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -a_r & 0 \\
t_1 & t_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -a_t \\
0 & 0 & 0 & p_1 & p_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & q_1 & q_2 & 1 & 0 & 0 & 0 & -b_q & 0 & 0 \\
0 & 0 & 0 & r_1 & r_2 & 1 & 0 & 0 & 0 & 0 & -b_r & 0 \\
0 & 0 & 0 & t_1 & t_2 & 1 & 0 & 0 & 0 & 0 & 0 & -b_t \\
0 & 0 & 0 & 0 & 0 & 1 & p_1 & p_2 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & q_1 & q_2 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & r_1 & r_2 & 1 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & t_1 & t_2 & 1 & 0 & 0 & -1
\end{pmatrix}.$$

Лемма 12.4. *Определитель матрицы A пропорционален произведению площадей треугольников $\triangle PQR, \triangle PRT, \triangle PQT$ и $\triangle Q'R'T'$. Имеет место равенство*:

$$|\det A| = 16 S_{PQR} S_{PRT} S_{PQT} S_{Q'R'T'}.$$

Доказательство. Прямой подсчет показывает, что определитель матрицы A равен:

$$\det A = (q_2p_1 - q_1p_2 - r_1q_2 - r_2p_1 + r_2q_1 + r_1p_2) \times$$
$$\times(-t_2p_1 + r_2p_1 + t_2r_1 - t_1r_2 + t_1p_2 - r_1p_2) \times (q_2p_1 - q_1p_2 - t_1q_2 - t_2p_1 + t_2q_1 + t_1p_2) \times$$
$$\times (b_ta_q - b_ta_r + b_qa_r + b_ra_t - b_ra_q - b_qa_t).$$

Рис. 12.3:

Заметим, что абсолютные величины сомножителей в правой части этого равенства равны удвоенным площадям следующих треугольников:

$$|q_2p_1 - q_1p_2 - r_1q_2 - r_2p_1 + r_2q_1 + r_1p_2| = 2S_{PQR},$$
$$|-t_2p_1 + r_2p_1 + t_2r_1 - t_1r_2 + t_1p_2 - r_1p_2| = 2S_{PRT},$$
$$|q_2p_1 - q_1p_2 - t_1q_2 - t_2p_1 + t_2q_1 + t_1p_2| = 2S_{PQT},$$
$$|b_ta_q - b_ta_r + b_qa_r + b_ra_t - b_ra_q - b_qa_t| = 2S_{Q'R'T'}.$$

Докажем, например, последнее равенство (см. рис. 12.3):

$$S_{R'Q'T'} = |S_{R'Q'DF} + S_{T'R'FE} - S_{T'Q'DE}| =$$
$$= \left|(Q'D + R'F)\,\frac{b_q - b_r}{2} + (T'E + R'F)\,\frac{b_r - b_t}{2} - (T'E + Q'D)\,\frac{b_q - b_t}{2}\right| =$$
$$= \left|\frac{Q'D}{2}\,(b_t - b_r) + \frac{R'F}{2}\,(b_q - b_t) + \frac{T'E}{2}\,(b_r - b_q))\right| =$$
$$= \left|\frac{a_q}{2}\,(b_t - b_r) + \frac{a_r}{2}\,(b_q - b_t) + \frac{a_t}{2}\,(b_r - b_q)\right| =$$
$$= \frac{1}{2}|b_ta_q - b_ta_r + b_qa_r + b_ra_t - b_ra_q - b_qa_t|.$$

Остальные три равенства доказываются аналогично. ■

Замечание. В утверждении леммы 12.4 точка P занимает выделенное положение в четверке точек P, Q, R, T. Это связано с тем, что при составлении системы уравнений (12.9) мы ввели дополнительное ограничение на однородные координаты точки $F(P)$ с целью фиксировать нормировку матрицы F линейного преобразования, соответствующего искомому проективному преобразованию, см. (12.7), (12.8). В качестве точки P может быть выбрана любая из точек рассматриваемой четверки общего положения. Однако при этом нормировка матрицы F, являющейся решением системы (12.9), будет, вообще говоря, зависеть от выбора этой точки.

Из доказанной леммы следует, что при вычислении проективного преобразования надо стремиться выбирать такие конфигурации сопряженных

точек, чтобы произведение площадей построенных на них треугольников было наибольшим. В этом случае будет достигаться максимальная устойчивость результата к погрешностям исходных данных.

12.5 Вычисление матрицы проективного преобразования и оценка ее устойчивости (Е. С. Скрипка)

12.5.1 Прямой алгоритм вычисления проективного преобразования

Опишем более подробно *прямой* алгоритм вычисления проективного преобразования по двум сопряженным 4-х точечным множествам прообразов и образов. Он состоит в непосредственном решении приведенной выше системы линейных уравнений (12.8) методами вычислительной линейной алгебры, откуда его название — "прямой алгоритм". Нашей задачей будет получение оценки погрешности определения проективного преобразования с помощью прямого алгоритма — в зависимости от точности задания сопряженных точек, а также от взаимного расположения точек-прообразов.

Как и выше, будем рассматривать трехмерное евклидово пространство $\mathbb{R}^3$ и соответствующую ему проективную плоскость $\mathbb{R}P^2$.

Пусть даны две четверки точек P, Q, R, T и P', Q', R', T' на проективной плоскости $\mathbb{R}P^2$. Требуется вычислить проективное преобразование $\widehat{\mathcal{A}}$, переводящее первую четверку точек во вторую. Будем считать, что нам известны неоднородные координаты всех сопряженных точек, которые мы обозначим следующим образом:

$$P = \left\{\frac{p_1}{p_3}, \frac{p_2}{p_3}\right\}, \quad Q = \left\{\frac{q_1}{q_3}, \frac{q_2}{q_3}\right\}, \quad R = \left\{\frac{r_1}{r_3}, \frac{r_2}{r_3}\right\}, \quad T = \left\{\frac{t_1}{t_3}, \frac{t_2}{t_3}\right\},$$
$$P' = \{a_p, b_p\}, \quad Q' = \{a_q, b_q\}, \quad R' = \{a_r, b_r\} \quad T' = \{a_t, b_t\}.$$

Без ограничения общности будем считать, что $p_3 = q_3 = r_3 = t_3 = p'_3 = 1$. Тогда однородные координаты этих точек будут следующие:

$$\vec{P} = (p_1, p_2, 1), \quad \vec{Q} = (q_1, q_2, 1), \quad \vec{R} = (r_1, r_2, 1), \quad \vec{T} = (t_1, t_2, 1),$$
$$\vec{P'} = (p'_1, p'_2, 1), \quad \vec{Q'} = (q'_1, q'_2, q'_3), \quad \vec{R'} = (r'_1, r'_2, r'_3), \quad \vec{T'} = (t'_1, t'_2, t'_3),$$
$$a_p = p'_1, \quad b_p = p'_2, \qquad a_q = \frac{q'_1}{q'_3}, \quad b_q = \frac{q'_2}{q'_3},$$
$$a_r = \frac{r'_1}{r'_3}, \quad b_r = \frac{r'_2}{r'_3}, \quad a_t = \frac{t'_1}{t'_3}, \quad b_t = \frac{t'_2}{t'_3}.$$

Задача состоит в вычислении проективного преобразования $\widehat{\mathcal{A}}$ такого, что $\widehat{\mathcal{A}}(P) = P'$, $\widehat{\mathcal{A}}(Q) = Q'$, $\widehat{\mathcal{A}}(R) = R'$, $\widehat{\mathcal{A}}(T) = T'$.

Известно, что проективному преобразованию $\widehat{\mathcal{A}}$ проективной плоскости $\mathbb{R}P^2$ соответствует определенное с точностью до ненулевого множителя линейное преобразование $\mathcal{A}$ трехмерного евклидова пространства $\mathbb{R}^3$. Будем считать, что в $\mathbb{R}^3$ заранее фиксированы некоторые декартовы координаты. Тогда преобразованию $\mathcal{A}$ однозначно ставится в соответствие матрица преобразования $A = (a_{ij})$, $i, j \in \{1, 2, 3\}$: $A(\vec{P}) = \vec{P}'$, $A(\vec{Q}) = \vec{Q}'$, $A(\vec{R}) = \vec{R}'$, $A(\vec{T}) = \vec{T}'$. Но, поскольку преобразование $\mathcal{A}$ определено в нашем случае лишь с точностью до множителя, то, естественно, и матрица A будет определена лишь с точностью до произвольного ненулевого множителя. Чтобы исключить возникающую неоднозначность в определении матрицы A, будем считать выполненным условие нормировки $p'_3 = 1$. Тогда для элементов матрицы $A = (a_{ij})$, $i, j \in \{1, 2, 3\}$, будут справедливы следующие системы линейных уравнений:

$$\begin{cases} a_{11}p_1 + a_{12}p_2 + a_{13} = p'_1 = a_p, \\ a_{21}p_1 + a_{22}p_2 + a_{23} = p'_2 = b_p, \\ a_{31}p_1 + a_{32}p_2 + a_{33} = 1; \end{cases} \qquad \begin{cases} a_{11}q_1 + a_{12}q_2 + a_{13} = q'_1 = a_q q'_3, \\ a_{21}q_1 + a_{22}q_2 + a_{23} = p'_2 = b_q q'_3, \\ a_{31}q_1 + a_{32}q_2 + a_{33} = q'_3; \end{cases}$$

$$\begin{cases} a_{11}r_1 + a_{12}r_2 + a_{13} = r'_1 = a_r r'_3, \\ a_{21}r_1 + a_{22}r_2 + a_{23} = r'_2 = b_r r'_3, \\ a_{31}r_1 + a_{32}r_2 + a_{33} = r'_3; \end{cases} \qquad \begin{cases} a_{11}t_1 + a_{12}t_2 + a_{13} = t'_1 = a_t t'_3, \\ a_{21}t_1 + a_{22}t_2 + a_{23} = t'_2 = b_t t'_3, \\ a_{31}t_1 + a_{32}t_2 + a_{33} = t'_3. \end{cases}$$

Мы приходим к следующей системе уравнений (аналогичной системе (12.9)), в которой неизвестными являются коэффициенты матрицы A и величины q'_3, r'_3, t'_3:

$$\underbrace{\begin{pmatrix} p_1 & p_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ q_1 & q_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -a_q & 0 & 0 \\ r_1 & r_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -a_r & 0 \\ t_1 & t_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -a_t \\ 0 & 0 & 0 & p_1 & p_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_1 & q_2 & 1 & 0 & 0 & 0 & -b_q & 0 & 0 \\ 0 & 0 & 0 & r_1 & r_2 & 1 & 0 & 0 & 0 & 0 & -b_r & 0 \\ 0 & 0 & 0 & t_1 & t_2 & 1 & 0 & 0 & 0 & 0 & 0 & -b_t \\ 0 & 0 & 0 & 0 & 0 & 0 & p_1 & p_2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_1 & q_2 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & r_1 & r_2 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & t_1 & t_2 & 1 & 0 & 0 & -1 \end{pmatrix}}_{C} \underbrace{\begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \\ a_{31} \\ a_{32} \\ a_{33} \\ q'_3 \\ r'_3 \\ t'_3 \end{pmatrix}}_{D} = \underbrace{\begin{pmatrix} a_p \\ 0 \\ 0 \\ 0 \\ b_p \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}}_{B}.$$

Прямой алгоритм определения проективного преобразования состоит в решении этой системы уравнений численными методами линейной алгебры. Искомое проективное преобразование будет представлено в виде матрицы A. Рассмотрим вопрос об оценке устойчивости матрицы A к возмущениям начальных данных.

12.5.2 Оценка погрешности матрицы проективного преобразования при возмущенных начальных данных

Будем считать, что все неоднородные координаты точек P, Q, R, T и P', Q', R', T' на плоскости известны с некоторой погрешностью, ограниченной по модулю заданной константой δ (определяемой, в частности, разрешением имеющихся в нашем распоряжении изображений), а также заданной погрешностью вычислений. Другими словами, будем считать, что нам вместо координат точек P, Q, R, T и P', Q', R', T' заданы координаты возмущенных точек $\widetilde{P}$, $\widetilde{Q}$, $\widetilde{R}$, $\widetilde{T}$ и $\widetilde{P'}$, $\widetilde{Q'}$, $\widetilde{R'}$, $\widetilde{T'}$:

$$\begin{gathered}
\widetilde{P} = (p_1 + \Delta p_1, p_2 + \Delta p_2), \quad \widetilde{Q} = (q_1 + \Delta q_1, q_2 + \Delta q_2), \\
\widetilde{R} = (r_1 + \Delta r_1, r_2 + \Delta r_2), \quad \widetilde{T} = (t_1 + \Delta t_1, t_2 + \Delta t_2), \\
|\Delta p_i| < \delta, \quad |\Delta q_i| < \delta, \quad |\Delta r_i| < \delta, \quad |\Delta t_i| < \delta, \qquad i = 1, 2; \\
\widetilde{P'} = (a_p + \Delta a_p, b_p + \Delta b_p), \qquad \widetilde{Q'} = (a_q + \Delta a_q, b_q + \Delta b_q), \\
\widetilde{R'} = (a_r + \Delta a_r, b_r + \Delta b_r), \qquad \widetilde{T'} = (a_t + \Delta a_t, b_t + \Delta b_t),
\end{gathered}$$

удовлетворяющие условиям:

$$|\Delta a_j| < \delta, \qquad |\Delta b_j| < \delta, \quad j \in \{p, q, r, t\}.$$

Поскольку, как правило, приходится иметь дело с изображениями заранее ограниченного размера, будем считать, что координаты заданных точек $\{\widetilde{P}, \widetilde{Q}, \widetilde{R}, \widetilde{T}\}$ и $\{\widetilde{P'}, \widetilde{Q'}, \widetilde{R'}, \widetilde{T'}\}$ ограничены по модулю заданным числом $N/2$ (что соответствует квадратному экрану со стороной, равной N и началом координат, помещенным в центр экрана):

$$\begin{gathered}
|a_j| < N/2, \qquad |b_j| < N/2, \qquad j \in \{p, q, r, t\}, \\
|p_i| < N/2, \quad |q_i| < N/2, \quad |r_i| < N/2, \quad |t_i| < N/2, \quad i = 1, 2.
\end{gathered}$$

Имеем: $(C+\Delta C)(D+\Delta D) = B+\Delta B$. Следовательно, в предположении невырожденности матрицы C, справедливо равенство $\Delta D = (C+\Delta C)^{-1}(\Delta B - \Delta C D) = (C + \Delta C)^{-1}(\Delta B - \Delta C C^{-1} B)$, где

$$\Delta B = \begin{pmatrix} \Delta a_p \\ 0 \\ 0 \\ 0 \\ \Delta b_p \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \Delta D = \begin{pmatrix} \Delta a_{11} \\ \Delta a_{12} \\ \Delta a_{13} \\ \Delta a_{21} \\ \Delta a_{22} \\ \Delta a_{23} \\ \Delta a_{31} \\ \Delta a_{33} \\ \Delta a_{33} \\ \Delta q'_3 \\ \Delta r'_3 \\ \Delta t'_3 \end{pmatrix}, \qquad \Delta C =$$

$$= \begin{pmatrix} \Delta p_1 & \Delta p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \Delta q_1 & \Delta q_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\Delta a_q & 0 & 0 \\ \Delta r_1 & \Delta r_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\Delta a_r & 0 \\ \Delta t_1 & \Delta t_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\Delta a_t \\ 0 & 0 & 0 & \Delta p_1 & \Delta p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Delta q_1 & \Delta q_2 & 0 & 0 & 0 & 0 & -\Delta b_q & 0 & 0 \\ 0 & 0 & 0 & \Delta r_1 & \Delta r_2 & 0 & 0 & 0 & 0 & 0 & -\Delta b_r & 0 \\ 0 & 0 & 0 & \Delta r_1 & \Delta r_2 & 0 & 0 & 0 & 0 & 0 & 0 & -\Delta b_t \\ 0 & 0 & 0 & 0 & 0 & 0 & \Delta p_1 & \Delta p_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Delta q_1 & \Delta q_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Delta r_1 & \Delta r_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Delta t_1 & \Delta t_2 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Пусть $\|\cdot\|$ — некоторая векторная норма. Соответствующую ей (в операторном смысле) матричную норму будем также обозначать через $\|\cdot\|$. Тогда справедлива следующая оценка (здесь использовано, что $D = C^{-1}B$):

$$\begin{aligned} \|\Delta D\| = \|(C+\Delta C)^{-1}(\Delta B - \Delta C D)\| \leqslant \\ \leqslant \|(C+\Delta C)^{-1}\Delta B\| + \|(C+\Delta C)^{-1}\Delta C D)\| = \\ = \|(C+\Delta C)^{-1}\Delta B\| + \|(C+\Delta C)^{-1}\Delta C C^{-1} B)\|. \end{aligned}$$

Нам будет удобно выбрать векторную норму следующим образом:

Определение 12.8. Для вектора $B = (b_1, b_2, \ldots, b_n)^{\top}$ положим по определению $\|B\| = \max\limits_{1 \leqslant i \leqslant n} |b_i|$. Соответственно, для матрицы $C = (c_{ij})_{n \times n}$ положим $\|C\| = \max\limits_{1 \leqslant i \leqslant n} \sum\limits_{j=1}^{n} |c_{ij}|$.

Из курса матричной алгебры известно, что это — действительно нормы в пространствах векторов и матриц соответственно, согласованные в операторном смысле.

Теорема 12.3. *Пусть даны возмущенные координаты двух четверок сопряженных точек $\{P, Q, R, T\}$ и $\{P', Q', R', T'\}$ на проективной плоскости $\mathbb{R}P^2$, принадлежащих одной и той же аффинной карте S_3. Предположим, что погрешность каждой из возмущенных координат по модулю не превосходит заданного числа $\delta > 0$, а сами эти координаты (как возмущенные так и невозмущенные) ограничены сверху по абсолютной величине заданным числом $N/2$. Пусть A — матрица проективного преобразования, переводящего возмущенную четверку точек $\{\widetilde{P}, \widetilde{Q}, \widetilde{R}, \widetilde{T}\}$ в возмущенную четверку $\{\widetilde{P}', \widetilde{Q}', \widetilde{R}', \widetilde{T}'\}$. При этом, для исключения неоднозначности в определении матрицы A будем считать выполненным условие нормировки $p'_3 = 1$. Тогда норма (в соответствии с определением 12.8) отклонения матрицы A от матрицы, проективного преобразования, связывающего невозмущенные четверки сопряженных точек, в предположении, что эта матрица удовлетворяет тому же условию нормировки, оценивается сверху следующей величиной, зависящей лишь от N, δ и произведения площадей треугольников*

$\Delta PQR, \Delta PRT, \Delta PQT, \Delta Q'R'T'$:

$$\|\Delta A\| \leqslant \frac{20,5\delta N^{16}}{|\det C| \cdot |\det(C+\Delta C)|} + \frac{1,8\delta N^{8}}{|\det(C+\Delta C)|} + o(N).$$

Замечание. Согласно лемме 12.4, $|\det C| = 16 S_{PQR} S_{PRT} S_{PQT} S_{Q'R'T'}$, а следовательно с ростом N, при подходящем размещении сопряженных точек по всему экрану, величина $|\det C|$ будет расти как N^8. Поэтому правая часть неравенства, составляющего утверждение теоремы, при разумном подборе конфигурации сопряженных точек не растет с ростом линейного размера экрана N. В частности, она инвариантна относительно выбора масштаба измерений.

Доказательство требует достаточно громоздких вычислений и поэтому мы его опускаем.

Сформулированная теорема позволяет оптимальным образом выбирать конфигурации сопряженных точек при компьютерной склейке проективного преобразованных изображений.

12.6 Математическая модель камеры и пары камер

12.6.1 Определение финитной камеры

С точки зрения математического моделирования, камера (т.е. фотоаппарат, кинокамера и т.п.) в каждый момент съемки представляет собой центральную проекцию пространства $\mathbb{R}^3$ из определенной точки пространства — *центра проекции*, который мы также будем называть *центром камеры*, — на *плоскость проектирования*. Плоскость проектирования — а в некоторых случаях ее определенную часть, — будем называть *экраном камеры*. *Осью камеры* при этом называется прямая, проходящая через центр камеры и перпендикулярная плоскости проектирования. Расстояние от центра проекции до плоскости проектирования будем называть *фокусным расстоянием* камеры, см. рис. 12.4.

Опишем это отображение в координатах. Будем считать, что в трехмерном пространстве задана некоторая декартова система координат и на плоскости проектирования также задана своя декартова система координат. Рассмотрим сначала случай, когда эти системы координат согласованы таким образом, что координаты на плоскости совпадают с первыми двумя пространственными координатами, см. рис. 12.4. Из подобия треугольников получаем, что в этом случае отображение, задаваемое камерой, выглядит следующим образом:

$$(X, Y, Z) \to \left(\frac{fX}{Z}, \frac{fY}{Z}\right),$$

Рис. 12.4: Математическая модель камеры

где через f обозначено фокусное расстояние камеры. То же отображение, записанное в однородных координатах, имеет вид:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \to \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \mathrm{diag}(f, f, 1)[I|0] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix},$$

где через $\mathrm{diag}(f, f, 1)$ обозначена диагональная матрица с элементами $f, f, 1$ на главной диагонали, а через $[I|0]$ — прямоугольная матрица 3×4, полученная путем добавления к единичной (3×3)-матрице нулевого столбца справа.

Рассмотрим теперь случай, когда начало координат на плоскости проектирования сдвинуто относительно оси камеры на вектор (p_x, p_y). Тогда отображение, задаваемое камерой, записывается в аналогичном виде, но с другой матрицей. В этом случае матрица будет иметь вид:

$$\begin{pmatrix} f & 0 & -p_x & 0 \\ 0 & f & -p_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Действительно, легко проверить, что при $Z \neq 0$ выполняется соотношение

$$\begin{pmatrix} f & 0 & -p_x & 0 \\ 0 & f & -p_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} fX - p_x Z \\ fY - p_y Z \\ Z \end{pmatrix} = Z \begin{pmatrix} \dfrac{fX}{Z} - p_x \\ \dfrac{fY}{Z} - p_y \\ 1 \end{pmatrix}.$$

Рассмотрим более общий случай. Пусть камера (т.е. центр проекции и плоскость проектирования) расположена относительно заданной системы

декартовых координат в трехмерном пространстве произвольным образом. Обозначим через $\mathbf{t} = (t_1, t_2, t_3)^\top$ пространственные координаты центра камеры. Тогда существует ортогональное преобразование R трехмерного пространства, такое что

$$\overline{X}_{cam} = R(\overline{X} - \mathbf{t}) = \begin{pmatrix} R & -R\mathbf{t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix},$$

где через $\overline{X}_{cam}$ обозначен образ точки $\overline{X} \in \mathbb{R}^3$ с координатами (X, Y, Z) при отображении, задаваемом камерой. Предполагается, что точка $\overline{X}_{cam} \in \mathbb{R}P^2$ задана в однородных координатах.

Обозначим

$$K = \begin{pmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix}$$

и умножим последнее равенство на матрицу K слева. Получим

$$K \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \overline{X}_{cam} = K \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & -R\mathbf{t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} =$$

$$= K[R| - R\mathbf{t}]\overline{X} = KR[I| - \mathbf{t}]\overline{X},$$

где через $[R|-R\mathbf{t}]$ и $[I|-\mathbf{t}]$, как и выше обозначены прямоугольные 3×4 матрицы, полученные добавлением правого столбца к соответствующим квадратным матрицам. А именно — в первом случае к матрице R добавляется вектор-столбец $-R\mathbf{t}$, а во втором — к единичной матрице I добавляется вектор-столбец $-\mathbf{t}$.

Пока мы считали, что масштаб по осям в пространственной и плоской (на плоскости проектирования) системах координат совпадает. Если это не так, то в качестве матрицы K надо брать матрицу

$$K = \begin{pmatrix} \alpha_x & 0 & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{pmatrix},$$

где α_x и α_y — коэффициенты, отвечающие за изменение масштаба по осям OX и OY соответственно.

Если к тому же система координат на плоскости проектирования не является декартовой (оси координат не ортогональны), то в качестве матрицы K будет иметь вид

$$K = \begin{pmatrix} \alpha_x & s & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{pmatrix}.$$

Любую 3×4 матрицу $P = KR[I| - \mathbf{t}]$, где K имеет указанный выше вид, будем называть *матрицей финитной проективной камеры* или просто *финитной камерой*. Мы будем рассматривать такую матрицу как математическую модель некоторой реальной камеры.

Финитная проективная камера имеет 11 независимых параметров (3 для R — ортогонального преобразования, 5 для K, 3 для $\mathbf{t}$). Такое же число степеней свободы имеет произвольная однородная матрица размера 3×4.

Заметим, что квадратная матрица KR, участвующая в определении финитной камеры, невырождена. Обратно, любая 3×4 матрица P вида $P = M[I| - \mathbf{t}]$, где M — невырожденная (3×3)-матрица, является матрицей некоторой финитной камеры. Это следует из того, что невырожденная матрица M может быть разложена в произведение KR, где K — верхнетреугольная, а R — ортогональная матрица.

12.6.2 Восстановление характеристик финитной камеры по ее матрице

Пусть дана матрица камеры P. Как узнать геометрические характеристики этой камеры?

Поскольку, согласно приведенным выше формулам, матрица P действует на трехмерных векторах, представленных в виде четырехмерных векторов, последняя компонента которых равна 1, то удобно считать что P — это линейный оператор из $\mathbb{R}P^3$ в $\mathbb{R}P^2$. Другими словами, исходное трехмерное пространство $\mathbb{R}^3$, в котором расположена камера, мы будем отождествлять с фиксированной аффинной картой в $\mathbb{R}P^3$, точки которого могут быть заданы либо своими однородными координатами (при этом считаем что, в аффинной карте, моделирующей исходное $\mathbb{R}^3$, четвертая координата отлична от нуля), либо — неоднородными. В последнем случае неоднородные координаты получаются из однородных путем такой нормировки (умножения на общий множитель), что четвертая координата становится равной 1.

Центр камеры. Из чисто геометрических соображений легко усмотреть, что нулевое подпространство оператора P в исходном $\mathbb{R}^3$ одномерно. Значит, существует такая точка $C \in \mathbb{R}^3$, для которой справедливо равенство $PC = 0$. При этом, если C задана в однородных координатах (четырьмя числами), то ее четвертая координата отлична от нуля. Рассмотрим отрезок, соединяющий точку C с любой другой точкой $A \in \mathbb{R}^3$:

$$X(\lambda) = \lambda A + (1 - \lambda)C \qquad (0 \leqslant \lambda \leqslant 1).$$

Легко видеть, что при отображении P этот отрезок переходит в точку, которая в однородных координатах $\mathbb{R}P^3$ записывается в виде прямой, проходящей через начало координат:

$$PX(\lambda) = \lambda PA + (1 - \lambda)PC = \lambda PA.$$

Поскольку это верно для любой точки $A \in \mathbb{R}^3$, то C — это и есть центр проекции (центр камеры). Таким образом, центр камеры восстанавливается

Рис. 12.5: Полюса и полярные линии двух камер

по матрице P как ее нулевое подпространство (в однородных координатах $\mathbb{R}P^3$).

Плоскость, параллельная плоскости проектирования и проходящая через центр камеры. Из геометрических соображений легко усмотреть, что эта плоскость состоит из таких точек $X \in \mathbb{R}^3$, которые при отображении P переходят в бесконечно удаленные точки плоскости проектирования (рассматриваемой как аффинная карта в $\mathbb{R}P^2$). Последнее условие выражается формулой: $PX = (x, y, 0)^\top$. Таким образом, условие принадлежности точки $X \in \mathbb{R}^3$ искомой плоскости равносильно равенству нулю скалярного произведения векторов X и P^3, где P^3 — вектор, стоящий в третьей строке матрицы P. Это означает, что вектор P^3 ортогонален искомой плоскости и, следовательно, полностью ее определяет (поскольку плоскость по условию проходит через начало координат в $\mathbb{R}^3$).

12.6.3 Геометрия двух камер. Фундаментальная матрица

Пусть даны две камеры P и P' с центрами в точках C и C' соответственно. Предположим, что некоторая точка $X \in \mathbb{R}^3$ проецируется на экран первой камеры в точку $\mathbf{x}$, а на экран второй камеры — в точку $\mathbf{x}'$. Какова связь между координатами точек $\mathbf{x}$ и $\mathbf{x}'$? Чтобы ответить на этот вопрос, рассмотрим следующее отображение F. Зафиксируем точку $\mathbf{x}$ на первом экране. Множество возможных прообразов этой точки при центральном проецировании — это прямая, выходящая из центра первой камеры и проходящая через $\mathbf{x}$. Образ этой прямой на втором экране — прямая, которую мы обозначим через l', см. рис. 12.5.

Введем несколько новых понятий, связанных с парой камер P и P'. Прямую l' будем называть *полярной линией*, соответствующей точке $\mathbf{x}$. Возникающее при этом отображение $F\colon \mathbf{x} \to l'$ называется *фундаментальным*

Рис. 12.6: Фундаментальное отображение

отображением пары камер P и P'. Пересечения прямой CC', соединяющей центры камер, с первым экраном (точка $\mathbf{e}$) и со вторым экраном (точка $\mathbf{e}'$), называются *полюсами* данной пары камер.

Легко проверить, что все полярные линии на втором экране пересекаются в полюсе $\mathbf{e}'$.

Как задать фундаментальное отображение в координатах? Поступим следующим образом. Фиксируем некоторую плоскость π, которая не пересекает прямую CC', проходящую через центры камер (и, в частности, не содержит сами точки C и C'), рис. 12.6. Пусть точка $\mathbf{x}$ принадлежит первому экрану и луч $C\mathbf{x}$, идущий из центра первой камеры через эту точку, пересекает плоскость π в точке X. Спроецируем точку X при помощи второй камеры. Получим некоторую новую точку $\mathbf{x}'$ — уже на втором экране. Легко показать, что возникающее при этом отображение из первого экрана во второй $H_\pi : \mathbf{x} \to \mathbf{x}'$ будет проективным преобразованием.

Определим теперь понятие *координат прямой* на экране. Заметим, что каждой прямой на экране камеры однозначно соответствует некоторая плоскость в трехмерном пространстве, проходящая через эту прямую и центр камеры. Эта плоскость, в свою очередь, однозначно задается перпендикулярной ей прямой, проходящей через центр камеры — т.е. точкой на проективной плоскости $\mathbb{R}P^2$ (при этом экран камеры, напомним, является аффинной картой этой проективной плоскости). Итак, прямая на экране — или, что то же самое, на аффинной карте проективной плоскости, — может быть однозначно связана с некоторой точкой этой проективной плоскости. Однородные координаты этой точки мы и будем называть *координатами данной прямой на экране камеры*.

Вернемся к заданию фундаментального преобразования, определяемого парой камер. Вновь обратимся к рис. 12.6. Чтобы получить координаты прямой l', соединим точки $\mathbf{e}'$ и $\mathbf{x}'$, рис. 12.6. В координатах эта операция записывается так: $l' = \mathbf{e}' \times \mathbf{x}' = [\mathbf{e}']_\times \mathbf{x}'$, где через $\times$ обозначена операция

векторного произведения в $\mathbb{R}^3$, а через $[\mathbf{a}]_\times$ мы будем обозначать матрицу линейного оператора в $\mathbb{R}^3$, действующего по формуле: $[\mathbf{a}]_\times(\mathbf{x}) = \mathbf{a} \times \mathbf{x}$. Итак, справедлива следующая формула, выражающая искомое фундаментальное отображение F через определенное выше проективное преобразование H_π:

$$l' = F\mathbf{x} = [\mathbf{e}']_\times H_\pi \mathbf{x}.$$

Поскольку проективное преобразование $\mathbb{R}P^2 \to \mathbb{R}P^2$ определяется однородной (3×3)-матрицей, то, следовательно, фундаментальное отображение F можно задать однородной матрицей 3×3, которую мы также будем обозначать через F. Матрица F имеет ранг 2, поскольку $\operatorname{rank}(H_\pi) = 3$, а $\operatorname{rank}([\mathbf{e}']_\times) = 2$.

Выражение фундаментального отображения через матрицы камер. Вычислим матрицу фундаментального отображения F через матрицы камер P и P'. Для этого рассмотрим уравнение $PX = \mathbf{x}$ относительно X. Решениями этого уравнения в $\mathbb{R}^3$ является прямая, проходящая через центр камеры C и точку $\mathbf{x}$. Можно показать, что эта прямая может быть записана в виде:

$$X(\lambda) = P^+\mathbf{x} + \lambda C,$$

где P^+ — правая псевдообратная матрица к P (т.е. матрица, удовлетворяющая соотношению $PP^+ = I$), а C — нуль-вектор матрицы P (т.е. $PC = 0$). Мы обозначаем нуль-вектор матрицы P той же буквой, что и центр камеры, поскольку, как было показано выше, однородные координаты центра камеры, рассматриваемой как точка в $\mathbb{R}P^3$, являются нуль-вектором матрицы камеры P (мы считаем здесь, что $C \neq 0$).

Возьмем две несовпадающие точки на прямой $X(\lambda)$, а именно: $P^+\mathbf{x}$ и $P^+\mathbf{x} + C$. Вторая камера спроецирует их в точки с координатами $P'P^+\mathbf{x}$ и $P'(P^+\mathbf{x} + C)$ на втором экране. Поскольку плоскость π по условию была выбрана так, чтобы она не пересекала прямую CC', то эти точки на втором экране будут также различны. Следовательно, они определяют прямую $l' = F\mathbf{x}$ как прямую, которая через них проходит. В качестве координат этой прямой можно взять координаты векторного произведения векторов $P'P^+\mathbf{x}$ и $P'(P^+\mathbf{x} + C)$, которое ортогонально им обоим и, следовательно, ортогонально плоскости, проходящей через прямую l' и центр второй камеры. Таким образом, справедливо соотношение

$$l' = (P'C) \times P'(P^+\mathbf{x} + C) = (P'C) \times P'P^+\mathbf{x}.$$

Но поскольку C — это центр первой камеры, то $P'C = \mathbf{e}'$, и, следовательно, фундаментальная матрица может быть записана в виде:

$$F = [\mathbf{e}']_\times P'P^+.$$

Пример 12. Пусть система координат в пространстве $\mathbb{R}^3$ выбрана так, что первая камера имеет матрицу $P = K[I|0]$, а вторая — матрицу $P' = K'[R|\mathbf{t}]$. Тогда

$$P^+ = \begin{pmatrix} K^{-1} \\ 0^\top \end{pmatrix}, \qquad C = (0, 0, 0, 1)^\top,$$

и, следовательно,

$$F = [P'C]_\times P'P^+ = [K'\mathbf{t}]_\times K'RK^{-1}.$$

12.6.4 Свойства фундаментальной матрицы

Теорема 12.4. 1) *Матрица F является фундаментальной матрицей для пары камер (P, P') тогда и только тогда, когда для любых сопряженных точек $\mathbf{x} \leftrightarrow \mathbf{x}'$ выполнено соотношение*:

$$\mathbf{x}'^\top F\mathbf{x} = 0.$$

2) *Если F — фундаментальная матрица, соответствующая паре камер (P, P'), то $F^\top$ — фундаментальная матрица и для пары (P', P).*

3) *Для любой точки $\mathbf{x}$ на первом экране, кроме полюса $\mathbf{e}$, соответствующая полярная линия $l' = F\mathbf{x}$ проходит через второй полюс $\mathbf{e}'$.*

4) *Полюс $\mathbf{e}'$ (находящийся на экране второй камеры) лежит в левом нулевом пространстве матрицы F, т.е. имеет место равенство $\mathbf{e}'^\top F = 0$. Аналогично, полюс $\mathbf{e}$ лежит в правом нулевом пространстве матрицы F, т.е. $F\mathbf{e} = 0$.*

Доказательство. Доказательство утверждений 1) – 3) получается прямой проверкой. Утверждение 4) вытекает из утверждения 3), поскольку, согласно 3), $\mathbf{e}'^\top(F\mathbf{x}) = (\mathbf{e}'^\top F)\mathbf{x} = 0$ для любого $\mathbf{x}$ на первом экране, а следовательно $\mathbf{e}'^\top F = 0$. Аналогично $F\mathbf{e} = 0$. ■

12.6.5 Восстановление пары камер по фундаментальной матрице

Предположим, что пара камер, по которой построена матрица фундаментального отображения F, нам неизвестна. Известно лишь само отображение F. Зададимся вопросом — какие характеристики этой пары камер мы можем восстановить, зная матрицу F? Другими словами, наша цель — обратить (насколько это возможно) построенное выше отображение

$$(P, P') \longrightarrow F.$$

Будем называть любую пару камер, порождающую данную матрицу фундаментального отображения F *восстановлением* для F. Докажем ряд предложений.

Лемма 12.5. *Пусть H — проективное преобразование пространства $\mathbb{R}P^3$, в которое вложено исходное пространство $\mathbb{R}^3$ как аффинная карта, выделяемая условием неравенства нулю четвертой однородной координаты. Тогда фундаментальная матрица для пары камер (P, P') совпадает с фундаментальной матрицей для преобразованной пары камер (PH, PH').*

Доказательство. Для любых сопряженных точек $\mathbf{x}$ и $\mathbf{x}'$ на первом и втором экранах соответственно, имеем:

$$\begin{gathered}\mathbf{x}'^{\top} F\mathbf{x} = 0 \qquad \Leftrightarrow \qquad (P'X)^{\top} FPX = 0 \quad \forall X \in \mathbb{R}P^3 \Leftrightarrow \\ \Leftrightarrow \qquad \big((P'H)(H^{-1}X)\big)^{\top} F(PH)(H^{-1}X) = 0 \Leftrightarrow \\ \Leftrightarrow (P'HY)^{\top} F(PHY) = 0 \quad \forall Y \in \mathbb{R}P^3.\end{gathered}$$

Последнее соотношение означает, что F — фундаментальная матрица для преобразованной пары камер $(PH, P'H)$. Лемма доказана. ■

Покажем, что верно и обратное, т.е. любые два восстановления, построенные для одной и той же фундаментальной матрицы F, связаны некоторым проективным преобразованием в $\mathbb{R}P^3$. С этой целью докажем ряд вспомогательных утверждений.

Лемма 12.6. *Для любой финитной камеры P существует проективное преобразование H такое, что $PH = [I|0]$.*

Доказательство. Дополним матрицу P размера (3×4) до невырожденной матрицы P^* размера (4×4). Это можно сделать, поскольку $\operatorname{rank}(P) = 3$. Тогда в качестве матрицы искомого преобразования H можно взять матрицу $H = (P^*)^{-1}$. Лемма доказана. ■

Лемма 12.7. *Фундаментальная матрица для пары камер $P = [I|0]$, $P' = [M|\mathbf{m}]$ равна $[\mathbf{m}]_\times M$.*

Доказательство. Имеем: $P^+ = \begin{pmatrix} I \\ 0^{\top} \end{pmatrix}$, $C = (0, 0, 0, 1)^{\top}$. Следовательно,

$$F = [P'C]_\times P'P^+ = [\mathbf{m}]_\times M.$$

Лемма доказана. ■

Следующая теорема является основной в данном разделе. Она утверждает, что любые два восстановления одной и той же матрицы фундаментального преобразования связаны некоторым проективным преобразованием. Другими словами, *проективное восстановление матрицы фундаментального преобразования единственно.*

Теорема 12.5 ([51]). *Пусть F — матрица фундаментального преобразования как для пары камер (P, P'), так и для пары камер $(\widetilde{P}, \widetilde{P}')$. Тогда существует проективное преобразование H, переводящее одну пару камер в другую: $\widetilde{P} = PH$ и $\widetilde{P}' = P'H$.*

Доказательство. По лемме 12.5 существуют проективные преобразования H_1 и H_2 такие, что $PH_1 = [I|0]$ и $\widetilde{P}H_2 = [I|0]$. Пусть $P'H_1 = [A|\mathbf{a}]$ и $\widetilde{P}'H_2 = [\widetilde{A}|\widetilde{\mathbf{a}}]$. Тогда по лемме 12.7 фундаментальная матрица F имеет вид

$$F = [\mathbf{a}]_\times A = [\widetilde{\mathbf{a}}]_\times \widetilde{A}.$$

Покажем, что $\widetilde{\mathbf{a}} = k\mathbf{a}$ и $\widetilde{A} = \dfrac{1}{k}(A + \mathbf{a}\mathbf{v}^\top)$ для некоторого вещественного числа k и некоторой вектор-строки $\mathbf{v} \in \mathbb{R}^3$. Из равенств $F^\top \mathbf{a} = 0$, $F^\top \widetilde{\mathbf{a}} = 0$ и $\operatorname{rank} F = 2$ следует, что $\widetilde{\mathbf{a}} = k\mathbf{a}$. Имеем

$$[\mathbf{a}]_\times A = [\widetilde{\mathbf{a}}]_\times \widetilde{A} \Leftrightarrow [\mathbf{a}]_\times (k\widetilde{A} - A) = 0 \Leftrightarrow \mathbf{a} \times (k\widetilde{A} - A)\mathbf{e}_i = 0 \quad (i = 1, 2, 3),$$

где через $\mathbf{e}_i$ обозначены базисные вектора в $\mathbb{R}^3$. Отсюда получаем, что вектор $\mathbf{a}$ сонаправлен с векторами, которые стоят в столбцах матрицы $(k\widetilde{A} - A)$, поэтому $k\widetilde{A} - A = \mathbf{a}\mathbf{v}^\top$, $\widetilde{A} = \dfrac{1}{k}(A + \mathbf{a}\mathbf{v}^\top)$.

Имеем: $P'H_1 = [A|\mathbf{a}]$ и $\widetilde{P}'H_2 = [1/k(A + \mathbf{a}\mathbf{v}^\top)|k\mathbf{a}]$. Возьмем $H = \begin{pmatrix} \frac{1}{k}I & 0 \\ \frac{1}{k}\mathbf{v}^\top & k \end{pmatrix}$. Тогда $PH_1H = \dfrac{1}{k}[I|0] = \dfrac{1}{k}\widetilde{P}H_2$, т.е.

$$PH_1HH_2^{-1} = \frac{1}{k}\widetilde{P},$$

$$P'H_1H = [A|\mathbf{a}]H = [\frac{1}{k}(A + \mathbf{a}\mathbf{v}^\top)|k\mathbf{a}] = \widetilde{P}'H_2.$$

Следовательно,

$$P'H_1HH_2^{-1} = \widetilde{P}'.$$

В качестве искомого проективного преобразования возьмем $H_1HH_2^{-1}$. Теорема доказана. ■

Таким образом, если мы найдем хотя бы одну пару камер, соответствующую заданной матрице фундаментального отображения F, то, применяя всевозможные проективные преобразования $\mathbb{R}P^3$ мы сможем получить все такие пары. Но как найти эту пару? Ответ на этот вопрос дает следующий раздел.

12.6.6 Канонические пары камер, порожденные фундаментальной матрицей

Начнем с ряда вспомогательных лемм. Первое из них является стандартным утверждением линейной алгебры.

Лемма 12.8. *Квадратная матрица A является кососимметрической тогда и только тогда, когда $\mathbf{x}^\top A\mathbf{x} = 0$ для любого вектора $\mathbf{x}$.*

Доказательство. *Необходимость.* Имеем

$$\mathbf{x}^\top A\mathbf{x} = (\mathbf{x}^\top A\mathbf{x})^\top = \mathbf{x}^\top A^\top x = -\mathbf{x}^\top A\mathbf{x} \Rightarrow \mathbf{x}^\top A\mathbf{x} = 0.$$

Достаточность. Возьмем $\mathbf{x} = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ (единица стоит на i-ом месте), тогда получим $a_{ii} = 0$. Теперь возьмем $\mathbf{x} = (0, 0 \ldots, 1, 0, \ldots 1, 0, \ldots, 0)^\top$ (единицы стоят на i-ом и j-ом месте), получим $a_{ii} + a_{ij} + a_{ji} + a_{jj} = 0$, т.е. $a_{ij} = -a_{ji}$. Лемма доказана. ■

Лемма 12.9. *Ненулевая* (3×3)*-матрица* F *является фундаментальной матрицей, соответствующей некоторой паре камер* (P, P') *тогда и только тогда, когда матрица* $(P')^\top FP$ *— кососимметрическая.*

Доказательство. $(P')^\top FP$ — кососимметрическая $\Leftrightarrow$ $X^\top (P')^\top FPX = 0\,(\forall X) \Leftrightarrow (P'X)^\top F(PX) = 0 \quad (\forall X) \Leftrightarrow \mathbf{x}'F\mathbf{x} = 0$ для любых пар сопряженных точек $\mathbf{x}$ и $\mathbf{x}'$. Лемма доказана. ■

Теорема 12.6. *Пусть* F *— некоторая* (3×3)*-матрица ранга* 2, S *— кососимметрическая* (3×3)*-матрица. Определим пару камер соотношениями* $P = [I|0]$ *и* $P' = [SF|\mathbf{e}']$, *где* $\mathbf{e}'^\top F = 0$, *и предположим, что* P' *имеет ранг* 3 *(т.е. корректно задает матрицу камеры). Тогда* F *— фундаментальная матрица, соответствующая паре камер* (P, P').

Доказательство. Имеем

$$(P')^\top FP = [SF|\mathbf{e}']^\top F[I|0] = \begin{pmatrix} F^\top S^\top F & 0 \\ \mathbf{e}'^\top F & 0 \end{pmatrix} = \begin{pmatrix} F^\top S^\top F & 0 \\ 0 & 0 \end{pmatrix}.$$

По лемме 12.9 получаем отсюда, что F — фундаментальная матрица, соответствующая паре камер (P, P'). Теорема доказана. ■

В теореме 12.6 используется предположение о том, что матрица P' имеет ранг 3. Дадим достаточные условия для того, чтобы это предположение было выполнено. Для этого представим матрицу S из формулировки теоремы в следующем виде:

$$S = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} = [\mathbf{s}]_\times,$$

где $\mathbf{s} = (a, b, c)^\top$. В этих обозначениях справедлива следующая

Лемма 12.10. *Если* $\mathbf{s}^\top \mathbf{e}' \neq 0$, *то матрица* P' *из формулировки теоремы* 12.6 *будет иметь ранг* 3.

Доказательство. Из того, что $\mathbf{e}'^\top F = 0$, следует, что вектор $\mathbf{e}'$ ортогонален линейной оболочке $\operatorname{span}(\mathbf{f}^1, \mathbf{f}^2, \mathbf{f}^3)$ вектор-столбцов матрицы F. Далее, поскольку $\mathbf{s}^\top \mathbf{e}' \neq 0$, то вектор $\mathbf{s}$ не перпендикулярен вектору $\mathbf{e}'$. Следовательно, $\mathbf{s}$ не принадлежит $\operatorname{span}(\mathbf{f}^1, \mathbf{f}^2, \mathbf{f}^3)$. Из свойств векторного произведения следует, что линейная оболочка векторов $(\mathbf{s} \times \mathbf{f}^1, \mathbf{s} \times \mathbf{f}^2, \mathbf{s} \times \mathbf{f}^3)$ — это плоскость, перпендикулярная вектору $\mathbf{s}$. Но поскольку вектор $\mathbf{e}'$ не перпендикулярен $\mathbf{s}$, то матрица $[[\mathbf{s}]_\times F|\mathbf{e}']$ имеет ранг 3. Лемма доказана. ■

Замечание. Из доказанной леммы следует, что в качестве матрицы S в теореме 12.6 можно взять $[\mathbf{e}']_\times$.

Замечание. Из леммы 12.7 следует, что имеется четырехпараметрическое семейство пар камер: $\{(P = [I|0], P' = [A + \mathbf{a}\mathbf{v}^\top|k\mathbf{a}])\,|\,\mathbf{a} \in \mathbb{R}^3,\ k \in \mathbb{R}\}$, которые имеют одну и ту же фундаментальную матрицу. Отсюда получаем

формулы **канонических камер**, соответствующих данной фундаментальной матрице F:

$$P = [I|0], \qquad P' = [[\mathbf{e}']_\times F + \mathbf{e}'\mathbf{v}^\top | \lambda \mathbf{e}'].$$

12.7 ПЛА и НЛА алгоритмы приближенного вычисления проективного преобразования. Оценка их устойчивости (Г. В. Носовский, А. А. Толченников)

12.7.1 Вычисление проективного преобразования по точно заданным сопряженным точкам

Пусть имеется два экземпляра проективной плоскости $M = \mathbb{R}P^2$ и $M' = \mathbb{R}P^2$, отнесенные к однородным координатам $(x_i : y_i : w_i)$ и $(x'_i : y'_i : w'_i)$ соответственно. Пусть даны два набора точек $X_i \in M$ и $X'_i \in M'$ $(i = 1, \ldots, n, n \geqslant 4)$, принадлежащие, соответственно, аффинным картам $\{w\} \neq 0$ и $\{w'\} \neq 0$. Пусть эти два набора точек связаны некоторым проективным преобразованием $H\colon M \to M'$, $H(X_i) = X'_i$ $(i = 1, \ldots, n)$. Решим задачу нахождения преобразования H. Будем считать, что точки X_i и X'_i заданы своими однородными координатами: $X_i = (x_i : y_i : w_i)$, $X'_i = (x'_i : y'_i : w'_i)$, а искомое проективное преобразование H представляется невырожденной (3×3)-матрицей, определенной с точностью до множителя.

Будем предполагать, что все точки в наборах $\{X_i\}_{i=1}^n$ и $\{X'_i\}_{i=1}^n$ — точки общего положения.

В однородных координатах равенство $X'_i = HX_i$ справедливо с точностью до множителя и равносильно равенству $(X'_i) \times (HX_i) = 0$. Обозначим j-ую строку матрицы H через $(\mathbf{h}^j)^\top$. Тогда $HX_i = \begin{pmatrix} (\mathbf{h}^1)^\top X_i \\ (\mathbf{h}^2)^\top X_i \\ (\mathbf{h}^3)^\top X_i \end{pmatrix}$. В этих обозначениях имеем:

$$X'_i \times HX_i = \begin{pmatrix} y'_i(\mathbf{h}^3)^\top X_i - w'_i(\mathbf{h}^2)^\top X_i \\ w'_i(\mathbf{h}^1)^\top X_i - x'_i(\mathbf{h}^3)^\top X_i \\ x'_i(\mathbf{h}^2)^\top X_i - y'_i(\mathbf{h}^1)^\top X_i \end{pmatrix} = \begin{pmatrix} y'_i(X_i^\top \mathbf{h}^3) - w'_i(X_i^\top \mathbf{h}^2) \\ w'_i(X_i^\top \mathbf{h}^1) - x'_i(X_i^\top \mathbf{h}^3) \\ x'_i(X_i^\top \mathbf{h}^2) - y'_i(X_i^\top \mathbf{h}^1) \end{pmatrix} =$$

$$= \begin{pmatrix} 0^\top & -w'_i X_i^\top & y'_i X_i^\top \\ w'_i X_i^\top & 0^\top & -x'_i X_i^\top \\ -y'_i X_i^\top & x'_i X_i^\top & 0^\top \end{pmatrix} \begin{pmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{pmatrix} = C_i \mathbf{h} \qquad (1 \leqslant i \leqslant n).$$

В последнем равенстве элементы матриц C_i — вектор-строки (1×3), т.е. матрицы C_i имеют размер (3×9). Из последнего равенства следует, что для каждого $i = 1, \ldots, n$ соотношение $X'_i = HX_i$ эквивалентно соотношению $C_i \mathbf{h} = 0$.

Утверждение 12.1. *Для любого $1 \leqslant i \leqslant n$ матрица C_i имеет ранг 2.*

Доказательство. Фиксируем i и рассмотрим несколько случаев:

1) если $x'_i \neq 0, y'_i \neq 0$, то 3-я строка матрицы C_i пропорциональна сумме первой строки, умноженной на x'_i, и второй, умноженной на y'_i, а строки 1 и 2 линейно независимы. Это означает, что $\operatorname{rank}(C_i) = 2$.

2) если $x'_i = 0, y'_i \neq 0$, то строки 1 и 2 матрицы C_i линейно независимы, а 3-я строка пропорциональна 2-ой. Это означает, что $\operatorname{rank}(C_i) = 2$.

3) если $x'_i \neq 0, y'_i = 0$, то 1 и 2 строки матрицы C_i линейно независимы, а 3-я строка пропорциональна 1-ой. Это означает, что $\operatorname{rank}(C_i) = 2$.

4) если $x'_i = 0, y'_i = 0$, то строки 1 и 2 матрицы C_i линейно независимы, а 3-я строка нулевая. Это означает, что $\operatorname{rank}(C_i) = 2$. Утверждение доказано. ■

Из доказательства утверждения 12.1 следует, что третья строка матрицы C_i является линейной комбинацией первых двух ее строк. Таким образом, для нахождения матрицы проективного преобразования H достаточно решить систему линейных уравнений $A\mathbf{h} = 0$, где A — матрица размера $(2n \times 9)$, составленная из матриц A_i размера (2×9), которые, в свою очередь, получены из матриц C_i исключением третьей строки:

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix}, \qquad A_i = \begin{pmatrix} 0^\top & -w'_i X_i^\top & y'_i X_i^\top \\ w'_i X_i^\top & 0^\top & -x'_i X_i^\top \end{pmatrix}. \tag{12.10}$$

Известно, что проективное преобразование однозначно задается при помощи четырех пар соответствий точек общего положения (см., напр. [43]). Следовательно, согласно предположению о том, что $n \geqslant 4$, система $A\mathbf{h} = 0$ будет иметь единственное с точностью до множителя решение. Это означает, что матрица A имеет ранг 8.

Докажем предложение которое будет использовано ниже:

Утверждение 12.2. *Пусть A — матрица, определенная в* (12.10). *Тогда нулевые подпространства матриц A и $A^\top A$ совпадают.*

Доказательство. Пусть для некоторого вектора $\mathbf{h}$ выполнено: $A^\top A\mathbf{h} = 0$. Тогда $\langle A^\top A\mathbf{h}, \mathbf{h}\rangle = 0$. Транспонируя, получим: $\langle A\mathbf{h}, A\mathbf{h}\rangle = 0$, т.е. $A\mathbf{h} = 0$. Следовательно, нулевое подпространство матрицы $A^\top A$ лежит в нулевом подпространстве матрицы A. Обратное включение очевидно. Утверждение доказано. ■

Описанный выше метод нахождения H как решения системы линейных уравнений $A\mathbf{h} = 0$ относится к случаю, когда координаты точек измерены абсолютно точно. Теперь перейдем к алгоритмам нахождения проективного преобразования в случае, когда координаты точек X_i и X'_i измерены с некоторой ошибкой.

12.7.2 Простой линейный алгоритм (ПЛА) вычисления проективного преобразования по приближенно заданным сопряженным точкам

Исходные данные: координаты точек $\widetilde{X}_i = (\widetilde{x}_i, \widetilde{y}_i, 1) = (x_i + \Delta x_i, y_i + \Delta y_i, 1)$, $\widetilde{X}'_i = (\widetilde{x}'_i, \widetilde{y}'_i, 1) = (x'_i + \Delta x'_i, y'_i + \Delta y'_i, 1)$, где $1 \leqslant i \leqslant n$ $(n \geqslant 4)$. Известно, что точки $X_i = (x_i, y_i, 1)$ и $X'_i = (x'_i, y'_i, 1)$ связаны некоторым проективным преобразованием: $H(X_i) = X'_i$.

Требуется: найти (приближенно) преобразование H. Будем обозначать искомое приближение к H через $\widetilde{H}$.

Алгоритм решения поставленной задачи (простой линейный алгоритм [51]):

1) По соответствиям $\widetilde{X}_i \leftrightarrow \widetilde{X}'_i$ строим матрицы $\widetilde{A}_i$ с помощью формулы 12.10. Из них собираем матрицу $\widetilde{A}$ размера $(2n \times 9)$ по формуле 12.10.

2) Ищем вектор $\widetilde{\mathbf{h}}$, на котором достигается минимум функции $||\widetilde{A}\mathbf{h}||$ при условии $||\mathbf{h}|| = 1$.

3) Полагаем $\widetilde{H} = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix}$.

Поясним, почему полученный таким способом вектор $\widetilde{\mathbf{h}}$ (однозначно определяющий искомое преобразование $\widetilde{H}$) близок к истинному $\mathbf{h}$ (определяющему неизвестное преобразование H). А также почему $\widetilde{\mathbf{h}}$ определен однозначно и как его вычислять.

Утверждение 12.3. *Пусть A — матрица размера $m \times n$, $\mathbf{g}$ — вектор-столбец размера $(n \times 1)$. Пусть минимум нормы $||A\mathbf{g}||$ при условии $||\mathbf{g}|| = 1$, достигается на векторе $\widetilde{\mathbf{g}}$ (где $||\cdot||$ — евклидова норма вектора). Тогда $\widetilde{\mathbf{g}}$ является нормированным собственным вектором матрицы $A^{\top}A$, отвечающим ее минимальному собственному значению.*

Доказательство. Воспользуемся методом множителей Лагранжа (см., например [2]). Определим функцию Лагранжа

$$F(g_1, \ldots g_n, \lambda) = ||A\mathbf{g}||^2 - \lambda(||\mathbf{g}||^2 - 1) = (a_{11}g_1 + \ldots + a_{1n}g_n)^2 + \ldots$$
$$\ldots + (a_{m1}g_1 + \ldots + a_{mn}g_n)^2 - \lambda(g_1^2 + \ldots + g_n^2 - 1).$$

Тогда необходимым условием того, что $\widetilde{\mathbf{g}}$ является точкой экстремума является условие:

$$\begin{cases} F'_{g_1}(\widetilde{\mathbf{g}}, \lambda) = 0, \\ \ldots\ldots\ldots \\ F'_{g_n}(\widetilde{\mathbf{g}}, \lambda) = 0, \\ F'_{\lambda}(\widetilde{\mathbf{g}}, \lambda) = 0. \end{cases} \tag{12.11}$$

Имеем

$$F'_{g_1} = 2a_{11}(a_{11}g_1 + \ldots + a_{1n}g_n) + \ldots + 2a_{m1}(a_{m1}g_1 + \ldots + a_{mn}g_n) - 2\lambda g_1,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$F'_{g_n} = 2a_{1n}(a_{11}g_1 + \ldots + a_{1n}g_n) + \ldots + 2a_{mn}(a_{m1}g_1 + \ldots + a_{mn}g_n) - 2\lambda g_n,$$

$$F'_{\lambda} = -(g_1^2 + \ldots + g_n^2 - 1).$$

Условие (12.11) в данном случае приобретает вид

$$\begin{cases} a_{11}(A\widetilde{\mathbf{g}})_1 + \ldots + a_{m1}(A\widetilde{\mathbf{g}})_m = \lambda\widetilde{g}_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}(A\widetilde{\mathbf{g}})_1 + \ldots + a_{mn}(A\widetilde{\mathbf{g}})_m = \lambda\widetilde{g}_n, \\ g_1^2 + \ldots + g_n^2 = 1 \end{cases}$$

или

$$\begin{cases} (a_{11}, \ldots, a_{m1})A\widetilde{\mathbf{g}} = \lambda\widetilde{g}_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ (a_{1n}, \ldots, a_{mn})A\widetilde{\mathbf{g}} = \lambda\widetilde{g}_n, \\ g_1^2 + \ldots + g_n^2 = 1. \end{cases}$$

Последние равенства можно записать в виде $A^{\top}A\widetilde{\mathbf{g}} = \lambda\widetilde{\mathbf{g}}$. Следовательно, $||A\widetilde{\mathbf{g}}||^2 = \langle A\widetilde{\mathbf{g}}, A\widetilde{\mathbf{g}}\rangle = \langle A^{\top}A\widetilde{\mathbf{g}}, \widetilde{\mathbf{g}}\rangle = \langle\lambda\widetilde{\mathbf{g}}, \widetilde{\mathbf{g}}\rangle = \lambda||\widetilde{\mathbf{g}}||^2 = \lambda$. Отсюда следует, что собственный вектор $\mathbf{g} = \widetilde{\mathbf{g}}$, на котором достигается минимум $||A\mathbf{g}||$, соответствует минимальному собственному значению матрицы $A^{\top}A$. Утверждение доказано. ■

Согласно утверждению 12.2, истинное значение $\mathbf{h}$ есть собственный вектор $A^{\top}A$, соответствующий нулевому собственному значению. Найденный с помощью ПЛА вектор $\widetilde{\mathbf{h}}$, по утверждению 12.3 является собственным вектором $\widetilde{A}^{\top}\widetilde{A}$, соответствующим минимальному собственному значению этой матрицы. При небольших ошибках в измерениях, кратность $\lambda_{min}(\widetilde{A}^{\top}\widetilde{A})$ останется равна 1 (в силу непрерывной зависимости собственных значений от элементов матрицы), поэтому собственное подпространство матрицы $\widetilde{A}^{\top}\widetilde{A}$, соответствующее минимальному собственному значению, останется одномерным. Следовательно, вектор $\widetilde{\mathbf{h}}$ в однородных координатах будет определен однозначно.

12.7.3 Вычисление вектора, соответствующего минимальному собственному значению матрицы

Для применения НЛА необходимо уметь находить собственный вектор матрицы, соответствующий ее минимальному собственному значению. Рассмотрим эту задачу.

Пусть дана матрица $B_{n\times n}$, все собственные значения которой вещественны и положительны (например, матрица B симметрична и положительно определена). Упорядочим все собственные значения матрицы B по убыванию: $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_{n-1} > \lambda_n$. Опишем часто применяемый алгоритмический способ нахождения минимального собственного значения матрицы B и

соответствующего ему собственного вектора. Выберем число $\lambda \in \mathbb{R}$ так, что $\lambda \geqslant \lambda_1$ (например, можно взять $\lambda = ||B||_\infty$). Тогда матрица $B' = \lambda E - B$ будет иметь базис из собственных векторов $\mathbf{e}_i$, $i = 1, \ldots, n$: $B'\mathbf{e}_i = (\lambda - \lambda_i)\mathbf{e}_i$, причем $\lambda - \lambda_n > \lambda - \lambda_{n-1} \geqslant \ldots \geqslant \lambda - \lambda_1$. Следовательно, для матрицы $B' = \lambda E - B$ применима сформулированная ниже теорема 12.7, а собственное значение $\lambda - \lambda_n$ (а значит, и λ_n) и соответствующий ему собственный вектор $\mathbf{e}_n$ могут быть найдены с помощью итерационного процесса, описанного в этой теореме.

Теорема 12.7 (Степенной метод [4]). *Пусть матрица $B_{n\times n}$ имеет базис из собственных векторов $\mathbf{e}_i$, $i = 1, \ldots, n$: $B\mathbf{e}_i = \lambda_i \mathbf{e}_i$, причем $|\lambda_1| > |\lambda_2| \geqslant |\lambda_3| \geqslant \ldots \geqslant |\lambda_n|$ (т.е. вектора занумерованы в порядке убывания модуля собственного значения, причем собственное значение с максимальным модулем единственно). Тогда для всякого вектора $\mathbf{x}^{(0)} \in \mathbb{C}^n$ такого, что $\langle \mathbf{x}^{(0)}, \mathbf{e}_1 \rangle \neq 0$, итерационный процесс*

$$\mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)}, \qquad \lambda_1^{(k)} = \frac{\langle \mathbf{x}^{(k+1)}, \mathbf{x}^{(k)} \rangle}{\langle \mathbf{x}^{(k)}, \mathbf{x}^{(k)} \rangle}$$

сходится к собственному значению λ_1, причем

$$\lambda_1^{(k)} = \lambda_1 + O\left(\lambda_1 \left|\frac{\lambda_2}{\lambda_1}\right|^k\right) \qquad (k \to \infty),$$

а величины $\mathbf{e}_1^{(k)} = \dfrac{\mathbf{x}^{(k)}}{||\mathbf{x}^{(k)}||}$ сходятся к собственному вектору, соответствующему λ_1:

$$\mathbf{e}_1^{(k)} = \mathbf{e}^{i\phi_k}\mathbf{e}_1 + O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right),$$

где $\mathbf{e}^{i\phi_k}$ — комплексное число, по модулю равное 1.

Замечание. В реальных вычислениях итерационный процесс всегда сходится, даже если условие $\langle \mathbf{x}^{(0)}, \mathbf{e}_1 \rangle \neq 0$ не выполнено. Дело в том, что из-за присутствия округлений для некоторого k_0 будет выполнено $\langle \mathbf{x}^{(k_0)}, \mathbf{e}_1 \rangle \neq 0$, даже если $\langle \mathbf{x}^{(0)}, \mathbf{e}_1 \rangle = 0$. После этого описанный выше процесс работает, как если бы его начали с $\mathbf{x}^{(0)} = \mathbf{x}^{(k_0)}$.

12.7.4 Неинвариантность ПЛА относительно движений плоскости

Определим понятие *инвариантного относительно замен декартовых координат* (*движений плоскости*) алгоритма нахождения проективного преобразования.

Пусть имеется движение двумерной плоскости, задаваемое ортогональной (2×2)-матрицей R и вектором $\mathbf{v}$: $\mathbf{x}' = R\mathbf{x} + \mathbf{v}$. Если рассматривать

$\mathbb{R}^2$ как аффинную карту в $\mathbb{R}P^2$, то движение $R\mathbf{x} + \mathbf{v}$ можно записать в однородных координатах следующим образом:

$$\begin{pmatrix} x_1' \\ x_2' \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha & v_1 \\ \sin\alpha & \cos\alpha & v_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}.$$

Таким образом, любая пара замен декартовых координат в M и M' соответственно, задается формулами $\widetilde{X} = TX$ и $\widetilde{X}' = T'X'$. Здесь T и T' — преобразования вида $\begin{pmatrix} sR & \mathbf{v} \\ 0^\top & 1 \end{pmatrix}$, где R — некоторая ортогональная (2×2)-матрица, описывающая поворот системы координат, s — скаляр, отвечающий изменению масштаба, а $\mathbf{v} = (a, b)^\top$ — вектор-столбец, отвечающий сдвигу начала системы координат.

Обозначение. Будем использовать запись: X_i, X_i' и H для обозначения измеренных с ошибкой координат сопряженных точек и приближенного проективного преобразования в исходной системе координат. Такие же обозначения с волной: $\widetilde{X}_i$, $\widetilde{X}_i'$ и $\widetilde{H}$ будем использовать для тех же величин в новой системе координат.

Определение 12.9. Алгоритм нахождения проективного преобразования называется *инвариантным относительно замен координат*, задаваемых матрицами T и T', если приближенные преобразования H и $\widetilde{H}$, найденные в результате работы алгоритма, примененного к соответствиям $X_i \leftrightarrow X_i'$ и к $\widetilde{X}_i = TX_i \leftrightarrow \widetilde{X}_i' = T'X_i'$, связаны соотношением $\widetilde{H} = T'HT^{-1}$.

Покажем, что простой линейный алгоритм не является инвариантным относительно всевозможных замен декартовых координат. Пусть на $\mathbf{h}_0$ достигается минимум $||A\mathbf{h}||$ при условии $||\mathbf{h}|| = 1$, и на $\widetilde{\mathbf{h}}_0$ достигается минимум $||\widetilde{A}\widetilde{\mathbf{h}}||$ при условии $||\widetilde{\mathbf{h}}|| = 1$. Нам надо показать, что для соответствующих матриц преобразований H_0 и $\widetilde{H}_0$ соотношение

$$\widetilde{H}_0 = T'H_0T^{-1} \tag{12.12}$$

вообще говоря, неверно. В силу невырожденности T и T' можно ввести параметризацию области изменения вектора $\widetilde{\mathbf{h}}$ через область изменения вектора $\mathbf{h}$ по формуле (12.12), которая определяет зависимость $\widetilde{\mathbf{h}} = \widetilde{\mathbf{h}}(\mathbf{h})$. Тогда задачу можно сформулировать так: пусть на $\mathbf{h}_0$ достигается минимум $||A\mathbf{h}||$ при условии $||\mathbf{h}|| = 1$, и на $\mathbf{h}_1$ достигается минимум $||\widetilde{A}\widetilde{\mathbf{h}}(\mathbf{h})||$ при условии $||\widetilde{\mathbf{h}}(\mathbf{h})|| = 1$. Будет ли верным соотношение $\mathbf{h}_0 = \mathbf{h}_1$? Ответ, вообще говоря, отрицателен. Чтобы это показать, докажем

Утверждение 12.4. *Пусть* $T' = \begin{pmatrix} sR & \mathbf{v} \\ 0^\top & 1 \end{pmatrix}$ *— замена декартовых координат в* M', *а* T *— произвольная замена координат в* M. *Тогда* $s||A\mathbf{h}|| = ||\widetilde{A}\widetilde{\mathbf{h}}(\mathbf{h})||$.

Доказательство. Фиксируем $1 \leqslant i \leqslant n$. По условию координаты точек в M и M' изменяются по формулам: $\widetilde{X}_i = TX_i$, $\widetilde{X}'_i = T'X'_i$. Рассмотрим вектор $\mathbf{e}_i = X'_i \times HX_i$. Тогда по определению матриц A_i, вектор $A_i\mathbf{h}$ состоит из первых двух компонент вектора $\mathbf{e}_i$. Аналогично, вектор $\widetilde{A}_i\widetilde{\mathbf{h}}$ состоит из первых двух компонент вектора $\widetilde{\mathbf{e}}_i$, где $\widetilde{\mathbf{e}}_i = \widetilde{X}'_i \times \widetilde{H}\widetilde{X}_i$. Имеем

$$\widetilde{\mathbf{e}}_i = \widetilde{X}'_i \times \widetilde{H}\widetilde{X}_i = T'X'_i \times (T'HT^{-1})TX_i = T'X'_i \times T'HX_i.$$

Докажем вспомогательное

Утверждение 12.5. *Для любой матрицы* $T = \begin{pmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \mathbf{t}_3 \end{pmatrix}$ *размера* (3×3) *(здесь* $\mathbf{t}_1$, $\mathbf{t}_2$, $\mathbf{t}_3$ *— столбцы матрицы* T*) и для любых векторов* $X, Y \in \mathbb{R}^3$ *выполнено соотношение* $TX \times TY = T^*(X \times Y)$*, где матрица* T^* *имеет вид* $\begin{pmatrix} \mathbf{t}_2 \times \mathbf{t}_3 & \mathbf{t}_3 \times \mathbf{t}_1 & \mathbf{t}_1 \times \mathbf{t}_2 \end{pmatrix}$.

Доказательство. Пусть $X = (x_1, x_2, x_3)$, $Y = (y_1, y_2, y_3)$,

$$T = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix}.$$

Сначала покажем, что $TX \times TY$ можно представить в виде $B(X \times Y)$, где B — некоторая матрица размера 3×3. Выпишем первую компоненту $TX \times TY$:

$$\begin{aligned}(t_{31}y_1 + t_{32}y_2 + t_{33}y_3)(t_{21}x_1 + t_{22}x_2 + t_{23}x_3) - (t_{21}y_1 + t_{22}y_2 + t_{23}y_3) \times \\ \times (t_{31}x_1 + t_{32}x_2 + t_{33}x_3) = (t_{33}t_{22} - t_{23}t_{32})(y_3x_2 - y_2x_3) + (t_{31}t_{23} - t_{21}t_{33}) \times \\ \times (y_1x_3 - x_1y_3) + (t_{32}t_{21} - t_{22}t_{31})(x_1y_2 - y_1x_2) = \\ = (B)_{11}(X \times Y)_1 + (B)_{12}(X \times Y)_2 + (B)_{13}(X \times Y)_3\end{aligned}$$

Аналогичные равенства можно написать и для двух остальных компонент. Таким образом, для некоторой матрицы B выполнено требуемое соотношение $TX \times TY = B(X \times Y)$. Первый столбец матрицы B равен

$$B\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = B\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \times \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = T\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \times T\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \mathbf{t}_2 \times \mathbf{t}_3.$$

Аналогично находятся остальные два столбца матрицы B. Отсюда получаем, что $B = T^*$. Утверждение доказано. ■

Продолжим доказательство утверждения 12.4.

Из утверждения 12.5 следует, что

$$\widetilde{\mathbf{e}}_i = T'X'_i \times T'HX_i = T'^*(X'_i \times HX_i) = T'^*\mathbf{e}_i. \tag{12.13}$$

Найдем T'^*. Имеем по условию:

$$T' = \begin{pmatrix} sR & \mathbf{v} \\ 0^\top & 1 \end{pmatrix},$$

где R — ортогональная матрица, $\mathbf{v} = (a, b)^\top$, s — скаляр. Следовательно, столбцы матрицы T' имеют вид

$$\mathbf{t}_1 = \begin{pmatrix} s\cos\alpha \\ s\sin\alpha \\ 0 \end{pmatrix}, \qquad \mathbf{t}_2 = \begin{pmatrix} -s\sin\alpha \\ s\cos\alpha \\ 0 \end{pmatrix}, \qquad \mathbf{t}_3 = \begin{pmatrix} a \\ b \\ 1 \end{pmatrix}.$$

Имеем

$$\mathbf{t}_2 \times \mathbf{t}_3 = \begin{pmatrix} s\cos\alpha \\ s\sin\alpha \\ -sb\sin\alpha - sa\cos\alpha \end{pmatrix}, \qquad \mathbf{t}_3 \times \mathbf{t}_1 = \begin{pmatrix} -s\sin\alpha \\ s\cos\alpha \\ -sb\cos\alpha + sa\sin\alpha \end{pmatrix},$$

$$\mathbf{t}_1 \times \mathbf{t}_2 = \begin{pmatrix} 0 \\ 0 \\ s^2 \end{pmatrix}.$$

Из последних трех равенств получаем, что

$$T'^* = s\begin{pmatrix} R & 0 \\ -\mathbf{v}^\top R & s \end{pmatrix}.$$

Подставим полученное выражение для T'^* в (12.13). Первые две компоненты получившегося векторного равенства можно записать в виде

$$\widetilde{A}_i\widetilde{\mathbf{h}}(\mathbf{h}) = (\widetilde{\mathbf{e}}_{i1}, \widetilde{\mathbf{e}}_{i2})^\top = sR(\mathbf{e}_{i1}, \mathbf{e}_{i2})^\top = sRA_i\mathbf{h}$$

Откуда $||\widetilde{A}\widetilde{\mathbf{h}}(\mathbf{h})|| = s||A\mathbf{h}||$. Утверждение 12.4 доказано. ■

Теперь все готово, чтобы отрицательно ответить на вопрос об инвариантности простого линейного алгоритма. Пусть минимум функции $||A\mathbf{h}||$ на поверхности $||\mathbf{h}|| = 1$ достигается на $\mathbf{h}_0$, а минимум функции $||\widetilde{A}\widetilde{\mathbf{h}}(\mathbf{h})||$ на поверхности $||\widetilde{\mathbf{h}}(\mathbf{h})|| = 1$ достигается на $\mathbf{h}_1$. По утверждению 12.4 имеем: $s||A\mathbf{h}|| = ||\widetilde{A}\widetilde{\mathbf{h}}(\mathbf{h})||$, т.е. в первом случае минимум функции ищется на сфере, а во втором случае ищется минимум той же функции, умноженной на заданную константу — но уже по другой поверхности. При этом вектор $\mathbf{h}_0$ не обязан, вообще говоря, совпадать с $\mathbf{h}_1$. Поясним это на примере.

Пример 13. Пусть происходит изменение масштаба на втором экране, а на первом экране координаты не меняются:

$$T' = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Поверхность $M(T, T')$, на которой ищется минимум, задается уравнением $||\widetilde{\mathbf{h}}(\mathbf{h})|| = 1$. Из формулы (12.12) имеем: $\widetilde{h}_1 = sh_1, \ldots, \widetilde{h}_6 = sh_6$, $\widetilde{h}_7 = h_7$, $\widetilde{h}_8 = h_8$, $\widetilde{h}_9 = h_9$. Следовательно, уравнение поверхности $M(T, T')$ можно записать в виде: $s^2(h_1^2 + \ldots + h_6^2) + h_7^2 + h_8^2 + h_9^2 = 1$. Предположим, что

вектор, на котором достигается минимум функции $||A\mathbf{h}||$ на сфере совпал с вектором, на котором достигается минимум по поверхности $M(T, T')$. Тогда, проводя вычисления, аналогичные вычислениям из утверждения 12.3, получаем, что для вектора $\mathbf{h}$, на котором достигается минимум, выполнены уравнения

$$\begin{gathered} A^\top A\mathbf{h} = \lambda\mathbf{h}, \\ A^\top A\mathbf{h} = \nu \begin{pmatrix} sh_1 & sh_2 & sh_3 & sh_4 & sh_5 & sh_6 & h_7 & h_8 & h_9 \end{pmatrix}^\top \end{gathered} \tag{12.14}$$

для некоторых λ и ν. Поскольку X_i и X_i' $(i = 1, \ldots, n,\ n \geqslant 4)$ — это координаты сопряженных точек, которые измерены с некоторой ошибкой, то проективного преобразования такого, что для любого i: $H(X_i) = (X_i')$, вообще говоря, не существует. Следовательно, вообще говоря, $\lambda \neq 0$ и $\nu \neq 0$. Поэтому уравнения (12.14) будут выполняться только если или первые шесть координат вектора $\mathbf{h}$ равны нулю, или последние три координаты вектора $\mathbf{h}$ равны нулю. Но для произвольных соответствий $X_i \leftrightarrow X_i'$ такое условие на H не обязано выполняться. Получили противоречие.

Утверждение 12.6. *Простой линейный алгоритм инвариантен относительно преобразований вида* $T = \begin{pmatrix} R & 0 \\ 0^\top & 1 \end{pmatrix}$, $T' = \begin{pmatrix} R' & 0 \\ 0^\top & 1 \end{pmatrix}$, *где* R *и* R' *— ортогональные матрицы.*

Доказательство. Для того, чтобы это доказать, достаточно показать, что минимум функции $||A\mathbf{h}||$ на сфере $||\mathbf{h}|| = 1$ и на поверхности $M(T, T')$, задаваемой уравнением $||\widetilde{\mathbf{h}}(\mathbf{h})|| = 1$ (где функция $\widetilde{\mathbf{h}}(\mathbf{h})$ задана формулой (12.12)), достигается в одной и той же точке. Условие $||\widetilde{\mathbf{h}}(\mathbf{h})||^2 = 1$ означает, что сумма квадратов элементов матрицы $\widetilde{H}$ равна 1. Обозначим $B = HT^{-1}$, и пусть $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ — столбцы матрицы B. Тогда по формуле (12.12) имеем:

$$\widetilde{H} = T'B = T' \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{pmatrix} = \begin{pmatrix} T'\mathbf{b}_1 & T'\mathbf{b}_2 & T'\mathbf{b}_3 \end{pmatrix}.$$

Но, по условию, матрица T' — ортогональная, поэтому суммы квадратов элементов у матриц $\widetilde{H}$ и B будут совпадать. Аналогично, поскольку T — ортогональная матрица, то суммы квадратов элементов у матриц B и H совпадают. Получаем, что условие $||\mathbf{h}|| = 1$ равносильно условию $||\widetilde{\mathbf{h}}(\mathbf{h})|| = 1$. Таким образом, поверхность $M(T, T')$ в этом случае является единичной сферой. Отсюда следует инвариантность простого линейного алгоритма относительно данного класса замен координат. Утверждение доказано. ■

12.7.5 Нормализованный линейный алгоритм (НЛА)

Существует эвристическое предположение [51], что не все системы координат одинаково хороши для использования простого линейного алгоритма. В качестве обоснования этого предположения позднее будет приведен пример. В этом разделе будет рассмотрена модификация ПЛА, инвариантная относительно замен координат.

Нормализованный линейный алгоритм (НЛА) нахождения проективного преобразования H.

Исходные данные: координаты точек $\widetilde{X}_i = (\widetilde{x}_i, \widetilde{y}_i, 1) = (x_i + \Delta x_i, y_i + \Delta y_i, 1)$, $\widetilde{X}'_i = (\widetilde{x}'_i, \widetilde{y}'_i, 1) = (x'_i + \Delta x'_i, y'_i + \Delta y'_i, 1)$, где $1 \leqslant i \leqslant n$ $(n \geqslant 4)$. Известно, что точки $X_i = (x_i, y_i, 1)$ и $X'_i = (x'_i, y'_i, 1)$ связаны некоторым неизвестным проективным преобразованием $H(X_i) = X'_i$.

Требуется: найти (приближенно) преобразование H. Будем обозначать искомое приближение к H через $\widetilde{H}$.

Алгоритм решения поставленной задачи (нормализованный линейный алгоритм [51]):

1) Выбираем преобразование T_0 такое, что для точек $T_0\widetilde{X}_i$ выполнены условия: центр масс точек $T_0\widetilde{X}_1, T_0\widetilde{X}_2, \ldots, T_0\widetilde{X}_n$ находится в начале координат, и среднее расстояние от этих точек до начала координат равно $\sqrt{2}$.

2) Выбираем преобразование T'_0 такое, что для точек $T'_0\widetilde{X}'_i$ выполнены аналогичные условия.

3) Применяем ПЛА для соответствий $T_0\widetilde{X}_i \leftrightarrow T'_0\widetilde{X}'_i$. Получим преобразование $\widetilde{H}_0$.

4) Полагаем $\widetilde{H} = (T'_0)^{-1}\widetilde{H}_0 T_0$.

Первые два шага НЛА будем называть *нормализацией*.

Замечание. Хотя преобразования T_0 и T'_0 определены с точностью до поворотов плоскостей R и R' соответственно, но на результат алгоритма НЛА это не влияет. Действительно, пусть нормализация происходит при помощи других преобразований RT_0 и $R'T'_0$, отличающихся от прежних на повороты R и R'. Применяя ПЛА к соответствиям $RT_0\widetilde{X}_i \leftrightarrow R'T'_0\widetilde{X}'_i$, получаем, в силу утверждения 12.6, преобразование $R'\widetilde{H}_0R^{-1}$. Тогда результатом НЛА является преобразование $(R'T'_0)^{-1}R'\widetilde{H}_0R^{-1}RT_0 = (T'_0)^{-1}\widetilde{H}_0T_0 = \widetilde{H}$.

Известно [51], что алгоритм ПЛА менее устойчив к возмущению начальных данных, чем НЛА.

Утверждение 12.7. *Нормализованный линейный алгоритм является инвариантным относительно любых замен координат T и T' в M и M' соответственно.*

Доказательство. 1) Пусть имеются соответствия сопряженных точек $\widetilde{X}_i \leftrightarrow \widetilde{X}'_i$. Пусть T_0 — такое преобразование, что центр масс точек $T_0\widetilde{X}_i$ находится в начале координат, и среднее расстояние от начала координат до этих точек равно $\sqrt{2}$. Пусть T'_0 — такое преобразование, что для точек $T'_0\widetilde{X}'_i$ выполнены аналогичные условия. Пусть после применения ПЛА к соответствиям $T_0\widetilde{X}_i \leftrightarrow T'_0\widetilde{X}'_i$ получается преобразование H_0. Тогда результатом НЛА будет преобразование $H_1 = (T'_0)^{-1}H_0T_0$.

2) Пусть имеются соответствия $T\widetilde{X}_i \leftrightarrow T'\widetilde{X}'_i$. Согласно приведенному выше замечанию, нормализацию можно провести, например, при помощи преобразований T_0T^{-1} и $T'_0T'^{-1}$. Следовательно, в результате применения НЛА получим преобразование $H_2 = T'(T'_0)^{-1}H_0T_0T^{-1}$.

Преобразования, полученные в пункте 1) и 2), связаны соотношением $H_2 = T' H_1 T^{-1}$, что и означает инвариантность НЛА. Утверждение доказано. ■

12.7.6 Оценка ошибки для алгоритмов ПЛА и НЛА

Определим, насколько устойчив алгоритм ПЛА к ошибкам в измерениях координат точек на обоих экранах. Поскольку вычислительная процедура как ПЛА, так и НЛА сводится, по сути, к вычислению собственного вектора некоторой матрицы, то необходимо оценить насколько изменится собственный вектор матрицы при возмущении ее элементов.

Основным утверждением настоящего раздела является теорема 12.8.

Следующие две леммы доказываются в курсах матричной алгебры.

Введем следующие нормы на линейном пространстве вещественных матриц $n \times m$:

$$||A||_\infty = \max_{1 \leqslant i \leqslant n} \sum_{j=1}^{m} |a_{ij}|, \qquad ||A||_1 = \max_{1 \leqslant j \leqslant m} \sum_{i=1}^{n} |a_{ij}|, \qquad ||A||_2 = \sup_{\neq 0} \frac{||A\mathbf{x}||}{||\mathbf{x}||},$$

где $||\mathbf{x}||$ — евклидова норма вектора $\mathbf{x}$. Известно, что эти формулы действительно задают нормы в пространстве матриц [4].

Лемма 12.11. *Матричная норма $||A||_1$ подчинена векторной норме $||\mathbf{x}||_1 = \sum\limits_{i=1}^{m} |x_i|$, т.е.*

$$||A||_1 = \sup_{\neq 0} \frac{||A\mathbf{x}||_1}{||\mathbf{x}||_1}.$$

Аналогично, матричная норма $||A||_\infty$ подчинена векторной норме $||\mathbf{x}||_\infty = \max\limits_{1 \leqslant i \leqslant m} |x_i|$, т.е.

$$||A||_\infty = \sup_{\neq 0} \frac{||A\mathbf{x}||_\infty}{||\mathbf{x}||_\infty}.$$

Нам понадобится следующее следствие из теоремы Гершгорина, см. например [4].

Лемма 12.12 ([4]). *Пусть $A_{n\times n}$ — диагонализируемая матрица, т.е. существует такая невырожденная матрица C и такая диагональная матрица $\Lambda = \mathrm{diag}(\lambda_1, \ldots \lambda_n)$, что $A = C\Lambda C^{-1}$. Пусть имеется матрица $K_{n\times n}$. Тогда для любого собственного значения $\widehat{\lambda}$ матрицы $A + K$ существует такое собственное значение матрицы A, что*

$$|\widehat{\lambda} - \lambda| \leqslant ||C^{-1}||_\infty ||C||_\infty ||K||_\infty.$$

Введем необходимые обозначения. Пусть $A_{2n\times 9}$ — матрица, составленная из матриц A_i, построенных по точным соответствиям $X_i \leftrightarrow X_i'$ $(i = 1, \ldots, n)$ согласно формуле (12.10). Обозначим $B_{9\times 9} = A^\top A$. По утверждению 12.2 матрица B имеет ранг 8. Докажем вспомогательную

Лемма 12.13. *Пусть* $\mathbf{w}$ *— единичный вектор, перпендикулярный нулевому подпространству симметричной, неотрицательно определенной матрицы* B. *Упорядочим собственные значения матрицы* B: $\lambda_1(B) = 0 < \lambda_2(B) \leqslant \ldots \leqslant \lambda_9(B)$. *Тогда справедливо неравенство*:

$$\|B\mathbf{w}\| \geqslant \lambda_2(B).$$

Доказательство. Найдем вектор $\mathbf{w}_0$, на котором достигается минимум функции $\|B\mathbf{w}\|$ на 7-мерной сфере, задаваемой уравнениями: $\|\mathbf{w}\| = 1$ и $\langle \mathbf{w}, \mathbf{h}\rangle = 0$ (где $\mathbf{h}$ — вектор из нулевого подпространства матрицы B). Воспользуемся методом множителей Лагранжа. Составим функцию Лагранжа:

$$F(w_1, \ldots, w_9, \alpha, \beta) = \|B\mathbf{w}\|^2 - \frac{1}{2}\alpha\big(\|\mathbf{w}\|^2 - 1\big) - \beta\langle \mathbf{w}, \mathbf{h}\rangle.$$

Проводя стандартные вычисления производных функции Лагранжа, получим необходимое условие на $\mathbf{w}_0$:

$$\begin{cases} B^2\mathbf{w}_0 = \alpha\mathbf{w}_0 + \beta\mathbf{h}, \\ \langle \mathbf{w}_0, \mathbf{h}\rangle = 0, \quad \|\mathbf{w}_0\| = 1. \end{cases} \tag{12.15}$$

Раскладывая вектор $\mathbf{w}_0$ по базису из собственных векторов матрицы B^2 (который существует в силу симметричности матрицы B^2), и подставляя это разложение в первое равенство системы (12.15), получаем, что $\beta = 0$. Следовательно, $\mathbf{w}_0$ — собственный вектор матрицы B^2, отвечающий некоторому собственному значению $\lambda_i(B^2)$. Имеем

$$\|B\mathbf{w}_0\| = \sqrt{\langle B\mathbf{w}_0, B\mathbf{w}_0\rangle} = \sqrt{\langle B^2\mathbf{w}_0, \mathbf{w}_0\rangle} = \sqrt{\lambda_i(B^2)}\|\mathbf{w}_0\| = \lambda_i(B).$$

Отсюда следует, что $\|B\mathbf{w}\| \geqslant \lambda_2(B)$. А поскольку это неравенство достигается на нормированном собственном векторе, соответствующем $\lambda_2(B)$, то минимум функции $\|B\mathbf{w}\|$ в точности равен $\lambda_2(B)$. Лемма доказана. ■

Пусть имеются координаты точек на M и M', которые измерены с некоторой ошибкой: $\widetilde{X}_i = (\widetilde{x}_i, \widetilde{y}_i, 1) = (x_i + \Delta x_i, y_i + \Delta y_i, 1)$, $\widetilde{X}'_i = (\widetilde{x}'_i, \widetilde{y}'_i, 1) = (x'_i + \Delta x'_i, y'_i + \Delta y'_i, 1)$. Следуя ПЛА, составляем матрицу $\widetilde{A}_{2n\times 9}$. Пусть $\widetilde{B}_{9\times 9} = \widetilde{A}^\top\widetilde{A}$. Обозначим $\Delta B = \widetilde{B} - B$ и $\Delta A = \widetilde{A} - A$.

Лемма 12.14. $\|\Delta B\|_\infty \leqslant 2n(2\|A\|_\infty\|\Delta A\|_\infty + \|\Delta A\|_\infty^2)$.

Доказательство. По определению, имеем

$$\widetilde{B} = \widetilde{A}^\top\widetilde{A} = (A + \Delta A)^\top(A + \Delta A) = A^\top A + \Delta A^\top A + A^\top\Delta A + \Delta A^\top\Delta A.$$

Следовательно,

$$\Delta B = \widetilde{B} - B = \widetilde{B} - A^\top A = \Delta A^\top A + A^\top\Delta A + \Delta A^\top\Delta A.$$

Отсюда получаем

$$\|\Delta B\|_\infty \leqslant \|\Delta A^\top A\|_\infty + \|A^\top\Delta A\|_\infty + \|\Delta A^\top\Delta A\|_\infty. \tag{12.16}$$

Заметим, что для любого n-мерного вектора выполнено неравенство: $\frac{1}{n}||\mathbf{x}||_1 \leqslant ||\mathbf{x}||_\infty \leqslant ||\mathbf{x}||_1$. Отсюда и из леммы 12.11 получаем

$$||A^\top \Delta A||_\infty = \sup_{\neq 0} \frac{||A^\top \Delta A\mathbf{x}||_\infty}{||\mathbf{x}||_\infty} = \sup_{\neq 0} \frac{||A^\top \Delta A\mathbf{x}||_\infty}{||\Delta A\mathbf{x}||_\infty} \frac{||\Delta A\mathbf{x}||_\infty}{||\mathbf{x}||_\infty} \leqslant$$
$$\leqslant ||\Delta A||_\infty \sup_{\neq 0} \frac{||A^\top \mathbf{y}||_\infty}{||\mathbf{y}||_\infty} \leqslant ||\Delta A||_\infty \sup_{\neq 0} \frac{2n||A^\top \mathbf{y}||_1}{||\mathbf{y}||_1} =$$
$$= 2n||\Delta A||_\infty ||A^\top||_1 = 2n||\Delta A||_\infty ||A||_\infty.$$

Из последнего неравенства и из (12.16) следует, что

$$||\Delta B||_\infty \leqslant 2n\big(2||A||_\infty ||\Delta A||_\infty + ||\Delta A||_\infty^2\big). \qquad \blacksquare$$

Теперь оценим сверху $||A||_\infty$ и $||\Delta A||_\infty$. Эти оценки даются следующими двумя леммами.

Лемма 12.15. *Пусть для невозмущенных координат точек $X_i = (x_i, y_i, 1)$ и $X'_i = (x'_i, y'_i, 1)$ выполнено условие: $|x_i| < C$, $|y_i| < C$ и $|x'_i| < C'$, $|y'_i| < C'$. Тогда*

$$||A||_\infty \leqslant 2C + 1 + 2C'C + C'.$$

Доказательство. По определению, матрица $A_{2n\times 9}$ состоит из матриц-блоков A_i размера (2×9), определенных по формуле (12.10). Подставим в формулу (12.10) координаты невозмущенных точек $X_i = (x_i, y_i, 1)$ и $X'_i = (x'_i, y'_i, 1)$:

$$A_i = \begin{pmatrix} 0 & 0 & 0 & -x_i & -y_i & -1 & y'_i x_i & y'_i y_i & y'_i \\ x_i & y_i & 1 & 0 & 0 & 0 & -x'_i x_i & -x'_i y_i & -x'_i \end{pmatrix}.$$

Отсюда непосредственно следует оценка:

$$||A||_\infty \leqslant \max_{i=1,\dots,n} \Big\{ \big(|x_i| + |y_i| + 1\big)\big(1 + \max(|x'_i|, |y'_i|)\big) \Big\}.$$

Следовательно,

$$||A||_\infty \leqslant 2C + 1 + 2C'C + C'. \qquad \blacksquare$$

Лемма 12.16. *Пусть для невозмущенных координат точек $X_i = (x_i, y_i, 1)$ и $X'_i = (x'_i, y'_i, 1)$ выполнено условие: $|x_i| < C$, $|y_i| < C$ и $|x'_i| < C'$, $|y'_i| < C'$. Пусть $\varepsilon = \max\limits_i(\Delta x_i, \Delta y_i, \Delta x'_i, \Delta y'_i)$, где $\Delta x_i = |x_i - \widetilde{x}_i|$, $\Delta y_i = |y_i - \widetilde{y}_i|$, $\Delta x'_i = |x'_i - \widetilde{x}'_i|$, $\Delta y'_i = |y'_i - \widetilde{y}'_i|$. Тогда*

$$||\Delta A||_\infty \leqslant \varepsilon(2C + 2C' + 3 + 2\varepsilon).$$

Доказательство. По условию, имеем: $\Delta A_i = \widetilde{A}_i - A_i$. Выпишем соответствующие (транспонированные) матрицы:

$$\widetilde{A}_i^\top = \begin{pmatrix} 0 & x_i + \Delta x_i \\ 0 & y_i + \Delta y_i \\ 0 & 1 \\ -x_i - \Delta x_i & 0 \\ -y_i - \Delta y_i & 0 \\ -1 & 0 \\ (y_i' + \Delta y_i')(x_i + \Delta x_i) & -(x_i' + \Delta x_i')(x_i + \Delta x_i) \\ (y_i' + \Delta y_i')(y_i + \Delta y_i) & -(x_i' + \Delta x_i')(y_i + \Delta y_i) \\ y_i' + \Delta y_i' & -x_i' - \Delta x_i' \end{pmatrix},$$

$$\Delta A_i^\top = \begin{pmatrix} 0 & \Delta x_i \\ 0 & \Delta y_i \\ 0 & 0 \\ -\Delta x_i & 0 \\ -\Delta y_i & 0 \\ 0 & 0 \\ \Delta y_i' x_i + \Delta x_i y_i' + \Delta x_i \Delta y_i' & -\Delta x_i' x_i - \Delta x_i x_i' - \Delta x_i' \Delta x_i \\ \Delta y_i y_i' + \Delta y_i' y_i + \Delta y_i' \Delta y_i & -\Delta x_i' y_i - \Delta y_i x_i' - \Delta x_i' \Delta y_i \\ \Delta y_i' & -\Delta x_i' \end{pmatrix}.$$

Отсюда непосредственно получаем оценку

$$\|\Delta A\|_\infty \leqslant \varepsilon \max_{i=1,\dots,n} \bigl\{3 + |x_i| + |y_i| + 2\varepsilon + 2\max(|x_i'|, |y_i'|)\bigr\}.$$

Окончательно имеем

$$\|\Delta A\|_\infty \leqslant \varepsilon(2C + 2C' + 3 + 2\varepsilon). \qquad \blacksquare$$

Обозначим

$$K = 2n\bigl(2(2C+1)(C'+1)(2C+2C'+3+0,2) + 0,1(2C+2C'+3+0,2)^2\bigr).$$

Заметим, что при $C = C' = \sqrt{2}$, $K = 343n$. Тогда при условии, что $\varepsilon < 0,1$, из последних трех предложений будем иметь

$$\|\Delta B\|_\infty \leqslant \varepsilon K.$$

Пусть $\mathbf{h}$ — нормированный собственный вектор матрицы B, соответствующий нулевому собственному значению, $\widetilde{\mathbf{h}}$ — нормированный собственный вектор матрицы $\widetilde{B}$, соответствующий ее минимальному собственному значению. Обозначим $\widetilde{\mathbf{h}} = \mathbf{h} + \Delta\mathbf{h}$. Имеет место

Теорема 12.8. *Если ошибки в измерениях столь малы, что*

$$\varepsilon \leqslant \min\left\{\frac{\lambda_2(B)}{12K}; 0,1\right\},$$

то справедлива оценка:

$$\|\Delta\mathbf{h}\| \leqslant \frac{12\varepsilon K}{\sqrt{\lambda_2^2(B) - (12\varepsilon K)^2}}.$$

Доказательство. Рассмотрим вектор $\Delta\mathbf{h}' \perp \mathbf{h}$ такой, что вектор $\mathbf{h} + \Delta\mathbf{h}'$ коллинеарен вектору $\widetilde{\mathbf{h}}$. Если α — угол между $\mathbf{h}$ и $\widetilde{\mathbf{h}}$, то $\|\Delta\mathbf{h}'\| = \operatorname{tg}\alpha$, $\|\Delta\mathbf{h}\| = 2\sin\dfrac{\alpha}{2}$. Тогда имеем

$$\|\Delta\mathbf{h}\| \leqslant \|\Delta\mathbf{h}'\|, \qquad \|\mathbf{h} + \Delta\mathbf{h}'\| = \sqrt{1 + \|\Delta\mathbf{h}'\|^2}. \tag{12.17}$$

Пусть $\lambda_{\min}(\widetilde{B}) = \widetilde{\lambda}$ и $\Delta\lambda = \lambda_{\min}(\widetilde{B}) - \lambda_{\min}(B) = \widetilde{\lambda}$. Поскольку вектор $\mathbf{h}+\Delta\mathbf{h}'$ коллинеарен вектору $\widetilde{\mathbf{h}}$, являющемуся собственным вектором матрицы $\widetilde{B}$, то

$$(\widetilde{B} - \widetilde{\lambda}E)(\mathbf{h} + \Delta\mathbf{h}') = 0.$$

Преобразуем последнее равенство:

$$(B + \Delta B - \Delta\lambda E)(\mathbf{h} + \Delta\mathbf{h}') = 0 \Leftrightarrow (B + (\Delta B - \Delta\lambda E))(\mathbf{h} + \Delta\mathbf{h}') = 0 \Leftrightarrow$$
$$\Leftrightarrow B\Delta\mathbf{h}' = -(\Delta B - \Delta\lambda E)(\mathbf{h} + \Delta\mathbf{h}').$$

Отсюда и из (12.17) получаем неравенство:

$$\|B\Delta\mathbf{h}'\|^2 \leqslant \|\Delta B - \Delta\lambda E\|_2^2 \cdot \|\mathbf{h} + \Delta\mathbf{h}'\|^2 =$$
$$= \big(\|\Delta B\|_2 + |\Delta\lambda|\big)^2\big(1 + \|\Delta\mathbf{h}'\|^2\big). \tag{12.18}$$

Применим лемму 12.12 к определенной выше симметричной — а следовательно, диагонализируемой — матрице B. Пусть B представлена в виде $O\Lambda O^{-1}$, где Λ — диагональная, а O — ортогональная матрица. Тогда по лемме 12.12 имеем оценку:

$$|\Delta\lambda| \leqslant \|O^{-1}\|_\infty \|O\|_\infty \|\Delta B\|_\infty. \tag{12.19}$$

Докажем вспомогательную

Лемма 12.17. *Для произвольной ортогональной матрицы $O_{n\times n}$ имеет место неравенство*: $\|O\|_\infty \leqslant \sqrt{n}$.

Доказательство. Найдем $\max\big(|a_1| + \ldots + |a_n|\big)$ при условии $\sum\limits_{i=1}^{n} a_i^2 = 1$. Воспользуемся методом множителей Лагранжа. Функция Лагранжа будет иметь вид

$$F(a_1, \ldots, a_n, \lambda) = a_1 + \ldots + a_n - \lambda(a_1^2 + \ldots + a_n^2).$$

Отсюда получаем условия на условный экстремум:

$$F'_{a_i} = 1 - 2\lambda a_i = 0.$$

Следовательно, условный экстремум достигается при $a_1 = \ldots = a_n = \dfrac{1}{\sqrt{n}}$ и $\max\big(|a_1| + \ldots + |a_n|\big) = \sqrt{n}$. Лемма доказана. ■

Применяя эту лемму к правой части неравенства (12.19), получаем, что $|\Delta\lambda| \leqslant 9\|\Delta B\|_\infty$. Отсюда и из неравенства (12.18) следует, что

$$\|B\Delta\mathbf{h}'\|^2 \leqslant \big(\|\Delta B\|_2 + 9\|\Delta B\|_\infty\big)^2\big(1 + \|\Delta\mathbf{h}'\|^2\big). \tag{12.20}$$

Докажем еще одну вспомогательную

Лемма 12.18. *Для определенных выше норм матрицы $A_{n\times n}$ имеют место оценки*:

$$\frac{1}{\sqrt{n}}\|A\|_\infty \leqslant \|A\|_2 \leqslant \sqrt{n}\|A\|_\infty.$$

Доказательство. Поскольку для любого $\mathbf{x}$: $\|\mathbf{x}\|_\infty \leqslant \|\mathbf{x}\| \leqslant \sqrt{n}\|\mathbf{x}\|_\infty$, то

$$\frac{1}{\sqrt{n}}\|A\|_\infty = \frac{1}{\sqrt{n}}\sup_{\neq 0}\frac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} \leqslant \|A\|_2 = \sup_{\neq 0}\frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \leqslant$$
$$\leqslant \sqrt{n}\sup_{\neq 0}\frac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \sqrt{n}\|A\|_\infty. \qquad \blacksquare$$

Из леммы 12.17 и из (12.20) следует, что

$$\|B\Delta\mathbf{h}'\|^2 \leqslant \big(12\|\Delta B\|_\infty\big)^2\big(1 + \|\Delta\mathbf{h}'\|^2\big).$$

Но поскольку $\|\Delta B\|_\infty \leqslant \varepsilon K$, то

$$\|B\Delta\mathbf{h}'\|^2 \leqslant \big(12\varepsilon K\big)^2\big(1 + \|\Delta\mathbf{h}'\|^2\big). \tag{12.21}$$

Обозначим $\delta = \dfrac{\Delta\mathbf{h}'}{\|\Delta\mathbf{h}'\|}$. Тогда из леммы 12.13 следует неравенство:

$$\|B\Delta\mathbf{h}'\| = \|B\delta\| \cdot \|\Delta\mathbf{h}'\| \geqslant \lambda_2(B)\|\Delta\mathbf{h}'\|. \tag{12.22}$$

Из (12.21) и (12.22) получаем неравенство:

$$\lambda_2^2(B)\|\Delta\mathbf{h}'\|^2 \leqslant \big(12\varepsilon K\big)^2\big(1 + \|\Delta\mathbf{h}'\|^2\big).$$

Или:

$$\|\Delta\mathbf{h}'\|^2\big(\lambda_2^2(B) - (12\varepsilon K)^2\big) \leqslant (12\varepsilon K)^2.$$

Поскольку, по условию, в левой части последнего неравенства стоит положительная величина, то:

$$\|\Delta\mathbf{h}'\| \leqslant \frac{12\varepsilon K}{\sqrt{\lambda_2^2(B) - (12\varepsilon K)^2}}.$$

Остается применить неравенство (12.17) для получения искомой оценки. Теорема доказана. $\blacksquare$

Основные обозначения, принятые в книге

$\mathbb{N}$ — множество натуральных чисел
$\mathbb{Z}$ — множество целых чисел
$\mathbb{R}$ — множество действительных чисел
$\mathbb{C}$ — множество комплексных чисел
$\emptyset$ — пустое множество
$\langle a, b\rangle$ или $a \cdot b$ — скалярное произведение векторов a, b
$a \times b$ — векторное произведение векторов a, b
(a, b, c) — смешанное произведение векторов a, b, c
$x \in A$ — элемент x принадлежит множеству A
$A \subset B$ — множество A содержится в множестве B
$\overline{A}$ — замыкание множества
$A \cup B$ — объединение множеств A и B
$A \cap B$ — пересечение множеств A и B
$A \setminus B$ — вычитание из множества A множество B
$\operatorname{supp} f$ — носитель функции f
C^k — множество непрерывных функций, которые имеют непрерывные до порядка k производные
$f \circ g$ — композиция отображений g и f (сначала применяется отображение g, а потом f)
$\blacksquare$ — конец доказательства

Библиография

[1] Александров А. Д. Внутренняя геометрия выпуклых поверхностей. —М.: Гостехиздат, 1948.

[2] Бляшке В. Введение в дифференциальную геометрию. — М.: Мир, 1957.

[3] Бишоп Р., Криттенден Р. Геометрия многообразий. — М.: Мир, 1967.

[4] Богачев К. Ю. Практикум на ЭВМ. Методы решения линейных систем и нахождения собственных значений. — Москва, 1999.

[5] Боресков А. В., Шикин Е. В. Компьютерная графика. Полигональные модели. — Москва, "Диалог–МИФИ", 2001.

[6] Буземан Г. Геометрия геодезических. — М.: Физматгиз, 1962.

[7] Васильев А. М., Соловьев Ю. П. Дифференциальная геометрия (методические указания). — М.: Изд-во Моск. ун-та, 2004.

[8] Вен Дем А., Фоли Дж. Основы интерактивной машинной графики. Т. 1, Т. 2. — М.: Мир, 1985, (James D. Foley, Andries Van Dam. Fundamentals Interactive Computer Graphics. — Addison-Wesley Publishing Company Reading, Massachusetts, Menlo Park, California, London, Amsterdam, Don Mills, Ontario, Sydney, 1982).

[9] Гильберт Д., Кон-Фоссен С. Наглядная геометрия. — М.: Гостехиздат, 1961.

[10] Громол Д., Клингенберг В., Мейер В. Риманова геометрия в целом. — М.: Мир, 1971.

[11] Дубровин Б. А., Новиков С. П., Фоменко А. Т. Современная геометрия. — М.: Наука, 1986.

[12] Ефимов Н. В. Высшая геометрия. — М.: Наука, 1971.

[13] Зейферт Т., Трельфалль В. Топология. — М.; Л.: ГОНТИ, 1938.

[14] Зейферт Т., Трельфалль В. Вариационное исчисление в целом. — М.: ИЛ, 1947.

[15] Кованцов Н. И. и др. Дифференциальная геометрия, топология, тензорный анализ. — Киев, 1989.

[16] Ласло М. Вычислительная геометрия и компьютерная графика на C++. — М.: Бином, 1997, (Laszlo M.J. Computational Geometry and Computer Graphics in C++. — Prentice Hall, 1996).

[17] Лефшец С. Алгебраическая топология. — М.: ИЛ, 1949.

[18] Мантуров О. В. Элементы тензорного анализа. — М.: Просвещение, 1991.

[19] Милнор Дж. Теория Морса. — М.: Мир, 1965.

[20] Мищенко А. С., Соловьев Ю. П., Фоменко А. Т. Сборник задач по дифференциальной геометрии и топологии. — М.: Изд-во физ.-мат. лит., 2004.

[21] Мищенко А. С., Фоменко А. Т. Курс дифференциальной геометрии и топологии. — М.: Факториал Пресс, 2000.

[22] Новиков С. П., Фоменко А. Т. Элементы дифференциальной геометрии и топологии. — М.: Наука, 1987.

[23] Номидзу К. Группа Ли и дифференциальная геометрия. —М.: ИЛ, 1960.

[24] Норден А. П. Теория поверхностей. — М.: Гостехиздат, 1952.

[25] Носовский Г. В., Фоменко А. Т. Об определении исходных структур в перемешанных последовательностях // Труды семинара по векторному и тензорному анализу с их приложениями к геометрии, механике и физике. Вып. 22. Изд-во Московского университета. — М., 1985. Сс. 119–131.

[26] Носовский Г. В., Фоменко А. Т. Некоторые методы и результаты анализа перемешанных последовательностях // Труды семинара по векторному и тензорному анализу с их приложениями к геометрии, механике и физике. Вып. 23. Изд-во Московского университета. — М., 1988. Сс. 104–121.

[27] Плисс А. И., Шикин Е. В. Кривые и поверхности на экране компьютера. Руководство по сплайнам для пользователя. — Москва, “Диалог–МИФИ”, 2001.

[28] Погорелов А. В. Дифференциальная геометрия. —М.: Наука, 1974.

[29] Погорелов А. В. Внешняя геометрия выпуклых поверхностей. —М.: Наука, 1969.

[30] Понтрягин Л. С. Гладкие многообразия и их применения в теории гомотопий. —М.: Наука, 1985.

[31] Понтрягин Л. С. Непрерывные группы. —М.: Наука, 1984.

[32] Препарата Ф., Шеймос М. Вычислительная геометрия. — М.: Мир, 1989, (Preparata F.P., Shamos M. Computational Geometry: An Introduction. — Apringer-Vertag. New York Berlin Tokyo, 1985).

[33] Рашевский П. К. Риманова геометрия и тензорный анализ. — М.: Наука, 1967.

[34] Рашевский П. К. Курс дифференциальной геометрии. — М.: Гостехиздат, 1956.

[35] Розендорн Э. Р. Задачи по дифференциальной геометрии. — М.: Наука. 1971.

[36] Рохлин В. А., Фукс Д. Б. Начальный курс топологии (геометрический главы). — М.: Наука. 1977.

[37] Спрингер Дж. Введение в теорию римановых поверхностей. — М.: ИЛ, 1960.

[38] Стернберг С. Лекции по дифференциальной геометрии. — М.: Мир. 1970.

[39] Стинрод Н. Топология косых произведений. — М.: ИЛ, 1953.

[40] Торп Дж. Начальные главы дифференциальной геометрии. — М.: Мир. 1982.

[41] Фиников С. П. Курс дифференциальной геометрии. — М.: Гостехиздат, 1952.

[42] “Фотограмметрия”. Ред. Кель Л. Н. и др. — М.: Недра. 1989.

[43] Хартсхорн Р. Основы проективной геометрии. — М.: Мир. 1970.

[44] Хелгасон С. Дифференциальная геометрия и симметрические пространства. — М.: Мир, 1964.

[45] Хирш М. Дифференциальная топология. — М.: Мир. 1979.

[46] Хорн Б. П. Зрение роботов. — М.: Мир. 1989.

[47] Шварц Дж. Дифференциальная геометрия и топология. — М.: Мир, 1970.

[48] Christoph M. Hoffman. Geometric and Solid Modeling. — Morgan Kaufmann Publishers Inc., San ateo, California, 1989.

[49] Glenn A. Kramer. Solving Geometric Constraint System. — The MIT Press, Cambridge, Massachusetts, London, England, 1992.

[50] Cyberworlds 2004. Proceedings of the International Conference on Cyberworlds 2004, November 18–20, 2004, Tokyo, Japan. — IEEE Computer Society, USA, 2004.

[51] Hartley R., Zisserman A. Multiple View Geometry in Computer Vision. — Cambridge Univ. Press, 2000.

[52] Hierarchical and Geometrical Methods in Scientific Visualisation. — Springer–Verlag, Berlin, Heidelberg, New York, 2003.

[53] Hosaka M. Modelling of Curves and Surfaces in CAD/CAM. — Springer–Verlag, Berlin, Heidelberg, 1992.

[54] Hosaka M. Theory of curve and Surface synthesis and their smooth fitting // J. IPS Japan 10(3): 121–123, 1969.

[55] Level of Detail for 3D Graphics. David Luebke, Martin Reddy, Jonathan D. Cohen, Amitabh Varshney, Benjamin Watson, Robert Huebner. — Morgan Kaufmann Publishers, Amsterdam, Boston, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo, 2003.

[56] Nosovskiy G. V. Computer Gluing of 2D Projective Images. — “Contemporary Geometry and Related Topics. Proceedings of the Workshop in Belgrade, Yougoslavia, 15–21 May 2002.” World Scientific, New Jersey, London, Singapore, 2004.

Предметный указатель

B-кривая, 268, 280
    неоднородная, 280
B-сплайн
    ненормированный, 275
    нормированный, 275

Абсолют, 42
Атлас
    карт, 107
    ориентированный, 139
Аффинная карта, 313

Базис Бернштейна, 234
Бинормаль, 56
Бутылка Клейна, 141

Вектор
    времениподобный, 30
    изотропный, 30
    пространственноподобный, 30
    световой, 30
    скорости кривой, 22, 125
Вложение, 132
Внешнее умножение, 168
Внешняя дифференциальная форма, 169
Внутренность множества, 86, 87
Вторая квадратичная форма, 66

Геодезическая, 187
Геометрия
    Лобачевского, 42
    гиперболическая, 42
    эллиптическая, 40
Главная нормаль, 52, 56
Главные
    кривизны, 71
    направления, 68
    оси, 68
Гомеоморфизм, 15, 89
Гомеоморфные пространства, 89
Гомотопическая эквивалентность, 91
Гомотопия, 91
Гомотопные отображения, 91
Градиент внешней дифференциальной формы, 196
Группа
    изометрий, 154
    специальная линейная, 119

Движение, 65, 153
Декартово произведение, 90
Диффеоморфизм, 114
Диффеоморфные многообразия, 114
Дифференциал, 129
Дифференцирование функции по направлению вектора, 126
Длина кривой, 22, 27, 30

Замыкание множества, 86, 87

Изометрия, 65, 153
Имя, 315
Индекс
    ковариантный, 159
    контравариантный, 159

Каноническая камера, 340
Канонический базис, 62
Касательная плоскость к поверхности, 62
Касательное пространство к многообразию, 125

Касательный вектор
кривой, 22, 125
многообразия, 124
поверхности, 62, 122
Катеноид, 81
Квадратичная кривая Безье, 236
Ковариантное дифференцирование, 178
вдоль кривой, 186
Координаты
аффинные, 322
барицентрические, 308
декартовы, 13
евклидовы, 181
касательного вектора, 123, 124
конформные, 83
криволинейные, 16
неоднородные, 313, 322
однородные, 313, 321
полярные, 13
прямой, 334
сферические, 20
точки, 15
цилиндрические, 19
Координаты на поверхности, 60
Кривая
Безье, 235
бирегулярная, 56
гладкая, 22
регулярная, 51
Кривизна, 52, 56
гауссова, 67
нормального сечения, 70
средняя, 67
Кручение, 56
Кубическая кривая Безье, 236
Кубическая поверхность Кунса, 299, 300

Линейная поверхность Кунса, 297
Линия
ломаная, 221
уровня, 116
Локсодрома, 35

Матрица Якоби, 15
Матрица финитной проективной камеры, 332
Метрика, 85
Лобачевского, 44
евклидова, 28
индефинитная, 29
индуцированная, 32
индуцированная на поверхности, 62
конформная, 34
конформно-евклидовая, 63
риманова, 27, 135
Мировая линия, 30
Многообразие, 107
вещественно аналитическое, 112
гладкое, 111, 112
замкнутое, 138
комплексно аналитическое, 113
ориентируемое, 139
с краем, 138
уровня, 116
Многоугольник нулевого уровня, 243
Множество
замкнутое, 86, 87
открытое, 86, 87
Модель Пуанкаре, 42

Натуральная параметризация кривой, 51
Натуральное уравнение, 57
Нормализация, 349
Носитель функции, 102

Область, 14, 60
фундаментальная, 141
Обобщенная поверхности Гордона, 304
Обобщенная поверхность Кунса, 299
Объем области, 171
Окрестность
множества, 86, 87
точки, 86, 87
Операторное поле
кососимметричное, 167

симметричное , 167
Операция
дифференцирования, 127
рационального деления отрезка, 253
Ось камеры, 329
Отображение
гладкое, 114
непрерывное, 88
непрерывное в точке, 88, 90

Параллельное векторное поле, 186
Параллельный вектор, 186
Параметр
местный, 221
натуральный, 51
Параметрическая длина, 221
Первая квадратичная форма поверхности, 62
Первая фундаментальная форма поверхности, 62
Плоскость проектирования, 329
Площадь поверхности, 63
Поверхность, 60
Безье, 247
Бельтрами, 76
Гордона, 304
Лагранжа, 303
Эрмита, 301
билинейная, 295
вложенная, 62
гармоническая, 83
гладкая, 60
линейчатая, 295
минимальная, 80
невырожденная, 122
нулевой кривизны, 75
отрицательной кривизны, 75
перехода, 302
положительной кривизны, 75
постоянной кривизны, 75
регулярная, 60
секториальная, 295
тензорного произведения, 306
треугольная, 296
уровня, 116
Погружение, 132
Подмногообразие, 132
Покрытие
открытое, 96
Полюс, 334
Полярная линия, 333
Преобразование
дробно-линейное, 46
конформное, 35
проективное, 314
Признак, 315
Проективная кривая Безье, 255
Проективная плоскость, 313
Проективное пространство, 313
Производная функции по направлению вектора, 126
Пространство
Минковского, 29
евклидово, 13
компактное, 97
линейно связное, 93
метрическое, 86
нормальное, 100
связное, 94
стягиваемое, 91
топологическое, 87
хаусдорфово, 95
Прямой геликоид, 81
Псевдосфера, 37, 76
Псевдоупругая энергия, 225

Равномерная параметризация, 221
Радиус кривизны, 52, 56
Радиус-вектор
гармонический, 83
минимальный, 83
Разбиение единицы, 103
Разделенная разность, 269
Размерность многообразия, 107
Разнесение, 316
Расстояние между множествами, 86
Расширенное множество узлов, 280
Рациональная кривая Безье, 253, 255

Рациональная поверхность Безье с весами, 261
Рациональное деление отрезка, 253
Репер Френе, 56
Род поверхности, 152

Связанные имена, 316
Связанные ячейки, 316
Связность
аффинная, 179
риманова, 181
симметричная, 179
согласованная с метрикой, 181
Сетка, 303
Сечение
главное нормальное, 71
нормальное, 70
плоское, 69
Система координат
локальная, 17
непрерывная, 15
регулярная , 15
Система точек общего положения, 322
Сопряженные точки, 312, 315
Сплайн, 221
Лагранжа, 230
Ньютона, 231
Эрмита, 221
Стереографическая проекция, 33
Сходимость
последовательности, 90
равномерная, 100

Тензор
деформаций, 162
кручения аффинной связности, 179
метрический, 162
моментов инерции, 162
напряжений, 163
Тензорное поле, 159
кососимметричное, 166
симметричное, 166
Тензорной закон преобразования координат, 124
Топология, 87
Тор, 121
Точка
внутренняя, 86, 87, 138
граничная, 138
критическая, 72
опорная, 221
прикосновения, 86, 87
регулярная, 131, 132
характеристическая, 221
Точки общего положения, 314

Узел, 274
Узел сплайна, 221
Узловое значение, 221
Уравнения геодезических, 188
Усеченная степенная функция, 274

Факторпространство, 92
Фактортопология, 92
Финитная камера, 332
Фокусное расстояние, 329
Формула
Гаусса–Остроградского, 205
Грина, 205
Ньютона–Лейбница, 206
Стокса, 205
Фундаментальное отображение, 334
Функции
замены координат, 110
перехода, 110
смещения линейной поверхности Кунса, 297
Функция непрерывно дифференцируемая на многообразии, 111

Центр
камеры, 329
проекции, 329

Четырехугольный лист Безье, 248

Эквивалентные атласы, 112
Экран камеры, 329

Якобиан, 15