

Парадигма развития науки

А. Е. Кононюк

Истины и информация

**(Фундаментальная теория
представления истин и информации)**

Книга 6

**Отображение информации
средствами топологии**

**Киев
«Освіта України»**

2016



Кононюк Анатолий Ефимович

Не познав информацию не познаешь истину



Структурная схема парадигмы развития науки



УДК 51 (075.8)
ББК В161.я7
К65

Рецензент:

Н.К.Печурин - д-р техн. наук, проф. (Национальный авиационный университет).

Кононюк А. Е.

К213 Истины и информация. — В 16-и кн. Кн.6. — К.:Освіта України. 2016.—576 с.

ISBN 978-966-373-693-8 (многотомное издание)

ISBN 978-966-373-694-10 (книга 6)

Многотомная работа посвящена систематическому изложению общих формализмов, математических моделей и алгоритмических методов, которые могут быть используемых при моделировании и исследованиях математических моделей истин и информации.

Развиваются представления и методы решения, основанные на теориях эвристического поиска и автоматическом доказательстве теорем, а также процедуральные методы, базирующиеся на классе проблемно-ориентированных языков, сочетающих свойства языков программирования и автоматических решателей задач отображения истин и информации различными математическими средствами.

В работе излагаются основы теории формализованного отображения истин и информации такими средствами математики как: множества, отношения, поверхности, пространства, алгебраические системы, матрицы, графы, математическая логика и др.

Для бакалавров, специалистов, магистров, аспирантов, докторантов всех специальностей.

УДК 51 (075.8)
ББК В161.я7

ISBN 978-966-373-693-8 (многотомное издание) © Кононюк А. Е., 2016
ISBN 978-966-373-694-10 (книга 6) © Освіта України, 2016

Оглавление

Предисловие	7
1. Геометрические объекты как источники визуальной информации..	10
1.1. Описание геометрических объектов.....	10
1.2 Преобразование декартовых прямоугольных координат	13
1.3. Модификации векторов и точек.....	20
1.4. Однородные координаты.....	28
1.5. Геометрия кривых линий.....	32
1.6. Геометрия двухмерных кривых.....	46
1.7 Геометрия поверхностей	50
1.8. Кривизна линий на поверхности	64
1.9. Тензоры поверхности.....	76
1.10. Криволинейные координаты.. ..	90
1.11 Тензоры в криволинейных координатах.....	104
1.12. Ортогональные криволинейные координаты.....	113
1.13. Математическая модель геометрии объектов	116
2. Введение в топологию	118
2.1. Топология линий	118
2.1.1. Идея непрерывности.....	118
2.1.2. Чем занимается топология?	122
2.1.3. Простейшие топологические инварианты.....	125
2.1.4. Эйлерова характеристика графа.....	129
2.1.5. Индекс пересечения.....	133
2.1.6. Теорема Жордана.....	139
2.1.7. Что такое линия?.....	143
2.1.8. Кривая Пеано.....	150
2.2. Топология поверхностей.....	155
2.2.1. Теорема Эйлера.....	155
2.2.2. Поверхности.....	157
2.2.3. Эйлерова характеристика поверхности.....	165
2.2.4. Классификация замкнутых ориентируемых поверхностей.....	169
2.2.5. Классификация замкнутых неориентируемых поверхностей.....	177
2.2.6. Векторные поля на поверхностях.....	185
2.2.7. Проблема четырех красок.....	193
2.2.8. Раскрашивание карт на поверхностях.....	195
2.2.9. «Дикая сфера».....	200
2.2.10. Узлы.....	206
2.2.11. Коэффициент зацепления.....	213
2.3. Гомотопии и гомологии.....	219
2.3.1. Периоды многозначных функций.....	219
2.3.2. Фундаментальная группа.....	224

2.3.3. Клеточные разбиения и полиэдры.....	230
2.3.4. Накрытия.....	235
2.3.5. Степень отображения и основная теорема алгебры.....	242
2.3.6. Группа узла.....	247
2.3.7. Циклы и гомологии.....	253
2.3.8. Топологическое произведение.....	266
2.3.9. Расслоения.....	271
2.3.10. Теория Морса.....	276
2.3.11. Использование топологических методов в физике.....	282
3. ТОПОЛОГИЯ ОБОЛОЧЕК.....	294
3.1. Топологические объекты.....	294
3.2. Эйлерова характеристика оболочек.....	298
3.3. Связность оболочек.....	303
3.4. Ориентируемость оболочек.....	307
3.5. Оболочки для моделирования тел.....	311
3.6. Поверхностное и твердотельное моделирование.....	318
4. МОДЕЛИРОВАНИЕ ТЕЛ.....	319
4.1. Математическая модель тел.....	319
4.2. Простейшие тела.....	325
4.3. Тела, полученные движением плоского контура.....	338
4.4. Построение тела по плоским сечениям.....	348
4.5. Тело в форме листа.....	350
4.6. Булевы операции над телами.....	352
4.7. Резка тела поверхностью.....	374
4.8. Построение симметричного тела.....	377
4.9. Построение эквидистантной оболочки тела.....	382
4.10. Построение тонкостенного тела.....	389
4.11. Скругление ребер тела.....	395
4.12. Построение фасок ребер тела.....	405
4.13. Некоторые способы построения тел.....	407
4.14. Последовательность моделирования тел.....	410
5. ВАРИАЦИОННЫЕ СВЯЗИ ГЕОМЕТРИЧЕСКИХ ОБЪЕКТОВ.....	412
5.1. Наложение вариационных связей.....	412
5.2. Фиксирующие связи.....	417
5.3. Вариационные связи точек в пространстве.....	418
5.4. Вариационные связи точек на кривых и поверхностях.....	429
5.5. Алгебраические связи.....	431
5.6. Минимизация изменения параметров.....	433
5.7. Условный экстремум функции изменения параметров.....	438
5.8. Вариационный метод определения изменений параметров.....	446
5.9. Геодезические линии.....	460
5.10. Вариационные связи двухмерных точек.....	465

5.11. Вариационные связи двухмерных линий	474
5.12. Формирование и решение системы уравнений связей.....	487
6. КОМПЬЮТЕРНАЯ ГРАФИКА	493
6.1. Визуализация геометрических объектов.....	493
6.2. Параллельные проекции на плоскость.....	497
6.3. Центральные проекции на плоскость.....	500
6.4. Полигоны кривых и поверхностей.....	507
6.5. Линии очерка.....	512
6.6. Определение видимой части геометрических объектов.....	520
6.7. Триангуляция поверхностей и тел	525
6.8. Моделирование света.....	546
6.9. Описание цвета.....	554
6.10. Формирование реалистических изображений.....	557
СПИСОК ЛИТЕРАТУРЫ.....	569

ПРЕДИСЛОВИЕ

Простейшие идеи топологии возникают из непосредственного наблюдения за окружающим миром. Интуитивно ясно, что высказывания о геометрических свойствах фигур не вполне исчерпываются сведениями об их «метрических» свойствах (размерах, углах и т. д.). Остается еще «кое-что» за пределами старой геометрии. Какой бы длинной ни была линия (веревка, провод, длинная молекула), она может быть замкнутой или нет; если линия замкнута, то она может сложным образом «заузляться». Две (или более) замкнутые линии могут «зацепляться» одна с другой и притом различными способами. Тела, их поверхности, могут иметь «дырки». Эти свойства тел характеризуются тем, что они не меняются при деформациях, допускающих любые растяжения без разрывов. Такие свойства и называются топологическими. Кроме элементарных геометрических фигур, топологическими свойствами обладают многие чисто математические объекты, и именно это определяет их важность.

Однако легче подметить существование топологических свойств фигур, чем создать их «исчисление», т. е. раздел математики, обладающий точными понятиями, строгими законами и методами, математическими формулами, изображающими топологические величины.

Первые важные наблюдения и точные топологические соотношения были найдены еще Эйлером, Гауссом и Риманом. Тем не менее, без преувеличения можно сказать, что топология как раздел науки основана в конце XIX века А. Пуанкаре. Процесс построения топологии и решения ее внутренних задач оказался трудным и длительным: он продолжался не менее 70—80 лет, наполненных глубокими открытиями и, в ряде случаев, даже пересмотром основ. В нем принял участие ряд наиболее выдающихся математиков своего времени. На протяжении многих лет, приблизительно до конца 50-х годов, топология рассматривалась даже математиками других областей как красивая, но бесполезная игрушка. Автор этих строк должен откровенно признаться, что еще в студенческие времена, в 50-х годах XX века, выбрал себе эту область для будущей деятельности, увлеченный ее красотой и необычностью по сравнению с более традиционными разделами математики, и при этом значительное время (до конца 60-х годов) испытывал неудовлетворенность итогами развития этой области именно из-за недостатка приложений. Следует заметить, что ряд красивых топологических закономерностей в других разделах математики был уже к тому времени обнаружен — в теории функций и комплексном анализе, в качественной теории динамических

систем и уравнений с частными производными, в теории операторов и даже в алгебре.

Однако лишь с начала 70-х годов XX века началось интенсивное проникновение методов топологии в аппарат современной физики. Сейчас важность топологических методов для различных разделов физики уже не вызывает сомнений — для теории поля и общей теории относительности, физики анизотропных сплошных сред и низких температур, современной квантовой теории и т. д.

Предлагаемая книга предназначена для тех, кто изучает, использует в работе или разрабатывает системы геометрического моделирования. Она посвящена математическому аппарату этих систем.

Компьютеры позволяют создавать численные модели различных объектов. С их помощью можно увидеть еще не существующий объект, получить его геометрические характеристики, выполнить исследование его физических свойств путем постановки численных экспериментов, внести необходимые изменения, подготовить производство и, наконец, изготовить объект. Инструментом для всего этого служат CAD/CAM/CAE системы. Общим элементом таких систем является математическая модель геометрии проектируемого объекта.

В работе с системами геометрического моделирования используются различные области знания. Теоретической основой геометрического моделирования являются дифференциальная и аналитическая геометрии, вариационное исчисление, топология и разделы вычислительной математики. Геометрическое моделирование изучает методы построения кривых линий, поверхностей и твердых тел, методы выполнения над ними различных операций и методы управления численными моделями.

1. Геометрические объекты как источники визуальной информации

1.1. Описание геометрических объектов

Геометрические объекты. Нас будет интересовать информация о форме окружающих предметов (истин), их размерах и взаимном расположении, не вдаваясь в подробности физических свойств. Другими словами, мы будем изучать и моделировать информацию о геометрических свойствах реальных или воображаемых объектов (истин). Нашей конечной целью является построение математических моделей геометрии этих объектов (истин). Эти модели нужны для принятия решений, для проведения исследований, для производства материальных ценностей.

Геометрическое моделирование изучает методы построения математической модели, описывающей геометрические свойства предметов окружающего мира. Оно базируется на аналитической и дифференциальной геометрии, вычислительной математике, вариационном исчислении, топологии и разрабатывает собственные математические методы моделирования.

Инструментом для геометрического моделирования служат математические методы решения тех или иных задач. Используемые методы позволяют описывать геометрические свойства предметов, создавать их математические модели и исследовать их путем проведения различных расчетов и численных экспериментов. При необходимости мы сможем редактировать моделируемые объекты и строить их графические отображения.

Для описания геометрических свойств окружающих предметов мы будем строить твердые тела или просто тела. Тела в свою очередь мы будем описывать точками, линиями и поверхностями. Все они обладают определенными общими свойствами, поэтому ими можно оперировать как объектами (истинами). Точки, линии, поверхности и тела будем называть геометрическими объектами (рис. 1.1).

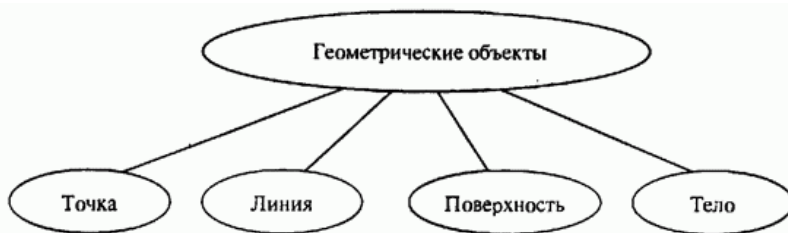


Рис.1.1

Геометрические объекты будут служить основными элементами математической модели геометрии реальных или воображаемых объектов. Мы будем строить их в трехмерном евклидовом пространстве считая их неизменными во времени.

В большинстве случаев мы будем использовать декартовы прямоугольные системы координат. В декартовой системе координат базисные векторы имеют одинаковую длину и постоянное направление в любом месте пространства. Это упрощает описание объектов, так как базисные векторы при дифференцировании выступают в роли констант. Мы рассмотрим также описание геометрических объектов в криволинейных системах координат.

Обозначения. Для количественных характеристик геометрических объектов мы будем использовать скалярные величины, векторы, а также тензоры. Скалярные величины будем обозначать строчными буквами латинского или греческого алфавита. Векторы в пространстве будем обозначать строчными буквами латинского алфавита, выделенными жирным шрифтом. Двухмерные векторы будем обозначать строчными буквами латинского алфавита, выделенными жирным курсивом. Точки будем обозначать прописными латинскими буквами. Тензоры и матрицы будем обозначать прописными буквами латинского алфавита, выделенными жирным шрифтом.

Систему координат с началом в точке O и базисными векторами $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ будем обозначать через $Oe_1e_2e_3$.

Для векторов в пространстве мы будем использовать записи типа

$$\mathbf{r} = r_1\mathbf{e}_1 + r_2\mathbf{e}_2 + r_3\mathbf{e}_3 = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = [r_1 \ r_2 \ r_3]^T,$$

где r_1, r_2, r_3 — компоненты вектора \mathbf{r} . В данной главе нам удобно использовать обозначения компонент векторов с индексами, равными ее номеру. В других главах мы будем также использовать обозначения компонент векторов через x, y, z .

Для описания сложных геометрических объектов нам потребуются векторы как в трехмерном, так и в двухмерном пространстве, например, на области параметров поверхности. Для двухмерных векторов мы будем использовать запись типа

$$\mathbf{p} = p_1\mathbf{e}_1 + p_2\mathbf{e}_2 = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = [p_1 \ p_2]^T,$$

где p_1, p_2 — компоненты вектора \mathbf{p} .

Операцию скалярного произведения векторов будем обозначать точкой: например,

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

Символом \times будем обозначать операцию векторного произведения векторов: например,

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

Запись двух векторов рядом $\mathbf{a}\mathbf{b}$ будет означать операцию диадного произведения векторов \mathbf{a} и \mathbf{b} , например,

$$\mathbf{ab} = \begin{bmatrix} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \\ a_3b_1 & a_3b_2 & a_3b_3 \end{bmatrix}.$$

Точка. Геометрическое моделирование мы начнем изучать с простых объектов, переходя постепенно к более сложным. Точка R пространства в общем случае описывается координатами u^1, u^2, u^3 некоторой системы координат. В декартовой прямоугольной системе координат точку можно описать с помощью радиус-вектора $\mathbf{r} = [r_1, r_2, r_3]^T$. Радиус-вектор определяет преобразование переноса, переводящее начальную точку декартовой системы координат в заданную точку пространства. Компоненты радиус-вектора точки равны ее координатам. Радиус-вектор в отличие от просто вектора связан с началом координат. Эта разница сказывается на формулах преобразования координат и на формулах изменения положения в пространстве.

1.2 Преобразование декартовых прямоугольных координат

Преобразования координат точек в пространстве. Рассмотрим, как изменятся декартовы прямоугольные координаты точки в пространстве при переходе от одной системы координат к другой. Пусть в пространстве заданы две декартовы прямоугольные системы координат: $Oe_1e_2e_3$ и $O_1i_1i_2i_3$

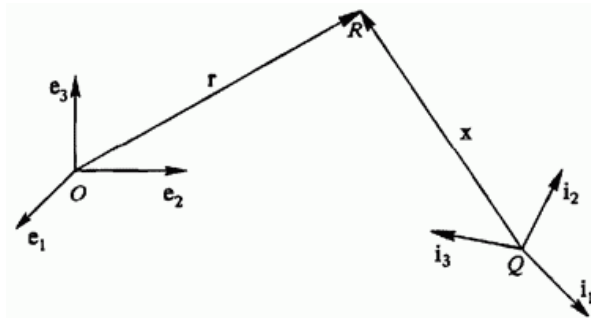


Рис. 2.1

Пусть положение некоторой точки R в первой системе описывается координатами r_1, r_2, r_3 которым соответствует вектор \mathbf{r} , а во второй системе положение этой же точки описывается координатами x_1, x_2, x_3 которым соответствует вектор \mathbf{x} (рис. 2.1).

Заметим, что равенствами

$$\begin{aligned}\mathbf{r} &= r_1 \mathbf{e}_1 + r_2 \mathbf{e}_2 + r_3 \mathbf{e}_3, \\ \mathbf{x} &= x_1 \mathbf{i}_1 + x_2 \mathbf{i}_2 + x_3 \mathbf{i}_3\end{aligned}$$

точка R описывается в разных системах координат. Обозначим через

$$\mathbf{q} = q_1 \mathbf{e}_1 + q_2 \mathbf{e}_2 + q_3 \mathbf{e}_3$$

вектор, направленный из точки O в точку Q. Если к вектору \mathbf{q} добавить вектор \mathbf{x} , то мы получим вектор \mathbf{r} . Сложение векторов мы выполним в первой системе координат, так как именно в ней рассматриваемая точка описывается вектором \mathbf{r} . Для этого разложим орты $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$ по ортам $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$:

$$\begin{aligned}\mathbf{i}_1 &= a_{11} \mathbf{e}_1 + a_{12} \mathbf{e}_2 + a_{13} \mathbf{e}_3, \\ \mathbf{i}_2 &= a_{21} \mathbf{e}_1 + a_{22} \mathbf{e}_2 + a_{23} \mathbf{e}_3, \\ \mathbf{i}_3 &= a_{31} \mathbf{e}_1 + a_{32} \mathbf{e}_2 + a_{33} \mathbf{e}_3,\end{aligned}\quad (2.3)$$

где $a_{ij}, i, j=1, 2, 3$ — компоненты ортов $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$ в системе координат $Oe_1e_2e_3$. Сложив векторы \mathbf{q} и \mathbf{x} получим равенство

$$\begin{aligned}r_1 \mathbf{e}_1 + r_2 \mathbf{e}_2 + r_3 \mathbf{e}_3 &= (a_{11}x_1 + a_{21}x_2 + a_{31}x_3 + q_1)\mathbf{e}_1 + \\ &+ (a_{12}x_1 + a_{22}x_2 + a_{32}x_3 + q_2)\mathbf{e}_2 + (a_{13}x_1 + a_{23}x_2 + a_{33}x_3 + q_3)\mathbf{e}_3,\end{aligned}$$

откуда следуют формулы преобразования координат точки при переходе из системы $Qi_1i_2i_3$ в систему координат $Oe_1e_2e_3$:

$$\begin{aligned}r_1 &= a_{11}x_1 + a_{21}x_2 + a_{31}x_3 + q_1, \\ r_2 &= a_{12}x_1 + a_{22}x_2 + a_{32}x_3 + q_2, \\ r_3 &= a_{13}x_1 + a_{23}x_2 + a_{33}x_3 + q_3.\end{aligned}\quad (2.4)$$

В матричном представлении (2.4) имеет вид

$$\mathbf{r} = \mathbf{A}^T \cdot \mathbf{x} + \mathbf{q}, \quad (2.5)$$

где \mathbf{A}^T — транспонированная матрица поворота:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Строки матрицы \mathbf{A} составлены из компонент векторов $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$. Умножив каждое из равенств (2.3) скалярно на $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ получим еще одно выражение для компонент матрицы \mathbf{A} :

$$\begin{aligned} a_{11} &= \mathbf{i}_1 \cdot \mathbf{e}_1, & a_{12} &= \mathbf{i}_1 \cdot \mathbf{e}_2, & a_{13} &= \mathbf{i}_1 \cdot \mathbf{e}_3, \\ a_{21} &= \mathbf{i}_2 \cdot \mathbf{e}_1, & a_{22} &= \mathbf{i}_2 \cdot \mathbf{e}_2, & a_{23} &= \mathbf{i}_2 \cdot \mathbf{e}_3, \\ a_{31} &= \mathbf{i}_3 \cdot \mathbf{e}_1, & a_{32} &= \mathbf{i}_3 \cdot \mathbf{e}_2, & a_{33} &= \mathbf{i}_3 \cdot \mathbf{e}_3. \end{aligned} \quad (2.6)$$

Из (2.6) видно, что базисные векторы $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ в системе координат с базисными векторами $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$ выражаются через те же коэффициенты, которые присутствуют в (2.3):

$$\begin{aligned} \mathbf{e}_1 &= a_{11}\mathbf{i}_1 + a_{21}\mathbf{i}_2 + a_{31}\mathbf{i}_3, \\ \mathbf{e}_2 &= a_{12}\mathbf{i}_1 + a_{22}\mathbf{i}_2 + a_{32}\mathbf{i}_3, \\ \mathbf{e}_3 &= a_{13}\mathbf{i}_1 + a_{23}\mathbf{i}_2 + a_{33}\mathbf{i}_3. \end{aligned} \quad (2.7)$$

Действительно, если каждое из равенств (2.7) умножить скалярно на $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$, то получим равенства (2.6).

Обозначим через $\mathbf{o} = o_1\mathbf{i}_1 + o_2\mathbf{i}_2 + o_3\mathbf{i}_3$ вектор, направленный из точки Q в точку O . Выполним сложение векторов \mathbf{o} и \mathbf{r} во второй системе координат, мы получим вектор \mathbf{x} . Приравняв соответствующие компоненты векторов, получим формулы преобразования координат

рассматриваемой точки при переходе из системы $Oe_1e_2e_3$ в систему координат $Oq_1q_2q_3$:

$$\begin{aligned}x_1 &= a_{11}r_1 + a_{12}r_2 + a_{13}r_3 + o_1, \\x_2 &= a_{21}r_1 + a_{22}r_2 + a_{23}r_3 + o_2, \\x_3 &= a_{31}r_1 + a_{32}r_2 + a_{33}r_3 + o_3.\end{aligned}\quad (2.8)$$

Если формулы (2.8) применить для точки Q , мы получим формулы, выражающие компоненты o_1, o_2, o_3 вектора \mathbf{o} через компоненты q_1, q_2, q_3 вектора \mathbf{q} :

$$\begin{aligned}o_1 &= -a_{11}q_1 - a_{12}q_2 - a_{13}q_3, \\o_2 &= -a_{21}q_1 - a_{22}q_2 - a_{23}q_3, \\o_3 &= -a_{31}q_1 - a_{32}q_2 - a_{33}q_3.\end{aligned}$$

Подставив последние в (2.8), получим

$$\begin{aligned}x_1 &= a_{11}(r_1 - q_1) + a_{12}(r_2 - q_2) + a_{13}(r_3 - q_3), \\x_2 &= a_{21}(r_1 - q_1) + a_{22}(r_2 - q_2) + a_{23}(r_3 - q_3), \\x_3 &= a_{31}(r_1 - q_1) + a_{32}(r_2 - q_2) + a_{33}(r_3 - q_3).\end{aligned}$$

В матричном представлении преобразования (2.8) имеют вид

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{r} + \mathbf{o} = \mathbf{A} \cdot (\mathbf{r} - \mathbf{q}).\quad (2.9)$$

Умножив равенство (2.9) слева на \mathbf{A}^T и прибавив вектор \mathbf{q} , в соответствии с формулой (2.5) мы должны получить вектор \mathbf{r} . Отсюда следует, что

$$\mathbf{A}^T \cdot \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \equiv \mathbf{E},$$

$$\mathbf{A}^T = \mathbf{A}^{-1}. \quad (2.10)$$

Мы видим, что транспонированная матрица поворота равна своей обратной матрице, т. е. матрица поворота системы координат является ортогональной. Отсюда же следует, что определитель матрицы \mathbf{A} равен единице:

$$|\mathbf{A}^T| = |\mathbf{A}| = 1. \quad (2.11)$$

Преобразования компонент векторов в пространстве. Пусть некоторый пространственный вектор в системе координат $Oe_1e_2e_3$ описывается выражением $r_1e_1 + r_2e_2 + r_3e_3$, а в системе координат $Qj_1j_2j_3$ этот же вектор описывается выражением $x_1j_1 + x_2j_2 + x_3j_3$. С учетом формул (2.3) и (2.7) получим, что преобразования компонент векторов описываются формулами (2.4) и (2.8), в которых компоненты q_1, q_2, q_3 и o_1, o_2, o_3 следует положить равными нулю.

Преобразования координат двумерных точек. Пусть на плоскости заданы две декартовы прямоугольные системы координат: Oe_1e_2 и Qj_1j_2

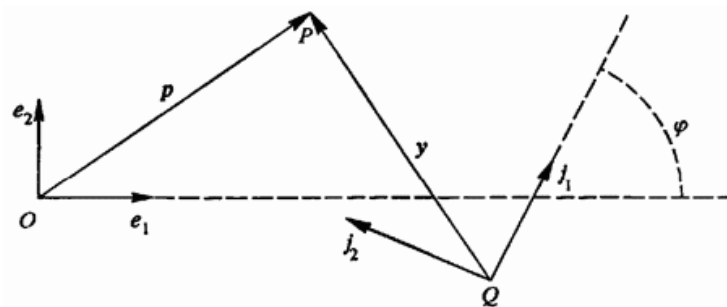


Рис. 2.2

Пусть положение некоторой точки Р в первой системе описывается вектором $\mathbf{p}=p_1\mathbf{e}_1+p_2\mathbf{e}_2$, а во второй системе положение этой же точки описывается вектором $\mathbf{y}=y_1\mathbf{j}_1+y_2\mathbf{j}_2$ (рис. 2.2).

Разложим вектор \mathbf{q} , направленный из точки О в точку Q, а также орты $\mathbf{j}_1, \mathbf{j}_2$ по ортам $\mathbf{e}_1, \mathbf{e}_2$:

$$\begin{aligned}\mathbf{q} &= q_1\mathbf{e}_1 + q_2\mathbf{e}_2, \\ \mathbf{j}_1 &= a_{11}\mathbf{e}_1 + a_{12}\mathbf{e}_2, \\ \mathbf{j}_2 &= a_{21}\mathbf{e}_1 + a_{22}\mathbf{e}_2.\end{aligned}\quad (2.12)-(2.13)$$

Сложив векторы \mathbf{q} и \mathbf{y} , получим

$$p_1\mathbf{e}_1 + p_2\mathbf{e}_2 = (a_{11}y_1 + a_{21}y_2 + q_1)\mathbf{e}_1 + (a_{12}y_1 + a_{22}y_2 + q_2)\mathbf{e}_2.$$

Отсюда следуют формулы преобразования координат точки:

$$\begin{aligned}p_1 &= a_{11}y_1 + a_{21}y_2 + q_1, \\ p_2 &= a_{12}y_1 + a_{22}y_2 + q_2.\end{aligned}\quad (2.14)$$

Матричное представление (2.14) совпадает с (2.5): $\mathbf{p}=\mathbf{A}^T\cdot\mathbf{y}+\mathbf{q}$. Так как базисные векторы \mathbf{j}_1 и \mathbf{j}_2 имеют единичную длину, равенствам (2.13) можно придать вид

$$\begin{aligned}\mathbf{j}_1 &= \cos\varphi\mathbf{e}_1 + \sin\varphi\mathbf{e}_2, \\ \mathbf{j}_2 &= -\sin\varphi\mathbf{e}_1 + \cos\varphi\mathbf{e}_2,\end{aligned}\quad (2.15)$$

если обе координатные системы правые или левые, и

$$\begin{aligned}\mathbf{j}_1 &= \cos\varphi\mathbf{e}_1 + \sin\varphi\mathbf{e}_2, \\ \mathbf{j}_2 &= \sin\varphi\mathbf{e}_1 - \cos\varphi\mathbf{e}_2,\end{aligned}$$

если одна из координатных систем правая, а другая — левая. Угол φ между векторами и \mathbf{e}_1 и \mathbf{j}_1 отсчитывается от \mathbf{e}_1 в сторону вектора \mathbf{e}_2 . Пусть обе системы координат являются правыми (для систем разной

ориентации следует изменить знак a_{21} и a_{22} на противоположный). Матрица поворота систем координат, выраженная через угол φ имеет вид

$$A \equiv \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix}.$$

Преобразование координат (2.14) при переходе от системы с базисными векторами \mathbf{e}_1 и \mathbf{e}_2 к системе координат с базисными векторами \mathbf{j}_1 и \mathbf{j}_2 примет вид

$$\begin{aligned} p_1 &= y_1 \cos \varphi - y_2 \sin \varphi + q_1, \\ p_2 &= y_1 \sin \varphi + y_2 \cos \varphi + q_2. \end{aligned} \quad (2.16)$$

Решив систему уравнений (2.16) относительно y_1 и y_2 получим обратное преобразование

$$\begin{aligned} y_1 &= (p_1 - q_1) \cos \varphi + (p_2 - q_2) \sin \varphi, \\ y_2 &= -(p_1 - q_1) \sin \varphi + (p_2 - q_2) \cos \varphi. \end{aligned} \quad (2.17)$$

Если в (2.17) положить $p_1=0, p_2=0$, то получим координаты точки O в системе с базисными векторами \mathbf{j}_1 и \mathbf{j}_2 , выраженные через компоненты q_1 и q_2

$$\begin{aligned} o_1 &= -q_1 \cos \varphi - q_2 \sin \varphi, \\ o_2 &= q_1 \sin \varphi - q_2 \cos \varphi. \end{aligned} \quad (2.18)$$

В системе координат с базисными векторами \mathbf{j}_1 и \mathbf{j}_2 координаты o_1 и o_2 равны компонентам вектора \mathbf{o} , построенного из точки Q в точку O . С учетом формул (2.18) преобразование координат точки (2.17) будет иметь вид

$$\begin{aligned} y_1 &= a_{11}p_1 + a_{12}p_2 + o_1 = p_1 \cos \varphi + p_2 \sin \varphi + o_1, \\ y_2 &= a_{21}p_1 + a_{22}p_2 + o_2 = -p_1 \sin \varphi + p_2 \cos \varphi + o_2. \end{aligned}$$

Выразим из (2.18) координаты q_1 и q_2 через координаты o_1 и o_2 и подставим их в (2.16). В результате получим

$$\begin{aligned} p_1 &= (y_1 - o_1) \cos \varphi - (y_2 - o_2) \sin \varphi, \\ p_2 &= (y_1 - o_1) \sin \varphi + (y_2 - o_2) \cos \varphi. \end{aligned} \quad (2.19)$$

Матричные записи преобразований координат точки при переходе от одной декартовой прямоугольной системы координат к другой в двумерном пространстве и трехмерном пространстве совпадают.

Преобразования компонент двумерных векторов. Пусть некоторый двумерный вектор в системе координат Oe_1e_2 описывается выражением $p_1 e_1 + p_2 e_2$, а в системе координат Oj_1j_2 этот же вектор описывается выражением $y_1 j_1 + y_2 j_2$. С учетом формул (2.15) получим, что преобразования компонент двумерных векторов описываются формулами (2.16) и (2.17), в которых компоненты q_1, q_2 следует положить равными нулю.

1.3. Модификации векторов и точек

Модификациями будем называть изменения положения и формы геометрических объектов. Многие линии, поверхности и тела описываются определенным образом связанным набором точек, векторов и скаляров. При изменении положения геометрического объекта в пространстве требуется выполнять соответствующие модификации радиус-векторов точек и векторов, описывающих данный объект.

Сдвиг точки в пространстве. Простейшей модификацией точки является ее сдвиг в пространстве на вектор сдвига \mathbf{t} . Положение точки до модификации будем называть исходным и описывать радиус-вектором \mathbf{r}_0 , положение точки после модификации будем называть новым и описывать радиус-вектором \mathbf{r} . Положение точки после модификации будет описываться радиус-вектором, равным сумме радиус-вектора ее исходного положения \mathbf{r}_0 и вектора сдвига \mathbf{t}

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{t}. \quad (3.1)$$

Компоненты вектора \mathbf{r} равны сумме соответствующих компонент векторов \mathbf{r}_0 и \mathbf{t} .

Поворот точки в пространстве вокруг оси. Рассмотрим, как изменится радиус-вектор точки при ее повороте вокруг некоторой оси. Пусть начальное положение точки описывается радиус-вектором \mathbf{r}_0 , а ось вращения определяется точкой Q и ортом \mathbf{v} .

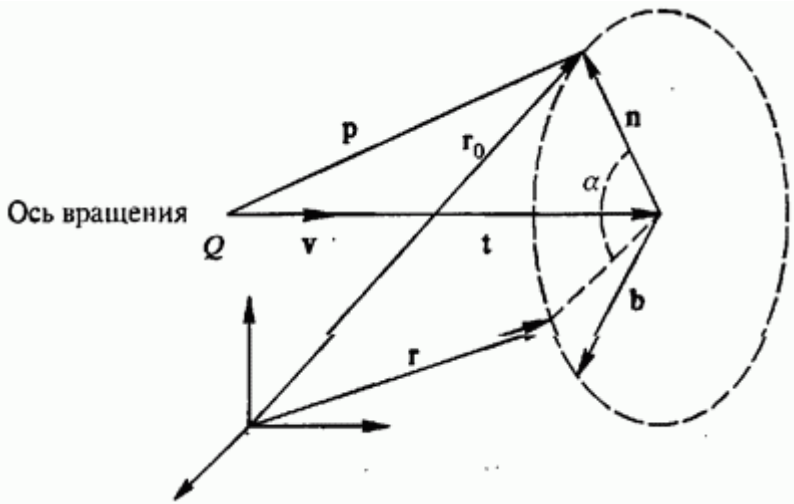


Рис. 3.1. Поворот точки вокруг оси

Пусть \mathbf{q} есть радиус-вектор точки Q . Выполним поворот точки вокруг оси на угол α против часовой стрелки, если взгляд направить навстречу вектору \mathbf{v} (рис. 3.1).

Построим вектор $\mathbf{p} = \mathbf{r}_0 - \mathbf{q}$. Разложим вектор \mathbf{p} на две составляющие:

$$\mathbf{p} = \mathbf{t} + \mathbf{n},$$

$$\mathbf{t} = (\mathbf{p} \cdot \mathbf{v})\mathbf{v}, \quad \mathbf{n} = \mathbf{p} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}, \quad (3.2)$$

где вектор \mathbf{t} параллелен вектору \mathbf{v} , а вектор \mathbf{n} ортогонален вектору \mathbf{v} . При вращении вектор \mathbf{t} не изменится, а вектор \mathbf{n} повернется на угол α в сторону вектора

$$\mathbf{b} = \mathbf{v} \times \mathbf{n} = \mathbf{v} \times (\mathbf{p} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}) = \mathbf{v} \times \mathbf{p}. \quad (3.3)$$

Так как вектор \mathbf{v} имеет единичную длину, то вектор \mathbf{b} будет иметь длину, равную длине вектора \mathbf{n} . Кроме того, он ортогонален векторам \mathbf{v} и \mathbf{n} . После поворота на угол α вектор \mathbf{n} станет равным вектору $\mathbf{n} \cos \alpha + \mathbf{b} \sin \alpha$. Следовательно, после поворота рассматриваемая точка будет определяться радиус-вектором

$$\begin{aligned} \mathbf{r} &= \mathbf{q} + \mathbf{t} + \mathbf{n} \cos \alpha + \mathbf{b} \sin \alpha = \\ &= \mathbf{q} + (\mathbf{p} \cdot \mathbf{v})\mathbf{v} + (\mathbf{p} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}) \cos \alpha + \mathbf{v} \times \mathbf{p} \sin \alpha, \end{aligned} \quad (3.4)$$

где $\mathbf{p} = \mathbf{r}_0 - \mathbf{q}$. Преобразуем выражение

$$(\mathbf{p} \cdot \mathbf{v})\mathbf{v} = \begin{bmatrix} v_1 v_1 p_1 + v_1 v_2 p_2 + v_1 v_3 p_3 \\ v_2 v_1 p_1 + v_2 v_2 p_2 + v_2 v_3 p_3 \\ v_3 v_1 p_1 + v_3 v_2 p_2 + v_3 v_3 p_3 \end{bmatrix} = (\mathbf{v}\mathbf{v}) \cdot \mathbf{p}. \quad (3.5)$$

С учетом (3.5) выражение (3.4) примет вид

$$\mathbf{r} = \mathbf{q} + (\mathbf{v}\mathbf{v}) \cdot \mathbf{p} + \cos \alpha (\mathbf{E} - \mathbf{v}\mathbf{v}) \cdot \mathbf{p} + \sin \alpha \mathbf{v} \times \mathbf{p} = \mathbf{q} + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{q}). \quad (3.6)$$

Матрица поворота определяется равенством

$$\mathbf{A} = (1 - \cos \alpha) \mathbf{v}\mathbf{v} + \cos \alpha \mathbf{E} + \sin \alpha \mathbf{v} \times,$$

где

$$\mathbf{v}\mathbf{v} = \begin{bmatrix} v_1 v_1 & v_1 v_2 & v_1 v_3 \\ v_2 v_1 & v_2 v_2 & v_2 v_3 \\ v_3 v_1 & v_3 v_2 & v_3 v_3 \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{v} \times = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}.$$

Матрица \mathbf{A} является ортогональной. При транспонировании матрицы \mathbf{A} изменится только знак перед последним ее слагаемым, что соответствует повороту точки на угол $-\alpha$.

Симметрия точки относительно плоскости. Определим координаты точки \mathbf{r} , симметричной точке \mathbf{r}_0 относительно плоскости. Пусть

плоскость симметрии определяется точкой Q и двумя ортами \mathbf{u} и \mathbf{v} (рис. 3.2).

Пусть \mathbf{q} есть радиус-вектор точки Q . Построим вектор $\mathbf{p} = \mathbf{r}_0 - \mathbf{q}$ и представим его в виде суммы трех векторов — проекции на орт \mathbf{u} и, проекции на орт \mathbf{v} и перпендикулярной плоскости составляющей \mathbf{n} :

$$\mathbf{p} = (\mathbf{p} \cdot \mathbf{u})\mathbf{u} + (\mathbf{p} \cdot \mathbf{v})\mathbf{v} + \mathbf{n}, \quad (3.7)$$

где $\mathbf{n} = \mathbf{p} - (\mathbf{p} \cdot \mathbf{u})\mathbf{u} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}$. После зеркального отражения вектора \mathbf{p} его нормальная к плоскости составляющая изменит знак на противоположный.

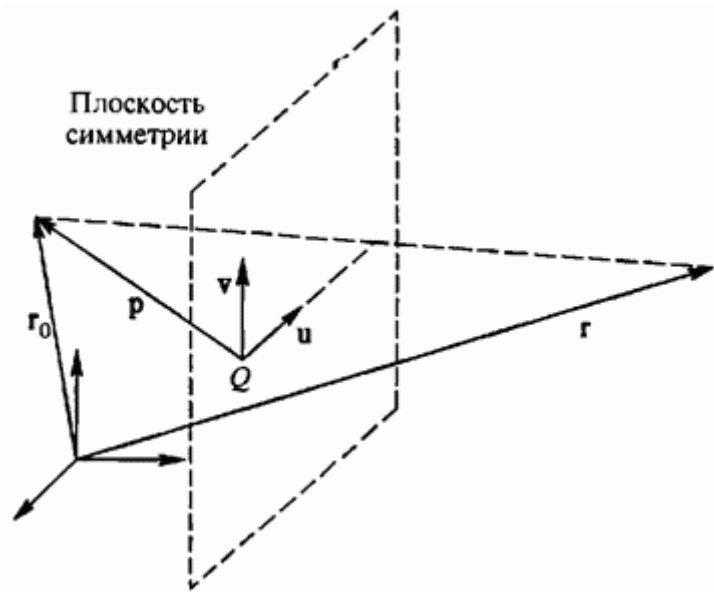


Рис. 3.2. Симметрия точки относительно плоскости

Положение симметричной точки будет описываться радиус-вектором

$$\begin{aligned} \mathbf{r} &= \mathbf{q} + (\mathbf{p} \cdot \mathbf{u})\mathbf{u} + (\mathbf{p} \cdot \mathbf{v})\mathbf{v} - (\mathbf{p} - (\mathbf{p} \cdot \mathbf{u})\mathbf{u} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}) = \\ &= \mathbf{q} + (2\mathbf{u}\mathbf{u} + 2\mathbf{v}\mathbf{v} - \mathbf{E}) \cdot \mathbf{p} = \mathbf{q} + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{q}), \quad (3.8) \end{aligned}$$

где матрица $\mathbf{A} = 2\mathbf{uu} + 2\mathbf{vv}$ — матрица симметрии, \mathbf{uu} , \mathbf{vv} — диадные произведения векторов.

Масштабирование в пространстве. Рассмотрим масштабирование проекций на координатные оси расстояния до точки r_0 относительно некоторой другой точки Q , остающейся неподвижной после масштабирования. Пусть \mathbf{q} есть радиус-вектор точки Q . В общем случае при масштабировании проекции на координатные оси вектора $\mathbf{p} = r_0 - \mathbf{q}$ могут изменяться в различное число раз, т. е. масштабирование может быть ортотропным. Пусть проекция вектора \mathbf{p} на орт \mathbf{e}_1 при масштабировании увеличивается в m_1 раз, проекция вектора \mathbf{p} на орт \mathbf{e}_2 увеличивается в m_2 раз, проекция вектора \mathbf{p} на орт \mathbf{e}_3 увеличивается в m_3 раз. Тогда положение рассматриваемой точки после модификации будет описываться радиус-вектором

$$\mathbf{r} = \mathbf{q} + \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \mathbf{q} + \mathbf{A} \cdot (r_0 - \mathbf{q}), \quad (3.9)$$

где \mathbf{A} — матрица масштабирования.

Модификация векторов в пространстве. Формулы модификации свободного вектора в пространстве получим из формул модификации радиус-вектора, положив в (3.1) $\mathbf{t} = 0$, а в (3.6), (3.8), (3.9) — $\mathbf{q} = 0$. Вектор в отличие от радиус-вектора не привязан ни к какой точке пространства и поэтому модификации вектора можно выполнить в местной системе координат, начало которой находится в точке Q , а координатные оси параллельны исходным координатным осям. После переноса начала местной системы координат в точку Q ее радиус-вектор будет равен нулю. Этим отличаются модификации вектора и радиус-вектора.

Сдвиг двухмерной точки. Рассмотрим модификации двухмерных точек. Векторная формула сдвига двухмерной точки на вектор \mathbf{t} совпадает с (3.1)

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{t}. \quad (3.10)$$

Поворот двухмерной точки вокруг точки. Повороты двумерной точки выполняются вокруг оси, перпендикулярной плоскости, в которой лежит точка. Пусть начальное положение точки описывается радиус-вектором \mathbf{r}_0 , а неподвижная точка Q имеет радиус-вектор \mathbf{q} .

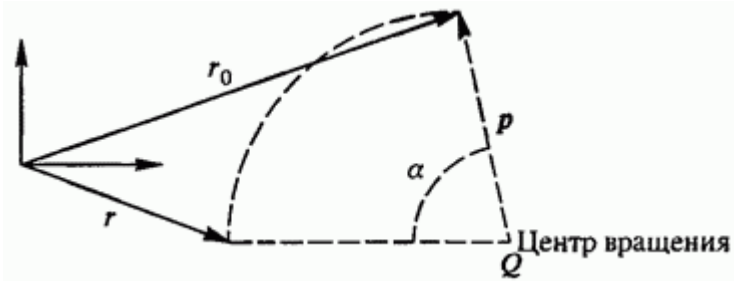


Рис. 3.3. Вращение двухмерной точки

Выполним поворот точки в плоскости на угол α против часовой стрелки, если взгляд направить на плоскость (рис. 3.3).

Построим вектор $\mathbf{p} = \mathbf{r}_0 - \mathbf{q}$ и вектор \mathbf{b} , который имеет длину вектора \mathbf{p} и повернут относительно него на прямой угол против часовой стрелки. Вектор \mathbf{b} получен с помощью преобразования

$$\mathbf{b} = \mathbf{N} \cdot \mathbf{p}, \quad (3.11)$$

где двумерная матрица N имеет вид

$$\mathbf{N} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \quad (3.12)$$

После поворота на угол α вектор \mathbf{p} станет равным вектору $\mathbf{p} \cos \alpha + \mathbf{b} \sin \alpha$. Следовательно, после поворота рассматриваемая точка будет определяться радиус-вектором

$$\mathbf{r} = \mathbf{q} + \mathbf{p} \cos \alpha + \mathbf{b} \sin \alpha = \mathbf{q} + \mathbf{p} \cos \alpha + \mathbf{N} \cdot \mathbf{p} \sin \alpha = \mathbf{q} + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{q}), \quad (3.13)$$

где $\mathbf{A} = \cos\alpha \mathbf{E} + \sin\alpha \mathbf{N}$ — матрица поворота. Матрица \mathbf{A} является ортогональной. При транспонировании матрицы \mathbf{A} изменится только знак перед последним ее слагаемым, что соответствует повороту точки на угол — α .

Симметрия двумерной точки относительно линии. Определим координаты точки r , симметричной точке r_0 относительно линии. Пусть линия симметрии определяется точкой Q и ортом \mathbf{v} (рис. 3.4).

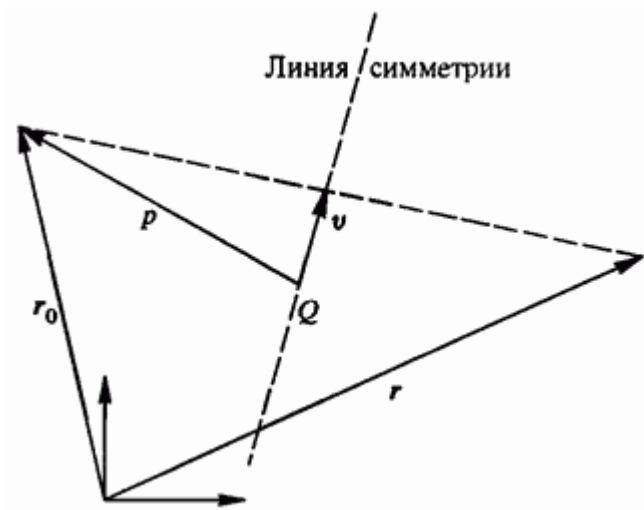


Рис. 3.4. Симметрия точки относительно линии

Пусть \mathbf{q} есть радиус-вектор точки Q . Построим вектор $\mathbf{p} = \mathbf{r}_0 - \mathbf{q}$ и представим его в виде суммы двух векторов — проекции на орт \mathbf{v} и перпендикулярной ему составляющей \mathbf{n} :

$$\mathbf{p} = (\mathbf{p} \cdot \mathbf{v})\mathbf{v} + \mathbf{n}, \quad (3.14)$$

где $\mathbf{n} = \mathbf{p} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}$. Преобразуем выражение

$$(\mathbf{p} \cdot \mathbf{v})\mathbf{v} = \begin{bmatrix} v_1 v_1 p_1 + v_1 v_2 p_2 \\ v_2 v_1 p_1 + v_2 v_2 p_2 \end{bmatrix} = (\mathbf{v}\mathbf{v}) \cdot \mathbf{p}, \quad (3.15)$$

$$\mathbf{v}\mathbf{v} = \begin{bmatrix} v_1 v_1 & v_1 v_2 \\ v_2 v_1 & v_2 v_2 \end{bmatrix}$$

где $\mathbf{v}\mathbf{v}$ — диадное произведение векторов. После зеркального отражения вектора \mathbf{p} его нормальная к линии составляющая изменит знак на противоположный. Положение симметричной точки будет описываться радиус-вектором

$$\mathbf{r} = \mathbf{q} + (\mathbf{p} \cdot \mathbf{v})\mathbf{v} - (\mathbf{p} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}) = \mathbf{q} + (2\mathbf{v}\mathbf{v} - \mathbf{E})\mathbf{p} = \mathbf{q} + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{q}), \quad (3.16)$$

где матрица $\mathbf{A} = 2\mathbf{v}\mathbf{v} - \mathbf{E}$ — матрица симметрии.

Масштабирование в двухмерном пространстве. Пусть задана неподвижная точка \mathbf{q} и требуется масштабировать относительно нее положения других точек. Положение точки с радиус-вектором \mathbf{r}_0 после масштабирования по координатным осям относительно неподвижной точки \mathbf{q} будет описываться радиус-вектором

$$\mathbf{r} = \mathbf{q} + \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \mathbf{q} + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{q}), \quad (3.17)$$

где p_1 и p_2 — компоненты вектора $\mathbf{p} = \mathbf{r}_0 - \mathbf{q}$, m_1 — коэффициент увеличения компоненты p_1 , m_2 — коэффициент увеличения компоненты p_2 .

Модификация двухмерных векторов. Модификации свободного двухмерного вектора получим из модификаций радиус-вектора, положив в (3.10) $\mathbf{t} = 0$, а в (3.14), (3.16), (3.17) — $\mathbf{q} = 0$. Вектор в отличие от радиус-вектора не привязан ни к какой точке двухмерного пространства и поэтому модификации вектора можно выполнить в местной системе координат, начало которой находится в точке \mathbf{Q} , а

координатные оси параллельны исходным координатным осям. Этим отличаются модификации вектора и радиус-вектора.

1.4. Однородные координаты

Рассмотренные выше модификации поворота (3.6), симметрии (3.8), масштабирования (3.9) радиус-вектора точки описываются формулами, имеющими одинаковый вид:

$$\mathbf{r} = \mathbf{q} + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{q}) = \mathbf{A} \cdot \mathbf{r}_0 + (\mathbf{q} - \mathbf{A} \cdot \mathbf{q}) = \mathbf{A} \cdot \mathbf{r}_0 + \mathbf{t}, \quad (4.1)$$

где $\mathbf{t} = \mathbf{q} - \mathbf{A} \cdot \mathbf{q}$ — преобразованный вектор сдвига. Сдвиг (3.1) точки описывается этой же формулой с единичной матрицей \mathbf{A} и вектором \mathbf{t} , равным вектору сдвига. Аналогичный вид имеет преобразование координат точки (2.5). Преобразованиям координат и модификациям точки можно придать единый простой вид

$$\mathbf{R} = \mathbf{A}_t \cdot \mathbf{R}_0. \quad (4.2)$$

Для этого нужно увеличить размерность векторов и матриц на единицу. Вектор, дополненный еще одной компонентой, называется расширенным вектором. Компоненты расширенного вектора называются однородными координатами.

Представим каждый радиус-вектор в расширенном виде

$$\mathbf{R} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} = [r_1 \quad r_2 \quad r_3 \quad 1]^T \quad (4.3)$$

Матрица \mathbf{A}_t представляет собой матрицу \mathbf{A} , окаймленную снизу нулями, а справа — вектором сдвига \mathbf{t} в расширенном виде

$$\mathbf{A}_t = \begin{bmatrix} a_{11} & a_{12} & a_{13} & t_1 \\ a_{21} & a_{22} & a_{23} & t_2 \\ a_{31} & a_{32} & a_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (4.4)$$

Индекс матрицы \mathbf{A}_t говорит о том, что она является расширенной, и включает вектор сдвига \mathbf{t} . Для преобразований (4.1) будем использовать запись

$$\mathbf{R} = \mathbf{Q} + \mathbf{A}_0 \cdot (\mathbf{R}_0 - \mathbf{Q}) = \mathbf{A}_t \cdot \mathbf{R}_0, \quad (4.5)$$

считая, что расширенная матрица \mathbf{A}_t включает преобразование по матрице \mathbf{A} и сдвиг по вектору $\mathbf{t} = \mathbf{q} - \mathbf{A} \cdot \mathbf{q}$. Расширенный вектор \mathbf{Q} описывает точку, остающуюся неподвижной при преобразовании.

В некоторых случаях, например для построения рациональных кривых и поверхностей, наряду с координатами для точек необходим дополнительный параметр — так называемый вес (значимость) точки. Этот параметр точки в вычислениях преобразуется так же, как и координаты, поэтому его считают дополнительной координатой.

При наличии у точки дополнительной координаты w запись ее радиус-вектора в однородных координатах имеет вид

$$\mathbf{R} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{bmatrix} = \begin{bmatrix} wr_1 \\ wr_2 \\ wr_3 \\ w \end{bmatrix} \equiv \begin{bmatrix} w\mathbf{r} \\ w \end{bmatrix} \equiv [w\mathbf{r} \quad w]^\top \quad (4.6)$$

При использовании однородных координат вычисления производятся для однородных компонент R_i , $i=1, 2, 3$, без выделения декартовых

координат. Декартовы координаты точки получают на конечном этапе вычислений делением

$$r_1 = \frac{R_1}{R_4}, \quad r_2 = \frac{R_2}{R_4}, \quad r_3 = \frac{R_3}{R_4}.$$

Следует заметить, что векторы \mathbf{R} и $k\mathbf{R}$ в однородных координатах представляют одну и ту же точку. Для точек, имеющих вес, сдвиг, поворот, симметрия, масштабирование и другие модификации описываются общей формулой (4.2).

С помощью однородных координат и расширенных матриц удобно описывать преобразования координат. Например, переход из системы координат $Q_1i_1z_1$ в систему $O_1e_1z_1$ вместо равенств (2.4) может быть описан матричным равенством

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & a_{31} & q_1 \\ a_{12} & a_{22} & a_{32} & q_2 \\ a_{13} & a_{23} & a_{33} & q_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} = \mathbf{A}_q^T \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix}. \quad (4.7)$$

Обратное преобразование (2.8) может быть описано матричным равенством

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & o_1 \\ a_{21} & a_{22} & a_{23} & o_2 \\ a_{31} & a_{32} & a_{33} & o_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} = \mathbf{A}_o \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} \quad (4.8)$$

Компоненты o_1, o_2, o_3 связаны с компонентами q_1, q_2, q_3 соотношениями

$$\begin{aligned} o_1 &= -a_{11}q_1 - a_{12}q_2 - a_{13}q_3, \\ o_2 &= -a_{21}q_1 - a_{22}q_2 - a_{23}q_3, \\ o_3 &= -a_{31}q_1 - a_{32}q_2 - a_{33}q_3. \end{aligned} \quad (4.9)$$

Матрицы преобразований (4.7) и (4.8) удовлетворяют равенству $\mathbf{A}_o \cdot \mathbf{A}_q^T = \mathbf{A}_q^T \mathbf{A}_o = \mathbf{E}$

Двухмерное пространство. Двухмерные точки в однородных координатах имеют аналогичный вид

$$P = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} wp_1 \\ wp_2 \\ w \end{bmatrix} \quad (4.10)$$

Преобразования двухмерных расширенных векторов аналогичны преобразованиям пространственных расширенных векторов. Например, преобразование координат (2.16) может быть описано матричным равенством

$$\begin{bmatrix} p_1 \\ p_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \varphi & -\sin \varphi & q_1 \\ \sin \varphi & \cos \varphi & q_2 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ 1 \end{bmatrix}. \quad (4.11)$$

Обратное преобразование (2.17) может быть описано матричным равенством

$$\begin{bmatrix} y_1 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \varphi & \sin \varphi & o_1 \\ -\sin \varphi & \cos \varphi & o_2 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ 1 \end{bmatrix}, \quad (4.12)$$

где компоненты o_1, o_2 связаны с компонентами q_1, q_2 соотношениями (2.18).

Использование однородных координат оказывается полезным и там, где дополнительная компонента точки отсутствует. В этом случае ее полагают равной единице.

1.5. Геометрия кривых линий

Кривой линией или просто кривой будем называть геометрическое место точек, координаты которых описываются непрерывными и однозначными функциями $r_1(t)$, $r_2(t)$, $r_3(t)$ параметра t , принимающего значения на отрезке $t_{\min} \leq t \leq t_{\max}$.

В декартовой прямоугольной системе координат $Oe_1e_2e_3$ кривую можно описать радиус-вектором

$$\mathbf{r}(t) = \sum_{i=1}^3 r_i(t) \mathbf{e}_i, \quad t_{\min} \leq t \leq t_{\max}. \quad (5.1)$$

Представление кривой в виде (5.1) называется параметрическим. Далее будем предполагать, что координатные функции $r_i(t)$ имеют непрерывные производные до любого порядка, который нам потребуется. Точку кривой $\mathbf{r}(t)$ будем называть обыкновенной, если в этой точке не обращается в нуль длина вектора первой производной по параметру $\mathbf{r}' = d\mathbf{r}/dt$. В противном случае точку кривой будем называть особой.

Положение точки кривой зависит от параметра t , который является ее внутренней координатой. Параметр t своими значениями однозначно определяет точки кривой. Но так как геометрически он ничем не связан с кривой, то для нее можно использовать другую параметризацию. Для этого достаточно ввести новый параметр v , который связан с параметром t функциональной зависимостью

$$t = t(v), \quad v_{\min} \leq v \leq v_{\max}, \quad (5.2)$$

где $t_{\min} = t(v_{\min})$, $t_{\max} = t(v_{\max})$. Будем предполагать при этом, что функция $t(v)$ является взаимно однозначной и монотонно возрастающей. Когда параметр v пробегает свою область изменения от v_{\min} до v_{\max} , параметр t пробегает свою область изменения от t_{\min} до t_{\max} . Кривая (5.1), выраженная через параметр v , будет иметь вид

$$\mathbf{r}(v) = \mathbf{r}(v(t)), \quad v_{\min} \leq v \leq v_{\max}.$$

При этом форма кривой останется прежней. Как параметр t , так и параметр v одинаково годятся, чтобы характеризовать точки рассматриваемой кривой, и выбор параметра зависит от нас. Мы будем пользоваться этим свойством кривых, чтобы согласовать параметрические длины кривых. Пусть имеются две кривые: первая $\mathbf{a}(t)$, $t_{\min} \leq t \leq t_{\max}$ вторая $\mathbf{c}(v)$, $v_{\min} \leq v \leq v_{\max}$, и нам нужно, чтобы параметрические длины кривых были одинаковыми.

Изменим параметризацию второй кривой так, чтобы пределы изменения ее параметров совпадали с пределами первой кривой. Для этого у второй кривой введем параметр t , связанный параметром v зависимостью

$$v(t) = v_{\min} \frac{t_{\max} - t}{t_{\max} - t_{\min}} + v_{\max} \frac{t - t_{\min}}{t_{\max} - t_{\min}}. \quad (5.3)$$

Вторая кривая теперь будет иметь вид $\mathbf{c}=\mathbf{c}(t)=\mathbf{c}(v(t))$,

$t_{\min} \leq t \leq t_{\max}$. В частном случае параметром кривой может служить длина ее дуги, отсчитываемая от начальной точки. В общем случае параметр t , как любая координатная система, может быть определен удобным для нас способом.

Для векторной функции, как и для скалярной, определяются производные. Производные векторной функции также представляют собой векторы. Для любых векторных функций $\mathbf{a}(t)$ и $\mathbf{c}(t)$ и любой скалярной функции $\lambda(t)$ справедливы правила дифференцирования

$$\begin{aligned}\frac{d(\mathbf{a} + \mathbf{c})}{dt} &= \frac{d\mathbf{a}}{dt} + \frac{d\mathbf{c}}{dt}, \\ \frac{d(\lambda\mathbf{a})}{dt} &= \mathbf{a} \frac{d\lambda}{dt} + \lambda \frac{d\mathbf{a}}{dt}, \\ \frac{d(\mathbf{a} \cdot \mathbf{c})}{dt} &= \mathbf{c} \cdot \frac{d\mathbf{a}}{dt} + \mathbf{a} \cdot \frac{d\mathbf{c}}{dt}, \\ \frac{d(\mathbf{a} \times \mathbf{c})}{dt} &= \frac{d\mathbf{a}}{dt} \times \mathbf{c} + \mathbf{a} \times \frac{d\mathbf{c}}{dt}.\end{aligned}$$

В декартовой прямоугольной системе координат производные векторной функции n -го порядка по ее параметру имеют простой вид

$$\mathbf{r}^{(n)} \equiv \frac{d^n \mathbf{r}}{dt^n} = \sum_{i=1}^3 \frac{d^n r_i(t)}{dt^n} \mathbf{e}_i. \quad (5.4)$$

В криволинейной системе координат базисные векторы изменяют свою длину и направление при переходе от одной точки к другой и формулы для производных имеют более сложный вид.

Если координатные функции кривой в некоторой точке достаточное число раз дифференцируемы, то векторную функцию кривой в окрестности этой точки можно разложить в ряд Тейлора

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{r}'(t)\Delta t + \mathbf{r}''(t)\frac{(\Delta t)^2}{2} + \dots + \mathbf{r}^{(n)}(t)\frac{(\Delta t)^n}{n!} + \mathbf{q}_n \frac{(\Delta t)^{n+1}}{(n+1)!}, \quad (5.5)$$

$$\mathbf{q}_n = \sum_{i=1}^3 \frac{d^{n+1} r_i}{dt^{n+1}} \mathbf{e}_i.$$

где Производные $\frac{d^{n+1} r_i}{dt^{n+1}}$ вычислены при некоторых значениях параметра $t \leq t_i \leq t + \Delta t$. Вектор \mathbf{q}_n представляет собой остаточный член ряда Тейлора. Из свойств рядов Тейлора координатной функции $r_i(t)$ следует, что длина вектора \mathbf{q}_n не превосходит некоторое положительное число, постоянное для всех t из его области изменения для кривой, и что предел длины остаточного

вектора стремится к нулю при увеличении числа членов усеченного ряда

$$\lim_{n \rightarrow \infty} |\mathbf{q}_n| = 0.$$

При изменении параметризации кривой $\mathbf{r}(t)$, $t(v)$ производные по новому параметру выражаются через производные по старому параметру следующим образом:

$$\begin{aligned} \frac{d\mathbf{r}}{dv} &= \frac{d\mathbf{r}}{dt} \frac{dt}{dv}, \\ \frac{d^2\mathbf{r}}{dv^2} &= \frac{d^2\mathbf{r}}{dt^2} \left(\frac{dt}{dv}\right)^2 + \frac{d\mathbf{r}}{dt} \frac{d^2t}{dv^2}, \\ \frac{d^3\mathbf{r}}{dv^3} &= \frac{d^2\mathbf{r}}{dt^2} \left(\frac{dt}{dv}\right)^3 + 3 \frac{d^2\mathbf{r}}{dt^2} \frac{dt}{dv} \frac{d^2t}{dv^2} + \frac{d\mathbf{r}}{dt} \frac{d^3t}{dv^3}. \end{aligned}$$

Исследуем поведение кривой на бесконечно малом участке вблизи обыкновенной точки. Если существуют производные по параметру координатных функций, то кривая также имеет производные по параметру соответствующего порядка. Рассмотрим геометрический смысл производной векторной функции. Пусть при каком-либо значении параметра t радиус-вектор $\mathbf{r}(t)$ указывает на некоторую точку R . Перейдем к другому параметру $t_1 = t + \Delta t$ при котором векторная функция $\mathbf{r}(t + \Delta t)$ указывает на некоторую другую точку R_1 . Разность этих двух значений векторной функции $\Delta \mathbf{r} = \mathbf{r}(t + \Delta t) - \mathbf{r}(t)$ описывает хорду RR_1 (рис. 5.1).

Вектор

$$\mathbf{t} = \frac{\Delta \mathbf{r}}{\Delta t} \quad (5.6)$$

параллелен хорде RR_1 , но в общем случае не равен ей по длине. Устремим Δt к нулю, тогда точка R_1 будет приближаться к точке R , вектор (5.6) будет стремиться к касательной к кривой в точке R .

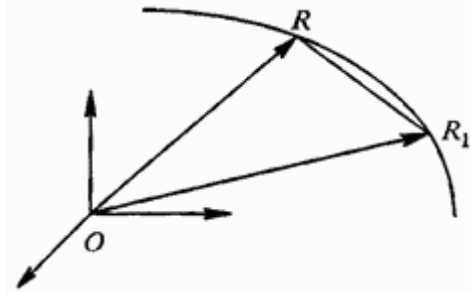


Рис. 5.1

Предел отношения (5.6) при $\Delta t \rightarrow 0$ является первой производной векторной функции

$$\mathbf{r}' \equiv \sum_{i=1}^3 \frac{dr_i(t)}{dt} \mathbf{e}_i = \lim_{\Delta t \rightarrow 0} \sum_{i=1}^3 \frac{r_i(t + \Delta t) - r_i(t)}{\Delta t} \mathbf{e}_i. \quad (5.7)$$

Таким образом, производная $\mathbf{r}'(t)$ кривой $\mathbf{r}(t)$ есть вектор, направленный по касательной к кривой в точке, определяемой параметром t . Заметим, что производная всегда направлена в сторону возрастания параметра. Зная первую производную радиус-вектора кривой, можно вычислить длину кривой. Длина кривой равна пределу, к которому стремится длина ломаной, вписанной в кривую. Таким образом, длина кривой равна интегралу

$$s = \int_{t_{\min}}^{t_{\max}} \sqrt{\mathbf{r}' \cdot \mathbf{r}'} dt. \quad (5.8)$$

Натуральная параметризация кривой.

Длина вектора производной зависит от способа параметризации кривой. Существует способ параметризации, при котором длина вектора производной равна единице. Для этого используется параметризация, геометрически связанная с кривой, а именно: за параметр принимается длина дуги кривой, отсчитываемая от некоторой

ее точки (например, начальной). Действительно, в этом случае длина хорды RR_1 стремится к длине дуги Δt , а длина вектора \mathbf{r}' стремится к единице, когда Δt стремится к нулю. Векторная зависимость

$$\mathbf{r} = \mathbf{r}(s), \quad (5.9)$$

где s — длина дуги, называется уравнением кривой с натуральной параметризацией.

Натуральная параметризация удобна в теоретических исследованиях, так как она упрощает формулы. С практической точки зрения она не всегда удобна, так как требует знание длины дуги.

Сопровождающий трехгранник.

В каждой точке кривой можно построить плоскость, перпендикулярную ее первой производной. Такая плоскость называется нормальной плоскостью кривой. Плоскость, в которой лежат и первая производная кривой и ее вторая производная, называется соприкасающейся плоскостью. Если вторая производная кривой параллельна первой производной или ее длина равна нулю, то в качестве соприкасающейся плоскости можно взять любую плоскость, в которой лежит первая производная кривой. Точка кривой, в которой векторы первой и второй производных кривой коллинеарны, называется точкой распрямления. Точки распрямления не зависят от способа параметризации кривой. Название соприкасающейся плоскости обусловлено тем, что она проходит через заданную точку кривой с наивысшим порядком касания, и ее можно определить как предельное положение плоскости, построенной по трем бесконечно близким точкам кривой.

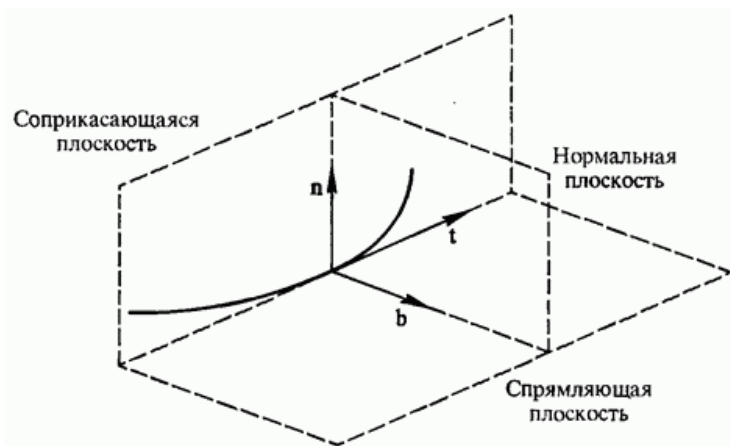


Рис. 5.2. Сопровождающий трехгранник кривой

Плоскость, перпендикулярная нормальной и соприкасающейся плоскостям, называется спрямляющей плоскостью (рис. 5.2).

Единичный вектор, направленный вдоль первой производной кривой, называется касательным вектором кривой в данной точке. Единичный вектор, направленный вдоль линии пересечения нормальной и соприкасающейся плоскостей и направленный в сторону второй производной (в сторону вогнутости кривой), называется главной нормалью кривой в данной точке. Единичный вектор, направленный вдоль линии пересечения нормальной спрямляющей плоскостей и образующий с касательным и нормальным вектором правую тройку векторов, называется бинормалью кривой в данной точке.

Таким образом, с каждой точкой кривой связаны три взаимно перпендикулярные плоскости: нормальная, соприкасающаяся и спрямляющая, а также три взаимно ортогональных вектора единичной длины: касательный t , нормаль n и бинормаль $b = t \times n$. Совокупность трех перечисленных плоскостей и трех векторов t, n, b называется сопровождающим трехгранником кривой.

Тройка векторов t, n, b выражается через производные векторной функции кривой. Они помогут нам в исследовании строения кривой в

бесконечно малой окрестности каждой ее точки. Тройка единичных векторов связана соотношениями

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}, \quad \mathbf{t} = \mathbf{n} \times \mathbf{b}, \quad \mathbf{n} = \mathbf{b} \times \mathbf{t}. \quad (5.10)$$

Формулы Френе-Серре.

Предположим, что нам известно уравнение кривой с натуральной параметризацией (5.9). В этом случае

$$\mathbf{t} = \frac{d\mathbf{r}}{ds}. \quad (5.11)$$

Из того, что длина первой производной равна единице и не меняется вдоль кривой, следует

$$\frac{d(\mathbf{t} \cdot \mathbf{t})}{ds} = 2\mathbf{t} \cdot \frac{d\mathbf{t}}{ds} = 2\mathbf{t} \cdot \frac{d^2\mathbf{r}}{ds^2} = 0, \quad (5.12)$$

т. е. векторы первой и второй производной ортогональны. Следовательно, вторая производная кривой с натуральной параметризацией направлена по главной нормали:

$$\frac{d\mathbf{t}}{ds} = \frac{d^2\mathbf{r}}{ds^2} = k\mathbf{n}. \quad (5.13)$$

Коэффициент k называется кривизной кривой линии. Ниже мы покажем, что обратная ей величина равна радиусу окружности, соприкасающейся с кривой в рассматриваемой точке. Бинормаль по определению ортогональна касательному вектору и главной нормали. Из этого следует, что

$$\begin{aligned} \frac{d(\mathbf{b} \cdot \mathbf{t})}{ds} &= \mathbf{t} \cdot \frac{d\mathbf{b}}{ds} + \mathbf{b} \cdot \frac{d\mathbf{t}}{ds} = \mathbf{t} \cdot \frac{d\mathbf{b}}{ds} + k\mathbf{b} \cdot \mathbf{n} = \mathbf{t} \cdot \frac{d\mathbf{b}}{ds} = 0, \\ \frac{d(\mathbf{b} \cdot \mathbf{b})}{ds} &= 2\mathbf{b} \cdot \frac{d\mathbf{b}}{ds} = 0. \end{aligned} \tag{5.14}$$

$$\frac{d\mathbf{b}}{ds}$$

Таким образом, вектор $\frac{d\mathbf{b}}{ds}$ ортогонален векторам \mathbf{t} и \mathbf{b} и, следовательно, он параллелен главной нормали. Это принято записывать в виде

$$\frac{d\mathbf{b}}{ds} = -\chi\mathbf{n}. \tag{5.15}$$

Коэффициент χ называется кручением кривой линии. Равенства (5.13) и (5.15) определяют производные ортов \mathbf{t} и \mathbf{b} по длине дуги. Найдем производную нормали по длине дуги

$$\frac{d\mathbf{n}}{ds} = \frac{d(\mathbf{b} \times \mathbf{t})}{ds} = \frac{d\mathbf{b}}{ds} \times \mathbf{t} + \mathbf{b} \times \frac{d\mathbf{t}}{ds} = -\chi\mathbf{n} \times \mathbf{t} + k\mathbf{b} \times \mathbf{n} = \chi\mathbf{b} - k\mathbf{t}. \tag{5.16}$$

Нами получены дифференциальные зависимости для векторов \mathbf{t} , \mathbf{n} и \mathbf{b} :

$$\frac{d\mathbf{t}}{ds} = k\mathbf{n}, \quad \frac{d\mathbf{n}}{ds} = \chi\mathbf{b} - k\mathbf{t}, \quad \frac{d\mathbf{b}}{ds} = -\chi\mathbf{n}, \tag{5.17}$$

которые известны как формулы Френе-Серре. Они выражают производные векторов сопровождающего трехгранника в виде разложения по самим этим векторам. Используя (5.10), (5.11) и формулы Френе-Серре, выразим векторы \mathbf{t} , \mathbf{n} и \mathbf{b} , кривизну и кручение кривой через производные радиус-вектора кривой по ее длине дуги следующим образом:

$$\mathbf{t} = \frac{d\mathbf{r}}{ds}, \quad \mathbf{n} = \frac{1}{k} \frac{d^2\mathbf{r}}{ds^2}, \quad \mathbf{b} = \frac{d\mathbf{r}}{ds} \times \frac{1}{k} \frac{d^2\mathbf{r}}{ds^2},$$

$$k = \left| \frac{d^2\mathbf{r}}{ds^2} \right|, \quad \chi = \frac{1}{k^2} \left(\frac{d\mathbf{r}}{ds} \times \frac{d^2\mathbf{r}}{ds^2} \right) \cdot \frac{d^3\mathbf{r}}{ds^3}.$$

Соприкасающаяся окружность.

Поясним, почему коэффициент k называют кривизной кривой. Построим окружность радиусом равным некоторой величине ρ . Ее уравнение с натуральной параметризацией в некоторой местной системе координат с осями $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$ имеет вид

$$\mathbf{r}(s) = \rho \cos\left(\frac{s}{\rho}\right) \mathbf{i}_1 + \rho \sin\left(\frac{s}{\rho}\right) \mathbf{i}_2.$$

Нормаль к окружности и производные радиус-вектора окружности равны

$$\mathbf{n} = -\cos\left(\frac{s}{\rho}\right) \mathbf{i}_1 - \rho \sin\left(\frac{s}{\rho}\right) \mathbf{i}_2,$$

$$\frac{d\mathbf{r}}{ds} = -\sin\left(\frac{s}{\rho}\right) \mathbf{i}_1 + \rho \cos\left(\frac{s}{\rho}\right) \mathbf{i}_2,$$

$$\frac{d^2\mathbf{r}}{ds^2} = \frac{1}{\rho} \left(-\cos\left(\frac{s}{\rho}\right) \mathbf{i}_1 - \rho \sin\left(\frac{s}{\rho}\right) \mathbf{i}_2 \right) = \frac{\mathbf{n}}{\rho}.$$

Очевидно, что коэффициент пропорциональности k для окружности в формуле (5.13) равен $1/\rho$. Таким образом, если мы построим для

кривой соприкасающуюся с ней окружность (с таким же $\frac{d^2\mathbf{r}}{ds^2}$ как у кривой), то радиус этой окружности будет равен ρ (рис. 5.3).

Величина $\rho = 1/k$ называется радиусом кривизны кривой. Для произвольной кривой ее кривизна k и кручение χ являются функциями параметра t .

Поясним, почему коэффициент χ называют кручением кривой. Построим в некоторой точке кривой сопровождающий трехгранник и посмотрим, как он будет себя вести при движении вдоль кривой.

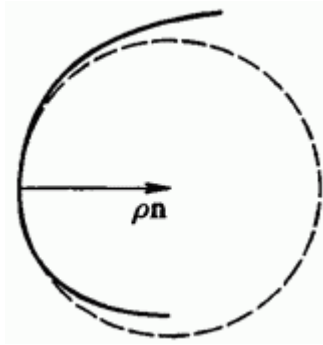


Рис. 5.3. Соприкасающаяся с кривой окружность

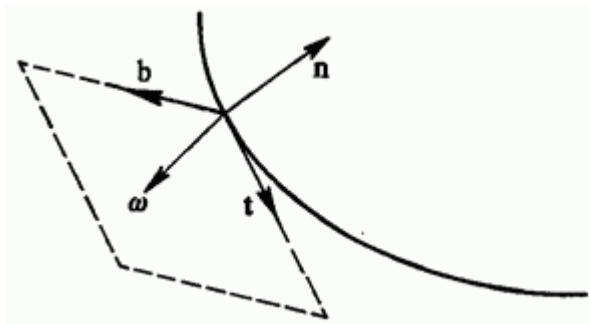


Рис. 5.4. Вектор Дарбу

Из (5.17) получим, что при увеличении параметра на небольшую величину Δt касательный вектор t повернется в сторону главной нормали n на угол k , а бинормаль b повернется в сторону, противоположную главной нормали n , на угол χ (рис. 5.4).

Если наблюдать этот процесс, «сидя на кончике вектора \mathbf{t} », то мы увидим, что главная нормаль \mathbf{n} и соприкасающаяся плоскость повернулись в сторону бинормали \mathbf{b} на угол χ . Теперь представим, что точка движется по кривой, проходя единицу длины ее дуги за единицу времени. В этом случае угловая скорость вращения сопровождающего трехгранника вокруг касательного вектора будет равна кручению кривой χ . Если кручение $\chi = 0$, то кривая является плоской. Справедливо и обратное утверждение. Если кручение кривой равно нулю, то соприкасающаяся плоскость во всех точках кривой одна и та же, все бинормали параллельны друг другу, а кривая является плоской.

Полный вектор угловой скорости вращения сопровождающего трехгранника по отношению к пути, проходимому по кривой, называется вектором Дарбу. Он равен

$$\boldsymbol{\omega} = k\mathbf{b} + \chi\mathbf{t}. \quad (5.19)$$

Вектор Дарбу придает механический смысл формулам Френе-Серре (5.17), с использованием которого последние имеют вид

$$\frac{d\mathbf{t}}{ds} = \boldsymbol{\omega} \times \mathbf{t}, \quad \frac{d\mathbf{n}}{ds} = \boldsymbol{\omega} \times \mathbf{n}, \quad \frac{d\mathbf{b}}{ds} = \boldsymbol{\omega} \times \mathbf{b}.$$

Натуральные уравнения кривой.

Кривизна и кручение полностью характеризуют форму кривой, если не считать ее положение и ориентацию в пространстве. Они также являются определенными функциями длины дуги как параметра

$$k = k(s) > 0, \quad \chi = \chi(s). \quad (5.20)$$

Так как длина дуги, кривизна и кручение не зависят от способа параметризации и от выбора координат в пространстве, то и функции $k(s)$ и $\chi(s)$ также от этого выбора не зависят. Уравнения (5.20) называются натуральными уравнениями кривой. Если две кривые имеют одинаковые натуральные уравнения, то они идентичны и отличаются только положением и ориентацией в пространстве.

Функции $s(t)$, $k(t)$, $\chi(t)$ однозначно определяют кривую с точностью до положения и ориентации в пространстве.

Как правило, параметризация кривой не является натуральной. Определим орты \mathbf{t} , \mathbf{n} и \mathbf{b} и коэффициенты k и χ для произвольной параметризации. Производные по параметру будем обозначать штрихами. Первую, вторую и третью производные радиус-вектора кривой по ее параметру представим следующим образом:

$$\mathbf{r}' \equiv \frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}}{ds} \frac{ds}{dt} = s' \mathbf{t}, \quad (1.5.1)$$

$$\mathbf{r}'' \equiv \frac{d^2\mathbf{r}}{dt^2} = \frac{d(s'\mathbf{t})}{dt} = \frac{d^2s}{dt^2} \mathbf{t} + \frac{dt}{ds} \left(\frac{ds}{dt} \right)^2 = s'' \mathbf{t} + s'^2 k \mathbf{n}, \quad (1.5.2)$$

$$\begin{aligned} \mathbf{r}''' \equiv \frac{d^3\mathbf{r}}{dt^3} &= \frac{d(s''\mathbf{t} + s'^2 k \mathbf{n})}{dt} = \frac{d^3s}{dt^3} \mathbf{t} + 3k \mathbf{n} \frac{ds}{dt} \frac{d^2s}{dt^2} + \mathbf{n} \left(\frac{ds}{dt} \right)^3 \frac{dk}{ds} + (\chi \mathbf{b} - k \mathbf{t}) k \left(\frac{ds}{dt} \right)^3 \\ &= (s''' - k^2 s'^3) \mathbf{t} + \left(3k s'' s' + s'^3 \frac{dk}{dt} \right) \mathbf{n} + k \chi s'^3 \mathbf{b}, \end{aligned} \quad (1.5.3)$$

(5.21)-(5.23)

Где s' , s'' , s''' — первая, вторая и третья производные длины дуги по параметру кривой. Из (5.21) получим формулу для вычисления производной длины дуги по параметру и формулу для вычисления касательного вектора \mathbf{t} :

$$s' = |\mathbf{r}'| = \sqrt{\frac{d\mathbf{r}}{dt} \cdot \frac{d\mathbf{r}}{dt}},$$

$$\mathbf{t} = \frac{d\mathbf{r}/dt}{|d\mathbf{r}/dt|} \equiv \frac{d\mathbf{r}/dt}{ds/dt} = \frac{\mathbf{r}'}{s'}. \quad (5.24)-(5.25)$$

Умножив (5.21) векторно на (5.22), получим формулу для определения кривизны кривой и направления бинормали

$$k \mathbf{b} = \frac{\mathbf{r}' \times \mathbf{r}''}{s'^3}. \quad (5.26)$$

Умножив (5.26) векторно на (5.25) и используя равенство для двойного векторного произведения $(\mathbf{r}' \times \mathbf{r}'') \times \mathbf{t} = (\mathbf{t} \cdot \mathbf{r}') \mathbf{r}'' - (\mathbf{t} \cdot \mathbf{r}'') \mathbf{r}'$, получим формулу для вычисления направления вектора главной нормали

$$k\mathbf{n} = \frac{(\mathbf{r}' \times \mathbf{r}'') \times \mathbf{r}'}{s'^4} = \frac{\mathbf{r}''}{s'^2} - \frac{\mathbf{r}' \cdot \mathbf{r}''}{s'^4} \mathbf{r}'.$$
 (5.27)

Из правой части равенства (5.27) видно, что вектор $k\mathbf{n}$ является составляющей вектора $\mathbf{r}''/(s')^2$ перпендикулярной к касательному вектору \mathbf{t} (составляющая, параллельная \mathbf{t} , вычитается). Кривизна кривой равна длине вектора, стоящего в правой части

$$k = \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|^3}.$$
 (5.28)

Соответственно, при $|\mathbf{r}' \times \mathbf{r}''| \neq 0$ радиус соприкасающейся окружности определится формулой

$$\rho = \frac{|\mathbf{r}'|^3}{|\mathbf{r}' \times \mathbf{r}''|}.$$
 (5.28.1)

Умножив (5.26) скалярно на (5.23), получим формулу для определения кручения кривой

$$\chi = \frac{(\mathbf{r}' \times \mathbf{r}'') \cdot \mathbf{r}'''}{k^2 s'^6} = \frac{(\mathbf{r}' \times \mathbf{r}'') \cdot \mathbf{r}'''}{|\mathbf{r}' \times \mathbf{r}''|^2}.$$
 (5.29)

Если кривизна кривой в данной точке не равна нулю, то, поделив обе части равенств (5.26) и (5.27) на кривизну (5.28), получим бинормаль и нормаль:

$$\mathbf{b} = \frac{\mathbf{r}' \times \mathbf{r}''}{|\mathbf{r}' \times \mathbf{r}''|^2},$$
$$\mathbf{n} = \frac{\mathbf{r}''(\mathbf{r}' \cdot \mathbf{r}') - \mathbf{r}'(\mathbf{r}' \cdot \mathbf{r}'')}{|\mathbf{r}' \times \mathbf{r}''|^2 s'}.$$

Если известна векторная функция линии (5.1), то формулы (5.24)-(5.31) позволяют получить всю геометрическую информацию о кривой линии.

Из формул (5.28) и (5.29) видно, что кривизна всегда неотрицательна (в числителе и знаменателе стоят квадратные корни), а кручение может иметь любой знак. Если кривизна равна нулю, то направление главной нормали, бинормали и кручение не определены. Если кривизна равна нулю в каждой точке кривой, то она является прямой линией. Вектор главной нормали в этом случае может иметь произвольное направление в нормальной плоскости. Если векторы \mathbf{r}' , \mathbf{r}'' , \mathbf{r}''' коллинеарны, то кручение кривой равно нулю и кривая является плоской.

1.6. Геометрия двумерных кривых

Кроме пространственных линий, для нас практическое значение будут иметь двумерные кривые на плоскости. В частности, двумерным пространством будет служить область параметров поверхностей.

Рассмотрим случай двумерной кривой. Пусть на плоскости определена декартова прямоугольная система координат с началом в точке O и базисными векторами e_1 и e_2 . Компоненты двумерных векторов будем обозначать через x и y .

Большинство формул для двумерных кривых можно получить из соответствующих формул для пространственных кривых, положив в них третью координату равной нулю и опустив ее и все векторы, ортогональные плоскости кривой (в их числе бинормаль). Радиус-вектор двумерной кривой $\mathbf{r}(t) = [x(t) \ y(t)]^T$ будет описываться выражением

$$\mathbf{r}(t) = x(t)\mathbf{e}_1 + y(t)\mathbf{e}_2. \quad (6.1)$$

Для двумерной кривой $\mathbf{r}(s)$ формулы Френе-Серре имеют вид

$$\frac{d\mathbf{t}}{ds} = k\mathbf{n}, \quad \frac{d\mathbf{n}}{ds} = -k\mathbf{t}, \quad (6.2)$$

где $\mathbf{t} = d\mathbf{r}/ds$ — касательный вектор кривой, s — длина дуги кривой. Кривизна двумерной кривой определяется равенством

$$k = \left| \frac{d^2\mathbf{r}}{ds^2} \right|. \quad (6.3)$$

Натуральное уравнение двумерной кривой выражает ее кривизну как функцию дуги вдоль кривой и имеет вид

$$k = k(s).$$

Чтобы сохранить справедливость всех приведенных выше формул, следует переопределить операцию векторного произведения для двумерных векторов. Результатом векторного произведения двумерных векторов $\mathbf{r}_1 = [x_1 \ y_1]^T$ и $\mathbf{r}_2 = [x_2 \ y_2]^T$ будем считать скалярную величину, равную

$$\mathbf{r}_1 \times \mathbf{r}_2 = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = (x_1 y_2 - x_2 y_1). \quad (6.4)$$

Выразим через координаты x и y производную длины дуги и кривизну двумерной кривой, используя формулы (5.24) и (5.26),

$$s' = |\mathbf{r}'| = \sqrt{\mathbf{r}' \cdot \mathbf{r}'} = \sqrt{x'^2 + y'^2},$$

$$k = \frac{\mathbf{r}' \times \mathbf{r}''}{|\mathbf{r}'|^3} = \frac{y''x' - x''y'}{(\sqrt{x'^2 + y'^2})^3}. \quad (6.5)-(6.6)$$

Эволюта и эвольвента. Для каждой обыкновенной точки кривой можно указать центр кривизны. Геометрическое место центров кривизны всех точек данной кривой называется эволютой этой кривой. Выражение для радиус-вектора эволюты $\mathbf{a}(t)$ получим, добавив к радиус-вектору кривой $\mathbf{r}(t)$ вектор нормали, деленный на кривизну кривой:

$$\mathbf{a}(t) = \mathbf{r}(t) + \frac{1}{k} \mathbf{n}(t) = \mathbf{r} + \frac{1}{k^2} \frac{d^2 \mathbf{r}}{ds^2}. \quad (6.7)$$

При произвольной параметризации кривой нормаль выразим с помощью равенства (5.27) и получим

$$\mathbf{a}(t) = \mathbf{r} + \frac{\mathbf{r}''(\mathbf{r}' \cdot \mathbf{r}') - \mathbf{r}'(\mathbf{r}' \cdot \mathbf{r}'')}{k^2 s'^4} = \mathbf{r} + \frac{\mathbf{r}''(\mathbf{r}' \cdot \mathbf{r}') - \mathbf{r}'(\mathbf{r}' \cdot \mathbf{r}'')}{|\mathbf{r}' \times \mathbf{r}''|^2} (\mathbf{r}' \cdot \mathbf{r}'). \quad (6.8)$$

Для двумерной кривой $\mathbf{r}(t) = [x(t) \ y(t)]^T$ запишем векторное равенство (6.8) отдельно для каждой координаты эволюты $\mathbf{a}(t) = [x_a(t) \ y_a(t)]^T$

$$x_a(t) = x + \frac{x''(x'^2 + y'^2) - x'(x'x'' + y'y'')}{(y''x' - x''y')^2} (x'^2 + y'^2) = x + y' \frac{x'^2 + y'^2}{x''y' - y''x'},$$

$$y_a(t) = y + \frac{y''(x'^2 + y'^2) - y'(x'x'' + y'y'')}{(y''x' - x''y')^2} (x'^2 + y'^2) = y + x' \frac{x'^2 + y'^2}{y''x' - x''y'}. \quad (6.9)$$

В точках распрямления радиус-вектор соответствующей точки эволюты стремится к бесконечности. Если точка распрямления является точкой перегиба, то эволюта терпит в соответствующей точке разрыв. Эволюта в каждой своей точке касается нормали к исходной кривой в соответствующей точке. Эволюту можно определить как огибающую семейства нормалей.

Исходная кривая по отношению к своей эволюте является эвольвентой (разверткой). Для кривой $a(s)$ эвольвента описывается радиус-вектором

$$r(s) = a - (s - s_0) \frac{da}{ds} = a - (s - s_0)t, \quad (6.10)$$

где s — длина дуги кривой $a(s)$, $s_0 = \text{const}$. t — касательная к кривой. Для заданной плоской кривой можно построить множество эвольвент, в зависимости от s_0 (или от того, в какой точке кривой принять длину дуги равную нулю). На рис. 6.1 показана кривая $a(s)$ и ее эвольвента.

Эвольвента строится следующим образом. Положим, что длина дуги кривой $a(s)$ отсчитывается от точки C_0 .

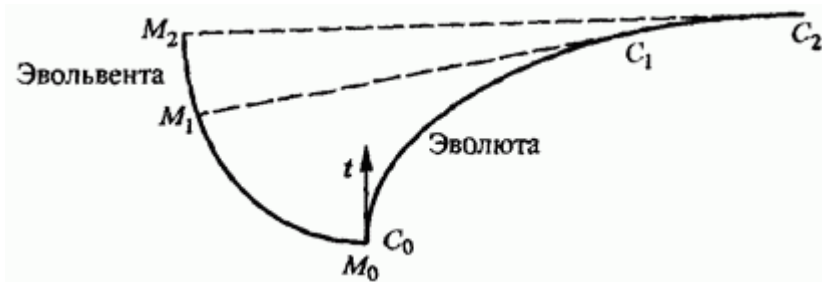


Рис. 6.1. Эволюта и эвольвента

Для получения точки M эвольвенты, соответствующей некоторой точке C_i исходной кривой, построим в точке C_i касательную и отложим на ней с учетом знака отрезок длиной, равной длине дуги $C_i C_0$, взятой с отрицательным знаком, если значение параметра в точке C_i , больше значения параметра в точке C_0 , и взятой с положительным знаком в противном случае. Можно сказать, что эвольвента представляет собой развертку исходной кривой.

Для доказательства равенства (6.10) покажем, что эволюта кривой есть кривая $a(s)$. Заметим, что параметр s является длиной дуги кривой $a(s)$ но не является длиной дуги для эвольвенты $r(s)$, поэтому радиус-вектора эвольвенты выразится правой частью формулы (6.8). Подставим в

формулу эволюты (6.8) значения векторной функции \mathbf{r} и ее производных, выраженные через векторную функцию \mathbf{a} и ее производные по параметру s :

$$\begin{aligned} \mathbf{r} &= \mathbf{a} - (s - s_0)\mathbf{t}, \\ \mathbf{r}' &= \mathbf{t} - (s - s_0)k\mathbf{n} - \mathbf{t} = -(s - s_0)k\mathbf{n}, \\ \mathbf{r}'' &= (s - s_0)k^2\mathbf{t} - k\mathbf{n}, \\ \mathbf{r}' \cdot \mathbf{r}' &= (s - s_0)^2k^2, \\ \mathbf{r}' \cdot \mathbf{r}'' &= (s - s_0)k^2, \\ \mathbf{r}' \times \mathbf{r}'' &= |-(s - s_0)k\mathbf{n} \times ((s - s_0)k^2\mathbf{t} - k\mathbf{n})| = (s - s_0)^2k^3, \end{aligned}$$

где s — длина дуги кривой $\mathbf{a}(s)$, k — ее кривизна, \mathbf{n} — главная нормаль и \mathbf{t} — касательная кривой $\mathbf{a}(s)$.

В результате получим

$$\begin{aligned} \mathbf{r} + \frac{\mathbf{r}''(\mathbf{r}' \cdot \mathbf{r}') - \mathbf{r}'(\mathbf{r}'' \cdot \mathbf{r}')}{|\mathbf{r}' \times \mathbf{r}''|^2}(\mathbf{r}' \cdot \mathbf{r}') &= \mathbf{a} - (s - s_0)\mathbf{t} + \\ &+ \frac{((s - s_0)k^2\mathbf{t} - k\mathbf{n})(s - s_0)^2k^2 + (s - s_0)k\mathbf{n}(s - s_0)k^2}{(s - s_0)^4k^6}(s - s_0)^2k^2 = \mathbf{a}, \end{aligned}$$

что и требовалось доказать.

1.7 Геометрия поверхностей

Поверхностью будем называть геометрическое место точек, координаты которых описываются непрерывными и однозначными функциями $r_1(u, v)$, $r_2(u, v)$, $r_3(u, v)$ двух параметров u и v , принадлежащих непрерывной и связной двумерной области Ω . В декартовой прямоугольной системе координат $Oxyz$ поверхность можно описать радиус-вектором

$$\mathbf{r}(u, v) = \sum_{i=1}^3 r_i(u, v)\mathbf{e}_i, \quad u, v \in \Omega. \tag{7.1}$$

Представление поверхности в виде (7.1) называется параметрическим. В общем случае область изменения параметров представляет собой

часть двумерной плоскости. В простейшем случае область изменения параметров поверхности представляет собой прямоугольник, который можно описать четырьмя числами u_{\min} , u_{\max} , v_{\min} , v_{\max} . Область изменения параметров для простейшего случая определяется неравенствами

$$u_{\min} \leq u \leq u_{\max}, \quad v_{\min} \leq v \leq v_{\max}.$$

Такую область изменения параметров могут иметь очень многие поверхности, но для построения математической модели детали сложной формы необходимы, поверхности общего вида. Два параметра u и v определяют точку поверхности. Если эти параметры сами являются функциями двух других параметров u_0 и v_0 , $u = u(u_0, v_0)$, $v = v(u_0, v_0)$ то поверхность будет описываться радиус-вектором

$$\mathbf{r}(u_0, v_0) = \mathbf{r}(u(u_0, v_0), v(u_0, v_0)), \quad u_0, v_0 \in \Omega_0,$$

зависящим от параметров u_0, v_0 . Новые параметры u_0, v_0 имеют свою область определения Ω_0 . Форма поверхности не зависит от способа ее параметризации и мы имеем возможность использовать удобную для нас параметризацию и область изменения параметров. Параметры поверхности являются координатами внутренней двумерной системы координат, в качестве которой для простоты мы будем выбирать декартову прямоугольную систему координат на плоскости.

Частные производные радиус-вектора поверхности $\mathbf{r}(u, v)$, по параметрам u и v представляют собой векторы. Будем обозначать частные производные радиус-вектора поверхности численными нижними индексами, соответствующими номеру параметра, по которому выполнено дифференцирование

$$\begin{aligned} \mathbf{r}_1 &= \frac{\partial \mathbf{r}}{\partial u}, & \mathbf{r}_2 &= \frac{\partial \mathbf{r}}{\partial v}, \\ \mathbf{r}_{11} &= \frac{\partial^2 \mathbf{r}}{\partial u^2}, & \mathbf{r}_{22} &= \frac{\partial^2 \mathbf{r}}{\partial v^2}, \\ \mathbf{r}_{12} &= \frac{\partial^2 \mathbf{r}}{\partial u \partial v} = \mathbf{r}_{21} = \frac{\partial^2 \mathbf{r}}{\partial v \partial u}. \end{aligned} \quad (7.2)$$

Далее будем предполагать, что координатные функции $r_i(u, v)$ имеют непрерывные производные по каждому параметру любого порядка, который нам потребуется. В дальнейшем ограничимся рассмотрением поверхностей, для которых векторы \mathbf{r}_1 и \mathbf{r}_2 не коллинеарны и не обращаются в нуль. Такую точку поверхности $\mathbf{r}(u, v)$ будем называть обыкновенной. В противном случае точку поверхности будем называть особой. В обыкновенной точке существует единственная касательная к поверхности плоскость. Параметры u и v можно рассматривать как координаты двухмерной точки в плоскости параметров, тогда радиус-вектор поверхности можно рассматривать как функцию точки плоскости параметров.

При изменении параметризации поверхности $\mathbf{r}(u, v)$, где $u = u(u_0, v_0)$, $v = v(u_0, v_0)$ производные ее радиус-вектора по новым параметрам выражаются через производные по старым параметрам следующим образом:

$$\begin{aligned} \frac{\partial \mathbf{r}}{\partial u_0} &= \frac{\partial \mathbf{r}}{\partial u} \frac{\partial u}{\partial u_0} + \frac{\partial \mathbf{r}}{\partial v} \frac{\partial v}{\partial u_0} = \mathbf{r}_1 \frac{\partial u}{\partial u_0} + \mathbf{r}_2 \frac{\partial v}{\partial u_0}, \\ \frac{\partial \mathbf{r}}{\partial v_0} &= \frac{\partial \mathbf{r}}{\partial u} \frac{\partial u}{\partial v_0} + \frac{\partial \mathbf{r}}{\partial v} \frac{\partial v}{\partial v_0} = \mathbf{r}_1 \frac{\partial u}{\partial v_0} + \mathbf{r}_2 \frac{\partial v}{\partial v_0}, \\ \frac{\partial^2 \mathbf{r}}{\partial u_0^2} &= \mathbf{r}_{11} \left(\frac{\partial u}{\partial u_0} \right)^2 + 2\mathbf{r}_{12} \frac{\partial u}{\partial u_0} \frac{\partial v}{\partial u_0} + \mathbf{r}_{22} \left(\frac{\partial v}{\partial u_0} \right)^2 + \mathbf{r}_1 \frac{\partial^2 u}{\partial u_0^2} + \mathbf{r}_2 \frac{\partial^2 v}{\partial u_0^2}, \\ \frac{\partial^2 \mathbf{r}}{\partial v_0^2} &= \mathbf{r}_{11} \left(\frac{\partial u}{\partial v_0} \right)^2 + 2\mathbf{r}_{12} \frac{\partial u}{\partial v_0} \frac{\partial v}{\partial v_0} + \mathbf{r}_{22} \left(\frac{\partial v}{\partial v_0} \right)^2 + \mathbf{r}_1 \frac{\partial^2 u}{\partial v_0^2} + \mathbf{r}_2 \frac{\partial^2 v}{\partial v_0^2}, \\ \frac{\partial^2 \mathbf{r}}{\partial u_0 \partial v_0} &= \mathbf{r}_{11} \frac{\partial u}{\partial u_0} \frac{\partial u}{\partial v_0} + \mathbf{r}_{12} \frac{\partial u}{\partial u_0} \frac{\partial v}{\partial v_0} + \\ &\quad + \mathbf{r}_{21} \frac{\partial u}{\partial v_0} \frac{\partial v}{\partial u_0} + \mathbf{r}_{22} \frac{\partial v}{\partial u_0} \frac{\partial v}{\partial v_0} + \mathbf{r}_1 \frac{\partial^2 u}{\partial u_0 \partial v_0} + \mathbf{r}_2 \frac{\partial^2 v}{\partial u_0 \partial v_0}. \end{aligned}$$

Если зафиксировать один из параметров, а другой изменять в некоторых пределах, то мы получим кривую линию, которая лежит на

поверхности. Такие кривые называются координатными линиями поверхности. Будем называть u -линиями поверхностные кривые, вдоль которых меняется только параметр u , а v -линиями — поверхностные кривые, вдоль которых меняется только параметр v . Производные \mathbf{r}_1 и \mathbf{r}_2 радиус-вектора поверхности представляют собой векторы, касательные к соответствующим координатным линиям.

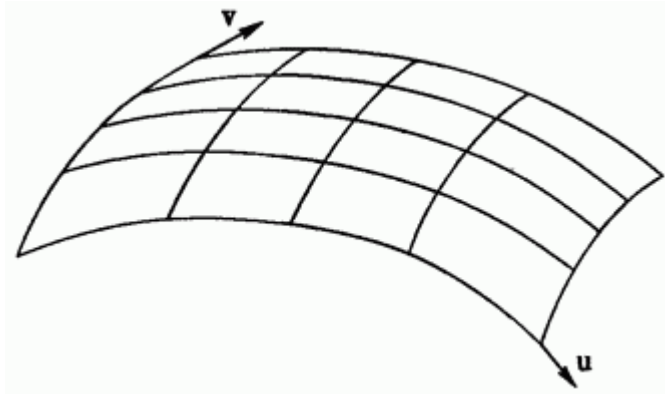


Рис. 7.1. Координатная сетка поверхности

Для поверхности можно построить семейство u -кривых при различных значениях v , и семейство v -кривых при различных значениях u . Два таких семейства кривых образуют координатную сетку на поверхности (рис. 7.1). На координатной сетке поверхности можно построить криволинейную систему координат в пространстве.

Произвольную линию на поверхности можно построить, если ввести зависимость параметров поверхности u и v от некоторого общего для них параметра t

$$\mathbf{u} = \mathbf{u}(t), \quad \mathbf{v} = \mathbf{v}(t). \quad (7.3)$$

Радиус-вектор линии на поверхности будет описываться зависимостью

$$\mathbf{r}(t) = \mathbf{r}(\mathbf{u}(t), \mathbf{v}(t)). \quad (7.4)$$

Через каждую точку поверхности можно провести множество различных кривых (7.4). Найдем дифференциал радиус-вектора (7.4) вдоль некоторой кривой на поверхности:

$$d\mathbf{r} = \mathbf{r}_1(u, v) du + \mathbf{r}_2(u, v) dv, \quad (7.5)$$

где $du = u'dt$, $dv = v'dt$. В дальнейшем будем считать, что u' и v' не обращаются в нуль одновременно, так как в противном случае точка перестает быть обыкновенной. Вектор производной (7.4)

$$\frac{d\mathbf{r}}{dt} = \mathbf{r}_1 \frac{du}{dt} + \mathbf{r}_2 \frac{dv}{dt} \equiv \mathbf{r}_1 u' + \mathbf{r}_2 v' \quad (7.6)$$

лежит в плоскости, определяемой векторами \mathbf{r}_1 и \mathbf{r}_2 . Если зафиксировать точку на поверхности и провести через нее всевозможные кривые, изменяя зависимости (7.3), то производные всех этих кривых в рассматриваемой точке будут лежать в плоскости, определяемой векторами \mathbf{r}_1 и \mathbf{r}_2 . Эта плоскость называется касательной плоскостью поверхности в данной точке (рис. 7.2).

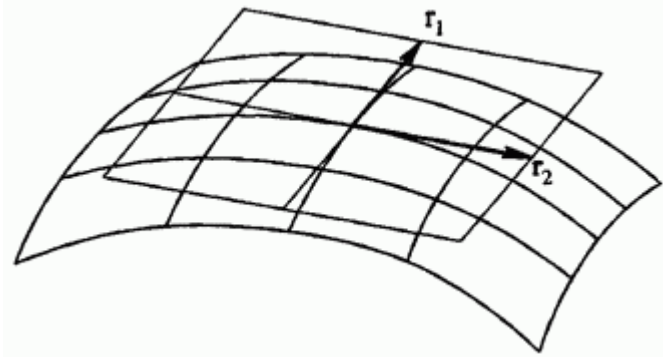


Рис. 7.2. Касательная плоскость и производные радиус-вектора поверхности

Дифференциал (7.5) имеет то же направление, что и производная (7.6). Из (7.5) видно, что направление дифференциала и производной (7.6) зависят от отношения $du:dt = u':v'$.

Пространственной кривой (7.4) соответствует двухмерная кривая

$$l_{uv}(t) = [u(t) \ v(t)]^T$$

на параметрической плоскости поверхности.

Первая квадратичная форма поверхности.

Метрические свойства поверхности выражаются через метрические свойства кривых на них. Мы будем исследовать геометрию поверхности в малой окрестности некоторой ее точки R , определяемой параметрами u и v . Сместимся из точки R по какой-нибудь кривой на поверхности (7.3) в бесконечно близкую ей точку R_1 , определяемую параметрами $u+du$, $v+dv$ и вычислим длину дуги RR_1 . С точностью до слагаемых, линейно зависящих от бесконечно малых величин, длина дуги RR_1 равна длине дифференциала (7.5)

$$ds = |\mathbf{r}_1 du + \mathbf{r}_2 dv|.$$

Квадрат длины бесконечно малой дуги равен

$$ds^2 = (\mathbf{r}_1 du + \mathbf{r}_2 dv) \cdot (\mathbf{r}_1 du + \mathbf{r}_2 dv) = \mathbf{r}_1 \cdot \mathbf{r}_1 du^2 + 2\mathbf{r}_1 \cdot \mathbf{r}_2 du dv + \mathbf{r}_2 \cdot \mathbf{r}_2 dv^2. \quad (7.7)$$

Скалярные произведения векторов \mathbf{r}_1 и \mathbf{r}_2 , как и сами векторы, зависят только от выбора точки R и являются некоторой характеристикой поверхности в этой точке. Введем обозначения

$$\begin{aligned} g_{11} &= \mathbf{r}_1 \cdot \mathbf{r}_1, \\ g_{12} &= g_{21} = \mathbf{r}_1 \cdot \mathbf{r}_2 = \mathbf{r}_2 \cdot \mathbf{r}_1, \\ g_{22} &= \mathbf{r}_2 \cdot \mathbf{r}_2. \end{aligned} \quad (7.8)$$

Тогда квадрат длины бесконечно малой дуги кривой на поверхности будет определяться формулой

$$ds^2 = g_{11} du^2 + 2g_{12} du dv + g_{22} dv^2. \quad (7.9)$$

Выражение в правой части (7.9) является квадратичной формой дифференциалов du, dv . Оно называется первой основной квадратичной формой поверхности. Первая основная квадратичная форма является метрической характеристикой поверхности и служит для бесконечно малых измерений вдоль поверхности. Выражение (7.9) можно записать в матричном виде

$$ds^2 = [du \ dv] \cdot \mathbf{G} \cdot [du \ dv]^T, \quad (7.10)$$

где $\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$ — матрица первой квадратичной формы.

С помощью первой квадратичной формы можно вычислить длину дуги кривой на поверхности. Пусть задан участок кривой на поверхности $u=u(t), v=v(t), t_0 \leq t \leq t_1$. Интегрируя дифференциал длины дуги ds в пределах от t_0 до t_1 , получим длину соответствующего участка кривой

$$s = \int_{t_0}^{t_1} \sqrt{g_{11}u'u' + 2g_{12}u'v' + g_{22}v'v'} dt. \quad (7.11)$$

Первая квадратичная форма поверхности позволяет вычислять углы между кривыми на поверхности. Пусть имеются две кривые на поверхности, проходящие через общую точку M . Обозначим через du, dv дифференциалы параметров поверхности, соответствующие бесконечно малому смещению вдоль первой кривой на поверхности, а через $\delta u, \delta v$ — дифференциалы параметров поверхности, соответствующие бесконечно малому смещению вдоль второй кривой. Эти бесконечно малые смещения определяются векторами

$$d\mathbf{r} = \mathbf{r}_1 du + \mathbf{r}_2 dv, \quad \delta\mathbf{r} = \mathbf{r}_1 \delta u + \mathbf{r}_2 \delta v.$$

Данные дифференциалы направлены по касательным к кривым и к поверхности и поэтому угол между кривыми равен углу между векторами дифференциалов. Найдем косинус угла между кривыми на

поверхности как скалярное произведение касательных к ним векторов, деленное на произведение длин этих векторов:

$$\begin{aligned} \cos(\widehat{d\mathbf{r} d\mathbf{r}}) &= \frac{d\mathbf{r} \cdot \delta\mathbf{r}}{|d\mathbf{r}||\delta\mathbf{r}|} = \\ &= \frac{g_{11} du du + g_{12} du dv + g_{21} dv du + g_{22} dv dv}{\sqrt{(g_{11} du du + 2g_{12} du dv + g_{22} dv dv)(g_{11} \delta u \delta u + 2g_{12} \delta u \delta v + g_{22} \delta v \delta v)}}. \end{aligned} \quad (7.12)$$

Выражение (7.12) позволяет найти угол φ между координатной

u -линией и v -линией в некоторой точке поверхности, если мы положим в ней $du \neq 0, dv = 0, \delta u = 0, \delta v \neq 0$:

$$\cos \varphi = \frac{g_{12}}{\sqrt{g_{11}g_{22}}}. \quad (7.13)$$

Знак косинуса угла мы взяли совпадающим со знаком g_{12} следовательно, исходя из обозначения (7.8), угол φ мы измеряем между положительными направлениями координатных u -линии и v -линии в рассматриваемой точке. Если $g_{12} = 0$, то координатные линии в данной точке ортогональны.

Первая квадратичная форма используется и для вычисления площади поверхности или ее части. Пусть нам требуется вычислить площадь части поверхности, которая на параметрической плоскости имеет область Ω . Разобьем область Ω конечным числом

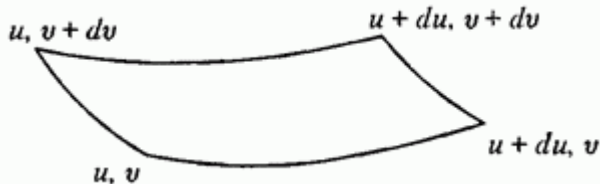


Рис. 7.3. Бесконечно малый элемент поверхности

координатных u -линий и v -линий. Рассматриваемая часть поверхности разобьется на конечное число так называемых криволинейных

четыреугольников. Один из таких четырехугольников показан на рис. 7.3.

Пусть стороны этого четырехугольника равны du , dv . В первом приближении площадь четырехугольника равна

$$dS = |\mathbf{r}_1 du \times \mathbf{r}_2 dv| = |\mathbf{r}_1 \times \mathbf{r}_2| du dv.$$

Представим квадрат длины вектора $\mathbf{r}_1 \times \mathbf{r}_2$ следующим образом:

$$\begin{aligned} |\mathbf{r}_1 \times \mathbf{r}_2|^2 &= |\mathbf{r}_1|^2 |\mathbf{r}_2|^2 \sin^2 \varphi = \\ &= |\mathbf{r}_1|^2 |\mathbf{r}_2|^2 (1 - \cos^2 \varphi) = g_{11} g_{22} \left(1 - \frac{g_{12}^2}{g_{11} g_{22}}\right) = g_{11} g_{22} - g_{12}^2. \end{aligned} \quad (7.14)$$

Таким образом, площадь четырехугольника в первом приближении определится формулой

$$dS = \sqrt{g_{11} g_{22} - g_{12}^2} du dv. \quad (7.15)$$

Будем теперь измельчать разбиение, увеличивая число координатных линий, таким образом, чтобы наибольшие из значений du , dv стремились к нулю. За искомую площадь примем предел, к которому стремится сумма площадей криволинейных четырехугольников при бесконечном измельчении координатной сетки. Из теории кратных интегралов известно, что если функция $g_{11} g_{22} - g_{12}^2$ непрерывна по u и v , то предел указанной суммы существует, не зависит от способа разбиения и равен двойному интегралу от указанной функции по области Ω изменения переменных u и v

$$S = \iint_{\Omega} \sqrt{g_{11} g_{22} - g_{12}^2} du dv. \quad (7.16)$$

Заметим, что так как первая квадратичная форма в (7.9) определяет квадрат длины дуги, который в обыкновенной точке больше нуля при всех du , dv одновременно не равных нулю, то определитель матрицы G положителен:

$$g = |\mathbf{G}| = g_{11}g_{22} - g_{12}^2 > 0. \quad (7.17)$$

Первая квадратичная форма поверхности используется для вычислений длин кривых на поверхности, углов между ними, площади поверхности. Те геометрические свойства поверхности, которые можно установить из знания первой квадратичной формы, называются внутренней геометрией поверхности.

Вторая квадратичная форма поверхности.

Найдем нормаль к поверхности. Так как векторы \mathbf{r}_1 и \mathbf{r}_2 лежат в касательной к поверхности плоскости, то их векторное произведение перпендикулярно к касательной плоскости. Поделив вектор $\mathbf{r}_1 \times \mathbf{r}_2$ на его длину, определяемую выражением (7.14), получим формулу для единичной нормали к поверхности в рассматриваемой точке

$$\mathbf{m} = \frac{\mathbf{r}_1 \times \mathbf{r}_2}{\sqrt{g_{11}g_{22} - g_{12}^2}}. \quad (7.18)$$

Продолжим исследование поверхности вблизи некоторой ее точки R путем исследования кривых на ней. Рассмотрим некоторую кривую на

поверхности $\mathbf{u} = u(t), \quad \mathbf{v} = v(t)$. Вычислим приращение радиус-вектора кривой, которое он получит при бесконечно малом приращении ее параметра dt . Для этого используем разложение радиус-вектора в ряд Тейлора (5.5). Приращение радиус-вектора кривой $\Delta \mathbf{r}$ с точностью до второго порядка малости относительно dt равно

$$\Delta \mathbf{r} = \frac{d\mathbf{r}}{dt} dt + \frac{1}{2} \frac{d^2\mathbf{r}}{dt^2} dt^2. \quad (7.19)$$

Найдем проекцию приращения радиус-вектора кривой $\Delta \mathbf{r}$ на нормаль к поверхности в данной точке. Для этого используем производные радиус-вектора кривой

$$\begin{aligned} \mathbf{r}' &= \mathbf{r}_1 u' + \mathbf{r}_2 v', \\ \mathbf{r}'' &= \mathbf{r}_{11} u'^2 + 2\mathbf{r}_{12} u'v' + \mathbf{r}_{22} v'^2 + \mathbf{r}_1 u'' + \mathbf{r}_2 v''. \end{aligned} \quad (7.20)$$

Вектор \mathbf{r}' ортогонален вектору нормали, поэтому основную роль в проекции на нормаль приращения радиус-вектора $\Delta \mathbf{r}$ будет играть вторая производная. Проекция вектора \mathbf{r}'' на нормаль к поверхности \mathbf{m} характеризует искривление поверхности (именно поверхности, а не кривой). Если мы имеем кривую на плоскости, то как бы она искривлена ни была, проекция вектора \mathbf{r}'' на нормаль к плоскости будет равна нулю. Умножив скалярно (7.19) на нормаль к поверхности, получим

$$\Delta \mathbf{r} \cdot \mathbf{m} = \frac{1}{2} \mathbf{r}'' \cdot \mathbf{m} dt^2 = \frac{1}{2} (\mathbf{r}_{11} \cdot \mathbf{m} u'^2 + 2\mathbf{r}_{12} \cdot \mathbf{m} u'v' + \mathbf{r}_{22} \cdot \mathbf{m} v'^2) dt^2.$$

Скалярные произведения векторов $\mathbf{r}_{11} \cdot \mathbf{m}$, $\mathbf{r}_{12} \cdot \mathbf{m}$, $\mathbf{r}_{22} \cdot \mathbf{m}$ зависят только от выбора точки на поверхности и являются некоторой ее характеристикой в этой точке. Введем обозначения

$$\begin{aligned} b_{11} &= \mathbf{r}_{11} \cdot \mathbf{m}, \\ b_{12} &= b_{21} = \mathbf{r}_{12} \cdot \mathbf{m} = \mathbf{r}_{21} \cdot \mathbf{m}, \\ b_{22} &= \mathbf{r}_{22} \cdot \mathbf{m}. \end{aligned} \quad (7.21)$$

Тогда для главной части отклонения кривой на поверхности от касательной плоскости получим значение

$$\Delta \mathbf{r} \cdot \mathbf{m} = \frac{1}{2} (b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2). \quad (7.22)$$

В (7.22) использовались обозначения (7.21). Выражение в скобках правой части (7.22) является квадратичной формой от дифференциалов du , dv . Оно называется второй основной квадратичной формой поверхности. Также как и первая квадратичная форма, вторая квадратичная форма является характеристикой поверхности и служит для определения искривленности поверхности. Выражение (7.22) можно записать в матричном виде

$$2\Delta\mathbf{r} \cdot \mathbf{m} = [du \ dv] \cdot \mathbf{B} \cdot [du \ dv]^T, \quad (7.23)$$

где $\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ — матрица второй квадратичной формы.

Компоненты второй квадратичной формы можно выразить несколько иначе. Используем тот факт, что вектор нормали \mathbf{m} всегда ортогонален векторам \mathbf{r}_1 и \mathbf{r}_2 . Дифференцируя равенства $\mathbf{m} \cdot \mathbf{r}_1 = 0$, $\mathbf{m} \cdot \mathbf{r}_2 = 0$ по u и v , получим

$$\begin{aligned} \mathbf{m}_1 \cdot \mathbf{r}_1 + \mathbf{m} \cdot \mathbf{r}_{11} &= 0, \\ \mathbf{m}_2 \cdot \mathbf{r}_1 + \mathbf{m} \cdot \mathbf{r}_{12} &= 0, \\ \mathbf{m}_1 \cdot \mathbf{r}_2 + \mathbf{m} \cdot \mathbf{r}_{21} &= 0, \\ \mathbf{m}_2 \cdot \mathbf{r}_2 + \mathbf{m} \cdot \mathbf{r}_{22} &= 0, \end{aligned}$$

где \mathbf{m}_1 , \mathbf{m}_2 — частные производные вектора нормали по параметрам u и v . Подставим в эти равенства обозначения (7.21):

$$\begin{aligned} b_{11} &= -\mathbf{m}_1 \cdot \mathbf{r}_1, \\ b_{12} &= -\mathbf{m}_2 \cdot \mathbf{r}_1, \\ b_{21} &= -\mathbf{m}_1 \cdot \mathbf{r}_2, \\ b_{22} &= -\mathbf{m}_2 \cdot \mathbf{r}_2. \end{aligned} \quad (7.24)$$

Далее вычислим дифференциалы

$$\begin{aligned} d\mathbf{r} &= \mathbf{r}_1 du + \mathbf{r}_2 dv, \\ d\mathbf{m} &= \mathbf{m}_1 du + \mathbf{m}_2 dv \end{aligned}$$

и перемножим их скалярно. В результате придем к равенству

$$\begin{aligned} d\mathbf{r} \cdot d\mathbf{m} &= \mathbf{m}_1 \cdot \mathbf{r}_1 du^2 + (\mathbf{m}_2 \cdot \mathbf{r}_1 + \mathbf{m}_1 \cdot \mathbf{r}_2) du dv + \mathbf{m}_2 \cdot \mathbf{r}_2 dv^2 = \\ &= -(b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2). \end{aligned} \quad (7.25)$$

Дифференциалы $d\mathbf{r}$, $d\mathbf{m}$ берутся в одной и той же точке поверхности при одних и тех же du , dv .

Деривационные формулы Вейнгартена.

Для построения эквидистантных поверхностей нам необходимо будет уметь определять производные вектора нормали. Вектор \mathbf{m} характерен тем, что он всегда имеет единичную длину, и поэтому его производные по параметрам u и v не содержат составляющих, параллельных \mathbf{m} , т. е. его производные лежат в касательной плоскости. Действительно,

$$\frac{\partial(\mathbf{m} \cdot \mathbf{m})}{\partial u} = 2\mathbf{m}_1 \cdot \mathbf{m} = 0, \quad \frac{\partial(\mathbf{m} \cdot \mathbf{m})}{\partial v} = 2\mathbf{m}_2 \cdot \mathbf{m} = 0.$$

Используя это, представим производные нормали в виде

$$\begin{aligned} \mathbf{m}_1 &= -b_1^1 \mathbf{r}_1 - b_1^2 \mathbf{r}_2, \\ \mathbf{m}_2 &= -b_2^1 \mathbf{r}_1 - b_2^2 \mathbf{r}_2, \end{aligned} \quad (7.26)$$

где $b^1_1, b^2_1, b^1_2, b^2_2$ — пока неизвестные коэффициенты разложения.

Знак минус выбран из соображений удобства в дальнейшем. Для определения этих коэффициентов умножим каждое из равенств (7.26) скалярно на \mathbf{r}_1 и \mathbf{r}_2 . Используя (7.8), получим две системы уравнений

$$\begin{aligned} b_{11} &= b_1^1 g_{11} + b_1^2 g_{12}, \\ b_{12} &= b_1^1 g_{21} + b_1^2 g_{22}, \\ b_{21} &= b_2^1 g_{11} + b_2^2 g_{12}, \\ b_{22} &= b_2^1 g_{21} + b_2^2 g_{22}. \end{aligned} \quad (7.27)$$

Решая эти уравнения относительно искомым коэффициентов, получим

$$\begin{aligned} b_1^1 &= \frac{b_{11}g_{22} - b_{12}g_{12}}{g}, & b_1^2 &= \frac{b_{12}g_{11} - b_{11}g_{21}}{g}, \\ b_2^1 &= \frac{b_{21}g_{22} - b_{22}g_{12}}{g}, & b_2^2 &= \frac{b_{22}g_{11} - b_{21}g_{21}}{g}, \end{aligned} \quad (7.28)$$

где $g = g_{11}g_{22} - g_{12}g_{21} = \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} = |G|$ — определитель матрицы

G . Формулы (7.26) с учетом (7.28) называются деривационными формулами Вейнгартена. Они выражают частные производные нормали к поверхности через коэффициенты первой и второй квадратичных форм в рассматриваемой точке. Запишем эти формулы в матричном виде

$$\begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} = - \begin{bmatrix} b_1^1 & b_1^2 \\ b_2^1 & b_2^2 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} = -\mathbf{B}' \cdot \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix}, \quad (7.29)$$

где \mathbf{B}' — матрица коэффициентов (7.28).

Деривационные формулы Вейнгартена можно записать более кратко, если использовать матрицу, обратную матрице G . Для компонент матрицы, обратной матрице G , введем следующие обозначения:

$$\mathbf{G}^{-1} = \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix}, \quad (7.30)$$

где

$$g^{11} = \frac{g_{22}}{g}, \quad g^{12} = -\frac{g_{12}}{g}, \quad g^{21} = -\frac{g_{21}}{g}, \quad g^{22} = \frac{g_{11}}{g}.$$

Матрица \mathbf{B}' с учетом (1.7.30) равна $\mathbf{B}' = \mathbf{B} \cdot \mathbf{G}^{-1}$, а деривационные формулы Вейнгартена примут вид

$$\begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} = -\mathbf{B} \cdot \mathbf{G}^{-1} \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix}. \quad (7.31)$$

1.8. Кривизна линий на поверхности

Теорема Менье.

Продолжим исследование поведения кривых на поверхности с помощью второй квадратичной формы. Установим зависимость кривизны кривой на поверхности от ориентации ее касательного вектора на соприкасающейся плоскости. Рассмотрим равенство

$$\mathbf{r}'' \cdot \mathbf{m} = \mathbf{r}_{11} \cdot \mathbf{m}u'^2 + 2\mathbf{r}_{12} \cdot \mathbf{m}u'v' + \mathbf{r}_{22} \cdot \mathbf{m}v'^2 = b_{11}u'^2 + 2b_{12}u'v' + b_{22}v'^2. \quad (8.1)$$

Вторая производная радиус-вектора кривой на поверхности согласно (5.22) равна

$$\mathbf{r}'' = \frac{d^2\mathbf{r}}{dt^2} = \frac{d(s't)}{ds} = s''\mathbf{t} + s'^2k\mathbf{n}, \quad (8.2)$$

где \mathbf{n} — главная нормаль кривой на поверхности. Касательный вектор \mathbf{t} кривой лежит в касательной плоскости и ортогонален нормали поверхности \mathbf{m} , поэтому равенство (8.1) перепишем в следующем виде

$$ds^2k\mathbf{n} \cdot \mathbf{m} = b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2. \quad (8.3)$$

Квадрат дифференциала длины дуги кривой определяется равенством (7.9). Обозначим угол между нормалью к поверхности и нормалью к кривой на этой поверхности через ω . Разделим последнее равенство на квадрат дифференциала длины дуги и получим выражение, связывающее кривизну кривой, угол между нормальями ω и коэффициенты первой и второй квадратичных форм поверхности:

$$k \cos \omega = \frac{b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2}{g_{11} du^2 + 2g_{12} du dv + g_{22} dv^2}. \quad (8.4)$$

Для уяснения геометрического смысла последнего соотношения рассмотрим рис. 8.1, где показаны некоторая кривая на поверхности и соответствующее ей нормальное сечение поверхности, проходящее через точку M кривой. Нормальное сечение поверхности есть кривая пересечения поверхности и плоскости, проходящей через нормаль к поверхности и касающейся кривой в точке M . Так как нормальное сечение лежит как на поверхности, так и на секущей плоскости, то нормаль к нему также лежит в этой плоскости и, следовательно, для него $|\cos \omega|=1$.

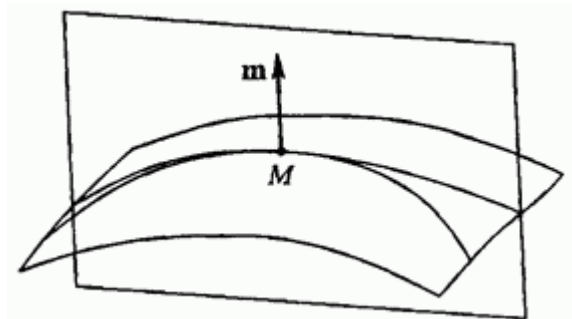


Рис. 8.1. Нормальное сечение поверхности

Коэффициенты первой и второй квадратичных форм зависят только от положения точки M . А вот дифференциалы du, dv зависят от направления кривой на поверхности, поэтому величина $k \cos \omega$ зависит как от положения точки, так и от направления кривой на поверхности, определяемого отношением $du:dv$. Любая другая кривая на поверхности, проходящая через ту же точку M и имеющая общую касательную с рассматриваемым нормальным сечением, будет иметь одно и то же значение $k \cos \omega$, несмотря на то, что у нее другая кривизна. Кривизна такой кривой будет не меньше кривизны нормального сечения, так как нормальное сечение имеет максимальное значение $|\cos \omega|$. Обозначим кривизну нормального сечения через μ . Таким образом, кривизна нормального сечения определяется равенством

$$\mu = \frac{b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2}{g_{11} du^2 + 2g_{12} du dv + g_{22} dv^2}.$$

Угол ω — угол между нормалью к поверхности и нормалью к кривой, он же равен углу между нормальной к поверхности плоскостью и соприкасающейся плоскостью кривой. Таким образом, нормальное сечение имеет минимальную кривизну из всех кривых, проходящих через заданную точку в заданном направлении, и его кривизна является некоторой характеристикой поверхности. Кривизна нормального сечения называется нормальной кривизной поверхности в заданной точке и в заданном направлении. Если известна кривизна нормального сечения, то можно определить кривизну линии на поверхности, касательной к этому нормальному сечению, при условии, что известен угол между нормалью поверхности и главной нормалью кривой.

Этот факт констатирует

Теорема Менье. Радиус кривизны $\rho=1/k$ в заданной точке кривой на поверхности равен произведению радиуса кривизны $\rho_m=1/\mu$ соответствующего нормального сечения в этой точке на косинус угла между нормалью к поверхности и главной нормалью кривой:

$$\frac{1}{k} = \frac{1}{\mu} \cos \omega = \frac{1}{\mu} \mathbf{n} \cdot \mathbf{m}. \quad (8.6)$$

В заданной точке поверхности можно построить бесчисленное множество нормальных сечений, которые отличаются направлением, определяемым отношением $du:dv$. Направление нормального сечения, для которого кривизна нормального сечения равна нулю, называется асимптотическим направлением в рассматриваемой точке. В каждой точке поверхности существует не более двух асимптотических направлений, если не считать те случаи, когда в точке все коэффициенты второй квадратичной формы равны нулю.

Мы рассмотрели проекцию вектора кривизны $k\mathbf{n}$ произвольной кривой на поверхности на нормаль к поверхности \mathbf{m} . Теперь рассмотрим оставшуюся часть вектора кривизны — его проекцию на касательную плоскость, равную

$$\mathbf{h} = k\mathbf{n} - \mathbf{m}(k\mathbf{n} \cdot \mathbf{m}).$$

Длина этого вектора равна $k\sin\phi$ и называется геодезической кривизной линии на поверхности. Вектор \mathbf{h} совпадает с вектором кривизны кривой, являющейся ортогональной проекцией на касательную плоскость рассматриваемой кривой на поверхности. Геодезическая кривизна нормального сечения равна нулю. Нормальная кривизна является характеристикой поверхности, а геодезическая кривизна является характеристикой линии на ней.

К поверхностям применяют такой термин, как изгибание. Изгибание — это изменение формы поверхности, не вызывающее ее деформации. При изгибаниях поверхности ее первая квадратичная форма не меняется, т.е. изгибания не меняют внутреннюю геометрию поверхности. Пусть мы нарисовали некоторую линию на поверхности. При изгибаниях поверхности в общем случае изменяется кривизна этой линии и нормальная кривизна поверхности вдоль линии. Геодезическая кривизна линий на поверхности при изгибаниях остается неизменной. Геодезической линией на поверхности называется кривая на поверхности, геодезическая кривизна которой в каждой точке равна нулю. Длина дуги геодезической линии, проходящей через две заданные точки, меньше длины дуги любой другой кривой на поверхности, соединяющей эти точки.

Рассмотрим случаи, когда кривизна нормального сечения не равна нулю. Если нормальное сечение касательно к координатной u -линии, то $dv=0$ и

$$\mu_u = \frac{b_{11}}{g_{11}},$$

где через μ_u обозначена нормальная кривизна поверхности в u -направлении. Аналогично нормальная кривизна поверхности μ_v в

v -направлении равна

$$\mu_v = \frac{b_{22}}{g_{22}}.$$

Главные кривизны поверхности.

В заданной точке поверхности кривизна нормального сечения зависит от выбранного направления на ней. Выражение (8.5) перепишем в другом виде (используя равенство (7.25))

$$\mu(d\mathbf{r} \cdot d\mathbf{r}) = -(d\mathbf{r} \cdot d\mathbf{m}), \quad (8.7)$$

Обратим внимание на то, что кривизна нормального сечения зависит от направлений векторов $d\mathbf{r}$ и $d\mathbf{m}$ относительно друг друга. Зададимся целью найти такое направление движения по поверхности, при котором векторы $d\mathbf{r}$ и $d\mathbf{m}$ были бы коллинеарными.

Другими словами, попробуем найти направление на поверхности, определяемое отношением $du:dv$, для которого выполняется равенство

$$d\mathbf{m} = -\lambda d\mathbf{r}$$

или

$$\begin{aligned} \mathbf{r}_1 du + \mathbf{r}_2 dv &= -\frac{1}{\lambda}(\mathbf{m}_1 du + \mathbf{m}_2 dv) = \frac{1}{\lambda}(b_1^1 \mathbf{r}_1 + b_1^2 \mathbf{r}_2) du + \frac{1}{\lambda}(b_2^1 \mathbf{r}_1 + b_2^2 \mathbf{r}_2) dv = \\ &= \frac{1}{\lambda}(b_1^1 du + b_2^1 dv)\mathbf{r}_1 + \frac{1}{\lambda}(b_1^2 du + b_2^2 dv)\mathbf{r}_2, \quad (1)(8.8) \end{aligned}$$

где λ — неизвестный пока коэффициент. Мы можем считать \mathbf{r}_1 и \mathbf{r}_2 базисными векторами, по которым разложены векторы $d\mathbf{r}$ и $d\mathbf{m}$. Для их коллинеарности нужно, чтобы коэффициенты при \mathbf{r}_1 и \mathbf{r}_2 в правой и левой частях (8.8) были равны. Это равенство выразится следующим образом:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} du \\ dv \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} b_1^1 & b_2^1 \\ b_1^2 & b_2^2 \end{bmatrix} \cdot \begin{bmatrix} du \\ dv \end{bmatrix}$$

или

$$\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \cdot \begin{bmatrix} du \\ dv \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \cdot \begin{bmatrix} du \\ dv \end{bmatrix}.$$

Для перехода к последнему равенству мы использовали соотношение $B'=B \cdot G^{-1}$. Итак, для определения искомого направления мы пришли к системе линейных алгебраических уравнений для du, dv

$$(B - \lambda G) \begin{bmatrix} du \\ dv \end{bmatrix} = 0. \quad (8.9)$$

Данная система является однородной и имеет ненулевое решение, если определитель ее матрицы равен нулю. Раскрыв определитель, приходим к квадратному уравнению относительно λ , откуда в общем случае найдем два корня: λ_1, λ_2 . Подставив каждый из корней в любое из уравнений (1.8.9), получим два направления на поверхности, определяемые отношениями $du_1:dv_1, du_2:dv_2$.

Направления движения на поверхности, для которых векторы $d\mathbf{r}$ и $d\mathbf{m}$ коллинеарны, называются главными направлениями поверхности. Сравним соотношения (1.8.7) и (1.8.8) и увидим, что λ_1, λ_2 равны кривизне нормальных сечений в главных направлениях, которые обозначим через μ_1, μ_2 . Нормальные сечения в данной точке поверхности, касательные к которым идут по главным направлениям, называются главными сечениями, а их кривизны называются главными кривизнами в данной точке поверхности. Запишем квадратное уравнение, из которого определяются главные кривизны

$$(g_{11}g_{22} - g_{12}^2)\lambda^2 - (b_{11}g_{22} + b_{22}g_{11} - 2b_{12}g_{12})\lambda + (b_{11}b_{22} - b_{12}^2) = 0. \quad (8.10)$$

Из (8.10) легко получить сумму и произведение корней уравнения, т.е. сумму и произведение главных кривизн:

$$\begin{aligned}\mu_1 + \mu_2 &= \frac{b_{11}g_{22} + b_{22}g_{11} - 2b_{12}g_{12}}{g}, \\ \mu_1\mu_2 &= \frac{b_{11}b_{22} - b_{12}b_{21}}{g}.\end{aligned}\tag{8.12}$$

Полусумма главных кривизн называется средней кривизной поверхности в данной точке, а произведение главных кривизн называется гауссовой кривизной поверхности в данной точке.

Обозначим через $\mathbf{t}_1, \mathbf{t}_2$ касательные векторы главных сечений. Покажем, что главные направления поверхности ортогональны друг другу. Выразим главные направления через производные радиус-вектора

$$\begin{aligned}\mathbf{t}_1 &= \mathbf{r}_1 du_1 + \mathbf{r}_2 dv_1, \\ \mathbf{t}_2 &= \mathbf{r}_1 du_2 + \mathbf{r}_2 dv_2.\end{aligned}$$

Их скалярное произведение равно

$$\mathbf{t}_1 \cdot \mathbf{t}_2 = [du_1 \ dv_1] \cdot \mathbf{G} \cdot [du_2 \ dv_2]^T.\tag{8.13}$$

Покажем, что оно в общем случае равно нулю. Для этого систему двух уравнений (8.9) запишем для первого главного направления, первое из этих уравнений умножим на du_2 , второе уравнение умножим на dv_2 и сложим с первым, в результате получим равенство

$$[du_2 \ dv_2] \cdot (\mathbf{B} - \mu_1\mathbf{G}) \cdot [du_1 \ dv_1]^T = 0.$$

Аналогично получим второе равенство, поменяв местами главные направления,

$$[du_1 \ dv_1] \cdot (\mathbf{B} - \mu_2\mathbf{G}) \cdot [du_2 \ dv_2]^T = 0.$$

Вычтем последние два равенства одно из другого и получим равенство:

$$[du_1 \ dv_1] \cdot ((\mu_2 - \mu_1)\mathbf{G}) \cdot [du_2 \ dv_2]^T = 0, \quad (8.14)$$

из которого следует, что если главные кривизны различны, то выражение (8.13) равно нулю и главные направления ортогональны. Если главные кривизны поверхности равны, то за главные могут быть выбраны любые два ортогональных направления (такую ситуацию мы имеем на сфере и плоскости). Точка, в которой $\mu_2 = \mu_1$, называется точкой закругления.

Так как главные направления в общем случае ортогональны, то производные радиус-вектора поверхности и ее нормали в любом направлении можно разложить по единичным векторам $\mathbf{t}_1, \mathbf{t}_2$ главных направлений:

$$\begin{aligned} \frac{d\mathbf{r}}{ds} &= \mathbf{t}_1 \cos \varphi + \mathbf{t}_2 \sin \varphi, \\ \frac{d\mathbf{m}}{ds} &= -\mu_1 \mathbf{t}_1 \cos \varphi - \mu_2 \mathbf{t}_2 \sin \varphi, \end{aligned}$$

где угол φ отсчитывается в касательной плоскости от первого главного направления ко второму. Кривизна нормального сечения в произвольно выбранном направлении с учетом последних равенств и формул (7.25) и (8.4) определится равенством

$$\mu = -\frac{d\mathbf{r}}{ds} \cdot \frac{d\mathbf{m}}{ds} = \mu_1 \cos^2 \varphi + \mu_2 \sin^2 \varphi. \quad (8.15)$$

Формула (8.15) называется формулой Эйлера. Она выражает кривизну произвольного нормального сечения в точке через главные кривизны и угол между нормальным сечением и первым главным направлением. Из этого равенства мы видим, что главные кривизны поверхности μ_1, μ_2 являются максимальной и минимальной кривизнами соответственно.

За определение главных направлений поверхности можно принять следующее: направления, для которых кривизна нормального сечения

принимает максимальное и минимальное значение, называются главными направлениями поверхности.

Гауссова кривизна поверхности (8.12) может быть использована для определения поведения поверхности в некоторой ее точке M . Так как знаменатель в (8.12) больше нуля, то знак гауссовой кривизны зависит от знака числителя, т. е. от знака определителя матрицы \mathbf{B} . Если $|\mathbf{B}| > 0$, то точка M называется эллиптической. Поведение поверхности в эллиптической точке показано на рис. 8.2.

При движении от точки M в любом направлении поверхность изгибается или в сторону нормали или в противоположную сторону в зависимости от знаков главных кривизн.

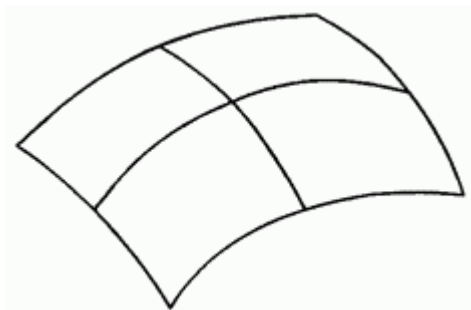


Рис. 8.2. Эллиптическая точка поверхности

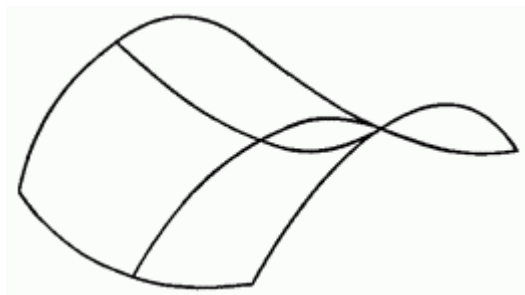


Рис. 8.3. Гиперболическая точка поверхности

Если $|\mathbf{B}| < 0$, то точка М называется гиперболической. Поведение поверхности в гиперболической точке показано на рис. 8.3. Так как в такой точке главные кривизны имеют разные знаки, то согласно (8.15) существуют такие нормальные сечения, для которых выполняется равенство

$$\mu_1 \cos^2 \varphi + \mu_2 \sin^2 \varphi = 0. \quad (8.16)$$

Касательные к нормальным сечениям под углами

$$\varphi = \pm \arctg \left(\sqrt{-\frac{\mu_1}{\mu_2}} \right) \quad (8.17)$$

расположены в касательной плоскости симметрично относительно главных направлений и определяют асимптотические направления в точке М. Если в точке $|\mathbf{B}| = 0$, то такая точка называется параболической. Поведение поверхности в параболической точке показано на рис. 8.4.

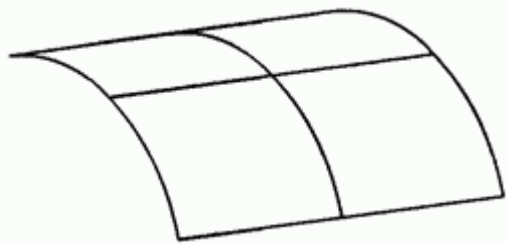


Рис. 8.3. Параболическая точка поверхности

В случае $\mu_1 = \mu_2 = 0$ каждое из направлений является асимптотическим. В противном случае главным направлением является асимптотическое направление, для которого кривизна равна нулю. Соответствующее нормальное сечение в точке М имеет точку распрямления.

Кривая на поверхности называется линией кривизны, если касательная в каждой точке к ней параллельна одному из главных направлений в этой точке поверхности. Линиями кривизны часто являются

координатные линии. Пусть координатные u -линии и v -линии являются линиями кривизны. В этом случае в каждой точке поверхности выполняются равенства

$$g_{12} = g_{21} = b_{12} = b_{21} = 0 \quad (8.18)$$

в силу ортогональности главных направлений.

Справедливо и обратное утверждение: если в каждой точке поверхности выполняются равенства (8.18), то координатные линии являются линиями кривизны. Действительно, в этом случае согласно (7.28) коэффициенты b_1^2, b_2^1 в разложении (7.26) равны нулю и, следовательно, вдоль координатных линий производные нормалей коллинеарны производным радиус-вектора.

Третья квадратичная форма поверхности. Нормаль к поверхности, как и ее радиус-вектор, есть функция параметров u, v . Модуль дифференциала нормали к поверхности равен углу между нормалью в двух бесконечно близких точках, связанных параметрическим смещением du, dv . Квадрат этого угла определяется равенством

$$d\vartheta^2 = d\mathbf{m} \cdot d\mathbf{m} = \mathbf{m}_1 \cdot \mathbf{m}_1 du^2 + 2\mathbf{m}_2 \cdot \mathbf{m}_1 du dv + \mathbf{m}_2 \cdot \mathbf{m}_2 dv^2. \quad (8.19)$$

Введем обозначения

$$\begin{aligned} f_{11} &= \mathbf{m}_1 \cdot \mathbf{m}_1, \\ f_{12} &= f_{21} = \mathbf{m}_1 \cdot \mathbf{m}_2 = \mathbf{m}_2 \cdot \mathbf{m}_1, \\ f_{22} &= \mathbf{m}_2 \cdot \mathbf{m}_2. \end{aligned} \quad (8.20)$$

Равенство (8.19) примет вид

$$d\mathbf{m} \cdot d\mathbf{m} = f_{11} du^2 + 2f_{12} du dv + f_{22} dv^2. \quad (8.21)$$

В правой части (8.21) мы получили квадратичную форму от du, dv . Эта квадратичная форма называется третьей основной квадратичной

формой поверхности. Так же как первая и вторая квадратичные формы она является характеристикой поверхности в заданной точке. Выражение (8.21) можно записать в матричном виде

$$d\mathbf{m} \cdot d\mathbf{m} = [du \ dv] \cdot \mathbf{F} \cdot [du \ dv]^T,$$

где $F = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}$ — матрица третьей квадратичной формы.

Производные вектора нормали по параметрам поверхности ортогональны вектору нормали. Дифференцируя равенства $\mathbf{m} \cdot \mathbf{m}_1 = 0$, $\mathbf{m} \cdot \mathbf{m}_2 = 0$ по параметрам, получим еще одно выражение для коэффициентов третьей квадратичной формы поверхности

$$\begin{aligned} \mathbf{m} \cdot \mathbf{m}_{11} &= -\mathbf{m}_1 \cdot \mathbf{m}_1 = -f_{11}, \\ \mathbf{m} \cdot \mathbf{m}_{12} &= -\mathbf{m}_1 \cdot \mathbf{m}_2 = -f_{12}, \\ \mathbf{m} \cdot \mathbf{m}_{21} &= -\mathbf{m}_2 \cdot \mathbf{m}_1 = -f_{21}, \\ \mathbf{m} \cdot \mathbf{m}_{22} &= -\mathbf{m}_2 \cdot \mathbf{m}_2 = -f_{22}. \end{aligned} \quad (8.22)$$

Таким образом, коэффициенты третьей квадратичной формы отражают проекции на нормаль вторых производных вектора нормали.

Полученные нами три квадратичные формы связаны друг с другом уравнением. Получим его. Для этого выразим производные радиус-вектора и нормали по длине дуги в произвольном направлении, определяемом в касательной плоскости углом φ относительно первого главного направления, через касательные векторы главных сечений $\mathbf{t}_1, \mathbf{t}_2$:

$$\begin{aligned} \frac{d\mathbf{r}}{ds} &= \mathbf{t}_1 \cos \varphi + \mathbf{t}_2 \sin \varphi, \\ \frac{d\mathbf{m}}{ds} &= -\mu_1 \mathbf{t}_1 \cos \varphi - \mu_2 \mathbf{t}_2 \sin \varphi. \end{aligned}$$

Векторы

$$\frac{d\mathbf{m}}{ds} + \mu_1 \frac{d\mathbf{r}}{ds} = (\mu_1 - \mu_2)\mathbf{t}_2 \sin \varphi,$$

$$\frac{d\mathbf{m}}{ds} + \mu_2 \frac{d\mathbf{r}}{ds} = (\mu_2 - \mu_1)\mathbf{t}_1 \cos \varphi$$

коллинеарны главным направлениям, и, следовательно, ортогональны. Перемножив скалярно эти векторы, получим уравнение, связывающее квадратичные формы поверхности

$$d\mathbf{m} \cdot d\mathbf{m} + (\mu_1 + \mu_2) d\mathbf{m} \cdot d\mathbf{r} + \mu_1 \mu_2 d\mathbf{r} \cdot d\mathbf{r} = 0$$

или

$$[du \ dv] \cdot (\mathbf{F} + (\mu_1 + \mu_2)\mathbf{B} + \mu_1 \mu_2 \mathbf{G}) \cdot [du \ dv]^T = 0. \quad (8.23)$$

Так как du, dv в общем случае не равны нулю, то для выполнения соотношения (8.23) должно выполняться матричное равенство

$$\mathbf{F} + (\mu_1 + \mu_2)\mathbf{B} + \mu_1 \mu_2 \mathbf{G} = 0. \quad (8.24)$$

Это и есть уравнение, связывающее коэффициенты трех квадратичных форм поверхности. Из (8.24) следует, что коэффициенты третьей квадратичной формы поверхности выражаются через коэффициенты первой и второй квадратичных форм поверхности соотношением

$$\mathbf{F} = -(\mu_1 + \mu_2)\mathbf{B} - \mu_1 \mu_2 \mathbf{G}. \quad (8.25)$$

1.9. Тензоры поверхности

Далее мы рассмотрим уже известные нам квадратичные формы поверхности несколько с другой точки зрения, а также получим новые связывающие их уравнения. Квадратичные формы являются характеристикой поверхности в некоторой ее точке. Если изменить

параметризацию поверхности не изменяя поверхность, то изменятся коэффициенты квадратичных форм, но останутся неизменными формулы определения площади, главных направлений, главных кривизн, других характеристик поверхности и уравнения, связывающие коэффициенты квадратичных форм. При изменении параметризации квадратичные формы ведут себя аналогично тому, как ведут себя векторы при изменении пространственной системы координат, т. е. меняются описывающие их числа, но не меняются связывающие их соотношения. По аналогии с векторами можно считать, что коэффициенты квадратичных форм являются характеристиками некоторых объектов, связанных с поверхностью. Объекты, математические свойства которых могут быть описаны упорядоченной совокупностью чисел, преобразующихся по определенному закону при переходе от одной координатной системы к другой, называются тензорами. Операции над тензорами не зависят от системы координат, поэтому тензоры называют инвариантными геометрическими объектами. Они описываются своими компонентами, которых может быть довольно много, поэтому прежде чем уточнить определение тензора, мы договоримся о символике записи сумм компонентов объектов.

Соглашение о суммировании по повторяющимся индексам.

Многие из приведенных выше формул примут компактный вид, если мы будем использовать соглашение о суммировании по повторяющимся индексам. Для этого наряду с нижними индексами для некоторых величин будем использовать верхние индексы и смешанные индексы. Так нижние индексы использовались нами для обозначения производных радиус-вектора поверхности и коэффициентов основных квадратичных форм в выражениях (7.9), (7.22), (8.21); верхние индексы использовались для обозначения компонент матрицы G^{-1} в (7.30); смешанные индексы использовались в формулах (7.28).

Соглашение о суммировании заключается в том, что в выражениях выполняется суммирование по повторяющемуся один раз снизу и один раз сверху индексу, при котором индекс пробегает значения от единицы до размерности пространства. Например,

$$a_i b_k^i \equiv \sum_{i=1}^3 a_i b_k^i, \quad a_i b^{ijk} c_k \equiv \sum_{i=1}^3 \sum_{k=1}^3 a_i b^{ijk} c_k.$$

Индексы, по которым производится суммирование, называют немymi, а остальные индексы называют свободными. Немые индексы отсутствуют в одной из частей равенства. Свободные индексы могут принимать значение номера одного из измерений пространства. Чтобы не путаться в индексах при большом их количестве, следует помнить о трех правилах:

- свободные индексы по обе стороны знака равенства занимают одинаковые позиции,
- индексы, по которым производится суммирование, встречаются один раз сверху и один раз внизу и только в одной части равенства,
- индекс параметра, по которому выполняется дифференцирование, является нижним.

Тензоры.

Далее параметры поверхности u и v будем обозначать с помощью верхних индексов через u^1 и u^2 , а их дифференциалы — через du^1 и du^2 соответственно. Использование верхних или нижних индексов для компонент некоторого объекта отражает зависимости, по которым изменяются компоненты этого объекта при переходе от одной системы координат к другой. Так при переходе от параметров u^1 и u^2 к параметрам $u^{1'}$ и $u^{2'}$ (u^1, u^2) производные радиус-вектора поверхности по новым параметрам будут связаны с производными радиус-вектора поверхности по старым параметрам соотношениями

$$\Gamma_{k'} = \sum_{i=1}^2 \frac{\partial u^i}{\partial u^{k'}} \Gamma_i, \quad k = 1, 2, \quad (9.1)$$

тогда как бесконечно малые приращения параметров будут связаны соотношениями

$$du^{k'} = \sum_{i=1}^2 \frac{\partial u^{k'}}{\partial u^i} du^i, \quad k = 1, 2. \quad (9.2)$$

Закон преобразования компонент объекта при изменении системы координат и обуславливает место написания его индексов. Обозначим коэффициенты матриц преобразования параметров через

$$A^{i k'} = \frac{\partial u^i}{\partial u^{k'}},$$
$$A^{k' i} = \frac{\partial u^{k'}}{\partial u^i} \quad (9.3)-(9.4)$$

и запишем равенства (9.1) и (9.2), опустив знак суммы и используя вместо него соглашение о суммировании

$$\mathbf{r}_{k'} = A^{i k'} \mathbf{r}_i,$$
$$du^{k'} = A^{k' i} du^i, \quad (9.5)$$

где k', i принимают значения 1 и 2.

Якобиан матриц преобразования (9.3) и (9.4) должен быть отличен от нуля. Матрицы с компонентами (9.3) и (9.4) описывают однородные линейные преобразования.

При переходе к другой параметризации поверхности компоненты матриц G , V и G^{-1} изменяются по формулам

$$\begin{aligned} g_{k'n'} &= A^i_{k'} A^j_{n'} g_{ij}, \\ b_{k'n'} &= A^i_{k'} A^j_{n'} b_{ij}, \\ g^{k'n'} &= A^{k'}_i A^{n'}_j g^{ij}. \end{aligned} \quad (9.7)-(9.9)$$

Тензорами называются объекты, компоненты которых при изменении системы координат подвергаются линейным однородным преобразованиям, причем матрица преобразования для каждого нижнего индекса составлена из производных (9.3), а матрица преобразования для каждого верхнего индекса составлена из производных (9.4). Коэффициенты первой основной квадратичной формы играют особую роль — они являются компонентами метрического тензора двумерного пространства поверхности. Этот метрический тензор может быть представлен своими ковариантными компонентами в виде матрицы G (7.10) или своими контравариантными компонентами в виде матрицы G^{-1} (7.30). Тензор, представленный своими ковариантными компонентами, называется ковариантным тензором. Ковариантные компоненты имеют нижние индексы. Тензор, представленный своими контравариантными компонентами, называется контравариантным тензором. Контравариантные компоненты имеют верхние индексы. По количеству индексов у компонент тензору приписывают ранг. Так метрический тензор является тензором второго ранга.

С помощью коэффициентов метрического тензора производится «поднятие» и «опускание» индексов компонент других объектов этого пространства, что уже осуществлялось равенствами (7.28) и (7.27), связывающим коэффициенты b_i^k и b_{ij} :

$$\begin{aligned} b_i^k &= b_{ij} g^{jk}, \\ b_{ij} &= b_i^k g_{jk}. \end{aligned} \quad (9.10)$$

В первом равенстве суммирование производится по индексу j , и объект выражается через смешанные компоненты b_i^k которые получаются из b_{ij} заменой индекса j на индекс, стоящий рядом с индексом j в компоненте метрического тензора g^{ik} . Во втором равенстве мы

получили ковариантные компоненты b_{ij} объекта через его смешанные компоненты b^k , а суммирование выполняется по индексу k .

Верхние и нижние индексы являются равноправными. Векторы \mathbf{g}_i образуют местный базис в заданной точке поверхности, по которому можно разложить любой другой вектор в этой точке.

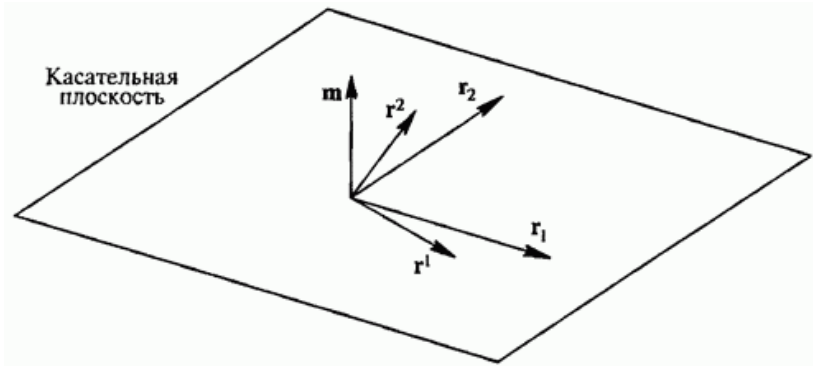


Рис. 9.1. Касательный и взаимный базисы поверхности в точке

Базис $\mathbf{r}_1, \mathbf{r}_2, \mathbf{m}$ называется касательным, так как его первые два вектора направлены по касательным к координатным линиям поверхности. Введем еще один базис, по которому также можно разложить любой другой вектор в этой точке. Этот базис представлен векторами $\mathbf{r}^1, \mathbf{r}^2, \mathbf{m}$.

Он называется взаимным упомянутому базису, и определяется равенствами

$$\begin{aligned} \mathbf{r}^1 \cdot \mathbf{r}_1 &= 1, & \mathbf{r}^1 \cdot \mathbf{r}_2 &= 0, & \mathbf{r}^1 \cdot \mathbf{m} &= 0, \\ \mathbf{r}^2 \cdot \mathbf{r}_1 &= 0, & \mathbf{r}^2 \cdot \mathbf{r}_2 &= 1, & \mathbf{r}^2 \cdot \mathbf{m} &= 0. \end{aligned}$$

Векторы $\mathbf{r}^1, \mathbf{r}^2$ лежат в касательной плоскости, их ориентация показана на рис. 9.1.

Векторы $\mathbf{r}^1, \mathbf{r}^2$ определим равенствами

$$\begin{aligned} \mathbf{r}^1 &= \frac{\mathbf{r}_2 \times \mathbf{m}}{\sqrt{g}} = \frac{\mathbf{r}_2 \times (\mathbf{r}_1 \times \mathbf{r}_2)}{g} = \frac{(\mathbf{r}_2 \cdot \mathbf{r}_2)\mathbf{r}_1 - (\mathbf{r}_2 \cdot \mathbf{r}_1)\mathbf{r}_2}{g} = \\ &= \frac{g_{22}\mathbf{r}_1 - g_{12}\mathbf{r}_2}{g} = g^{11}\mathbf{r}_1 + g^{12}\mathbf{r}_2 = g^{1i}\mathbf{r}_i, \\ \mathbf{r}^2 &= -\frac{\mathbf{r}_1 \times \mathbf{m}}{\sqrt{g}} = -\frac{\mathbf{r}_1 \times (\mathbf{r}_1 \times \mathbf{r}_2)}{g} = -\frac{(\mathbf{r}_1 \cdot \mathbf{r}_2)\mathbf{r}_1 - (\mathbf{r}_1 \cdot \mathbf{r}_1)\mathbf{r}_2}{g} = \\ &= -\frac{g_{12}\mathbf{r}_1 - g_{11}\mathbf{r}_2}{g} = g^{21}\mathbf{r}_1 + g^{22}\mathbf{r}_2 = g^{2i}\mathbf{r}_i. \quad (1.9.12) \end{aligned} \quad (9.12)$$

Эти равенства можно получить другим способом: разложить векторы $\mathbf{r}^1, \mathbf{r}^2$ по векторам $\mathbf{r}_1, \mathbf{r}_2$ аналогично тому, как были разложены производные нормали $\mathbf{m}_1, \mathbf{m}_2$ в (7.26), и найти коэффициенты разложения, что в результате даст

$$\mathbf{r}^k = g^{ki}\mathbf{r}_i.$$

Умножив скалярно векторы взаимного базиса, получим выражение для контравариантных компонент метрического тензора поверхности

$$g^{ki} = \mathbf{r}^k \cdot \mathbf{r}^i. \quad (9.13)$$

Формулы Гаусса и Петерсона-Кодацци.

Найдем разложение вторых производных радиус-вектора поверхности \mathbf{r}_{ij} по местному базису поверхности в этой же точке, представленному векторами $\mathbf{r}_1, \mathbf{r}_2, \mathbf{m}$. Для этого запишем их в виде

$$\mathbf{r}_{ij} = \Gamma_{ij}^k \mathbf{r}_k + b_{ij} \mathbf{m}, \quad (9.14)$$

где Γ_{ij}^k — неизвестные пока коэффициенты разложения векторов \mathbf{r}_{ij} по векторам \mathbf{r}_k . Коэффициенты разложения векторов \mathbf{r}_{ij} по вектору \mathbf{m} равны коэффициентам второй квадратичной формы b_{ij} , так как при скалярном умножении (9.14) на \mathbf{m} мы должны получить равенства (7.21). Умножим (9.14) скалярно на $\mathbf{r}_1, \mathbf{r}_2$ и получим систему уравнений для Γ_{ij}^k

$$\begin{aligned}\Gamma_{ij}^1 g_{11} + \Gamma_{ij}^2 g_{12} &= \Gamma_{ij,1}, \\ \Gamma_{ij}^1 g_{21} + \Gamma_{ij}^2 g_{22} &= \Gamma_{ij,2},\end{aligned}\quad (9.15)$$

где введено обозначение $\Gamma_{ij,1} = \mathbf{r}_{ij} \cdot \mathbf{r}_1$, $\Gamma_{ij,2} = \mathbf{r}_{ij} \cdot \mathbf{r}_2$. Коэффициенты $\Gamma_{ij,k}$ могут быть выражены через производные коэффициентов первой основной квадратичной формы. Для этого выпишем известные равенства

$$\mathbf{r}_j \cdot \mathbf{r}_k = g_{jk}, \quad \mathbf{r}_k \cdot \mathbf{r}_i = g_{ki}, \quad \mathbf{r}_i \cdot \mathbf{r}_j = g_{ij}.$$

Эти равенства выражают одно и то же, просто в них циклически переставляются индексы i, j, k , которые могут принимать значения от единицы до размерности пространства. Продифференцируем первое равенство по u^i , второе — по u^j , третье — по u^k и получим

$$\mathbf{r}_{ji} \cdot \mathbf{r}_k + \mathbf{r}_j \cdot \mathbf{r}_{ki} = \frac{\partial g_{jk}}{\partial u^i}, \quad \mathbf{r}_{kj} \cdot \mathbf{r}_i + \mathbf{r}_k \cdot \mathbf{r}_{ij} = \frac{\partial g_{ki}}{\partial u^j}, \quad \mathbf{r}_{ik} \cdot \mathbf{r}_j + \mathbf{r}_i \cdot \mathbf{r}_{jk} = \frac{\partial g_{ij}}{\partial u^k}.$$

Если сложим первые два равенства и вычтем из них третье, то получим формулу для определения коэффициентов $\Gamma_{ij,k}$

$$\Gamma_{ij,k} = \frac{1}{2} \left(\frac{\partial g_{kj}}{\partial u^i} + \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k} \right). \quad (9.16)$$

Теперь решим систему уравнений (9.15) относительно $\Gamma_{ij,k}$ и получим равенства

$$\begin{aligned}\Gamma_{ij}^1 &= \frac{g_{22}}{g} \Gamma_{ij,1} - \frac{g_{12}}{g} \Gamma_{ij,2} = g^{11} \Gamma_{ij,1} + g^{12} \Gamma_{ij,2}, \\ \Gamma_{ij}^2 &= \frac{g_{11}}{g} \Gamma_{ij,2} - \frac{g_{21}}{g} \Gamma_{ij,1} = g^{12} \Gamma_{ij,1} + g^{22} \Gamma_{ij,2}.\end{aligned}$$

С использованием соглашения о суммировании последние равенства и равенства (9.15) примут вид

$$\begin{aligned}\Gamma_{ij}{}^n &= g^{nk}\Gamma_{ij,k}, \\ \Gamma_{ij,n} &= g_{kn}\Gamma_{ij}{}^k.\end{aligned}\quad (9.17)-(9.18)$$

Подставим в последние равенства значения (9.16) и получим окончательное выражение для коэффициентов в разложении (9.14)

$$\Gamma_{ij}{}^n = \frac{1}{2}g^{nk} \left(\frac{\partial g_{kj}}{\partial u^i} + \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k} \right).\quad (9.19)$$

Коэффициенты $\Gamma_{ij,k}$ называются символами Кристоффеля 1-рода, а коэффициенты $\Gamma_{ij}{}^k$ — символами Кристоффеля 2-рода. Они выражаются через коэффициенты первой основной квадратичной формы поверхности и их частные производные по параметрам. Символы Кристоффеля не являются компонентами тензора.

В (9.14) мы выразили производные векторов $\mathbf{r}_1, \mathbf{r}_2$ по параметрам. Формулы (9.14) называются деривационными формулами Гаусса. Совместно с деривационными формулами Вейнгартена (7.26) они представляют собой деривационные формулы для локального базиса $\mathbf{r}_1, \mathbf{r}_2, \mathbf{m}$ на поверхности. Перепишем их в виде

$$\begin{aligned}\frac{\partial \mathbf{r}_1}{\partial u^i} &= \Gamma_{1i}{}^1 \mathbf{r}_1 + \Gamma_{1i}{}^2 \mathbf{r}_2 + b_{1i} \mathbf{m}, \\ \frac{\partial \mathbf{r}_2}{\partial u^i} &= \Gamma_{2i}{}^1 \mathbf{r}_1 + \Gamma_{2i}{}^2 \mathbf{r}_2 + b_{2i} \mathbf{m}, \\ \frac{\partial \mathbf{m}}{\partial u^i} &= -b_i{}^1 \mathbf{r}_1 - b_i{}^2 \mathbf{r}_2.\end{aligned}\quad (9.20)$$

Эти формулы аналогичны формулам Френе-Серре (5.17) для кривых. Они выражают производные векторов локального базиса $\mathbf{r}_1, \mathbf{r}_2, \mathbf{m}$ в виде разложения по самим этим векторам. Дифференцируя формулы (8.9) еще раз и пользуясь имеющимися разложениями первых

производных, можно получить вторые производные локального базиса, а за ними — третьи производные и так далее.

Например, вторые производные вектора нормали равны

$$\begin{aligned} m_{ij} &= -\frac{\partial b_i^k}{\partial u^j} \mathbf{r}_k - b_i^k \mathbf{r}_{kj} = -\frac{\partial b_i^k}{\partial u^j} \mathbf{r}_k - b_i^k (\Gamma_{kj}^n \mathbf{r}_n + b_{kj} \mathbf{m}) = \\ &= -\left(\frac{\partial b_i^k}{\partial u^j} - b_i^n \Gamma_{nj}^k \right) \mathbf{r}_k - b_i^k b_{kj} \mathbf{m}. \end{aligned} \quad (9.21)$$

Найдем разложение производных векторов $\mathbf{r}^1, \mathbf{r}^2$ по взаимному базису поверхности в этой же точке, представленному векторами $\mathbf{r}^1, \mathbf{r}^2, \mathbf{m}$, дифференцированием формул (1.9.12)

$$\begin{aligned} \frac{\partial \mathbf{r}^1}{\partial u^i} &= \mathbf{r}_{2i} \times \mathbf{m} \frac{1}{\sqrt{g}} + \mathbf{r}_2 \times \mathbf{m}_i \frac{1}{\sqrt{g}} - \mathbf{r}_2 \times \mathbf{m} \frac{1}{2g\sqrt{g}} \frac{\partial g}{\partial u^i} = \\ &= \frac{1}{\sqrt{g}} \Gamma_{2i}^k \mathbf{r}_k \times \mathbf{m} - \frac{1}{\sqrt{g}} b_i^k \mathbf{r}_2 \times \mathbf{r}_k - \mathbf{r}^1 \Gamma_{ki}^k = \\ &= -\Gamma_{2i}^1 \mathbf{r}^2 + \Gamma_{2i}^2 \mathbf{r}^1 + b_i^1 \mathbf{m} - \mathbf{r}^1 (\Gamma_{1i}^1 + \Gamma_{2i}^2) = -\Gamma_{1i}^1 \mathbf{r}^1 - \Gamma_{2i}^1 \mathbf{r}^2 + b_i^1 \mathbf{m}, \\ \frac{\partial \mathbf{r}^2}{\partial u^i} &= -\mathbf{r}_{1i} \times \mathbf{m} \frac{1}{\sqrt{g}} - \mathbf{r}_1 \times \mathbf{m}_i \frac{1}{\sqrt{g}} + \mathbf{r}_1 \times \mathbf{m} \frac{1}{2g\sqrt{g}} \frac{\partial g}{\partial u^i} = \\ &= -\frac{1}{\sqrt{g}} \Gamma_{1i}^k \mathbf{r}_k \times \mathbf{m} + \frac{1}{\sqrt{g}} b_i^k \mathbf{r}_1 \times \mathbf{r}_k - \mathbf{r}^2 \Gamma_{ki}^k = \\ &= \Gamma_{1i}^1 \mathbf{r}^2 - \Gamma_{1i}^2 \mathbf{r}^1 + b_i^2 \mathbf{m} - \mathbf{r}^2 (\Gamma_{1i}^1 + \Gamma_{2i}^2) = -\Gamma_{1i}^2 \mathbf{r}^1 - \Gamma_{2i}^2 \mathbf{r}^2 + b_i^2 \mathbf{m}, \end{aligned}$$

где использовались дериивационные формулы (9.20) и равенство

$$\frac{1}{2g} \frac{\partial g}{\partial u^i} = \Gamma_{1i}^1 + \Gamma_{2i}^2 = \Gamma_{ki}^k, \quad (9.24)$$

Равенство (9.24) следует из соотношения

$$\begin{aligned} \frac{\partial g}{\partial u^i} &= \frac{\partial (g_{11}g_{22} - g_{12}g_{21})}{\partial u^i} = 2\mathbf{r}_{1i} \cdot (g_{22}\mathbf{r}_1 - g_{12}\mathbf{r}_2) + 2\mathbf{r}_{2i} \cdot (g_{11}\mathbf{r}_2 - g_{21}\mathbf{r}_1) = \\ &= 2g\mathbf{r}_{1i} \cdot (g^{11}\mathbf{r}_1 + g^{12}\mathbf{r}_2) + 2g\mathbf{r}_{2i} \cdot (g^{22}\mathbf{r}_2 + g^{21}\mathbf{r}_1) = 2g\mathbf{r}_{ji} \cdot \mathbf{r}_k g^{kj} = 2g\Gamma_{ki}^k. \end{aligned} \quad (9.25)$$

Мы получили дериивационные формулы для взаимного базиса $\mathbf{r}^1, \mathbf{r}^2$ на поверхности, выражающего производные векторов взаимного базиса $\mathbf{r}^1, \mathbf{r}^2, \mathbf{m}$ в виде разложения по самим этим векторам

$$\begin{aligned}\frac{\partial \mathbf{r}^1}{\partial u^i} &= -\Gamma_{1i}{}^1 \mathbf{r}^1 - \Gamma_{2i}{}^1 \mathbf{r}^2 + b_i{}^1 \mathbf{m}, \\ \frac{\partial \mathbf{r}^2}{\partial u^i} &= -\Gamma_{1i}{}^2 \mathbf{r}^1 - \Gamma_{2i}{}^2 \mathbf{r}^2 + b_i{}^2 \mathbf{m}, \\ \frac{\partial \mathbf{m}}{\partial u^i} &= -b_{i1} \mathbf{r}^1 - b_{i2} \mathbf{r}^2.\end{aligned}\tag{9.26}$$

Функции $g_{ij}(u^1, u^2)$, $b_{ij}(u^1, u^2)$ однозначно определяют поверхность с точностью до положения и ориентации в пространстве, аналогично тому, как функции $s(t)$, $k(t)$, $\chi(t)$ определяют кривую. Функции $g_{ij}(u^1, u^2)$, $b_{ij}(u^1, u^2)$ не являются независимыми, они связаны между собой уравнениями, которые мы получим ниже.

Дифференцируя равенство (9.14) по одному из параметров, получим вторую производную вектора \mathbf{r}_i :

$$\begin{aligned}\frac{\partial^2 \mathbf{r}_i}{\partial u^n \partial u^j} &= \frac{\partial \Gamma_{ij}{}^k}{\partial u^n} \mathbf{r}_k + \Gamma_{ij}{}^k \mathbf{r}_{kn} + \frac{\partial b_{ij}}{\partial u^n} \mathbf{m} + b_{ij} \mathbf{m}_n = \\ &= \left(\frac{\partial \Gamma_{ij}{}^m}{\partial u^n} + \Gamma_{ij}{}^k \Gamma_{kn}{}^m - b_{ij} b_n{}^m \right) \mathbf{r}_m + \left(\frac{\partial b_{ij}}{\partial u^n} + \Gamma_{ij}{}^k b_{kn} \right) \mathbf{m}.\end{aligned}\tag{9.27}$$

Изменим последовательность дифференцирования вектора \mathbf{r}_i в (9.27), что приведет к смене мест индексов n, j :

$$\begin{aligned}\frac{\partial^2 \mathbf{r}_i}{\partial u^j \partial u^n} &= \frac{\partial \Gamma_{in}{}^k}{\partial u^j} \mathbf{r}_k + \Gamma_{in}{}^k \mathbf{r}_{kj} + \frac{\partial b_{in}}{\partial u^j} \mathbf{m} + b_{in} \mathbf{m}_j = \\ &= \left(\frac{\partial \Gamma_{in}{}^m}{\partial u^j} + \Gamma_{in}{}^k \Gamma_{kj}{}^m - b_{in} b_j{}^m \right) \mathbf{r}_m + \left(\frac{\partial b_{in}}{\partial u^j} + \Gamma_{in}{}^k b_{kj} \right) \mathbf{m}.\end{aligned}\tag{9.28}$$

Левые части равенств (9.27) и (9.28) равны, следовательно, должны быть равны и правые их части. Вычитая (9.28) из (9.27), получим равенство

$$\begin{aligned}\left(\frac{\partial \Gamma_{ij}{}^m}{\partial u^n} - \frac{\partial \Gamma_{in}{}^m}{\partial u^j} + \Gamma_{ij}{}^k \Gamma_{kn}{}^m - \Gamma_{in}{}^k \Gamma_{kj}{}^m + b_{in} b_j{}^m - b_{ij} b_n{}^m \right) \mathbf{r}_m + \\ + \left(\frac{\partial b_{ij}}{\partial u^n} - \frac{\partial b_{in}}{\partial u^j} + \Gamma_{ij}{}^k b_{kn} - \Gamma_{in}{}^k b_{kj} \right) \mathbf{m} = 0.\end{aligned}\tag{9.29}$$

Так как в обыкновенной точке векторы $\mathbf{r}_1, \mathbf{r}_2, \mathbf{m}$ некопланарны, то для выполнения векторного равенства (9.29) должны выполняться следующие соотношения

$$\begin{aligned} \frac{\partial \Gamma_{ij}^m}{\partial u^n} - \frac{\partial \Gamma_{in}^m}{\partial u^j} + \Gamma_{ij}^k \Gamma_{kn}^m - \Gamma_{in}^k \Gamma_{kj}^m &= b_{ij} b_n^m - b_{in} b_j^m, \\ \frac{\partial b_{ij}}{\partial u^n} - \Gamma_{in}^k b_{kj} &= \frac{\partial b_{in}}{\partial u^j} - \Gamma_{ij}^k b_{kn}. \end{aligned} \quad (9.30)-(9.31)$$

Эти уравнения связывают между собой коэффициенты первой и второй основных квадратичных форм поверхности. Уравнение (9.30) называется уравнением Гаусса. Уравнения (9.31) называются уравнениями Петперсона-Кодацци. Правая часть равенства (9.30) является тензором, следовательно, и его левая часть также представляет собой тензор. Этот тензор называется тензором кривизны или тензором Римана и обозначается

$$R_{jni}^m = \frac{\partial \Gamma_{ij}^m}{\partial u^n} - \frac{\partial \Gamma_{in}^m}{\partial u^j} + \Gamma_{kn}^m \Gamma_{ij}^k - \Gamma_{kj}^m \Gamma_{in}^k. \quad (9.32)$$

Умножим обе части уравнения Гаусса на δ_{in} и, произведя суммирование, опустим в нем верхний индекс

$$\frac{\partial \Gamma_{ij,l}}{\partial u^n} - \frac{\partial \Gamma_{in,l}}{\partial u^j} + \Gamma_{kn,l} \Gamma_{ij}^k - \Gamma_{kj,l} \Gamma_{in}^k = b_{ij} b_{nl} - b_{in} b_{jl}. \quad (9.33)$$

Аналогично получим ковариантные компоненты тензора кривизны

$$R_{jnli} = \frac{\partial \Gamma_{ij,l}}{\partial u^n} - \frac{\partial \Gamma_{in,l}}{\partial u^j} + \Gamma_{kn,l} \Gamma_{ij}^k - \Gamma_{kj,l} \Gamma_{in}^k. \quad (9.34)$$

В силу симметрии символов Кристоффеля по первым двум индексам тензор кривизны (9.34) обладает определенной симметрией, он кососимметричен по первой паре индексов и по второй паре индексов: $R_{jnli} = R_{iljn} = -R_{nijl} = -R_{jnli}$. Для поверхностей равенство (9.34) эквивалентно одному равенству

$$\frac{\partial \Gamma_{11,2}}{\partial u^2} - \frac{\partial \Gamma_{12,2}}{\partial u^1} + \Gamma_{k2,2} \Gamma_{11}^k - \Gamma_{k1,2} \Gamma_{12}^k = b_{11} b_{22} - b_{12} b_{21} = \mu_1 \mu_2 \mathcal{E}. \quad (9.35)$$

Мы использовали равенство (8.12) для гауссовой кривизны поверхности. Таким образом, гауссова кривизна поверхности может быть определена через коэффициенты первой квадратичной формы и их производные. Уравнения Петерсона-Кодацци (9.31) сводятся к двум уравнениям

$$\begin{aligned} \frac{\partial b_{11}}{\partial u^2} - \Gamma_{12}^k b_{k1} &= \frac{\partial b_{12}}{\partial u^1} - \Gamma_{11}^k b_{k2}, \\ \frac{\partial b_{22}}{\partial u^1} - \Gamma_{21}^k b_{k2} &= \frac{\partial b_{21}}{\partial u^2} - \Gamma_{22}^k b_{k1}. \end{aligned} \quad (9.36)-(9.37)$$

Если над тензором кривизны (9.32) выполнить операцию свертки по индексам n, m , то мы получим ковариантные компоненты тензора Риччи

$$R_{ji} = R_{j1i}^1 + R_{j2i}^2 = R_{jni}^n = b_{ij}b_n^n - b_{in}b_j^n = (b_{ij}b_{nk} - b_{in}b_{jk})g^{kn}. \quad (9.38)$$

Операция свертывания заключается в том, у тензора выбираются два индекса, один верхний, другой нижний, и выполняется суммирование всех компонент тензора, в которых два выбранных индекса имеют одинаковые значения. Так как для поверхностей индексы принимают значения 1 и 2, то ковариантные компоненты тензора Риччи поверхности равны

$$R_{ij} = \mu_1 \mu_2 g_{ij}.$$

Найдем уравнения, которым должна удовлетворять геодезическая кривая на поверхности $u^1 = u^1(s)$, $u^2 = u^2(s)$, где s — длина дуги кривой. По определению, проекция вектора кривизны геодезической кривой \mathbf{k}_n на касательную плоскость к поверхности должна быть равна нулю. Найдем вектор кривизны произвольной кривой на поверхности

$$\begin{aligned} \mathbf{k}_n &= \frac{d^2 \mathbf{r}}{ds^2} = r_{11} \left(\frac{du^1}{ds} \right)^2 + 2r_{12} \frac{du^1}{ds} \frac{du^2}{ds} + r_{22} \left(\frac{du^2}{ds} \right)^2 + r_1 \frac{d^2 u^1}{ds^2} + r_2 \frac{d^2 u^2}{ds^2} \equiv \\ &\equiv r_{ij} \frac{du^i}{ds} \frac{du^j}{ds} + r_k \frac{d^2 u^k}{ds^2} = r_k \Gamma_{ij}^k \frac{du^i}{ds} \frac{du^j}{ds} + r_k \frac{d^2 u^k}{ds^2} + m b_{ij} \frac{du^i}{ds} \frac{du^j}{ds}. \end{aligned} \quad (9.39)$$

По определению, вектор геодезической кривизны равен проекции вектора кривизны кривой на касательную плоскость к поверхности, т. е. он равен вектору

$$\mathbf{q} = \left(\frac{d^2 u^k}{ds^2} + \Gamma_{ij}^k \frac{du^i}{ds} \frac{du^j}{ds} \right) \mathbf{r}_k,$$

откуда следует, что координатные функции $u^i(s)$, $i=1, 2$ геодезической кривой должны удовлетворять уравнениям

$$\frac{d^2 u^k}{ds^2} + \Gamma_{ij}^k \frac{du^i}{ds} \frac{du^j}{ds} = 0. \quad (9.40)$$

Сама геодезическая кривизна линии на поверхности равна длине вектора геодезической кривизны \mathbf{q} , приведенного выше.

Для моделируемых нами поверхностей будет известна их векторная функция. От нас будет требоваться умение получить всю необходимую информацию о поверхности по ее радиус-вектору. Выпишем полученные выше равенства, выразив все величины через радиус-вектор поверхности и его производные. Итак, задана векторная функция \mathbf{r} двух параметров u^1, u^2 принимающих значение на связной области Ω :

$$\mathbf{r} = \mathbf{r}(u^1, u^2), \quad u^1, u^2 \in \Omega.$$

По ее производным найдем вектор нормали, взаимный базис и коэффициенты квадратичных форм:

$$\begin{aligned} \mathbf{r}_i &\equiv \frac{\partial \mathbf{r}}{\partial u^i}, & g_{ij} &= \mathbf{r}_i \cdot \mathbf{r}_j, & g &= |\mathbf{G}| = g_{11}g_{22} - g_{12}g_{21}, \\ \mathbf{m} &= \frac{\mathbf{r}_1 \times \mathbf{r}_2}{\sqrt{g}}, & \mathbf{r}^1 &= \frac{\mathbf{r}_2 \times \mathbf{m}}{\sqrt{g}}, & \mathbf{r}^2 &= \frac{\mathbf{m} \times \mathbf{r}_1}{\sqrt{g}}, \\ g^{ij} &= \mathbf{r}^i \cdot \mathbf{r}^j, & b_{ij} &= \mathbf{r}_{ij} \cdot \mathbf{m}, & b_i^j &= -\mathbf{m}_i \cdot \mathbf{r}^j. \end{aligned} \quad (9.41)$$

По вторым и третьим производным найдем символы Кристоффеля и дериационные формулы базисов поверхности в заданной точке

$$\begin{aligned}
 \mathbf{r}_{ik} &\equiv \frac{\partial^2 \mathbf{r}}{\partial u^i \partial u^k}, & \mathbf{r}_{ijk} &\equiv \frac{\partial^3 \mathbf{r}}{\partial u^i \partial u^j \partial u^k}, & \Gamma_{ij,k} &= \mathbf{r}_{ij} \cdot \mathbf{r}_k, & \Gamma_{ij}{}^k &= \mathbf{r}_{ij} \cdot \mathbf{r}^k, \\
 \frac{\partial g_{ij}}{\partial u^k} &= \mathbf{r}_{ik} \cdot \mathbf{r}_j + \mathbf{r}_i \cdot \mathbf{r}_{jk}, & \frac{\partial b_{ij}}{\partial u^k} &= \mathbf{r}_{ijk} \cdot \mathbf{m} + \mathbf{r}_{ij} \cdot \mathbf{m}_k, \\
 \frac{\partial \Gamma_{ij,k}}{\partial u^n} &= \mathbf{r}_{ijn} \cdot \mathbf{r}_k + \mathbf{r}_{ij} \cdot \mathbf{r}_{kn}, & \frac{\partial \Gamma_{ij}{}^k}{\partial u^n} &= \mathbf{r}_{ijn} \cdot \mathbf{r}^k + \mathbf{r}_{ij} \cdot \mathbf{r}^{kn}, \\
 \mathbf{r}_{ik} &\equiv \frac{\partial \mathbf{r}_i}{\partial u^k} = \Gamma_{ik}{}^j \mathbf{r}_j + b_{ik} \mathbf{m}, & \mathbf{r}^i{}_k &\equiv \frac{\partial \mathbf{r}^i}{\partial u^k} = -\Gamma_{kj}{}^i \mathbf{r}^j + b_k{}^i \mathbf{m}, \\
 \mathbf{m}_i &\equiv \frac{\partial \mathbf{m}}{\partial u^i} = -b_{ik} \mathbf{r}^k = -b_i{}^k \mathbf{r}_k, \\
 \mathbf{m}_{ij} &\equiv \frac{\partial^2 \mathbf{m}}{\partial u^i \partial u^j} = -(\mathbf{r}_{ijk} \cdot \mathbf{m} + \mathbf{r}_{ik} \cdot \mathbf{m}_j + \mathbf{r}_{jk} \cdot \mathbf{m}_i) \mathbf{r}^k - b_{ik} b_j{}^k \mathbf{m}.
 \end{aligned} \tag{9.42}$$

В равенствах (9.41) и (9.42) верхние и нижние индексы принимают значения от 1 до 2. Вся необходимая информация о поверхности может быть получена на основе равенств (9.41) и (9.42) при известной векторной функции (7.1). Из них мы видим, что для описания поверхности требуются производные ее радиус-вектора до третьего порядка включительно.

1.10. Криволинейные координаты

Наряду с декартовыми прямоугольными системами координат можно использовать и другие в общем случае криволинейные координатные системы. Такие координатные системы нужны также в неевклидовых пространствах. В криволинейной системе координат при дифференцировании векторных функций приходится учитывать изменение базисных векторов системы, что значительно усложняет описание геометрических объектов. Но в определенных случаях криволинейные координаты удобнее, чем декартовы. Мы уже сталкивались с двухмерной криволинейной системой координат, которая строилась по параметрическим линиям на поверхности. Мы вынуждены были ими пользоваться, так как в общем случае на криволинейной поверхности невозможно построить единую двухмерную декартову прямоугольную систему координат.

Мы рассмотрим криволинейные системы координат, чтобы знать, какие трудности нас ожидают при их использовании и чтобы уметь эти

трудности преодолевать. Так как геометрические объекты не зависят от системы координат, то для их описания можно построить и использовать математический аппарат, инвариантный по отношению к системе координат.

Пусть мы имеем в распоряжении эталон единицы длины и инструмент для измерения углов. С их помощью в той точке пространства, в которой мы находимся, можно построить декартову прямоугольную систему координат с началом в этой точке и базисом e_1, e_2, e_3 . Координаты точек в ней обозначим через x_1, x_2, x_3 . Пусть имеются непрерывные дифференцируемые и однозначные функции

$$u^1 = u^1(x^1, x^2, x^3), \quad u^2 = u^2(x^1, x^2, x^3), \quad u^3 = u^3(x^1, x^2, x^3), \quad (10.1)$$

такие, что обратные им функции

$$x^1 = x^1(u^1, u^2, u^3), \quad x^2 = x^2(u^1, u^2, u^3), \quad x^3 = x^3(u^1, u^2, u^3) \quad (10.2)$$

также являются непрерывными дифференцируемыми и однозначными. Потребуем, чтобы определитель матрицы Якоби системы (10.2) был отличен от нуля:

$$\begin{vmatrix} \frac{\partial x^1}{\partial u^1} & \frac{\partial x^1}{\partial u^2} & \frac{\partial x^1}{\partial u^3} \\ \frac{\partial x^2}{\partial u^1} & \frac{\partial x^2}{\partial u^2} & \frac{\partial x^2}{\partial u^3} \\ \frac{\partial x^3}{\partial u^1} & \frac{\partial x^3}{\partial u^2} & \frac{\partial x^3}{\partial u^3} \end{vmatrix} \neq 0. \quad (10.3)$$

Если один из параметров в (1.10.2) зафиксировать ($u^i = \text{const}$), то мы получим некоторую поверхность в пространстве, которую назовем поверхностью i -группы. Функции (10.2) должны быть такими, чтобы поверхности одной группы не пересекались друг с другом. В этом случае параметры u^1, u^2, u^3 могут служить координатами точек в рассматриваемом пространстве. Поверхность i -группы называют i -й

координатной поверхностью, а линии пересечения координатных поверхностей разных групп называют координатными линиями.

Метрический тензор.

В каждой точке $R(u^1, u^2, u^3)$ пространства можно построить локальную систему координат с началом в данной точке и базисными функциями $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$, определенными следующим образом:

$$\begin{aligned} \mathbf{r}_1 &= \lim_{\delta u^1 \rightarrow 0} \frac{R(u^1 + \delta u^1, u^2, u^3) - R(u^1, u^2, u^3)}{\delta u^1}, \\ \mathbf{r}_2 &= \lim_{\delta u^2 \rightarrow 0} \frac{R(u^1, u^2 + \delta u^2, u^3) - R(u^1, u^2, u^3)}{\delta u^2}, \\ \mathbf{r}_3 &= \lim_{\delta u^3 \rightarrow 0} \frac{R(u^1, u^2, u^3 + \delta u^3) - R(u^1, u^2, u^3)}{\delta u^3}. \end{aligned} \tag{10.4}$$

В силу (10.3) векторы $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ некопланарны и могут быть использованы в качестве базиса для разложения по нему любого другого вектора в этой точке пространства. Этот базис будем называть касательным базисом. Базисные векторы $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ в каждой точке различны, и поэтому по ним можно производить разложение только тех векторов, которые вычислены в этой же точке пространства. Базисный вектор \mathbf{r}_k направлен по касательной к k -й координатной линии в данной точке, его длина в общем случае отлична от единицы.

Квадрат длины бесконечно малого отрезка, заданного бесконечно малыми приращениями координат du^1, du^2, du^3 , равен

$$ds^2 = \mathbf{r}_i \cdot \mathbf{r}_j du^i du^j = g_{ij} du^i du^j.$$

Коэффициенты

$$g_{ij} = g_{ij}(u^1, u^2, u^3) = \mathbf{r}_i \cdot \mathbf{r}_j, \quad i, j = 1, 2, 3 \tag{10.5}$$

являются ковариантными компонентами метрического тензора в точке $R(u^1, u^2, u^3)$.

При переходе к другой криволинейной системе координат $u^{1'}, u^{2'}, u^{3'}$ ковариантные компоненты метрического тензора $g_{k'n'}$ в новой системе координат связаны с компонентами g_{ij} соотношениями

$$g_{k'n'} = A^i_{k'} A^j_{n'} g_{ij},$$

где

$$A^i_{k'} = \frac{\partial u^i}{\partial u^{k'}}.$$

Аналогичными соотношениями связаны и базисные векторы криволинейных систем

$$\mathbf{r}_{k'} = A^i_{k'} \mathbf{r}_i. \quad (10.6)$$

Матрицу, составленную из ковариантных компонент метрического тензора, будем обозначать через G , как и матрицу коэффициентов первой квадратичной формы поверхности. Обратную ей матрицу будем обозначать через G^{-1} .

Для касательного базиса $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ в каждой точке можно построить взаимный базис $\mathbf{r}^1, \mathbf{r}^2, \mathbf{r}^3$ по правилу

$$\mathbf{r}_i \cdot \mathbf{r}^k = \delta_i^k, \quad (10.7)$$

где δ_i^k — символы Кронекера, они принимают значения δ_i^k при $i=k$ и $\delta_i^k=0$ при $i \neq k$. Для нахождения векторов взаимного базиса представим их в виде разложения по касательному базису с неизвестными коэффициентами. Из приведенных выше условий получим систему линейных алгебраических уравнений относительно искомым коэффициентов. После ее решения получим коэффициенты

разложения векторов взаимного базиса по векторам касательного базиса

$$\mathbf{r}^k = g^{ki} \mathbf{r}_i, \quad (1.10.8)$$

где g^{ki} — коэффициенты матрицы G^{-1} . Коэффициенты $g^{ki} = \mathbf{r}^k \cdot \mathbf{r}^i$ называются контравариантными компонентами метрического тензора, заданного в системе координат u^1, u^2, u^3 . При переходе к другой криволинейной системе координат $u^{1'}, u^{2'}, u^{3'}$ контравариантные компоненты метрического тензора в новой системе координат связаны с компонентами g^{ki} соотношениями

$$g^{n'j'} = A^{n'k} A^{j'i} g^{ki},$$

где

$$A^{j'i} = \frac{\partial u^{j'}}{\partial u^i}.$$

Аналогичными соотношениями связаны и векторы взаимных базисов криволинейных систем

$$\mathbf{r}^{i'} = A^{i'k} \mathbf{r}^k. \quad (10.9)$$

Так как G^{-1} является обратной матрицей G , то их произведение равно единичной матрице, откуда следует, что смешанные компоненты метрического тензора равны функции δ_i^k

$$g_{ij} g^{jk} = g_i^k = \delta_i^k.$$

Векторы взаимного базиса и касательного базиса связаны соотношениями

$$\begin{aligned} \mathbf{r}^1 &= \frac{\mathbf{r}_2 \times \mathbf{r}_3}{\sqrt{g}}, & \mathbf{r}^2 &= \frac{\mathbf{r}_3 \times \mathbf{r}_1}{\sqrt{g}}, & \mathbf{r}^3 &= \frac{\mathbf{r}_1 \times \mathbf{r}_2}{\sqrt{g}}, \\ \mathbf{r}_1 &= \mathbf{r}^2 \times \mathbf{r}^3 \sqrt{g}, & \mathbf{r}_2 &= \mathbf{r}^3 \times \mathbf{r}^1 \sqrt{g}, & \mathbf{r}_3 &= \mathbf{r}^1 \times \mathbf{r}^2 \sqrt{g}, \end{aligned} \quad (10.10)$$

где g равен определителю матрицы G . Последние соотношения можно записать в виде

$$\mathbf{r}^k = e^{ijk} \frac{\mathbf{r}_i \times \mathbf{r}_j}{\sqrt{g}}, \quad \mathbf{r}_k = e_{ijk} \mathbf{r}^i \times \mathbf{r}^j \sqrt{g},$$

или

$$\mathbf{r}_i \times \mathbf{r}_j = e_{ijk} \mathbf{r}^k \sqrt{g}, \quad \mathbf{r}^i \times \mathbf{r}^j = e^{ijk} \frac{\mathbf{r}_k}{\sqrt{g}},$$

где используются символы Леви-Чивита e^{ijk} , e_{ijk} . Величины e^{ijk} , e_{ijk} равны 0, если в среди индексов встречаются одинаковые; равны 1 для последовательности индексов 1, 2, 3 и получающейся из нее круговой перестановкой последовательностей 2, 3, 1 и 3, 1, 2; равны -1 при нарушении этого порядка (для последовательности индексов 3, 2, 1 и получающейся из нее круговой перестановкой последовательностей 2, 1, 3 и 1, 3, 2).

В криволинейной системе координат произвольный вектор \mathbf{a} может быть разложен по касательному или взаимному базису и представлен в виде

$$\mathbf{a} = a^j \mathbf{r}_j = a_i \mathbf{r}^i.$$

Результатом операции скалярного произведения векторов \mathbf{a} и \mathbf{b} является число, которое в зависимости от представления векторов равно

$$\mathbf{a} \cdot \mathbf{b} = a^i b^j g_{ij} = a_i b_j g^{ij} = a^i b_i = a_i b^i. \quad (10.11)$$

Их векторное произведение может быть представлено в виде

$$\mathbf{a} \times \mathbf{b} = a^i b^j e_{ijk} \sqrt{g} \mathbf{r}^k = a_i b_j e^{ijk} \frac{\mathbf{r}^k}{\sqrt{g}}, \quad (10.12)$$

где использовались равенства

$$e^{ijk} \frac{1}{\sqrt{g}} = (\mathbf{r}^i \times \mathbf{r}^j) \cdot \mathbf{r}^k, \quad e_{ijk} \sqrt{g} = (\mathbf{r}_i \times \mathbf{r}_j) \cdot \mathbf{r}_k.$$

Величины

$$e^{ijk} = e^{ijk} \frac{1}{\sqrt{g}} = (\mathbf{r}^i \times \mathbf{r}^j) \cdot \mathbf{r}^k, \quad \varepsilon_{ijk} = e_{ijk} \sqrt{g} = (\mathbf{r}_i \times \mathbf{r}_j) \cdot \mathbf{r}_k$$

являются соответственно контравариантными и ковариантными компонентами тензора Леви-Чивита.

Символы Кристоффеля. Найдем изменение векторов касательного базиса при переходе из точки $R(u^1, u^2, u^3)$ в бесконечно близкую к ней точку $R(u^1 + \delta u^1, u^2 + \delta u^2, u^3 + \delta u^3)$. С точностью до линейных слагаемых относительно δu^i , $i = 1, 2, 3$ изменение базисных векторов можно представить в виде

$$\delta \mathbf{r}_i = \frac{\partial \mathbf{r}_i}{\partial u^j} \delta u^j = \mathbf{r}_{ij} \delta u^j = \Gamma_{ij}{}^k \mathbf{r}_k \delta u^j, \quad (10.13)$$

где $\Gamma_{ij}{}^k$ — неизвестные пока коэффициенты разложения производных \mathbf{r}_{ij} касательного базиса по самому базису $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$

Умножив равенство

$$\mathbf{r}_{ij} \equiv \frac{\partial \mathbf{r}_i}{\partial u^j} = \Gamma_{ij}{}^k \mathbf{r}_k \quad (10.14)$$

скалярно на $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ получим систему уравнений для $\Gamma_{ij}{}^k$

$$\begin{aligned}\Gamma_{ij}^1 g_{11} + \Gamma_{ij}^2 g_{12} + \Gamma_{ij}^3 g_{13} &= \Gamma_{ij,1} \equiv \mathbf{r}_{ij} \cdot \mathbf{r}_1, \\ \Gamma_{ij}^1 g_{21} + \Gamma_{ij}^2 g_{22} + \Gamma_{ij}^3 g_{23} &= \Gamma_{ij,2} \equiv \mathbf{r}_{ij} \cdot \mathbf{r}_2, \\ \Gamma_{ij}^1 g_{31} + \Gamma_{ij}^2 g_{32} + \Gamma_{ij}^3 g_{33} &= \Gamma_{ij,3} \equiv \mathbf{r}_{ij} \cdot \mathbf{r}_3.\end{aligned}\quad (10.15)$$

Коэффициенты $\Gamma_{ij,k} \equiv \mathbf{r}_{ij} \cdot \mathbf{r}_k$ могут быть выражены через производные ковариантных ком понент метрического тензора. Для этого выпишем известные равенства

$$\mathbf{r}_j \cdot \mathbf{r}_k = g_{jk}, \quad \mathbf{r}_k \cdot \mathbf{r}_i = g_{ki}, \quad \mathbf{r}_i \cdot \mathbf{r}_j = g_{ij},$$

в которых циклически переставляются индексы i, j, k ; индексы могут принимать значения от единицы до размерности пространства. Продифференцируем первое равенство по u^i , второе — по u^j , третье — по u^k и получим

$$\mathbf{r}_{ji} \cdot \mathbf{r}_k + \mathbf{r}_j \cdot \mathbf{r}_{ki} = \frac{\partial g_{jk}}{\partial u^i}, \quad \mathbf{r}_{kj} \cdot \mathbf{r}_i + \mathbf{r}_k \cdot \mathbf{r}_{ij} = \frac{\partial g_{ki}}{\partial u^j}, \quad \mathbf{r}_{ik} \cdot \mathbf{r}_j + \mathbf{r}_i \cdot \mathbf{r}_{jk} = \frac{\partial g_{ij}}{\partial u^k}.$$

Если сложим первые два равенства и вычтем из них третье, то получим формулу для определения коэффициентов $\Gamma_{ij,k}$

$$\Gamma_{ij,k} = \frac{1}{2} \left(\frac{\partial g_{kj}}{\partial u^i} + \frac{\partial g_{ki}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k} \right).\quad (10.16)$$

Теперь решим систему уравнений (10.15) относительно Γ_{ij}^k и получим равенства

$$\Gamma_{ij}^n = g^{nk} \Gamma_{ij,k}.\quad (10.17)$$

Саму систему (10.15) перепишем в виде

$$\Gamma_{ij,n} = g_{nk} \Gamma_{ij}^k.\quad (10.18)$$

Подставим в (10.17) равенства (10.16) и получим окончательное выражение для коэффициентов в разложении (10.13)

$$\Gamma_{ij}{}^n = \frac{1}{2}g^{nk} \left(\frac{\partial g_{kj}}{\partial u^i} + \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k} \right). \quad (10.19)$$

Коэффициенты $\Gamma_{ij,k}$ называются символами Кристоффеля 1-го рода, а коэффициенты $\Gamma_{ij}{}^k$ называются символами Кристоффеля 2-го рода. Они выражаются через компоненты метрического тензора и их частные производные по криволинейным координатам. Символы Кристоффеля не являются компонентами тензора. Как можно заметить, соотношения (10.15)-(10.19) идентичны соотношениям (9.15)-(9.19) для оболочек с той лишь разницей, что размерность пространства параметров оболочки на единицу меньше.

Мы использовали те же обозначения для компонент метрического тензора, символов Кристоффеля, касательного и взаимного базиса. Все полученные для них соотношения справедливы для пространства любой размерности.

Векторная функция.

В декартовой прямоугольной системе координат компоненты радиус-вектора точки равны координатам этой точки пространства. В криволинейной системе понятие; радиус-вектора как преобразования, переводящего начальную точку координатной системы в заданную точку пространства, теряет смысл.

Линию в криволинейной системе координат можно описать как совокупность функций $u^1 = u^1(t)$, $u^2 = u^2(t)$, $u^3 = u^3(t)$ параметра t . Пусть приращению $\delta(t)$ параметра кривой соответствуют приращения координатных функций δu^1 , δu^2 , δu^3 . Тогда вектор $\mathbf{r}_1 \delta u^1 + \mathbf{r}_2 \delta u^2 + \mathbf{r}_3 \delta u^3$ будет направлен по касательной к кривой. Если поделить его на δt и устремить δt к нулю, то в пределе мы получим вектор производной кривой

$$\mathbf{a}(t) = \frac{du^i(t)}{dt} \mathbf{r}_i = a^i(t) \mathbf{r}_i. \quad (10.20)$$

Производная кривой представляет собой векторную функцию параметра t .

Поверхность в криволинейной системе координат может быть описана как совокупность функций $u^1 = u^1(v, w)$, $u^2 = u^2(v, w)$, $u^3 = u^3(v, w)$ двух параметров v и w . Касательные к v -линиям и w -линиям поверхности векторы обозначим через \mathbf{v} и \mathbf{w} . Эти векторы определяются равенствами

$$\mathbf{v}(v, w) = \frac{\partial u^i(v, w)}{\partial v} \mathbf{r}_i = v^i(v, w) \mathbf{r}_i,$$

$$\mathbf{w}(v, w) = \frac{\partial u^i(v, w)}{\partial w} \mathbf{r}_i = w^i(v, w) \mathbf{r}_i,$$

и представляют собой векторные функции параметров поверхности. Свойства поверхности могут быть определены с помощью кривых на ней, поэтому далее мы будем исследовать некоторую векторную функцию $\mathbf{a}(t)$ параметра t .

При вычислении производной векторной функции нужно учитывать то, что базисные векторы в каждой точке пространства в общем случае различные. В декартовой системе координат базисные векторы во всех точках пространства одинаковы по величине и направлению, поэтому при дифференцировании они выступают в роли констант. Пусть изменению параметра δt кривой соответствует приращение координат $\delta u^1, \delta u^2, \delta u^3$. Найдем приращение векторной функции $\mathbf{a}(t)$ связанной с этой кривой. Потребуем, чтобы максимальная из величин δu^i стремилась к нулю при $\delta t \rightarrow 0$. Приращение векторной функции в местном базисе точки $u^1(t), u^2(t), u^3(t)$ равно

$$\delta \mathbf{a} = \mathbf{a}(t + \delta t) - \mathbf{a}(t) = a^i(t + \delta t)(\mathbf{r}_i + \delta \mathbf{r}_i) - a^i(t)\mathbf{r}_i = a^i(t + \delta t)(\mathbf{r}_i + \Gamma_{ij}^k \mathbf{r}_k \delta u^j) - a^i(t)\mathbf{r}_i.$$

Поделив это равенство на δt и устремив δt к нулю, получим формулу для производной векторной функции

$$\frac{d\mathbf{a}}{dt} = \frac{da^i}{dt} \mathbf{r}_i + a^i \Gamma_{ij}^k \mathbf{r}_k \frac{du^j}{dt} = \frac{da^i}{dt} \mathbf{r}_i + a^n \Gamma_{nj}^i \mathbf{r}_i \frac{du^j}{dt} = \left(\frac{da^i}{dt} + a^n \Gamma_{nj}^i \frac{du^j}{dt} \right) \mathbf{r}_i. \quad (10.21)$$

Выражение

$$\frac{Da^i}{dt} = \frac{da^i}{dt} + a^n \Gamma_{nj}^i \frac{du^j}{dt} \quad (10.22)$$

называется абсолютной или ковариантной производной контравариантных компонент векторной функции в направлении кривой $u^1(t), u^2(t), u^3(t)$.

Касательный базис и взаимный базис в некоторой точке пространства равноправны, поэтому вектор $\mathbf{a}(t)$ кривой может быть разложен также по взаимному базису в точке кривой

$$\mathbf{a}(t) = a_i \mathbf{r}^j, \quad (10.23)$$

где a_i — ковариантные компоненты вектора \mathbf{a} . Для вывода производной $\mathbf{a}(t)$ в данном представлении нужно знать, как меняются векторы взаимного базиса при переходе из точки $R(u^1, u^2, u^3)$ в бесконечно близкую к ней точку $R_1(u^1 + \delta u^1, u^2 + \delta u^2, u^3 + \delta u^3)$. Дифференцируя равенства (1.10.7), получим

$$\mathbf{r}_i \cdot \frac{\partial \mathbf{r}^k}{\partial u^j} = -\frac{\partial \mathbf{r}_i}{\partial u^j} \cdot \mathbf{r}^k = -\Gamma_{ij}^n \mathbf{r}_n \cdot \mathbf{r}^k = -\Gamma_{ij}^k, \quad (10.24)$$

откуда следует, что $-\Gamma_{ij}^k$ являются коэффициентами разложения производных $\partial \mathbf{r}^k / \partial u^j$ по взаимному базису

$$\mathbf{r}^k_j \equiv \frac{\partial \mathbf{r}^k}{\partial u^j} = -\Gamma_{ij}^k \mathbf{r}^i. \quad (10.25)$$

Действительно, если мы скалярно умножим равенство (10.25) на \mathbf{r}_i то получим равенства (10.24). Таким образом, с точностью до линейных слагаемых относительно δu^j изменение векторов взаимного базиса равно

$$\delta \mathbf{r}^k = \frac{\partial \mathbf{r}^k}{\partial u^j} \delta u^j = -\Gamma_{ij}^k \mathbf{r}^i \delta u^j. \quad (10.26)$$

Используя покомпонентное представление (10.23), получим формулу для производной векторной функции кривой

$$\frac{d\mathbf{a}}{dt} = \frac{da_i}{dt} \mathbf{r}^i - a_i \Gamma_{kj}^i \mathbf{r}^k \frac{du^j}{dt} = \frac{da_i}{dt} \mathbf{r}^i - a_n \Gamma_{ij}^n \mathbf{r}^i \frac{du^j}{dt} = \left(\frac{da_i}{dt} - a_n \Gamma_{ij}^n \frac{du^j}{dt} \right) \mathbf{r}^i. \quad (10.27)$$

Выражение

$$\frac{Da_i}{dt} = \frac{da_i}{dt} - a_n \Gamma_{ij}^n \frac{du^j}{dt} \quad (10.28)$$

называют абсолютной или ковариантной производной ковариантных компонент векторной функции в направлении кривой $u^1(t)$, $u^2(t)$, $u^3(t)$.

Как можно видеть, производная векторной функции представляет собой также векторную функцию. Дифференцируя представления

$$\frac{d\mathbf{a}}{dt} = \frac{Da_i}{dt} \mathbf{r}^i, \quad \frac{d\mathbf{a}}{dt} = \frac{Da^i}{dt} \mathbf{r}_i$$

и, используя формулы (10.21) и (10.27), можно получить для векторной функции $\mathbf{a}(t)$ производные более высокого порядка.

Параметр t кривой определяет некоторую точку пространства, поэтому компоненты векторной функции зависят от координат точек, через которые проходит кривая. Представим, что в каждой точке пространства задана векторная функция $\mathbf{a}(u^1, u^2, u^3)$. Такая векторная функция определяет в пространстве векторное поле.

Ковариантные производные.

Найдем изменение векторного поля при переходе из точки $R(u^1, u^2, u^3)$ в бесконечно близкую к ней точку вдоль одной из координатных

линий. Пусть этому переходу соответствует изменение координаты u^i на бесконечно малую величину δu^i .

Приращение векторного поля в представлении через контравариантные компоненты $\mathbf{a}(u^1, u^2, u^3) = a^i \mathbf{r}_i$ с точностью до линейных по δu^i членов равно

$$\delta \mathbf{a} = \frac{\partial a^i}{\partial u^j} \delta u^j \mathbf{r}_i + a^i \frac{\partial \mathbf{r}_i}{\partial u^j} \delta u^j.$$

Поделив обе части этого равенства на δu^j и устремив δu^i к нулю, получим формулу для производной векторного поля в представлении через контравариантные компоненты

$$\frac{\partial \mathbf{a}}{\partial u^j} = \frac{\partial a^i}{\partial u^j} \mathbf{r}_i + a^i \Gamma_{ij}^k \mathbf{r}_k = \left(\frac{\partial a^i}{\partial u^j} + a^n \Gamma_{nj}^i \right) \mathbf{r}_i. \quad (10.29)$$

Выражение

$$\nabla_j a^i = \frac{\partial a^i}{\partial u^j} + a^n \Gamma_{nj}^i \quad (10.30)$$

называется ковариантной производной контравариантных компонент векторного поля.

Приращение векторного поля в представлении через ковариантные компоненты $\mathbf{a}(u^1, u^2, u^3) = a_i \mathbf{r}^i$ с точностью до линейных по δu^i членов равно

$$\delta \mathbf{a} = \frac{\partial a_i}{\partial u^j} \delta u^j \mathbf{r}^i + a_i \frac{\partial \mathbf{r}^i}{\partial u^j} \delta u^j.$$

Отсюда получим формулу для производной векторного поля в представлении через ковариантные компоненты

$$\frac{\partial \mathbf{a}}{\partial u^j} = \frac{\partial a_i}{\partial u^j} \mathbf{r}^i - a_i \Gamma_{kj}^i \mathbf{r}^k = \left(\frac{\partial a_i}{\partial u^j} - a^n \Gamma_{ij}^n \right) \mathbf{r}^i. \quad (10.31)$$

Выражение

$$\nabla_j a_i = \frac{\partial a_i}{\partial u^j} - a_n \Gamma_{ij}^n \quad (10.32)$$

называют ковариантной производной ковариантных компонент векторного поля.

Итак, нами получены формулы (10.22) и (10.28) для вычисления производных вектора кривой $\mathbf{a}(t)$ по параметру t в произвольной криволинейной системе координат. Формулы (10.22), (10.28), (10.30), (10.32) представляют собой аппарат абсолютного дифференцирования, который дает формулы вычисления производных векторных функций в любой системе координат. Ценой универсальности этого аппарата является необходимость вычисления производных базисных векторов по координатам выбранной системы координат. Для вычисления производной вектора нужно знать его разложение по касательному или взаимному базису точки, в которой он задан.

Покажем, что компоненты метрического тензора можно выносить из-под знака ковариантной производной ∇_j . Дифференцируя равенства $g_{ki} = \mathbf{r}_k \cdot \mathbf{r}_i$ и $g^{ki} = \mathbf{r}^k \cdot \mathbf{r}^i$ и используя (10.14) и (10.25), получим выражения для производных компонент метрического тензора

$$\begin{aligned} \frac{\partial g_{ki}}{\partial u^j} &= g_{kn} \Gamma_{ij}^n + g_{ni} \Gamma_{kj}^n, \\ \frac{\partial g^{ki}}{\partial u^j} &= -g^{kn} \Gamma_{nj}^i - g^{ni} \Gamma_{nj}^k. \end{aligned} \quad (10.34)$$

Подставив в равенство (10.30) соотношения $a^i = g^{ik} a_k$ и (10.34), получим

$$\begin{aligned}\nabla_j a^i &= \nabla_j (g^{ik} a_k) = g^{ik} \frac{\partial a_k}{\partial u^j} + a_k \frac{\partial g^{ki}}{\partial u^j} + a_k g^{nk} \Gamma_{nj}{}^i = \\ &= g^{ik} \frac{\partial a_k}{\partial u^j} - a_k g^{ni} \Gamma_{nj}{}^k = g^{ik} \left(\frac{\partial a_k}{\partial u^j} - a_n \Gamma_{kj}{}^n \right) = g^{ik} \nabla_j a_k.\end{aligned}\quad (10.35)$$

Во время преобразований мы меняли обозначение немых индексов. Это вполне допустимо, так как немой индекс не входит в результат и, следовательно, может быть обозначен любой буквой. Аналогично подставив в равенство (10.32) соотношения $a_i = g_{ik} a^k$ и (10.33), получим

$$\begin{aligned}\nabla_j a_i &= \nabla_j (g_{ik} a^i) = g_{ik} \frac{\partial a^k}{\partial u^j} + a^k \frac{\partial g_{ik}}{\partial u^j} - a^k g_{kn} \Gamma_{ij}{}^n = \\ &= g_{ik} \frac{\partial a^k}{\partial u^j} + a^k g_{ni} \Gamma_{kj}{}^n = g_{ik} \left(\frac{\partial a^k}{\partial u^j} + a^n \Gamma_{nj}{}^k \right) = g_{ik} \nabla_j a^k.\end{aligned}\quad (10.36)$$

Из равенств (10.35) и (10.36) следует вывод о том, что компоненты метрического тензора можно выносить из-под знака ковариантной производной.

1.11 Тензоры в криволинейных координатах

В криволинейных координатах мы рассмотрим понятие тензора — объекта, операций над которым не зависят от координатной системы. Одним из тензоров является метрический тензор. Компоненты первой квадратичной формы поверхности являются компонентами метрического тензора параметрического пространства поверхности. Примерами тензоров могут служить тензор деформации и тензор напряжений сплошной среды, тензор инерции твердого тела, тензор кривизны пространства. Тензоры, заданные для каждой точки некоторой области пространства, образуют тензорное поле. Тензоры описываются своими компонентами.

Криволинейная поверхность представляет собой двухмерный аналог искривленного пространства. Продолжим аналогию между трехмерным пространством и поверхностью — представителем двухмерных пространств. Если нам известны компоненты метрического тензора как функции используемой системы координат, то мы можем по ним выполнить исследование некоторых свойств самого пространства, в частности его кривизны. Выше мы рассмотрели тензор кривизны поверхности. Компоненты аналогичного тензора можно вычислить и для трехмерного пространства.

Тензор кривизны.

В искривленных пространствах последовательность дифференцирования по координатам в общем случае играет роль. При изменении последовательности дифференцирования появляются дополнительные слагаемые. Пусть дана некоторая векторная функция $\mathbf{a} = a_i(u^1, u^2, u^3) = \mathbf{r}^i$. Рассмотрим некоторую поверхность $u^i = u^i(t, w)$, $i=1, 2, 3$ в пространстве. Всегда можно выбрать поверхность так, чтобы она проходила через любую заданную точку или линию. Вычислим абсолютный дифференциал $D_t \mathbf{a}$ векторной функции при бесконечно малом смещении из точки u^1, u^2, u^3 в точку

$u^1 + \frac{\partial u^1}{\partial t} dt, u^2 + \frac{\partial u^2}{\partial t} dt, u^3 + \frac{\partial u^3}{\partial t} dt$ вдоль поверхности в направлении первого ее параметра

$$D_t \mathbf{a} = \left(\left(\frac{\partial a_i}{\partial w^j} - a_n \Gamma_{ij}^n \right) \mathbf{r}^i \frac{\partial u^j}{\partial t} \right) dt = \left(\frac{\partial a_i}{\partial t} - a_n \Gamma_{ij}^n \frac{\partial u^j}{\partial t} \right) \mathbf{r}^i dt \equiv (D_t a)_i \mathbf{r}^i dt.$$

Аналогично вычислим абсолютный дифференциал $D_w \mathbf{a}$ векторной функции при бесконечно малом смещении из той же точки u^1, u^2, u^3 в

точку $u^1 + \frac{\partial u^1}{\partial w} dw, u^2 + \frac{\partial u^2}{\partial w} dw, u^3 + \frac{\partial u^3}{\partial w} dw$ вдоль поверхности в направлении второго ее параметра

$$D_w \mathbf{a} = \left(\left(\frac{\partial a_i}{\partial w^j} - a_n \Gamma_{ij}^n \right) \mathbf{r}^i \frac{\partial u^j}{\partial w} \right) dw = \left(\frac{\partial a_i}{\partial w} - a_n \Gamma_{ij}^n \frac{\partial u^j}{\partial w} \right) \mathbf{r}^i dw \equiv (D_w a)_i \mathbf{r}^i dw.$$

Полученные дифференциалы можно также рассматривать как векторные функции тех же параметров в той же точке. Вычислим теперь дифференциал векторной функции $D_t \mathbf{a}$ вдоль поверхности в направлении второго ее параметра и дифференциал векторной функции $D_w \mathbf{a}$ вдоль поверхности в направлении первого ее параметра

$$\begin{aligned}
 D_w D_t \mathbf{a} &= \left(\frac{\partial(D_t \mathbf{a})_i}{\partial w} - (D_t \mathbf{a})_n \Gamma_{ij}^n \frac{\partial w^j}{\partial w} \right) \mathbf{r}^i dw = \\
 &= \left(\frac{\partial^2 a_i}{\partial t \partial w} - \frac{\partial a_n}{\partial w} \Gamma_{ij}^n \frac{\partial w^j}{\partial t} - a_n \frac{\partial \Gamma_{ij}^n}{\partial w} \frac{\partial w^j}{\partial t} - a_n \Gamma_{ij}^n \frac{\partial^2 w^j}{\partial t \partial w} - \right. \\
 &\quad \left. - \frac{\partial a_n}{\partial t} \Gamma_{ir}^n \frac{\partial u^r}{\partial w} + a_m \Gamma_{nj}^m \frac{\partial w^j}{\partial t} \Gamma_{ir}^n \frac{\partial u^r}{\partial w} \right) \mathbf{r}^i dt dw, \\
 D_t D_w \mathbf{a} &= \left(\frac{\partial(D_w \mathbf{a})_i}{\partial t} - (D_w \mathbf{a})_n \Gamma_{ij}^n \frac{\partial w^j}{\partial t} \right) \mathbf{r}^i dt = \\
 &= \left(\frac{\partial^2 a_i}{\partial w \partial t} - \frac{\partial a_n}{\partial t} \Gamma_{ij}^n \frac{\partial w^j}{\partial w} - a_n \frac{\partial \Gamma_{ij}^n}{\partial t} \frac{\partial w^j}{\partial w} - a_n \Gamma_{ij}^n \frac{\partial^2 w^j}{\partial w \partial t} - \right. \\
 &\quad \left. - \frac{\partial a_n}{\partial w} \Gamma_{ir}^n \frac{\partial u^r}{\partial t} + a_m \Gamma_{nj}^m \frac{\partial w^j}{\partial w} \Gamma_{ir}^n \frac{\partial u^r}{\partial t} \right) \mathbf{r}^i dt dw.
 \end{aligned}$$

Полученные выражения отличаются тем, что в них переставлены местами производные по t и по w . Первые и четвертые слагаемые обоих выражений равны. Второе слагаемое первого выражения равно пятому слагаемому второго выражения (если немой индекс r переобозначить через j). Вычтем из первого выражения второе и получим (используя возможность переобозначения немых индексов)

$$\begin{aligned}
 D_w D_t \mathbf{a} - D_t D_w \mathbf{a} &= \left(-a_n \frac{\partial \Gamma_{ij}^n}{\partial w} \frac{\partial w^j}{\partial t} + a_m \Gamma_{nj}^m \frac{\partial w^j}{\partial t} \Gamma_{ir}^n \frac{\partial u^r}{\partial w} + \right. \\
 &\quad \left. + a_n \frac{\partial \Gamma_{ij}^n}{\partial t} \frac{\partial w^j}{\partial w} - a_m \Gamma_{nj}^m \frac{\partial w^j}{\partial w} \Gamma_{ir}^n \frac{\partial u^r}{\partial t} \right) \mathbf{r}^i dw dt = \\
 &= \left(\left(-\frac{\partial \Gamma_{ij}^n}{\partial u^r} + \Gamma_{qj}^n \Gamma_{ir}^q \right) \frac{\partial u^j}{\partial t} \frac{\partial u^r}{\partial w} + \left(\frac{\partial \Gamma_{ir}^n}{\partial w^j} - \Gamma_{qr}^n \Gamma_{ij}^q \right) \frac{\partial w^j}{\partial t} \frac{\partial u^r}{\partial w} \right) a_n \mathbf{r}^i dw dt = \\
 &= \left(-\frac{\partial \Gamma_{ij}^n}{\partial u^r} + \Gamma_{kj}^n \Gamma_{ir}^k + \frac{\partial \Gamma_{ir}^n}{\partial w^j} - \Gamma_{kr}^n \Gamma_{ij}^k \right) \frac{\partial w^j}{\partial t} \frac{\partial u^r}{\partial w} a_n \mathbf{r}^i dw dt = \\
 &= R_{rji}^n \frac{\partial w^j}{\partial t} \frac{\partial u^r}{\partial w} a_n dw dt \mathbf{r}^i, \quad (1.11)
 \end{aligned}$$

где введено обозначение

$$R_{rji}^n = \frac{\partial \Gamma_{ir}^n}{\partial w^j} - \frac{\partial \Gamma_{ij}^n}{\partial u^r} + \Gamma_{kj}^n \Gamma_{ir}^k - \Gamma_{kr}^n \Gamma_{ij}^k. \quad (1.12)$$

Коэффициенты R_{rji}^n являются компонентами тензора кривизны пространства. Тензор кривизны называют тензором Римана. Он зависит лишь от точки пространства, в которой вычислен.

Если в качестве поверхности $u^i = u^i(t, w)$ взять одну из координатных поверхностей $u^1 = u^1(u^2, u^3)$, $u^2 = u^2(u^3, u^1)$, $u^3 = u^3(u^1, u^2)$,

проходящую через рассматриваемую точку, то выражение (11.1) примет вид

$$D_r D_j \mathbf{a} - D_j D_r \mathbf{a} = R_{rji}{}^n a_n \mathbf{r}^i du^r du^j, \quad (11.3)$$

так как $dw=du^r$, $dt=du^j$. Левая часть (11.3) называется вторым альтернативным дифференциалом векторной функции. Если векторную функцию взять в виде разложения по касательному базису $\mathbf{a}=a^i(u^1, u^2, u^3) = \mathbf{r}_i$, то проведя аналогичный вывод, получим следующее выражение для второго альтернативного дифференциала векторной функции

$$D_r D_j \mathbf{a} - D_j D_r \mathbf{a} = -R_{rji}{}^n a^i \mathbf{r}_n du^r du^j. \quad (11.4)$$

Из формулы (11.3) видно, что тензор кривизны кососимметричен по первым двум индексам

$$R_{rji}{}^n = -R_{jri}{}^n. \quad (11.5)$$

Компоненты тензора кривизны (11.2) являются трижды ковариантными и один раз контравариантными. С помощью компонент метрического тензора можно получить полностью ковариантные компоненты тензора кривизны

$$R_{rjik} = g_{nk} R_{jri}{}^n.$$

Если над тензором кривизны выполнить операцию свертки по индексам n, r , то мы получим тензор Риччи, ковариантные компоненты которого равны

$$R_{ji} = R_{jni}{}^n. \quad (11.6)$$

В евклидовом пространстве все компоненты тензора кривизны равны нулю:

$$R_{rj i}^n = \frac{\partial \Gamma_{ir}^n}{\partial u^j} - \frac{\partial \Gamma_{ij}^n}{\partial u^r} + \Gamma_{kj}^n \Gamma_{ir}^k - \Gamma_{kr}^n \Gamma_{ij}^k = 0.$$

Двухмерное пространство на некоторой поверхности в общем случае не является евклидовым и имеет ненулевой тензор кривизны. Плоская поверхность представляет собой пример евклидова двухмерного пространства. Хотя координатные линии на плоскости могут быть кривыми, все компоненты тензора кривизны плоского пространства равны нулю. Пространства, для которых компоненты тензора кривизны везде равны нулю, по аналогии с поверхностями называются плоскими. Изучением пространств с ненулевым тензором кривизны занимается риманова геометрия.

Ковариантные производные компонент тензора.

Метрический тензор является тензором второго ранга, а векторы являются тензорами первого ранга. По аналогии с записью векторов в виде $\mathbf{a} = a_i \mathbf{e}^i = a^i \mathbf{e}_i$ можно записать тензор второго ранга, например, метрический в виде суммы диад базисных векторов

$$\mathbf{G} = g_{ik} \mathbf{r}^i \mathbf{r}^k = g^{ik} \mathbf{r}_i \mathbf{r}_k. \quad (11.7)$$

Дифференцируя тензор в данном представлении по правилу дифференцирования суммы и произведения функций, получим выражение для ковариантных производных компонент тензора второго ранга

$$\begin{aligned} \frac{\partial (g_{ik} \mathbf{r}^i \mathbf{r}^k)}{\partial u^j} &= \frac{\partial g_{ik}}{\partial u^j} \mathbf{r}^i \mathbf{r}^k + g_{ik} \frac{\partial \mathbf{r}^i}{\partial u^j} \mathbf{r}^k + g_{ik} \mathbf{r}^i \frac{\partial \mathbf{r}^k}{\partial u^j} = \\ &= \frac{\partial g_{ik}}{\partial u^j} \mathbf{r}^i \mathbf{r}^k - g_{ik} \Gamma_{nj}^i \mathbf{r}^n \mathbf{r}^k - g_{ik} \mathbf{r}^i \Gamma_{nj}^k \mathbf{r}^n = \\ &= \left(\frac{\partial g_{ik}}{\partial u^j} - g_{nk} \Gamma_{ij}^n - g_{in} \Gamma_{kj}^n \right) \mathbf{r}^i \mathbf{r}^k = \nabla_j g_{ik} \mathbf{r}^i \mathbf{r}^k, \end{aligned}$$

$$\begin{aligned} \frac{\partial (g^{ik} \mathbf{r}_i \mathbf{r}_k)}{\partial u^j} &= \frac{\partial g^{ik}}{\partial u^j} \mathbf{r}_i \mathbf{r}_k + g^{ik} \frac{\partial \mathbf{r}_i}{\partial u^j} \mathbf{r}_k + g^{ik} \mathbf{r}_i \frac{\partial \mathbf{r}_k}{\partial u^j} = \\ &= \frac{\partial g^{ik}}{\partial u^j} \mathbf{r}_i \mathbf{r}_k + g^{ik} \Gamma_{ij}^n \mathbf{r}_n \mathbf{r}_k + g^{ik} \mathbf{r}_i \Gamma_{kj}^n \mathbf{r}_n = \\ &= \left(\frac{\partial g^{ik}}{\partial u^j} + g^{nk} \Gamma_{nj}^i + g^{in} \Gamma_{kj}^k \right) \mathbf{r}_i \mathbf{r}_k = \nabla_j g^{ik} \mathbf{r}_i \mathbf{r}_k, \end{aligned} \quad (11.8)-(11.9)$$

где

$$\begin{aligned}\nabla_j g_{ki} &= \frac{\partial g_{ki}}{\partial u^j} - g_{ni} \Gamma_{kj}^n - g_{kn} \Gamma_{ij}^n, \\ \nabla_j g^{ki} &= \frac{\partial g^{ki}}{\partial u^j} + g^{ni} \Gamma_{nj}^k + g^{kn} \Gamma_{nj}^i\end{aligned}\quad (11.10)$$

— ковариантные производные компонент метрического тензора. Как и следовало ожидать, из соотношений (10.33) и (10.34) ковариантные производные компонент метрического тензора равны нулю. Равенства (11.10) и (11.11) представляют собой определения ковариантных производных ковариантных и контравариантных компонент тензоров второго ранга (вместо компонент метрического тензора могут быть подставлены компоненты любого другого тензора второго ранга, так как при выводе никакие свойства компонент не использовались).

Аналогично записи векторов в виде $\mathbf{a} = a_i \mathbf{r}^i = a^i \mathbf{r}_i$ можно записать тензор произвольного ранга в виде суммы произведений его компонент на соответствующие базисные векторы, например, $\mathbf{T} = t_m^{ki} \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m$. Каждому верхнему индексу компоненты тензора соответствует вектор касательного базиса \mathbf{r}^i , а каждому нижнему индексу — вектор взаимного базиса \mathbf{r}^i . Используя аналогию с векторами, можно получить формулы для ковариантных производных компонент тензора произвольного ранга. Например, дифференцируя правую часть выражения \mathbf{r}^i по правилам дифференцирования суммы и произведения функций

$$\begin{aligned}\frac{\partial (t_m^{ki} \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m)}{\partial u^j} &= \frac{\partial t_m^{ki}}{\partial u^j} \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m + t_m^{ki} \frac{\partial \mathbf{r}_k}{\partial u^j} \mathbf{r}_i \mathbf{r}^m + t_m^{ki} \mathbf{r}_k \frac{\partial \mathbf{r}_i}{\partial u^j} \mathbf{r}^m + t_m^{ki} \mathbf{r}_k \mathbf{r}_i \frac{\partial \mathbf{r}^m}{\partial u^j} = \\ &= \frac{\partial t_m^{ki}}{\partial u^j} \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m + t_m^{ki} \Gamma_{kj}^n \mathbf{r}_n \mathbf{r}_i \mathbf{r}^m + t_m^{ki} \Gamma_{ij}^n \mathbf{r}_k \mathbf{r}_n \mathbf{r}^m - t_m^{ki} \Gamma_{nj}^m \mathbf{r}_k \mathbf{r}_i \mathbf{r}^n = \\ &= \frac{\partial t_m^{ki}}{\partial u^j} \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m + t_m^{ni} \Gamma_{nj}^k \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m + t_m^{kn} \Gamma_{nj}^i \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m - t_n^{ki} \Gamma_{mj}^n \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m = \\ &= \left(\frac{\partial t_m^{ki}}{\partial u^j} + t_m^{ni} \Gamma_{nj}^k + t_m^{kn} \Gamma_{nj}^i - t_n^{ki} \Gamma_{mj}^n \right) \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m = \nabla_j t_m^{ki} \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m\end{aligned}$$

и взяв компоненты при одинаковых триадах базисных векторов результата, получим выражения для ковариантных производных смешанных компонент тензора третьего ранга

$$\nabla_j t_m^{ki} = \frac{\partial t_m^{ki}}{\partial u^j} + t_m^{ni} \Gamma_{nj}^k + t_m^{kn} \Gamma_{nj}^i - t_n^{ki} \Gamma_{mj}^n. \quad (11.12)$$

Из полученного результата можно сформулировать общее правило вычисления ковариантной производной компонент тензора произвольного ранга. Кроме частной производной данной компоненты

тензора $\frac{\partial t_{***}^{***}}{\partial u^j}$ (индексы заменены точками) в ковариантной производной добавляются еще несколько сумм, каждая сумма соответствует одному из индексов. Так каждому нижнему индексу соответствует сумма вида $- t_{***}^{***} \Gamma_{*j}^n$, а каждому верхнему индексу соответствует сумма вида $+ t_{***}^{***} \Gamma_{nj}^*$, где значение соответствующего индекса (обозначенного точкой) переходит от компоненты к символам Кристоффеля, а по индексу n роизводится суммирование. В символической записи это выглядит так:

$$\nabla_j t_{***}^{***} = \frac{\partial t_{***}^{***}}{\partial u^j} + \dots + t_{***}^{***} \Gamma_{nj}^* + \dots - \dots - t_{***}^{***} \Gamma_{*j}^n - \dots$$

Оператор Гамильтона.

Ковариантные производные компонент тензора в свою очередь являются компонентами некоторого нового тензора $\mathbf{T} = \nabla_j t_m^{ki} \mathbf{r}^j \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m$. Символ ∇ обозначает набла-оператор или оператор Гамильтона, формальная запись которого представляется в виде

$$\nabla = \frac{\partial}{\partial u^j} \mathbf{r}^j = \frac{\partial}{\partial u^1} \mathbf{r}^1 + \frac{\partial}{\partial u^2} \mathbf{r}^2 + \frac{\partial}{\partial u^3} \mathbf{r}^3. \quad (11.13)$$

Оператору Гамильтона приписывают атрибуты вектора: он может действовать на тензоры тензорно (результатом является тензор, ранг которого на единицу выше ранга исходного тензора), скалярно (при этом ранг результата на единицу меньше ранга исходного тензора) и векторно (при этом ранг результата равен рангу исходного тензора):

$$\begin{aligned}\nabla \mathbf{a} &= \mathbf{r}^j \frac{\partial \mathbf{a}}{\partial u^j} = \mathbf{r}^j \mathbf{r}^i \left(\frac{\partial a_i}{\partial u^j} - a_n \Gamma_{ij}^n \right) = \nabla_j a_i \mathbf{r}^j \mathbf{r}^i = \text{grad } \mathbf{a}, \\ \nabla \cdot \mathbf{a} &= \mathbf{r}^j \cdot \frac{\partial \mathbf{a}}{\partial u^j} = \mathbf{r}^j \cdot \mathbf{r}^i \left(\frac{\partial a_i}{\partial u^j} - a_n \Gamma_{ij}^n \right) = \nabla_j a^j = \text{div } \mathbf{a}, \\ \nabla \times \mathbf{a} &= \mathbf{r}^j \times \frac{\partial \mathbf{a}}{\partial u^j} = \mathbf{r}^j \times \mathbf{r}^i \left(\frac{\partial a_i}{\partial u^j} - a_n \Gamma_{ij}^n \right) = \frac{1}{\sqrt{g}} e^{jik} \mathbf{r}_k \nabla_j a_i = \text{rot } \mathbf{a}.\end{aligned}\quad (11.14)-(11.16)$$

Например, в результате действия его на скаляр f (тензор нулевого ранга) получим градиент скаляра

$$\nabla f = \mathbf{r}^j \frac{\partial f}{\partial u^j} = \mathbf{r}^1 \frac{\partial f}{\partial u^1} + \mathbf{r}^2 \frac{\partial f}{\partial u^2} + \mathbf{r}^3 \frac{\partial f}{\partial u^3} = \text{grad } f.$$

Для тензора $\mathbf{T} = t_m^{ki} \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m$ результат действия оператора Гамильтона имеет вид

$$\begin{aligned}\nabla \mathbf{T} &= \mathbf{r}^j \frac{\partial \mathbf{T}}{\partial u^j} = \nabla_j t_m^{ki} \mathbf{r}^j \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m = \text{grad } \mathbf{T}, \\ \nabla \cdot \mathbf{T} &= \mathbf{r}^j \cdot \frac{\partial \mathbf{T}}{\partial u^j} = \nabla_j t_m^{ji} \mathbf{r}_i \mathbf{r}^m = \text{div } \mathbf{T}, \\ \nabla \times \mathbf{T} &= \mathbf{r}^j \times \frac{\partial \mathbf{T}}{\partial u^j} = \frac{e^{jqn}}{\sqrt{g}} g_{kq} \nabla_j t_m^{ki} \mathbf{r}_n \mathbf{r}_i \mathbf{r}^m = \frac{e^{jkn}}{\sqrt{g}} \nabla_j t_m^{ki} \mathbf{r}_n \mathbf{r}_i \mathbf{r}^m = \text{rot } \mathbf{T}.\end{aligned}$$

Оператор Гамильтона может воздействовать несколько раз на один и тот же объект. Так скалярное произведение операторов Гамильтона называется оператором Лапласа

$$\nabla^2 = \nabla \cdot \nabla = g^{ij} \frac{\partial^2}{\partial u^i \partial u^j} - g^{ik} \Gamma_{ik}^j \frac{\partial}{\partial u^j}.\quad (11.17)$$

Например, действие оператора Лапласа на скаляр записывается в виде

$$\nabla \cdot \nabla f = \nabla^2 f = \text{div grad } f.$$

В декартовой прямоугольной системе координат оператор Лапласа имеет вид

$$\nabla^2 = \frac{\partial^2}{(\partial x^1)^2} + \frac{\partial^2}{(\partial x^2)^2} + \frac{\partial^2}{(\partial x^3)^2}. \quad (11.18)$$

Оператор Гамильтона используется при описании тензорных (в частом случае скалярных или векторных) полей. Приведенные формулы позволяют записывать уравнения для математических объектов безотносительно к какой-либо системе координат. Если некоторый вектор $\mathbf{a}=\mathbf{a}(u^1, u^2, u^3)$ является функцией своего положения в пространстве, то дифференциал этого вектора $d\mathbf{a}$ с использованием оператора Гамильтона можно выразить следующим образом:

$$d\mathbf{a} = \frac{\partial \mathbf{a}}{\partial u^j} du^j = du^n \mathbf{r}_n \cdot \mathbf{r}^j \frac{\partial \mathbf{a}}{\partial u^j} = d\mathbf{r} \cdot \nabla \mathbf{a} = (\nabla \mathbf{a})^\top \cdot d\mathbf{r}, \quad (11.19)$$

где мы использовали обозначение $d\mathbf{r}=du^n \mathbf{r}_n$.

Формулы для векторных функций.

Итак, мы получили формулы для работы с геометрическими объектами в криволинейной системе координат. Для моделируемых нами кривых и поверхностей всегда будем считать известными определяющие их координатные функции. Умея вычислять производные векторных функций в криволинейной системе координат, мы сможем получить всю геометрическую информацию об объекте. Для этого необходимо знать зависимости (10.2), которые по формулам (10.4) позволят вычислить касательный базис. Всю остальную геометрическую информацию можно описать в терминах криволинейной системы координат. Полученные выше формулы позволяют представлять векторные функции и их производные в виде разложения их по касательному или взаимному базису с помощью соотношений

$$\begin{aligned} g_{ij} &= \mathbf{r}_i \cdot \mathbf{r}_j, \quad \mathbf{r}^k = g^{ki} \mathbf{r}_i, \quad \frac{\partial g_{ij}}{\partial u^k} = \mathbf{r}_{ik} \cdot \mathbf{r}_j + \mathbf{r}_i \cdot \mathbf{r}_{jk}, \\ \Gamma_{ij}^n &= \frac{1}{2} g^{nk} \left(\frac{\partial g_{kj}}{\partial u^i} + \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k} \right), \quad \mathbf{a} = a_k \mathbf{r}^k = a^k \mathbf{r}_k, \\ \mathbf{a}'(t) &= \frac{d\mathbf{a}}{dt} = \left(\frac{da^i}{dt} + a^n \Gamma_{nj}^i \frac{du^j}{dt} \right) \mathbf{r}_i = \left(\frac{da_i}{dt} - a_n \Gamma_{ij}^n \frac{du^j}{dt} \right) \mathbf{r}^i. \end{aligned} \quad (11.20)$$

1.12. Ортогональные криволинейные координаты

Описание геометрии объектов в криволинейных координатах является достаточно сложным. На практике в большинстве случаев используются ортогональные криволинейные координаты. В ортогональной системе координат координатные линии различного семейства взаимно ортогональны. Взаимно ортогональны и векторы $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ касательного базиса, так как они направлены по касательным к соответствующим координатным линиям.

В ортогональной криволинейной системе координат векторы $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ касательного базиса и соответствующие им векторы $\mathbf{r}^1, \mathbf{r}^2, \mathbf{r}^3$ взаимного базиса совпадают по направлению, но длина их в общем случае различна; недиагональные компоненты метрического тензора равны нулю: т. е. $g_{ij}=0, g^{ij}=0$ при $i \neq j$, а $g^{ii}=1/g^i$.

Часть символов Кристоффеля в ортогональной криволинейной системе координат равна нулю и, следовательно, многие формулы упрощаются.

Цилиндрическая система координат.

В качестве примера рассмотрим цилиндрическую систему координат. Параметрами цилиндрической системы являются полярный радиус $u^1=r$ полярный угол $u^2=\varphi$ и вертикальная ось $u^3=z$. Цилиндрические координаты r, φ, z связаны с декартовыми координатами x^1, x^2, x^3 равенствами

$$r = \sqrt{x^1 x^1 + x^2 x^2}, \quad \varphi = \arctg \frac{x^2}{x^1}, \quad z = x^3.$$

Обратные зависимости имеют вид

$$x^1 = r \cos \varphi, \quad x^2 = r \sin \varphi, \quad x^3 = z.$$

Матрица Якоби перехода от декартовой прямоугольной системы координат к цилиндрической системе координат равна:

$$\begin{bmatrix} \frac{\partial x^1}{\partial u^1} & \frac{\partial x^1}{\partial u^2} & \frac{\partial x^1}{\partial u^3} \\ \frac{\partial x^2}{\partial u^1} & \frac{\partial x^2}{\partial u^2} & \frac{\partial x^2}{\partial u^3} \\ \frac{\partial x^3}{\partial u^1} & \frac{\partial x^3}{\partial u^2} & \frac{\partial x^3}{\partial u^3} \end{bmatrix} = \begin{bmatrix} \cos \varphi & -r \sin \varphi & 0 \\ \sin \varphi & r \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Компоненты метрического тензора в новой системе координат определяются по формуле

$$g_{kn} = \frac{\partial x^i}{\partial u^k} \frac{\partial x^j}{\partial u^n} \delta_j^i,$$

где δ_j^i — символы Кронекера (10.7). Цилиндрическая система координат является ортогональной, компоненты метрического тензора и отличные от нуля символы Кристоффеля в ней равны

$$\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{G}^{-1} = \begin{bmatrix} g^{11} & g^{12} & g^{13} \\ g^{21} & g^{22} & g^{23} \\ g^{31} & g^{32} & g^{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\Gamma_{12,2} = \Gamma_{21,2} = r, \quad \Gamma_{22,1} = -r,$$

$$\Gamma_{12}^2 = \Gamma_{21}^2 = \frac{1}{r}, \quad \Gamma_{22}^1 = -r.$$

Как можно заметить, в цилиндрических координатах длины первого и третьего векторов касательного базиса равны единице, а длина второго вектора, соответствующего полярному углу, равна r .

Векторные функции в цилиндрической системе координат будем выражать с помощью векторов касательного и взаимного базисов следующим образом:

$$\mathbf{a} = a^r \mathbf{e}_r + a^\varphi \mathbf{e}_\varphi + a^z \mathbf{e}_z = a_r \mathbf{e}^r + a_\varphi \mathbf{e}^\varphi + a_z \mathbf{e}^z. \quad (12.1)$$

Частные производные векторов касательного базиса (10.14) по цилиндрическим координатам равны

$$\begin{array}{lll}
 \frac{\partial \mathbf{e}_r}{\partial r} = 0, & \frac{\partial \mathbf{e}_r}{\partial \varphi} = \frac{1}{r} \mathbf{e}_\varphi, & \frac{\partial \mathbf{e}_r}{\partial z} = 0, \\
 \frac{\partial \mathbf{e}_\varphi}{\partial r} = \frac{1}{r} \mathbf{e}_\varphi, & \frac{\partial \mathbf{e}_\varphi}{\partial \varphi} = -r \mathbf{e}_r, & \frac{\partial \mathbf{e}_\varphi}{\partial z} = 0, \\
 \frac{\partial \mathbf{e}_z}{\partial r} = 0, & \frac{\partial \mathbf{e}_z}{\partial \varphi} = 0, & \frac{\partial \mathbf{e}_z}{\partial z} = 0.
 \end{array} \quad (12.2)$$

Частные производные векторов взаимного базиса (10.25) по цилиндрическим координатам равны

$$\begin{array}{lll}
 \frac{\partial \mathbf{e}^r}{\partial r} = 0, & \frac{\partial \mathbf{e}^r}{\partial \varphi} = r \mathbf{e}^\varphi, & \frac{\partial \mathbf{e}^r}{\partial z} = 0, \\
 \frac{\partial \mathbf{e}^\varphi}{\partial r} = -\frac{1}{r} \mathbf{e}^\varphi, & \frac{\partial \mathbf{e}^\varphi}{\partial \varphi} = -\frac{1}{r} \mathbf{e}^r, & \frac{\partial \mathbf{e}^\varphi}{\partial z} = 0, \\
 \frac{\partial \mathbf{e}^z}{\partial r} = 0, & \frac{\partial \mathbf{e}^z}{\partial \varphi} = 0, & \frac{\partial \mathbf{e}^z}{\partial z} = 0.
 \end{array}$$

Мы видим, что некоторые векторы касательного и взаимного базисов цилиндрической системы координат изменяются при переходе от одной точки пространства к другой. Все это в общем случае усложняет описание геометрических объектов, но в некоторых частных случаях использование криволинейных систем является оправданным. Формула (10.21) для производной векторной функции в цилиндрической системе координат имеет вид

$$\frac{d\mathbf{a}}{dt} = \left(\frac{da^r}{dt} - r a^\varphi \frac{d\varphi}{dt} \right) \mathbf{e}_r + \left(\frac{da^\varphi}{dt} + \frac{a^\varphi}{r} \frac{dr}{dt} + \frac{a^r}{r} \frac{d\varphi}{dt} \right) \mathbf{e}_\varphi + \frac{da^z}{dt} \mathbf{e}_z. \quad (12.3)$$

Пример кривой.

Рассмотрим векторную функцию, описывающую один виток цилиндрической спирали в цилиндрической системе координат. Пусть ось спирали параллельна \mathbf{e}_z и проходит через начало координат, ее радиус равен ρ , а шаг равен h . Опишем спираль функциями координат:

$$r(t) = \rho, \quad \varphi(t) = t, \quad z(t) = h \frac{t}{2\pi}, \quad 0 \leq t \leq 2\pi.$$

Касательная спирали в соответствии с (10.20) описывается векторной функцией

$$\mathbf{a}(t) = \mathbf{e}_\varphi + \frac{h}{2\pi} \mathbf{e}_z.$$

Производная векторной функции спирали в соответствии с формулой (12.3) равна

$$\frac{d\mathbf{a}}{dt} = -\rho \mathbf{e}_r.$$

В каждой точке спирали она направлена к ее оси и имеет длину ρ .

В дальнейшем при построении кривых линий и поверхностей мы будем использовать декартову прямоугольную систему координат, как наиболее удобную для вычисления производных векторных функций в евклидовом пространстве. При этом будем использовать нижние индексы, так как в декартовой прямоугольной системе координат ковариантные и контравариантные компоненты равны.

1.13. Математическая модель геометрии объектов

Моделирование реального или воображаемого объекта представляет собой совокупность действий, которые позволяют создавать его математическую модель, редактировать ее, изменять ее положение и ориентацию в пространстве и обеспечивают взаимодействие с другими моделями. Взаимодействием мы называем выполнение различных операций над моделями: установление зависимости параметров одной модели от параметров других моделей, определение взаимного положения моделей. Для выполнения этих действий необходима информация об объекте. Геометрическая информация об объекте может храниться в виде структуры данных или может вычисляться. Определим математическую модель реального или воображаемого объекта как совокупность данных и функций, позволяющих получить

необходимую информацию об объекте и изменять его модель требуемым образом (рис. 13.1).

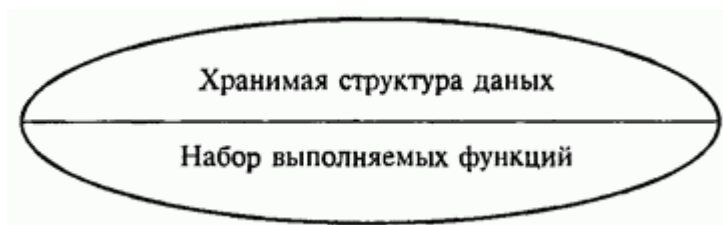


Рис. 13.1. Математическая модель

Программную реализацию структуры данных и функций называют классом. Геометрические объекты будут иметь свои данные и свои функции. Для построения точки в структуре ее данных достаточно хранить три координаты и иметь функции выполнения операций над радиус-векторами. Для построения произвольной линии нужно знать зависимость ее радиус-вектора $\mathbf{r}(t)$ от параметра, область изменения параметра и иметь функции вычисления производных радиус-вектора. Для построения поверхности нужно знать зависимость ее радиус-вектора $\mathbf{r}(u, v)$ от параметров, область изменения параметров и иметь функции вычисления частных производных радиус-вектора. Тело мы будем описывать совокупностью ограничивающих его поверхностей, дополненной информацией об их связях друг с другом.

С математической точки зрения все геометрические объекты равноправны. Для них существует ряд общих выполняемых функций. Все геометрические объекты могут быть подвержены модификациям сдвига, поворота, масштабирования, симметрии, поэтому они должны иметь функции, соответствующие этим действиям. Кроме того, для работы с геометрическим объектом нужны функции создания объекта (конструкторы), удаления объекта (деструктор), функция создания копии объекта, функции доступа к данным объекта, функции редактирования данных объекта. Математическая модель должна быть дополнена функциями, обеспечивающими взаимодействие объектов и выполнение над ними различных операций.

Мы рассмотрели теоретические основы геометрического моделирования. Дальнейшее изложение будет посвящено конкретным моделям и методам их построения.

2. Введение в топологию

2.1. Топология линий

2.1.1. Идея непрерывности

В основе, в истоках каждого отдела математики можно видеть основную идею, которая пронизывает всё его здание и определяет его лицо. Основной идеей топологии является идея *непрерывности*. Она встречается уже в математическом анализе, но, подчиненная другим идеям анализа, не получает там заметного развития. Свое полное и всестороннее развитие идея непрерывности получает в топологии. Приведем два примера применения этой идеи.

Пример 1. Покажем, что кубическое уравнение

$$x^3 + ax^2 + bx + c = 0 \quad (1)$$

с произвольными действительными коэффициентами a, b, c имеет, по крайней мере, один действительный корень. Запишем уравнение (1) (при $x \neq 0$) в виде

$$x^3 \left(1 + \frac{a}{x} + \frac{b}{x^2} + \frac{c}{x^3} \right) = 0. \quad (2)$$

При очень большом $|x|$ слагаемые $\frac{a}{x}, \frac{b}{x^2}, \frac{c}{x^3}$ очень малы по модулю, а выражение в скобках мало отличается от единицы и потому положительно. Следовательно, при очень большом $|x|$ левая часть уравнения (2) имеет тот же знак, что и x^3 , т. е. тот же знак, что и x . Иначе говоря, при большом отрицательном x (точка x_0 на рис. 1) левая часть уравнения (1) отрицательна, а при большом положительном x (точка x_1) она положительна.

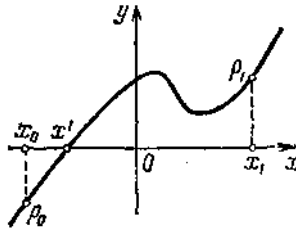


Рис. 1

Так как график является *непрерывной кривой*, то, переходя с одной стороны оси абсцисс на другую (из точки p_0 в точку p_1), он пересечет

эту ось хотя бы в одной точке. Точка x' пересечения графика с осью абсцисс и дает корень уравнения (1).

Задачи

1. Докажите, что всякое уравнение нечетной степени с действительными коэффициентами имеет хотя бы один действительный корень.
2. Докажите, что если в уравнении (1) свободный член с отрицателен, то это уравнение имеет хотя бы один положительный корень.

Пример 2. Покажем, что вокруг всякой замкнутой кривой K можно описать квадрат.

Проведем две параллельные прямые l, l' так, чтобы кривая K была расположена в полосе между ними, и будем перемещать сначала прямую l , а затем l' параллельно себе, до тех пор, пока они не коснутся кривой K . Полученные прямые m, m' (рис. 2, а) называются параллельными между собой *опорными прямыми* линии K . Проведем еще две опорные прямые, перпендикулярные l (рис. 2, б).

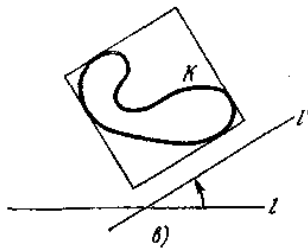
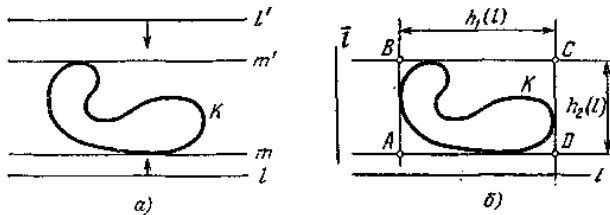


Рис. 2.

Мы получим описанный вокруг линии K прямоугольник $ABCD$. Докажем, что при некотором направлении прямой l прямоугольник $ABCD$ превратится в квадрат.

Обозначим длину стороны AD , параллельной l , через $h_1(l)$, а длину стороны AB , перпендикулярной l , — через $h_2(l)$. Описанный прямоугольник будет квадратом при

$$h_1(l) - h_2(l) = 0.$$

Пусть \bar{l} — прямая, перпендикулярная l . Описанный прямоугольник со сторонами, параллельными и перпендикулярными \bar{l} , совпадает с тем же прямоугольником $ABCD$, но параллельной \bar{l} является сторона AB , а перпендикулярной \bar{l} — сторона AD , т. е.

$$h_1(\bar{l}) = |AB| = \hat{h}_2(l), \quad h_2(\bar{l}) = |AD| = \hat{h}_1(l). \quad \text{Таким образом,}$$

$$h_1(\bar{l}) - h_2(\bar{l}) = -(h_1(l) - h_2(l)). \quad (3)$$

Будем теперь поворачивать прямую l , пока она не совпадет с \bar{l} . При этом описанный прямоугольник будет плавно меняться, а разность $h_1(l) - h_2(l)$ будет непрерывно зависеть от l . Но при переходе от l к \bar{l} эта разность меняет знак (см. (3)). Следовательно, при своем непрерывном изменении она (при некотором положении прямой l) обращается в нуль, т. е. описанный прямоугольник превращается в квадрат (рис. 2, в).

Задачи

3. Докажите, что вокруг всякой замкнутой кривой K можно описать ромб с углом 60° .

4. Докажите, что если диаметр плоской фигуры не превосходит d (т. е. расстояние между любыми ее точками не больше d), то существует правильный шестиугольник, содержащий эту фигуру, у которого расстояние между противоположными сторонами равно d .

5. Докажите, что если диаметр пространственной фигуры не превосходит d , то существует правильный октаэдр, содержащий эту фигуру, у которого расстояние между противоположными гранями равно d .

В топологии рассматриваются функции наиболее общего вида. Задать функцию — это значит каждой точке x некоторого множества A (области определения функции) поставить в соответствие определенную точку $f(x)$ некоторого другого множества B . Говорят в этом случае также, что задано *отображение* f множества A в множество B . Кратко это записывают в виде $f: A \rightarrow B$.

Пример 3. Обозначим через A контур равностороннего треугольника, а через B — описанную вокруг него окружность (рис. 3).

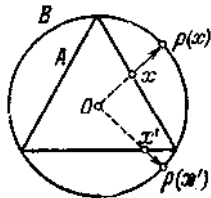


Рис. 3.

Тогда центральное проектирование p точек множества A на окружность является отображением $p: A \rightarrow B$. Функция $f: A \rightarrow B$ называется непрерывной в точке $x_0 \in A$, если для x , «мало» отличающихся от x_0 , значения $f(x)$ и $f(x_0)$ тоже «мало» отличаются друг от друга.

Более точно, функция $f(x)$ непрерывна в точке x_0 , если для любого числа $\varepsilon > 0$ можно подобрать такое число $\delta > 0$, что для любого x , отстоящего от x_0 менее чем на δ , соответствующее значение $f(x)$ отстоит от $f(x_0)$ менее чем на ε . Это определение имеет смысл, если и в множестве A , и в множестве B определено расстояние между точками.

Чтобы яснее понять, что такое непрерывность отображения, рассмотрим пример разрыва, т. е. нарушения непрерывности отображения. Возьмем обычную резинку, которую применяют в аптеке для упаковки лекарств. Она имеет форму окружности. Будем ее осторожно деформировать, и вдруг она в некоторой своей точке a разорвется. Что это значит? Некоторая ее часть B (рис. 4, а) которая была «близка» к a (т. е. находилась от нее на нулевом расстоянии), после разрыва (теперь ее обозначим через B') оказывается вовсе не близкой к a' (новое положение точки a).

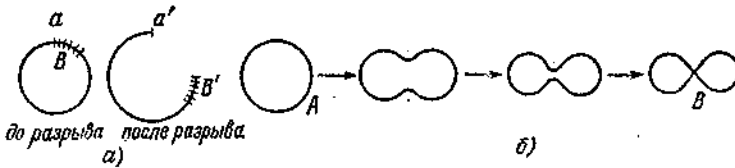


Рис. 4.

Итак, разрыв в точке a — это такое событие, когда некоторая часть B фигуры, бывшая раньше близкой к a (пишем: Bda), становится неблизкой к новому положению a' точки a .

Теперь понятно следующее определение:

Отображение $f: x \rightarrow x'$ называется непрерывным в точке a , если всякая часть B отображаемой фигуры, бывшая близкой к a (т. е. Bda), после отображения переходит в положение B' , близкое к точке

$a' = f(a)$ (т. е. $B'da'$).

Можно доказать, что это определение эквивалентно приведенному выше определению.

Если отображение $f: A \rightarrow B$ непрерывно в каждой точке x_0 множества A , то говорят просто, что отображение f непрерывно. Наглядно непрерывность отображения можно себе представить, сказав,

что «близкие» точки множества A переходят в «близкие» точки множества B , т. е. при отображении f не происходит разрывов, нарушения цельности множества A . Заметим, что при этом различные точки множества A могут образовывать «спайки» (рис. 4, б), «складки» и т. п.

2.1.2. Чем занимается топология?

Отображение $f: A \rightarrow B$ называется *взаимно однозначным*, если в каждую точку множества B отображается точно одна точка множества A . Это означает, что, во-первых, никакие две различные точки множества A не переходят в одну и ту же точку множества B (не «склеиваются» при отображении f) и, во-вторых, каждая точка множества B поставлена в соответствие некоторой точке множества A (т. е. A отображается на всё множество B , а не на его часть). Для взаимно однозначного отображения $f: A \rightarrow B$ можно определить *обратное отображение* $f^{-1}: B \rightarrow A$ (которое каждой точке $y \in B$ ставит в соответствие точку множества A , переходящую в y при отображении f).

Отображение $f: A \rightarrow B$ называется *гомеоморфным отображением* (или *гомеоморфизмом*), если оно, во-первых, взаимно однозначно и, во-вторых, взаимно непрерывно, т. е. не только само отображение f непрерывно, но и обратное отображение f^{-1} также непрерывно.

Наглядно гомеоморфизм можно представлять себе как такое отображение одного множества на другое, которое происходит и без разрывов, и без склеиваний. Например, будем считать, что фигуры A, B «изготовлены» из очень прочного и эластичного материала, и будем допускать любые растяжения и искривления этого материала без разрывов и без образования складок и склеек; если мы сможем при этих условиях «наложить» фигуру A на B , то они гомеоморфны. Так, контур треугольника (или, вообще, любого многоугольника) гомеоморфен окружности.

Пример 4. Поверхность шара, поверхность куба, цилиндра — все гомеоморфны между собой. Однако эти поверхности не гомеоморфны *тору* (который можно наглядно представить себе как поверхность баранки или автомобильной камеры, рис. 5). Поверхность гири (рис. 6) гомеоморфна тору.

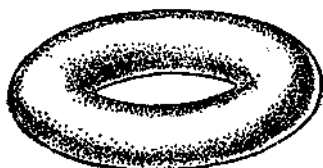


Рис. 5.



Рис. 6,

Пример 5. Будем представлять себе буквы русского алфавита в виде линий. Буквы **Г, Л, М, П, С** гомеоморфны между собой. Буквы **Е, У, Т, Ч, Ш, Ц, Э** также гомеоморфны между собой, но не гомеоморфны указанным ранее буквам. Буква **О** не гомеоморфна никакой другой букве русского алфавита.

Пример 6. Пусть A — полуокружность с центром o , из которой исключены концевые точки m, n , а B — касательная к полуокружности, параллельная диаметру mn (рис. 7). Центральное проектирование $p: A \rightarrow B$ из центра o является гомеоморфизмом. Таким образом, прямая гомеоморфна полуокружности без концевых точек.

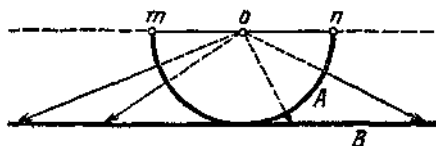


Рис. 7.

В свою очередь полуокружность гомеоморфна отрезку (ее можно распрямить). Следовательно, *прямая гомеоморфна открытому отрезку* (т. е. отрезку, из которого выброшены концевые точки).

Задачи

8. Докажите, что фигура, являющаяся объединением боковой поверхности цилиндра и его нижнего основания («стакан»), гомеоморфна кругу.

9. Докажите, что плоскость гомеоморфна открытому кругу (т. е. кругу, к которому не причисляются точки ограничивающей его окружности), а также сфере, из которой «выколота» (удалена) одна точка.

10. Докажите, что фигуры, изображенные на рис. 8 (лента, гомеоморфная боковой поверхности цилиндра, и дважды перекрученная лента) гомеоморфны между собой.

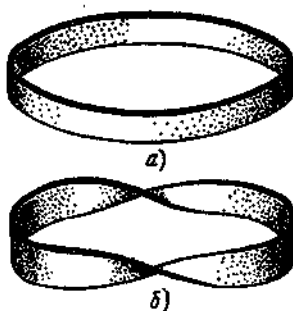


Рис. 8

Поучительно сравнить понятие гомеоморфизма и понятие конгруэнтности фигур. В геометрии рассматриваются отображения, сохраняющие расстояния между точками. Они называются *движениями* (или перемещениями). В результате движения каждая фигура перекладывается на новое место как твердое целое, без изменения расстояний. Две фигуры, которые переводятся одна в другую («совмещаются») с помощью движения, называются *конгруэнтными* и рассматриваются как одинаковые, как не отличающиеся (с геометрической точки зрения) друг от друга. В топологии рассматриваются отображения, более общие, чем движения, а именно гомеоморфные отображения. Две гомеоморфные между собой фигуры рассматриваются (с топологической точки зрения) как одинаковые, не отличающиеся друг от друга. *Те свойства фигур, которые не изменяются при гомеоморфных отображениях, называются топологическими свойствами фигур, или топологическими инвариантами* (от латинского слова *invariant* — неизменный). Изучением топологических свойств фигур и занимается *топология*.

Задачи

11. Если фигура A состоит лишь из конечного числа точек, то через $n(A)$ обозначим число ее точек; если же фигура A содержит бесконечно много точек, то условимся писать $n(A)=\infty$. Является ли $n(A)$ топологическим инвариантом?

12. Фигура A называется *вложимой в плоскость*, если она гомеоморфна некоторой фигуре, лежащей в плоскости. Например, «стакан» (задача 8) вложим в плоскость. Является ли свойство фигуры быть вложимой в плоскость топологическим инвариантом?

Не следует думать, что любые две гомеоморфные между собой фигуры можно перевести одну в другую, изгибая, растягивая и перемещая их (без разрывов и склеиваний) в пространстве. Например, фигуры,

изображенные на рис. 8, нельзя перевести одну в другую таким способом, они «неодинаково расположены» в пространстве. Верхнюю ленту нужно разрезать и затем, дважды перекрутив, снова склеить по тем же самым точкам; лишь после этого ее удастся совместить со второй лентой. Этот прием (разрезание фигуры, а затем, после надлежащего растяжения, перемещения ее частей — обратное склеивание) часто используется в топологии для доказательства гомеоморфности двух фигур.

«Одинаковость расположения» двух фигур в пространстве (или в объемлющей их фигуре) уточняется понятием *изотопии*. Говорят, что две гомеоморфные фигуры A и B *изотопны* в объемлющей их фигуре P (или, иначе, *топологически одинаково расположены в P*), если существует гомеоморфное отображение фигуры P на себя, при котором фигура A переходит в B . Ленты на рис. 8 гомеоморфны, но не изотопны друг другу в пространстве (доказательство этого будет дано ниже). О свойствах расположения можно говорить, если задана *пара фигур*: фигура A и объемлющая ее фигура P . Изучением свойств расположения (т. е. изучением топологических инвариантов пары фигур) также занимается топология.

Задачи

13. Линия A (рис. 9) не разрезает тор T на две части, а линия C разрезает. Изотопны ли A и C в фигуре T ? Изотопны ли A и C в трехмерном пространстве?



Рис. 9.



Рис. 10.

14. Докажите, что меридиан A и параллель B тора T (см. рис. 9) изотопны в T .

15. Докажите, что в фигуре восьмерки (рис. 10) любые две точки, отличные от x , изотопны.

2.1.3. Простейшие топологические инварианты

Выше (пример 4) мы говорили, что поверхность шара (т. е. сфера) не гомеоморфна тору, и вряд ли у читателя возникло сомнение в этом. Но как доказать, что две фигуры не являются гомеоморфными? Ведь из

того, что мы не сумели найти гомеоморфного отображения одной фигуры на другую, не вытекает еще с достоверностью, что такого гомеоморфного отображения не существует.

Для доказательства того, что две фигуры не гомеоморфны друг другу, пользуются *топологическими инвариантами*. Пусть, например, с помощью некоторого правила каждой фигуре ставится в соответствие определенное число, причем так, что числа, соответствующие двум гомеоморфным фигурам, всегда оказываются равными. Тогда это число выражает некоторое свойство фигуры, сохраняющееся при гомеоморфных отображениях, т. е. являющееся топологическим инвариантом. Если теперь две фигуры A и B таковы, что соответствующие им числа оказались различными, то эти фигуры не могут быть гомеоморфными.

Пример 7. Буква **bl** представляет собой фигуру, состоящую из двух «кусков», из двух не связанных между собой частей. Остальные буквы русского алфавита, кроме **Й, Ё**, состоят из одного связного куска. Число связных «кусков», из которых состоит фигура (говорят также: *число компонент* фигуры), является топологическим инвариантом; если две фигуры гомеоморфны, то обе они состоят из одинакового числа компонент. Поэтому буква **bl** не гомеоморфна, например, букве **О**, букве **П**, букве **Ц** и т. д.

Пример 8. На фигуре восьмерки (см. рис. 10) имеется такая точка x , что после удаления из восьмерки точки x вместе с близлежащими точками (рис. 11) мы получаем несвязную фигуру (содержащую более одной компоненты).



Рис. 11.

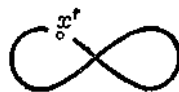


Рис. 12.

Точку, обладающую этим свойством, называют *разбивающей точкой* фигуры. Никакая отличная от x точка x' восьмерки не является разбивающей (рис. 12). Понятия «разбивающая точка», «неразбивающая точка» топологически инвариантны: если x есть разбивающая точка фигуры A , а $f: A \rightarrow B$ — гомеоморфное отображение, то $f(x)$ есть разбивающая точка фигуры B . Поэтому число разбивающих точек данной фигуры есть, ее топологический инвариант, число неразбивающих точек — также топологический инвариант.

Задачи

16. Для каждой из букв русского алфавита укажите, сколько разбивающих и неразбивающих точек она содержит. Докажите, что буквы **О**, **Г**, **Т**, **Б** попарно не гомеоморфны.

17. Докажите, что для любого натурального n существует фигура, содержащая ровно n разбивающих точек, а также фигура, содержащая ровно n неразбивающих точек.

Пример 9. Пусть A — некоторая фигура, составленная из конечного числа дуг, и x — ее точка. Число дуг фигуры A , сходящихся в точке x , называется *индексом* точки x в фигуре A . В фигуре буквы **Ж** (рис. 13) точка a имеет индекс 1, точка b имеет индекс 2, точка c — индекс 3, а точка d — индекс 4.

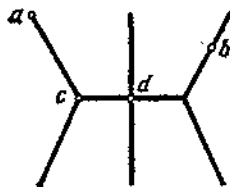


Рис. 13

Число точек индекса 1, содержащихся в фигуре A , число точек индекса 3, индекса 4 и т. д.— все это различные топологические инварианты фигуры A .

Задачи

18. Докажите, что буквы Ю и Ф не гомеоморфны.

19. Сформулируйте необходимое и достаточное условие, при выполнении которого фигура, составленная из конечного числа дуг, гомеоморфна окружности.

Фигуры, состоящие из конечного числа дуг, называют в топологии *конечными графами*. В конечном графе имеется конечное число *вершин* и некоторые из этих точек соединяются непересекающимися дугами (*ребрами* графа); при этом две вершины графа можно соединять несколькими ребрами и, кроме того, допускаются замкнутые ребра («петли»), которые начинаются и кончаются в одной и той же вершине.

20. Пусть G — конечный граф. Через $a_k(G)$ обозначим число вершин этого графа, имеющих индекс k . Докажите, что число ребер графа G равно $\frac{1}{2}(a_1(G) + 2a_2(G) + 3a_3(G) + \dots)$.

21. Докажите, что во всяком конечном графе число вершин, имеющих нечетный индекс, четно.

Пример 10. Граф называется *уникурсальным*, если его можно «нарисовать одним росчерком», т. е. пройти его весь непрерывным движением, не проходя одно и то же ребро дважды. Свойство графа быть уникурсальным является, очевидно, топологически инвариантным. Однако этот топологический инвариант не является новым, а выражается через понятие индекса точки (см. задачу 24).

Задачи

22. Докажите, что если каждая вершина конечного графа имеет индекс, не меньший двух, то этот граф содержит линию, гомеоморфную окружности и составленную из ребер графа.

23. Докажите, что если все вершины конечного связного графа имеют четный индекс, то этот граф можно «нарисовать одним росчерком», начав движение из произвольно заданной его вершины и возвращаясь в ту же вершину.

24. Докажите, что связный граф тогда и только тогда уникурсален, когда он содержит не более двух вершин нечетного индекса.

С уникурсальными графами тесно связана задача о кенигсбергских мостах, рассмотренная еще Эйлером. В то время в Кенигсберге (ныне Калининград областной) было 7 мостов (рис. 14) через реку Прегель. Вопрос состоит в том, можно ли, прогуливаясь по городу, пройти

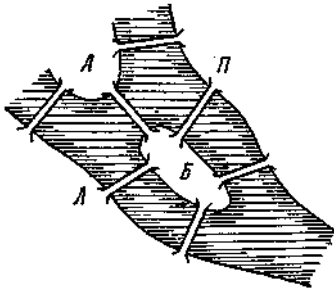


Рис. 14.

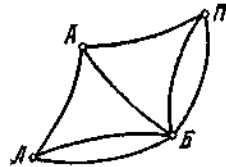


Рис. 15.

через каждый мост точно по одному разу. Сопоставим с планом города некоторый граф: вершина $Л$ обозначает левый берег, $П$ — правый берег, A и B острова, а ребра графа соответствуют мостам (рис. 15). В этом графе все четыре вершины имеют нечетный индекс. Следова-

тельно, граф не уникарсален, и потому требуемого маршрута прогулки не существует.

Задачи

25. Докажите, что, добавив еще один мост (где угодно на плане рис. 14), мы получим схему города, в котором можно пройти через, каждый мост трчно по одному разу.

26. *Полным графом* называется конечный граф без петель, у которого любые две вершины соединены точно одним ребром. В каком случае полный граф уникарсален?

2.1.4. Эйлерова характеристика графа

Всякий граф можно постепенно «построить», добавляя одно ребро за другим. Например, можно в требуемом графе перенумеровать заранее все ребра, а затем вычерчивать последовательно первое ребро, второе, третье и т. д.

Пример 11. На рис. 16 показан граф, который мы хотим построить, и выбрана нумерация его ребер (некоторые ребра прерваны, чтобы показать наглядно их возможное расположение в пространстве).

Нумерация ребер на рис. 16 выбрана так, что при последовательном вычерчивании графа в соответствии с этой нумерацией все время получаются связные графы. Если бы, однако, мы занумеровали ребра в обратном порядке, то при вычерчивании у нас получился бы несвязный граф, содержащий три отдельно взятых ребра, и лишь потом, при вычерчивании новых ребер, мы получили бы связный граф.

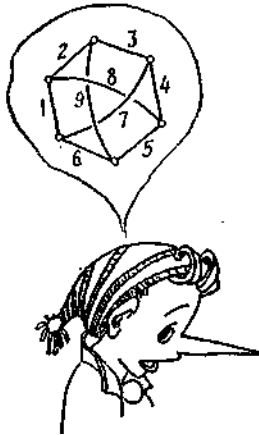


Рис. 16.

Естественно возникает вопрос: во всяком ли связном графе существует такая нумерация ребер, что при последовательном вычерчивании графа (в соответствии с этой нумерацией) все время получаются связные графы?

Ответ на этот вопрос утвердителен (см. задачу 28). Иначе говоря, *всякий связный граф может быть получен следующим образом: мы берем одно ребро, затем присоединяем к нему еще одно ребро так, чтобы снова получился связный граф, затем присоединяем еще одно ребро (так, чтобы снова получился связный граф) и т. д.* Это утверждение можно назвать «теоремой о вычерчивании связных графов».

Задача

27. Докажите, что любой связный граф можно вычертить «одним росчерком», если разрешить проходить каждое ребро точно два раза.

28. Выведите из предыдущей задачи доказательство теоремы о вычерчивании связного графа.

29. Докажите, что любые две вершины связного графа G можно соединить в G *простой цепочкой ребер*, т. е. такой, что объединение ребер этой цепочки гомеоморфно отрезку.

Указание: если цепочка ребер, соединяющая a и b , дважды проходит через некоторую вершину c , то она содержит «замкнутую цепочку» (начинающуюся и кончающуюся в c), которую можно удалить.

30. Докажите, что если любые две вершины графа G можно соединить не менее чем двумя различными простыми цепочками, то граф G не имеет вершин индекса 1. Верна ли обратная теорема?

Контуром в графе называется замкнутая цепочка ребер, объединение которых представляет собой линию, гомеоморфную окружности (рис. 17). Связный граф, не содержащий ни одного контура называется *деревом* (рис. 18).

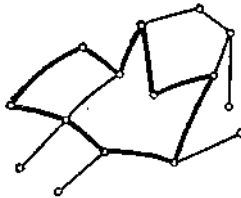


Рис. 17.

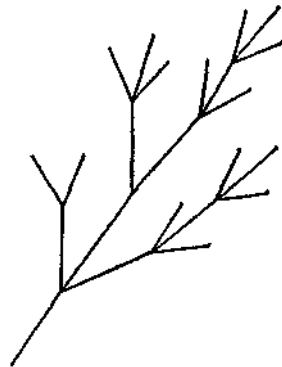


Рис. 18.

Мы докажем, что для любого дерева, имеющего V вершин и P ребер, справедливо соотношение

$$V - P = 1. \quad (4)$$

Для доказательства проведем индукцию по числу ребер P . При $P = 1$ (дерево имеет одно ребро и две вершины) соотношение (4) справедливо. Предположим, что для любого дерева, имеющего n ребер, соотношение (4) уже доказано, и пусть G — дерево, имеющее $n + 1$ ребро. Так как граф G связан, то его можно получить из некоторого связного графа G' добавлением одного ребра r (это вытекает из «теоремы о вычерчивании связного графа»). Граф G' содержит n ребер и тоже не содержит контуров т. е. является деревом. По предположению индукции для дерева G' соотношение (4) справедливо, и потому в G' имеется $n + 1$ вершина. Заметим теперь, что только один конец добавляемого ребра r является вершиной графа G' (в противном случае, взяв в G простую цепочку, соединяющую вершины a и b , и добавив к этой цепочке ребро r , мы получили бы контур в графе G ; рис. 19). Следовательно, при добавлении ребра r в графе G появляется одно новое ребро и одна новая вершина (рис. 20). Иначе говоря, граф G имеет $n + 2$ вершины и $n + 1$ ребро, и потому соотношение (4) для него справедливо.

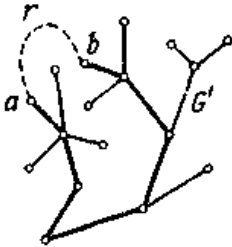


Рис. 19.

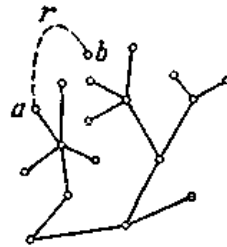


Рис. 20.

Проведенная индукция доказывает равенство (4) для любого дерева. Разность $V - P$, где V — число вершин, а P — число ребер графа G , называется *эйлеровой характеристикой* этого графа и обозначается через $\chi(G)$. Таким образом, *эйлерова характеристика любого дерева равна 1*.

Задачи

31. Граф, не содержащий контуров, называется *лесом*. Докажите, что если G — лес, то число деревьев, которые «растут» в нем (т. е. число компонент графа G), равно $\chi(G)$.

32. Докажите, что если G — дерево, то каждые две его вершины соединены только одной простой цепочкой. Верно ли обратное?

Пусть теперь G — связный граф, не являющийся деревом. Тогда в G имеется контур; пусть r_1 — какое-либо ребро, входящее в этот контур (рис. 21).

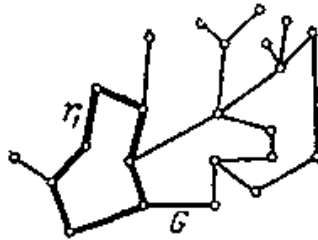


Рис. 21.

Удалив из G ребро r_1 , мы получим связный граф G' (поскольку концы выброшенного ребра r_1 соединены в G' простой цепочкой — оставшейся частью контура), причем вершины у графа G' — те же, что и у G . Если G' еще не является деревом, т. е. в G' также есть контур (рис. 22), то мы можем взять произвольное ребро r_2 этого контура и, выбросив его, получить связный граф G'' с теми же вершинами, что и у G , и т. д.

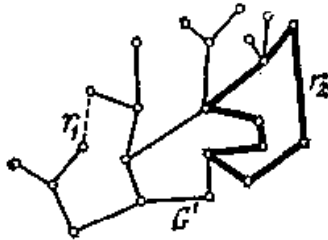


Рис. 22.

После выбрасывания какого-то ребра r_1 мы получим связный граф G^* , содержащий все вершины графа G и уже не имеющий контуров, т. е. являющийся деревом. Оно называется *максимальным деревом* графа G , а ребра r_1, r_2, \dots, r_k называются *перемычками*.

Если V — число вершин графа G , то максимальное дерево G^* имеет те же V вершин. Согласно (4) G^* имеет $V - 1$ ребро, и потому число ребер графа G равно $V - 1 + k$ (чтобы из G^* получить G , надо «возвратить» k выброшенных ребер-перемычек). Следовательно,

$$\chi(G) = V - (V - 1 + k) = 1 - k. \quad (5)$$

Так как $k \geq 1$, то $\chi(G) \leq 0$. Таким образом, для любого связного графа G справедливо соотношение $\chi(G) \leq 1$; равенство достигается в том и только в том случае, если G — дерево.

Далее, согласно (5) число перемычек $k = 1 - \chi(G)$. Иначе говоря, для получения графа G надо к его максимальному дереву добавить $1 - \chi(G)$ ребер-перемычек, каждое из которых соединяет две (возможно, совпадающие) вершины максимального дерева.

Задача

33. Если связный граф G получается из некоторого дерева добавлением нескольких замкнутых ребер («петель»), то в G имеется единственное максимальное дерево. Верно ли обратное?

34. Докажите, что если граф G содержит l компонент, то $\chi(G) \leq l$. В каком случае достигается равенство?

35. Будем говорить, что в графе G задана система токов, если каждому ребру сопоставлено направление и неотрицательное число (ток), причем выполняется правило Кирхгофа: для каждой вершины сумма входящих в нее токов равна сумме исходящих. Докажите, что если G — дерево, то в нем существует только тривиальная система токов (все токи равны нулю).

36. Пусть G — связный граф, G^* — его максимальное дерево, а r_1, r_2, \dots, r_k — перемычки. Докажите, что если произвольно задать токи на ребрах r_1, r_2, \dots, r_k , то их можно однозначно дополнить токами на остальных ребрах так, чтобы получилась система токов в G .

Указание. Для каждой перемычки r_i существует единственный контур, содержащий ее и не содержащий других перемычек. Если пустить по этому контуру ток такой величины и направления, как указано на перемычке r_i , а затем взять сумму всех этих «контурных токов», то мы и получим требуемую систему токов на графе G . Если бы существовали две различные системы токов, совпадающие на перемычках, то их разность была бы нетривиальной системой токов на дереве G^* .

2.1.5. Индекс пересечения

В следующих двух примерах рассматриваются графы, не вложимые в плоскость.

Пример 12 («домики и колодцы»). На плоскости даны шесть точек D_1, D_2, D_3 (домики) и K_1, K_2, K_3 (колодцы); можно ли на плоскости провести тропинки от каждого домика к каждому колодцу, так чтобы никакие две тропинки не пересекались? Ответ отрицательный: если мы

проведем все тропинки, кроме одной (рис. 23), то для последней тропинки уже «не будет места» на плоскости.

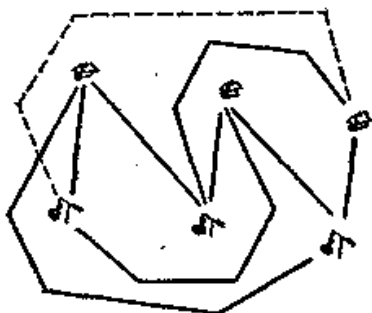


Рис. 23.

Таким образом, граф P_1 , изображенный на рис. 23, не вложим в плоскость.

Пример 13. Обозначим через P_2 полный граф с пятью вершинами. На рис. 24 одно ребро прервано: для него «нет места» на плоскости.

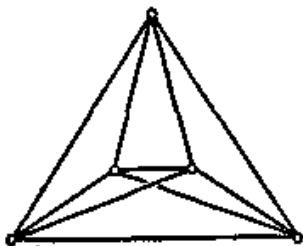


Рис. 24.

Таким образом, граф P_2 также не вложим в плоскость.

Интересно отметить, что построенные графы P_1 , P_2 являются «эталоном» графов, не вложимых в плоскость: *если граф не вложим в плоскость, то он обязательно содержит граф, гомеоморфный P_1 или P_2* . Это было доказано польским математиком К. Куратовским

Задачи

37. Докажите, что граф, рассмотренный в примере 11 (рис. 16), не вложим в плоскость.

38. Ребрами графа служат стороны и наименьшие диагонали правильного n -угольника. Докажите, что при четном n этот граф может быть вложен в плоскость, а при нечетном — нет.

39. Ребрами графа служат стороны и наибольшие диагонали правильного $2n$ -угольника. Докажите, что при $n \geq 3$ этот граф не вложим в плоскость, но его можно расположить на торе.

40. Ребрами графа служат стороны и наибольшие диагонали правильного $(2n + 1)$ -угольника. Докажите, что при $n \geq 2$ этот граф не вложим в плоскость. Можно ли его расположить на торе?

Рассуждения, приведенные в примерах 12 и 13 («нет места» на плоскости), конечно, являются не доказательствами, а лишь пояснениями. Ниже мы изложим строгое доказательство того, что графы P_1, P_2 не вложимы в плоскость.

Пусть a, b — два отрезка на плоскости, ни один из которых не содержит концевых точек другого отрезка. Если эти отрезки пересекаются, то будем писать $J(a, b) = 1$, а если нет, то $J(a, b) = 0$. Число $J(a, b)$ назовем *индексом пересечения отрезков a и b* .

Конечное множество отрезков на плоскости будем называть *цепью*. Отрезки, составляющие цепь, назовем ее *звеньями*, а точки, являющиеся концами звеньев, — *вершинами*.

Пусть x и y — две такие цепи, что ни одна из них не содержит ни одной вершины другой цепи. Отрезки, составляющие цепь x , обозначим через a_1, \dots, a_m , а отрезки, составляющие цепь y , — через b_1, \dots, b_n .

Если сумма $\sum_{i,j} J(a_i, b_j)$ (т. е. сумма индексов пересечения каждого

из отрезков a_1, \dots, a_m с каждым из отрезков b_1, \dots, b_n) *четна*, то будем писать $J(x, y) = 0$, а если эта сумма *нечетна*, то $J(x, y) = 1$. Число $J(x, y)$ назовем *индексом пересечения цепей x и y* (точнее, индексом пересечения по модулю 2).

Цепь, в каждой вершине которой сходится четное число звеньев, будем называть *циклом* (по модулю два). Мы докажем, что *индекс пересечения двух циклов на плоскости всегда равен нулю*.

В самом деле, так как каждая вершина цикла имеет индекс, не меньший двух, то этот цикл содержит ломаную, гомеоморфную окружности (см. задачу 17). Если из цикла выбросить эту замкнутую ломаную, то останется снова цикл (каждая вершина имеет четный индекс). В оставшемся цикле можно снова выделить ломаную, гомеоморфную окружности, и т. д. Итак, *каждый цикл можно представить как объединение конечного числа ломаных, гомеоморфных окружности (причем эти ломаные попарно не имеют общих отрезков)*.

Поэтому для доказательства того, что индекс пересечения двух циклов x, y на плоскости всегда равен нулю, достаточно установить это в случае, когда каждый из циклов x, y представляет собой ломаную, гомеоморфную окружности. Сдвинув чуть-чуть вершины циклов x и y

(что не изменит их индекс пересечения), мы можем добиться того, чтобы звенья, составляющие циклы x и y , были попарно не параллельными. Выберем теперь прямую l , не параллельную ни одной прямой, соединяющей какую-либо вершину цикла x с какой-либо вершиной цикла y .

Будем непрерывно перемещать цикл x (как твердое целое) параллельно прямой l (рис. 25).

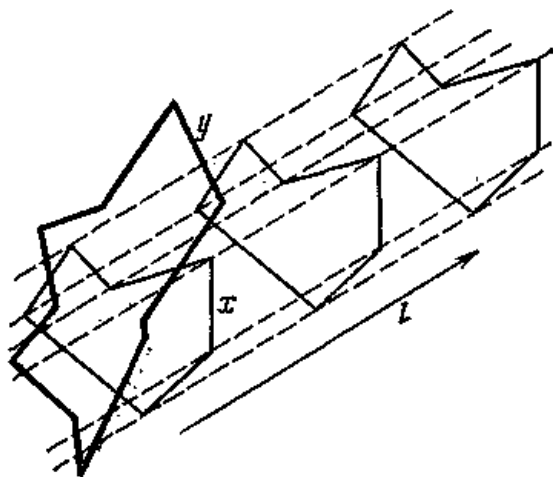


Рис. 25.

Индекс пересечения $J(x, y)$ мог бы изменяться лишь в те моменты, когда вершины одного из циклов x, y попадают на стороны другого (вершины цикла x не могут попасть в вершины цикла y в силу выбора прямой l). Однако в момент, когда некоторое звено a цикла x проходит через вершину q цикла y , число точек пересечения не меняет своей четности (рис. 26—28).



Рис. 26.

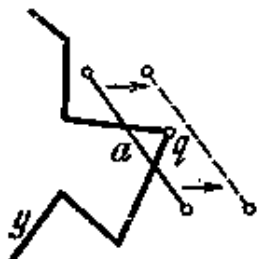


Рис. 27.

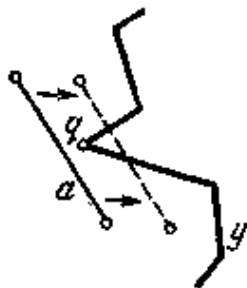


Рис. 28.

То же происходит и при прохождении вершин цикла x через стороны цикла y . Поэтому индекс пересечения $J(x, y)$ не меняется. Но в конце концов цикл x попадает в положение, в котором он не имеет общих точек с y (рис. 25), так что индекс пересечения становится равным нулю. Следовательно, и первоначально было $J(x, y) = 0$.

Теперь мы в состоянии доказать, что граф P_1 (пример 12), не вложим в плоскость. Две тропинки, ведущие от разных домиков к разным колодцам, условимся называть *несмежными*. Проведем (в виде ломаных линий) все требуемые тропинки (возможно, с пересечениями) и обозначим через I число точек пересечения по всем парам несмежных тропинок. Мы покажем, что *при любом способе проведения тропинок число I нечетно*.

Предположим, что мы меняем положение одной тропинки, скажем, тропинки D_1K_1 . Первоначальное ее положение обозначим через x , а новое — через x' (рис. 29).

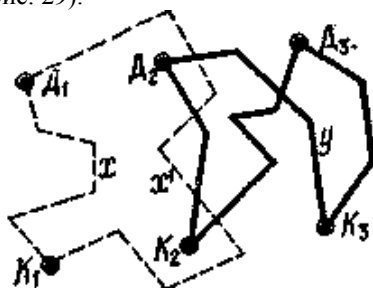


Рис. 29.

Несмежными с D_1K_1 являются четыре тропинки, соединяющие два оставшихся домика D_2, D_3 с двумя оставшимися колодцами K_2, K_3 . Они образуют цикл $D_2K_2D_3K_3D_2$, который мы обозначим через y .

Ломаные x и x' , вместе взятые, также образуют цикл. Так как индекс пересечения любых двух циклов равен нулю, то $J(x, y) = J(x', y)$. Иначе говоря, число точек пересечения тропинки x с циклом y (т. е. со всеми несмежными ей тропинками) имеет ту же четность, что и число точек пересечения тропинки x' с циклом y . Таким образом, при замене тропинки x тропинкой x' число I не меняет своей четности.

Но тогда ясно, что для любых двух расположений тропинок на плоскости число I имеет одну и ту же четность. Действительно, последовательно заменяя сначала одну тропинку первого расположения соответствующей тропинкой второго расположения, затем еще одну и т. д., мы постепенно заменим первое расположение тропинок вторым, а четность числа I , в силу доказанного, меняться не будет.

На рис. 23 имеется только одна точка пересечения тропинок, а потому для любого расположения тропинок число I нечетно. Следовательно, провести все тропинку без пересечений (т. е. так, чтобы было $I = 0$) невозможно и потому граф P_1 не вложим в плоскость.

Задачи

41. Докажите, что граф P_2 , рассмотренный в при мере 13, не вложим в плоскость.

42. Докажите, что на сфере (как и на плоскости) индекс пересечения любых двух циклов равен нулю. Покажите, что на торе существуют два цикла, индекс пересечения которых равен 1.

Пусть теперь a и b — два направленных отрезка, ни один из которых не содержит концов другой отрезка. Если, идя по направлению первого отрезка a мы увидим, что второй отрезок b пересекает его справа налево, то будем писать $J(a, b) = 1$; если слева направо, то $J(a, b) = -1$; наконец, если a и b не пересекаются, то $J(a, b) = 0$. Число $J(a, b)$ будем называть индексом пересечения направленных отрезков a и b .

Цепью (точнее, «целочисленной цепью» — в отличие от цепей по модулю 2, рассматривавшихся ранее) будем теперь называть конечное множество направленных отрезков на плоскости. *Индекс пересечения целочисленных цепей* x, y (рассматриваемых в определенном порядке: x — первая цепь; y — вторая) определяется как и прежде;

$$J(x, y) = \sum_{i,j} J(a_i, b_j),$$

где a_1, \dots, a_m — направленные отрезки, составляющие цепь x , а b_1, \dots, b_n — направленные отрезки, составляющие цепь y .

Наконец, цепь условимся называть *циклом* (точнее, целочисленным циклом), если для каждой вершины число входящих в нее направленных отрезков равно числу исходящих.

Задачи

43. *Направленным контуром* условимся называть замкнутую ломаную, гомеоморфную окружности, на звеньях которой отмечено стрелками некоторое направление обхода (так, что в каждой вершине имеется одно входящее звено и одно исходящее). Направленный контур является циклом. Докажите, что всякий цикл (целочисленный) можно представить как объединение конечного числа направленных контуров, которые попарно не имеют общих звеньев.

44. Докажите, что индекс пересечения любых двух целочисленных циклов равен нулю.

45. На рис. 30 цикл x состоит из двух отдельных направленных контуров.

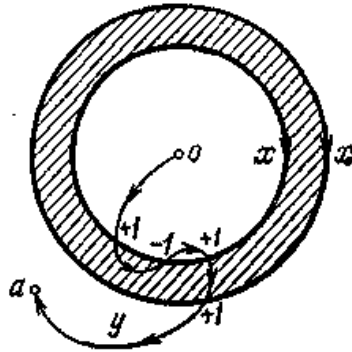


Рис. 30

Докажите, что точка a в том и только в том случае принадлежит внешней области кольца, если для любой направленной ломаной y , идущей от o к a , выполнено условие $J(x, y) = 2$. В каком случае точка a лежит внутри заштрихованной кольцевой области?

2.1.6. Теорема Жордана

Выше (см. рис. 25—28) мы доказали, что индекс пересечения любых двух циклов на плоскости равен нулю. Возможно, читатель готов предложить более простое доказательство: в каждой точке пересечения замкнутая ломаная x либо входит во внутреннюю область замкнутой ломаной y , либо выходит из внутренней области во внешнюю; так как число точек входа равно числу точек выхода (поскольку они чередуются), то общее число точек пересечения четно.

Однако это доказательство можно признать корректным, если уже выяснен смысл понятия «внутренняя область», а это понятие вовсе не

является таким простым, как кажется на первый взгляд. Выяснению его и посвящен этот пункт.

Замкнутая линия, гомеоморфная окружности, называется *простой замкнутой линией*. Теорема Жордана состоит в том, что *всякая простая замкнутая линия, расположенная на плоскости, разбивает эту плоскость на две области* (внутреннюю и внешнюю). Поясним смысл этой теоремы. Возьмем две точки, p и q , не лежащие на простой замкнутой линии l . Если p и q можно соединить ломаной, не пересекающей l , то считают, что точки p и q лежат в одной и той же области относительно линии l . Если же любая ломаная, соединяющая p и q , пересекает l , то считают, что p и q лежат в разных областях. Теорема Жордана утверждает, что линия l определяет на плоскости две области. Кажущаяся «очевидность» теоремы Жордана объясняется лишь тем, что мы имеем в виду очень простые линии (окружность, контур выпуклого многоугольника и т. ц.).

Пример 14. На рис. 31 изображена простая замкнутая ломаная.

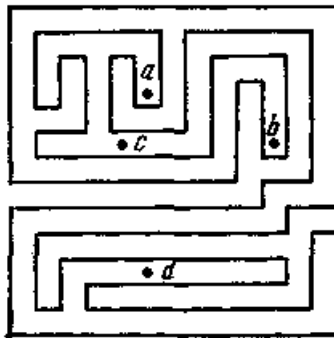


Рис. 31.

Однако вовсе не «очевидно», что она разрезает плоскость на две области; не сразу можно понять, в какой области (внутренней или внешней) лежат точки a, b, c, d .

Приведем доказательство теоремы Жордана. При этом ограничимся случаем, когда l — не произвольная простая замкнутая линия на плоскости, а простая замкнутая ломаная.

Пусть b_1, b_2, \dots, b_n — последовательные звенья ломаной l . Возьмем точки p, p' , симметричные относительно звена b_1 . Через точку p проведем отрезок, параллельный звену b_1 , до точки пересечения с биссектрисой угла между звеньями b_1 и b_2 (рис. 32).

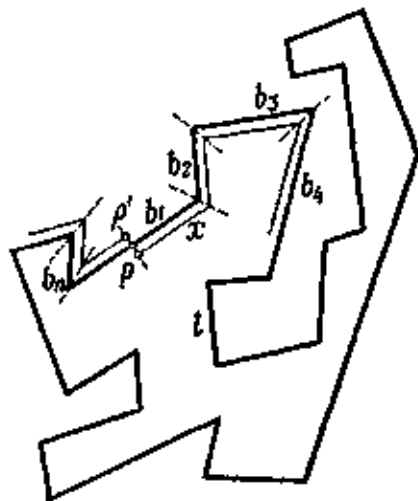


Рис. 32.

Из этой точки проведем отрезок, параллельный b_2 , до пересечения с биссектрисой угла между b_2 и b_3 и т. д. В результате мы получим ломаную x , звенья которой находятся на одном и том же расстоянии от соответствующих звеньев ломаной l . Если при этом расстояние $|pp'|$ достаточно мало, то линия x не пересекает линии l и, обойдя вокруг нее, должна вернуться либо в точку p , либо в p' . Однако в точку p' ломаная x прийти не может: если бы она соединяла точки p и p' , то, присоединив к x отрезок pp' , мы получили бы цикл, который с циклом l пересекается в единственной точке, т. е. индекс пересечения этих двух циклов был бы равен 1, что невозможно. Итак, x представляет собой замкнутую ломаную, один раз обходящую ломаную l . Аналогично получается ломаная x' , выходящая из p' один раз обходящая l и возвращающаяся в p' .

Пусть теперь c — произвольная точка, не лежащая на линии l . Тогда ее можно соединить, не пересекая l , либо с p , либо с p' мы можем провести из точки c луч, пересекающий линии x , x' , и от точки c пройти до первой точки пересечения этого луча с какой-либо из линий x , x' , а затем по этой линии дойти до точки p или p' .

Нетрудно понять, что если из точки c проведены две различные ломаные y , z , не пересекающие l и оканчивающиеся в p или p' , то обе они оканчиваются в одной и той же точке. В самом деле, если бы они оканчивались в разных точках (рис. 33), то ломаная $y \cup z$ вместе с

отрезком pp' составляла бы цикл, индекс пересечения которого с циклом l равен 1, что невозможно.

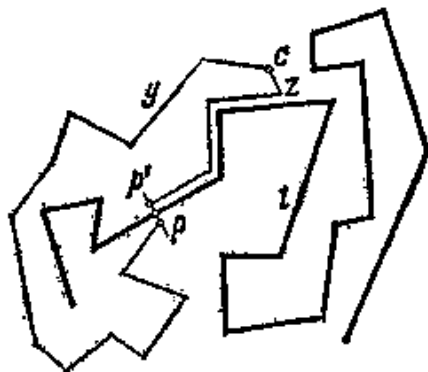


Рис. 33

Обозначим теперь через U множество всех точек плоскости, которые можно, не пересекая l , соединить с точкой p , а через V — множество точек, которые можно, не пересекая l , соединить с p' . Тогда U , V и будут теми двумя областями, на которые, как утверждает теорема Жордана, линия l разбивает плоскость. В самом деле, если точки c_1 , c_2 принадлежат одной области (скажем, U), то существуют ломаные y_1 , y_2 , не пересекающие l , которые соединяют c_1 , c_2 с точкой p . Объединение их представляет собой ломаную, соединяющую c_1 и c_2 и не пересекающую l . Итак, две точки, принадлежащие одной области, можно соединить ломаной, не пересекающей l . Если же точки c_1 , c_2 принадлежат различным областям, то их нельзя соединить ломаной, не пересекающей l (иначе, как и выше, мы получили бы цикл, имеющий с l индекс пересечения 1).

Заметим, что все «далекие» точки плоскости расположены в одной и той же области относительно линии l . Поэтому одна из двух областей, определяемых линией l , — неограниченная, а другая — ограниченная. Неограниченную область называют *внешней*, а ограниченную — *внутренней*.

Задачи

46. Если ломаная l сложная (см. рис. 31), то трудно определить «на глаз», во внутренней или внешней области лежит точка c (т. е. можно ли, отправляясь из c , выйти из «лабиринта», образованного линией l). Докажите, что если луч, исходящий из точки c и не проходящий через

вершины ломаной l , пересекает в четном числе точек, то s лежит во внешней области, а если в нечетном, то во внутренней.

47. Докажите, что всякая простая замкнутая линия на сфере разбивает сферу на две области.

48. На плоскости проведены k ломаных линий, каждая из которых соединяет две заданные точки p и q . Докажите, что если других общих точек ломаные попарно не имеют, то плоскость разбита на k областей.

Укажем (без доказательства), что *любые две простые замкнутые линии l_1, l_2 на плоскости изотопны между собой*, т. е. существует гомеоморфное отображение плоскости на себя, которое переводит l_1 в l_2 . Это предложение отличается от теоремы Жордана, оно утверждает нечто большее. В самом деле, пусть l_1 — окружность, а l_2 — произвольная простая замкнутая линия на плоскости. Гомеоморфное отображение плоскости на себя, переводящее l_1 в l_2 , переводит внешнюю область окружности l_1 во внешнюю область линии l_2 , а внутреннюю область окружности l_1 (т. е. открытый круг) — во внутреннюю область линии l_2 . Таким образом, *объединение простой замкнутой линии l_2 и ее внутренней области гомеоморфно кругу*. Теорема Жордана об этом ничего не говорит, утверждая лишь существование двух областей, внутренней и внешней.

2.1.7. Что такое линия?

Евклид определяет линию как «длину без ширины». Это, конечно, не определение, а лишь наглядное описание линий. Следующий пример показывает, однако, что это описание вряд ли можно считать удачным.

Пример 15. Возьмем квадрат площади 1 (рис. 34, *а*) и выбросим из него крест (рис. 34, *б*), причем ширину полосок креста подберем так, чтобы площадь креста была равна $1/4$.

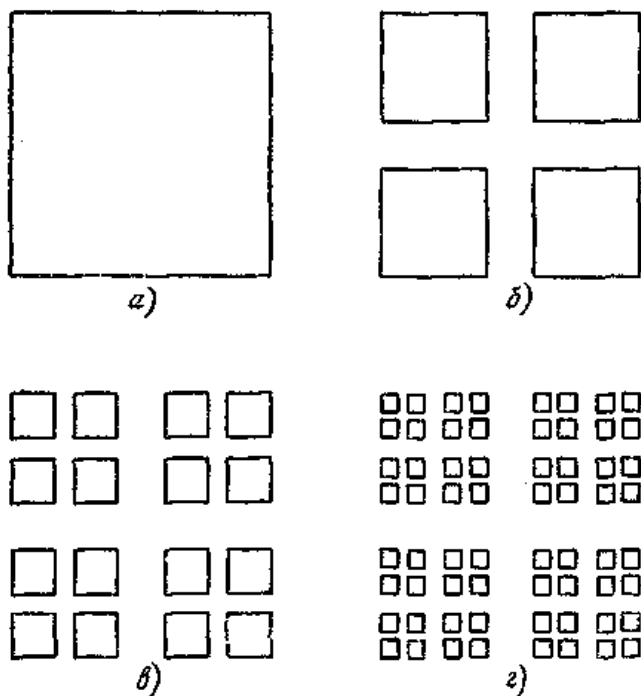


Рис. 34.

В каждом из оставшихся квадратов снова вырежем по кресту (рис. 34, в), причем так, чтобы сумма площадей крестов была равна $1/8$. В каждом из оставшихся 16 маленьких квадратов вновь выбросим по кресту (рис. 34, г) так, чтобы сумма площадей выбрасываемых кусков была, равна $1/16$, и т. д. Обозначим через A «предельную фигуру», т. е. пересечение $A_1 \cap A_2 \cap \dots \cap A_n \cap \dots$, где A_n — фигура, которая остается после проведения n этапов построения. Фигура A как бы «рассыпается» на отдельные точки (ибо остающиеся квадратики делаются все меньше) и тем не менее имеет положительную площадь. В самом деле, сначала мы выбросили из квадрата $1/4$ его площади, затем $1/8$, затем $1/16$ и т. д. В пределе у нас остается фигура A , имеющая площадь $1 - \left(\frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots \right)$. Так как сумма бесконечной убывающей геометрической прогрессии, записанной в скобках, равна $1/2$, то площадь предельной фигуры A равна $1/2$.

Построим теперь *простую дугу* (т. е. фигуру, гомеоморфную отрезку), которая проходит через все точки множества A . Для этого возьмем изогнутую полосу, содержащую четыре квадрата, полученных на первом этапе построения (рис. 35, а).

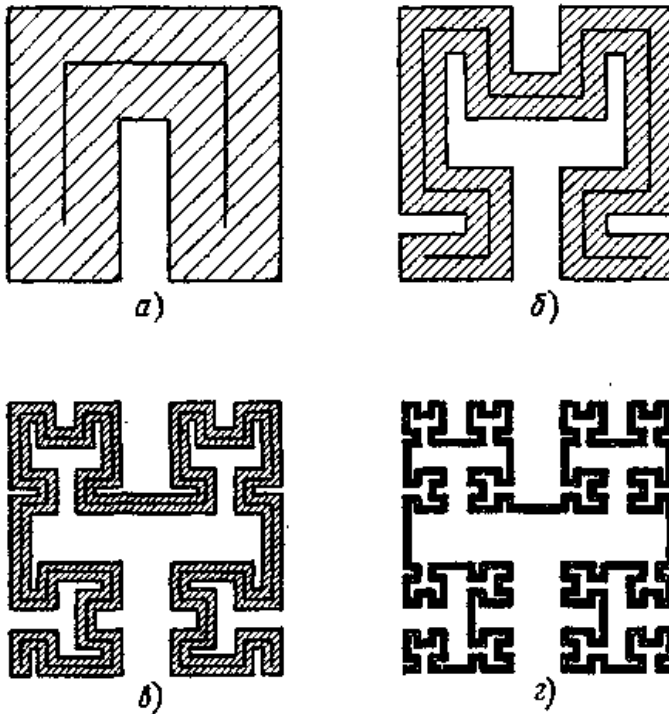


Рис. 35.

Затем сделаем полосу более узкой и изогнутой, так что она будет содержать все квадраты, полученные на втором этапе (рис. 35, б), затем на третьем (рис. 35, в) и т. д.

После n этапов этого построения мы получаем полосу B_n , которая содержится в предыдущих полосках и содержит фигуру A_n (а следовательно, и фигуру A). Пересечение $B_1 \cap B_2 \cap \dots$ этих полосок, т. е. их «предельную фигуру», обозначим через B ; она также содержит A , и потому площадь фигуры B не меньше $1/2$. Рис. 35 наглядно показывает, что B является чрезвычайно «извилистой» линией (простой дугой). Эта линия имеет положительную площадь, т. е. вряд ли может быть названа «длиной без ширины».

Евклид дает также описание линии как «границы поверхности». Однако и понятие «граница», как мы сейчас увидим, таит в себе много неожиданного. Мы привыкли считать, что к каждому участку линии плоскость примыкает «с двух сторон». Например, если l — простая замкнутая линия, то обе области, U и V , определяемые линией l , примыкают к ней на всем ее протяжении (т. е. как угодно близко к любой точке $x \in l$ имеются и точки области U и точки области V).

Кажется «наглядно очевидным», что линия не может быть совместной границей более чем двух областей на плоскости, которые примыкают к этой линии на всем ее протяжении. Однако здесь интуиция нас обманывает.

Пример 16. Покажем, что на плоскости существует линия, являющаяся *совместной границей трех областей*. Такие линии обнаружил японский математик Вада.

Предположим, что имеется окруженная морем земля и на ней два озера: теплое и холодное. Для подведения воды от озер и моря к суше проводятся каналы. В первый день от теплого озера отводится канал (не сообщающийся с морской водой и водой холодного озера) так, чтобы не далее чем на расстоянии l от каждой точки суши была вода теплого озера (рис. 36).

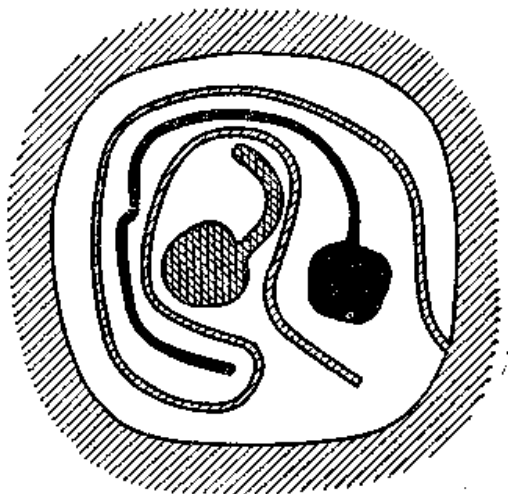


Рис. 36

Во второй день канал отводится от холодного озера, причем он нигде не сообщается с морем, теплым озером и построенным на день раньше каналом, и работа продолжается до тех пор, пока от каждой точки

оставшейся суши не далее чем на расстоянии 1 будет вода холодного озера. В третий день канал таким же порядком отводится от моря. В следующие три дня каналы продолжают далее, причем так, чтобы на расстоянии, меньшем $1/2$, от каждой точки оставшейся суши была вода обоих озер и морская вода. В следующие за ними три дня густота сети каналов увеличивается, так что любая вода будет не далее чем на $1/4$ от каждой точки оставшейся суши, и т. д. Заметим, что после каждого дня работы оставшаяся суша будет связным куском, так что мы можем покрывать ее на следующий день еще более плотной сетью каналов.

В пределе мы получим сеть теплой, холодной и морской вод, которые нигде вместе не сливаются. То, что останется от суши, будет уже «линией», причем как угодно близко от любой точки этой линии будет холодная, теплая и морская вода. Иначе говоря, на всем протяжении этой линии к ней будут «примыкать» три области: море с его каналами, холодное озеро с его каналами и теплое озеро с его каналами.

Евклид дает еще и третье описание линии: «поверхность имеет два измерения, линия имеет одно измерение, точка не имеет ни одного измерения». Определить, что такое размерность (число измерений) фигуры, пытались многие математики. Окончательное выяснение смысла этого понятия и создание теории размерности является заслугой замечательного советского математика П. С. Урысона, безвременно погибшего в возрасте 23 лет в 1924 году.

Говорят, что множество A , расположенное в фигуре X , отделяет точку a от точки b , если не существует в фигуре X связного множества, которое содержит точки a и b и не пересекается с A . Например, поверхность шара (сфера) отделяет в пространстве внутренние точки шара от внешних (рис. 37, а).

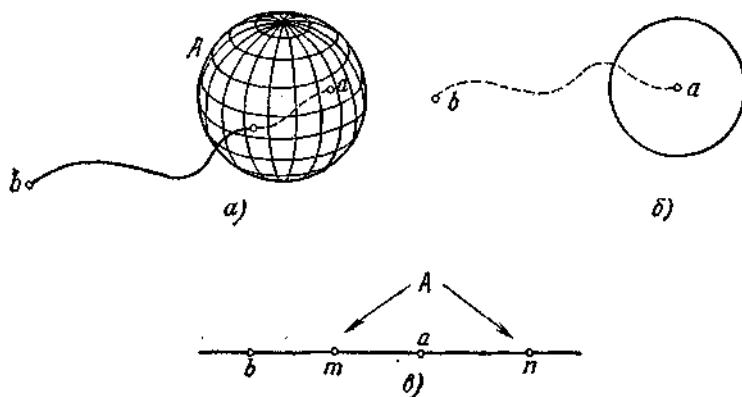


Рис. 37.

Таким образом, в трехмерном пространстве отделение точек можно производить с помощью двумерных фигур. На плоскости (которая представляет собой двумерную фигуру) точку вместе с близкими к ней точками можно от остальных точек отделить с помощью одномерной фигуры (т. е. линии, рис. 37, б). Наконец, на прямой (т. е. одномерной фигуре) точку a вместе с близкими к ней точками можно от остальных точек прямой отделить с помощью фигуры A , состоящей из двух точек m, n (см. рис. 37, в), т. е. с помощью нульмерной фигуры.

Итак, в фигуре, имеющей n измерений (или, как говорят, в n -мерной фигуре), отделение точки вместе с близкими к ней точками от остальной части фигуры можно производить с помощью фигур, имеющих на одно измерение меньше, чем вся фигура. Возникает мысль дать определение нульмерных фигур, через них определить одномерные фигуры (т. е. линии), затем с помощью одномерных определить двумерные фигуры и т. д.

Будем говорить⁴ что *фигура X нульмерна, если в ней не существует никакой связной фигуры, содержащей более одной точки.* Например, фигура, состоящая из конечного числа точек, нульмерна. Фигура A в примере 15 также нульмерна.

Если определено уже, какие фигуры считаются $(n - 1)$ -мерными, то n -мерная фигура определяется как фигура, не являющаяся $(n - 1)$ -мерной, в которой любую точку вместе с близкими к ней точками можно отделить от остальной части фигуры с помощью множества размерности $n - 1$ (или меньше). Это и есть урысоновское определение размерности.

Пример 17. Любой граф является одномерной фигурой, т. е. линией. Действительно, точку a вместе с близкими к ней точками можно отделить от остальной части графа конечным (т. е. нульмерным) множеством, отделяющее множество содержит две точки, если a — внутренняя точка ребра (a_1 на рис. 38), и k точек, если a — вершина индекса k (a_2 на рис.38).

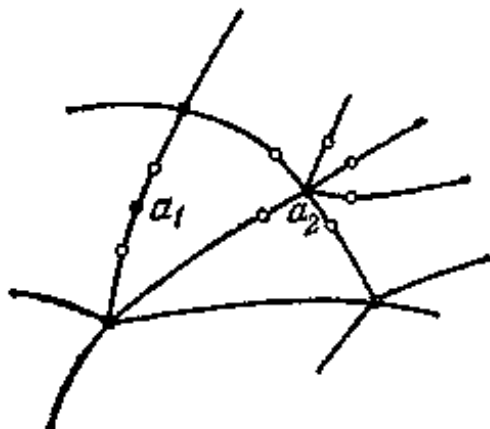


Рис. 38.

Пример 18. Интересный пример линии был построен польским математиком Серпинским. Разделим квадрат на девять квадратов и выбросим средний из них (рис. 39, а).

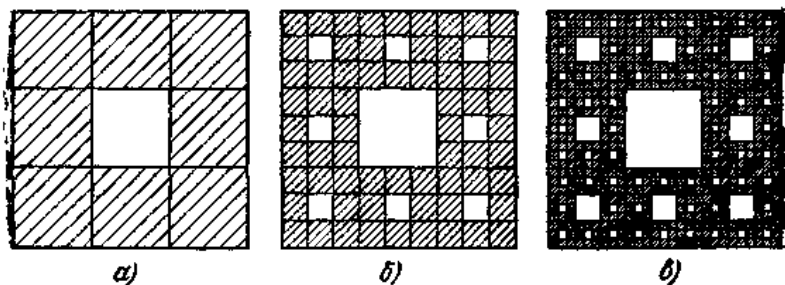


Рис. 39.

Каждый из восьми оставшихся квадратов снова разделим на девять квадратов и выбросим средний (рис. 39, б). Затем так же поступим с каждым из оставшихся квадратов (рис. 39, в) и т. д. В пределе мы получим некоторую одномерную фигуру S , т. е. линию («ковёр Серпинского»). Фигура S является универсальной плоской линией: *если линия l вложима в плоскость, то она вложима в ковер Серпинского, т. е. существует линия $l' \subset S$, гомеоморфная l* . Ясно, что линии, не вложимые в плоскость, не могут быть вложены и в ковер Серпинского. Однако существует в пространстве линия (аналог ковра

Серпинского, рис. 40), в которую, как доказал австрийский математик Менгер, можно вложить любую линию.

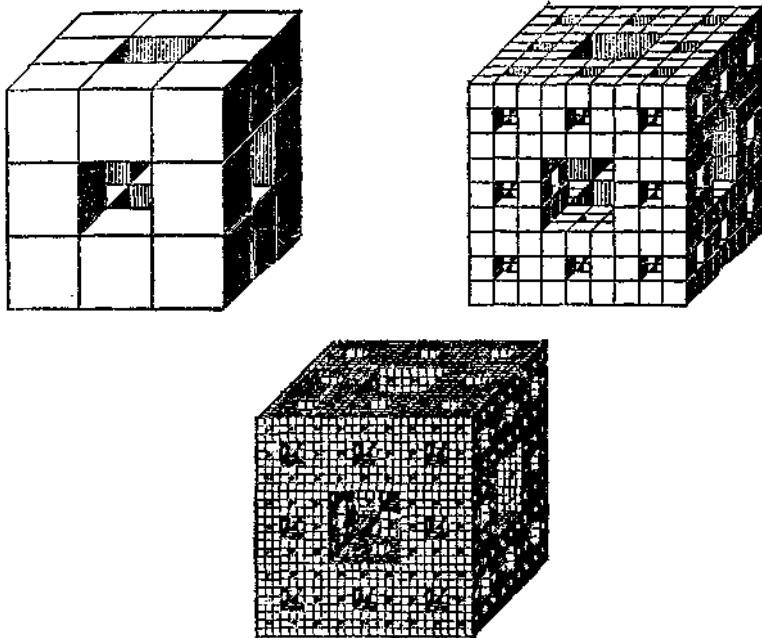


Рис. 40

Задача

49. Существует ли на плоскости линия, являющаяся совместной границей двадцати областей?

50. Докажите, что диагональ квадрата, в котором построен ковер Серпинского S , пересекает S по нульмерному множеству. Выведите отсюда, что ковер Серпинского является одномерной фигурой, т. е. линией.

51. Докажите, что свойство фигуры быть линией, является топологическим инвариантом.

2.1.8. Кривая Пеано

Часто дают еще одно наглядное описание «линия есть след движущейся точки».

Пример 19. Пусть движущаяся точка пробегает фигуру буквы Φ двумя способами, показанными на рис. 41 (сплошной линией указан путь, пройденный в некоторый момент, а штриховой — дальнейшее движение).

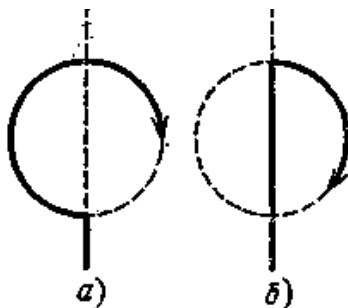


Рис. 41

В обоих случаях точка пробегает одно и то же множество, т. е. «след» движущейся точки одинаков, но пути различны.

Дадим точное определение понятия пути. Пусть в некоторой фигуре A движется точка, начиная от момента $t = 0$ до момента $t = 1$. Для каждого момента t , где $0 \leq t \leq 1$, известно положение $a(t)$ движущейся точки, т. е. каждой точке t отрезка $[0; 1]$ поставлена в соответствие точка $a(t) \in A$. Получается отображение отрезка $[0; 1]$ в фигуру A , причем отображение непрерывное, так как точка $a(t)$ «непрерывно» перемещается с изменением t . Это отображение и представляет собой путь. Мы приходим к следующему определению: *всякое непрерывное отображение отрезка $[0; 1]$ в фигуру A называется путем* (в этой фигуре).

Любую простую дугу можно представлять себе как путь (ведь простая дуга получается с помощью гомеоморфного отображения отрезка, а гомеоморфное отображение непрерывно). В частности, линию, рассмотренную в примере 15 (имеющую «площадь»), можно рассматривать как «след движущейся точки». Уже это показывает, что понятие пути является не слишком простым. Следующий пример еще более подтверждает это.

Пример 20. Покажем, что можно построить путь, который проходит через каждую точку квадрата. Иными словами, существует непрерывное отображение отрезка на весь квадрат; такие пути называются *кривыми Пеано*. Для получения кривой Пеано построим в квадрате Q все более извивающиеся «полоски-лабиринты»: будем делить квадрат на $4, 16, 64, \dots, 4^n, \dots$ конгруэнтных квадратиков (рис.

42), а затем уберем некоторые из их сторон (рис. 43), причем перегородки, оставленные на каком-то этапе построения, сохраняются и на всех последующих. Средние линии этих полосок (штриховая линия на рис. 43) и дадут в пределе *путь, заполняющий весь квадрат Q*, т. е. кривую Пеано.

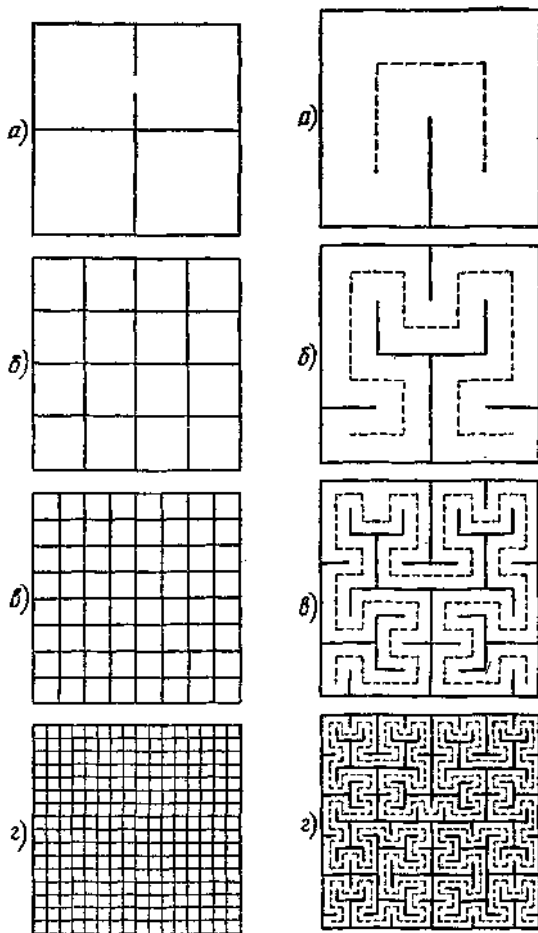


Рис. 42.

Рис. 43.

Более точно этот путь можно определить следующим образом. Рассмотрим непрерывное отображение отрезка $[0; 1]$ на первую штриховую ломаную линию (рис. 43, а), при котором отрезок $[0; 1/4]$ отображается на часть этой ломаной, лежащую в левой нижней

четверти большого квадрата, отрезок $[1/4, 1/2]$ — на часть, лежащую в левом верхнем квадрате, а отрезки $[1/2, 3/4]$ и $[3/4, 1]$ — на части, лежащие в правых (верхнем в нижнем) квадратах. Это отображение обозначим через $f_1(t)$ (где $0 \leq t \leq 1$). Далее, через $f_2(t)$ обозначим отображение отрезка $[0; 1]$ на вторую штриховую ломаную (рис. 43, б), при котором отрезки $[0, 1/16]$ $[1/16, 2/16], \dots, [15/16, 1]$ отображаются на последовательные части этой ломаной, лежащие в шестнадцати квадратах второго этапа. Аналогично, $f_3(t)$ будет отображением отрезка $[0; 1]$ на пунктирную ломаную третьей этапа (рис. 43, в) и т. д. Предел последовательности функций $f_1(t), f_2(t), f_3(t), \dots$ представляет собой отображена $f: [0; 1] \rightarrow Q$, т. е. некоторый путь в квадрате Q ; это и есть кривая Пеано. Легко пояснить, что этот предел существует. Возьмем, например, точку $1/3 \in [0; 1]$. Так как $1/3$ лежит во второй четверти отрезка $[0; 1]$, т. е. $1/3 \in [1/4, 1/2]$, то точка $f_3(1/3)$ лежит в левом верхнем квадрате на рис. 42, а. Далее, так как $1/3 \in [5/16, 6/16]$, то $f_2(1/3)$ лежит в шестом по порядку квадрате, пробегаемом штриховой ломаной на рис. 43, б (т. е. в левом верхнем квадрате на рис. 42, б). Так как $1/3 \in [21/64, 22/64]$, то $f_3(1/3)$ лежит в 22-м квадрате, пробегаемом штриховой ломаной на рис. 43, в (т. е. в левом верхнем квадрате на рис. 42, в), и т. д. Предел этой последовательности уменьшающихся квадратов (вложенных последовательно один в другой), т. е., в данном случае, левая верхняя вершина квадрата и есть точка $f(1/3)$. Таким же образом определяется точка $f(t)$ для любого $t \in [0; 1]$.

Заметим, что кривая Пеано не является простой дугой: она имеет бесконечно много точек «склеивания» (т. е. в квадрате имеется бесконечно много точек, через которые построенный путь $f(t)$ проходит более, чем один раз).

Задачи

52. Докажите, что в квадрате Q имеются точки, через которые построенная кривая Пеано $f(t)$ проходит четыре раза, но нет точек, через которые она проходит пять раз.

53. Существует ли «пространственная кривая Пеано», т. е. путь в кубе, заполняющий весь этот куб?

54. Расположим в горизонтальной плоскости квадрат Q и рассмотрим путь $f(t)$, представляющий собой кривую Пеано в этом квадрате. Через $g(t)$ обозначим точку в пространстве, расположенную над точкой $f(t)$ на высоте t (рис. 44).

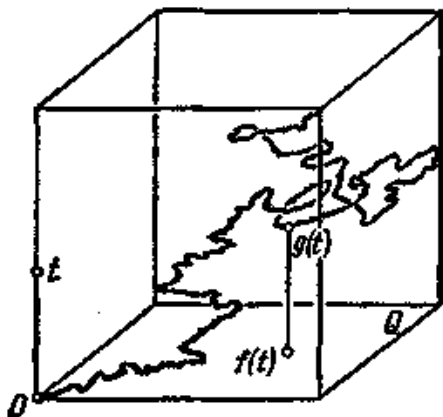


Рис 44.

Докажите, что когда t пробегает отрезок $[0; 1]$, точка $g(t)$ пробегает путь в пространстве, представляющий собой простую дугу. Докажите, что проекция этой простой дуги на горизонтальную плоскость заполняет весь квадрат Q . Иными словами, построенная линия (простая дуга) представляет собой замысловатую «крышу» над всем квадратом Q .

Этот пример показывает, что не только понятие пути, но даже понятие простой дуги не является таким простым, каким оно интуитивно кажется с первого взгляда.

52. Докажите, что в квадрате Q имеются точки, через которые построенная кривая Пеано $f(t)$ проходит четыре раза, но нет точек, через которые она проходит пять раз.

53. Существует ли «пространственная кривая Пеано», т. е. путь в кубе, заполняющий весь этот куб? 54. Расположим в горизонтальной

плоскости квадрат Q и рассмотрим путь $f(t)$, представляющий собой кривую Пеано в этом квадрате. Через $g(t)$ обозначим точку в пространстве, расположенную над точкой $f(t)$ на высоте t (рис. 44). Докажите, что когда пробегает отрезок $[0; 1]$, точка $g(t)$ пробегает путь в пространстве, представляющий собой простую дугу. Докажите, что проекция этой простой дуги на горизонтальную плоскость заполняет весь квадрат Q . Иными словами, построенная линия (простая дуга) представляет собой замысловатую «крышу» над всем квадратом Q .

Этот пример показывает, что не только понятие пути, но даже понятие простой дуги не является таким простым, каким оно интуитивно кажется с первого взгляда.

2.2. Топология поверхностей

2.2.1. Теорема Эйлера

В следующей таблице указано число вершин, ребер и граней пяти правильных многогранников.

Название многогранника	Число вершин	Число ребер	Число граней
Тетраэдр	4	6	4
Куб	8	12	6
Октаэдр	6	12	8
Додекаэдр	20	30	12
Икосаэдр	12	30	20

Из рассмотрения этой таблицы видно, что для каждого правильного многогранника имеет место соотношение:

$$V - P + G = 2, \quad (6)$$

где V — число вершин многогранника, P — число его ребер, G — число граней. Соотношение (6) легко проверяется также для пирамид, призм и других многогранников. Эйлер впервые подметил и доказал это замечательное свойство многогранников.

Уточним формулировку теоремы Эйлера. Прежде всего заметим, что любая грань каждого из рассмотренных многогранников гомеоморфна кругу. Далее, поверхность каждого из рассмотренных многогранников (или, вообще, любого выпуклого многогранника) гомеоморфна сфере: если o — произвольная внутренняя точка многогранника, а S — сфера с центром o , содержащая внутри себя этот многогранник, то проекция поверхности многогранника на сферу S из центра o представляет собой искомым гомеоморфизм. Таким образом, теорема Эйлера в уточненной формулировке принимает следующий вид: *для всякого многогранника, поверхность которого гомеоморфна сфере, а каждая грань гомеоморфна кругу, справедливо соотношение (6).*

Можно придать этой теореме чисто топологическую формулировку. Для этого заметим, что все вершины и ребра многогранника образуют связный граф, который разбивает поверхность многогранника на отдельные грани (т. е. куски, гомеоморфные кругу). Мы получаем следующее (более общее, чем теорема Эйлера) утверждение:

Пусть на сфере (или гомеоморфной ей поверхности) начерчен связный граф G , имеющий V вершин и P ребер и разбивающий сферу на Γ областей («граней»), тогда справедливо соотношение (6). Идея доказательства этой теоремы содержится в задаче 55.

Задачи

55. Пусть G — связный граф, начерченный на сфере, G^* — его максимальное дерево в k — число перемычек (т. е. ребер графа G , не содержащихся в G^*). Докажите, что граф G^* определяет на сфере лишь одну область (грань), и потому для него соотношение (6) справедливо. Докажите, что добавление каждой перемычки увеличивает число граней на одну, и получите отсюда доказательство теоремы Эйлера.

56. Докажите, что для любого связного графа, расположенного на плоскости, справедливо соотношение (6) (к числу «граней» надо причислять и наружную, неограниченную область).

57. Пусть G — граф, вложимый в плоскость. Докажите, что при любом способе его вложения в плоскость он разбивает плоскость на $r = V + P + 1$ областей, где r — число компонент графа G , а V и P — число его вершин и ребер.

58. Выпуклый n -угольник разбит на треугольники, примыкающие друг к другу целыми сторонами (рис. 45), причем на сторонах n -угольника расположены m вершин разбиения, а внутри него p вершин.

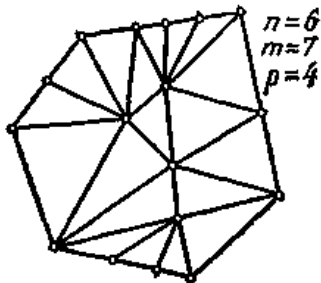


Рис. 45.

Докажите, что n -угольник разбит на $m + n + 2p - 2$ треугольников.

59. Обозначим через n_3 число треугольных граней выпуклого многогранника, через n_4 — число его четырехугольных граней и т. д. Докажите, что

$$3n_3 + 2n_4 + n_5 \geq 12 + n_7 + 2n_8 + 3n_9 + 4n_{10} + \dots$$

В каком случае имеет место равенство?

60. Говорят, что связный граф, расположенный на сфере, определяет *топологически правильное разбиение* сферы, если каждая грань этого разбиения является n -угольником (т. е. ограничена замкнутой цепочкой из n ребер) и в каждой вершине сходятся k граней. Докажите, что в этом случае

$$\frac{1}{n} + \frac{1}{k} = \frac{1}{2} + \frac{1}{P},$$

где P — число ребер. Выведите отсюда, что кроме разбиений, топологически эквивалентных пяти правильным многогранникам, существуют лишь два типа топологически правильных разбиений, которые показаны на рис. 46.

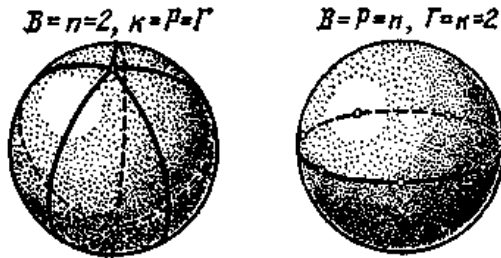


Рис. 46.

2.2.2. Поверхности

Пример 21. На рис. 47 изображена «книжка с тремя листами».

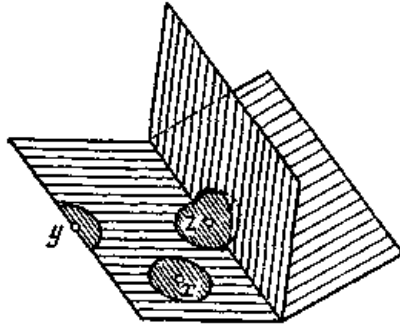


Рис. 47.

Вблизи точек x , y , z эта фигура устроена по-разному. Окрестность точки y имеет вид полукруга, причем точка y лежит на его границе. В этом случае говорят, что точка y лежит на крае фигуры. Окрестность точки z состоит из трех полукругов, соединенных до общего диаметру; говорят, что в этом месте фигура *разветвляется* (т. е. к некоторой линии примыкает три или более «листов» рассматриваемой фигуры). Наконец, точка x имеет окрестность в виде круга, причем точка x лежит внутри этого круга; здесь фигура не имеет ни края, ни разветвления. Фигура, у которой каждая точка x имеет окрестность, гомеоморфную кругу (внутри которого лежит точка x), называется *поверхностью*. Поверхность не имеет краев и разветвлений. Сфера и тор являются поверхностями. Рассматривают также *поверхности с краем*; они имеют края, но не имеют разветвлений. Круг — поверхность с краем. Сфера, в которой вырезаны несколько круглых отверстий (рис. 48), также является поверхностью с краем.

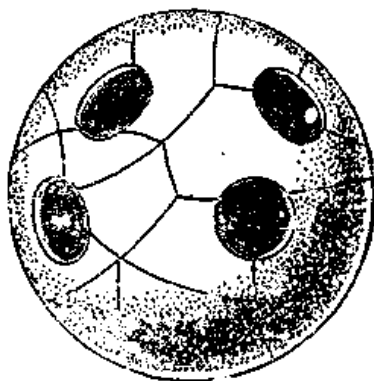


Рис. 48.

Пример 22. Если на торе вырезать круглую дыру, то мы получим поверхность с краем, которая называется *ручкой* (рис. 49).

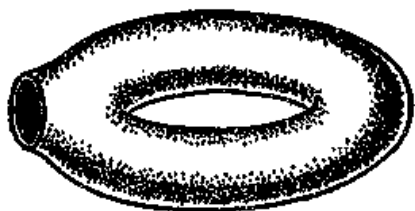


Рис. 49.

Пример 23. Интересный пример поверхности с краем был описан в 1862—1865 годах в работах немецких математиков Мёбиуса и Листинга. Она получается следующим образом. Лента прямоугольной формы (рис. 50, а) один раз перекручивается (рис. 50, б, в) и затем ее концы склеиваются.

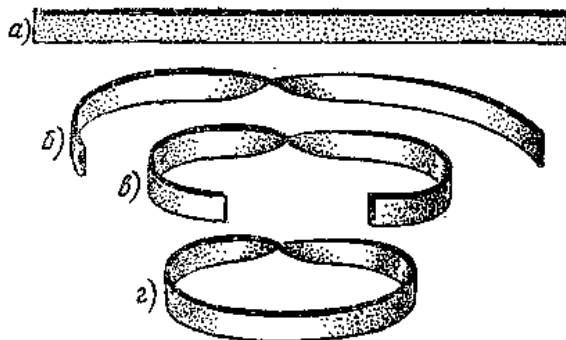


Рис. 50.

Полученная поверхность с краем (рис. 50, в) называется *лентой Мёбиуса*. Эта поверхность имеет лишь одну сторону. Например, перемещая кисточку по ленте Мёбиуса (рис. 51), мы придем к тому же месту, с которого начинали закрашивание, но с обратной стороны.

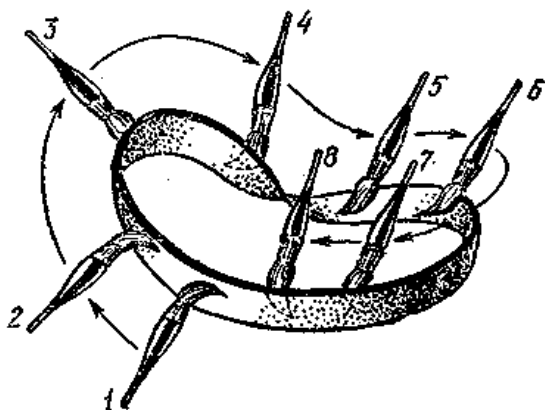


Рис. 51.

Перемещая кисточку дальше, мы закрасим всю ленту Мёбиуса и убедимся, что у нее нет «второй стороны».

Разумеется, наглядное описание односторонней поверхности с помощью «окрашивания» возможно лишь для «толстой поверхности», изготовленной из некоторого материала; математически же поверхность не имеет толщины. Поэтому приведем другое описание «односторонности». В каждой точке a ленты Мёбиуса можно провести два противоположных вектора, перпендикулярных к ней в этой точке (рис. 52, a).

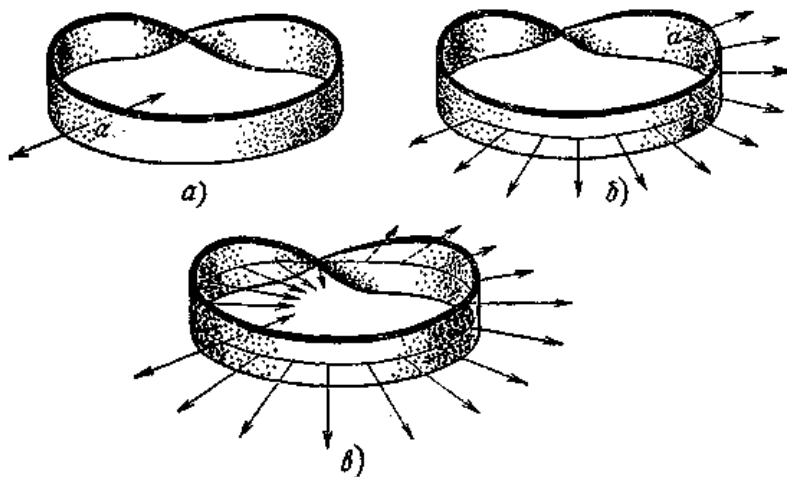


Рис. 52.

Эти векторы называют *нормальми* к ленте Мёбиуса в точке a . Выберем одну из них и начнем перемещать точку a вместе с нормалью по ленте Мёбиуса (рис. 52, б). Когда точка a обойдет всю ленту Мёбиуса, перемещающаяся нормаль перейдет не в свое первоначальное положение, а в противоположное (рис. 52, в). Итак, *на ленте Мёбиуса существует такой замкнутый путь (обход), что при прохождении этого пути нормаль к поверхности приходит в положение, противоположное первоначальному*. Поверхности, обладающие такими обходами, и называются *односторонними*.

Однако, говоря о нормалях, мы изучаем не только саму поверхность, но и ее расположение в пространстве.

Поэтому приведем «внутреннее» определение односторонних поверхностей. Условимся вокруг точки a , из которой проведена нормаль» описывать небольшую окружность и на ней отмечать стрелкой направление, которое из конца проведенной нормали наблюдается как направление против часовой стрелки (рис. 53, а).



Рис. 53.

Если точка a перемещается, то вместе с ней перемещается и нормаль, а также окружность с имеющимся на ней направлением. Когда мы обведем окружность по всей ленте Мёбиуса, направление на окружности изменится на противоположное (так как нормаль изменит свое направление, рис. 53, б).

Итак, на ленте Мёбиуса имеется такой замкнутый путь (обход), что при перемещении окружности вдоль этого пути направление на окружности меняется на противоположное. Такие обходы называются *обращающими ориентацию*.

Если на поверхности нет обращающих ориентацию обходов, то она называется *ориентируемой* (или *двусторонней*, если есть — *неориентируемой* (или *односторонней*). С наглядной точки зрения ориентируемость означает, что всю поверхность можно покрыть маленькими окружностями и выбрать на них такие направления, что близкие окружности будут ориентированы одинаково. Пусть теперь Q_1 и Q_2 — две поверхности, у каждой из которых имеется край, гомеоморфный окружности (рис. 54).

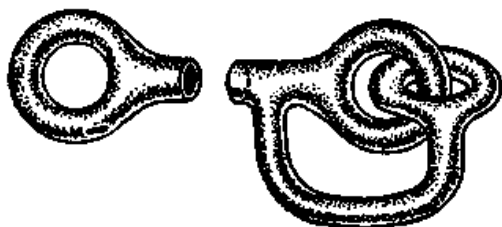


Рис. 54

Соединив («склеив») края этих поверхностей, мы получим одну новую поверхность. Говорят, что дыра, имеющаяся в поверхности Q_1 , заклеивается поверхностью Q_2 (или наоборот).

Пример 24. Рассмотрим сферу, в которой вырезано p круглых дыр, и заклеим каждую из дыр ручкой. Полученная поверхность (рис. 55, а) называется *сферой с p ручками*.

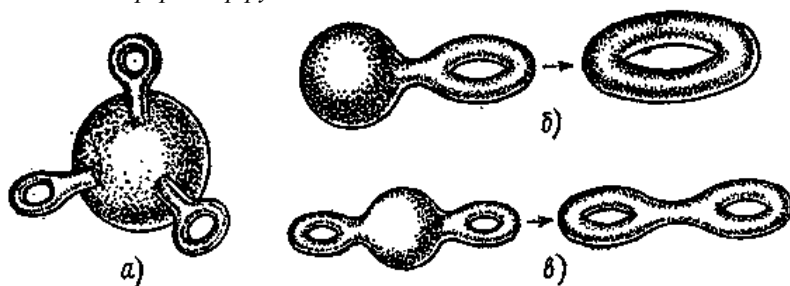


Рис. 55.

Сфера с одной ручкой гомеоморфна тору (рис. 55, б), а сфера с двумя ручками — поверхности «кренделя» (получающейся склеиванием двух ручек, рис. 55, в).

Задачи

- 61. Докажите, что граф «домики и колодцы» (пример 12) можно расположить (без самопересечений) на ленте Мебиуса.
- 62. У «зубчатой» фигуры, изображенной на рис. 56, а, склеиваются с перекручиванием каждые два отрезка, обозначенные одинаково (рис. 56, б).

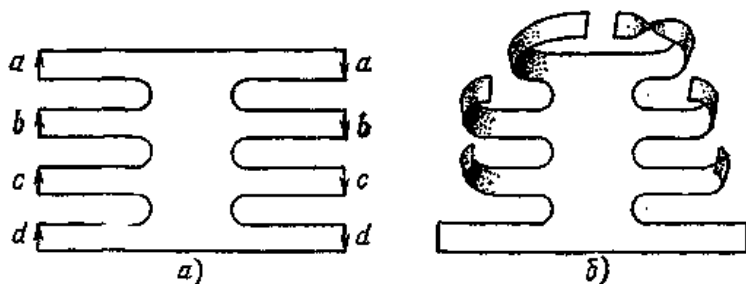


Рис. 56,

Докажите, что получающаяся поверхность является односторонней, а ее край гомеоморфен окружности.

63. В шаре высверлены три сквозных цилиндрических отверстия, не соединяющихся между собой. Докажите, что поверхность получившегося тела гомеоморфна сфере с тремя ручками.

64. В шаре высверлены три сквозных цилиндрических отверстия, оси которых проходят через центр шара. Докажите, что поверхность получившегося тела гомеоморфна сфере с пятью ручками.

65. Если попарно склеить противоположные стороны квадрата с учетом указанных на рис. 57, а направлений, то получится тор (рис. 57, б, в, г).

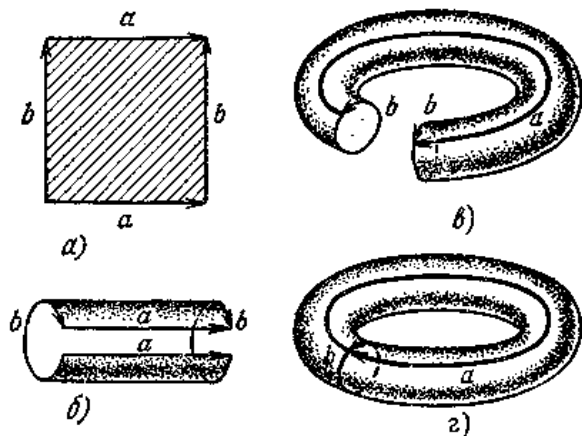


Рис. 57.

Какая поверхность получится, если склеивание произвести с учетом направлений на рис. 58 (сторона *c* остается не склеенной)?

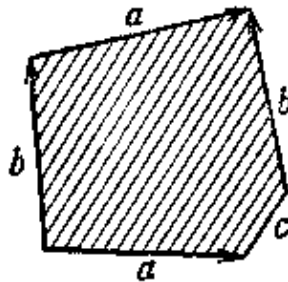


Рис. 58.

66. Какая поверхность получится, если в $4k$ -угольнике, показанном на рис. 59, попарно склеить одинаково обозначенные стороны с учетом направлений?

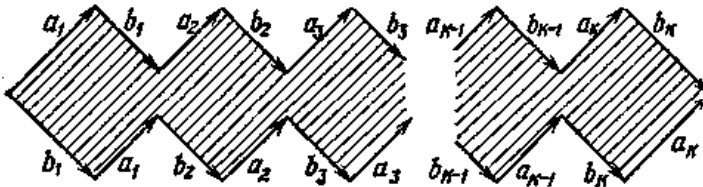


Рис. 59.

Мы подходим к формулировке замечательной теоремы о *топологической классификации поверхностей*, полученной в прошлом столетии Мёбиусом и французским математиком Жорданом. Условимся рассматривать только *замкнутые поверхности* (которые не имеют края и допускают разбиение на конечное число многоугольников). Плоскость, например, не является замкнутой поверхностью: конечный граф, начерченный на плоскости не разбивает ее на области, которые все гомеоморфны кругу. Задача топологической классификации поверхностей заключается в том, чтобы *указать такие попарно не гомеоморфные замкнутые поверхности, что любая замкнутая поверхность гомеоморфна одной из них*. Иначе говоря, нужно перечислить *все топологически различные замкнутые поверхности*.

Решение этой задачи рассмотрим сначала для *ориентируемых* поверхностей. Обозначим через P_0 сферу, а через P_k — сферу с k ручками. Оказывается, что *поверхности*

$$P_0, P_1, P_2, \dots, P_k, \dots \quad (7)$$

и дают полную топологическую классификацию замкнутых ориентируемых поверхностей, т. е. здесь перечислены все топологически различные типы таких поверхностей. Доказательство будет дано в следующих двух пунктах.

2.2.3. Эйлерова характеристика поверхности

Пусть Q — поверхность (с краем или без края, двусторонняя или односторонняя), которая допускает разбиение на многоугольники; это означает, что на поверхности можно «нарисовать» граф, разбивающий ее на конечное число кусков, гомеоморфных кругу. Обозначим число вершин и ребер графа через V и P , а число многоугольников, на которые Q разбивается этим графом, — через Γ . Число

$$\chi(Q) = V - P + \Gamma \quad (8)$$

называется *эйлеровой характеристикой* поверхности Q . Строго говоря, число (8) определяется не самой поверхностью Q , а выбором ее разбиения на многоугольники. Однако теорема Эйлера показывает, что для поверхности Q , гомеоморфной сфере, эйлерова характеристика не зависит от выбора разбиения на многоугольники: $\chi(Q) = 2$ (см. (6)). Мы докажем, что и для любой поверхности Q ее эйлерова характеристика $\chi(Q)$ не зависит от выбора разбиения на многоугольники, а определяется самой поверхностью, является ее топологическим инвариантом.

В самом деле, пусть на поверхности Q «нарисованы» два графа G_1, G_2 , каждый из которых задает разбиение на многоугольники. Числа вершин, ребер и граней разбиения, определяемого графом G_1 , обозначим через V_1, P_1, Γ_1 , а соответствующие числа для разбиения, определяемого графом G_2 , — через V_2, P_2, Γ_2 . Вообще говоря, графы G_1 и G_2 могут пересекаться в бесконечном числе точек. Однако, «пошевелив» граф G_1 , мы сможем добиться того, чтобы G_1 и G_2 пересекались лишь в конечном числе точек.

Далее, если граф $G_1 \cup G_2$ несвязен, то, «пошевелив» графы G_1, G_2 , можно добиться того, чтобы они имели общие точки и, следовательно, их объединение было связным. Итак, мы можем предполагать, что графы G_1 и G_2 пересекаются лишь в конечном числе точек и имеют связное объединение $G_1 \cup G_2$. Считая новыми вершинами все точки пересечения графов G_1 и G_2 , а также все вершины этих графов, мы найдем, что $G_1 \cup G_2$ является конечным связным графом (его ребрами являются куски ребер графов G_1 и G_2 , на которые они разбиваются вершинами графа $G_1 \cup G_2$).

Обозначим через V и P число вершин и ребер графа $G_1 \cup G_2$, а через Γ — число граней, на которые он разбивает поверхность Q . Идея состоит в том, чтобы доказать равенства

$$\left. \begin{aligned} V_1 - P_1 + \Gamma_1 &= V - P + \Gamma, \\ V_2 - P_2 + \Gamma_2 &= V - P + \Gamma, \end{aligned} \right\} \quad (9)$$

из которых и будет следовать, что $V_1 - P_1 + \Gamma_1 = V_2 - P_2 + \Gamma_2$. Оба равенства (9) доказываются одинаково; докажем первое.

Пусть M — некоторый многоугольник («грань»), определяемый графом G_1 . Обозначим число вершин и ребер графа $G_1 \cup G_2$, расположенных внутри M (не на контуре), через V' и P' , а число вершин (а значит, и ребер) этого графа, расположенных на контуре многоугольника M , через q . Далее, число граней, определяемых графом $G_1 \cup G_2$ и содержащихся в M , обозначим через Γ' . На рис. 60 имеем $V' = 4$, $P' = 12$, $\Gamma' = 9$, $q = 15$.

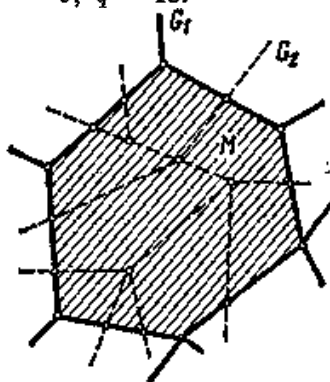


Рис. 60.

Вырежем теперь многоугольник M (вместе с имеющейся на нем частью графа $G_1 \cup G_2$) из поверхности Q . Так как M гомеоморфен кругу и, значит, полусфере, то его можно второй («нижней») полусферой дополнить до поверхности, гомеоморфной сфере (рис. 61).

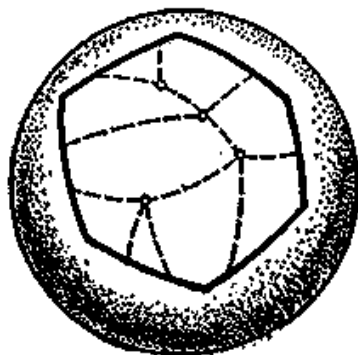


Рис. 61.

На этой сфере расположен связный граф, имеющий $V' + q$ вершин, $P' + q$ ребер и определяющий $\Gamma' + 1$ граней (Γ' граней содержится в M и еще одной гранью является нижняя полусфера). Следовательно, согласно (6), $(V' + q) - (P' + q) + (\Gamma' + 1) = 2$, т. е.

$$V' - P' + \Gamma' = 1. \quad (10)$$

Если теперь (возвращаясь к поверхности Q , на которой начерчен граф $G_1 \cup G_2$) мы выбросим из графа $G_1 \cup G_2$ его часть, расположенную внутри M , то получится новый граф, для которого, однако, число $V - P + \Gamma$ останется таким же, как и для графа $G_1 \cup G_2$. В самом деле, вместо V' вершин, P' ребер и Γ' граней, имевшихся внутри M , мы теперь будем иметь 0 вершин, 0 ребер и одну грань (сам многоугольник M), т. е. число $V' - P' + \Gamma'$ заменится на $0 - 0 + 1$, а это, согласно (10), ничего не меняет.

Теперь ясно, что если мы из графа $G_1 \cup G_2$ выбросим его части, расположенные внутри всех многоугольников, определяемых графом G_1 , то получим новый граф G^* , для которого число $V - P + \Gamma$ будет таким же, как и для графа $G_1 \cup G_2$. Иначе говоря,

$$V^* - P^* + \Gamma^* = V - P + \Gamma, \quad (11)$$

где V^* и P^* — число вершин и ребер графа G^* , а Γ^* — число определяемых им граней.

Заметим, наконец, что граф G^* получается из G_1 добавлением нескольких новых вершин на ребрах. Добавление каждой новой вершины увеличивает число ребер на 1 (поскольку добавленная вершина разбивает одно из ребер на два). Следовательно, если переход от графа G_1 к G^* осуществляется добавлением k новых вершин, то

$V^* = V_1 + k$, $P^* = P_1 + k$. Кроме того, $\Gamma^* = \Gamma_1$ (так как граф G^* определяет те же грани, что и граф G_1). Таким образом,

$$\begin{aligned} V^* - P^* + \Gamma^* &= (V_1 + k) - (P_1 + k) + \Gamma_1 = \\ &= V_1 - P_1 + \Gamma_1, \end{aligned}$$

а это, согласно (11), и дает первое из соотношений (9).

Итак, эйлерова характеристика поверхности не зависит от ее разбиения на многоугольники, а определяется самой поверхностью. Кроме того, эйлерова характеристика является *топологическим инвариантом*: если поверхности Q_1 и Q_2 гомеоморфны, то $\chi(Q_1) = \chi(Q_2)$. В самом деле, при гомеоморфизме $f: Q_1 \rightarrow Q_2$ граф G_1 , начерченный на поверхности Q_1 , переходит в граф $G_2 = f(G_1)$, начерченный на Q_2 , причем вершин, ребер и граней на поверхности Q_2 будет столько же, сколько и на поверхности Q_1 .

Задачи

67. Докажите, что сфера с вырезанными в ней q дырами имеет эйлерову характеристику $2 - q$.

68. Пусть Q_1 и Q_2 — две поверхности, каждая из которых имеет край, гомеоморфный окружности. Докажите, что, склеивая эти края (см. рис. 54), мы получим поверхность, эйлерова характеристика которой равна $\chi(Q_1) + \chi(Q_2)$.

69. Чему равна эйлерова характеристика круга? Ручки? Ленты Мёбиуса?

70. Докажите, что эйлерова характеристика поверхности P_k равна $2 - 2k$.

71. На торе осуществлено топологически правильное разбиение (см. задачу 60). Докажите, что каждая грань является либо треугольником, либо четырехугольником, либо шестиугольником. Приведите примеры правильных разбиений каждого из этих типов.

72. На замкнутой поверхности Q нарисован граф с V вершинами и P ребрами. Он разбивает поверхность Q на Γ областей (среди которых, возможно, есть не гомеоморфные кругу). Докажите, что

$$V - P + \Gamma \geq \chi(Q).$$

Указание. Для того чтобы разрезать поверхность Q на многоугольники (гомеоморфные кругу), достаточно последовательно применить одну или несколько операций следующего вида: а) добавление новой вершины на одном из ребер графа; б) добавление ребра, имеющего лишь одну общую вершину с начерченным графом; в) добавление ребра, соединяющего две вершины уже начерченного графа. Проверьте, что при каждой из этих операций число $V - P + \Gamma$ может лишь уменьшиться.

73. На замкнутой поверхности Q начерчен граф с V вершинами и P ребрами; поверхность Q разбивается этим графом на Γ областей. Докажите, что если каждая из областей имеет на своей границе не менее k ребер, то $(k - 2) P \leq kV - k\chi(Q)$.

2.2.4. Классификация замкнутых ориентируемых поверхностей

Поверхности (7) попарно не гомеоморфны, так как имеют разные эйлеровы характеристики (задача 70). Таким образом, для доказательства теоремы, сформулированной ранее, остается установить, что любая замкнутая ориентируемая поверхность гомеоморфна одной из поверхностей (7). Доказательство проведем в несколько этапов.

А) Пусть Q — некоторая связная замкнутая ориентируемая поверхность. Начертим на ней связный граф G , разбивающий ее на грани, гомеоморфные кругу. Для каждой вершины графа G возьмем на поверхности Q маленький кружок, содержащий внутри себя эту вершину. Эти кружки будем называть *шапочками*. Далее, для каждого ребра графа G возьмем узкую *полосу*, идущую вдоль этого ребра и соединяющую шапочки, которые соответствуют концам взятого ребра. Если удалить из поверхности Q все шапочки и все полосы, то от каждой грани останется кусок, гомеоморфный кругу; этот кусок будем называть *сердцевиной* грани. На рис. 62, на котором изображен кусок поверхности Q , шапочки заштрихованы, полосы иокрыты точками, а сердцевины оставлены белыми.

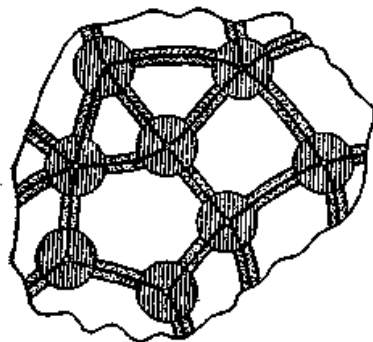


Рис. 62.

Идея заключается в том, чтобы разрезать поверхность Q на шапочки, полоски, сердцевинки, а затем снова склеить ее из этих кусков, прослеживая шаг за шагом, что получается при склеивании.

Прежде всего вырежем из поверхности Q все сердцевинки граней; оставшуюся часть поверхности обозначим через Q_0 . Ее край состоит из всех контуров сердцевин.

Б) Выделим в графе G максимальное дерево (на рис. 63 оно вычерчено жирно), и все полоски, соответствующие перемычкам (т. е. ребрам графа G , не входящим в это дерево), рассечем в середине.

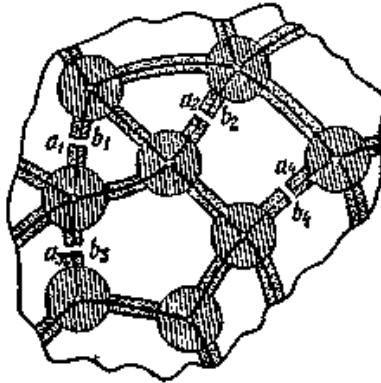


Рис. 63.

Отрезки $a_1 b_1, a_2 b_2, \dots, a_p b_p$, по которым производится рассечение полосок, назовем *хордами*. Рассечение будем производить постепенно: рассечение по хорде $a_1 b_1$ превращает Q_0 в поверхность Q_1 , если в Q_1 произвести рассечение по хорде $a_2 b_2$, получим поверхность Q_2 ; ...; наконец, рассекая Q_{p-1} по хорде $a_p b_p$, получим поверхность Q_p . Для получения из Q_p поверхности Q_0 нужно вновь произвести склеивание по хордам.

В) Прежде чем осуществлять эти обратные склеивания, заметим, что поверхность Q_p гомеоморфна кругу. В самом деле, будем вычерчивать максимальное дерево графа G , беря одно ребро, еще одно, еще одно и т. д.— так, чтобы все время получалось дерево. Полоска и две шапочки, соответствующие первому ребру и его концам, составляют поверхность, гомеоморфную кругу (рис. 64, а).

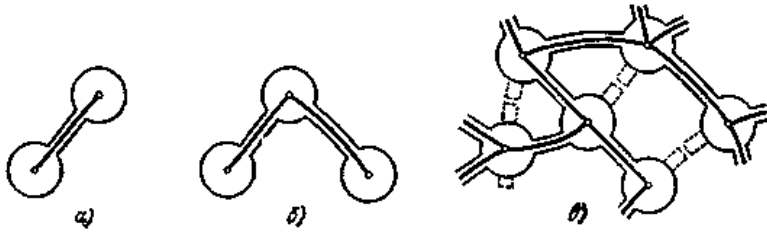


Рис. 64.

Добавление полоски и шапочки, соответствующих второму ребру, снова дает поверхность, гомеоморфную кругу (рис. 64, б).

Вообще, при каждом проведении ребра, к уже имевшейся поверхности, гомеоморфной кругу, приклеиваются одна полоска и одна шапочка, что вновь дает поверхность, гомеоморфную кругу. В конце концов, вычертив максимальное дерево графа G , мы получим гомеоморфную кругу поверхность, составленную из всех шапочек и тех полосок, которые соответствуют ребрам максимального дерева. Для получения поверхности Q_p остается приклеить *полуполоски*, образовавшиеся из оставшихся полосок после рассечения по хордам (пунктир на рис. 64, в). Но каждое приклеивание полуполоски оставляет поверхность гомеоморфной кругу.

Мы покажем теперь, что *каждая из поверхностей $Q_p, Q_{p-1}, \dots, Q_1, Q_0$ гомеоморфна сфере, в которой вырезаны несколько дыр и часть из них заклеена ручками*. Относительно поверхности Q_p это очевидно: она гомеоморфна кругу, т. е. сфере, в которой вырезана одна дыра и не вклеено ни одной ручки.

Г) Рассмотрим при каждом $i = 1, \dots, p$ переход от поверхности Q_{i-1} к поверхности Q_i (т. е. рассечение по хорде $a_i b_i$) и обратный переход от Q_i к Q_{i-1} . Здесь могут представиться две возможности: точки a_i и b_i расположены на одной и той же компоненте края поверхности Q_{i-1} или на равных компонентах.

Если a_i и b_i расположены на равных компонентах края поверхности Q_{i-1} , то рассечение по хорде $a_i b_i$ приводит к уменьшению числа дыр на одну (рис. 65, а и б).

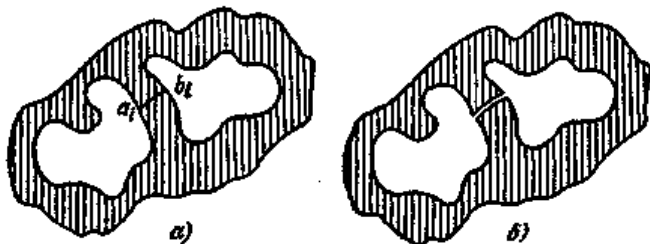


Рис. 65.

Следовательно, обратный переход (от Q_i к Q_{i-1}) сводится к вырезанию одной новой дыры. Поэтому, если Q_i получалась из сферы вырезанием нескольких дыр и заклеиванием части из них ручками, то это же справедливо и для Q_{i-1} .

Д) Пусть теперь концы хорды $a_i b_i$ принадлежат одной компоненте края поверхности Q_{i-1} (рис. 66).



Рис. 66.

Получающаяся при рассечении поверхность Q_i (рис. 67) гомеоморфна поверхности (рис. 68), получающейся из Q_{i-1} двумя разрезами: сначала по замкнутой линии l , не пересекающейся с краем поверхности Q_{i-1} (это дает промежуточную поверхность Q_i^* , (рис. 69), а затем по хорде $a_i c_i$, концы которой лежат на разных компонентах края поверхности Q_i^* . Обратный переход от Q_i (см. рис. 68) к Q_i^* (рис. 69), как мы видели в пункте Г), сводится к вырезанию одной дыры. Остается рассмотреть переход от Q_i^* к Q_{i-1} .

Итак, пусть Q_i^* получается из Q_{i-1} разрезанием по нтуру l не пересекающемуся с краем поверхности Q_{i-1} .



Рис. 67.

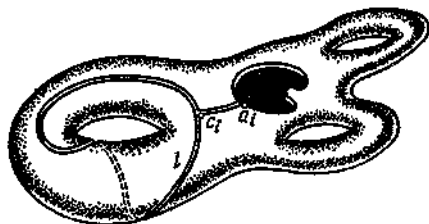


Рис. 68.



Рис. 69.



Рис. 70.

Если вместо того, чтобы производить этот разрез, мы вырежем из Q_{i-1} узкую полосу L , заключающую внутри себя линию l (рис. 70), то получится поверхность, гомеоморфная Q_i^* . Полоска L гомеоморфна либо ленте Мёбиуса, либо боковой поверхности цилиндра. В самом деле, если эту полосу разрезать (рис. 71), то ее можно распрямить в узкую прямоугольную ленту; следовательно, L можно получить склеиванием концов прямоугольной ленты, и надо лишь проследить, происходит это склеивание с перекручиванием или нет.



Рис. 71.

Но ленте Мёбиуса полоска L гомеоморфна быть не может, так как на ленте Мёбиуса имеется обход, обращающий ориентацию, а исходная поверхность Q (и все поверхности Q_0, Q_1, \dots, Q_p) ориентируемы. Итак, L гомеоморфна боковой поверхности цилиндра. Следовательно, после разрезания по линии l она распадается на две части, и потому поверхность Q_i^* имеет по сравнению с Q_{i-1} две новые компоненты края l_1, l_2 . Обратный переход от Q_i^* к Q_{i-1} заключается в склеивании двух контуров дыр l_1, l_2 , имеющих на поверхности Q^* . Окружим контуры l_1, l_2 узкими кольцевыми полосками и соединим их полоской друг с другом. Мы получим на поверхности Q_i^* фигуру («очки», рис. 72), гомеоморфную кругу с двумя дырами (рис. 73).

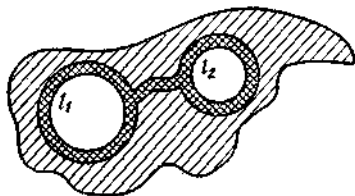


Рис. 72.

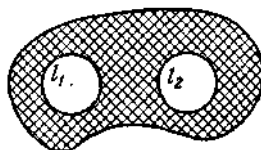


Рис. 73.

Склеивание контуров l_1 и l_2 должно производиться с учетом противоположной ориентации на них, так как иначе полоска, заштрихованная на рис. 74, превратилась бы при склеивании в ленту Мёбиуса, что невозможно в силу ориентируемости поверхности Q_{i-1} .

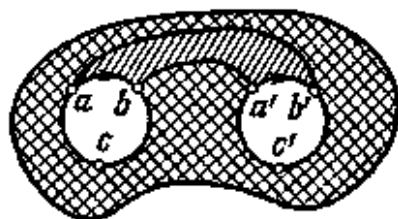


Рис. 74.

Следовательно, склеивание контуров l_1 и l_2 равносильно вклеиванию в поверхность Q_i^* ручки (рис. 75).



Рис. 75,

Итак, переход от Q_i к Q_i^* сводится к вырезыванию одной дыры, а переход от Q_i^* к Q_{i-1} — к уменьшению числа дыр и вклеиванию одной ручки. Поэтому, если Q_i получалось из сферы вырезанием нескольких дыр и вклеиванием части из них ручками, то это же справедливо и для Q_{i-1} .

Е) Проведенная индукция показывает, что Q_0 получается из сферы вырезанием $k+r$ дыр и заклеиванием k из цих ручками ($k \geq 0, r \geq 0$). Остается заметить, что при переходе от Q_0 к исходной поверхности Q в поверхность Q_0 вновь вклеиваются все сердцевины, т. е. каждая из r дыр, имеющихя в поверхности Q_0 , заклеивается кругом. Таким образом, Q получается из сферы вырезанием k дыр и вклеиванием их всех ручками, т. е. Q гомеоморфна одной из поверхностей (7).

Задачи

74. Сформулируйте в докажете теорему о топологической классификации ориентируемых поверхностей с краем.

75. На поверхности P_k проведено q контуров, не пересекающихся друг с другом, причем после разрезания по всем этим контурам по верхность остается связной. Докажите, что $q \leq k$.

76. На замкнутой поверхности Q осуществлено топологически правильное разбиение: каждая грань — пятиугольник, в каждой вершине сходятся по четыре грани. Докажите, что если число граней не кратно восьми, то поверхность Q неориентируема.

77. На замкнутой поверхности Q проведены три линии p, q, r , гомеоморфные отрезку, которые имеют общие концы и попарно не имеют других общих точек. Докажите, что если разрез по одной из линий $p \cup q, p \cup r, q \cup r$ оставляет поверхность связной, то хотя бы одна из двух других также обладает этим свойством.

78. Если на одной из граней правильного додекаэдра (рис. 76, а) продолжить все стороны до пересечения, то мы получим правильную пятиконечную звезду (рис. 76, б). Две такие звезды, построенные на смежных гранях (рис. 76, в) имеют общий отрезок ad . Условимся, однако, считать, что эти звезды примыкают друг к другу только по отрезкам ab и cd , а отрезок bc будем считать «лишним» пересечением этих звезд, происходящим из-за «неудачного» расположения этих звезд в пространстве. После построения аналогичных звезд для всех граней додекаэдра (рис. 76, г) мы получим некоторую поверхность Q , расположенную в пространстве с самопересечениями («лишними» линиями пересечения будут ребра исходного додекаэдра).

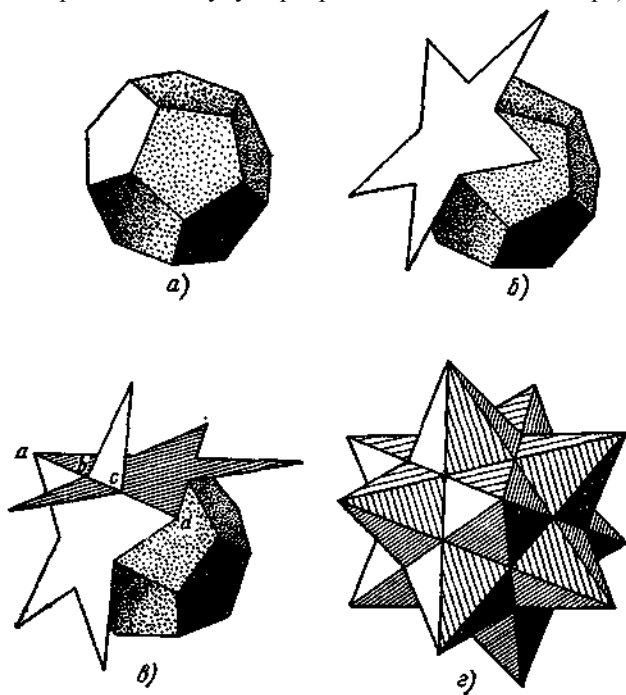


Рис. 76.

Докажите, что эта поверхность является ориентируемой и имеет эйлерову характеристику

$$\chi(Q) = -16,$$

а значит, гомеоморфна сфере с девятью ручками.

Можно построить и иную поверхность, связанную с додекаэдром. Добавим к контуру каждой звезды такие отрезки, как bc (рис. 76, e), так что получится замкнутая пятизвенная ломаная (самопересекающаяся). Затем расправим эти ломаные, устранив самопересечения (так что сторонами станут отрезки типа ad), и на каждую из них натянем грань (пятиугольник). Тогда получится поверхность, состоящая из двенадцати пятиугольных граней, причем число вершин (таких, как a , d) будет также равно 12, а число ребер 30. Докажите, что эта поверхность ориентируема и имеет эйлерову характеристику $\chi=6$, а значит, гомеоморфна сфере с четырьмя ручками.

79. Построим на каждой грани куба «четырёхконечную звезду» (с искривленными лучами; см. рис. 77, a) так, чтобы соседние звезды соприкасались краями лучей (рис. 77, b).

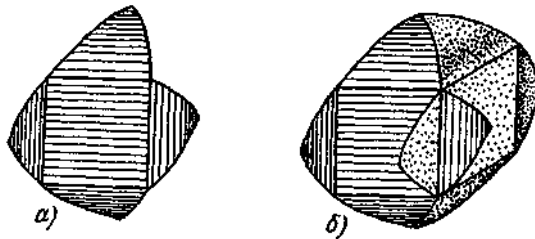


Рис. 77.

После построения таких звезд на всех гранях куба получится поверхность, расположенная в пространстве с самопересечениями («лишними» линиями пересечения будут ребра куба). Докажите, что эта поверхность гомеоморфна P_3 .

80. Какие поверхности получатся, если построить «трехконечные звезды» (аналогично тому, как описано в условии задачи 79) на гранях тетраэдра, октаэдра, икосаэдра?

2.2.5. Классификация замкнутых неориентируемых поверхностей

Замкнутую неориентируемую поверхность можно расположить в пространстве лишь с самопересечениями.

Пример 25. На рис. 78, a изображена поверхность с краем l , а на рис. 78, b — ее разрез через «горлышко».

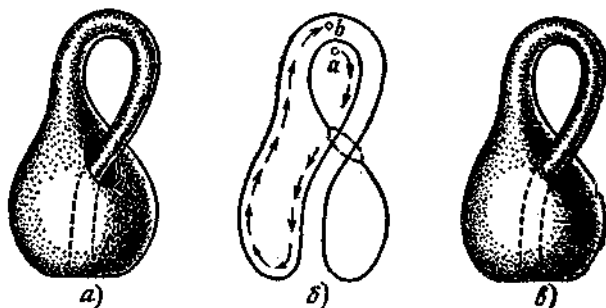


Рис. 78.

Заклеивание дыры l кругом дает замкнутую поверхность (рис. 78, в), которая, однако, пересекает себя. В действительности самопересечения не должно быть — мы хотим считать, что оно возникает лишь из-за «неудачного» расположения поверхности в пространстве. Полученная поверхность называется *бутылкой Клейна*. Она одно-сторонняя: начав движение от точки, расположенной на внешней поверхности горлышка, можно пройти внутрь горлышка (рис. 78, б).

Пример 26. Так как *край ленты Мёбиуса гомеоморфен окружности* (рис. 79), то можно попытаться приклеить ленту Мёбиуса своим краем к краю дыры, вырезанной в некоторой поверхности.



Рис. 79.

На рис. 80, а изображена лента Мёбиуса (кольцо с перекручиванием), а на рис. 80, б — кусок поверхности Q с вырезанной в ней дырой. Если «разогнуть» внутреннюю «лопасть» поверхности Q , то легко увидеть (рис. 80, в), что в ней вырезана дыра, гомеоморфная кругу.

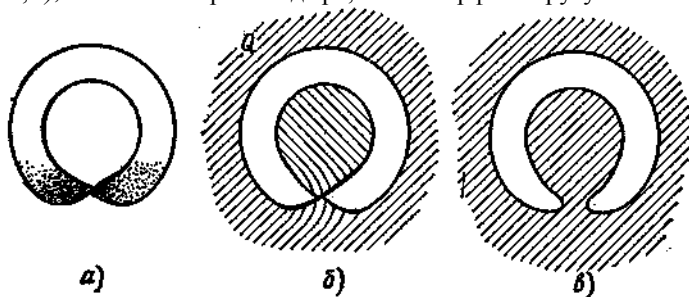


Рис. 80.

Так как поверхности, изображенные на рис. 80, *а* и *б*, имеют одинаковый край, то можно склеить их краями, т. е. *вклеить ленту Мёбиуса в круглую дыру, вырезанную в поверхности Q*. Правда, при этом лента Мёбиуса окажется пересекающей с поверхностью *Q*, но мы будем считать, что пересечение возникает лишь из-за «неудачного» расположения в пространстве.

Заклеивание дыры лентой Мёбиуса можно описать и иначе. Разрежем ленту Мёбиуса по ее средней линии. Для этого мы должны сначала склеить боковые стороны прямоугольника (с перекручиванием — чтобы получилась лента Мёбиуса), а затем произвести разрез по линии *mnp* (рис. 81, *а*). Но можно выполнить действия в обратном порядке: сначала разрезать прямоугольник по линии *mnp* (рис. 81, *б*), а затем произвести склеивание боковых отрезков (с учетом направления стрелок). Для склеивания повернем нижнюю половину прямоугольника «наизнанку» (рис. 81, *в*) и расположим половинки, как на рис. 81, *г*. Теперь нетрудно произвести необходимые склеивания (рис. 81, *д*).

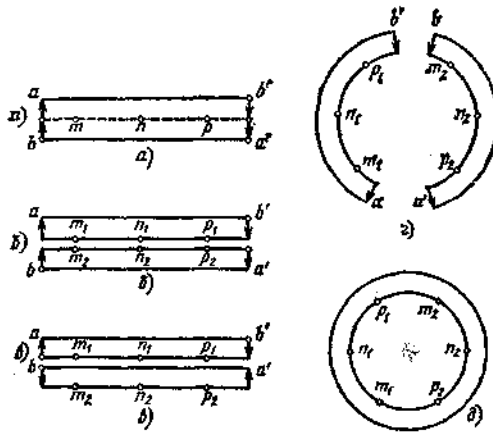


Рис. 81.

Мы видим, что *разрезание ленты Мёбиуса по средней линии дает фигуру, гомеоморфную кольцу*. На рис. 81 показаны точки, получившиеся из точек *m, n, p* при разрезании по средней линии. На рис. 81, *д* одинаково обозначенные точки являются диаметрально противоположными. Обратное склеивание снова превращает в ленту Мёбиуса. Следовательно, *если на одной окружности кольца склеить между собой каждые две диаметрально противоположные точки, то мы получим ленту Мёбиуса*.

Пусть теперь l — контур круглой дыры на некоторой поверхности Q . Вырежем из поверхности узкую полосу (кольцо) вокруг дыры l и обозначим через V наружный контур этого кольца (рис. 82).

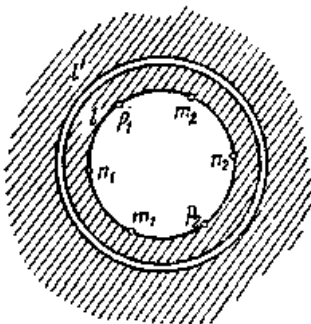


Рис. 82

Тогда получится поверхность, гомеоморфная Q (только с несколько большей дырой V), и отдельно кольцо. Склеим теперь на контуре l отрезанного кольца каждые две диаметрально противоположные точки; тогда кольцо превратится в ленту Мёбиуса. Эту ленту Мёбиуса мы и вклеим в дыру V . В результате мы вклеим в поверхность Q (точнее, в поверхность, гомеоморфную ей) ленту Мёбиуса. Но разрезание поверхности по контуру l и обратное склеивание этого разреза можно было и не делать: достаточно было просто склеить на контуре l каждые две диаметрально противоположные точки. Итак, *склеивание каждой двух диаметрально противоположных точек на контуре круглой дыры равносильно вклеиванию в эту дыру ленты Мёбиуса.*

Пример 27. В проективной геометрии к точкам обычной (евклидовой) плоскости присоединяются *несобственные* (бесконечно удаленные) точки. Присоединение бесконечно удаленных точек производится таким образом, что к каждой прямой, проходящей в евклидовой плоскости, присоединяется одна бесконечно удаленная точка, причем для всех параллельных между собой прямых эта бесконечно удаленная точка одна и та же (т. е. параллельные прямые «пересекаются в бесконечности»), а для

непараллельных прямых бесконечно удаленные точки различны. Плоскость, пополненная бесконечно удаленными точками, называется *проективной плоскостью*. Чтобы выяснить топологическую структуру проективной плоскости, рассмотрим полусферу с центром o , касающуюся плоскости и расположенную так, что диаметрально плоскость полусферы параллельна плоскости (рис. 83).

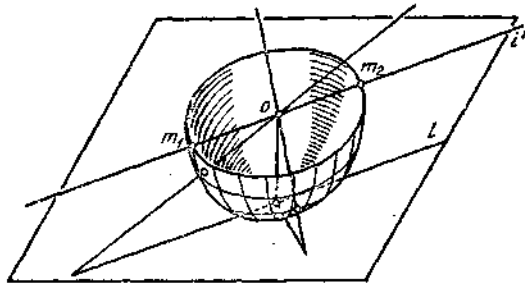


Рис. 83.

Центральное проектирование из точки o является гомеоморфным отображением открытой полусферы (получающейся выбрасыванием из полусферы всех точек ограничивающей ее окружности) на всю евклидову плоскость.

Проведем теперь через точку касания полусферы прямую l на плоскости, а через точку o — прямую l' , параллельную l . Прямые l и l' «пересекаются в бесконечности» так, что точки m_1 и m_2 , в которых прямая l' пересекается с краем полусферы, «проектируются» (вдоль прямой Γ) в одну и ту же точку — в бесконечно удаленную точку прямой l . Следовательно, отображение полусферы с краем на плоскость, пополненную бесконечно удаленными точками (т. е. на проективную плоскость), не является взаимно однозначным: двум различным точкам m_1, m_2 на крае полусферы соответствует одна и та же точка проективной плоскости. Чтобы это отображение стало взаимно однозначным (и гомеоморфным), нужно склеить между собой каждые две диаметрально противоположные точки на крае полусферы. Иначе говоря, проективная плоскость гомеоморфна полусфере, к краю которой приклеена лента Мёбиуса (или сфере с одной дырой, которая заклеена лентой Мёбиуса). Из этого следует, что проективная плоскость (в отличие от евклидовой) является односторонней поверхностью.

Теперь мы можем сформулировать вторую половину теоремы Мёбиуса—Жордана о классификации поверхностей, а именно, дать перечисление всех топологически различных типов замкнутых неориентируемых поверхностей. Обозначим через N_q поверхность, получающуюся из сферы вырезыванием в ней q дыр и заклеиванием их всех лентами Мёбиуса. Оказывается, что *поверхности*

$$N_1, N_2, \dots, N_q, \dots \quad (12)$$

и дают полную топологическую классификацию замкнутых неориентируемых поверхностей.

Задачи

81. Докажите, что если в поверхности N_q вырезать круглую дыру, то получающаяся поверхность с краем может быть расположена в трехмерном пространстве без самопересечений.

Указание: получающаяся поверхность гомеоморфна рассмотренной в задаче 62.

82. Докажите, что эйлериона характеристики шпирхюсти N_q равна $2 - q$.

83. На сфере вырезаны $m + n + p$ дыр; m из них заклеены ручками, а n — лентами Мёбиуса. Докажите, что эйлера характеристика получившейся поверхности с краем равна $2 - 2m - n - p$.

84. Докажите, что граф «4 домика и 4 колодца» (ребрами которого являются соединяющие тропинки — по одной от каждого домика к каждому колодцу) не может быть размещен без самопересечений на проективной плоскости, но может быть расположен на торе.

85. На поверхности Q удалось начертить граф « m домиков и n колодцев». Докажите, что $\chi(Q) \leq m + n - \frac{mn}{2}$.

86. Какой из поверхностей (12) гомеоморфна бутылка Клейна? Проективная плоскость?

87. Какие поверхности получаются, если на рис. 84, $a, б, в, г$ склеить (с учетом направлений) стороны, помеченные одинаковыми буквами?

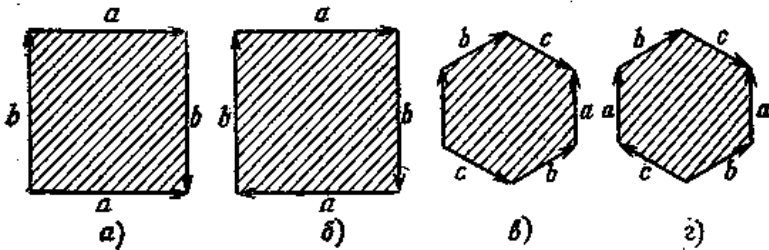


Рис. 84.

88. В трехмерном пространстве R^3 взята лента Мёбиуса, а в четырехмерном пространстве R^4 , содержащем R^3 , взята точка $p \notin R^3$. К ленте Мёбиуса добавлены все прямолинейные отрезки, соединяющие p с точками, лежащими на крае ленты Мёбиуса. Докажите, что получившаяся поверхность гомеоморфна проективной плоскости. Докажите, что любая поверхность N_p может быть расположена в R^4 без самопересечений.

Поверхности (12) попарно не гомеоморфны, так как имеют различные эйлеровы характеристики (задача 82). Поэтому для доказательства того, что поверхности (12) дают полную топологическую классификацию замкнутых неориентируемых поверхностей, остается установить, что каждая замкнутая неориентируемая поверхность гомеоморфна одной из поверхностей (12). Это доказывается точно так же, как и в пункте 12. Разница будет, во-первых, в том, что лента L (см. рис. 70) теперь может оказаться гомеоморфной ленте Мёбиуса (поскольку рассматриваются неориентируемые поверхности). В этом случае поверхность Q^*_i , получившаяся после вырезания полосы L , будет иметь одну новую компоненту края (так как край полосы L , т. е. ленты Мёбиуса, гомеоморфен окружности). Наоборот, Q_{i+1} получается из Q^*_i приклеиванием полосы L к одному из контуров, принадлежащих краю поверхности Q^*_i , т. е. Q_{i-1} получается из Q^*_i заклеиванием одной дыры лентой Мёбиуса. Во-вторых, разница будет в том, что склеивание контуров l_1 и l_2 на рис. 72—73 теперь может производиться с учетом не обязательно противоположной ориентации на них (что равносильно вклеиванию ручки, рис. 75), но, возможно, с учетом одинаковой ориентации контуров. В этом случае поверхность Q_{i-1} получается из Q^*_i заклеиванием двух дыр лентами Мёбиуса (задача 89). Таким образом, рассуждение, проведенное в пункте 1.2.4, показывает, что *любая замкнутая неориентируемая поверхность Q получается из сферы вырезанием $k + q$ дыр и заклеиванием k дыр ручками, а q дыр — лентами Мёбиуса*. При этом $q \geq 1$, так как при $q = 0$ мы получили бы ориентируемую поверхность P_k , а по условию поверхность Q была неориентируемой. Остается заметить, что если в поверхность вклеена хотя бы одна лента Мёбиуса, то вклеивание ручки равносильно вклеиванию двух лент Мёбиуса (задача 90). Поэтому поверхность Q , получающаяся из сферы вырезанием $k+g$ дыр и вклеиванием k ручек и q лент Мёбиуса (где $q \geq 1$), гомеоморфна сфере, в которой вырезаны $2k + q$ дыр и все они заклеены лентами Мёбиуса. Иначе говоря, Q гомеоморфна одной из поверхностей (12).

Задачи

89. В круге вырезаны две дыры, и их контуры l_1, l_2 склеены с учетом одинаковой ориентации на них. Докажите, что это эквивалентно заклеиванию обеих дыр лентами Мёбиуса.

Указание. Произведем дополнительные разрезы по линиям $amna'$, $erqe'$ (рис. 85, а) и перевернем «наизнанку» отрезанный кусок (рис. 85, б). Теперь склеивание контуров l_1 и l_2 осуществляется непосредственно (рис. 85, в) и остается вновь склеить проведенные разрезы, т. е. склеить «диаметрально противоположные» точки на двух контурах.

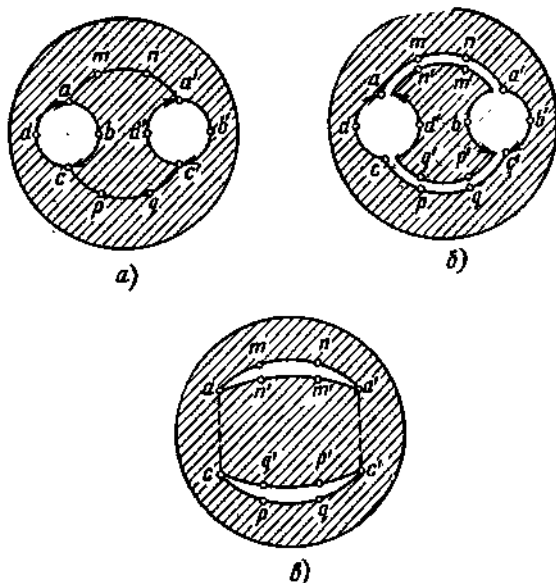


Рис. 85.

90. В круге вырезаны три дыры и одна из них заклеена лентой Мёбиуса, а контуры двух других дыр склеены с учетом противоположной ориентации на них (ручка). Докажите, что это равносильно заклеиванию всех трех дыр лентами Мёбиуса.

Указание. Проведем дополнительный разрез $tabm_2$, а затем вывернем отрезанный кусок «наизнанку» (рис. 86). Мы получаем «серповидную» дыру, на контуре которой склеиваются каждые две «диаметрально противоположные» точки, и еще два контура, которые надо склеить с учетом одинаковой ориентации на них.

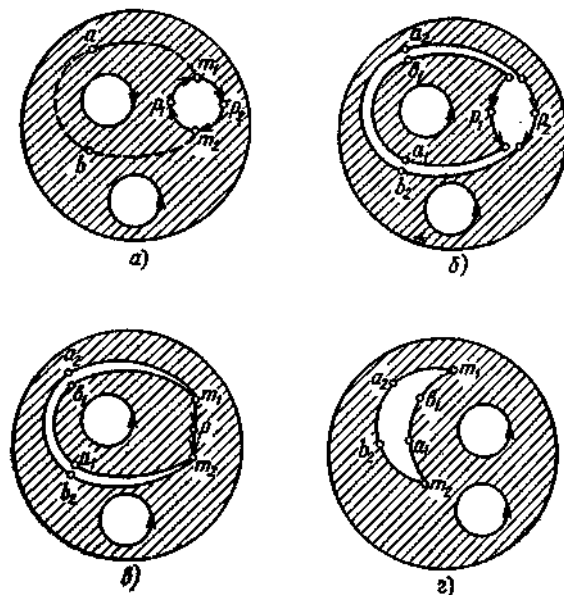


Рис. 86.

91. Сформулируйте и докажите теорему о топологической классификации неориентируемых поверхностей с краем.

2.2.6. Векторные поля на поверхностях

Вопрос, который рассматривается в этом пункте, заключается в следующем. Можно ли на заданной ориентируемой поверхности Q построить непрерывное поле направлений, т. е. выбрать в каждой ее точке такой ненулевой касательный вектор, что при переходе от точки к точке вектор меняется непрерывно?

Пример 28. На сфере направление с севера на юг (рис. 87, а) имеет особые точки в полюсах; в этих точках векторы направлены в разные стороны и непрерывность нарушается. То же можно сказать о направлении с запада на восток (рис. 87, б).

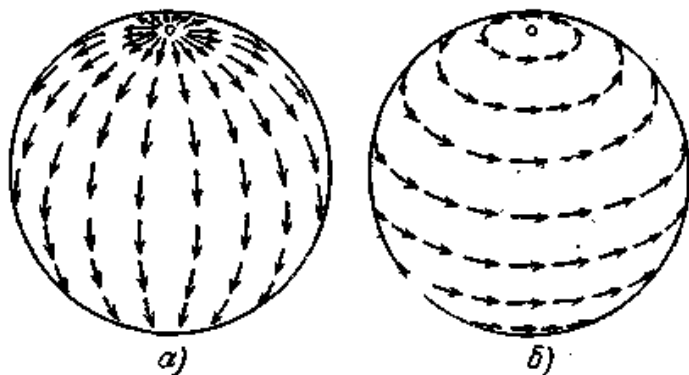


Рис. 87.

Вообще, как мы увидим дальше, на всей сфере не существует непрерывного поля направлений. Это иногда формулируют в виде «теоремы о еже»: если из каждой точки поверхности сферы растет «колючка» (ненулевой вектор, не обязательно касающийся сферы) и направления «колючек» от точки к точке меняются непрерывно, то найдется хотя бы одна «колючка», перпендикулярная к сфере.

Действительно, в противном случае, спроектировав каждую «колючку» $\vec{a}q$ на касательную плоскость, проведенную в точке a , из которой эта «колючка» растет (рис. 88), мы получили бы на всей сфере непрерывное поле ненулевых касательных векторов, а это невозможно.

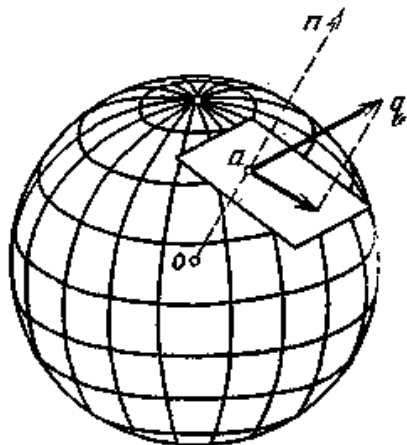


Рис. 88.

На рис. 89, *а, б* показан вид векторных полей, рассмотренных в примере 28, вблизи северного полюса, а на рис. 89, *в* — более сложная особая точка (так называемое *седло*).

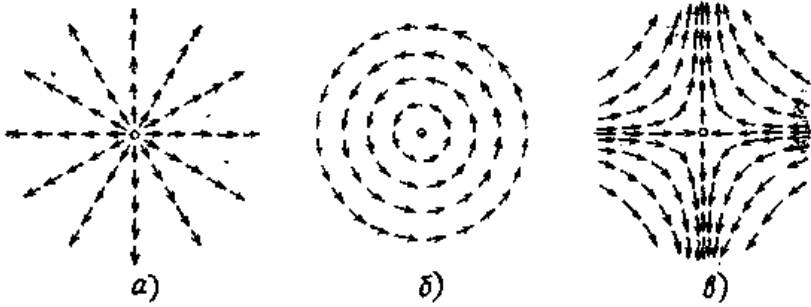


Рис. 89.

Когда мы один раз обойдем особую точку (например, против часовой стрелки), направления векторов совершат в случаях, изображенных на рис. 89, *а*) и *б*) один поворот в том же направлении (рис. 90, *а, б*), а в случае рис. 89, *в*) — один поворот, но уже в противоположном направлении (рис. 90, *в*).

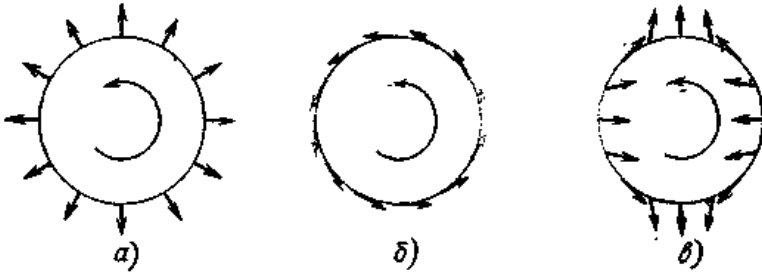


Рис. 90.

В связи с этим говорят, что особая точка на рис. 89, *а*) (и на рис. 89, *б*) имеет индекс $+1$, а особая точка на рис. 89, *в*) — индекс -1 . Выдающийся французский математик Анри Пуанкаре (1854—1912) доказал, что *если на замкнутой ориентируемой поверхности Q задано поле ненулевых касательных векторов, непрерывное всюду, кроме конечного числа особых точек, то сумма индексов всех особых точек равна $\chi(Q)$.*

Пример 29. Так как $\chi(P_k) = 2 - 2k$, то $\chi(P_k) \neq 0$ при $k \neq 1$. Следовательно, на ориентируемой поверхности, отличной от тора P_1 , не существует непрерывного поля ненулевых касательных векторов без особых точек. На торе же векторное поле существует (например,

можно взять векторы, направленные вдоль параллелей). Доказательство теоремы Пуанкаре проведем в два этапа: сначала докажем, что для любых двух векторных полей сумма индексов одинакова, а затем построим поле, для которого эту сумму легко вычислить.

Пусть на поверхности Q заданы два ненулевых векторных поля с конечным числом особых точек. Вектор первого поля в точке x обозначим через $v_1(x)$, а вектор второго поля — через $v_2(x)$. Разобьем Q на маленькие многоугольники так, чтобы в каждом многоугольнике было не более одной особой точки каждого поля и все особые точки лежали внутри многоугольников.

Заметим, что если x не является особой точкой поля v_1 , то вблизи точки x мы можем повернуть векторы этого поля, оставляя его непрерывным (рис. 91: по мере увеличения радиуса окружности векторы поворачиваются все меньше).

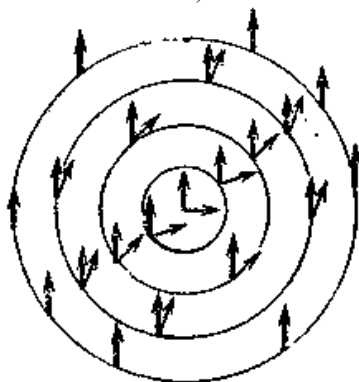


Рис. 91.

Пользуясь этим, повернем векторы поля v_1 вблизи вершин таким образом, чтобы в каждой вершине x_0 векторы $v_1(x_0)$ и $v_2(x_0)$ совпали (рис. 92).

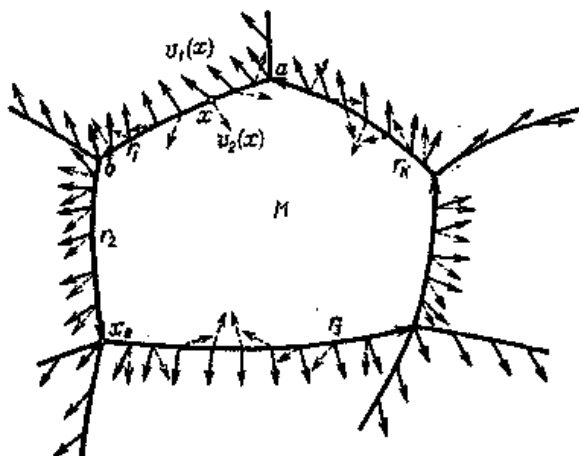


Рис. 92.

Так как поверхность Q ориентируема, то на ней можно указать положительное направление отсчета углов (скажем, против часовой стрелки, если смотреть на поверхность с внешней стороны).

Возьмем теперь некоторое ребро r_1 (рис. 92) и выберем на нем направление, (например, от вершины a к b). Будем сначала, двигаясь от a к b в этом направлении, следить за вектором $v_1(x)$, а потом, возвращаясь от b к a , следить за вектором $v_2(x)$. Когда мы пробежим ребро r_1 туда и обратно, вектор, за которым мы следим, непрерывно перемещаясь, вернется к прежнему положению (поскольку $v_1(a) = v_2(a)$ и $v_1(b) = v_2(b)$). Число оборотов, которое совершает этот вектор (учитывая выбранное направление отсчета углов), обозначим через $d(r_1)$. На рис. 92 имеем $d(r_1) = 1$, $d(r_2) = 0$, $d(r_3) = -1$. Заметим, что если на ребре r_2 взять противоположное направление (от b к a), то $d(r_1)$ изменит знак (так как вектор, за которым мысленно будем поворачиваться в обратном направлении).

Пусть M — один из многоугольников. Когда мы обойдем его контур (в положительном направлении), вектор $v_1(x)$ совершит некоторое число оборотов — обозначим это число через $z_1(M)$, — а вектор $v_2(x)$ совершит $z_2(M)$ оборотов.

Обозначим через r_1, r_2, \dots, r_k стороны многоугольника M и зададим на них направления, соответствующие положительному обходу его контура (см. рис. 92). Обойдем контур (начиная от точки a) в

положительном направлении, следя за вектором $v_1(x)$, а затем (после возвращения в a) обойдем контур в противоположном направлении, наблюдая за вектором $v_2(x)$. В результате наблюдаемый вектор совершит $z_1(M) - z_2(M)$ оборотов.

Но мы можем следить за поворотами векторов «по частям»: наблюдая $v_1(x)$ при движении по ребру r_1 и $v_2(x)$ при обратном движении по ребру r_1 , далее, наблюдая $v_1(x)$ при прохождении ребра r_2 и $v_2(x)$ при обратном движении по ребру r_2 и т. д. В этом случае мы насчитаем $d(r_1) + d(r_2) + \dots + d(r_k)$ оборотов. Так как суммарный поворот не зависит от того, в каком порядке складывать углы поворота вектора на каждом из ребер, то

$$z_1(M) - z_2(M) = d(r_1) + d(r_2) + \dots + d(r_k). \quad (13)$$

Из формулы (13) нетрудно вывести соотношение

$$\sum z_1(M) = \sum z_2(M), \quad (14)$$

в котором суммирование производится по всем многоугольникам. В самом деле, просуммируем равенства (13) по всем многоугольникам. В правой части получившейся суммы каждое ребро r встретится дважды, так как к нему примыкают два многоугольника M_1 и M_2 (рис. 93).

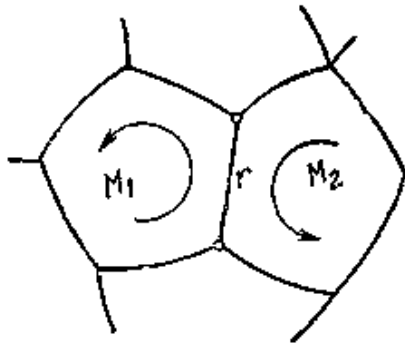


Рис. 93.

Но при положительном обходе контура M_1 ребро r получит одно направление, а при положительном обходе контура M_2 — противоположное направление. Следовательно, в правой части один раз встретится $d(r)$, а второй раз $-d(r)$. Так как это произойдет с каждым ребром, то $\sum z_1(M) - \sum z_2(M) = 0$.

Пусть M — некоторый многоугольник и x_0 — особая точка поля $v_1(x)$, расположенная в нем. Построим систему простых замкнутых линий, соединяющую контур многоугольника M с окружностью, обходящей точку x_0 (рис. 94).

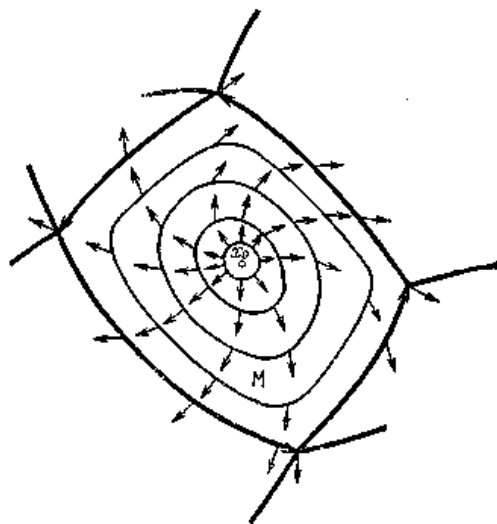


Рис. 94.

При переходе от одной линии к близкой ей число оборотов вектора $v_1(x)$ должно измениться мало, поскольку поле $v_1(x)$ непрерывно. Но число оборотов является целым и потому «мало» измениться не может, т. е. остается постоянным при переходе от линии к линии. Но при обходе по контуру многоугольника M число оборотов равно $z_1(M)$, а при обходе по окружности вокруг точки x_0 получается индекс этой точки. Таким образом, число $z_1(M)$ равно индексу особой точки x_0 (если внутри M нет особых точек, то $z_1(M) = 0$). Из этого вытекает, что число $\sum z_1(M)$ равно сумме индексов всех особых точек поля $v_1(x)$.

Аналогично, число $\sum z_2(M)$ равно сумме индексов поля $v_2(x)$. Из этого в силу (14) вытекает, что у обеих полей сумма индексов одинакова. Этим завершён первый этап.

Выберем теперь внутри каждого многоугольника «центр», а на каждом ребре — «середину», и построим векторное поле, как показано на рис.95: вдоль ребер векторы направлены от вершин к «середине», из вершин векторы выходят, а к «центрам» многоугольников приходят.

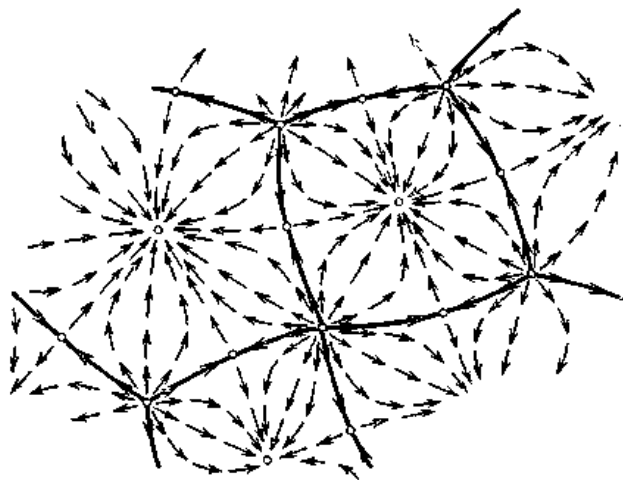


Рис. 95.

Особыми точками этого поля на поверхности Q будут вершины, «середины» и «центры». При этом (рис. 96) индекс каждой особой точки в вершине и в «центре» равен $+1$, а индекс «середины» ребра равен -1 (седло).



Рис. 96.

Следовательно, для этого поля (а потому и для любого другого) сумма индексов всех особых точек равна

$$B \cdot (+1) + P \cdot (-1) + \Gamma \cdot (+1) = \chi(Q).$$

Задачи

92. Докажите, что на всякой замкнутой поверхности существует векторное поле с единственной особой точкой.

93. Докажите, что на всякой поверхности с краем существует векторное поле без особых точек (направление векторов в точках края должно касаться поверхности, но может не быть касательным к краю).

94. Докажите, что теорема Пуанкаре остается справедливой для ориентируемой поверхности с краем, если в каждой точке края вектор направлен по касательной к этому краю.

95. Докажите теорему Брауэра: если $f: K \rightarrow K$ — произвольное непрерывное отображение круга K в себя, то существует (хотя бы одна) неподвижная точка, т. е. такая точка $x \in K$, которая переходит при отображении f в себя: $f(x) = x$.

Указание. Если бы неподвижных точек не было, то, построив векторы, идущие из каждой точки x в точку $f(x)$, мы получили бы ненулевое непрерывное векторное поле без особенностей.

2.2.7. Проблема четырех красок

Области, на которые конечный граф G разбивает плоскость, назовем «странами». На рис. 97 страны А и Б пограничны (примыкают друг к другу по общему ребру).

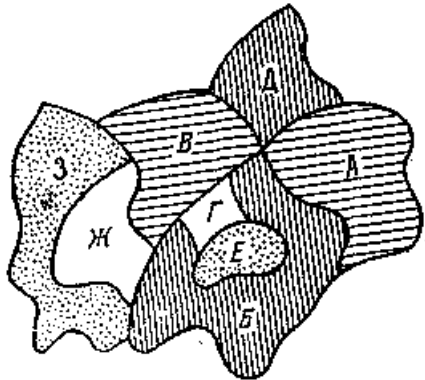


Рис. 97

Страны Б и В также пограничны (у них даже два общих ребра). Страны А и В не являются пограничными: хотя у них есть общая точка (вершина), но нет общих ребер.

Мы хотим раскрасить страны в разные цвета, чтобы получилась «политическая карта». Разумеется, чтобы страны были хорошо видны, необходимо пограничные страны окрашивать в разные цвета. Однако в целях экономии количества красок разрешается непограничные страны окрашивать одним цветом. Какое минимальное количество красок нужно иметь, чтобы можно было раскрасить любую карту на плоскости?

Эта задача была сформулирована в 1852 году лондонским студентом Гутри, который обнаружил, что для различения графств на карте Англии достаточно четырех красок, и выдвинул гипотезу о том, что четырех цветов достаточно для раскраски любой карты. Спустя почти сорок лет английский математик Хивуд доказал, что любую карту на плоскости можно раскрасить в пять цветов. Постепенно проблема четырех красок приобретала все больший интерес. В 1968 году Оре и Стемпл доказали, что любую карту, имеющую *не более 40 стран*, можно раскрасить в четыре цвета.

В настоящее время считается, что справедливость гипотезы четырех красок установлена. Слово «считается» употреблено в связи с тем, что известные сейчас решения этой проблемы базируются на применении ЭВМ и связаны с выполнением огромного, необозримого количества вычислений, причем проверка правильности вычислений практически невыполнима. Первое «машинное» решение было получено в 1976 году американскими математиками К. Аппелем и В. Хакеном. С помощью машины (которая «помогала» им постепенно усовершенствовать первоначальную программу) они разбили все возможные карты на почти 2000 (четко указанных) типов и составили машинную программу их исследования. Для каждого из этих типов (кроме трех, которые были исследованы «вручную», поскольку машина с ними не справилась) машина решала (по составленной программе) задачу: *может ли в рассматриваемом типе карт найтись такая, которая не раскрашивается в четыре цвета?* Когда, выполнив десятки миллиардов арифметических и логических операций, машина давала ответ «Нет», переходили к следующему типу карт и т. д. Получив ответ «Нет» для всех типов карт, Аппель и Хакен объявили, что ими получено машинное решение проблемы четырех красок.

Однако гарантии в правильности этого «машинного» решения все же нет. Ведь в каком-то (скажем, семнадцатом) типе карт машина могла ответить «Нет» не в результате безупречного анализа, а из-за мимолетного сбоя в электронной схеме (что бывает нередко). Не зная об этом, вычислители переходят к восемнадцатому, девятнадцатому ... типу карт, фактически пропустив исследование семнадцатого типа. Не будет гарантии в правильности решения даже в том случае, если мы, затратив много месяцев, повторим уникальный машинный эксперимент: может быть, где-то в многомиллиардной цепочке вычислений, связанных с тем же семнадцатым типом, и в нашей машине произойдет сбой?

Новое машинное решение было предложено в 1978 году Д. Коэном. Число типов карт было у него существенно меньшим, причем результат машинных вычислений по каждому типу и подтипу он

получал не в виде готового «Нет», а в форме, допускающей «ручную» проверку. Его не интересовало, каким путем шла машина при исследовании данного подтипа и сколько операций это потребовало; исследование подтипа считалось завершенным, когда машина находила достаточно короткий путь проверки окончательного «Нет». Решение проблемы четырех красок, найденное Коэном, изложено в книге среднего объема и формата. По его мнению, проверка этого решения может быть, при желании, выполнена вручную одним человеком в течение двух-трех лет (!) ежедневной восьмичасовой работы. Однако «скептики» считают, что и это решение малоприемлемо: вряд ли человек, занимавшийся с утра до вечера в течение двух лет нудным перебором вариантов, может гарантировать, что он нигде не допустил ни одной ошибки»

Задачи

96. На плоскости (или на сфере) начерчен граф, все вершины которого имеют четные индексы. Докажите, что получившаяся карта может быть раскрашена в два цвета.

Указание: воспользуйтесь понятием индекса пересечения.

97. Докажите, что для раскраски любой карты на плоскости (или на сфере) достаточно пяти цветов.

98. На некоторой поверхности начерчен такой граф, что из каждой двух пограничных стран хотя бы одна является треугольником. Докажите, что эту карту можно раскрасить в четыре цвета.

99. Проведено n «перегородок», идущих от одной на концентрических окружностей к другой. Сколько нужно цветов для раскрашивания получившейся карты?

2.2.8. Раскрашивание карт на поверхностях

Пример 30. Хивуд доказал, что *любую карту на торе можно раскрасить семью красками* (это вытекает из доказываемого ниже неравенства (16)). Он привел также пример, показывающий, что меньше, чем семью цветами, обойтись нельзя. При склеивании противоположных сторон прямоугольник (рис. 98) превращается в тор с семью странами на нем (рис. 99).

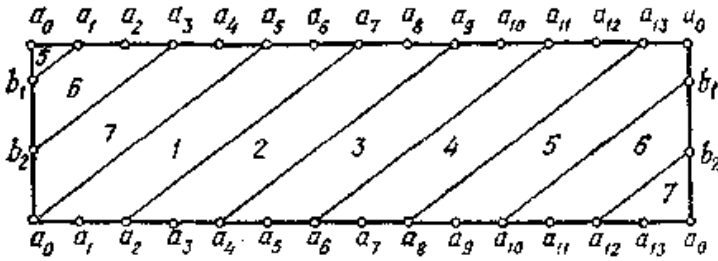


Рис. 98.

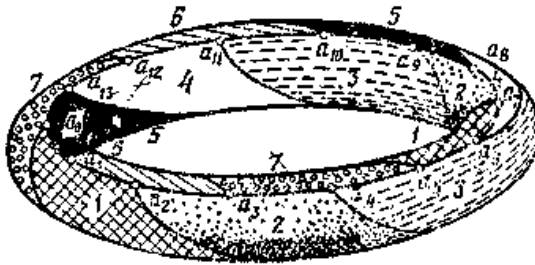


Рис. 99.

Каждые две страны пограничны, т. е. все семь стран должны быть раскрашены в разные цвета.

Если любая карта на поверхности Q допускает раскраску в n цветов, но существует карта, для раскраски которой нельзя обойтись меньшим числом цветов, то n называется *хроматическим числом* поверхности Q ; оно обозначается через $\text{col}(Q)$. Для сферы и тора имеем, согласно сказанному выше, $\text{col}(P_0) = 4$, $\text{col}(P_1) = 7$.

Вообще, для произвольной замкнутой поверхности Q , отличной от бутылки Клейна N_2 , хроматическое число дается *формулой Хивуда*:

$$\text{col}(Q) = \left\lceil \frac{7 + \sqrt{49 - 24\chi(Q)}}{2} \right\rceil, \quad (15)$$

где квадратные скобки означают *целую часть*, а для бутылки Клейна имеем $\text{col}(N_2) = 6$.

Эти результаты получены усилиями нескольких поколений математиков. Хивуду принадлежит доказательство неравенства

$$\text{col}(Q) \leq \left\lceil \frac{7 + \sqrt{49 - 24\chi(Q)}}{2} \right\rceil. \quad (16)$$

Оставалось установить, что на поверхности Q , отличной от бутылки Клейна, существует карта, для раскраски которой нельзя обойтись

меньшим, чем указано в (15), числом цветов. Вначале такие карты были построены для нескольких первых поверхностей в последовательностях (7), (12). Доказательство существования требуемой карты для любой неориентируемой поверхности было получено Рингелем (1954), а для ориентируемой — Рингелем и Янгсом (1968).

Приведем доказательство неравенства (16). Пусть на поверхности Q начерчена карта, требующая для своей раскраски s цветов, где $s = \text{col}(Q)$. Выберем внутри каждой страны точку («столицу»). Для каждых двух пограничных стран проведем по территории этих стран одну «железную дорогу», соединяющую их столицы (рис. 100), причем так, чтобы различные «железные дороги» не пересекались.

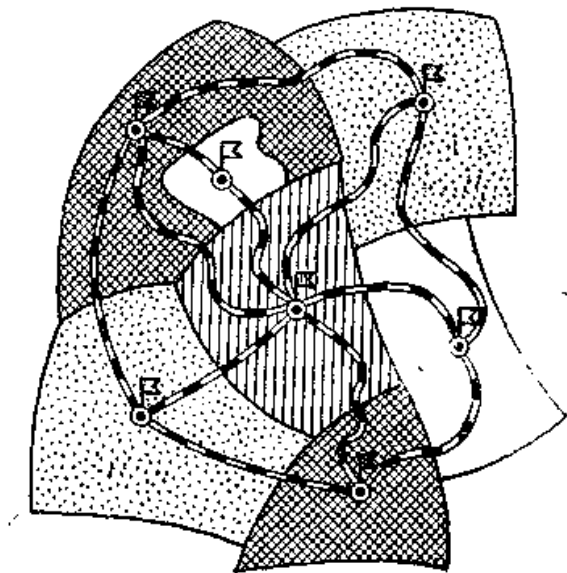


Рис. 100.

Вместо окраски страны в некоторый цвет мы можем водрузить в ее «столице» флаг, имеющий этот цвет; если при этом две «столицы» соединены «железной дорогой» (т. е. страны пограничны), то их флаги должны иметь разный цвет. Таким образом, надо «окрасить» вершины графа G^* (ребрами которого являются «железные дороги») таким образом, чтобы любые две смежные вершины (соединенные ребром) имели разный цвет. Ясно, что *хроматическое число* графа G^* , т. е. наименьшее число цветов, требуемых для такой его раскраски, равно s .

Выбросим из G^* некоторую вершину a и все примыкающие к ней ребра. Если при переходе к получившемуся графу G' хроматическое число не уменьшилось, то вместо G^* можно взять более простой граф G' . Возможно, в G' тоже можно провести выбрасывание вершины и т. д. В конце концов мы получим «неупрощаемый» граф G^{**} , содержащийся в G^* , т. е. хроматическое число графа G^{**} равно c , но выбрасывание любой вершины графа G^{**} и примыкающих к ней ребер приводит к уменьшению хроматического числа. Число вершин и ребер графа G^{**} обозначим через B и P , а число областей, на которые этот граф разбивает поверхность Q , — через Γ . Тогда (см. задачу 72)

$$B - P + \Gamma \geq \chi(Q). \quad (17)$$

К каждой вершине графа G^{**} примыкает не менее $c - 1$ ребер. В самом деле, допустим, что к вершине $b \in G^{**}$ примыкает k ребер $[bq_1], \dots, [bq_k]$, где $k < c - 1$.

Выбросив из G^{**} вершину b и эти ребра, мы получим граф G'' с меньшим, чем c , хроматическим числом. Раскрасим этот граф в $c - 1$ цвет. Так как $k < c - 1$, то из $c - 1$ цвета, в которые окрашен граф G'' , хотя бы один цвет не использован для окраски вершин q_1, \dots, q_k . Окрасив вершину b в этот неиспользованный цвет, мы получим окраску графа G^{**} в $c - 1$ цвет, что, однако, противоречит выбору графа G^{**} .

Итак, к каждой вершине графа G^{**} примыкает не менее $c - 1$ ребра. Из этого (см. задачу 20) вытекает неравенство

$$(c - 1) B \leq 2P. \quad (18)$$

Далее, каждая из областей, определяемых графом G^{**} , имеет не менее трех ребер. Действительно, наличие «одноугольника» (рис. 101, а) означало бы существование «железной дороги», идущей из столицы в ту же самую столицу (без захода в другие столицы), а наличие «двуугольника» (рис. 101, б) означало бы, что некоторые две «столицы» соединены двумя дорогами; но мы таких дорог не проводили.

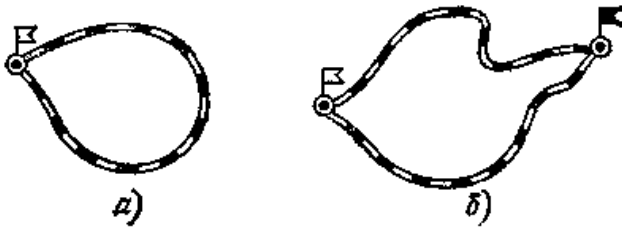


Рис 101.

Пересчитав теперь ребра по всем Γ областям, мы насчитаем не менее 3Γ ребер; при этом каждое ребро засчитывается дважды (так как к нему примыкают две области). Следовательно, $3\Gamma \leq 2P$, т. е. $\frac{2}{3}P - \Gamma \geq 0$. Прибавляя это неравенство к (17), получаем $B - \frac{1}{3}P \geq \chi(Q)$ или иначе, $2P \leq 6B - 6\chi(Q)$. Учитывая (18), получаем теперь $(c - 1)B \leq 6B - 6\chi(Q)$, т. е.

$$c - 1 \leq 6 - \frac{6\chi(Q)}{B}, \quad (19)$$

Теперь нетрудно завершить доказательство. Пусть сначала поверхность Q гомеоморфна сфере: $Q = P_0$. Тогда $\chi(Q) = 2$, т. е. доказываемое неравенство (16) принимает вид $\text{col}(Q) \leq 4$. Это неравенство справедливо, поскольку проблема четырех красок решена. Пусть теперь $Q = N_1$, т. е. $\chi(Q) = 1$. Доказываемое неравенство (16) принимает вид $\text{col}(Q) \leq 6$. Это неравенство справедливо, так как из (19) следует, что $c - 1 \leq 6 - \frac{6}{B}$ и потому $c - 1 \leq 5$ (поскольку число $c - 1$ — целое). Пусть, наконец, Q — замкнутая поверхность, отличная от P_0 и N_1 . Тогда $\chi(Q) \leq 0$ (задачи 70, 82). Так как $B \geq c$ (иначе граф G^{**} можно было бы раскрасить в $c - 1$ цвет), то $-\frac{6\chi(Q)}{B} \leq -\frac{6\chi(Q)}{c}$, и потому, согласно (19), $c - 1 \leq 6 - \frac{6\chi(Q)}{c}$, т. е. $c^2 - 7c + 6\chi(Q) \leq 0$. Это означает, что число c принадлежит отрезку, концами которого являются корни квадратного трехчлена $x^2 - 7x + 6\chi(Q)$ (корни действительны, поскольку $6\chi(Q) \leq 0$). Следовательно, c не превосходит большего из этих корней, т. е.

$$c \leq \frac{1}{2} (7 + \sqrt{49 - 24\chi(Q)}).$$

Таким образом, неравенство (16) справедливо и в этом случае.

Задачи

100. Поверхность Q получается из сферы вырезыванием $k + q$ дыр и заклеиванием k из них ручками. Докажите, что

$$\text{col}(Q) = \text{col}(P_k).$$

101. Поверхность Q получается из сферы вырезыванием $k + q$ дыр и заклеиванием q из них лентами Мёбиуса. Докажите, что $\text{col}(Q) = \text{col}(N_q)$. В частности, хроматическое число ленты Мёбиуса равно 6.

102. Приведите пример карты на проективной плоскости (или на ленте Мёбиуса), которую нельзя раскрасить пятью цветами.

103. Чему равно хроматическое число графа, вершинами и ребрами которого служат вершины и ребра n -угольника?
104. Каково хроматическое число графа « m домиков и n колодцев»?
105. Докажите, что существует граф, не вложимый в заданную поверхность.
106. Граф G с хроматическим числом 2 имеет n вершин. Какое максимальное число ребер может иметь этот граф?
107. Докажите, что если на заданной поверхности начерчена карта с достаточно маленькими странами, то ее можно раскрасить в 7 цветов.
108. Докажите, что если на поверхности Q можно начертить граф, имеющий хроматическое число c , то $\text{col}(Q) \geq c$.
109. Докажите, что на бутылке Клейна можно начертить полный граф с шестью вершинами. Выведите отсюда, что $\text{col}(N_2) \geq 6$.

2.2.9. «Дикая сфера»

Мы рассмотрим здесь вопросы, связанные с пространственной теоремой Жордана, но прежде поговорим об *индексе пересечения* в пространстве.

Пусть Q — поверхность (возможно, имеющая край), составленная из плоских многоугольников, и G — граф, ребрами которого являются прямолинейные отрезки. Будем говорить, что G и Q *находятся в общем положении*, если вершины графа не принадлежат Q , а его ребра не имеют общих точек с ребрами поверхности Q . Если при этом число точек пересечения графа с поверхностью *четно*, то будем писать $J(G, Q) = 0$, а если *нечетно*, то $J(G, Q) = 1$. Число $J(G, Q)$ назовем *индексом пересечения* графа G с поверхностью Q .

Как и в п. 5, доказывается, что *если поверхность Q не имеет края, а граф G представляет собой цикл (т. е. в каждой его вершине сходится четное число ребер), то*

$$J(G, Q) = 0.$$

Задачи

110. Условимся говорить, что объединение конечного числа многоугольников в пространстве представляет собой *двумерный цикл* (по модулю 2), если эти многоугольники не имеют общих внутренних точек и к каждому ребру (мы считаем «ребрами» стороны многоугольников) примыкает четное число многоугольников. Докажите, что равенство $J(G, Q) = 0$ остается справедливым, если G — одномерный, а Q — двумерный цикл (по модулю два), причем они находятся в пространстве в общем положении.

111. Пусть Q — двумерный цикл по модулю 2, который будем представлять себе «спаянным» из металлических многоугольников. Докажите, что области, на которые он разбивает пространство, можно «залить жидкостями» двух цветов, так, чтобы к каждому многоугольнику примыкали с двух сторон разные по цвету «жидкости».

112. Докажите, что существует «пространственная карта», требующая для своей раскраски (или «заливки разноцветными жидкостями») не менее 1982 цветов.

113. Конечное множество ориентированных многоугольников в пространстве, которые не имеют общих внутренних точек (но могут иметь общие ребра и вершины), образуют *двумерный целочисленный цикл*, если для каждого направленного ребра число положительно примыкающих к нему многоугольников (M_1 и M_4 на рис. 102) равно числу отрицательно примыкающих (M_2 и M_3 на рис. 102).

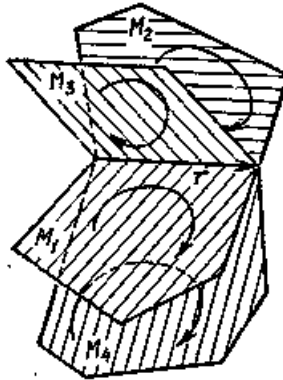


Рис. 102.

Докажите, что если G — одномерный целочисленный цикл, а Q — двумерный целочисленный цикл в пространстве, то их индекс пересечения

$$J(G, Q) = \sum_{i,j} J(r_i, M_j)$$

равен нулю; здесь суммирование распространено на все направленные отрезки r_1, \dots, r_k , составляющие цикл G , и все ориентированные многоугольники M_1, \dots, M_l , составляющие цикл Q , причем $J(r_i, M_j) = +1$ или $J(r_i, M_j) = -1$, в зависимости от

того, соответствуют ли ориентации *правилу буравчика* (рис. 103, а) или нет (рис. 103, б).

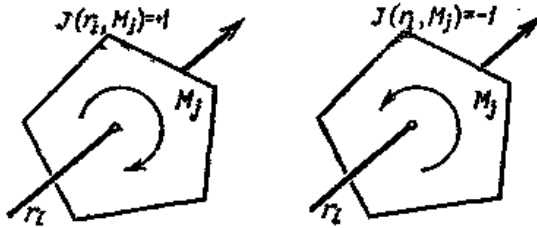


Рис. 103.

114. Пусть Q — двумерный целочисленный цикл, p и q — не принадлежащие ему точки. Проведем направленную ломаную x , идущую от точки p к q . Докажите, что индекс пересечения $J(x, Q)$ не зависит от выбора ломаной x , а определяется только точками p и q .

115. Пусть Q — двумерный цикл по модулю 2. Докажите, что его многоугольники можно ориентировать так, чтобы получился целочисленный цикл.

116. Пусть Q — двумерный целочисленный цикл. Докажите, что если существует направленная ломаная x (незамкнутая), для которой $J(x, Q) = n$, то Q разбивает пространство не менее, чем на $n+1$ областей. Верно ли обратное?

С помощью понятия индекса пересечения по модулю 2 доказывается (примерно так же, как в п. б) пространственный аналог теоремы Жордана: *всякая замкнутая поверхность Q , расположенная в трехмерном пространстве без самопересечений, разбивает пространство на две области; одна из них — ограниченная (она называется внутренней), а другая (внешняя) — неограниченная.* Заметим, что (хотя об этом не говорится в формулировке теоремы) поверхность, вложенная в трехмерное пространство без самопересечений, должна быть двусторонней: ведь она разбивает пространство на две области, т. е. имеет «две стороны». Этим подтверждается то, что замкнутая односторонняя поверхность не может быть вложена в трехмерное пространство без самопересечений.

Задачи

117. Из точки s проведен луч l , не встречающий ребер поверхности Q (составленной из плоских многоугольников). Докажите, что s в том и только в том случае принадлежит внутренней области, определяемой поверхностью Q , если луч l пересекает Q в нечетном числе точек.

118. Существует ли в пространстве множество, являющееся совместной границей трех областей?

В конце п. 6, где шла речь о «плоской» теореме Жордана, мы отметили, что объединение простой замкнутой линии и ее внутренней области гомеоморфно кругу. Справедливо ли аналогичное (на первый взгляд столь же «очевидное») пространственное утверждение: «объединение поверхности, гомеоморфной сфере, и ее внутренней области гомеоморфно шару»? Для наиболее простых поверхностей, гомеоморфных сфере (например, для выпуклых многогранников), это, действительно, верно. Однако в общем случае это «очевидное» утверждение неверно — интуиция здесь обманывает нас. Иными словами, *существует в трехмерном пространстве такая поверхность, гомеоморфная сфере, что объединение этой поверхности и ее внутренней области не гомеоморфно шару*. Построение такой «дикий сферы» связано с работами французского математика Антуана и американского математика Александра.

Прежде чем переходить к описанию «дикий сферы», рассмотрим вопрос о *стягиваемости линий*. Пусть l — простая замкнутая линия, расположенная в фигуре M . Будем говорить, что линия l *стягиваема* в M , если существует поверхность P , гомеоморфная кругу и имеющая l своим краем, которая расположена (возможно, с самопересечениями) в фигуре M . Термин «стягиваемость» связан с тем, что на «круге» P можно начертить систему «концентрических окружностей», по которым l может быть постепенно стянута в точку.

Задачи

119. Докажите, что если M — открытый шар, то в M любая простая замкнутая линия стягиваема.

120. Докажите, что если фигура A состоит из конечного числа точек, то окружность l , расположенная в пространстве вне A , может быть вне A стянута.

Указание. Если взять шар, содержащий l , то, «вдавливая» его, чтобы получались идущие внутрь трубки, можно построить тело, гомеоморфное шару, которое содержит внутри себя линию l , но не содержит точек множества A .

Утверждение, содержащееся в задаче 120, можно пояснить следующим образом. Если l — окружность, не проходящая через точки фигуры A (состоящей из конечного числа точек), то круг, натянутый на l (даже если он проходит через некоторые из точек фигуры A), может быть при помощи небольшой деформации «снят» с тех точек фигуры A , через которые он проходил. Может создаться впечатление, что в

дополнительном пространстве произвольного нульмерного множества A любая окружность стягиваема. Однако это неверно.

Пример 31. Рассмотрим замкнутую цепь A_1 , состоящую из нескольких сцепленных звеньев (рис. 104, а).

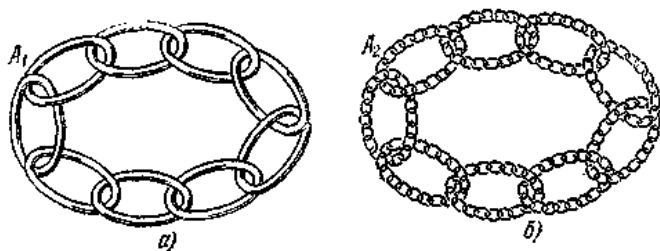


Рис. 104.

Далее, заменим каждое звено аналогичной замкнутой цепочкой из более мелких звеньев, проходящей внутри прежнего звена. Мы получим множество $A_2 \subset A_1$, состоящее из многих мелких звеньев (рис. 104, б). Теперь каждое звено, составляющее множество A_2 , заменим замкнутой цепочкой еще более мелких звеньшек (располагающейся внутри прежнего звена). Мы получим множество $A_3 \subset A_2$. Продолжая этот процесс, получим последовательность $A_1 \supset A_2 \supset A_3 \subset \dots$. Пересечение всех этих множеств представляет собой *антуановское множество* A^* .

Множество A^ нульмерно.* В самом деле, размеры звеньшек, составляющих множество A_n , неограниченно уменьшаются с ростом n , и потому в A^* не существует связных множеств, отличных от точки.

Пусть теперь l_1 — окружность, сквозь которую проходит первоначально взятая цепочка A_1 , и K_1 — круг с границей l_1 . На рис. 105, а круг K_1 пересекает тор T_1 , служащий поверхностью одного из звеньев цепи, по двум окружностям (меридианам); одна из них обозначена через l_2 .

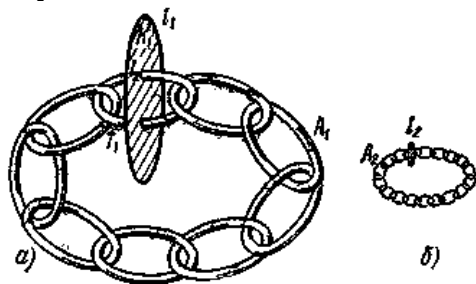


Рис. 105.

Часть круга K_1 , ограниченная линией l_2 , представляет собой меньший круг K_2 . Этот меньший круг по отношению к части множества A_2 , расположенной внутри тора T_1 , находится в таком же положении, что и круг K_1 по отношению ко всему множеству A_1 (рис. 105, б). Можно себе представить, что K_2 (а потому и K_1) пересекает один из торов (служащих границами звеньев множества A_2) по двум меридианам; один из них обозначим через l_3 и т. д. Таким образом, круг K_1 («пленка», натянутая на l_1) пересекается и с множеством A_1 , и с A_2 , и с A_3 и т. д., откуда вытекает, что K_1 пересекается и с предельным множеством A^* . Разумеется, на l_1 можно натянуть не только K_1 , но и иную пленку, являющуюся непрерывным образом круга; однако проведенное рассуждение поясняет, что любая такая пленка пересекается с A^* (путь доказательства этого факта обсуждается в п. 25). Итак, *окружность l_1 не стягиваема во внешнем пространстве антуановского множества A^** . Это нульмерное, как бы рассыпающееся на отдельные точки множество «мешает» провести пленку (являющуюся непрерывным образом круга), натянутую на l_1 .

Задачи

- 121. Докажите, что на l_1 можно натянуть пленку, гомеоморфную ручке и расположенную вне A^* .
- 122. Постройте простую замкнутую линию, проходящую через все точки антуановского множества.

Пример 32. Теперь мы в состоянии дать описание «дикой сферы». Пусть S — сфера, содержащая внутри себя множество A_1 и окружность l_1 (рис. 105, а). Продавим сферу S в нескольких местах, чтобы получились трубки, идущие внутрь и подходящие к каждому из торов, ограничивающих множество A_1 (рис. 106).

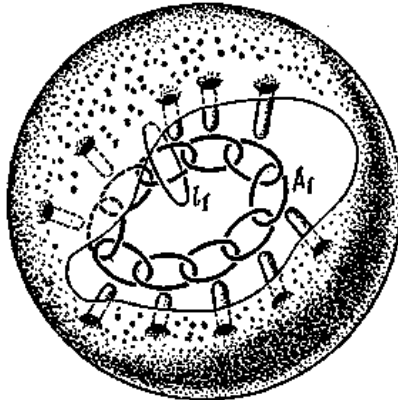


Рис. 106.

Мы можем это сделать так, чтобы трубки не подходили к окружности l_1 . Полученная поверхность S_1 (сфера со вдавленными трубками) гомеоморфна сфере, а ее внутренняя область U_1 гомеоморфна открытому шару, причем линия l_1 лежит внутри U_1 . Теперь от концов трубок вдавним внутрь более узкие трубки, идущие внутри торов, ограничивающих множество A_1 и подходящие к звеньям множества A_2 . Мы получим поверхность S_2 , также гомеоморфную сфере, причем ее внутренняя область U_2 гомеоморфна открытому шару и содержит l_1 . Затем мы вдавним еще более узкие трубки, подходящие к звеньям множества A_3 , и т. д. На каждом шаге получается поверхность S_n , гомеоморфная сфере, причем ее внутренняя область U_n гомеоморфна открытому шару и содержит l_1 . Постепенно трубки становятся все короче, поверхность видоизменяется все меньше, и поэтому предельная поверхность S^* остается гомеоморфной сфере. Так как при этом «щупальцы», выпускаемые поверхностями S_n , все ближе подходят к A^* , то предельная сфера S^* содержит множество A^* . Внутренняя область U^* сферы S^* не пересекается с антуановским множеством A^* (поскольку это множество лежит на границе S^* области U^*), т. е. вся область U^* целиком лежит вне A^* . Поэтому окружность l_1 , расположенная внутри U^* , не стягиваема в U^* (поскольку линия l_1 не стягиваема вне A^*). Из этого вытекает, что область U^* не гомеоморфна открытому шару (задача 119). Поэтому объединение поверхности S^* и ее внутренней области не гомеоморфно замкнутому шару.

2.2.10. Узлы

Если концы «нити», на которой завязан узел, не соединены, то узел можно «развязать». Поэтому в топологии рассматривают узлы только на замкнутых линиях.

Пример 33. На рис. 107 изображен *простой узел* (иногда его называют также «трилистником»).

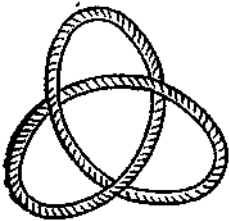


Рис. 107.

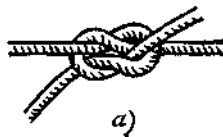


Рис. 108.

Пример 34. Обычный двойной узел (рис. 108, а) не следует смешивать с так называемым *морским узлом* (рис. 108, б); узел на рис. 108, а моряки пренебрежительно именуют «бабушкиным» узлом (он легче развязывается). На рис. 109 даны топологические схемы этих узлов. С топологической точки зрения *узел — это линия в трехмерном пространстве, гомеоморфная окружности*.

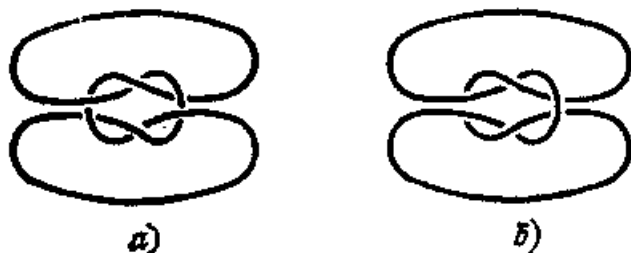


Рис. 109.

Разными узлами считаются такие, которые *неизотопны*. Представляется, например, наглядно очевидным, что «заузленная» и «незаузленная» линии на рис. 110 топологически неодинаково расположены (неизотопны).

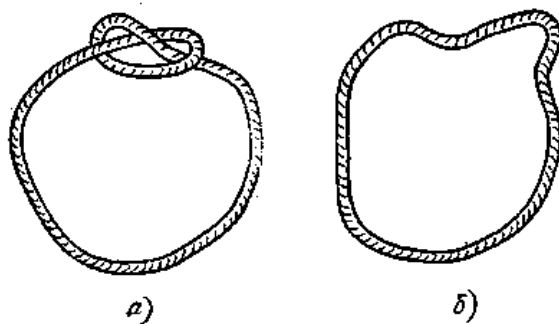


Рис. 110.

Доказательство этого факта мы рассмотрим позже. Будем представлять себе узел реализованным в виде простой замкнутой ломаной l в пространстве и спроектируем ее на «горизонтальную» плоскость. Проекция узла l может оказаться пересекающей себя; при этом мы можем предполагать (немного сдвинув, если нужно, некоторые из звеньев), что проекция имеет лишь *двойные* точки пересечения (т. е. проекции никаких трех звеньев не имеют общей точки). Мы условимся на чертежах прерывать то из двух звеньев, пересекающихся на проекции, которое проходит ниже. В

результате получится наглядный рисунок, так называемая нормальная проекция узла (рис. 111, а).

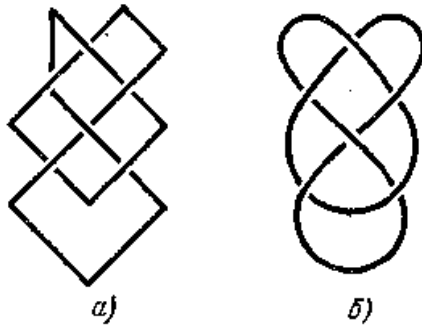


Рис. 111

Можно изображать узлы и в виде плавных линий на плоскости (без угловых точек), сохраняя те же соглашения о сплошных и прерванных участках (рис. 111,б).

Задачи

123. Докажите, что всякий целочисленный одномерный цикл можно представить в виде объединения нескольких *ориентированных узлов*, т. е. направленных контуров в пространстве, которые могут иметь общие вершины, но других общих точек попарно не имеют.

124. Докажите, что трижды перекрученная лента (рис. 112) гомеоморфна ленте Мёбиуса, а ее край изотопен простому узлу.

125. Докажите, что замысловатая «пряжка», изображенная на рис. 113, гомеоморфна ручке, а ее край изотопен простому узлу.



Рис. 112.

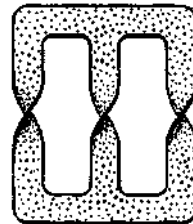


Рис. 113.

126. Докажите, что любая не гомеоморфная кругу поверхность (ориентируемая или неориентируемая), край которой гомеоморфен окружности, может быть так вложена в трехмерное пространство, что ее край будет простым узлом.

В связи с задачами 124—126 возникает вопрос, для *любого ли узла L существует «натянутая на него пленка», т. е. поверхность (без самопересечений), имеющая L своим краем?* Утвердительный ответ на этот вопрос дал Ф. Франкль (советский математик, приехавший к нам в 1934 году из Австрии) при помощи следующего изящного рассуждения. Нормальная проекция узла L разбивает плоскость на области, причем эту «карту» можно раскрасить в два цвета, например, белый и красный. Возможность такой «шахматной» раскраски вытекает из задачи 96, поскольку в каждой вершине получающегося в проекции графа сходятся четыре ребра. Мы можем при этом считать, что наружная (неограниченная) «страна» окрашена в белый цвет; на рис. 114, *а* контур узла показан без перерывов, чтобы ясно были видны «страны».

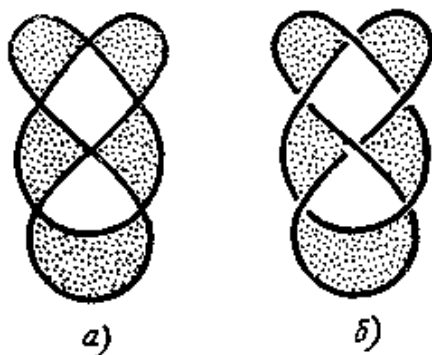


Рис. 114.

Если теперь вновь наметить на проекции узла L разрывы, как на нормальной проекции, то чертеж станет пространственным, т. е. «красные» области окажутся соединенными с перекручиваниями (рис 114, *б*). Это и дает требуемую поверхность, имеющую узел L своим краем. Заметим, что получаемая поверхность, вообще говоря, неориентируема (см. рис. 112). Однако более «осторожное» приклеивание стран позволяет для любого узла построить ориентируемую поверхность, имеющую этот узел своим краем (см. задачи 130—132).

Задачи

127. Для каждого из узлов на рис. 115 изобразите указанным способом поверхность, имеющую этот узел своим краем.

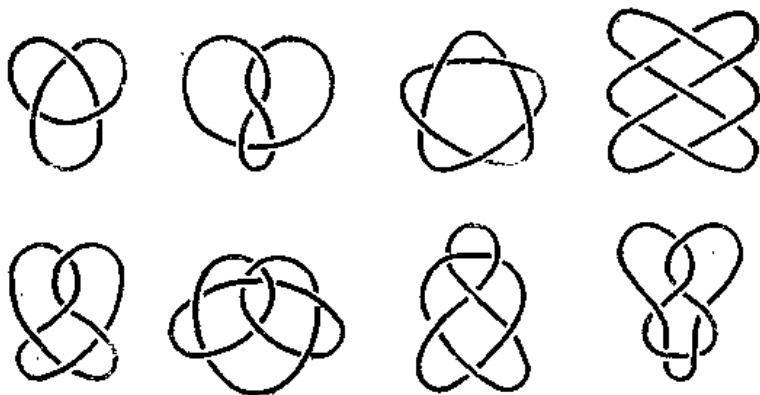


Рис. 115.

Какой из поверхностей, P_k или N_q , с вырезанной в ней круглой дырой гомеоморфны эти поверхности?

128. Докажите, что поверхность Франкля в том и только в том случае неориентируема, если существует простая замкнутая линия, проходящая по «красным» областям и нечетное число раз переходящая из страны в страну через двойные точки (рис. 116).



Рис. 116.

129. Докажите, что если имеется *переплетение*, т. е. объединение нескольких простых замкнутых линий в пространстве (рис. 117), которые попарно не пересекаются друг с другом, то существует связная поверхность без самопересечений, имеющая своим краем это переплетение.

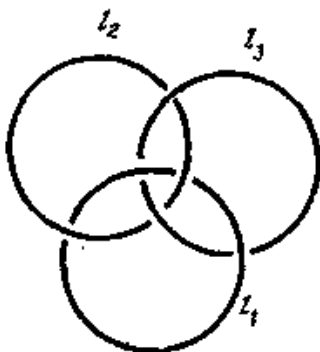


Рис 117.

130. Пусть L — переплетение. Выберем на каждой из его линий направление обхода и обозначим через z одномерный целочисленный цикл, являющийся нормальной проекцией переплетения L . Фиксируем во внешней области точку o и каждой стране M , на которые цикл z разбивает плоскость, припишем число $k(M)$, равное индексу пересечения $J(x, z)$, где x — направленная ломаная, идущая от точки o к некоторой внутренней точке страны M (рис. 118).

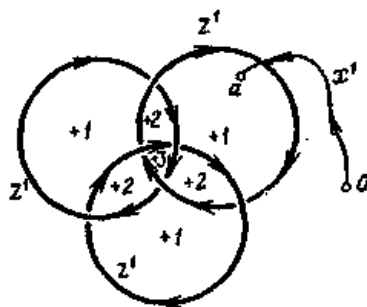


Рис. 118.

Докажите, что если M_I — страна, для которой $|k(M)|$ принимает наибольшее значение, то те части цикла z , которые составляют границу страны M_I , имеют направления, образующие обход контура страны M_I (по или против часовой стрелки).

131. В условиях задачи 130 выбросим из цикла z участки, составляющие контур страны M_I (с некоторым запасом) и вместо них добавим «перемычки» (пунктир на рис. 119, а), в результате чего получится цикл z' , имеющий меньшее число стран, чем z .

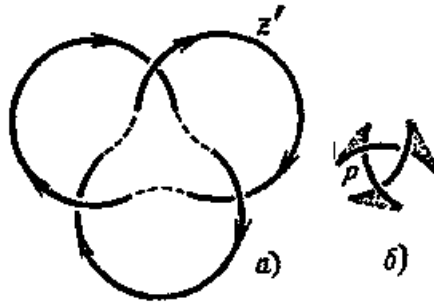


Рис. 119,

Аналогичное построение проведем и на переплетении L ; получится переплетение L' , проектирующееся в цикл z' . Выброшенные участки дополним такими же перемычками и натянем на полученный контур поверхность P , гомеоморфную кругу и имеющую вблизи перемычек перекрученные «лопасти» (см. рис. 119, б). Докажите, что если на L' натянута ориентируемая пленка Q , то при подклеивании к Q поверхности P (по перемычкам) мы получаем также ориентируемую поверхность, натянутую на цикл z .

132. Пользуясь результатами двух последних задач, докажите, что если L — переплетение в пространстве, на каждом контуре которого выбрано произвольное направление обхода, то существует такая ориентируемая поверхность, краем которой является переплетение L , что некоторая не ориентация согласована с направлением обхода на каждой из линий переплетения L .

Указание. Надо следить за тем, чтобы на каждом шаге построения весь край получающейся поверхности был виден сверху.

133. Докажите, что если для цикла z наибольшее значение числа $|k(M)|$, рассмотренного в задаче 130, равно n , то для поверхности, натянутой на переплетение L по методу задач 130—132, имеем $\chi(Q) \geq n - q$, где q — число двойных точек цикла z .

134. Докажите, что если переплетение L содержит l компонент, причем число двойных точек на его нормальной проекции z равно q , а наибольшее значение числа $|k(M)|$ для цикла z равно n , то на это переплетение можно натянуть пленку, гомеоморфную сфере с k ручками, в которой вырезано l круглых дыр, где

$$k \leq 1 + \frac{q - n - l}{2}.$$

135. Докажите, что на морской (а также на «бабушкин») узел можно натянуть пленку, гомеоморфную сфере с тремя дырами, две из которых заклеены ручками.

2.2.11. Коэффициент зацепления

Для двух не пересекающихся друг с другом ориентированных контуров x , y в пространстве (x — первый контур, y — второй) можно следующим образом определить некоторое целое число, называемое *коэффициентом зацепления* этих контуров. Рассмотрим нормальную проекцию переплетения $x \cup y$ на некоторую («горизонтальную») плоскость и пусть a — двойная точка на этой проекции, в которой контур x идет ниже, чем y . Если, двигаясь вблизи a по направлению контура x , мы увидим (в проекции), что y пересекает его слева направо (рис. 120, а), то точке a припишем число $+1$, а если справа налево (рис. 120, б), то -1 .

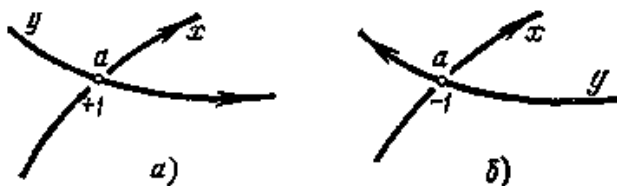


Рис. 120.

В остальных случаях (т. е. если пересекаются два участка одного и того же контура или если контур x проходит выше, чем y) двойной точке a припишем число 0. Сумма этих чисел по всем двойным точкам на проекции называется *коэффициентом зацепления* и обозначается через $w(x, y)$.

Пример 35. Для двух соседних звеньев обычной металлической цепи (рис. 121, а) коэффициент зацепления равен ± 1 (рис. 121, б).

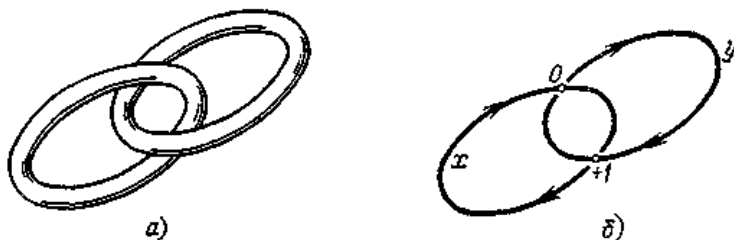


Рис. 121.

Для контуров, изображенных на рис. 122, имеем $w(x, y) = 3$.

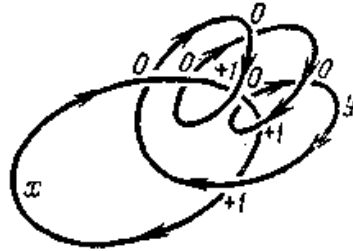


Рис. 122.

Как мы увидим дальше, коэффициент зацепления $w(x, y)$ зависит лишь от расположения самих контуров x, y , а не от способа проектирования. Далее, если контуры x, y непрерывно деформируются в пространстве (например, движутся, как шарнирные ломаные), оставаясь в каждый момент не пересекающимися, то их коэффициент зацепления $w(x, y)$ не меняется. Наконец, отметим, что коэффициент зацепления $w(x, y)$ является (с точностью до знака) *изотопическим инвариантом*. Иначе говоря, если f — гомеоморфное отображение трехмерного пространства на себя, то при этом отображения x, y переходят в такие контуры $f(x), f(y)$, что $w(f(x), f(y)) = \pm w(x, y)$.

Пример 36. Ранее упоминалось, что хотя дважды перекрученная лента гомеоморфна неперекрученной (см. рис. 8) эти фигуры не изотопны друг другу в пространстве. Теперь мы можем это доказать. В самом деле, коэффициент зацепления краев ленты равен в случае дважды перекрученной ленты ± 1 (в зависимости от того, в какую сторону перекручена лента), а в случае неперекрученной ленты — нулю (рис. 123).

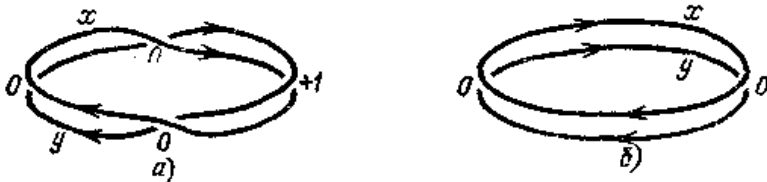


Рис. 123.

Поэтому при гомеоморфном отображении пространства на себя дважды перекрученная лента не может перейти в неперекрученную. Дважды перекрученная лента не может быть превращена в неперекрученную, как бы мы ни деформировали ее (оставляя

гомеоморфной себе); ведь при такой деформации края ленты перемещаются, не пересекаясь друг с другом, и потому коэффициент зацепления измениться не может.

Пример 37. Постоянный ток I , протекающий по бесконечному прямолинейному проводу P , создает магнитное поле, напряженность которого на расстоянии r от провода имеет величину $H \approx \frac{2I}{r}$. Как известно, потенциалом магнитного поля называется работа, которую надо затратить, чтобы из некоторой фиксированной точки x_0 (точки нулевого потенциала) переместить магнитный полюс, равный единице, в заданную точку поля. В рассматриваемом случае потенциал W магнитного поля многозначен. Действительно, если из точки x_0 перемещать магнитный полюс в точку a по двум путям, показанным на рис. 124, a , b , то первое перемещение требует дополнительной работы против силы $\frac{2I}{r}$ на пути $2\pi r$, т. е. дополнительной работы, численно равной $4\pi I$.

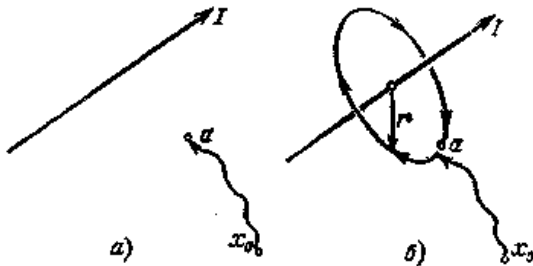


Рис. 124.

Мы видим, что один обход вокруг провода (не обязательно по окружности — можно идти по любому пути) меняет магнитный потенциал $W(a)$ на величину $4\pi I$. Вообще, после m обходов вокруг провода, где m — целое число (положительное, отрицательное или нуль), потенциал изменится на $4\pi Im$. То же выражение для изменения потенциала справедливо в случае любого (не обязательно прямолинейного) провода (рис. 125).

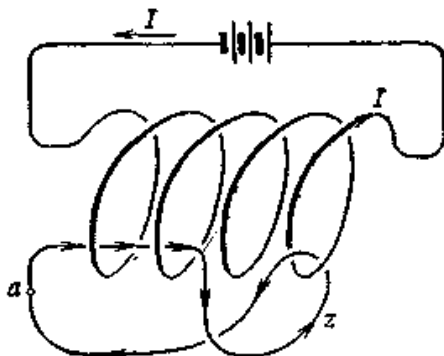


Рис. 125

Число обходов («витков») пути z вокруг проводника P равно взятому со знаком минус коэффициенту зацепления проводника P с путем z , т. е. при обходе вокруг проводника P по замкнутому пути z магнитный потенциал меняется на величину $4\pi Im$, где $m = -w(P, z)$. Величину Im иногда называют «числом ампервитков» (если ток I измеряется в амперах).

Задачи

136. Докажите, что при перестановке контуров их коэффициент зацепления не меняется; $w(x, y) = w(y, x)$.

137. Направленные контуры $x' y'$ симметричны контурам x, y относительно некоторой плоскости (учитывая направление обхода). Докажите, что $w(x', y') = -w(x, y)$.

138. Каков коэффициент зацепления края ленты Мёбиуса (см. рис. 50, a) и его средней линии?

139. Пусть Q — поверхность, гомеоморфная ленте Мёбиуса (рис. 126), x — ее край, а y — «средняя линия» (т. е. образ средней линии ленты Мёбиуса M , изображенного на рис. 50, z , при гомеоморфизме $f: M \rightarrow Q$). Докажите, что число $w(x, y)$ нечетно.

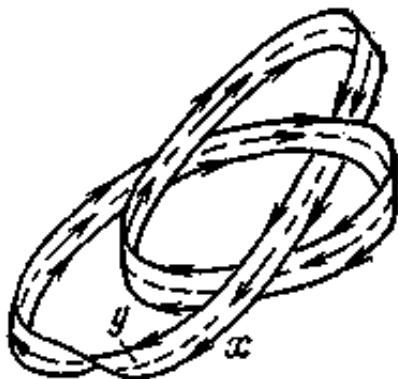


Рис. 126

Теперь мы дадим другое (эквивалентное) определение коэффициента зацепления. Будем представлять себе контуры x , y лежащими «почти целиком» в плоскости нормальной проекции, так что лишь вблизи двойных точек один из них проходит чуть ниже другого. Далее, рассмотрим ориентированную пленку Q , натянутую на контур y (так, как это было описано в задаче 132: контур y виден целиком, если смотреть на пленку Q сверху), причем будем считать, что эта пленка «провисает», располагаясь вблизи своего края почти вертикально (рис. 127).

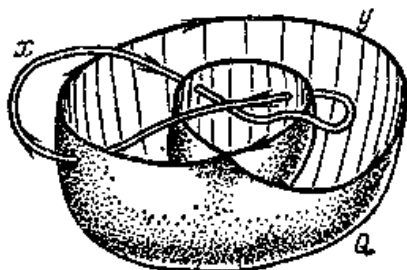


Рис. 127.

Тогда в тех двойных точках, в которых контур x идет выше y , он проходит и выше пленки Q , т. е. не пересекает ее. В тех же точках, где контур x проходит ниже y , он пересекает пленку Q ; при этом рассматриваемый участок контура x имеет с пленкой Q индекс пересечения $+1$, если, глядя по направлению линии x , мы видим, что

линия y пересекает ее слева направо (рис. 128, а) и -1 , если пересечение происходит справа налево (рис. 128, б).

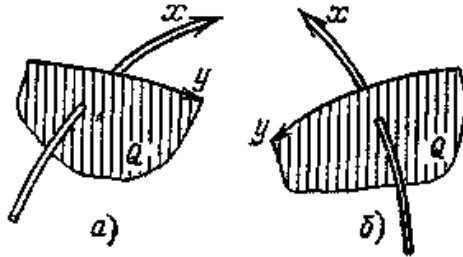


Рис. 128.

Из этого следует (если сравнить определение коэффициента зацепления и индекса пересечения), что справедливо равенство

$$w(x, y) = J(x, Q), \quad (20)$$

где Q — двумерная ориентируемая пленка, натянутая на контур y и согласованно с ним ориентированная.

Равенство (20) останется справедливым, если взять любую пленку Q (не обязательно построенную так, как это было сделано в задаче 132). В самом деле, пусть имеются две различные ориентируемые пленки Q, Q' , натянутые на контур y и ориентированные согласованно с ним. Рассмотрим разность пленок Q и Q' , т. е. объединение пленки Q с имеющейся на ней ориентацией и пленки Q' с противоположной ориентацией. Эта разность является двумерным целочисленным циклом (даже если Q и Q' пересекаются). Так как индекс пересечения целочисленного цикла x с этим двумерным циклом равен нулю, то $J(x, Q) = J(x, Q')$.

Из равенства (20) следует, что коэффициент зацепления (первоначально определенный с помощью нормальной проекции) не зависит от выбора плоскости проекции. Из (20) вытекают также и другие упоминавшиеся выше свойства коэффициента зацепления.

Задачи

140. Пленка, натянутая на контур y (и находящаяся в общем положении с контуром x), пересекается с x в единственной точке. Докажите, что пленка, натянутая на контур x , пересекается с y .

141. Пусть x, y — ориентированные контуры в пространстве. Из формулы (20) вытекает, что если существует ориентированная пленка, натянутая на y и не пересекающаяся с x (рис. 129), то $w(x, y) = 0$. Докажите обратную теорему.

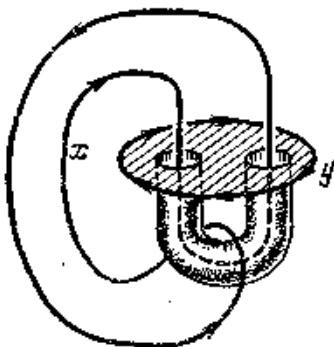


Рис. 129.

142. Докажите, что если $w(x, y)$ есть четное число, то существует пленка (возможно, не ориентируемая), натянутая на y и не имеющая общих точек с x .

143. Убедитесь, что три контура на рис. 117 имеют попарно коэффициенты зацепления, равные нулю. Постройте пленку, гомеоморфную ручке, которая натянута на один из этих контуров и не пересекается с двумя другими.

2.3. Гомотопии и гомологии

2.3.1. Периоды многозначных функций

Пусть h — некоторый путь в фигуре X , идущий от начальной точки x_0 до конечной точки x_1 . Иначе говоря, $h: [0; 1] \rightarrow X$ есть непрерывное отображение, удовлетворяющее условиям $h(0) = x_0$, $h(1) = x_1$. Будем этот путь *непрерывно деформировать* в фигуре X , оставляя концевые точки x_0 и x_1 неподвижными. На рис. 130 положения деформируемого пути изображены тонкими линиями.

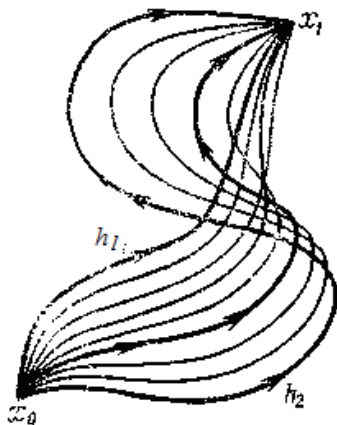


Рис. 130

Мы всегда будем рассматривать только такие деформации путей, при которых конечные точки не смещаются.

Два пути h_1, h_2 в фигуре X , имеющие одни и те же конечные точки, называются *гомотопными* в этой фигуре, если при помощи деформации (происходящей в фигуре X) h_1 может быть превращен в h_2 ; гомотопность путей обозначается записью $h_1 \sim h_2$.

Пример 38. В круге любые два пути, имеющие общие концы, гомотопны между собой. Наглядно это можно пояснить, вообразив, что путь представляет собой растянутую резиновую нить, извилистым образом расположенную внутри круга. Если мы отпустим резинку, закрепив конечные точки x_0, x_1 и позволив нити свободно перемещаться, то она начнет деформироваться и, сжимаясь, расположится по прямолинейному отрезку, соединяющему x_0 и x_1 . Таким образом, любой путь в круге гомотопен отрезку, соединяющему конечные точки x_0 и x_1 , и потому любые два пути, соединяющие x_0 и x_1 гомотопны.

Пример 39. Обозначим через X кольцо, ограниченное двумя окружностями с общим центром o . Выберем некоторую точку $x_0 \in X$ и для любой точки $x \in X$ обозначим через $\varphi(x)$ величину угла x_0ox (рис. 131).

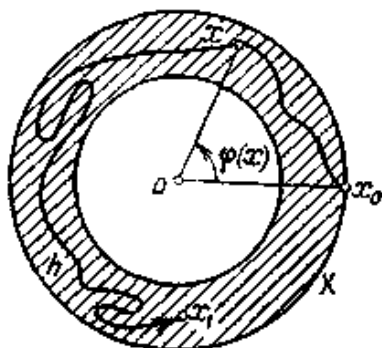


Рис. 131

Функция $\varphi(x)$ *многозначна* (она определена с точностью до слагаемого вида $2k\pi$, где k — целое число). Предположим, что в точке $x = x_0$ мы выбрали какое-либо одно значение φ_0 этой функции. При перемещении точки x в кольце X угол $\varphi(x)$ будет непрерывно меняться. Поэтому каждому пути h , ведущему в кольце X от точки x_0 к точке x_1 , соответствует вполне определенное значение φ_1 многозначной функции $\varphi(x)$, к которому мы приходим, взяв в точке x_0 значение φ_0 этой функции и непрерывно перемещаясь вдоль пути h от x_0 до x_1 . При этом гомотопным путям, ведущим в кольце X из точки x_0 в точку x_1 , соответствует одной то же значение функции. В самом деле, значение функции, к которому мы приходим, проходя путь h , будет при непрерывной деформации пути h само непрерывно меняться. Следовательно, это значение должно оставаться постоянным: непрерывно меняясь, оно не может «перескочить» от одного из возможных значений функции $\varphi(x)$ в точке x_1 к другому (отличающемуся от него на $2k\pi$).

Пример 40. Если в примере фиксировать точку нулевого потенциала x_0 , то во внешней области проводника P определится многозначная функция $W(x)$ (магнитный потенциал). Переместившись из точки x_0 в точку x_1 по некоторому пути h , мы придем ко вполне определенному значению магнитного потенциала в точке x_1 . Гомотопные между собой пути, ведущие из x_0 в x_1 , дадут в точке заодно и то же значение магнитного потенциала, а негомотопные могут привести к различным значениям функции $W(x)$ в точке x_1 .

Задачи

144. На множестве X (рис. 132) постройте многозначную функцию, принимающую в точке x_0 бесконечное множество значений, среди которых имеются значения $0, 1, \sqrt{5}$.

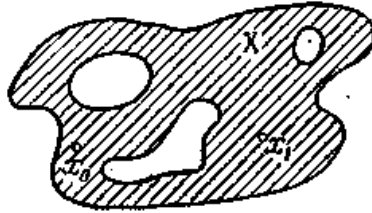


Рис. 132

На множестве X (рис. 132) постройте такую многозначную функцию, что два негомотопных пути, ведущих из x_0 к x_1 приводят к одному и тому же значению в точке x_1 .

Пример 41. На фигуре X (рис. 133) рассмотрим функцию $f(x) = \varphi_1(x) + \sqrt{2}\varphi_2(x) - \sqrt{3}\varphi_3(x)$, где $\varphi_1(x)$, $\varphi_2(x)$, $\varphi_3(x)$ — величины углов a_1o_1x , a_2o_2x , a_3o_3x .

Функция $f(x)$ многозначна. Если мы, начав движение из точки x_0 , пройдем путь h_1 (рис. 133), то (при возвращении в точку x_0) к значению функции $\varphi_1(x_0)$ прибавится 2π , а значения функций $\varphi_2(x_0)$ и $\varphi_3(x_0)$ не изменятся; следовательно, к значению функции $f(x)$ в точке x_0 прибавится 2π .

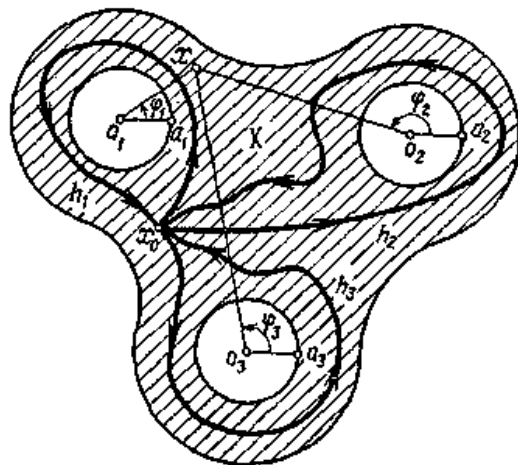


Рис. 133.

При обходе же пути h_2 к значению функции в точке x_0 прибавится $2\pi\sqrt{2}$, а при обходе пути h_3 прибавится $-2\pi\sqrt{3}$. Числа 2π , $2\pi\sqrt{2}$, $-2\pi\sqrt{3}$ можно назвать *периодами* функции $f(x)$ на множестве X соответствующими замкнутым путям h_1 , h_2 , h_3 .

Рассмотрим теперь путь, получающийся, если сначала пройти h_1 , а затем h_2 ; этот путь обозначается через h_1h_2 и называется *произведением* путей h_1 и h_2 . Ясно, что при прохождении пути h_1h_2 к значению функции $f(x)$ прибавится $2\pi + 2\pi\sqrt{2}$. Аналогично, при прохождении пути h_3h_1 к значению функции $f(x)$ прибавится $-2\pi\sqrt{3} + 2\pi$. Вообще, при перемножении двух путей (начинающихся и кончающихся в точке x_0) соответствующие этим путям периоды функции $f(x)$ складываются.

Так как значение функции $f(x)$, получаемое в результате прохождения некоторого пути, не изменяется при гомотопии этого пути, то мы можем не различать гомотопные пути. Иначе говоря, можно рассматривать не сами пути (начинающиеся и кончающиеся в точке x_0), а *классы путей*, объединяя в один класс все гомотопные между собой пути. Класс всех путей, гомотопных пути h , будем обозначать через $[h]$, а множество всех таких классов — через $\pi(X)$. Эти классы можно перемножать: берем путь h , принадлежащий первому классу, путь k , принадлежащий второму классу, и перемножаем их; тогда класс, который содержит путь hk , и называется *произведением* двух взятых классов: $[h] \cdot [k] = [hk]$.

Смысл введения классов путей понятен: каждому классу соответствует некоторый период многозначной функции $f(x)$, а при перемножении двух классов соответствующие периоды складываются.

2.3.2. Фундаментальная группа

Классы гомотопных путей и произведение этих классов можно рассмотреть для любой фигуры X . Будем рассматривать только такие пути в X , которые начинаются и кончаются в фиксированной точке $x_0 \in X$. Любые два из таких путей можно перемножить. Будем рассматривать *классы путей*, объединяя в один класс все гомотопные между собой пути. Если a — некоторый класс и h — какой-либо путь, принадлежащий этому классу, то будем говорить, что h — *представитель* класса a и писать $a = [h]$. Множество всех классов обозначим через $\pi(X)$. *Умножение* классов определим так же, как и в предыдущем пункте (см. рис. 133): если a и b — два класса путей (начинающихся и кончающихся в точке x_0), а h и k — какие-либо их представители, т. е. $a = [h]$, $b = [k]$, то класс, представителем которого является путь hk , мы объявляем *произведением* классов a и b , т. е. $ab = [hk]$. Заметим, что если вместо h и k взять другие представители h' и k' рассматриваемых классов a и b , то мы получим путь $h'k'$, гомотопный пути hk , т. е. определяющий тот же самый класс: $[h'k'] = [hk]$. Таким образом, произведение двух классов определяется именно этими классами, а от выбора представителей не зависит. Оказывается, что *относительно введенной операции умножения множество $\pi(X)$ является группой*.

Укажем вкратце, как это устанавливается. Если h — какой-либо путь, принадлежащий классу a , а q — путь, который может быть стянут в точку, то $qh \sim h$ (рис. 134) и $hq \sim h$.

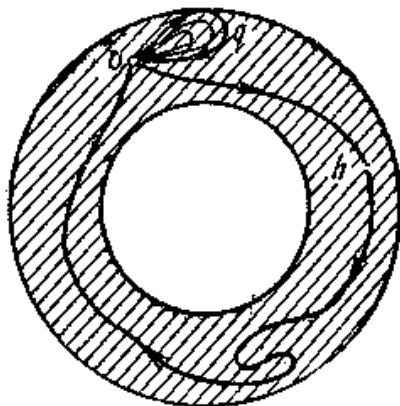


Рис. 134.

Поэтому, обозначая символом 1 класс всех путей, стягиваемых в точку, мы получаем $1a = a, a1 = a$ для любого класса $a \in \pi(X)$, т. е. класс 1 является *единицей* относительно умножения, введенного в $\pi(X)$.

Далее, если a — какой-либо класс и h — его представитель, то обозначим через h^{-1} путь h , пробегаемый в обратном направлении (рис. 135).



Рис. 135.

Тогда каждый из путей hh^{-1} и $h^{-1}h$ может быть стянут в точку (на рис. 136 показано стягивание в точку пути hh^{-1}).

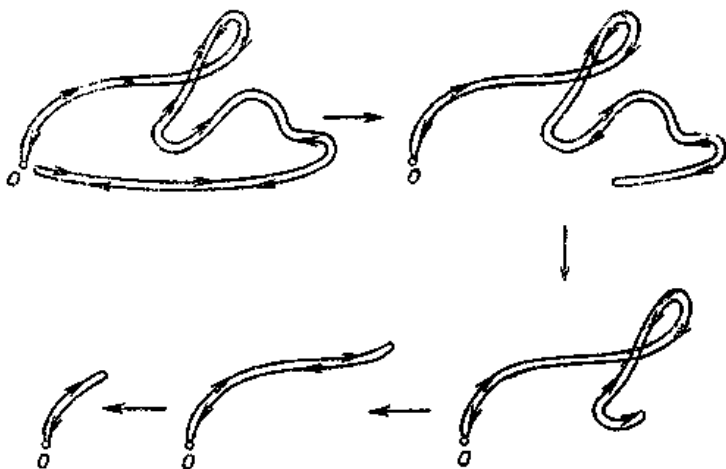


Рис. 136.

Поэтому, обозначив через a^{-1} класс, которому принадлежит путь h^{-1} , мы найдем, что $aa^{-1} = 1$, $a^{-1}a = 1$, т. е. в $\pi(X)$ для каждого элемента a существует обратный.

Несложно доказывается, что умножение в $\pi(X)$ ассоциативно. Таким образом, множество $\pi(X)$ есть группа. Она называется *фундаментальной группой* фигуры X (построенной в точке x_0).

Можно показать (см. задачу 148), что если любые две точки могут быть соединены путем в фигуре X , то фундаментальные группы фигуры X , построенные в разных точках x_0 и x'_0 , изоморфны. В этом случае (который мы только и будем рассматривать) можно просто говорить о фундаментальной группе фигуры X , не указывая, в какой точке она построена. *Фундаментальная группа является топологическим инвариантом*, т. е. если фигуры X и Y гомеоморфны, то их фундаментальные группы $\pi(X)$ и $\pi(Y)$ изоморфны. Заслуга открытия и изучения этого топологического инварианта принадлежит Пуанкаре.

Задачи

146. Если группа $\pi(X)$ тривиальна (т. е. состоит только из единичного элемента), то фигура X называется *односвязной*. Иначе говоря, фигура X односвязна, если любой замкнутый путь в X может быть стянут в точку. Докажите, что любая выпуклая фигура (в частности, прямая, плоскость, отрезок, круг, шар, выпуклый многоугольник или многогранник) односвязны.

147. Докажите, что сфера односвязна.

Указание: любой путь (даже заполняющий всю сферу, подобно кривой Пеано) может быть деформирован в «гладкий путь», не покрывающий всю сферу.

148. Пусть w — путь, соединяющий две точки x_0, x'_0 фигуры X . Каждому замкнутому пути h с начальной точкой x_0 поставим в соответствие путь $h^* = w^{-1}hw$ с начальной точкой x'_0 (рис. 137).

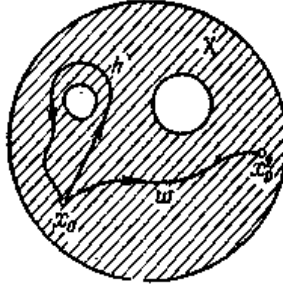


Рис. 137.

Докажите, что этим определяется изоморфизм фундаментальной группы фигуры X , построенной в точке x_0 , и группы, построенной в точке x'_0 .

Пример 42. Покажем, что *фундаментальная группа окружности является свободной циклической, т. е. изоморфна аддитивной группе целых чисел.*

В самом деле, обозначим путь, равномерно обходящий окружность B в некотором «положительном» направлении, через a , а обратный путь — через a^{-1} . Тогда a^n будет обозначать путь, $|n|$ раз обходящий окружность: в «положительном» направлении, если $n > 0$, и в «отрицательном», если $n < 0$ (путь a^0 оставляет точку покоящейся в начальной точке x_0).

Любому пути можно поставить в соответствие некоторый график: положение точки, пробегающей путь, задается значением параметра t (например, времени) на единичном отрезке $0 \leq t \leq 1$. С другой стороны, этому же положению точки соответствует ее угловая координата φ на окружности B (отсчитываемая от начальной точки x_0). Откладывая по оси абсцисс t , а по оси ординат угол φ , получим график зависимости $\varphi(t)$ (причем $\varphi(0)=0$). Если точка, движущаяся равномерно, обходит окружность n раз, мы получаем путь a^n ; его график — прямолинейный отрезок, соединяющий точки $(0; 0)$ и $(1; 2\pi)$. Однако точка может двигаться по окружности B , многократно

изменяя направление движения. На рис. 138,*а* показан график пути, схематически изображенного на рис. 138, б.

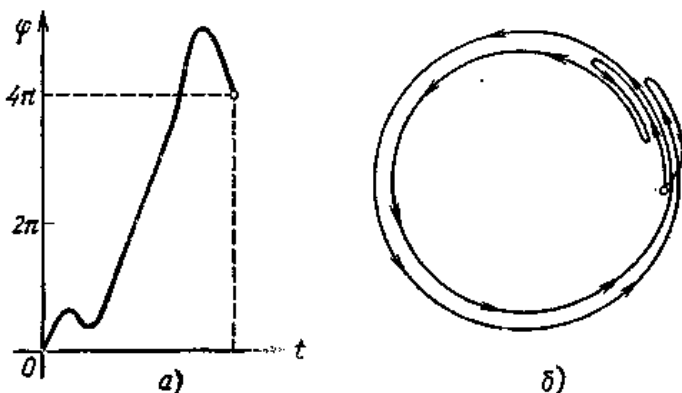


Рис 138.

Но каким бы ни был замкнутый путь на окружности, его график всегда соединяет точку $(0; 0)$ с точкой $(1; 2n\pi)$, где n — некоторое целое число: ведь, пройдя этот путь, мы возвращаемся в точку x_0 , угловая координата которой является числом, кратным 2π . Число n называется *числом обходов по окружности*.

Любой путь f , совершающий n обходов, гомотопен пути a^n : начертив на одном рисунке графики путей f и a^n , заставим каждую точку первого графика перемещаться параллельно оси ординат до графика пути a^n . Если такое перемещение производить одновременно для всех точек (рис. 139), то мы продеформируем график пути f в отрезок, являющийся графиком пути a^n .

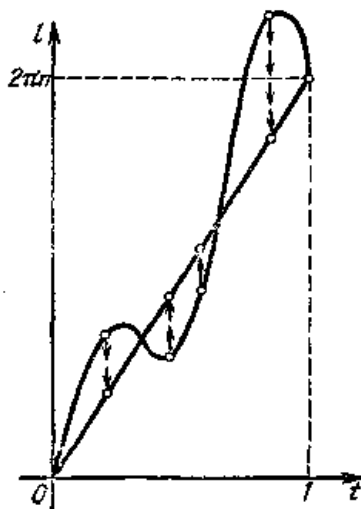


Рис 139.

Но если график деформируется, то и сам путь деформируется, откуда и следует, что пути f и a^n гомотопны. Значит, все пути, совершающие n обходов по окружности, гомотопны пути a^n , т. е. принадлежат одному классу путей. Пути же, для которых число обходов различно, не гомотопны между собой.

Итак, элементы фундаментальной группы окружности B находятся во взаимно однозначном соответствии с целыми числами. Так как при перемножении путей их числа обходов, очевидно, складываются, то из этого следует, что группа $\pi(B)$ изоморфна аддитивной группе целых чисел.

Задачи

149. Докажите, что фундаментальная группа кругового кольца является свободной циклической группой.

150. Фигура X представляет собой плоскость, из которой выколота (удалена) одна точка. Докажите, что группа $\pi(X)$ — свободная циклическая.

151. Докажите, что внутренняя область простой замкнутой линии l односвязна. Если же область G имеет границу, состоящую более чем из одного замкнутого контура (см. рис. 132), то G не односвязна.

2.3.3. Клеточные разбиения и полиэдры

Мы часто рассматривали поверхность Q , на которой начерчен граф G , разбивающий ее на части, гомеоморфные кругу. Это пример *клеточного разбиения*. Поверхность Q представляется в виде объединения попарно не пересекающихся *клеток*: *нульмерных*, *одномерных* и *двумерных*. Нульмерными клетками являются точки — вершины графа G . Одномерные клетки — ребра графа G (без концов). Каждая одномерная клетка гомеоморфна открытому отрезку (без концов). Двумерные клетки — куски поверхности, на которые она распадается, если ее разрезать по ребрам графа G . Каждая двумерная клетка гомеоморфна открытому кругу.

Можно также рассматривать клеточные разбиения, в которых к некоторому ребру (одномерной клетке) примыкают три, четыре или большее число двумерных клеток, а не обязательно две или одна, как было в случае поверхности с краем. К некоторым ребрам может не примыкать ни одной двумерной клетки (рис. 140).

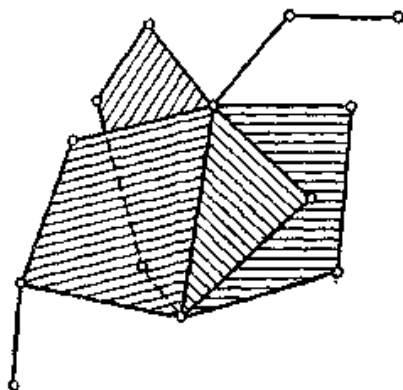


Рис. 140

Если клеточное разбиение состоит только из нульмерных и одномерных клеток, то оно представляет собой граф. В топологии рассматривают клеточные разбиения любого числа измерений. Например, трехмерное клеточное разбиение состоит из клеток размерностей 0,1,2,3 (причем если удалить все клетки размерностей 0,1,2, то оно распадается на трехмерные клетки, каждая из которых гомеоморфна открытому шару).

Фигура, которую можно представить в виде клеточного разбиения, называется *полиэдром*. Фигуры, рассмотренные в примерах 16, 18, 31, полиэдрами не являются.

Пример 43. Сферу P_0 можно представить в виде клеточного разбиения, состоящего из одной нульмерной и одной двумерной клетки. Действительно, если в сфере «выколоть» точку, то оставшаяся часть τ будет гомеоморфна открытому кругу. Одномерных клеток это клеточное разбиение не содержит.

Пример 44. На рис. 141 изображено клеточное разбиение, состоящее из круга и его границы, разбитой на две полуокружности r_1, r_2 .

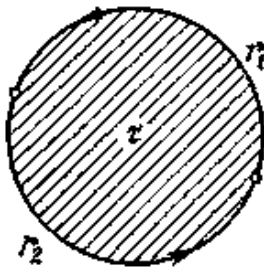


Рис. 141

Склеивание диаметрально противоположных точек окружности превращает круг в проективную плоскость, причем оба ребра, r_1, r_2 , склеиваются в одно ребро r . Мы получаем клеточное разбиение проективной плоскости, содержащее одну вершину, одно ребро r и одну двумерную клетку τ .

Пример 45. Начертим на торе параллель a и меридиан b , пересекающиеся в точке o (см. рис. 57, z). Мы получаем клеточное разбиение тора, состоящее из одной вершины o , двух ребер a, b и одной двумерной клетки τ .

Действительно, разрез по меридиану и параллели превращает тор в квадрат (см. задачу 65), т. е. в кусок, гомеоморфный кругу.

Задачи

152. Докажите, что ручку можно представить в виде клеточного разбиения, содержащего одну вершину, три ребра и одну двумерную клетку (см. рис. 58).

153. Докажите (см. рис. 59), что сферу с k ручками можно представить в виде клеточного разбиения, содержащего одну вершину, $2k$ ребер и одну двумерную клетку.

Нам понадобится говорить о направлении обхода на контуре грани (двумерной клетки). Смысл слов «направление обхода» в случае, когда грань гомеоморфна кругу, очевиден. В более сложных случаях направление обхода определяется следующим образом. При разрезании по всем ребрам грань (рис. 142, а) превращается в кусок, гомеоморфный кругу (рис. 142, б).

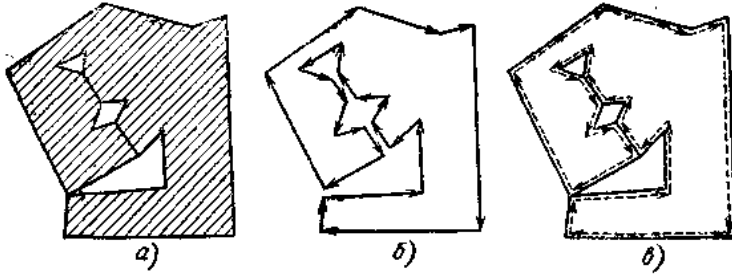


Рис. 142.

Обойдем один раз границу этого куска в некотором направлении. При обратном «склеивании» рассмотренного куска в грань этот обход и даст *обход по контуру грани*. Можно поступить и иначе: совершить обход «очень близко» к границе клетки, нигде ее не пересекая (рис. 142, в). Рассмотрим некоторую грань клеточного разбиения; все ребра, к которым она примыкает, как-либо ориентируем (т. е. выберем на них направления) и обозначим буквами a, b, c, \dots . Совершим теперь обход по контуру грани и одновременно с этим будем выписывать некоторый одночлен. Если мы, начиная обход, движемся сначала по ребру a , то мы напишем a или a^{-1} , смотря по тому, проходим мы (совершая обход) ребро a по направлению имеющейся на этом ребре стрелки или против нее. Если следующее ребро, которое мы проходим, обозначено, скажем, буквой d , то мы справа припишем d или d^{-1} , смотря по тому, в каком направлении мы пробегаем ребро d . Если вслед за тем мы проходим ребро m , то припишем справа m или m^{-1} и т. д. Совершив весь обход, мы выпишем некоторый одночлен, который называется *гомотопической границей* рассматриваемой грани.

Гомотопическую границу можно записать по-разному, в зависимости от того, в каком направлении обходить контур грани и с какого ребра начинать обход. Мы будем для каждой грани брать одну запись гомотопической границы (безразлично, какую именно). Например, обходя против часовой стрелки контуры клеток $\tau_1, \tau_2, \tau_3, \tau_4$ на рис. 143, мы получим их гомотопические границы:

$$adb\bar{c}; kh^{-1}g^{-1}fd^{-1}g; hl; l^{-1}k^{-1}.$$

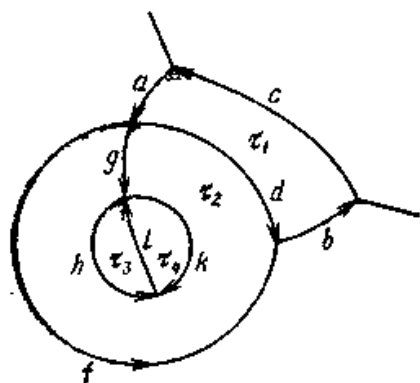


Рис. 143.

Заметим, что ребро g встречается в гомотопической границе грани τ_2 дважды.

Мы дадим теперь (без доказательства) способ вычисления фундаментальной группы связного полиэдра X . Возьмем какое-нибудь его клеточное разбиение и обозначим через G граф, образованный вершинами и ребрами. В графе G выберем максимальное дерево и все ребра, входящие в это дерево, пометим цифрой 1. Остальные ребра графа G (перемычки) как-либо ориентируем и пометим различными буквами a, b, c, \dots . Далее, для каждой грани рассматриваемого клеточного разбиения выпишем ее гомотопическую границу, не обращая внимания на ребра, помеченные цифрой 1. Наконец, построим группу, приняв буквы a, b, c, \dots , надписанные на перемычках, за ее образующие элементы, а за определяющие соотношения между этими образующими — равенства, получающиеся, если все выписанные гомотопические границы приравнять единице. *Эта группа изоморфна фундаментальной группе полиэдра X .*

Пример 46. Возьмем клеточное разбиение проективной плоскости, рассмотренное в примере 44. Максимальное дерево состоит из одной вершины. Поэтому ребро r является единственным образующим элементом фундаментальной группы. Далее, гомотопическая граница двумерной клетки τ (см. рис. 141) равна $r \cdot r$. Итак, фундаментальная группа проективной плоскости определяется одной образующей r с соотношением $r^2 = 1$, т. е. является группой второго порядка.

Пример 47. Рассмотрим клеточное разбиение тора, описанное в примере 45. Гомотопическая граница двумерной клетки τ равна

$aba^{-1}b^{-1}$ (надо обойти контур на рис. 57, a против часовой стрелки). Таким образом, фундаментальная группа тора имеет две образующие a, b , связанные единственным соотношением $aba^{-1}b^{-1} = 1$, т. е. $ab = ba$. Иначе говоря, она является свободной абелевой группой с двумя образующими.

Задачи

154. Используя клеточное разбиение, рассмотренное в задаче 152, докажите, что фундаментальная группа ручки представляет собой группу с тремя образующими a, b, c и единственным соотношением $ba = cab$. Эта группа некоммутативна: например, $ba \neq ab$.

155. Используя результат задачи 153, докажите, что группа $\pi(P_k)$ имеет $2k$ образующих $a_1, b_1, a_2, b_2, \dots, a_k, b_k$ с единственным соотношением $a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \dots a_k b_k a_k^{-1} b_k^{-1} = 1$. При $k \geq 2$ эта группа неабелева (например, $a_1 b_1 \neq b_1 a_1$).

156. Докажите, что группа $\pi(N_q)$ имеет q образующих c_1, c_2, \dots, c_q , связанных единственным соотношением $c_1^2 c_2^2 \dots c_q^2 = 1$.

157. Докажите, что две замкнутые поверхности (без края) в том и только в том случае гомеоморфны, если их фундаментальные группы изоморфны.

158. Букетом окружностей B^1_k называется объединение k простых замкнутых линий, которые все имеют общую точку o и больше общих точек попарно не имеют (рис. 144).

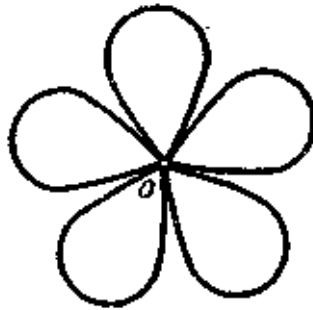


Рис. 144.

Докажите, что $\pi(B^1_k)$ есть свободная группа с k образующими.

159. Область X на плоскости ограничена одним внешним контуром и k внутренними (см. рис. 132). Докажите, что $\pi(X)$ есть свободная группа с k образующими.

2.3.4. Накрытия

Пример 48. На окружности B с центром o фиксируем начальную точку x_0 и для любой точки $x \in B$ обозначим через $\varphi(x)$ величину центрального угла $x_0 o x$. Функция $\varphi(x)$ определена с точностью до числа, кратного 2π . График E этой многозначной функции может быть построен на боковой поверхности бесконечного цилиндра; он имеет вид *винтовой линии* с шагом 2π (рис. 145).



Рис. 145.

Обозначим через p проектирование линии E на окружность B вдоль образующих цилиндра. Для произвольной точки $x \in B$ возьмем небольшую ее окрестность U . Часть линии E , проектирующаяся на U , состоит из отдельных кусков $\dots, V_{-1}, V_0, V_1, \dots$ (рис. 146).



Рис. 146,

Каждый из них с помощью проекции p гомеоморфно отображается на всю окрестность U . Указанное свойство подводит нас к понятию *накрытия*. Пусть p — непрерывное отображение фигуры E на B . Допустим, что p обладает таким же свойством, как и в примере 48: для каждой точки $x \in B$ можно подобрать такую ее окрестность U , что полный прообраз $p^{-1}(U)$ (т. е. множество всех точек фигуры E , переходящих при отображении p в точки окрестности U) распадается на части, каждая из которых при помощи p гомеоморфно отображается на U . При этих условиях E называется *накрытием* (или *накрывающей фигурой*) для B . Части полного прообраза $p^{-1}(U)$, гомеоморфно отображающиеся на U , называются *листами* накрытия. По числу листов различают двулистные, трехлистные и т. д. накрытия. Накрытие окружности винтовой линией бесконечнолистно.

Пример 49. Всякая односторонняя поверхность N имеет в качестве двулистной накрывающей некоторую двустороннюю поверхность P . Расположим поверхность N в пространстве (с самопересечениями) без точек излома и отложим на каждой нормали (в ту и другую сторону от точки $x \in N$) отрезки xx и xx' постоянной длины ε , где ε — малое положительное число. Если бы поверхность N была двусторонней, то точки x и x' описали бы две различные поверхности, «параллельные» N . Но так как поверхность N односторонняя, то мы получим одну поверхность P : ведь при обходе по некоторому замкнутому пути на односторонней поверхности нормаль xx меняет направление, т. е. переходит в xx' , так что точки \bar{x} и x' принадлежат одному куску поверхности P . Наглядно это построение можно описать так:

вообразим поверхность N изготовленной из «толстого» материала и окрасим ее всю краской. Если теперь «сжечь» поверхность N , считая, что краска негорючая, то оставшийся тонкий слой краски и образует поверхность P , двусторонне накрывающую N . При этом поверхность P двусторонняя: одна ее сторона обращена к сожженной поверхности N , а другая — наружу.

Например, если ленту Мёбиуса, изготовленную из «толстого» горючего материала, окрасить негорючей краской, а затем ленту Мёбиуса сжечь, то мы получим ленту, гомеоморфную боковой поверхности цилиндра (четырежды перекрученную, как легко убедиться на модели), которая двусторонне накрывает ленту Мёбиуса.

Задачи

160. Докажите, что если E является k -листным покрытием полиэдра B , то $\chi(E) = k\chi(B)$.

161. Докажите, что двусторонне накрывающая поверхности N_q является сферой с $q - 1$ ручками.

Пусть фигура E является накрывающей для B и $p: E \rightarrow B$ — соответствующая проекция. Пусть, далее, h — путь в фигуре B , исходящий из точки x_0 , а $\tilde{x}_0 \in E$ — некоторая точка, расположенная «над» x_0 , т. е. удовлетворяющая условию $p(\tilde{x}_0) = x_0$. Тогда в E существует (причем единственный) путь h , начинающийся в точке x_0 и переходящий в путь k при отображении p ; он называется *накрывающим путем*. В самом деле, пусть U — маленькая окрестность точки x_0 и \tilde{U} — тот лист накрытия, который содержит точку \tilde{x}_0 . Тогда, поскольку $p: \tilde{U} \rightarrow U$ — гомеоморфизм, мы однозначно сможем «поднять» кусочек пути h , находящийся в окрестности U , на лист \tilde{U} (рис. 147).

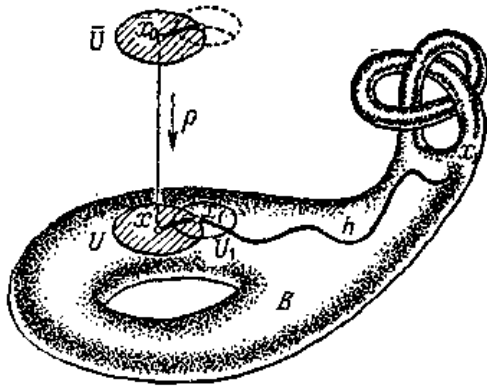


Рис. 147.

Если x_1 — концевая точка того участка пути, который мы уже «подняли», то можно рассмотреть окрестность U_1 точки x_1 и соответствующий лист накрытия, что позволит продолжить накрывающий путь \tilde{h} еще на один кусочек, и т. д.

Рассмотрение накрывающих путей позволяет установить теорему о связи между накрытиями и фундаментальной группой. Мы ее приведем (без доказательства) в следующей упрощенной формулировке. *Если связный полиэдр E является k -листным накрывающим над B и порядок группы $\pi(E)$ (т. е. число ее элементов) равен n , то порядок группы $\pi(B)$ равен kn .*

Накрытие E над B называется *универсальным*, если оно односвязно. В силу сказанного выше число листов универсального накрытия над B равно порядку группы $\pi(B)$; любое другое накрытие имеет меньшее число листов.

Накрытие проективной плоскости сферой (см. задачу 161) универсально в силу односвязности сферы. Сфера является также универсальной накрывающей для самой себя. Оказывается, что *для всех замкнутых поверхностей, кроме сферы и проективной плоскости, универсальной накрывающей является плоскость*. Доказательством этого факта мы и закончим этот пункт. Прежде всего, так как односторонняя поверхность N имеет своей двусторонней накрывающей некоторую двустороннюю поверхность P , то универсальное накрытие над P будет универсальным накрытием и над N . Поэтому достаточно рассмотреть двусторонние поверхности, отличные от сферы.

Разделим плоскость двумя системами параллельных прямых на конгруэнтные квадраты; склеивая каждый квадрат в тор, мы получим

отображение всей плоскости на тор, причем точкам, одинаково расположенным в различных квадратах (рис. 148), соответствует одна и та же точка тора (рис. 149).

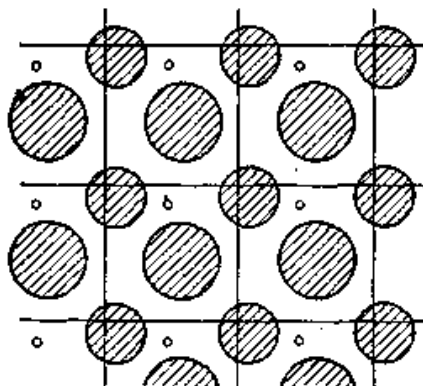


Рис.148.

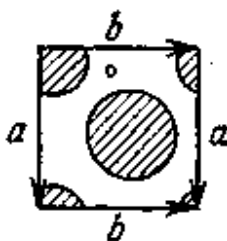


Рис 149.

Получающееся покрытие универсально, так как плоскость односвязна. Каждый из квадратов является так называемой *фундаментальной областью*, т. е. связным куском накрывающей (плоскости), который взаимно однозначно отображается на тор. Рис. 150 показывает, что фундаментальная область определена не однозначно.

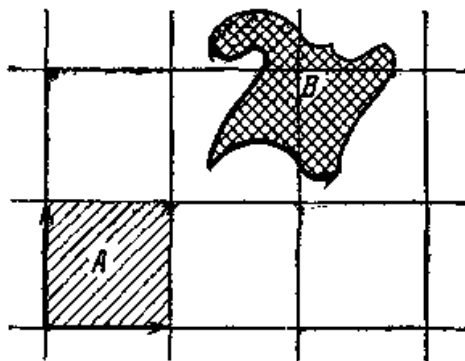


Рис. 150.

Опишем теперь разбиение на фундаментальные области, из которых склеиваются другие двусторонние поверхности, например, P_2 . Такое разбиение удобно произвести с помощью геометрии Лобачевского. В этой геометрии сумма углов многоугольника меньше, чем в евклидовой геометрии, причем сумма углов уменьшается при увеличении размеров многоугольника. Например, существует правильный восьмиугольник с углами $\frac{\pi}{4}$. Если такие восьмиугольники прикладывать друг к другу целыми сторонами, то ими можно заполнить всю плоскость Лобачевского, причем в вершинах будут сходиться по восемь многоугольников. На рис. 151 изображено такое разбиение для модели плоскости Лобачевского в «круге Пуанкаре».

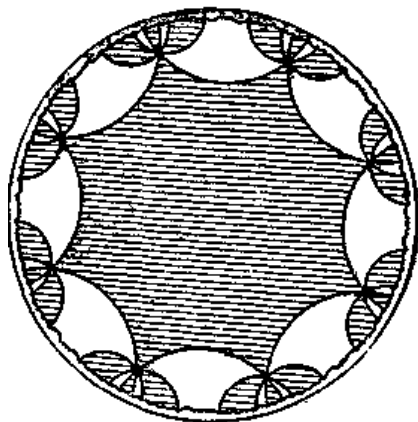


Рис. 151.

Это и есть разбиение плоскости Лобачевского (она гомеоморфна открытому кругу, а потому и плоскости Евклида) на фундаментальные области: оклеивание сторон каждого восьмиугольника дает P_2 (см. рис. 59) и получается накрывающее отображение плоскости Лобачевского на P_2 . Аналогичное разбиение плоскости Лобачевского можно построить и для любой поверхности P_k ($k \geq 2$).

Задачи

162. На рис. 152 изображена «плоскость с бесконечным числом ручек». Покажите, что при $k \geq 2$ она может служить накрытием над поверхностью P_k .

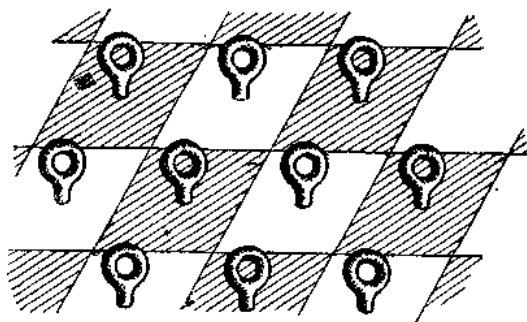


Рис. 152.

163. Покажите, что поверхность, изображенная на рис. 153, может служить накрывающей для любой поверхности P_k ($k \geq 2$).

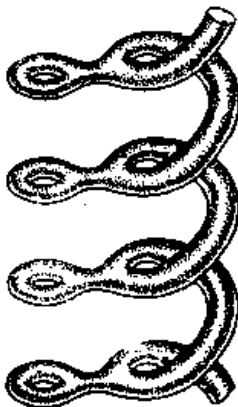


Рис. 153.

164. Постройте универсальную накрывающую для фигуры, состоящей из сферы и касающейся ее окружности.

2.3.5. Степень отображения и основная теорема алгебры

На рис. 154 изображено непрерывное отображение f окружности P на окружность Q .

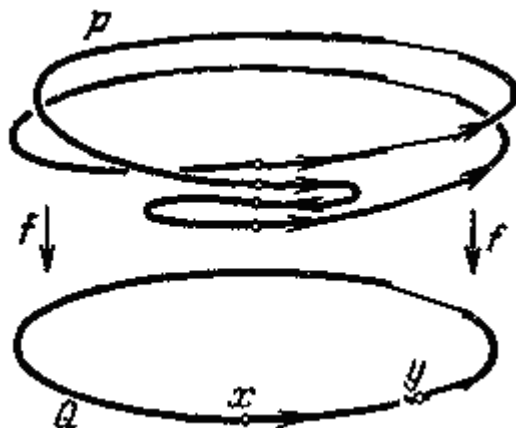


Рис. 154

На окрестность точки y отображаются два куска окружности P , причем отображаются положительно (т.е. с сохранением направления обхода). Говорят, что в точке y это отображение имеет степень 2. В точке x отображение также имеет степень 2: хотя на окрестность точки x отображаются четыре куска окружности P , но три из них отображаются положительно, а один — отрицательно. Если мы обозначим через p число листов, положительно отображающихся на окрестность некоторой точки $z \in Q$, а через n — число листов, отображающихся отрицательно, то степень отображения f в точке z будет число $p - n$. Во всех точках окружности Q степень отображения f одинакова (и равна двум); например, в точке x имеем $p - n = 3 - 1 = 2$. О степени отображения можно говорить и в случае отображения поверхностей. Пусть P и Q — две замкнутые ориентируемые поверхности, на каждой из которых задана ориентация. Пусть, далее, $f: P \rightarrow Q$ — некоторое непрерывное отображение; будем представлять себе, что поверхность P «наложена» на поверхность Q , располагая на ней несколькими «слоями» и образуя складки. Если на окрестность точки $z \in Q$ отображается несколько «листов» поверхности P , то некоторые из этих листов могут отображаться положительно (с сохранением ориентации, рис. 155, а), а некоторые — отрицательно (рис. 155, б).

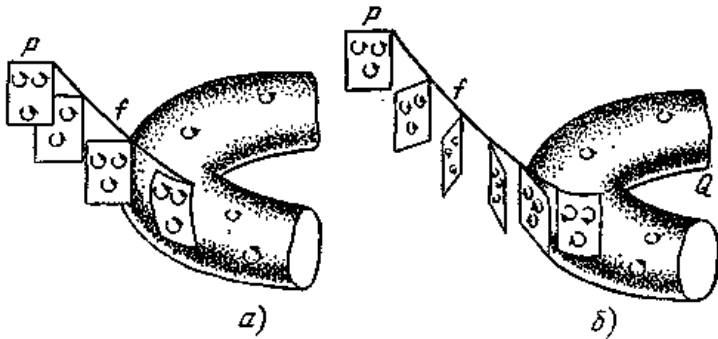


Рис. 155.

Если все «листы» поверхности P отображаются на окрестность точки z гомеоморфно, причем число листов, на которых отображение f положительно, равно p , а число листов, на которых оно отрицательно, равно n , то число $p - n$ называется *степенью* отображения f в точке z .

Нетрудно понять, что степень отображения f одинакова вблизи любой точки поверхности Q . Действительно, при перемещении точки z числа

p и n меняются плишь при прохождении через край складки, но разность $p - n$ остается неизменной (рис. 156).



Рис. 156.

Заметим еще, что когда отображение f непрерывно деформируется, степень его остается неизменной; это можно пояснить, заметив, что образование (или расправление) складок не меняет степени отображения.

С помощью понятия степени отображения можно дать изящное доказательство **основной теоремы алгебры**: *любой многочлен*

$$f(z) = z^m + a_1 z^{m-1} + \dots + a_{m-1} z + a_m,$$

степени $m \geq 1$ с комплексными (в частности, действительными) коэффициентами a_1, \dots, a_m имеет хотя бы один корень.

Возьмем сферу S , касающуюся плоскости в начале координат, и будем называть точку касания *южным полюсом*, а противоположную точку n — *северным полюсом* (рис. 157).

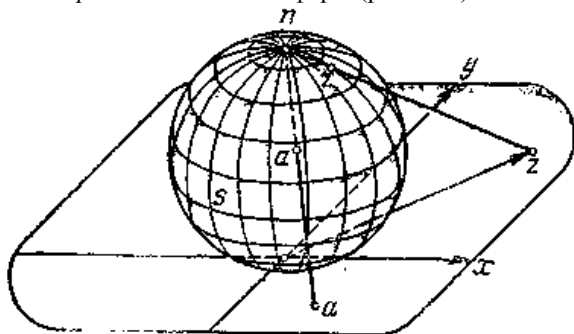


Рис. 157.

Будем изображать комплексное число $z = x + iy$ точкой в плоскости, считая x и y его координатами (рис. 158).

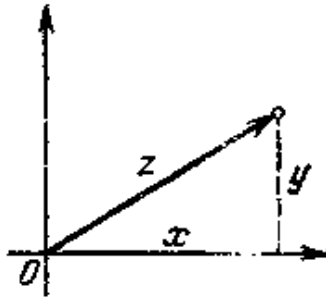


Рис. 158.

Отрезок n_z пересекает сферу S в некоторой точке, которую мы будем считать *изображением* комплексного числа z на сфере S . Обратно, имея на сфере точку a , легко узнать, какое комплексное число она изображает: прямая na при пересечении с плоскостью и даст искомое комплексное число. Однако северный полюс n не изображает никакого комплексного числа. Мы условимся считать, что точка n изображает «бесконечное» комплексное число, обозначаемое символом ∞ . Поводом для такого соглашения служит то, что при неограниченном удалении точки z на плоскости (в любую сторону) от начала координат изображающая ее точка на сфере приближается к n . Сфера S называется *комплексной сферой*, или *сферой Римана*. Отметим, что (в отличие от проективной плоскости; см. рис. 83) сфера S получилась из плоскости добавлением одной бесконечно удаленной точки ∞ .

Мы будем изображать значения z на одной комплексной сфере S_1 , а значения многочлена $f(z)$ — на другой такой же сфере S_2 . Каждой «конечной» точке $z = x + iy$ сферы S_1 соответствует «конечная» точка $f(z)$ сферы S_2 . При этом если z будет приближаться к ∞ , то $f(z)$ также будет приближаться к точке ∞ сферы S_2 . Действительно, мы имеем

$$f(z) = z^m \left(1 + \frac{a_1}{z} + \frac{a_2}{z^2} + \dots + \frac{a_{m-1}}{z^{m-1}} + \frac{a_m}{z^m} \right);$$

при $z \rightarrow \infty$ (т. е. при неограниченном увеличении числа $|z|$) выражение в скобках приближается к единице, а множитель z^m неограниченно увеличивается. Таким образом, дополнив определение многочлена условием $f(\infty) = \infty$, мы получаем непрерывное отображение f всей сферы S_1 на сферу S_2 .

Для доказательства основной теоремы алгебры нужно установить, что найдется точка $z \in S_1$, для которой $f(z) = 0$, т. е. что точка 0

сферы S_2 является образом хотя бы одной точки $z \in S_1$. Если бы это было не так, т. е. точка 0 сферы S_2 не покрывалась образом $f(S_1)$ сферы S_1 , то степень отображения $f: S_1 \rightarrow S_2$ вблизи точки $0 \in S_2$ была бы равна нулю, а так как степень одинакова вблизи любой точки, то просто степень отображения f была бы равна нулю. Поэтому для доказательства основной теоремы алгебры достаточно установить, что степень отображения f отлична от нуля. Мы покажем, что она равна m , т. е. совпадает со степенью многочлена $f(z)$ (это и послужило причиной введения термина «степень отображения»).

Будем изменять значения коэффициентов a_1, a_2, \dots, a_m , приближая их к нулю; многочлен $f(z)$ будет меняться, отображение $f: S_1 \rightarrow S_2$ будет непрерывно деформироваться. В результате мы получим многочлен $f_1(z) = z^m$. Но так как при деформации отображения его степень не меняется, то отображения f и f_1 имеют одинаковую степень. Степень же отображения f_1 легко подсчитать. Разобьем плоскость лучами, исходящими из точки 0, на m конгруэнтных углов (рис. 159).

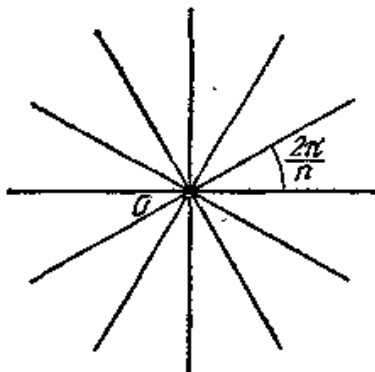


Рис. 159.

Так как при возведении комплексного числа z в степень m его аргумент увеличивается в m раз, то каждый из этих углов с помощью f_1 отображается («растягивается») на всю сферу S_2 . Таким образом, при отображении f_1 образ сферы S_1 покрывает m раз (причем положительно) сферу S_2 . Отсюда и вытекает, что степень отображения f_1 (а значит, и f) равна m . Теорема доказана. В настоящее время известно много различных доказательств основной теоремы алгебры, но все они являются топологическими, т. е. в той или иной форме используют идею непрерывности. Без привлечения идей топологии доказывать основную теорему алгебры невозможно; **можно сказать (хотя это**

звучит несколько странно), что основная теорема алгебры является *неалгебраической теоремой*.

Задачи

165. Докажите, что если $q \geq mk$, то существует отображение $f: P_q \rightarrow P_k$, имеющее степень m .

166. Докажите, что если P и Q — ориентируемые поверхности и отображение $f: P \rightarrow Q$ является n -листным накрытием, то степень отображения f равна $\pm k$.

167. Докажите, что если $f(z)$ — многочлен степени $m > 1$, то при некотором c (комплексном или действительном) уравнение $f(z) = c$ имеет не более $m - 1$ различных корней.

Указание. Если число корней равно m для любого c , то $f: S_1 \rightarrow S_1$ является накрытием и, следовательно, гомеоморфизмом.

168. Докажите, что при $k \geq 1$ всякое отображение $f: P_0 \rightarrow P_k$ стягиваемо (и потому имеет степень нуль).

Указание. Докажите, что для f существует покрывающее отображение $\tilde{f}: P_0 \rightarrow \tilde{E}$, где \tilde{E} — универсальное накрытие над P_k .

2.3.6. Группа узла

Пусть L_1 и L_2 — два узла в трехмерном пространстве. Обозначим через D_1 дополнительное пространство узла L_1 (т. е. множество всех точек пространства, не лежащих на линии L_1), а через D_2 — дополнительное пространство узла L_2 . Если узлы L_1, L_2 «одинаковы» (изотопны), т. е. существует гомеоморфное отображение f пространства на себя, при котором L_1 переходит в L_2 , то $f(D_1) = D_2$, т. е. дополнительные пространства гомеоморфны. Следовательно, группы $\pi(D_1)$ и $\pi(D_2)$ изоморфны, т. е. *фундаментальная группа дополнительного пространства является инвариантом узла*. Этот инвариант называется группой узла. Мы будем группу узла обозначать буквой G , т. е. $G(L_1) = \pi(D_1)$. Из сказанного ясно, что если группы $G(L)$ и $G(L')$ не изоморфны, то узлы L_1 и L_2 не изотопны.

Укажем теперь (без доказательства) способ вычисления группы узла. Пусть нормальная проекция узла L разбита на n дуг a_1, a_2, \dots, a_n , отделенных перерывами. Кроме того, выберем на L направление обхода и отметим его стрелками на дугах a_1, a_2, \dots, a_n (рис. 160).

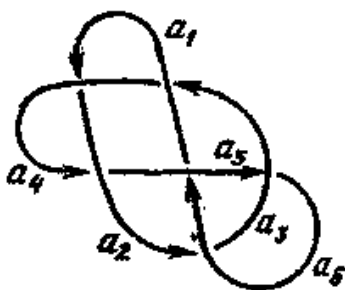


Рис. 160.

Теперь для описания группы $G(L)$ возьмем в пространстве точку o , расположенную выше линии L , и из нее проведем замкнутый путь x_k , охватывающий дугу a_k , расположенную над a_k , и обходящий ее в соответствии с правилом буравчика (рис. 161).

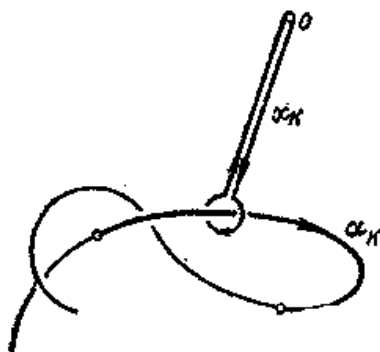


Рис. 161.

Гомотопические классы путей x_k ($k = 1, 2, \dots, n$) являются образующими группы узла.

Рассмотрим теперь какую-нибудь двойную точку проекции и обойдем вокруг нее по небольшой окружности l (по часовой стрелке), выписывая одновременно некоторый одночлен. Именно, если встретившаяся (при движении по l) дуга всходит внутрь окружности, то возьмем соответствующий символ x в степени $+1$, а если выходит из окружности, то в степени -1 . Обойдя вокруг двойной точки, мы выпишем, слева направо, произведение четырех множителей, которое приравняем единице; например, для двойной точки на рис. 162 получим соотношение

$$x_i x_k^{-1} x_j^{-1} x_k = 1.$$

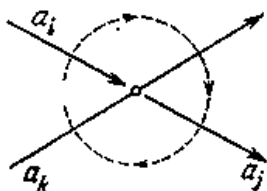


Рис. 162.

Нетрудно наглядно представить себе, что путь $x_i x_k^{-1} x_j^{-1} x_k$ действительно гомотопен нулю в дополнительном пространстве: на рис. 163 изображена пленка, гомеоморфная кругу, натянутая на этот путь.

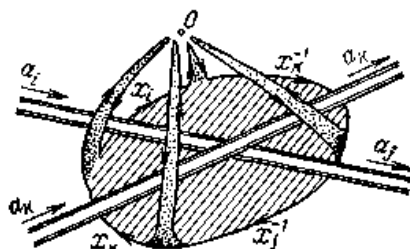


Рис. 163.

Оказывается, что, написав такие соотношения для всех двойных точек, мы и получаем полную систему соотношений между образующими x_1, x_2, \dots, x_n . Это описание группы узла применимо и к произвольным переплетениям.

Прежде чем переходить к рассмотрению конкретных узлов и переплетений, рассмотрим один алгебраический пример.

Пример 50. Докажем, что группа G , заданная тремя образующими x_1, x_2, x_3 и соотношениями

$$x_2 x_1 x_3^{-1} x_1^{-1} = 1, \quad x_3 x_2 x_1^{-1} x_2^{-1} = 1, \quad x_1 x_3 x_2^{-1} x_3^{-1} = 1,$$

неабелева. Для доказательства обозначим через G' группу самосовмещений равностороннего треугольника; она состоит из шести элементов: поворотов вокруг точки o на углы $0, \frac{2\pi}{3}, \frac{4\pi}{3}$ и трех осевых симметрии x'_1, x'_2, x'_3 , оси которых показаны на рис. 164.

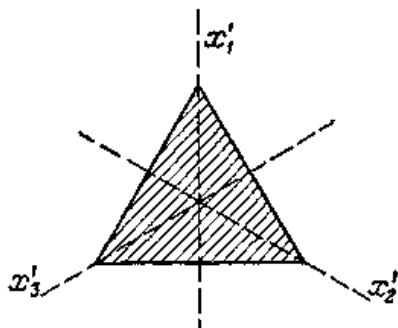


Рис. 164.

Без труда проверяется, что для элементов x'_1, x'_2, x'_3 указанные соотношения справедливы. При этом группа G' неабелева. Следовательно, группа G , заданная образующими x_1, x_2, x_3 и выписанными соотношениями, также неабелева (действительно, из этих соотношений не может вытекать, что группа G абелева, так как тогда и группа G' должна была бы быть абелевой, что неверно).

Пример 51. На рис. 165 изображена проекция простого узла L .

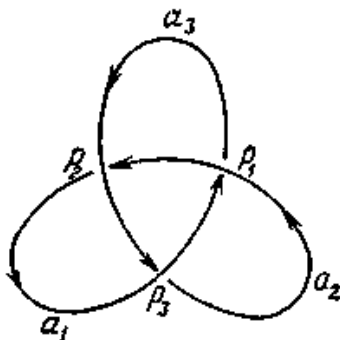


Рис. 165.

Соотношения между образующими x_1, x_2, x_3 (взяты в двойных точках p_1, p_2, p_3) совпадают с соотношениями, которые указаны в примере 50. Таким образом, группа $G(L)$ этого узла неабелева. Следовательно, узел L не изотопен окружности (у которой фундаментальная группа дополнительного пространства является свободной циклической и потому абелева). Таким образом, узел L не может быть развязан без разрезания нити.

Пример 52. На рис. 166 изображено переплетение L , образованное средними линиями торов, которые составляют множество A_1 на рис. 104, а.

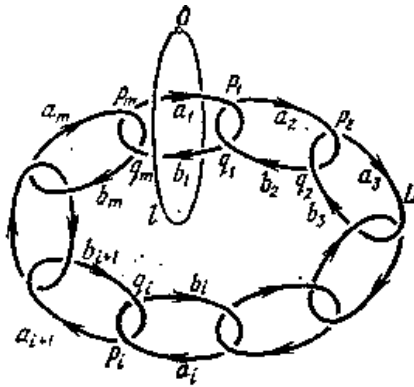


Рис. 166.

Группа $G(L)$ этого переплетения имеет $2m$ образующих $x_1, \dots, x_m, y_1, \dots, y_m$, которые получаются, если рассматривать пути, охватывающие дуги $a_1, \dots, a_m, b_1, \dots, b_m$, изображенные на рис. 166. Между этими образующими имеется $2m$ соотношений (выписанных для двойных точек p_i, q_i), которые имеют следующий вид:

$$x_i x_{i+1}^{-1} x_i^{-1} y_{i+1} = 1, \quad x_i y_{i+1}^{-1} y_i^{-1} y_{i+1} = 1; \quad i = 1, \dots, m$$

(где следует считать $x_{m+1} = x_1, y_{m+1} = y_1$). Окружность l , изображенная на рис. 166, представляет собой путь в дополнительном пространстве переплетения L , причем класс этого пути равен $x_1^{-1} y_1$ (рис. 167).

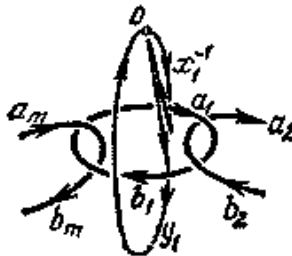


Рис. 167.

Докажем, что путь l негомотопен нулю в дополнительном пространстве, т. е. при стягивании окружности l в точку она непременно пересечет переплетение L .

Для доказательства обозначим через G' группу самосовмещений правильного m -угольника. Она состоит из поворотов вокруг точки o на углы $0, \frac{2\pi}{m}, \frac{4\pi}{m}, \dots, \frac{2(m-1)\pi}{m}$ и осевых симметрии x'_1, \dots

\dots, x'_m , оси которых показаны на рис. 168.

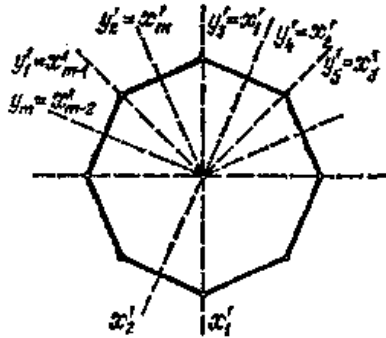


Рис. 168.

Положим, кроме того,

$$y'_1 = x'_{m-1}, y'_2 = x'_m, y'_3 = x'_1, \dots, y'_m = x'_{m-2}.$$

Без труда проверяется, что эти элементы группы G' удовлетворяют всем выписанным соотношениям (поскольку $(x'_i)^{-1} = x'_i$, а $x'_i x'_{i+1}$ есть поворот на угол $\frac{2\pi}{m}$). Кроме того, элемент $(x'_i)^{-1} y'_i$ (представляющий собой поворот на угол $\frac{4\pi}{m}$) отличен от единицы группы G' (т. е. от тождественного отображения). Следовательно, и в группе G элемент $(x'_i)^{-1} y'_i$ отличен от единицы. Иначе говоря, окружность l определяет в дополнительном пространстве путь, негомотопный нулю.

Аналогично можно доказать, что путь l негомотопен нулю и в дополнении переплетения, представляющего собой объединение средних линий торов, составляющих множество A_2 (рис. 104, б), и т. д. Это и дает обоснование свойств антуановского множества, рассмотренного ранее.

Задачи

169. Докажите, что переплетение, изображенное на рис. 117, невозможно «разнять», не разрывая ни одной из линий.

Указание. Докажите, что окружность l_1 определяет ненулевой элемент группы $G(L)$, где L — переплетение, образованное двумя другими окружностями. Для этого проверьте, что $G(L)$ есть свободная группа с двумя образующими.

170. Докажите, что окружность l , изображенную на рис. 169, невозможно «снять» с линии L и, следовательно, в дополнительном пространстве линии L не существует пленки, гомеоморфной кругу, которая «натянута» на l .

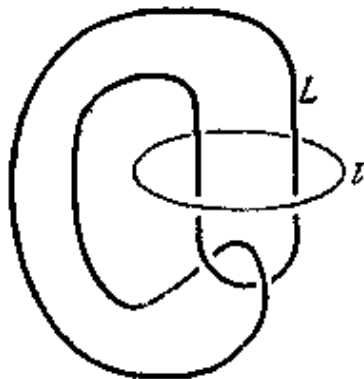


Рис. 169

Докажите также, что существует пленка, гомеоморфная ручке, которая «натянута» на l и расположена в дополнительном пространстве линии L .

2.3.7. Циклы и гомологии

На каждом из рисунков 170, 171 одномерный цикл z (он изображен в виде плавной линии, а не ломаной) ограничивает на поверхности некоторую область x .

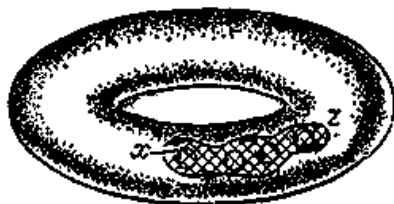


Рис. 170.

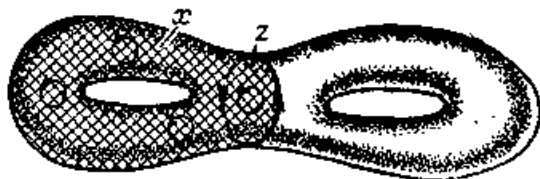


Рис. 171.

На рисунках эта область («пленка», натянутая на цикл z) ориентирована согласованно с этим циклом. Будем считать каждый *ограничивающий цикл* (т. е. цикл, на который можно натянуть пленку) несущественным, или, как говорят, *гомологичным нулю*.

На рис. 172, *а* изображены два цикла z_1 и z_2 ; объединение этих циклов обозначим через $z_1 + z_2$. На рис. 172, *б* показана разность $z_1 - z_2$ этих циклов (т. е. сумма цикла z_1 и цикла $-z_2$, получающегося из z_2 изменением ориентации).



Рис. 172. Задачи

На рисунке видно, что цикл $z_1 - z_2$ гомологичен нулю (он является границей пленки x). В этом случае говорят, что циклы z_1 и z_2 *гомологичны*.

В работах Пуанкаре были введены и изучены *группы гомологии*, являющиеся важными топологическими инвариантами. Идея их построения состоит в том, чтобы изучить, как велик в данной фигуре X запас циклов, которые попарно не гомологичны.

Задачи

171. Докажите, что на сфере каждый одномерный цикл гомологичен нулю.

172. Докажите, что в дополнительном пространстве множества A_1 , рассмотренного в примере 31 (а значит, и в дополнительном пространстве антуановского множества $A^* \subset A_1$) цикл l_1 (см. рис. 105, *а*) гомологичен нулю. Отсюда можно заключить, что

стягиваемость цикла является более тонким свойством, чем его гомологичность нулю.

173. Докажите, что каждый из циклов, изображенных замкнутыми контурами на рис. 173, 174, гомологичен нулю вне остальных контуров.

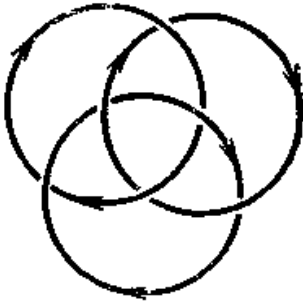


Рис. 173.

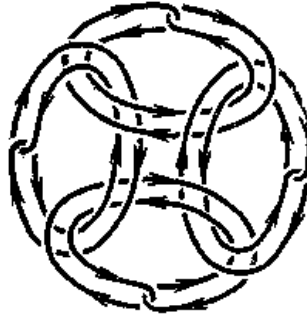


Рис. 174.

174. Докажите, что если коэффициент зацепления $\mathfrak{w}(z_1, z_2)$ отличен от нуля, то ни один из циклов z_1, z_2 не гомологичен нулю в дополнительном пространстве другого цикла.

Для построения групп гомологии следует обобщить понимание циклов и натягиваемых на них пленок. На рис. 175 каждый из циклов z_1, z_2 гомологичен нулю: цикл z_1 является границей круга $\tau_1 + \tau_2$, а цикл z_2 — границей круга $\tau_1 + \tau_3$.

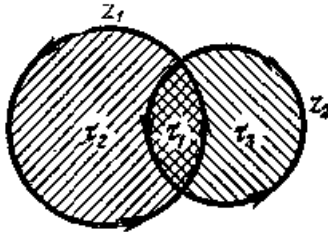


Рис. 175.

Сумма же $z_1 + z_2$ ограничивает «область» $(\tau_1 + \tau_2) + (\tau_1 + \tau_3) = 2\tau_1 + \tau_2 + \tau_3$, состоящую из «дважды взятой» клетки τ_1 и «один раз взятых» клеток τ_2 и τ_3 . Таким образом, чтобы убедиться, что цикл $z_1 + z_2$ гомологичен нулю, приходится брать клетки с определенными коэффициентами. Точно так же циклы могут состоять из клеток, взятых с некоторыми коэффициентами.

Например, на рис. 176 сумма $r_1 + r_2 + r_3 + 3r_4$ является циклом, поскольку (учитывая «трижды взятое» ребро r_4 , в каждой вершине число входящих и исходящих ребер одинаково.

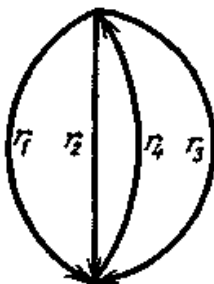


Рис. 176.

Отметим следующую теорему, вытекающую из закона двойственности Александра — Понтрягина (полную формулировку этого закона мы здесь не приводим). Пусть P — полиэдр, расположенный в трехмерном евклидовом пространстве, а Q — его дополнительное пространство; цикл z_1 , расположенный в одной из фигур P, Q , в том и только в том случае не гомологичен нулю в этой фигуре, если в другой из этих фигур найдется цикл z_2 , зацепленный с z_1 (т. е. $\mathbf{w}(z_1, z_2) \neq 0$).

Пример 53. На рис. 177 изображена линия P и цикл z' в дополнительном пространстве Q , причем цикл z' не зацеплен с одномерными циклами полиэдра P .

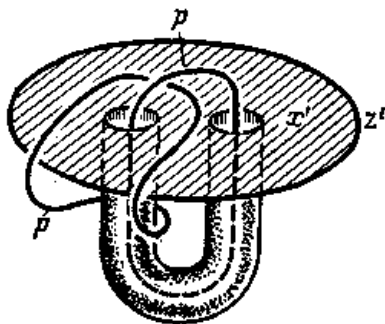


Рис. 177.

Следовательно, z гомологичен нулю в Q . На рисунке показана двумерная пленка $x' \subset Q$, границей которой является цикл z' .

Задачи

175. Для циклов m_1, m_2, m_3, m_4 , расположенных на поверхности кренделя (рис. 178), укажите в дополнительном пространстве такие циклы z_1, z_2, z_3, z_4 , что $w\{m_i, z_j\}$ равно единице при $i = j$ и нулю при $i \neq j$ ($i, j = 1, 2, 3, 4$).



Рис. 178.

176. Докажите, что для любого узла $l \subset R^3$ существует такой полиэдр $K \subset R^3$, гомеоморфный боковой поверхности цилиндра, что один край его совпадает с l , а другой край l' имеет с l нулевой коэффициент зацепления.

177. Постройте в фигуре P (см. рис. 132) циклы m_1, m_2, m_3 , а в ее дополнении (в пространстве) циклы z_1, z_2, z_3 так, что $w\{m_i, z_j\}$ равно единице при $i = j$ и нулю при $i \neq j$.

Будем рассматривать не сами одномерные циклы, а *классы гомологии*, объединяя в один класс все гомологичные между собой одномерные циклы рассматриваемой фигуры X . Операция сложения превращает множество всех классов в группу; это и есть одномерная *группа гомологии* $H_1(X)$.

Опишем способ вычисления одномерной группы гомологии. Прежде всего отметим, что *если два цикла z_1, z_2 гомотопны (т. е. могут быть получены один из другого с помощью деформации), то они гомологичны*. Наглядно это можно объяснить тем, что «след», который «замечает» цикл z_1 в процессе его деформации в z_2 , и есть та пленка, которая соединяет циклы z_1 и z_2 (рис. 179).

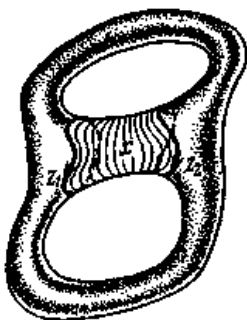


Рис. 179.

Обратное может не иметь места; гомологичные циклы на рис. 180 не гомотопны: перемещению цикла z_1 на поверхности Q в цикл z_2 мешают «дырки», имеющиеся на поверхности между этими циклами.

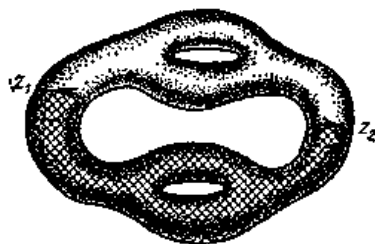


Рис. 180.

Таким образом, для гомологичности двух циклов достаточно (но не необходимо), чтобы они были гомотопными. Легко представить себе наглядно, что если задано клеточное разбиение полиэдра X , то любой одномерный цикл в этом полиэдре может быть при помощи деформации «сдвинут» в *одномерный остов*, т. е. в граф, состоящий из всех вершин и ребер (рис. 181).

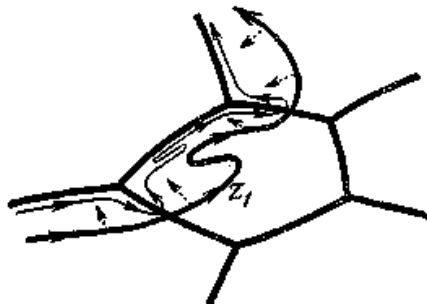


Рис. 181.

«Складки», которые могут возникнуть при деформации, можно распрямить. Следовательно, любой одномерный цикл гомотопен (а значит, и гомологичен) циклу, составленному из ребер, взятых с некоторыми коэффициентами. Таким образом, для вычисления группы гомологии $H_1(X)$ достаточно рассматривать одномерные циклы, составленные из ребер (с некоторыми целочисленными коэффициентами). Пленки же, натягиваемые на циклы, можно считать составленными из двумерных клеток, взятых с некоторыми коэффициентами.

Значит, нужно, во-первых, найти все одномерные циклы (составленные из ребер) и, во-вторых, научиться вычислять границы двумерных клеток, чтобы выяснить, какие одномерные циклы друг другу гомологичны. Первое не представляет труда: надо лишь проследить, чтобы в каждой вершине число входящих ребер было равно числу исходящих (учитывая коэффициенты). Второе фактически мы уже умеем делать; надо совершить обход по контуре клетки (в соответствии с ее ориентацией), но выписывать не произведение ребер (что мы делали при составлении *гомотопической границы* клетки), а их сумму с учетом знаков. Иначе говоря, в границу двумерной клетки τ (ее обозначают через $\partial\tau$) ребро r войдет с коэффициентом, равным сумме показателей степени, с которыми r входит в гомотопическую границу. Например, для клеток на рис. 143, ориентированных против часовой стрелки, мы имеем

$$\begin{aligned} \partial\tau_1 &= a + b + c + d; & \partial\tau_2 &= -d + f - h + k; \\ \partial\tau_3 &= h + l; & \partial\tau_4 &= -k - l. \end{aligned}$$

Пример 54. Клеточное разбиение двумерной сферы P_0 , рассмотренное в примере 43, содержит только две клетки: нульмерную и двумерную. Одномерных клеток это разбиение не содержит совсем, поэтому *группа $H_1(P_0)$ тривиальна* (ненулевых одномерных циклов в этом клеточном разбиении нет).

Пример 55. Рассмотренное в примере 44 клеточное разбиение проективной плоскости N_1 состоит из одной нульмерной клетки, одной одномерной клетки r и одной двумерной клетки τ . Любой одномерный цикл имеет вид kr (поскольку, кроме r , других ребер нет), причем цикл $2r$ гомологичен нулю (так как $2r = \partial\tau$; см. рис. 141).

Отсюда следует, что *одномерная группа гомологии $H_1(N_1)$ проективной плоскости является группой порядка 2.*

Заметим, что при вычислении групп гомологии в примерах 54, 55 мы использовали одно клеточное разбиение рассматриваемого полиэдра, но говорили не о «группе гомологии этого разбиения», а о группе гомологии самого полиэдра. В действительности, это оправдано,

поскольку группы гомологии полиэдров не зависят от выбора клеточных разбиений, а всецело определяются самими полиэдрами.

Задачи

178. На рис. 182 показано клеточное разбиение ленты Мёбиуса (обе полуокружности на внутреннем контуре склеиваются в одно ребро a).

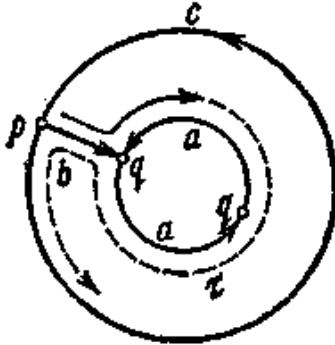


Рис. 182.

Проверьте, что $\partial z = c - 2a$, и выведите отсюда, что одномерная группа гомологии ленты Мёбиуса является свободной циклической.

179. Докажите, что для клеточного разбиения тора T , рассмотренного в примере 45, справедливо соотношение $\partial \tau = 0$. Выведите отсюда, что $H_1(T)$ есть свободная абелева группа с двумя образующими a, b .

180. Докажите, что цикл z , изображенный на рис. 183, гомогичен $\pm 3a \pm 2b$ (где знаки зависят от направлений, выбранных на параллели a и меридиане b тора T).

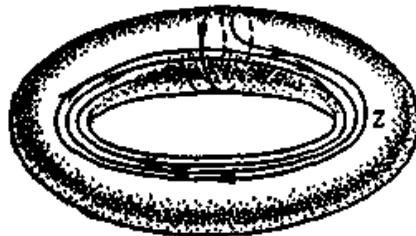


Рис. 183.

181. Докажите, что одномерная группа гомологии кренделя P_2 является свободной абелевой группой с четырьмя образующими m_1, m_2, m_3, m_4 (см. рис. 178).

182. Докажите, что одномерная группа гомологии поверхности P_k является свободной абелевой группой с $2k$ образующими.

183. Вычислите группу $H_1(N_q)$. (Ответ: абелева группа с q образующими c_1, c_2, \dots, c_q , связанными единственным соотношением $2c_1 + 2c_2 + \dots + 2c_q = 0$. Эту группу можно описать и иначе: прямая сумма группы второго порядка и свободной абелевой группы с $q - 1$ образующими.)

184. Докажите, что замкнутая поверхность Q в том и только в том случае неориентируема, если в группе $H_1(Q)$ имеется элемент порядка 2. Докажите также, что две замкнутые поверхности в том и только в том случае гомоморфны, если их одномерные группы гомологии изоморфны.

185. Докажите, что не существует клеточного разбиения тора, состоящего менее чем из четырех клеток.

Рассмотрим теперь нульмерные гомологии. Нульмерный цикл мы получаем, взяв вершины клеточного разбиения с некоторыми целочисленными коэффициентами. Далее, граница ребра равна разности его концов: на рис. 184 имеем $\partial r_1 = b - a$, $\partial r_2 = 0$.

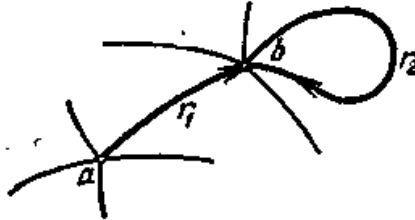


Рис 184.

Два нульмерных цикла гомологичны, если их разность является границей некоторой суммы одномерных клеток (с какими-то коэффициентами).

Наконец, будем рассматривать нульмерные классы гомологии, объединяя в один класс все гомологичные между собой нульмерные циклы рассматриваемого полиэдра X . Операция сложения превращает множество всех классов в группу; это и есть нульмерная группа гомологии $H_0(X)$.

Задачи

186. Докажите, что если r_1, r_2, \dots, r_k — простая цепочка направленных ребер, идущая от вершины a к вершине b , то

$$\partial(r_1 + r_2 + \dots + r_k) = b - a.$$

187. Докажите, что если X — связный полиэдр, то любой нульмерный цикл в X гомологичен одной точке, взятой с некоторым коэффицентом, т. е. группа $H_0(X)$ является свободной циклической.

188. Докажите, что если полиэдр X состоит из k компонент, то $H_0(X)$ есть свободная абелева группа с k образующими.

Аналогично определяется двумерная группа гомологии $H_2(X)$: нужно в X рассматривать двумерные циклы и натянутые на них трехмерные «пленки».

Пример 56. Пусть X состоит из всех точек трехмерного пространства, лежащих на торе и внутри него («полный тор»). Полиэдр X можно представить в виде клеточного разбиения, в котором, кроме клеток o , a , b, v , расположенных на торе, имеется еще одна двумерная клетка τ' («поперечное сечение» тора), границей которой служит меридиан b , и одна трехмерная клетка v — внутренности тора, рассеченная по клетке τ' . Границы клеток имеют следующие значения:

$$\partial a = 0, \quad \partial b = 0, \quad \partial \tau = 0, \quad \partial \tau' = b, \quad \partial v = \tau$$

(заметим, что трехмерная клетка v с двух сторон примыкает к двумерной клетке τ' , причем с одной стороны ориентация клетки τ' наблюдается, как ориентация по часовой стрелке, а с другой стороны — против втрелки; поэтому в ∂v клетка τ' не входит).

Одномерные циклы этого клеточного разбиения имеют вид $ka + lb$ (где k, l — целые), причем цикл b гомологичен нулю (на него натянута пленка τ'). Следовательно, любой одномерный цикл гомологичен ka , и потому группа $H_1(X)$ — свободная циклическая. Далее, так как

$\partial(m\tau + n\tau') = nb$, то $m\tau + n\tau'$ только в том случае является двумерным циклом (т. е. имеет границу, равную нулю), если $n=0$. Итак, двумерные циклы имеют вид $m\tau$. Но любой такой цикл гомологичен нулю (поскольку $\partial v = \tau'$, т. е. v есть «трехмерная пленка», натянутая на двумерный цикл τ). Следовательно, группа $H_2(X)$ тривиальна.

В ряде случаев удобно вместо вычисления всей группы гомологии $H_r(X)$ ограничиться лишь нахождением ранга этой группы; он называется r -мерным числом Бетти полиэдра X и обозначается через $p_r(X)$.

Определение чисел Бетти можно сформулировать и иначе. Говорят, что r -мерные циклы z_1, \dots, z_n в X гомологически независимы, если, каковы бы ни были целые числа k_1, \dots, k_n , хотя бы одно из которых отлично от нуля, цикл $k_1 z_1 + \dots + k_n z_n$ не гомологичен

нулю в X . Теперь r -мерное число Бетти $p_r(X)$ определяется как наибольшее число гомологически независимых r -мерных циклов в X . В качестве примера применения чисел Бетти приведем (без доказательства) следующую теорему о вычислении эйлеровой характеристики. Пусть X — некоторый полиэдр; рассмотрим какое-нибудь его клеточное разбиение и обозначим через a_r число r -мерных клеток этого разбиения ($r = 0, 1, 2, \dots$). Тогда эйлерова характеристика полиэдра X , т. е. число $\chi(X) = \sum (-1)^r a_r$, может быть вычислена по его числам Бетти: $\chi(X) = \sum (-1)^r p_r(X)$ (суммирование распространено на все значения $r = 0, 1, 2, \dots$ вплоть до наибольшей из размерностей клеток полиэдра X).

Пример 57. *Трехмерная сфера S^3 определяется как граница шара в четырехмерном пространстве R^4 ; в декартовых прямоугольных координатах x_1, x_2, x_3, x_4 она определяется уравнением*

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1.$$

Рассматривая «четырёхмерный аналог» рисунков 157 и 7, нетрудно доказать, что трехмерная сфера с выколотой точкой гомеоморфна трехмерному евклидову пространству, а потому — открытому трехмерному шару. Следовательно, трехмерную сферу можно представить в виде клеточного разбиения, состоящего только из двух клеток: 1 нульмерной клетки o и трехмерной клетки v . Поэтому (ср. пример 54) группы, гомологии $H_0(S^3)$ и $H_3(S^3)$ являются свободными циклическими, а в остальных размерностях группы гомологии трехмерной сферы тривиальны. Из этого следует, что

$$p_0(S^3) = p_3(S^3) = 1, \quad p_1(S^3) = p_2(S^3) = 0.$$

Пример 58. Подобно тому, как склеивание противоположных сторон квадрата дает тор, склеивание противоположных граней куба дает *трехмерный тор T^3* (его не следует путать с полным тором, см. пример 56). Например, на гранях $ABCD$ и $abcd$ (рис. 185) склеиваются между собой точки, которые являются концами отрезка, параллельного ребрам Aa, Bb, Cc, Dd .

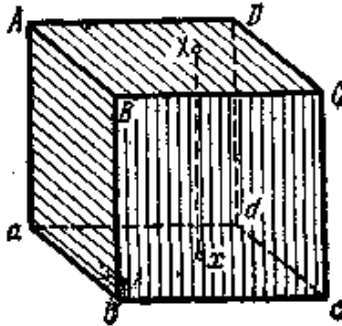


Рис. 185.

Все вершины куба склеиваются вместе, что дает одну нульмерную клетку. Далее, все параллельные ребра склеиваются вместе; после склеивания остаются три одномерные клетки. Парное склеивание противоположных граней дает три двумерные клетки. Наконец, мы имеем одну трехмерную клетку. Это и дает клеточное разбиение трехмерного тора T^3 . Каждая из клеток имеет границу, равную нулю, и потому группа гомологии $H_3(T^3)$ трехмерного тора T^3 является свободной циклической, а каждая из групп $H_1(T^3), H_2(T^3)$ является свободной абелевой группой с тремя образующими. Из этого следует, что

$$p_0(T^3) = p_2(T^3) = 1, p_1(T^3) = p_3(T^3) = 3.$$

Пример 59. Покажем, что из двух полных торов, склеивая их друг с другом границами, можно получить трехмерную сферу. Тор (см. рис. 5) разбивает пространство на две части: внутреннюю, представляющую собой полный тор, и внешнюю. При пополнении трехмерного пространства одной точкой (в результате чего получается трехмерная сфера) внешняя часть также превращается в полный тор (что и дает разбиение трехмерной сферы на два склеенных между собой полных тора). Это ясно из рассмотрения рис. 186.

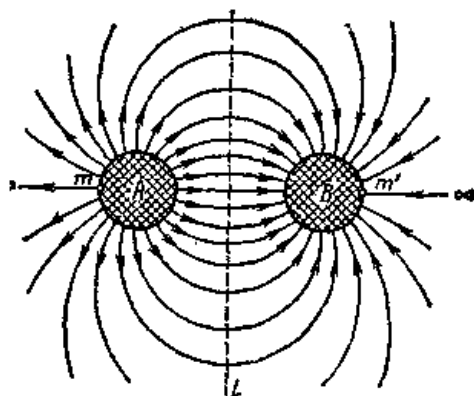


Рис. 186.

При вращении этого чертежа вокруг прямой l каждая «силовая линия», начинающаяся на «заряде» A и кончающаяся на B (включая и линию $t \infty t'$, которая представляет собой одну линию, ибо имеется только одна бесконечно удаленная точка), дает топологический круг, и все такие круги заполняют внешность тора, получающегося при вращении «зарядов» A и B . Этим кругам имеется столько, сколько точек на окружности «заряда» A . Таким образом, внешность тора (пополненная одной бесконечно удаленной точкой) гомеоморфна полному тору.

Задачи

189. Докажите, что для трехмерного шара X группы $H_1(X)$, $H_2(X)$, $H_3(X)$ тривиальны.

190. Рассмотрим часть трехмерного пространства, ограниченную двумя концентрическими сферами («трехмерное кольцо»), и отождествим (склеим) каждые две точки этих сфер, лежащие на одном радиусе. Докажите, что у получающегося полиэдра X все группы гомологии $H_0(X)$, $H_1(X)$, $H_2(X)$, $H_3(X)$ — свободные циклические.

191. Вычислите группы гомологии полиэдра, представляющего собой объединение поверхности P_k и ее внутренней области.

192. Вычислите гомологии *трехмерного проективного пространства*, которое получается из трехмерного шара, если на его границе склеить каждые две диаметрально противоположные точки.

193. Докажите, что любая замкнутая поверхность может быть без самопересечений расположена в трехмерном проективном пространстве.

194. Пусть X — полиэдр, рассматриваемый в некотором клеточном разбиении и α_r — число его r -мерных клеток ($r = 0, 1, \dots, n$, где n —

наибольшая из размерностей клеток). Докажите, что для любого $r = 0, 1, \dots, n - 1$ справедливо неравенство

$$\sum_{k=1}^r (-1)^{r-k} \alpha_k \geq \sum_{k=1}^r (-1)^{r-k} p_k(X).$$

Указание. Рассмотрите r -мерный остов X^r полиэдра X (состоящий из всех клеток рассматриваемого разбиения, имеющих размерность $\leq r$) и докажите соотношения

$$p_0(X^r) = p_0(X), \quad p_1(X^r) = p_1(X), \quad \dots, \\ p_{r-1}(X^r) = p_{r-1}(X), \quad p_r(X^r) \geq p_r(X).$$

Отметим в заключение, что при построении групп гомологии можно было бы в качестве коэффициентов брать не целые числа, а вычеты по модулю 2, вычеты по модулю m (или, вообще, элементы некоторой абелевой группы G). Получающиеся группы гомологии обозначаются через $H_r(X, Z_2), H_r(X, Z_m), H_r(X, G)$. Например, если коэффициентами являются вычеты по модулю 2, то все клетки можно рассматривать неориентированными. В случае, когда группой коэффициентов является циклическая группа порядка p , где p — простое число, группа гомологии $H_r(X, Z_p)$ является прямой суммой нескольких групп, изоморфных Z_p ; число слагаемых в этой прямой сумме называется r -мерным числом Бетти полиэдра X по модулю p .

Задачи

195. Докажите, что для проективной плоскости N_1 группы $H_0(N_1, Z_2), H_1(N_1, Z_2), H_2(N_1, Z_2)$ являются группами второго порядка.

196. Докажите, что поверхности P_k и N_{2k} имеют (во всех размерностях) одинаковые группы гомологии по модулю 2.

197. Вычислите для трехмерного проективного пространства (см. задачу 192) гомологии по модулю 2.

2.3.8. Топологическое произведение

Пример 60. Каждая точка цилиндра E (рис. 187) может быть задана парой точек (x, y) , где x лежит на нижнем основании B , а y — на образующей F_1 , проведя через x отрезок, параллельный F , а через y — круг, параллельный B , мы получим на их пересечении искомую точку цилиндра.

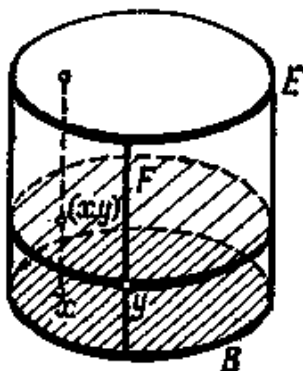


Рис. 187.

Таким образом, цилиндр E можно рассматривать как множество всех пар (x, y) , где x — точка одной фигуры B (круга), а y — точка другой фигуры F (отрезка).

Пример 61. На торе E (рис. 188) проведем меридиан B и параллель F .

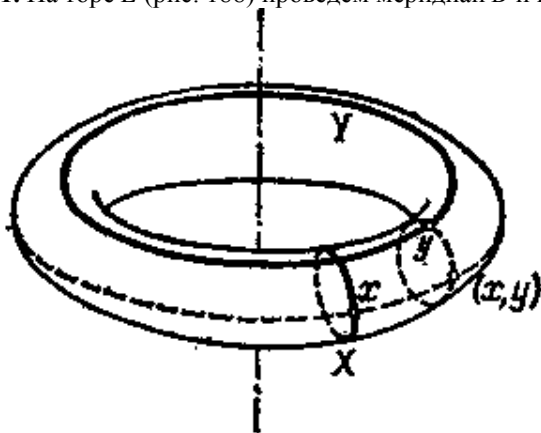


Рис. 188.

Для задания любой точки тора достаточно указать точку $x \in B$ и точку $y \in F$: проведя через x параллель, а через y меридиан, мы получим на их пересечении искомую точку тора. Таким образом, тор E можно рассматривать как множество всех пар (x, y) , где $x \in B$, $y \in F$.

В рассмотренных примерах мы имели *топологическое произведение* фигур B и F : цилиндр есть топологическое произведение круга и отрезка, тор — топологическое произведение двух окружностей. Вообще, *фигура E называется топологическим произведением фигур B и F , если E можно рассматривать как мнджество всевозможных пар (x, y) , где $x \in B, y \in F$* . Отметим, что здесь говорится лишь о том, из каких точек состоит фигура E , но следует еще указать топологию в фигуре E . Наглядно эту топологию можно описать, сказав, что точки (x_1, y_1) и (x_2, y_2) фигуры E будут «близкими», если x_1 и x_2 являются «близкими» в B , а y_1 и y_2 являются «близкими» в F .

Существенно, чтобы каждой точке фигуры E соответствовала некоторая пара (x, y) и чтобы различные пары соответствовали различным точкам фигуры E .

Пример 62. Рассмотрим на сфере *экватор B и нулевой меридиан F* . Для задания точки на сфере достаточно указать ее *географические координаты*, т. е. точки $x \in B, y \in F$: проведя через эти точки меридиан и параллель (рис. 189), мы получим на их пересечении искомую точку (x, y) сферы.

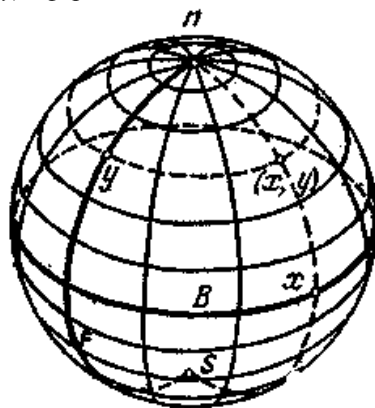


Рис. 189.

Но это не значит, что сфера — топологическое произведение экватора и меридиана; действительно, если x и x' — две различные точки экватора, а n — верхний конец меридиана (северный полюс), то различным парам (x, n) и (x', n) соответствует одна и та же точка n на сфере.

Задачи

198. Докажите, что кольцо, ограниченное двумя concentрическими окружностями, есть топологическое произведение окружности и отрезка.

199. Докажите, что полный тор (пример 56) является топологическим произведением круга и окружности.

200. Докажите, что полиэдр X , рассмотренный в задаче 190, является топологическим произведением сферы и окружности.

201. Докажите, что трехмерный тор T^3 является топологическим произведением обычного (двумерного) тора и окружности. Можно также сказать, что T^3 — топологическое произведение трех окружностей.

Рассмотрим теперь вопрос о гомологических свойствах топологического произведения, ограничиваясь для простоты рассмотрением не групп гомологии, а лишь чисел Бетти. Если в примере 61 считать меридиан и параллель одномерными циклами, то их топологическое произведение (тор) будет двумерным циклом. Вообще, если E — топологическое произведение полиэдров B и F , в которых соответственно взяты циклы z, z' размерностей r, r' , то можно рассмотреть произведение этих циклов, которое будет $(r + r')$ -мерным циклом в полиэдре E . Оказывается, что таким путем (перемножением циклов взятых в B и F) можно получить систему гомологически независимых циклов в полиэдре E . Для этого нужно сначала взять максимальное число гомологически независимых нульмерных циклов в B и r -мерных циклов в F ; перемножая эти циклы, мы получим $p_0(B) p_r(F)$ циклов в E , имеющих размерность r . Затем нужно взять максимальное число гомологически независимых одномерных циклов в B и $(r - 1)$ -мерных в F ; перемножая, мы получим еще $p_1(B) p_{r-1}(F)$ циклов в E , имеющих размерность r . Затем нужно взять двумерные циклы в B и $(r - 2)$ -мерные в F и т. д. Соединяя все полученные циклы вместе, мы и получим максимальное число r -мерных гомологически независимых циклов в полиэдре E . Таким образом, если E - топологическое произведение полиэдров B и F , то

$$p_r(E) = p_0(B) p_r(F) + p_1(B) p_{r-1}(F) + \\ + p_2(B) p_{r-2}(F) + \dots + p_r(B) p_0(F).$$

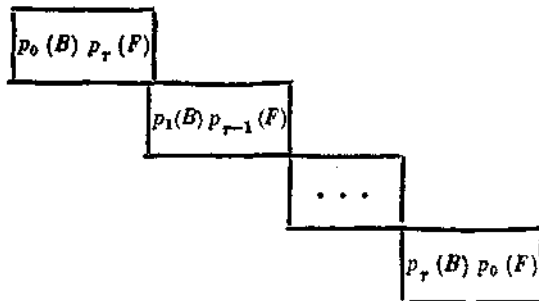
Этой формуле можно дать следующее «графическое» истолкование. Составим таблицу, у которой на пересечении i -го столбца и j -й строки стоит число $p_i(B) p_j(F)$:

...
$p_j(F)$	$p_0(B) p_j(F)$	$p_1(B) p_j(F)$...	$p_l(B) p_j(F)$...
...
$p_2(F)$	$p_0(B) p_2(F)$	$p_1(B) p_2(F)$...	$p_l(B) p_2(F)$...
$p_1(F)$	$p_0(B) p_1(F)$	$p_1(B) p_1(F)$...	$p_l(B) p_1(F)$...
$p_0(F)$	$p_0(B) p_0(F)$	$p_1(B) p_0(F)$...	$p_l(B) p_0(F)$...

$p_0(B)$	$p_1(B)$...	$p_l(B)$...
----------	----------	-----	----------	-----

То

гда суммирование чисел, стоящих на r -й диагонали этой таблицы, и дает r -мерное число Бетти полиэдра E :



Задачи

- 202. Составьте указанную таблицу для топологического произведения сферы и окружности; вычислите этим приемом числа Бетти полиэдра, рассмотренного в задаче 190 (см. также задачу 200).
- 203. Докажите, что если полиэдр E является топологическим произведением полиэдров B и F , то $\chi(E) = \chi(B) \cdot \chi(F)$.
- 204. Вычислите числа Бетти n -мерного тора (т. е. топологического произведения n окружностей).
- 205. Докажите, что трехмерная сфера не гомеоморфна топологическому произведению окружности и некоторой поверхности. То же докажите для трехмерного проективного пространства.

2.3.9. Расслоения

Вернемся к примеру 60 и обозначим через p проекцию цилиндра E на его основание B . Для каждой точки $x \in B$ прообраз представляет $p^{-1}(x)$ собой отрезок, параллельный F . Эти отрезки будем называть *слоями*. Над каждой точкой x базисной фигуры B расположен («растет») соответствующий слой, а весь цилиндр *расслаивается* и представляет собой объединение всех слоев (как бы связку стерженьков).

Вообще проекция p , ставящая в соответствие точке $(x, y) \in E$ точку $x \in B$, отображает топологическое произведение E фигур B и F на базу B , причем прообраз $p^{-1}(x)$ любой точки $x \in B$ (слой, «растущий над x »), гомеоморфен F . Это можно проследить в примере 61 и в задачах 198-201.

Рассмотрим теперь проекцию p винтовой линии E на окружность B (см. пример 48). Каждый прообраз $p^{-1}(x)$ («слой, растущий над точкой $x \in B$ ») гомеоморфен множеству F , состоящему из точек $\dots, -4\pi, -2\pi, 0, 2\pi, 4\pi, 6\pi, \dots$ числовой прямой. Этот пример отличается от топологического произведения окружности B и слоя F (которое состоит из бесконечного числа отдельных окружностей; рис. 190).

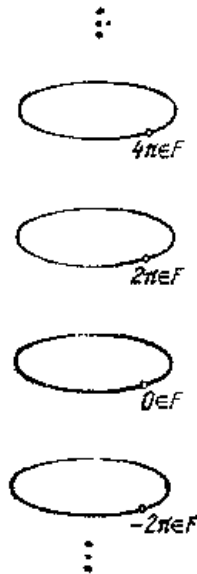


Рис. 190.

Однако прообраз $p^{-1}(U)$ окрестности U распадается на отдельные листы, т. е. $p^{-1}(U)$ есть топологическое произведение окрестности U и слоя F (см. рис. 146). Иначе говоря, локально (т. е. в окрестности каждой точки $x \in B$) E является топологическим произведением, но в целом — нет. В топологии в таких случаях используют термин *локально тривиальное расслоение* (т. е. топологическое произведение считается в топологии «тривиальным» расслоением). Вообще, *любое накрытие является локально тривиальным расслоением, причем слой F этого расслоения состоит из изолированных точек*. В примере 49 слой состоит из двух точек, накрытие E представляет собой ориентируемую поверхность, а база B — неориентируемую.

Пример 63. Обозначим через B среднюю линию ленты Мёбиуса E . Через каждую точку $x \in B$ проходит поперечный отрезок, который мы будем считать слоем, «растущим» над точкой x (эти поперечные отрезки получаютя из поперечных отрезков прямоугольной ленты при склеивании ее в ленту Мёбиуса). Отображая каждый поперечный отрезок в соответствующую точку x , мы получаем *проекцию* $p: E \rightarrow B$, причем $p^{-1}(x)$ есть слой над точкой $x \in B$. Это расслоение локально тривиально. В самом деле, если взять на окружности B небольшую дугу U , то ее прообраз $p^{-1}(U)$ представляет собой топологическое произведение дуги U и слоя F (рис. 191).



Рис. 191.

В целом же лента Мёбиуса E не является топологическим произведением окружности B и отрезка F (см. задачу 198).

Пример 64. Еще одним примером локально тривиального расслоения служит *нормированный касательный пучок* поверхности. Пусть B — ориентируемая поверхность. Обозначим через E множество всех векторов длины 1, касающихся поверхности B . Через $p: E \rightarrow B$ обозначим отображение, сопоставляющее каждому касательному вектору $x \in E$ ту точку $x \in B$, из которой этот вектор «растет». Слой $p^{-1}(x)$ над точкой $x \in B$ (состоящий из всех векторов длины 1, касающихся поверхности в точке x) гомеоморфен окружности. Отображение $p: E \rightarrow B$ является локально тривиальным расслоением. В самом деле, маленькую окрестность U точки x на поверхности B можно считать кусочком плоскости, и потому каждый вектор z ,

касающийся поверхности B в какой-либо точке $x \in U$, может быть задан как пара (x, y) , где y — точка единичной окружности F (рис. 192).



Рис. 192.

Иначе говоря, $p^{-1}(U)$ представляется в виде топологического произведения окрестности U и окружности F .

Пусть, в частности, E — нормированный касательный пучок сферы S^2 (т. е. множество всех единичных векторов, касающихся этой сферы). Пространство E можно представить в виде клеточного разбиения, состоящего из четырех клеток. В самом деле, пусть $x_0 \in S^2$ и F_0 — слой над точкой x_0 . Выберем точку $\tau^0 \in F_0$ и оставшуюся часть слоя F_0 (одномерную клетку) обозначим через τ^1 . Далее, пусть v — векторное поле на сфере S^2 , имеющее в точке x_0 единственную особенность (с индексом $+2$). Поле v можно рассматривать как двумерную клетку в E . Эта клетка проектируется на всю сферу S^2 с выколотой точкой x_0 , а с каждым слоем (кроме F_0) пересекается в одной точке. Наконец, выбросив из E клетки $\tau^0, \tau^1, \tau^2 = v$, мы получим множество τ^3 , гомеоморфное открытому трехмерному шару. Таким образом, E представляется в виде клеточного разбиения $\{\tau^0, \tau^1, \tau^2, \tau^3\}$. Заметим, что граница клетки $v = \tau^2$ представляет собой дважды обегаемый слой F_0 , т. е. $\partial\tau^2 = 2\tau^1$. Остальные клетки имеют нулевую границу: $\partial\tau^3 = 0, \partial\tau^1 = 0, \partial\tau^0 = 0$. Из этого нетрудно заключить, что числа Бетти пространства E имеют следующие значения: $p_0(E) = p_3(E) = 1, p_1(E) = p_2(E) = 0$. Заметим, что при рассмотрении гомологии по модулю 2 имеем $\partial\tau^2 = 0$, и потому числа Бетти пространства E по модулю 2 имеют вид $p_0 = p_1 = p_2 = p_3 = 1$.

Задачи

206. Докажите, что бутылку Клейна можно представить в виде локально тривиального расслоения, базой и слоем которого является окружность.

207. Докажите, что нормированный касательный пучок двумерного тора T гомеоморфен трехмерному тору.
208. Докажите, что если локально тривиальное расслоение имеет своей базой двумерную сферу, а слоем — окружность, то пространство E этого расслоения можно получить из двух полных торов, склеивая их границами.
209. Докажите, что если базой локально тривиального расслоения является окружность, а слоем — отрезок, то E гомеоморфно либо круговому кольцу, либо ленте Мёбиуса.

Французскому математику Жану Лере принадлежит важная теорема о гомологиях расслоенных пространств. Мы сформулируем ее здесь в упрощенном виде.

Пусть $p: E \rightarrow B$ — некоторое расслоение, базой которого является связный полиэдр, имеющий тривиальную фундаментальную группу, а слоем — произвольный полиэдр. Составим таблицу, которую мы имели для случая, когда E — топологическое произведение фигур B и F , и в этой таблице наметим стрелки «ходом коня» (рис. 193).

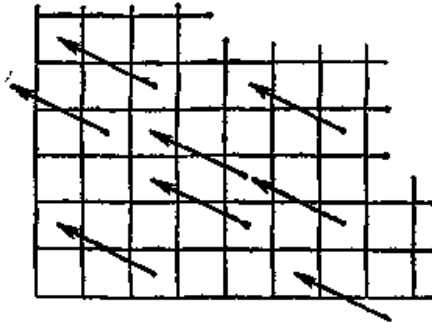


Рис. 193.

На каждой стрелке надпишем некоторое неотрицательное целое число, придерживаясь следующих правил: 1) число, стоящее в каждой клетке, не меньше, чем сумма чисел, надписанных на тех двух стрелках, одна из которых входит в данную клетку, а другая выходит из нее; 2) если начало или конец стрелки выходят за пределы таблицы, то на этой стрелке надписывается число 0. То, что получается, назовем *таблицей* E_2 .

Теперь составим новую таблицу. В каждой клетке поставим новое число, получающееся, если из числа, ранее стоявшего в этой клетке, вычтем сумму чисел, на входящей и исходящей стрелках. Затем

наметим стрелки «удлиненным ходом коня» (рис. 194) и на них надпишем числа по тем же правилам. Это дает *таблицу* E_3 .

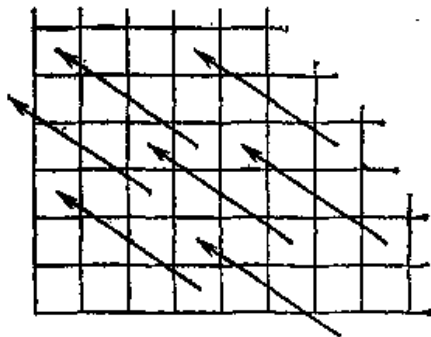


Рис. 194.

Таким же путем из таблицы E_3 получаем *таблицу* E_4 и т. д. Вообще в *таблицах* E_n стрелки ведут на n клеток влево и на $n - 1$ клетку вверх. Ясно, что какую бы мы ни взяли клетку, стоящие в ней числа (в таблицах E_2, E_3, E_4, \dots) в конце концов перестанут меняться, стабилизируются: стрелки становятся все длиннее и в конце концов будут выходить за пределы таблицы. Таблицу, составленную из стабилизировавшихся чисел, назовем *таблицей* E_∞ (стрелок в ней нет). Теорема Лере утверждает, что *существует такой способ надписывания чисел на стрелках, при котором суммирование чисел, стоящих на r -й диагонали таблицы E_∞ , дает r -мерное число Бетти пространства E* . Аналогичная теорема справедлива и для чисел Бетти по простому модулю p .

Пример 65. Пусть $p_1 E \rightarrow B^{\bar{}}$ — некоторое расслоение, базой которого служит сфера S^2 , а слоем — окружность S^1 . Тогда $p_0(B) = p_2(B) = 1, p_0(F) = p_1(F) = 1$, а остальные числа Бетти базы и слоя равны нулю. Поэтому таблица E_2 имеет вид, показанный на рис. 195 (во всех клетках, кроме указанных четырех, записано число 0; на всех стрелках, кроме указанной, надписано число 0).

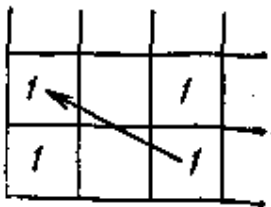


Рис. 195.

На указанной стрелке может быть надписано либо число 0, либо 1. Таблица E_3 совпадает с E_∞ (все стрелки выходят за пределы таблицы). Следовательно, по теореме Лере, пространство E рассматриваемого расслоения может иметь либо числа Бетти $p_0(E) = p_1(E) = p_2(E) = p_3(E) = 1$ (если на стрелке надписан 0), либо число Бетти $p_2(E) = p_3(E) = 1, p_1(E) = p_0(E) = 0$ (если надписана единица). Первая возможность реализуется, например, для топологического произведения сферы S^2 и окружности S^1 (см. задачу 200). Вторая возможность реализуется для касательного пучка сферы S^2 (см. пример 64).

2.3.10. Теория Морса

Наличие горизонтальной касательной — необходимое условие для того, чтобы дифференцируемая функция достигала максимума или минимума (локального) во внутренней точке x_0 своей области определения. Однако это условие не является достаточным: в точке перегиба с горизонтальной касательной функция не достигает ни максимума, ни минимума.

Заметим, что точки максимума и минимума устойчивы относительно малых «шевелений» графика (рис. 196, а). Точка перегиба (с горизонтальной касательной) устойчивостью не обладает: при «шевелении» графика она может пропасть (т. е. вблизи нее не будет точек с горизонтальной касательной); рис. 196, б).

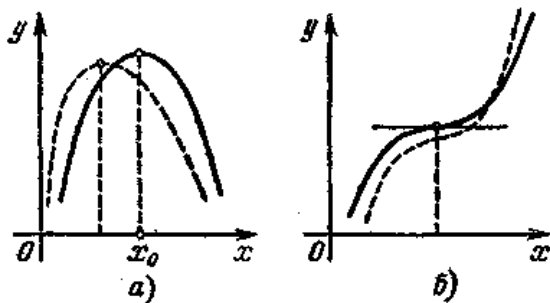


Рис. 196.

Для функций от двух переменных x, y (заданных в некоторой области на плоскости) можно указать аналогичное необходимое условие: для того чтобы функция $f(x, y)$ достигала локального максимума или минимума во внутренней точке (x_0, y_0) своей области определения, необходимо, чтобы эта точка была критической, т. е. чтобы график

функции имел в точке (x_0, y_0) горизонтальную касательную плоскость.

Пример 66. На рисунке 197 изображены графики функций

$$\begin{aligned} f_0(x, y) &= c + x^2 + y^2, & f_1(x, y) &= a + x^2 - y^2, \\ f_2(x, y) &= c - x^2 - y^2. \end{aligned} \quad (21)$$

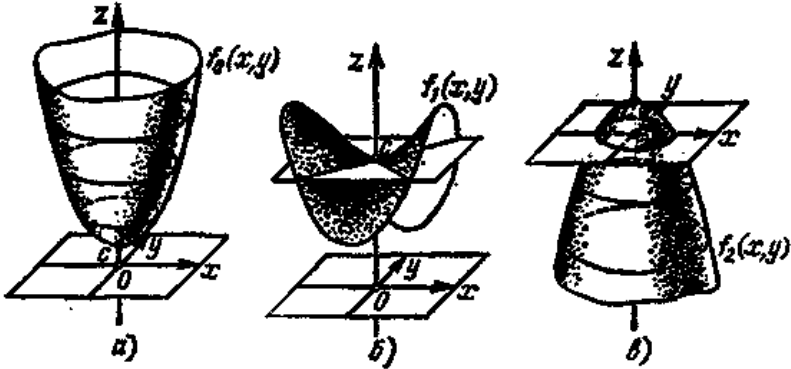


Рис.197

Точка $(0; 0)$ является для каждой из них критической: для $f_0(x, y)$ — это точка минимума, для $f_2(x, y)$ точка максимума, а функция $f_1(x, y)$ не имеет в точке $(0; 0)$ ни максимум ни минимума — это так называемая седловая точка. Все эти точки устойчивы относительно малых «шевелений» графика. Существуют и более сложные критические точки; например, функция $f(x, y) = x^3 - 3xy^2$ имеет в начале координат «седло третьего порядка» (три спуска и три подъема, а не два, как на рис. 197, б). Однако как угодно малым «шевелением» графика можно добиться, чтобы все критические точки стали невырожденными — такими, как на рис. 197.

Можно также рассматривать функции, заданные не на плоскости, а на поверхности — ведь вблизи каждой своей точки поверхность топологически устроена так же, как плоскость.

Пример 67. Для любой точки p , принадлежащей тору T , обозначим через $f(p)$ высоту точки p над горизонтальной плоскостью Π . Эта функция имеет (при показанном на рис. 198 расположении тора) одну (точку максимума a , одну точку минимума d и две седловые точки b, c .

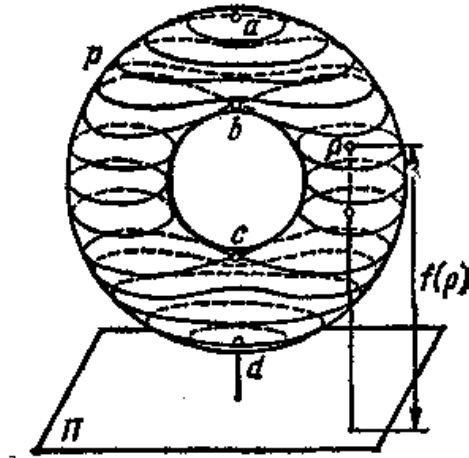


Рис.198.

Таким образом, если мы обозначим через C_0 число точек минимума, через C_1 — число седловых точек, а через C_2 — число точек максимума, то в этом примере $C_0 = 1, C_1 = 2, C_2 = 1$, и потому $C_0 - C_1 + C_2 = 0$.

Пример 69. Для сферы, расположенной обычным образом, та же функция $f(p)$ («высота» точек над горизонтальной плоскостью) имеет две критические точки: точку минимума («южный полюс») и точку максимума («северный полюс»). Таким образом, в этом случае $C_0 = 1, C_1 = 0, C_2 = 1$, т. е. $C_0 - C_1 + C_2 = 2$.

Рассмотренные примеры подводят к формулировке теоремы о критических точках, принадлежащей английскому математику Морсу. Условимся говорить, что точка минимума имеет индекс 0, седловая точка — индекс 1, а точка максимума — индекс 2. Теперь мы можем сформулировать «первую половину» теоремы Морса (для случая поверхности): *пусть на поверхности Q задана функция, имеющая только невырожденные критические точки; тогда*

$$C_0 - C_1 + C_2 = \chi(Q), \quad (22)$$

где C_0 — число критических точек индекса 0 (т. е. точек минимума), C_1 — число критических точек индекса 1 (седел), C_2 — число критических точек индекса 2 (точек максимума).

В самом деле, функция f определяет на поверхности Q *линии уровня* (вдоль каждой из которых функция f принимает постоянное значение). Кроме того, можно на Q наметить *линии наискорейшего спуска*, вдоль которых функция f наиболее быстро убывает; они перпендикулярны

линиям уровня. Векторы, касающиеся линий наискорейшего спуска, образуют векторное поле на поверхности Q . В точках, не являющихся критическими, это векторное поле не имеет особенностей. На рис. 199 показан вид векторного поля вблизи точки минимума (a), седловой точки (b), точки максимума (c).

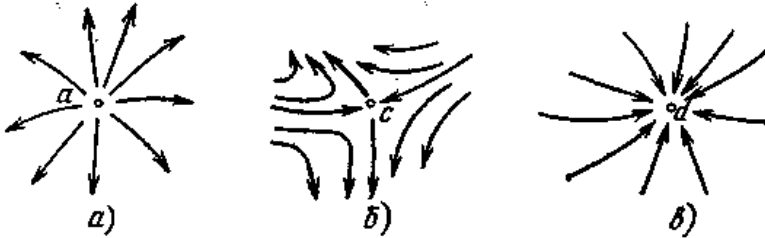


Рис. 199.

Легко проверяется, что $j = (-1)^k$, где j — индекс особенности векторного поля, а k — индекс критической точки (см. рис. 89). Следовательно, рассматриваемое векторное поле имеет C_0 особенностей с индексом $+1$ (минимумы), C_1 особенностей с индексом -1 (седловые точки) и еще C_2 особенностей с индексом $+1$ (максимумы). Из теоремы Пуанкаре о векторных полях теперь следует справедливость формулы (22) для любой замкнутой ориентируемой поверхности.

Задачи

210. Докажите, что формула (22) справедлива и для любой замкнутой неориентируемой поверхности.

211. Докажите, что если на поверхности P_k задана функция, все критические точки которой невырождены, то число критических точек не меньше $2k + 2$.

Рассмотрим теперь «вторую половину» теоремы Морса: пусть на поверхности Q задана функция, имеющая только невырожденные критические точки; тогда

$$C_0 \geq p_0(Q), \quad C_1 - C_0 \geq p_1(Q) - p_0(Q). \quad (23)$$

Для доказательства продолжим предыдущие рассуждения. Будем считать, что значения, которые функция f принимает в критических точках, попарно различны. Мы можем считать критические точки a_1, \dots, a_q перенумерованными таким образом, что $f(a_1) > f(a_2) > \dots > f(a_q)$.

Вблизи a_1 (точки наибольшего максимума, рис. 200, а) линии уровня замкнуты и окружают точку a_1 . Разрезав поверхность по одной из этих линий, мы получим двумерную клетку τ_1 и остаток Q (рис. 200, б).

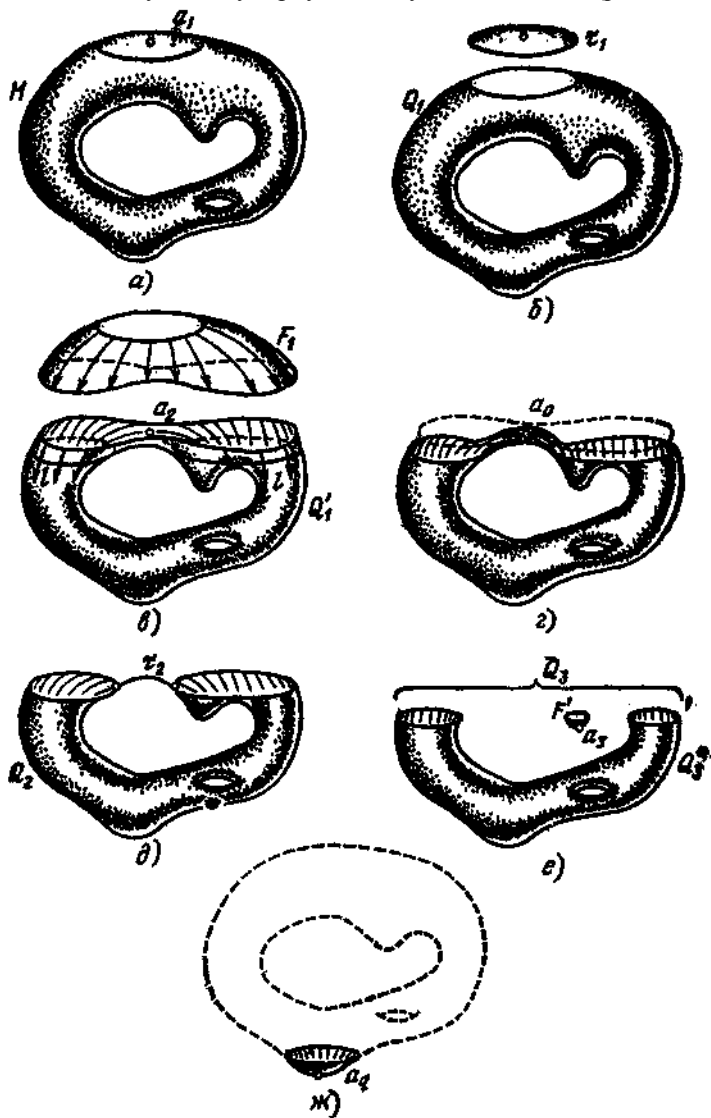


Рис. 200.

Теперь упростим фигуру Q_1 . Для этого рассмотрим линию уровня, проходящую чуть выше точки a_2 , и отрезем от Q_1 часть F_1 , расположенную выше этой линии уровня (рис. 200, в). Оставшуюся часть поверхности обозначим через Q'_1 . Так как между a_1 и a_2 критических точек нет, то вся часть F_1 заполнена «параллельно идущими» линиями наискорейшего спуска, и по этим линиям можно F_1 «сдвинуть» вниз в Q'_1 . Так как при этом сдвиге (как и при всякой гомотопии) любой цикл переходит в гомологичный, то фигура Q_1 имеет те же гомологии, что и Q'_1 .

Рассмотрим линию уровня l , проходящую чуть ниже точки a_2 , и обозначим через Q_2 часть поверхности, лежащую ниже этой линии. Пусть, например, a_2 является седлом. Мы можем сдвинуть (по линиям наискорейшего спуска) Q'_1 в Q_2 — всюду, за исключением окрестности точки a_2 (рис. 200, з). Затем можно сжать оставшуюся «перемычку» в одномерную клетку τ_2 , приклеенную к Q_2 (рис. 200, д). Эти деформации не изменяют гомологии. Следовательно» первоначальная поверхность Q имеет те же гомологии, что и фигура, получающаяся из Q_2 приклеиванием одномерной клетки τ_2 (соответствующей седлу), а затем двумерной клетки τ_1 (соответствующей максимуму).

Далее мы сдвинем Q_2 в часть Q'_2 , лежащую ниже линии уровня, проходящей чуть выше точки a_3 . Если a_3 есть точка минимума, то оставшаяся вблизи точки a_3 часть F' поверхности (т. е. круг) имеет те же гомологии, что и точка, т. е. нульмерная клетка τ_3 (рис. 200, е). Обозначив фигуру, получающуюся из Q_2 отбрасыванием куска F' , через Q'_3 , мы находим, что первоначальная поверхность Q имеет те же гомологии, что и фигура, получающаяся из Q'_3 добавлением нульмерной клетки τ_3 , одномерной τ_2 , двумерной τ_1 .

В конце концов, мы оставим от первоначальной поверхности последнюю точку минимума a_q (рис. 200, ж). Идя обратным путем, мы найдем, что фигуру, имеющую те же гомологии, что и поверхность Q , можно получить последовательным приклеиванием клеток: каждой точке минимума соответствует нульмерная клетка, седлу — одномерная клетка, а максимуму — двумерная. Иначе говоря, Q имеет те же гомологии, что и некоторый полиэдр, содержащий C_0 нульмерных, C_1 одномерных и C_2 двумерных клеток. Из этого и вытекает справедливость неравенств (23) (см. задачу 194).

Задачи

212. Докажите, что соотношения (22), (23) справедливы для чисел Бетти по любому простому модулю p .

213. Докажите, что $C_r \geq p_r(Q)$, $r = 0, 1, 2$.

Заметим в заключение, что можно рассматривать функции и на «многомерных поверхностях» (в задачах 190, 192 и примерах 57, 58 рассматриваются «трехмерные поверхности»). На n -мерной «поверхности» Q имеется $n+1$ типов невырожденных критических точек (в формулах, аналогичных (21), может быть в правой части 0, 1, . . ., n минусов). Теорема Морса (и рассмотренное ее доказательство) сохраняются и в этом случае. Например, неравенства (23) принимают вид

$$\sum_{k=0}^r (-1)^{r-k} C_k \geq \sum_{k=0}^r (-1)^{r-k} P_k(Q), \quad r=0, 1, \dots, n.$$

2.3.11. Использование топологических методов в физике

Многие математические понятия или даже целые теории, появившись на свет, долгие годы живут, не имея приложений вне самой математики. В качестве примера достаточно вспомнить историю с загадочными комплексными числами, вошедшими в физику и технику спустя несколько столетий после своего открытия. Хорошим примером подобного рода до недавнего времени могла служить и топология. Однако в течение последнего десятилетия в нескольких довольно далеких друг от друга областях физики возник ряд задач, получивших свою адекватную формулировку и решение на языке топологии, что позволило значительно продвинуться в соответствующих разделах физики.

Наглядную иллюстрацию к сказанному дает биофизика полимеров, имеющая дело с гигантскими молекулами белков и нуклеиновых кислот. Рассматривая положения, которые молекула может занимать в пространстве, мы сталкиваемся с ограничениями топологической природы. В самом деле, чисто математически длинная замкнутая молекула представляет собой замкнутую линию. Мы знаем, что такие линии образуют узлы. Различные узлы нельзя продеформировать друг в друга без разрыва линии и последующего склеивания концов. Условие неразрывности линий обеспечено тем, что для создания разрыва необходимо разорвать химические связи в данной точке полимерной цепи. Энергетические затраты такого процесса довольно значительны. Поэтому при достаточно низкой температуре вероятность разрывов мала и молекулы полимера могут существовать в состоянии с данной узельной конфигурацией практически неограниченно долго. Важный вопрос, какая часть

молекул из общего числа молекул заданной длины обладает определенной узельной конфигурацией, решается на основе перечисления типов топологически различных узлов, известном из алгебраической топологии.

В биофизике полимеров сами длинные молекулы образуют топологические объекты — узлы. В других областях физики мы приходим к объектам, обладающим топологическими свойствами, уже не столь непосредственным образом. Так, в теории поля фигурируют частицы, математически описываемые векторными полями, с топологическими особенностями. В физике конденсированного состояния с топологией оказалась связана устойчивость ряда дефектов структуры **упорядоченных веществ**: обычных и жидких кристаллов, сверхпроводников, сверхтекучих жидкостей и ферромагнетиков. Ниже мы познакомимся с простейшим веществом, устойчивостью дефектов которого имеет топологическую природу. Это нематический жидкий кристалл, часто называемый просто «нематик». Необходимые математические понятия: индекс векторного поля, фундаментальная группа, степень отображения и др., разъясняются весьма бегло. Точные определения и комментарии к ним можно найти в основном тексте книги.

2.3.11.1. Нематик

Нематический жидкий кристалл состоит из удлинённых молекул, взаимодействие между которыми стремится выстроить их параллельно друг другу. При высоких температурах тепловое движение препятствует этому и вещество представляет обычную жидкость (рис. 201, *a*).

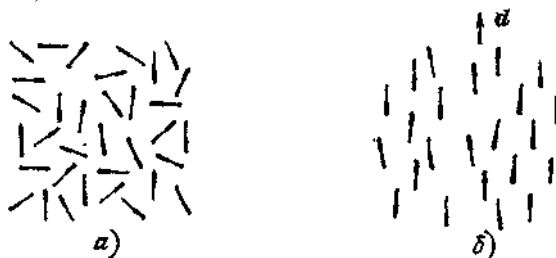


Рис. 201.

При температурах ниже некоторой критической (типичные значения температур перехода для нематиков порядка нескольких десятков градусов Цельсия) в жидкости появляется выделенное направление,

вдоль которого преимущественно ориентированы оси молекул. При этом, как и в обычной жидкости, распределение центров тяжести молекул нематика остается хаотическим (рис. 201, б). Небольшие отклонения осей молекул от параллельного друг другу направления связаны с тепловыми колебаниями. Математически направление преимущественной ориентации описывается с помощью единичного вектора d , называемого *директором*. Особое название вектора d связано с тем, что, хотя концы длинных молекул отличаются друг от друга, но их расположение неупорядоченно (см. рис. 201, б) и состояния нематика с противоположными направлениями вектора d (d и $-d$) физически неразличимы. Другими словами, векторы d можно представлять не как стрелочки, а как палочки, задающие лишь направление (direction), но не его знак.

Из-за влияния стенок сосуда и внешних полей (например, магнитного) состояния нематика всегда неоднородны. Это значит, что направление директора d постепенно меняется от точки к точке. Распределение d в пространстве называется *векторным полем* единичного вектора d .

2.3.11.2. Дисциплинации в тематике

Благодаря сильному рассеянию света нематический жидкий кристалл выглядит как мутная непрозрачная жидкость. Если разглядывать ее в микроскоп, то можно заметить длинные тонкие нити, плавающие в жидкости. Именно они и дали название этому типу жидких кристаллов (нема — по-гречески «нить»). Еще в начале XX века исследователи понимали, а теперь это твердо установленный факт, что нити в нематиках — не посторонние вкрапления, а особенности в расположении молекул.

В самом деле, в поле направлений директора d возможны особые линии, на которых направление d не определено (разрывно). Такие распределения d проще всего изобразить в случае плоских векторных полей, т. е., когда все векторы d в пространстве параллельны некоторой плоскости (см. рис. 202, где поле директора показано черточками).

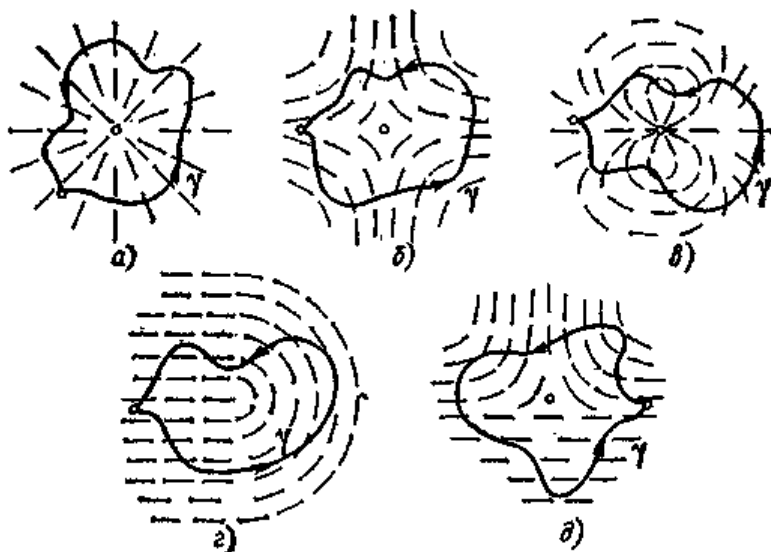


Рис. 202.

Известно, что особые точки векторных полей на плоскости характеризуются *индексами особой точки*, — числом ν полных оборотов в положительном направлении, совершаемых вектором d при обходе особой точки по замкнутому контуру γ . Так, особой точке, изображенной на рис. 202, *а*, соответствует индекс $\nu = 1$, на рис. 202, *б* — $\nu = -1$, на рис. 202, *в* — $\nu = 2$. Как мы помним, состояния, отличающиеся только знаком d , неразличимы, поэтому возможны особые точки, при обходе которых по замкнутому контуру γ вектор d совершает полуцелое число оборотов. Так, индекс особых точек, изображенных на рис. 202, *г* и *д*, равен соответственно $1/2$ и $-1/2$. Особые точки на рис. 202 представляют собой выход в плоскость рисунка особой линии в поле направлений d . Если взглянуть на особую линию рис. 202, *а* не «сверху», а «сбоку», то распределение d будет выглядеть, как показано на рис. 203, *а*. По предложению английского физика Франка линии разрыва в пояе направлений директора были названы *дисклинациями*.

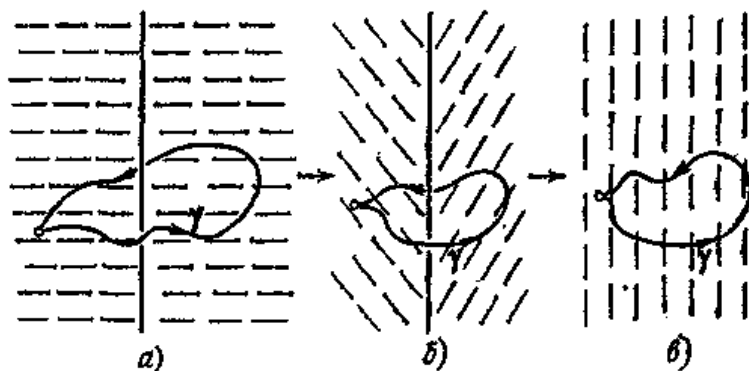


Рис. 203.

Так как взаимодействие между молекулами стремится выстроить их параллельно, наличие особых линий в распределении d энергетически невыгодно. Следовательно, в нематике должны возникать деформации распределения d , стремящиеся устранить особенности и перевести распределение в однородное, обладающее наинизшей энергией. Как это может произойти, легко увидеть на примере дисклинации, изображенной на рис. 203, *a*. Действительно, деформация, показанная на последовательности рисунков 203, *a* — *e*, переводит распределение в однородное, не обладающее никакими особенностями. Эта деформация поля d , напоминающая складывание зонтика, получила название «вытекание в третье измерение», так как направления d , первоначально лежащие в плоскости рис. 202, *a*, выстраиваются в перпендикулярном ей направлении (рис. 203, *в*). Таким образом, мы видим, что дисклинация, изображенная на рис. 202, *a*, 203, *a*, неустойчива по отношению к вытеканию в третье измерение. Спрашивается, устойчивы ли другие дисклинации? Как сформулировать признак, позволяющий отличать устойчивые дисклинации от неустойчивых? Разумеется, любую дисклинацию можно уничтожить, создавая разрывы в поле директора, как это показано на последовательности рисунков 204, *a* — *в*.

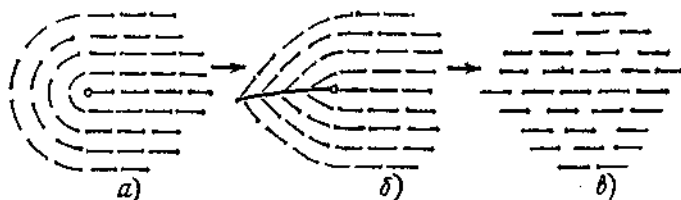


Рис. 204.

Заметим, однако, что в окрестности разрыва молекулы будут упорядочены не параллельно, как в нематике, а под углом друг к другу, как в обычной жидкости. Значит, создание такого разрыва эквивалентно плавлению нематического порядка на целой полуплоскости, опирающейся на особую линию, что требует больших затрат энергии. Иначе говоря, этот процесс обладает колоссальным энергетическим барьером. Поэтому при исследовании устойчивости дисклинаций необходимо сразу ограничиваться лишь непрерывными деформациями поля d . Раз так, здесь может оказаться полезной топология.

2.3.11.3. Дисклинации и топология

Пусть в объеме, заполненном нематиком, имеется какое-то распределение векторов d . Иными словами, каждой точке r нашего объема поставлен в соответствие вектор d — задано векторное поле $d(r)$. Возьмем векторы d из разных точек объема и перенесем параллельно самим себе так, чтобы они начинались из одной точки (рис. 205, а и б).

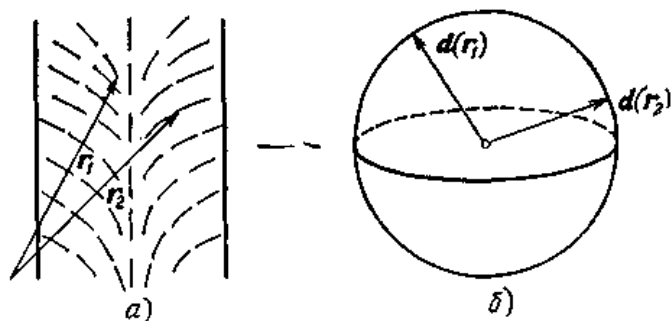


Рис. 205.

Тогда их концы будут лежать на поверхности сферы единичного радиуса, которая тем самым является областью изменения векторов d . Таким образом, мы получаем, что векторное поле $d(r)$ осуществляет отображение точек r нашего объема в точки на поверхности сферы единичного радиуса.

Заметим теперь, что сфера, по которой бегают концы вектора d , не совсем обычная. В самом деле, d не простой вектор, а вектор-директор, и состояния d и $-d$ физически неотличимы. Это значит, что

диаметрально противоположные точки сферы эквивалентны или, как говорят в топологии, сфера склеена своими диаметрально противоположными точками. Такая сфера называется *проективной плоскостью* и обозначается RP^2 . Разумеется, представить результат склеивания диаметрально противоположных точек сферы в трехмерном пространстве невозможно, да и не нужно; достаточно лишь помнить, что точки d и $-d$ — это не две разные, а одна и та же точка. Итак, векторное поле $d(r)$ осуществляет отображение точек r координатного пространства в точки проективной плоскости RP^2 . Посмотрим теперь, какое отношение имеет склейка к вопросу об устойчивости дисклинаций в нематике.

Пусть в поле директора $d(r)$ имеется дисклинационная линия L , т. е. линия, на которой поле $d(r)$ терпит разрыв (рис. 206, а).

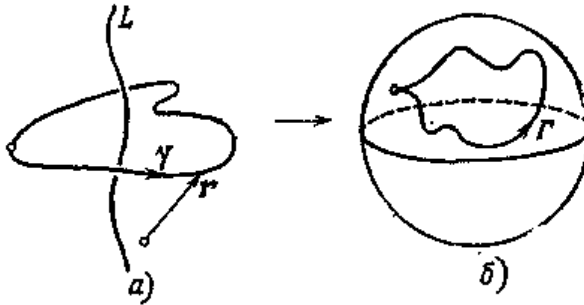


Рис. 206.

Окружим ее замкнутым контуром γ . Каждая точка r этого контура имеет своим образом точку $d(r)$ на поверхности RP^2 , а весь замкнутый контур γ отображается в замкнутый контур Γ на поверхности RP^2 (рис. 206, б). Очевидно, что любой непрерывной деформации поля $d(r)$ в окрестности контура γ соответствует деформация контура Γ на поверхности RP^2 . В частности, деформация распределения $d(r)$, названная вытеканием линии в третье измерение (см. рис. 203), сопровождается стягиванием контура Γ на поверхности RP^2 в точку (рис. 207, а — в).

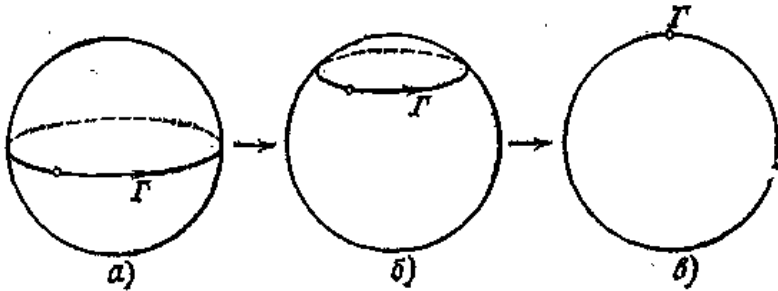


Рис. 207.

Понятно, что и в общем случае устранимым (неустойчивым) дисклинациям на поверхности RP^2 соответствуют контуры Γ , стягиваемые в точку. Будем обозначать класс таких контуров, а вместе с ним и дисклинаций, символом Γ_0 . Легко сообразить, что для плоских полей к этому классу относятся все дисклинации с целым индексом ν векторного поля $d(r)$ (см. рис. 202, а — в). Все контуры класса Γ_0 , а значит, и все векторные поля дисклинаций этого класса можно непрерывно продеформировать друг в друга.

С другой стороны, в нематиках имеются дисклинации, изображенные, например, на рис. 202, г, д, такие, что образами окружающих их контуров ν являются контуры типа $\Gamma_{1/2}$, соединяющие диаметрально противоположные точки сферы (рис. 208).

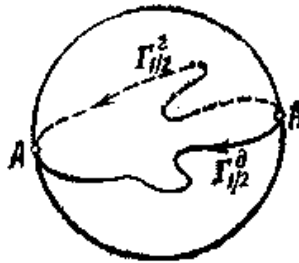


Рис. 208.

Мы знаем, что такие точки эквивалентны, а значит, контуры $\Gamma_{1/2}$ замкнуты. В отличие от контуров типа Γ_0 , контуры $\Gamma_{1/2}$ невозможно стянуть в точку по поверхности RP^2 . В этом отношении контуры типа $\Gamma_{1/2}$ напоминают контуры, оббегающие дырку у бублика. Это легко представить, деформируя и склеивая сферу только в двух

диаметрально противоположных точках A — начала и конца контура $\Gamma_{1/2}$.

Контур типа $\Gamma_{1/2}$ невозможно стянуть в точку, зато их можно продеформировать друг в друга. Дисклинации, изображенные на рис. 202, ε , δ , устойчивы. Соответствующие им распределения невозможно превратить в однородные никакой непрерывной деформацией поля $d(r)$, хотя легко преобразовать друг в друга. Полезно самостоятельно продеформировать поле рисунка 202, ε в поле рисунка 202, δ , для чего предварительно сделать деформацию соответствующих контуров $\Gamma_{1/2}$ на RP^2 , изображенных на рис. 208.

Подчеркнем, что существование замкнутых контуров $\Gamma_{1/2}$, не стягиваемых в точку, а значит, и топологически устойчивых дисклинаций, является исключительно следствием эквивалентности диаметрально противоположных точек d и $-d$. На обычной сфере без всяких склеек любой замкнутый контур можно стянуть в точку и, следовательно, если бы состояния d и $-d$ были отличимы, то в таком веществе (изотропный ферромагнетик) вообще не было бы устойчивых особых линий. Сфера и RP^2 локально (в окрестности каждой точки) эквивалентны, но глобально (в целом) имеют разные топологические свойства!

Итак, с точки зрения топологии в нематическом жидком кристалле имеется всего два типа линий дисклинаций: характеризующихся на проективной плоскости замкнутыми контурами типа Γ_0 , стягиваемыми по ней в точку, либо замкнутыми контурами типа $\Gamma_{1/2}$, не стягиваемыми по ней в точку. Устойчивые линии типа $\Gamma_{1/2}$ не могут заканчиваться внутри объема с нематиком. Они либо замкнуты, либо выходят на поверхность. Докажем это утверждение от противоположного. Если бы они могли заканчиваться в объеме, то тогда с них можно было бы снять и стянуть в точку охватывающий их контур γ , а значит, стянулся бы в точку и его образ $\Gamma_{1/2}$ на проективной плоскости, что невозможно для контуров типа $\Gamma_{1/2}$. Напротив, неустойчивым линиям топология не запрещает заканчиваться внутри объема, однако отрезкам особых линий энергетически выгодно сокращать свою длину и они или исчезают совсем, или, если они одним концом прикреплены к поверхности, сжимаясь, превращаются в поверхностную особую точку. Линии дисклинации взаимодействуют между собой, и в случае притяжения две линии могут, например, сливаться в одну. Каков будет результат слияния? Получится ли из двух устойчивых вновь устойчивая линия или она будет неустойчивой и исчезнет, т. е. исходные линии аннигилируют? Топология дает ответ и на этот

вопрос. На проективной плоскости образом контура $\psi = \gamma_1 + \gamma_2$ (см. рис. 209, а), охватывающего сразу две линии типа $\Gamma_{1/2}$, будет контур $\Gamma_{1/2}$, проходимый дважды, что обозначается как произведение $\Gamma_{1/2} \cdot \Gamma_{1/2}$. Как явствует из рис. 209, б, такой контур эквивалентен контуру Γ_0 и, следовательно, стягиваем в точку.

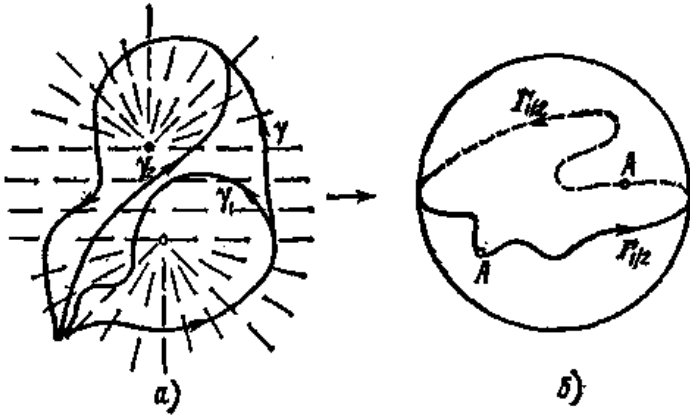


Рис. 209.

Это означает, что две устойчивые дисклинации, сливаясь, превращаются в неустойчивую дисклинацию типа Γ_0 , т. е. аннигилируют. В сказанном можно убедиться и непосредственно, деформируя распределение $d(r)$, изображенное на рис. 209, а, в распределение рис. 202, а. Напротив, слияние устойчивой и неустойчивой дисклинации всегда дает устойчивую, что можно коротко записать как $\Gamma_{1/2} \cdot \Gamma_0 = \Gamma_0 \cdot \Gamma_{1/2} = \Gamma_0$.

Сформулированные законы умножения означают, что множество классов контуров на RP^2 , состоящее из двух элементов Γ_0 и $\Gamma_{1/2}$, образует группу $\pi_1(RP^2)$, называемую *фундаментальной группой* проективной плоскости. Умножение элементов в этой группе можно заменить эквивалентным ему сложением индексов у контуров, при условии, что все целые числа эквивалентны нулю:

$$\Gamma_{1/2} \cdot \Gamma_{1/2} = \Gamma_0 \rightarrow \frac{1}{2} + \frac{1}{2} = 0; \quad \Gamma_{1/2} \cdot \Gamma_0 = \Gamma_{1/2} \rightarrow \frac{1}{2} + 0 = \frac{1}{2}.$$

2.3.11.4. Особые точки

Кроме особых линий в поле вектора $d(r)$, в объеме, занимаемом нема-
тиком, возможны особые точки — точки разрыва в поле $d(r)$.

Простейшим примером такой точки является точка, вокруг которой направления вектора d совпадают с направлениями радиус-вектора, отложенного из этой точки $\mathbf{d}(\mathbf{r}) = \frac{\mathbf{r}}{|\mathbf{r}|}$. Векторы $d(r)$, торчащие во все стороны вокруг особой точки, напоминают иголки свернувшегося в клубок ежа, поэтому эта особая точка получила название *еж*. Устойчив ли еж? Иными словами, можно ли путем непрерывной деформации поля $d(r)$ устранить эту особую точку и превратить поле $d(r)$ в однородное? Для ответа на этот вопрос окружим особую точку сферой σ . Образом сферы σ на RP^2 будет вся поверхность RP^2 , проходимая один раз. Таким образом, поле $d(r)$ вокруг ежа осуществляет *отображение степени единица* сферы σ на RP^2 . Будем непрерывно деформировать поле $d(r)$ — «причесывать» ежа. При этом образ сферы σ , который можно представлять как замкнутую пленку (т. е. пленку без края), обтягивающую RP^2 , будет также деформироваться и образовывать окладки. Но стянуть эту пленку в точку, оставаясь на RP^2 , нам не удастся, а значит, и не удастся устранить особую точку в поле $d(r)$. Степень отображения — топологический инвариант.

В общем случае, для исследования произвольной особой точки в поле $d(r)$ на устойчивость, надо окружить ее сферой σ и проследить за ее образом на RP^2 . Всем устойчивым особым точкам соответствуют пленки — образы σ на RP^2 , целое число раз окутывающие проективную плоскость. Неустойчивым особым точкам соответствуют замкнутые пленки, стягиваемые по RP^2 в точку с степенью отображения $N = 0$. Устойчивые особые точки, как и устойчивые особые линии, можно устранить только путем создания разрывов в поле $d(r)$, что требует преодоления колоссального энергетического барьера. Мы рассматривали такой процесс в случае особой линии (см. рис. 204). В отличие от особой линии, для устранения устойчивой особой точки необходимо создать разрыв в поле $d(r)$ на выходящей из нее линии.

Степень отображения — целочисленный индекс. Возникает вопрос: чем отличаются ежи, имеющие степень разного знака, например, ежи с $N = 1$ и $N = -1$. Они должны отличаться направлением иголок, т. е. разной ориентацией пленок — образов сферы σ . В первом случае иголки торчат наружу, а во втором — внутрь. Но мы знаем, что противоположные направления $d(r)$ неотличимы. Поэтому ежи с $N = 1$ и $N = -1$ — это одна и та же особая точка с $|N| = 1$. С другой стороны, слияние двух ежей, казалось бы, должно приводить к сложению индексов, и значит, если мы сложим два ежа с индексами $|N_1| = 1$ и $|N_2| = 1$, мы можем получить как ежа с индексом 2, так и ежа

с индексом 0, т. е. устранимую (неустойчивую) особую точку. Такое поведение кажется невероятным. А что же будет на самом деле?

Реальный процесс слияния ежей дает результат, зависящий от пути слияния! Этот факт следует из нетривиальности фундаментальной группы тематического жидкого кристалла $\pi_1(\mathbb{R}P^3)$. В топологии его называют *влиянием щ.* Например, слияние особых точек с $|N_1| = 1$ и $|N_2| = 1$ может происходить вдоль путей γ и $\tilde{\gamma}$, проходящих по разные стороны от линии устойчивой дисклинации. На рис. 210, а тонкими линиями показаны линии поля $d(r)$, линия дисклинации перпендикулярна плоскости рисунка и обозначена буквой O .

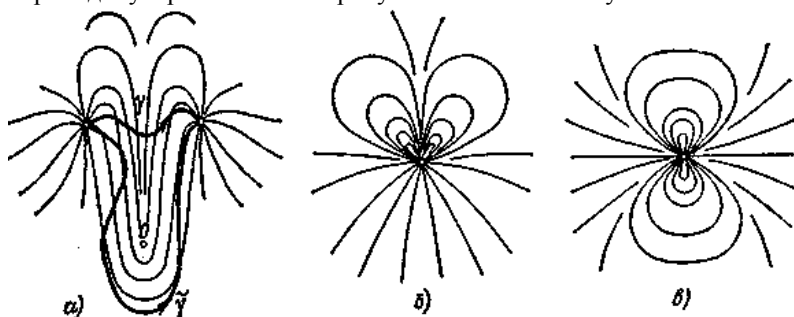


Рис. 210.

Очевидно, слияние вдоль пути γ дает точечную особенность с $|N|=2$, изображенную на рис. 210, б. Обратим внимание, что распределение поля на рис. 210, б не имеет оси симметрии, а только плоскость симметрии, перпендикулярную плоскости рисунка. Слияние ежей вдоль пути $\tilde{\gamma}$ дает точечную особенность с $|N| = 0$, которая показана на рис. 210, в. Здесь распределение $d(r)$ аксиально симметрично относительно горизонтальной оси, проходящей в плоскости рисунка.

Итак, наличие дисклинации в поле $d(r)$ приводит к неоднозначности результата слияния ежей. Разумеется, неоднозначность имеет место только в присутствии топологически устойчивых дисклинаций, соответствующих нетривиальным элементам фундаментальной группы.

Мы видели, что топологические свойства замкнутых контуров и пленок на проективной плоскости дают возможность проанализировать ряд вопросов, связанных с устойчивостью и слиянием дисклинаций и особых точек в тематическом жидком кристалле. Кроме особенностей поля $d(r)$, топология дает возможность расклассифицировать устойчивые неособые конфигурации поля $d(r)$ —

доменные стенки и солитоны, возникающие в поле $d(r)$ в присутствии внешних электрических и магнитных полей.

Нематик не одинок. Существует целый обширный класс упорядоченных веществ: обычные и жидкие кристаллы всех типов, ферро- и антиферромагнетики, сегнетоэлектрики, сверхпроводники и сверхтекучие жидкости, при изучении которых оказываются полезными топологические методы. Поведение особенностей в нематике определялось топологическими свойствами области изменения директора d — проективной плоскости RP^2 . В других упорядоченных веществах реализуются другие типы полей: поля векторов, поля матриц и, соответственно, другие области D изменения параметра порядка. Фундаментальная группа π_1 , вообще говоря, некоммутативна. Из существующих в природе веществ некоммутативную фундаментальную группу $\pi_1(D)$ имеет лишь двухосный (нематический жидкий кристалл, с которым имели дело мы, — одноосный нематический жидкий кристалл) нематический жидкий кристалл. Некоммутативность $\pi_1(D)$ приводит к ряду красивых, но пока еще экспериментально не обнаруженных следствий.

Наиболее впечатляющие приложения топологии появились в связи с открытием в 1972 г. сверхтекучих фаз легкого изотопа гелия — ^3He . Оказалось, что сами сверхтекучие свойства этих фаз в значительной мере диктуются топологией. Об этих свойствах, а также многих других приложениях топологии в физике можно прочесть в работах Г. Е. Воловика и В. П. Минеева.

3. ТОПОЛОГИЯ ОБОЛОЧЕК

3.1. Топологические объекты

Моделирование окружающих нас предметов требует привлечения более сложных геометрических объектов, чем точки, кривые и поверхности. Одной поверхностью в общем случае невозможно описать геометрическую форму некоторого заданного предмета, но это можно сделать с помощью нескольких поверхностей, связав их определенным образом. То, каким образом поверхности будут связаны друг с другом, составляет дополнительную информацию, которой мы снабдим новые объекты, построенные на основе уже рассмотренных геометрических объектов. Нашей целью является получение геометрической модели окружающих предметов. Все эти предметы занимают некоторую часть пространства, или, другими словами,

занимают конечный объем пространства. Для их моделирования нужно описать совокупность поверхностей, отделяющих внутренний объем предмета от остальной части пространства. Для этого потребуется набор определенным образом построенных и обрезанных поверхностей и информация о взаимной связи этих поверхностей — как одна поверхность переходит в другую.

Оболочки.

Поверхности могут быть замкнутыми по одному или двум параметрическим направлениям или незамкнутыми. Незамкнутые поверхности имеют границу. Границей будем называть линию на поверхности, соответствующую движению ее параметров по границе их области определения. Линию на замкнутой поверхности, по которой она замыкается сама на себя, будем называть швом. Поверхности могут стыковаться друг с другом по границам. Можно сказать, что по шву замкнутая поверхность стыкуется сама с собой. Совокупность стыкующихся по границам поверхностей будем называть оболочкой. Оболочка может состоять из одной поверхности или нескольких поверхностей. Также как и отдельная поверхность, оболочка может быть замкнутой и незамкнутой. Замкнутая оболочка не имеет границы. Незамкнутая оболочка имеет одну или несколько границ.

В предыдущих главах мы исследовали геометрические свойства кривых и поверхностей путем определения их количественных характеристик (длин и углов). В данной главе нас будут интересовать свойства геометрических объектов, не зависящие от количественных характеристик. Мы будем рассматривать непрерывную связь между точками геометрических объектов. Предположим, что оболочка выполнена из эластичного неразрываемого и не склеиваемого материала. Исследуем свойства этой оболочки, которые сохраняются при всевозможных ее деформациях. Деформацией будем называть изменение формы оболочки путем растяжения, сжатия, сдвига или изгиба ее поверхности, не приводящее к разрывам и не требующее склеивания поверхностей оболочки. Эластичная оболочка в виде куба может быть деформирована в сферу, или эллипсоид, или оболочку в виде тетраэдра, но не может быть деформирована в тороидальную оболочку.

Сфера, эллипсоид, оболочка в виде тетраэдра или куба могут быть преобразованы друг в друга путем непрерывных и обратимых

отображений. Свойства геометрических объектов, сохраняющиеся при непрерывных и обратимых отображениях одного пространства в другое, изучает топология. С топологической точки зрения сфера, эллипсоид, оболочка в виде тетраэдра или куба эквивалентны. Свойства, характеризующие непрерывность точек некоторой оболочки, являются топологическими свойствами. Несмотря на кажущуюся неопределенность, топологические свойства геометрических объектов связаны с фундаментальными математическими понятиями.

Топология изучает общий случай оболочек, которые могут самопересекаться, иметь или не иметь границы, уходить в бесконечность. Топология оперирует своими объектами, которые несут информацию о их взаимной связи друг с другом, и устанавливает между ними соотношения. При моделировании окружающих нас объектов мы будем строить оболочки из топологических объектов. Они будут нести и количественную геометрическую информацию и топологическую информацию. Количественная геометрическая информация топологического объекта содержится в его геометрическом носителе, которым может являться точка, кривая или поверхность. В данной главе мы сосредоточим внимание на топологических свойствах моделируемых объектов.

Вершины, ребра, циклы, грани.

Рассмотрим оболочки, построенные на основе поверхностей в трехмерном евклидовом пространстве. Для отслеживания связей составляющих оболочку поверхностей дополним поверхности информацией об этих связях и введем топологические объекты. Топологические объекты будут нести одновременно метрическую и топологическую информацию.

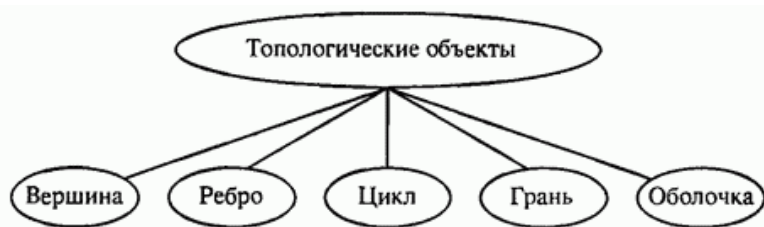


Рис. 1.1. Топологические объекты

Одним из топологических объектов является оболочка. При построении оболочки будем использовать такие топологические объекты, как грани, ребра, вершины и циклы (рис. 3.1). Все топологические объекты имеют общие принципы построения.

Гранью будем называть топологический объект, построенный на основе поверхности. Фактически грань представляет собой поверхность плюс информация о том, какая сторона поверхности является наружной стороной грани, и информация об ее положении в оболочке, т. е. информация об ее соседях. Информация о соседних гранях оформляется в виде циклов.

Цикл — это топологический объект, который описывает одну из границ грани, и содержит информацию о том, где и как к данной грани примыкают соседние грани. Так как вдоль одного цикла к данной грани могут примыкать несколько соседних граней, то цикл состоит из нескольких участков. Каждый участок цикла опирается на некоторое ребро.

Ребром будем называть топологический объект, построенный на основе линии стыковки соседних граней или на основе граничной линии оболочки. Грани стыкуются только по ребрам. Таким образом, каждая грань со всех сторон окружена ребрами. Вершиной будем называть топологический объект, построенный на основе точки, в которой стыкуются ребра. Вершины могут лежать только на краях ребер. Каждое ребро начинается и оканчивается в вершине. Если ребро замкнуто, то оно начинается и оканчивается в одной и той же вершине. Цикл состоит из ребер, образующих замкнутую линию вдоль одной из границ грани. Цикл всегда замкнут и ему приписывается определенное направление. Грань может содержать несколько циклов, причем один из них является внешним, а остальные — внутренними и целиком лежащими внутри внешнего цикла. За положительное направление цикла примем направление движения вдоль цикла, при котором грань всегда находится слева, если смотреть с наружной стороны грани.

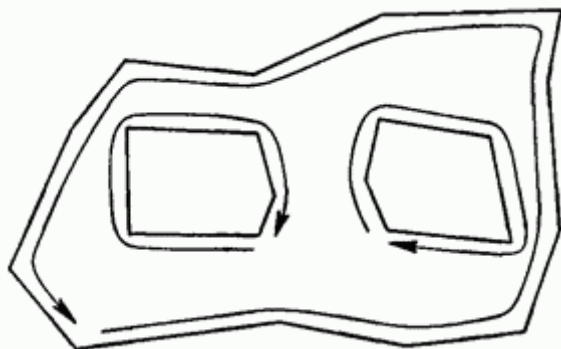


Рис.1.2. Ориентация внешнего и внутренних циклов грани

Таким образом, внешний цикл грани ориентирован против часовой стрелки, а внутренние циклы ориентированы по часовой стрелке, если смотреть с наружной стороны грани. Каждый цикл проходит по одной из границ поверхности. На рис. 5.1.2 приведен пример грани с ее циклами.

Грани, ребра и вершины строятся на базе известных геометрических объектов (точек, кривых и поверхностей) добавлением к ним информации о своих соседях и о взаимной ориентации. В результате геометрические объекты приобретают новое качество, чем и обусловлено введение топологических объектов.

3.2. Эйлерова характеристика оболочек

Топологические свойства оболочки могут быть выражены через количество ее граней, ребер, вершин и циклов. Пусть оболочка содержит F граней, E ребер, V вершин и L циклов. Число вершин, ребер, граней и циклов оболочки связаны между собой соотношением

$$F - E + V + (F - L) = H, \quad (2.1)$$

где величина H называется эйлеровой характеристикой оболочки. Формула (5.2.1) носит имя формулы Эйлера. Если каждая грань оболочки имеет один цикл, то $F-L=0$ и слагаемое в скобках в левой части (5.2.1) может быть опущено. Одну и ту же оболочку можно

построить из различного набора граней. Например, сферическую оболочку можно построить из двух полусфер или из нескольких сферических сегментов подобно футбольному мячу. Как будет видно далее, эйлерова характеристика оболочки не зависит от числа и формы составляющих ее граней, но зависит от природных характеристик оболочки, которые изучает топология.

Рассмотрим, как изменяется эйлерова характеристика оболочки при изменении составляющих ее элементов. Для этого будем изменять состав граней, ребер и вершин некоторого фрагмента оболочки, не изменяя состава остальной части оболочки.

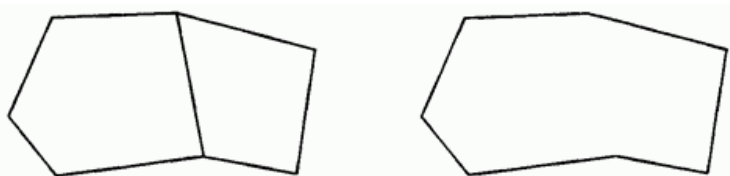


Рис. 2.1. Эйлерова характеристика не изменяется при ликвидации ребра оболочки

При желании оболочкой можно считать только показанный на рисунках фрагмент. На рис. 2.1 показан фрагмент оболочки, у которого ликвидируется одно ребро. Из рисунка видно, что при ликвидации одного ребра число граней, число ребер и число циклов уменьшается на единицу, а эйлерова характеристика оболочки не изменяется.

Если объединить два ребра, ликвидировав общую для них вершину, то число ребер и вершин оболочки уменьшатся на единицу. Если разрезать ребро на две части, вставив вершину, то число ребер и вершин оболочки увеличится на единицу. Эйлерова характеристика оболочки в обоих случаях не изменится. При введении одного дополнительного ребра между существующими вершинами число граней, число циклов и число ребер увеличится на единицу.

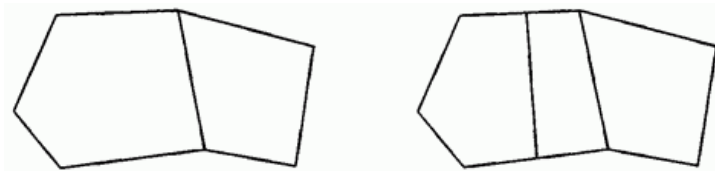


Рис. 2.2. Эйлерова характеристика не изменяется при добавлении ребра в оболочку

При введении дополнительного ребра с добавлением двух новых вершин на его концах число граней и циклов увеличивается на единицу, число ребер увеличивается на три из-за деления двух существующих ребер на две части и добавления нового ребра, а эйлерова характеристика оболочки не изменяется, что показано на рис. 2.2.

Эйлерова характеристика оболочки не изменяется и при введении одного дополнительного ребра, соединяющего существующую вершину и новую вершину, делящую существующее ребро на два ребра.

В приведенных примерах все грани не изменяли число ограничивающих их циклов, а все новые грани имели один цикл. В общем случае грань может иметь вырезы внутри. Грань с вырезами с топологической точки зрения отличается от грани без вырезов, так как первую нельзя преобразовать во вторую путем деформирования.

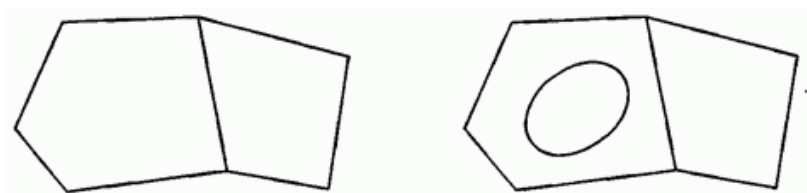


Рис. 2.3. Добавление грани и двух циклов в оболочку

Число вырезов в грани также играет существенную роль. Топологически эквивалентными являются грани, которые путем деформирования могут быть преобразованы одна в другую. Для этого грани должны иметь одинаковое число вырезов или одинаковое число

циклов. На рис. 2.3 показано добавление новой грани в оболочку путем введения замкнутого ребра, целиком лежащего внутри существующей грани.

Число граней, ребер и вершин при этом увеличится на единицу, а число циклов увеличится на два (цикл на добавленном замкнутом ребре должен быть посчитан дважды: один раз в одной грани, второй раз — в другой грани). Эйлера характеристика оболочки при этом также не изменится. Эйлера характеристика оболочки не изменится, если мы преобразуем грань с двумя циклами в грань с одним циклом, что приведено на рис. 5.2.4.

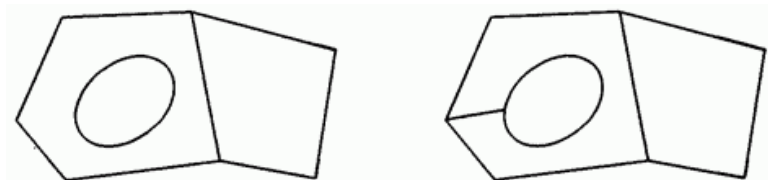


Рис. 2.4. Добавление ребра и ликвидация внутреннего цикла грани

Покажем это. Пусть новое ребро начинается и оканчивается в уже существующих вершинах, тогда число ребер увеличится на единицу, число циклов уменьшится на единицу, а эйлера характеристика не изменится.

Все перечисленные модификации оболочки не изменяют ее эйлеровой характеристики. Это иллюстрирует то, что эйлера характеристика не зависит от способа разбиения оболочки на грани, а зависит только от природы оболочки.

Рассмотрим еще один пример, показывающий, что эйлера характеристика зависит от топологии оболочки, и не зависит от способа раскроя ее на грани.

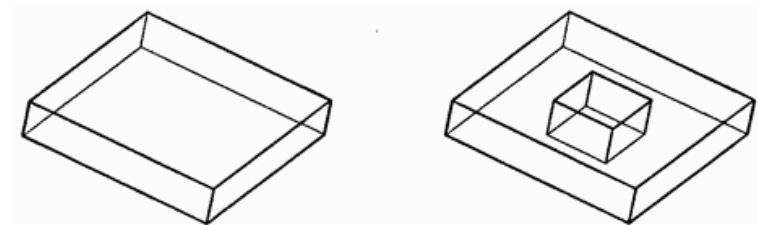


Рис. 2.5. Эйлера характеристика оболочек различна

Возьмем оболочку в форме четырех угольной призмы и превратим ее в оболочку в форме четырехугольной призмы с четырехугольным отверстием, что показано на рис. 2.5.

Исходная оболочки имела следующие числа граней, циклов, ребер и вершин: $F=6$, $L=6$, $E=12$, $V=8$ и ее эйлера характеристика равна $H=2$. Результирующая оболочки имеет: граней $F=10$, циклов $L=12$, ребер $E=24$, вершин $V=16$, а ее эйлера характеристика равна $H=0$. В двух гранях увеличилось число циклов (было по одному циклу, а стало по два).

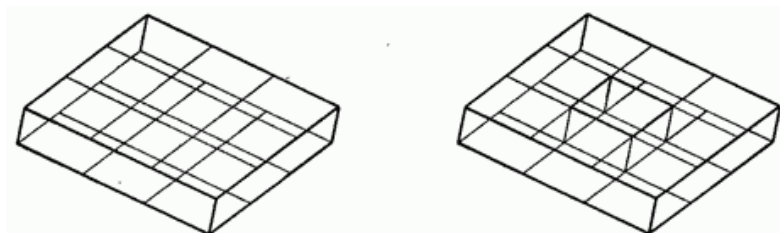


Рис. 2.6

Эйлера характеристика новой оболочки уменьшилась на две единицы. Повторим переход от той же призматической оболочки к призматической оболочке с вырезом, только верхнюю и нижнюю грань исходной оболочки представим в виде совокупности девяти граней, как показано на рис. 2.6.

Исходная оболочки имела следующие числа граней, циклов, ребер и вершин: $F=22$, $L=22$, $E=52$, $V=32$ и ее эйлера характеристика равна $H=2$. Результирующая оболочки имеет : граней $F=24$, циклов $L=24$,
302

ребер $E=56$, вершин $V=32$, а ее эйлерова характеристика равна $H=0$. Все грани в исходной и результирующей оболочках имеют по одному циклу. Эйлерова характеристика результирующей оболочки, как и предыдущем примере, уменьшилась на две единицы.

Это произошло в результате того, что мы получили оболочку, которая топологически неэквивалентна исходной оболочке. Действительно, представив, что обе оболочки, показанные на рис. 5.2.5, выполнены из легко деформируемого материала, скруглив углы, из первой оболочки получим сферу, а из второй — тор (рис. 2.7).



Рис. 2.7. Сфера и тор имеют различную топологию

Никакими деформациями невозможно из сферы получить тор — они имеют разную топологию.

Эйлерова характеристика оболочки отражает ее природные свойства, связанные с возможностью деформировать одну оболочку в другую или, другими словами, установить между ними взаимно однозначное соответствие. Для описания топологических свойств оболочек и их граней используется понятие связности. Оболочки одинаковой связности могут быть деформированы одна в другую при условии, что они имеют равное число границ и одинаковую ориентируемость.

3.3. Связность оболочек

Исследуем топологические свойства оболочек. Эти свойства не зависят от количественных характеристик (объема, площади поверхности и т.п.). Не будем обращать внимания на ребра и вершины, а будем воспринимать оболочку как единое целое. Представим, что оболочка выполнена из легко деформируемого неразрываемого и не склеиваемого материала. Мы можем изменить форму оболочки, но до

определенных пределов. Например, если оболочка имеет форму шара, то ей можно придать ему форму куба или тарелки, но нельзя придать ему форму тора или кувшина с ручкой. Данный факт является отражением того, что упомянутые шар, куб и тарелка имеют одинаковую топологию, которая отличается от топологии тора и кувшина с ручкой. Топологической характеристикой описанного свойства оболочек является связность. В приведенном примере оболочки с формой шара, куба или тарелки, имеют одну и ту же связность, которая отлична от связности оболочки с формой тора или кувшина с ручкой. Так как топология характеризует непрерывную связь точек объектов, то для определения топологических свойств следует изучить поведение топологических объектов при разделении их на части.

Связность — это топологическое понятие, характеризующее целостность некоторого объекта. С другой стороны связность отражает возможность разделить некоторый топологический объект на отдельные части. Простейшей является незамкнутая оболочка с одной границей (одним циклом) топологически эквивалентная кругу. Если представить, что оболочки выполнены из легко деформируемого материала, то простейшей является оболочка, которой путем деформирования можно придать плоскую форму с границей в виде окружности.

Прямоугольная, эллиптическая, любая плоская оболочка с одной границей является простейшей. Рассмотрим вопрос: какое минимальное число линий можно провести на поверхности оболочки, чтобы по этим линиям ее можно было бы разрезать на две простейшие? Простейшая оболочка делится на две отдельные части любой линией, проведенной от одной точки границы до другой. Простейшей оболочке приписывается связность равная единице. Ее называют односвязной.

Рассмотрим незамкнутые оболочки, имеющие конечное число циклов (границ), которые путем деформирования без склеивания и наложений можно сделать плоскими. На рис. 3.1 приведены примеры плоских оболочек с различным числом циклов.



Рис. 3.1. Плоские оболочки различной связности

Если плоская оболочка имеет L циклов (один внешний и $L-1$ внутренних), то на ней можно провести $L-1$ линий, не разрезающих оболочку на две отдельные части (например, от внешнего цикла к каждому внутреннему циклу). Любая следующая линия, начинающаяся и оканчивающаяся на границе оболочки, разрежет ее на две части. Плоская оболочка с вырезами обладает дополнительными связями, которые мы режем, а оболочка не теряет свою целостность. Целостность оболочки заключается в том, что из некоторой ее точки, двигаясь по ее поверхности, можно попасть в любую другую ее точку. Плоской оболочке с L циклами припишем связность равную L .

Связность оболочки определим минимальным числом линий, по которым ее можно разрезать на две простейшие оболочки (на две отдельные части). Если связность оболочки равна h , то $h-1$ разрезов достаточно, чтобы ими раскроить оболочку, превратив ее в простейшую. Связность обозначим через h .

Рассмотрим замкнутые оболочки. На оболочке как на поверхности можно свободно строить кривые линии. Предположим, что мы можем резать оболочку по кривым на ней. Если оболочка имеет топологию сферы, то по любой замкнутой кривой на ее поверхности оболочку можно разрезать на две отдельные части, представляющие собой односвязные оболочки. Если оболочка имеет топологию тора, то для раскроя ее в односвязную грань, потребуется, как минимум, две замкнутые кривые на ней. Способ раскроя тороидальной оболочки показан на рис. 3.2.

Используя определение, получим, что связность сферы равна единице, а связность тора равна трем.

На замкнутой оболочке связности h можно построить n замкнутых кривых, которые не нарушают ее целостности (которые не разрезают ее на отдельные части), но нельзя построить h таких кривых.



Рис. 3.2. Раскрой тороидальной оболочки



Рис. 3.3. Раскрой цилиндрической оболочки

Для того, чтобы замкнутую оболочку связности h раскроить в односвязную оболочку, ее нужно разрезать по $h-1$ замкнутым линиям. Эти линии не разрезают оболочку на две отдельные части, но любая совокупность, состоящая из h замкнутых кривых на оболочке, обязательно разрезает оболочку на две части.

Рассмотренные замкнутые оболочки имеют нечетную связность. Существуют оболочки четной связности. Четную связность могут иметь незамкнутые оболочки. Например, цилиндрическая оболочка конечной длины имеет связность равную двум. Действительно, линия, проходящая от одной границы до другой, не разрезает цилиндрическую оболочку на отдельные части, а только раскраивает ее в односвязную оболочку (рис 5.3.3). Любая другая линия, проходящая от одной границы до другой, разрежет цилиндрическую оболочку на части.

На рис. 3.4 показаны незамкнутые оболочки, полученные из замкнутых оболочек, если в последних сделать отверстия. Если в замкнутой оболочке связности h выполнить одно отверстие с одной замкнутой границей, то связность полученной незамкнутой оболочки будет равна связности исходной замкнутой оболочки. Каждое последующее

отверстие с одной замкнутой границей будет увеличивать связность полученной оболочки на единицу. Связность приведенных на рис. 3.4 незамкнутых оболочек соответственно равна 1, 4, 3, 6.

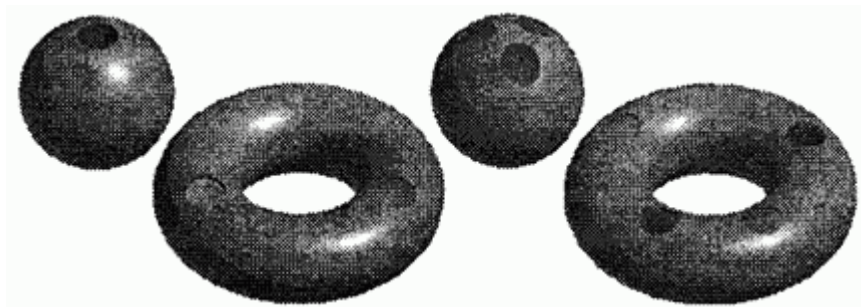


Рис. 3.4. Незамкнутые оболочки

Система разрезов на этих оболочках строится аналогично системе разрезов на соответствующей замкнутой оболочке, только часть линий проводится от одной границы к другой границе.

Связность не является единственной характеристикой оболочки. Две оболочки могут иметь одинаковую связность, но быть топологически различными. Кроме связности оболочка характеризуется ориентируемостью.

3.4. Ориентируемость оболочек

Для многих замкнутых оболочек одну из сторон можно определить как внутреннюю, а другую — как наружную. Для точек оболочек вводится такое топологическое понятие как ориентируемость. Представим, что вокруг всякой точки оболочки проведена окружность достаточно малого радиуса, расположенная на поверхности оболочки. Для каждой окружности определим такое направление обхода, что достаточно близкие точки всегда будут обходиться в одном и том же направлении. Если для некоторой оболочки это можно сделать, то такая оболочка называется ориентируемой. Существуют оболочки, для которых нельзя ввести единое направление обхода для окружностей близких точек. Такие оболочки называются неориентируемыми.

Лист Мёбиуса.

Примером неориентируемой оболочки является лист Мёбиуса, показанный на рис. 4.1. Лист Мёбиуса можно получить, взяв бумажную полосу и склеив дальние края, повернув предварительно их друг относительно друга на 180° . До склеивания краев полосы ее стороны можно покрасить двумя разными цветами. Если покраску проводить после склеивания, то окажется, что мы окрасим одним цветом обе стороны. При движении по листу Мёбиуса мы пройдем по обеим его сторонам.

Для точек листа Мёбиуса нельзя определить ориентацию. Действительно, задав для малой окружности некоторой точки ориентацию и двигаясь по оболочке, мы попадем в исходную точку, но с противоположным направлением. Результатом этого является невозможность окрасить разные стороны оболочки в разные цвета. У оболочки всего одна сторона. Лист Мёбиуса является односторонней оболочкой. Если оболочка является односторонней, то она не ориентируема. Справедливо и утверждение, что если оболочка является двусторонней, то она ориентируема.

Оболочка тогда и только тогда неориентируема, когда на ней можно построить такую замкнутую кривую s , что при движении вдоль этой кривой достаточно малой ориентированной окружности она придет в исходную точку ориентированной в противоположном направлении.



Рис. 4.1. Лист Мёбиуса — односторонняя незамкнутая оболочка



Рис. 4.2. Ориентируемая самопересекающаяся незамкнутая оболочка

Если двигаться вдоль кривой s на односторонней оболочке по одну сторону от этой кривой, то можно оказаться по другую сторону кривой, хотя при движении кривая не пересекалась.

Лист Мёбиуса является незамкнутой оболочкой. Существуют замкнутые односторонние оболочки. Односторонняя оболочка не может разбить пространство на внутреннюю и внешнюю части, поэтому односторонняя замкнутая оболочка всегда пересекает сама себя. Однако не всякая самопересекающаяся оболочка является односторонней. Оболочка, показанная на рис. 4.2, самопересекающаяся, но не односторонняя.

Бутылка Клейна.

Примером замкнутой односторонней оболочки является бутылка Клейна, которая показана на рис. 4.3. Бутылка Клейна имеет одну замкнутую линию самопересечения.

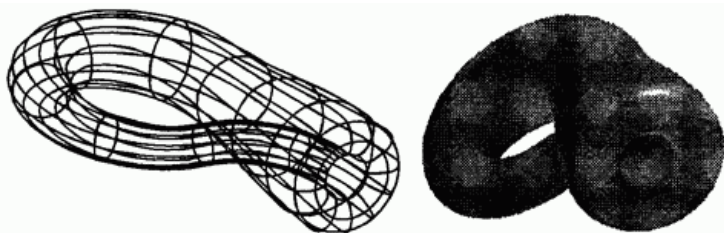


Рис. 4.3. Бутылка Клейна — односторонняя замкнутая оболочка

Она не может служить сосудом. Связность бутылки Клейна равна трем. Система линий, разрезающих бутылку Клейна на две односвязные части, аналогична системе линий тора.

Если бутылку Клейна разрезать плоскостью ее симметрии, то получим две незамкнутые самопересекающиеся оболочки, из которых путем деформирования можно получить два листа Мёбиуса.

Гептаэдр.

Еще одной односторонней оболочкой является гептаэдр. Его можно получить из октаэдра. Для этого удалим четыре грани из восьми: на верхней части октаэдра — левую переднюю и правую заднюю, а на нижней части — правую переднюю и левую заднюю грани, и добавим три квадратные взаимно ортогональные грани, построенные на диагоналях октаэдра. Грани гептаэдра по отдельности приведены на рис. 4.4. Оболочка гептаэдра состоит из семи граней. Ребра и вершины гептаэдра совпадают с ребрами и вершинами октаэдра (диагонали не считаются ребрами, а являются линиями самопересечения оболочки). В каждом ребре гептаэдра стыкуются только две грани. Гептаэдр является замкнутой односторонней оболочкой четной связности. Он показан на рис. 4.5. Если мы начнем движение по поверхности гептаэдра из точки А, то, двигаясь по указанной на рис. 4.5 траектории, попадем в точку В, расположенную на другой стороне поверхности.

Таким образом, кроме связности оболочки характеризуются еще и ориентируемостью. Тор и бутылка Клейна обладают одинаковой связностью, обе замкнутые оболочки, но тор является ориентируемой оболочкой, а бутылка Клейна — неориентируемой.

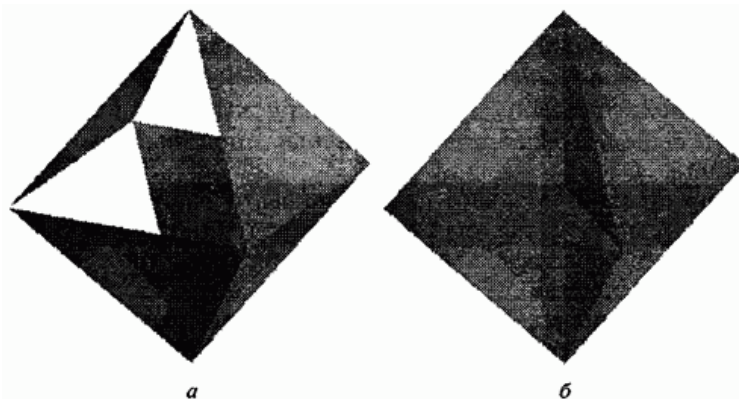


Рис. 4.4. Четыре грани октаэдра (а), грани на диагоналях октаэдра (б)

Известно еще несколько неориентируемых оболочек. Замкнутые неориентируемые оболочки пересекают сами себя, и поэтому не могут быть использованы для моделирования деталей.

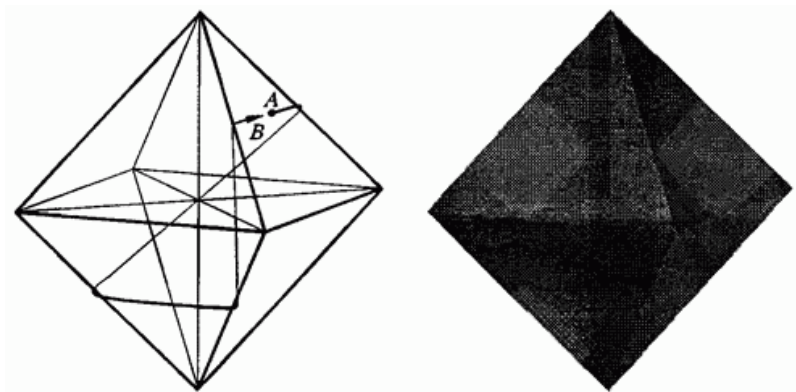


Рис. 4.5. Гептаэдр (семигранник) — самопересекающаяся односторонняя оболочка

Оболочки, с которыми мы сталкиваемся в реальности, являются ориентируемыми. Тем не менее, неориентируемые оболочки с математической точки зрения равноправны с ориентируемыми оболочками.

3.5. Оболочки для моделирования тел

Рассмотрим взаимно однозначное и непрерывное отображение одной оболочки на другую. При этом отображении соседние точки остаются соседними. Такое отображение может исказить оболочку, но при этом связанные части останутся связанными. Одним из видов отображения является деформация. При деформации топологический объект как целое непрерывно переходит сам в себя. Движение оболочки в пространстве является частным случаем деформации, тогда как зеркальное отражение оболочки относительно плоскости не является деформацией. При зеркальном отражении изменяется на обратное направление обхода всякой замкнутой кривой на оболочке, тогда как деформация сохраняет направление обхода неизменным.

Все многообразие оболочек можно классифицировать с топологической точки зрения. К одному и тому же типу относятся оболочки, которые топологически могут быть непрерывно и взаимно однозначно отображены одна на другую. Для этого должны быть выполнены следующие три условия:

- 1) оболочки должны иметь одинаковую связность;
- 2) оболочки должны быть либо ориентируемы, либо неориентируемы;
- 3) оболочки должны быть либо замкнуты, либо должны иметь одинаковое число границ.

Эти три необходимых условия непрерывного и взаимно однозначного отображения двух оболочек также являются и достаточными условиями.

Естественно, что оболочки, имеющие разное число границ, не могут быть отображены одна на другую.

Связность обуславливает существование системы разрезов оболочки, которая при топологическом отображении переходит в систему разрезов такой же структуры на другой оболочке. Следовательно, оболочки различной связности не могут быть отображены одна на другую.

Можно доказать, что всякая оболочка, которая может быть отображена на ориентируемую оболочку, также является ориентируемой. Это определяет третье условие принадлежности оболочек к одному типу.

Оболочки, имеющие одинаковую связность, ориентируемость и число границ, являются топологически эквивалентными.

Формула Эйлера-Пуанкаре.

Замкнутые, ориентируемые и не пересекающие сами себя оболочки имеют нечетную связность. Для таких оболочек эйлерова характеристика H связана с ее связностью h соотношением

$$H = 3 - h. \quad (5.1)$$

Используя это соотношение, получим формулу, связывающую число граней F , число циклов L , число ребер E и число вершин V оболочки с ее связностью h :

$$F - E + V + (F - L) = 3 - h. \quad (5.2)$$

Данная запись формулы Эйлера справедлива для замкнутых оболочек.

Связность не достаточно удобна для характеристики замкнутой оболочки. Введем еще одно понятие, которым можно заменить связность. Из тора путем его деформирования можно получить оболочку, по форме напоминающую гирию, которую будем называть сферой с ручкой. В общем случае любой замкнутой оболочке путем деформирования можно придать форму сферы с G ручками. Так, если взять толстую плиту, пробить в ней G отверстий и скруглить все ребра, то получим объект, оболочка которого топологически эквивалентна сфере с G ручками. На оболочке, топологически эквивалентной сфере с G ручками, можно провести $2G$ замкнутых кривых линий, по которым она раскраивается в простейшую оболочку. Любая следующая замкнутая линия разрежет оболочку на две простейшие оболочки. Сфера с G ручками имеет связность $h=1G+1$. Оболочка реальной детали топологически эквивалентна сфере с некоторым числом ручек. На рис. 5.1 приведена сфера с четырьмя ручками.

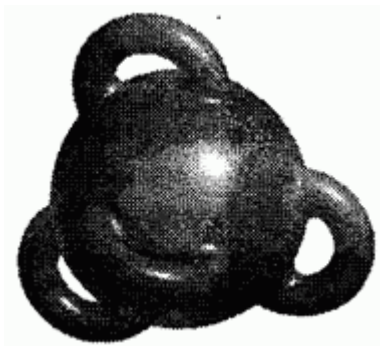


Рис. 5.1. Сфера с четырьмя ручками

Более наглядной, чем связность, характеристикой топологии оболочки может служить число ручек G сферы, к которой путем деформирования можно привести замкнутую оболочку. Примем во внимание, что сфера с G ручками имеет связность $h=1G+1$ и получим формулу, связывающую число граней F , ребер E , вершин V и циклов L с характерной величиной G ,

$$F - E + V + (F - L) - 2(1 - G) = 0. \quad (5.3)$$

Данная формула называется формулой Эйлера-Пуанкаре. Величина G (genus) характеризует топологический тип оболочки. Формула Эйлера-Пуанкаре позволяет определить топологический тип оболочки, если известно число ее граней, ребер, вершин и циклов:

$$G = 1 - F + \frac{E + L - V}{2}. \quad (5.4)$$

Реальные объекты могут иметь внутри пустоты. В этом случае объекты будут описываться несколькими оболочками. Одна из этих оболочек является внешней, остальные оболочки лежат внутри нее. Если моделируемый объект имеет m пустот, то он будет описываться $S=m+1$ оболочками. Потребуем, чтобы внутренние оболочки не пересекали друг друга и внешнюю оболочку. Для каждой оболочки справедлива формула (5.3), а для объекта с S замкнутыми оболочками формула Эйлера-Пуанкаре (5.3) примет вид

$$F - E + V + (F - L) - 2(S - G) = 0, \quad (5.5)$$

где F — общее число граней модели, E — общее число ребер модели, V — общее число вершин модели, L — общее число циклов модели, S — общее число оболочек моделируемого объекта, G — топологический тип моделируемого объекта, равный общему числу ручек всех описывающих его оболочек. Таким образом, топологически эквивалентными объектами будут являться два объекта, у которых равно число описывающих их оболочек и соответствующие внешние и внутренние оболочки имеют одинаковый топологический тип.

Если оболочка не является замкнутой, то говорить о ее топологическом типе G не имеет смысла. Для незамкнутой ориентируемой оболочки формула Эйлера имеет вид

$$F - E + V + (F - L) + h = 2. \quad (5.6)$$

Однородные оболочки.

Для моделирования деталей подходят замкнутые двусторонние не пересекающие сами себя оболочки. К ним относятся оболочки, топологически эквивалентные сфере с G ручками. Именно такие замкнутые оболочки мы и будем использовать для геометрического моделирования. К вершинам, ребрам, циклам и граням предъявим следующие требования. Грани не должны пересекать сами себя. Грани стыкуются только по ребрам, причем в каждом ребре стыкуются только две грани. Оболочка, приведенная на рис. 5.2, некорректна, так как в ребре АВ стыкуются четыре грани.

Такая оболочка топологически должна быть представлена одним из способов, приведенных на рис. 5.3 и 5.4. Ребра стыкуются только в вершинах. В каждой вершине может стыковаться любое конечное число ребер.

Каждую вершину можно обойти по поверхности оболочки, но при этом мы должны пересечь все стыкующиеся в вершине ребра и посетить все примыкающие к вершине грани. Такие вершины будем называть простыми.

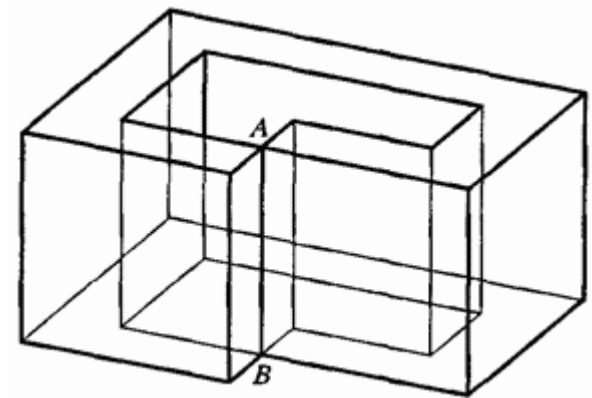


Рис. 5.2. В каждом ребре должны пересекаться только две грани

Оболочка, приведенная на рис. 5.5, также некорректна, так как вершину *A* можно обойти по поверхностям граней, не пересекая всех стыкующихся в ней ребер. Такая оболочка топологически должна быть представлена способом, приведенным на рис. 5.6.

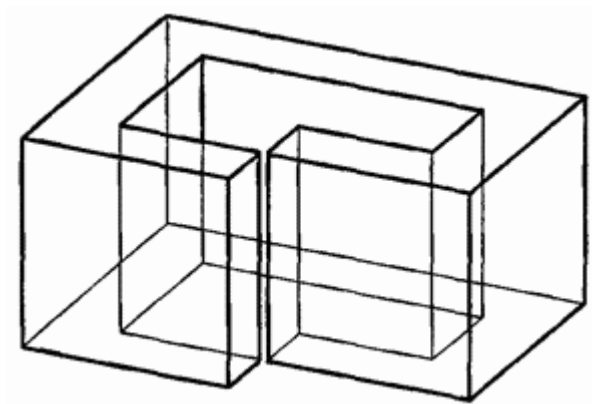


Рис. 5.3. Корректная оболочка

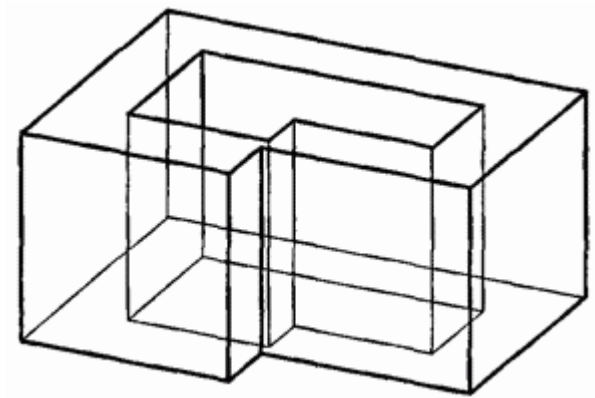


Рис. 5.4. Корректная оболочка

Таким образом, для моделирования мы будем использовать оболочки, вершины которых являются простыми, грани пересекаются по ребрам, причем в каждом ребре стыкуются только две грани.

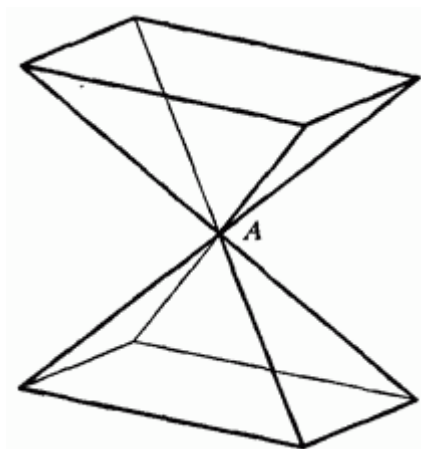


Рис. 5.5. Некорректная оболочка

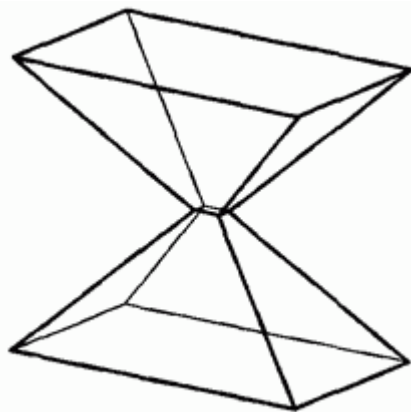


Рис. 5.6. Корректная оболочка

Такие оболочки будем называть однородными. У однородных оболочек все топологические элементы построены по единым правилам.

3.6. Поверхностное и твердотельное моделирование

В геометрическом моделировании используются термины «поверхностное моделирование» (моделирование поверхностей) и «твердотельное моделирование» (моделирование твердых тел). В обоих случаях результатом моделирования является некоторая оболочка (или несколько оболочек), описывающая поверхность моделируемого объекта. Но процесс моделирования в первом случае отличается от процесса моделирования во втором случае.

В поверхностном моделировании сначала создаются и модифицируются требуемым образом поверхности, описывающие отдельные элементы моделируемого объекта. Эти поверхности обрезают по линиям пересечения, сопрягают друг с другом поверхностями скругления или перехода, а также выполняют над ними другие операции. Затем из полученных поверхностей собирают оболочку. В поверхностном моделировании результирующая оболочка не обязательно должна быть замкнутой. Она может отражать лишь часть (главную часть) моделируемого объекта. Поверхностное

моделирование позволяет сосредоточить усилия на сложных формах объекта и широко применяется для проектирования кузовов автомобилей и планеров самолетов.

В твердотельном моделировании с самого начала работа идет с оболочками тел, а не с отдельными поверхностями. Оболочки полностью описывают поверхности моделируемых объектов, отделяющие их внутренний объем от остальной части пространства. Процесс построения оболочки тела в данном случае аналогичен процессу изготовления моделируемого объекта. Сначала создается оболочка некоторой заготовки простой формы. Далее оболочка заготовки изменяется необходимым образом. Для этого используются булевы операции над телами, операция построения тонкостенного тела из заготовки, операция скругления ребер, операция построения ребер жесткости и другие операции. С помощью операций оболочке тела придается требуемая форма.

Два подхода к моделированию имеют много общего и отличаются технологией создания модели. В обоих случаях выполняются аналогичные действия, но в разной последовательности. В следующей главе мы рассмотрим моделирование твердых тел, так как оно включает в себя все элементы поверхностного моделирования.

4. МОДЕЛИРОВАНИЕ ТЕЛ

4.1. Математическая модель тел

Точки, линии и поверхности являются математическими абстракциями. У них хотя бы один размер полагается равным нулю. Реальные объекты имеют все размеры, отличные от нуля, и занимают некоторый конечный объем в пространстве. Для геометрического моделирования предметов, занимающих конечный объем, в математике используются объекты, называемые твердыми телами или просто телами. Способ их описания отличается от способа описания кривых и поверхностей. При моделировании тел строятся поверхности, отделяющие занимаемую ими часть пространства от остальной части пространства. Существует несколько подходов к описанию тел.

Многие предметы можно смоделировать, используя только плоские поверхности. Такое представление тел называется плоскогранным

(Faceted representation или Faceted). Для описания криволинейных поверхностей плоскогранное представление может аппроксимировать их некоторым количеством пластин треугольной или четырехугольной формы. Использование плоских поверхностей значительно упрощает выполнение операций над телами. Плоскогранное представление широко применяется в строительстве и компьютерной графике для получения тоновых изображений.

Некоторые поверхности можно описать уравнениями в координатной форме (представить поверхности неявно). К ним относятся поверхности второго порядка, поверхность тора и другие. Используя для моделирования тел такие поверхности, мы приходим к конструктивной твердотельной геометрии (Constructive Solid Geometry или CSG). Конструктивная твердотельная геометрия оперирует примитивами, к которым, как правило, относят прямоугольную призму, треугольную призму, сферу, цилиндр, конус и тор. Над примитивами и полученными из них телами можно выполнять различные операции (в первую очередь булевы операции). Используемые конструктивной твердотельной геометрией поверхности (сферическая, цилиндрическая, коническая, поверхность тора и плоскость) делят пространство на две части и для них можно указать, с какой стороны поверхности находится внутренний объем тела. Неявное представление поверхностей дает возможность получить линии их пересечения в аналитической форме. Конструктивная твердотельная геометрия позволяет моделировать большинство промышленных деталей.

Наиболее общий подход к описанию тел состоит в представлении тела совокупностью ограничивающих его объем оболочек, грани и ребра которых заданы параметрически. Каждая оболочка строится из набора стыкующихся друг с другом поверхностей произвольной формы, содержащих полную информацию о своих границах и связях с соседями. Такое описание тел называется представлением с помощью границ (Bounded representation или B-rep).

Оно дает возможность выполнять над телами множество операций, сохраняя при этом единый способ их «внутреннего устройства». Представление тел с помощью границ позволяет моделировать объекты произвольной формы и сложности.

Все перечисленные подходы к описанию тел используют топологические объекты и удовлетворяют условиям связности, ориентируемости и замкнутости. Мы будем рассматривать представление тел с помощью границ, опираясь на такие топологические объекты, как вершина, ребро, грань и оболочка. Оболочки тела должны быть однородными, т. е. должны быть описаны по единым правилам. Оболочки состоят из набора граней. Каждая грань базируется на некоторой поверхности. Грань отличается от поверхности тем, что кроме поверхности она в структуре своих данных несет информацию о связях с соседними гранями и об ориентации по отношению к внутреннему объему тела. Там, где это не будет создавать путаницы, грани будем обозначать так же, как и поверхности, на которых они базируются.

Если взять любую деталь, то можно заметить, что ограничивающие ее поверхности делят пространство на две части: одну часть пространства занимает деталь и она находится внутри (вне) ограничивающих оболочек, а другая часть пространства лежит вне (внутри) детали. Перейти из одной части пространства в другую часть пространства, не пересекая ограничивающие оболочки, нельзя. Ограничивающие оболочки как бы изолируют одну часть пространства от другой. Тело может иметь одну или несколько оболочек. Если тело имеет пустоты, то его объем ограничен несколькими оболочками. Одна из этих оболочек является внешней, а остальные оболочки — внутренними. Внутренние оболочки ограничивают пустоты и целиком лежат внутри внешней оболочки. Оболочки тела не должны пересекать друг друга и сами себя. Для описания тел подходят двусторонние (ориентируемые) оболочки. Одной своей стороной каждая оболочка обращена внутрь тела, а другой — наружу. Для того чтобы отличать сторону оболочки, направленную наружу тела, от стороны, направленной внутрь тела, каждой точке оболочки приписывается нормаль, которая считается направленной наружу тела. Нормаль внешней оболочки тела направлена вне ограничиваемой ею части пространства, а нормаль внутренней оболочки тела направлена внутрь ограничиваемой ею части пространства. Таким образом, внутренние оболочки тела как бы вывернуты наизнанку.

С математической точки зрения одна внешняя оболочка и, возможно, несколько внутренних оболочек описывают тело. Другими словами, для создания математической модели тела достаточно смоделировать совокупность оболочек, ограничивающих его объем.

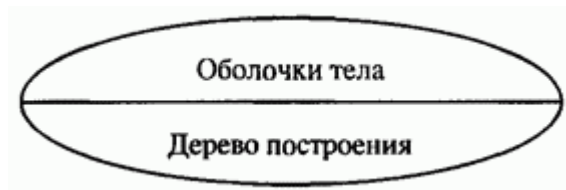


Рис. 1.1. Математическая модель тела

Но для редактирования тела необходима информация о последовательности и способах его построения, поэтому в модель тела включают еще и дерево построения (или протокол построения) тела (рис. 1.1).

В большинстве случаев оболочки, описывающие тело, являются замкнутыми, но тело можно описать и незамкнутой оболочкой, если ее края стянуты в точку. Привлечение топологии при моделировании тел необходимо для корректного выполнения операций над ними.

Топологические объекты были введены в предыдущей главе. Здесь мы остановимся на структуре данных этих объектов. Структура данных должна быть такой, чтобы при желании можно было найти всех соседей любого топологического объекта. Для этого они должны нести информацию о взаимной связи друг с другом. Будем строить оболочки тел из топологических объектов, которые будут нести и количественную геометрическую информацию, и топологическую информацию.

Оболочка тела состоит из набора граней. Грань строится на основе поверхности, которая входит в структуру данных грани. Поверхность является геометрическим носителем грани. Одна сторона грани направлена наружу оболочки, другая внутрь. Грань должна содержать признак ориентации нормали поверхности наружу или внутрь грани. Пусть признак принимает положительное значение, если нормаль поверхности направлена наружу грани, отрицательное значение — в противном случае.

Грани стыкуются между собой по ребрам, лежащим на линиях пересечения граней. Топологический объект ребро строится на основе линии пересечения поверхностей, стыкующихся в ребре граней. Пусть в ребре стыкуются грани, построенные на поверхностях $r(u, v)$ и $s(a, b)$.

Линия пересечения граней описывается двумя поверхностями и двумя двумерными линиями — каждая в пространстве параметров соответствующей поверхности. На грани, базирующейся на поверхности $\mathbf{r}(u, v)$ двумерную линию обозначим векторной функцией $l_{uv}(t)=[a(t)b(t)]^T$, а на грани, базирующейся на поверхности $\mathbf{s}(a, b)$ двумерную линию обозначим векторной функцией $l_{ab}(t)=[u(t)v(t)]^T$. Таким образом, линию пересечения граней, на которой базируется ребро, будем записывать в виде

$$\begin{aligned}
 l_{uv}(t) &= \begin{bmatrix} \mathbf{u}(t) \\ \mathbf{v}(t) \end{bmatrix} \equiv [\mathbf{u}(t) \quad \mathbf{v}(t)]^T, & l_{uv}(t) \in \mathbf{r}(u, v), \\
 l_{ab}(t) &= \begin{bmatrix} \mathbf{a}(t) \\ \mathbf{b}(t) \end{bmatrix} \equiv [\mathbf{a}(t) \quad \mathbf{b}(t)]^T, & l_{ab}(t) \in \mathbf{s}(a, b), \\
 & & t_{\min} \leq t \leq t_{\max}.
 \end{aligned}
 \tag{1.1}$$

Ребро может иметь совпадающее со своей линией направление или ей противоположное. Для этого в структуре данных ребра должен быть соответствующий признак. Пусть этот признак принимает положительное значение, если ребро направлено по кривой, и отрицательное значение — в противном случае.

Ребра, ограничивающие грань, входят в структуру данных грани в виде циклов. Цикл — это топологический объект, характеризующий границу грани. Цикл всегда замкнут и имеет определенное направление. Он состоит из списка ребер и их ориентации в цикле. Ориентацию ребра в цикле грани будем описывать признаком, который назовем флагом. Пусть ребру, направление которого совпадает с направлением цикла, приписывается положительный флаг, а ребру, направление которого не совпадает с направлением цикла, приписывается отрицательный флаг. Таким образом, цикл состоит из списка ребер в порядке их следования и списка соответствующих им флагов.

Грань может содержать несколько циклов, причем один из циклов является внешним, а остальные циклы — внутренними. Внутренние циклы должны целиком лежать внутри внешнего цикла.

Цикл будем считать направленным так, чтобы при движении вдоль него грань всегда находилась бы слева, если смотреть с наружной стороны грани. Таким образом, внешний цикл грани ориентирован против часовой стрелки, а внутренние циклы ориентированы по часовой стрелке, если смотреть навстречу нормали грани.

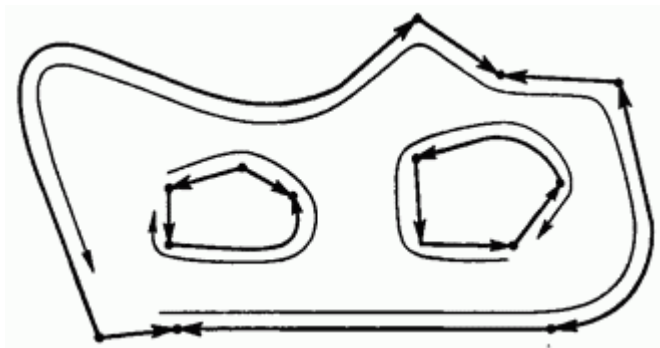


Рис. 1.2. Грань и ее структура данных: поверхность, ее ориентация и список циклов

По каждому циклу можно построить контур на поверхности, который описывает границу грани. На рис. 1.2 приведен пример плоской грани с указанием направлений ее циклов.

Ребро, разделяющее две грани, входит в два цикла: в одном цикле направление ребра совпадает с направлением цикла, а в другом — противоположно направлению цикла. Ребро входит в цикл грани слева от него с положительным флагом, а в цикл грани справа от него с отрицательным флагом.

Ребро начинается и оканчивается в вершинах. Каждая вершина базируется на точке в пространстве и содержит информацию о ребрах, стыкующихся в ней. В каждой вершине могут стыковаться несколько ребер. Если ребро замкнутое, то оно начинается и оканчивается в одной и той же вершине.



Рис. 1.3. Ребро АВ и его структура данных

Ребро должно содержать информацию о том, какая грань находится слева, а какая — справа от ребра, если смотреть с внешней стороны оболочки вдоль направления ребра, и какая вершина находится в начале ребра и какая — в конце. На рис. 1.3 показано ребро, начинающееся в вершине А и оканчивающееся в вершине В.

Описанная структура данных топологических объектов содержит двунаправленную связь между ними. Оказавшись на одном из топологических объектов, можно последовательно пройти всю оболочку, в которую входит этот объект.

Таким образом, математическая модель тела содержит количественную геометрическую информацию в виде поверхностей, линий и точек их стыковки, топологическую информацию в виде связей точек поверхности тела между собой и информацию о последовательности и способах построения в виде дерева построения.

4.2. Простейшие тела

Способы моделирования деталей часто повторяют технологический процесс их производства. Один из способов моделирования тел заключается в том, что берется некоторая заготовка тела и Затем путем удаления и добавления в определенных местах дополнительного объема (материала) получают тело требуемой формы. В качестве заготовок могут братья тела простейшей формы: прямоугольная призма, цилиндр, конус, шар, тор и другие. Простейшие тела состоят

из одной оболочки, построенной по общим правилам. Рассмотрим построение оболочки некоторых простейших тел.

Для каждого простейшего тела нам потребуется местная декартова прямоугольная система координат, радиус-вектор начала которой обозначим через \mathbf{p} , а базисные орты обозначим через i_x, i_y, i_z

Прямоугольная призма.

Начало местной системы координат поместим в одну из вершин призмы, а ее орты направим по ребрам, стыкующимся в этой вершине. Пусть в направлении орта i_x тело имеет длину, равную x в направлении орта i_y — длину, равную y , а в направлении орта i_z — длину, равную z . Прямоугольная призма состоит из шести граней. Каждая грань представляет собой часть плоскости, ограниченную прямоугольным контуром на ней, с признаком ориентации нормали плоскости наружу тела и одним циклом. Контуров состоят из отрезков прямых . Прямоугольная призма имеет 12 ребер. Каждое ребро состоит из линии пересечения поверхностей соседних граней и признака совпадения направления ребра с направлением линии пересечения. Каждая линия пересечения состоит из двух линий на поверхности: одна на поверхности одной грани, другая на поверхности второй грани. Обе линии на поверхности имеют одинаковую геометрическую и параметрическую длину и полностью совпадают в пространстве. Каждая линия на поверхности представляет собой совокупность поверхности и двухмерной линии на ней. Прямоугольная призма с ориентацией циклов граней показана на рис. 2.1.

Грани тела будут описываться поверхностями:

$$\begin{array}{lll}
 \mathbf{r}_1(u_1, v_1) = \mathbf{p} + u_1 \mathbf{i}_x + v_1 \mathbf{i}_y, & 0 \leq u_1 \leq x, & 0 \leq v_1 \leq y, \\
 \mathbf{r}_2(u_2, v_2) = \mathbf{p} + u_2 \mathbf{i}_y + v_2 \mathbf{i}_z, & 0 \leq u_2 \leq y, & 0 \leq v_2 \leq z, \\
 \mathbf{r}_3(u_3, v_3) = \mathbf{p} + u_3 \mathbf{i}_x + v_3 \mathbf{i}_z, & 0 \leq u_3 \leq x, & 0 \leq v_3 \leq z, \\
 \mathbf{r}_4(u_4, v_4) = \mathbf{p} + z \mathbf{i}_z + u_4 \mathbf{i}_x + v_4 \mathbf{i}_y, & 0 \leq u_4 \leq x, & 0 \leq v_4 \leq y, \\
 \mathbf{r}_5(u_5, v_5) = \mathbf{p} + x \mathbf{i}_x + u_5 \mathbf{i}_y + v_5 \mathbf{i}_z, & 0 \leq u_5 \leq y, & 0 \leq v_5 \leq z, \\
 \mathbf{r}_6(u_6, v_6) = \mathbf{p} + y \mathbf{i}_y + u_6 \mathbf{i}_x + v_6 \mathbf{i}_z, & 0 \leq u_6 \leq x, & 0 \leq v_6 \leq z.
 \end{array}$$

У третьей, четвертой и пятой граней нормаль поверхности направлена наружу тела, а у первой, второй и шестой — внутрь тела. Эта информация содержится в грани в виде признака совпадения нормалей.

Приведем описание одного из ребер.

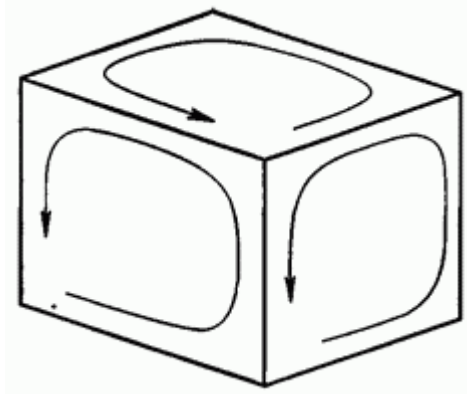


Рис. 2.1. Ориентация циклов граней призматического тела

Например, ребро между первой и второй гранями описывается линией пересечения этих граней, состоящей из двух двумерных отрезков:

$$\begin{aligned}
 l_{uv1}(t) &= [u_1(t) \ v_1(t)]^T = [0 \ yt]^T, & l_{uv1}(t) &\in \mathbf{r}_1(u_1, v_1), \\
 l_{uv2}(t) &= [u_2(t) \ v_2(t)]^T = [yt \ 0]^T, & l_{uv2}(t) &\in \mathbf{r}_2(u_2, v_2), \\
 & & 0 \leq t \leq 1, &
 \end{aligned} \tag{2.2}$$

и двух поверхностей. Выражение $l_{uv1}(t) \in \mathbf{r}_1(u_1, v_1)$ означает, что отрезок $l_{uv1}(t)$ лежит на поверхности $\mathbf{r}_1(u_1, v_1)$, а выражение $l_{uv2}(t) \in \mathbf{r}_2(u_2, v_2)$ означает, что отрезок $l_{uv2}(t)$ лежит на поверхности $\mathbf{r}_2(u_2, v_2)$. Пусть направление ребра совпадает с направлением кривой (2.2), что зафиксируем в признаке совпадения направлений. Признак совпадения направления ребра с направлением цикла грани мы назвали флагом. Флаг может принимать два значения: положительное и отрицательное. Ребро входит в цикл первой грани с положительным флагом, а в цикл второй грани — с отрицательным флагом. Если смотреть вдоль направления ребра снаружи тела, то слева от ребра лежит первая грань, а справа от ребра лежит вторая грань.

Для построения граней тела достаточно знать местную систему координат $\mathbf{p}, \mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z$ и стороны x, y, z призмы.

Цилиндрическое тело.

Другой заготовкой может служить цилиндрическое тело (рис. 2.2). Начало местной системы координат \mathbf{p} поместим в центр одного из торцевов цилиндра, а орт \mathbf{i}_z направим вдоль его оси. Пусть в цилиндр имеет радиус r и длину h .

Цилиндрическое тело имеет три грани. Торцевые грани состоят из частей плоскости, ограниченных окружностями на них, граничного цикла и признака ориентации нормали плоскости грани наружу тела. Цикл каждой торцевой грани состоит из одного замкнутого ребра. Геометрическим носителем такого ребра является линия пересечения, состоящая из двух кривых на поверхности. Одна кривая является окружностью на плоскости, а вторая — линией v_{\min} или v_{\max} на цилиндрической поверхности. Напомним, что обе линии на поверхности, составляющие линию пересечения, должны иметь одинаковую параметрическую длину.

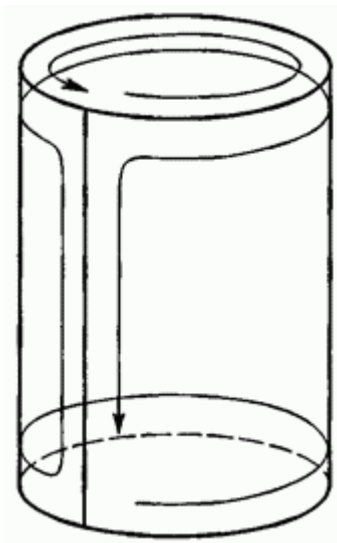


Рис. 2.2. Ориентация циклов граней цилиндрического тела

Боковая грань тела базируется на цилиндрической поверхности

$$\begin{aligned} \mathbf{r}_1(u_1, v_1) &= \mathbf{p} + r \cos u_1 \mathbf{i}_x + r \sin u_1 \mathbf{i}_y + h v_1 \mathbf{i}_z, \\ 0 \leq u_1 \leq 2\pi, \quad 0 \leq v_1 \leq 1. \end{aligned} \quad (2.3)$$

Нормали поверхности (2.3) и ее грани совпадают по направлению. Эта грань имеет один цикл. Цилиндрическая поверхность боковой грани является замкнутой по одному из параметров. Грани основания базируются на ограниченных окружностях плоскостях

$$\begin{aligned} \mathbf{r}_2(u_2, v_2) &= \mathbf{p} + u_2 \mathbf{i}_x + v_2 \mathbf{i}_y, \\ \mathbf{r}_3(u_3, v_3) &= \mathbf{p} + u_3 \mathbf{i}_x + v_3 \mathbf{i}_y + h \mathbf{i}_z, \end{aligned} \quad (2.4)-(2.5)$$

где параметры u_2, v_2 лежат внутри окружности $l_{uv2}(t) = [r \cos t \ r \sin t]^T$, $0 \leq t \leq 2\pi$, на плоскости $\mathbf{r}_2(u_2, v_2)$, а параметры u_3, v_3 лежат внутри окружности $l_{uv3}(t) = [r \cos t \ r \sin t]^T$, $0 \leq t \leq 2\pi$, на плоскости $\mathbf{r}_3(u_3, v_3)$.

Нормали поверхности (2.4) и ее грани противоположны по направлению, а нормали поверхности (2.5) и ее грани совпадают.

Ребро, построенное на линии замыкания оболочки, является швом. Шов, так же как и любое другое ребро, базируется на линии пересечения. В данном случае линия пересечения описывается двумерными кривыми

$$\begin{aligned} l_{uv01}(t) &= [u_1(t) \ v_1(t)]^T = [0 \ ht]^T, & l_{uv01}(t) &\in \mathbf{r}_1(u_1, v_1), \\ l_{uv02}(t) &= [u_1(t) \ v_1(t)]^T = [2\pi \ ht]^T, & l_{uv02}(t) &\in \mathbf{r}_1(u_1, v_1), \\ & & 0 \leq t \leq 1, \end{aligned}$$

являющимися линиями u_{\min} и u_{\max} на цилиндрической поверхности.

Ребро между боковой гранью и основанием (2.4) описывается линией пересечения этих граней, состоящей из двух двумерных кривых (отрезка и окружности)

$$\begin{aligned}
 l_{uv1}(t) &= [u_1(t) \ v_1(t)]^T = [t \ 0]^T, & l_{uv1}(t) &\in \mathbf{r}_1(u, v_1), \\
 l_{uv2}(t) &= [u_2(t) \ v_2(t)]^T = [r \cos t \ r \sin t]^T, & l_{uv2}(t) &\in \mathbf{r}_2(u_2, v_2), \\
 & & & 0 \leq t \leq 2\pi.
 \end{aligned}$$

Отрезок $l_{uv1}(t)$ лежит на поверхности $\mathbf{r}_1(u_1, v_1)$, окружность $l_{uv2}(t)$ лежит на поверхности $\mathbf{r}_2(u_2, v_2)$. Пусть направление ребра совпадает с направлением кривой (2.7), что отметим соответствующим признаком совпадения направлений. Это ребро входит в цикл первой грани с положительным флагом, а в цикл второй грани — с отрицательным флагом. Если смотреть вдоль направления ребра снаружи тела, то слева от ребра лежит первая грань, а справа от ребра лежит вторая грань.

Ребро между боковой гранью и основанием (2.5) описывается линией пересечения этих граней, состоящей из двух двумерных кривых (отрезка и окружности)

$$\begin{aligned}
 l_{uv1}(t) &= [u_1(t) \ v_1(t)]^T = [t \ h]^T, & l_{uv1}(t) &\in \mathbf{r}_1(u_1, v_1), \\
 l_{uv3}(t) &= [u_3(t) \ v_3(t)]^T = [r \cos t \ r \sin t]^T, & l_{uv3}(t) &\in \mathbf{r}_3(u_3, v_3), \\
 & & & 0 \leq t \leq 2\pi.
 \end{aligned} \tag{2.8}$$

Отрезок $l_{uv1}(t)$ лежит на поверхности $\mathbf{r}_1(u_1, v_1)$, окружность $l_{uv3}(t)$ лежит на поверхности $\mathbf{r}_3(u_3, v_3)$. Пусть направление ребра совпадает с направлением кривой (2.8), что отметим соответствующим признаком совпадения направлений. Это ребро входит в цикл первой грани с отрицательным флагом, а в цикл третьей грани — с положительным флагом. Если смотреть вдоль направления ребра снаружи тела, то слева от ребра лежит третья грань, а справа — первая грань.

Циклы граней основания содержат всего одно ребро. Цикл боковой грани состоит из списка ребер с соответствующими флагами:

ребро на базе кривой (2.7) — флаг положительный,

ребро на базе кривой (2.6) — флаг положительный,

ребро на базе кривой (2.8) — флаг отрицательный,

ребро на базе кривой (2.6) — флаг отрицательный.

Цилиндрическое тело и ориентация циклов его граней показаны на рис. 2.2.

Для построения граней тела достаточно знать местную систему координат, радиус и высоту h цилиндра.

Коническое тело.

Тело в форме усеченного конуса строится аналогично цилиндрическому телу с той лишь разницей, что в качестве поверхности боковой грани вместо цилиндрической используется коническая поверхность

$$\begin{aligned} \mathbf{r}_1(u_1, v_1) &= \mathbf{p} + (r + hv_1 \operatorname{tg} \gamma)(\cos u_1 \mathbf{i}_x + \sin u_1 \mathbf{i}_y) + hv_1 \mathbf{i}_z, \\ 0 &\leq u_1 \leq 2\pi, \quad 0 \leq v_1 \leq 1. \end{aligned} \quad (2.9)$$

Грани основания базируются на ограниченных окружностях поверхностей

$$\begin{aligned} \mathbf{r}_2(u_2, v_2) &= \mathbf{p} + u_2 \mathbf{i}_x + v_2 \mathbf{i}_y, \\ \mathbf{r}_3(u_3, v_3) &= \mathbf{p} + u_3 \mathbf{i}_x + v_3 \mathbf{i}_y + h \mathbf{i}_z, \end{aligned} \quad (2.10)-(2.11)$$

где параметры лежат внутри окружности

$$l_{uv2}(t) = [r \cos t \quad r \sin t]^\top, \quad 0 \leq t \leq 2\pi,$$

а параметры лежат внутри окружности

$$l_{uv3}(t) = [(r + h \operatorname{tg} \gamma) \cos t \quad (r + h \operatorname{tg} \gamma) \sin t]^\top, \quad 0 \leq t \leq 2\pi.$$

Ребро, построенное на боковой грани, будет являться швом. В общем случае коническое тело имеет три грани. Если конус не усеченный, то одна из торцевых граней стянута в точку. Стянутую в точку грань

можно исключить из модели тела, тогда оболочка тела с топологической точки зрения будет незамкнутой, хотя диаметр отверстия в ней равен нулю.

Для построения граней тела достаточно знать местную систему координат $\mathbf{p}, \mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z$, радиус r одного из оснований конуса, высоту h и угол конусности γ .

Сферическое тело.

Сферическая поверхность может служить оболочкой сферического тела, но эта оболочка не является замкнутой, так как имеет два отверстия нулевого радиуса в полюсах. Начало местной системы координат \mathbf{p} для сферического тела поместим в центр сферы. Оболочка, описываемая сферической поверхностью

$$\mathbf{r}(u, v) = \mathbf{p} + r \cos v \cos u \mathbf{i}_x + r \cos v \sin u \mathbf{i}_y + r \sin v \mathbf{i}_z,$$
$$0 \leq u \leq 2\pi, \quad -\frac{\pi}{2} \leq v \leq \frac{\pi}{2},$$

топологически эквивалентна цилиндрической поверхности.

Ребра в полюсах стянуты в точку, но двухмерные кривые на сферической поверхности в полюсах имеют ненулевую длину. Цикл грани сферического тела составляют три ребра, одно из которых является швом

$$l_{uv1}(t) = [u(t) \ v(t)]^T = [0 \ t]^T, \quad l_{uv1}(t) \in \mathbf{r}(u, v),$$
$$l_{uv2}(t) = [u(t) \ v(t)]^T = [2\pi \ t]^T, \quad l_{uv2}(t) \in \mathbf{r}(u, v),$$
$$-\frac{\pi}{2} \leq t \leq \frac{\pi}{2},$$

а два других описываются линиями на полюсах сферы:

$$\begin{aligned}
 l_{uv01}(t) &= [u(t) \ v(t)]^T = \left[t \ -\frac{\pi}{2} \right]^T, & l_{uv01}(t) &\in \mathbf{r}(u, v), \\
 l_{uv02}(t) &= [u(t) \ v(t)]^T = \left[t \ \frac{\pi}{2} \right]^T, & l_{uv02}(t) &\in \mathbf{r}(u, v), \\
 & & 0 &\leq t \leq 2\pi.
 \end{aligned}$$

Линии пересечения в полюсах состоят из двух одинаковых линий на сфере. Шов входит в список ребер цикла дважды, один раз с положительным флагом, второй раз — с отрицательным флагом. Сферическое тело показано на рис. 2.3.

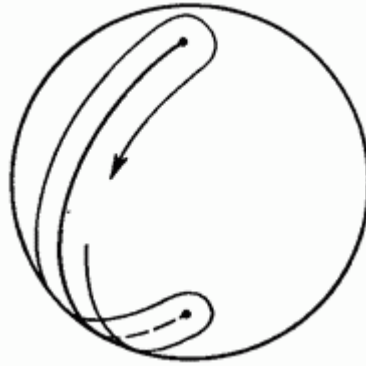


Рис. 2.3. Ориентация цикла грани сферического тела

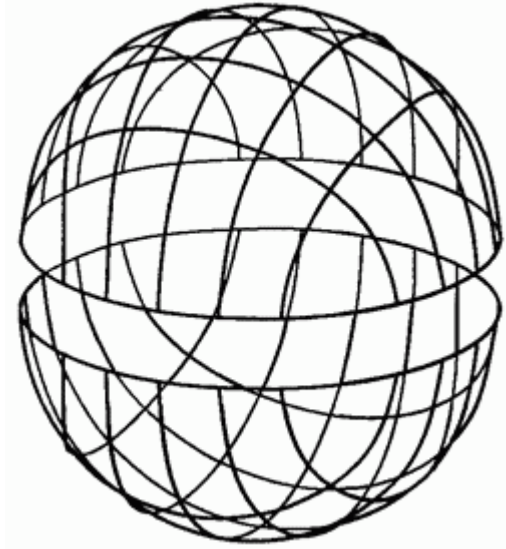


Рис. 2.4. Построение сферического тела по двум полусферам

Сферическое тело можно построить из двух полусфер, описываемых векторными функциями

$$\mathbf{r}_1(u, v) = \mathbf{p} + x\mathbf{i}_x + y\mathbf{i}_y + \sqrt{r^2 - x^2 - y^2}\mathbf{i}_z,$$

$$\mathbf{r}_2(u, v) = \mathbf{p} + x\mathbf{i}_x + y\mathbf{i}_y - \sqrt{r^2 - x^2 - y^2}\mathbf{i}_z, \quad (2.12)-(2.13)$$

где

$$x = x(u, v) = r \left(\cos u + \cos v - 1 + \frac{2u(1 - \cos v - \sin v)}{\pi} + \frac{2v(1 - \cos u - \sin u)}{\pi} \right),$$

$$y = y(u, v) = r \left(\sin u - \sin v - \frac{2u(1 - \cos v - \sin v)}{\pi} + \frac{2v(1 - \cos u - \sin u)}{\pi} \right),$$

$$0 \leq u \leq \frac{\pi}{2}, \quad 0 \leq v \leq \frac{\pi}{2},$$

r - радиус сферы, \mathbf{p} — радиус-вектор центра, $\mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z$ — орты, определяющие ориентацию сферы. Поверхность (2.12) описывает верхнюю полусферу, а поверхность (2.13) описывает нижнюю

полусферу. Нормали верхней поверхности и ее грани совпадают по направлению, а нормали нижней поверхности и ее грани противоположны по направлению. Оболочка сферического тела, состоящая из двух полусфер, является замкнутой. Ее полусферы стыкуются по четырем ребрам:

$$\begin{aligned}
 l_{uv11}(t) &= [u(t) \ v(t)]^T = \left[\frac{\pi t}{2} \ 0 \right]^T, & l_{uv11}(t) &\in r_1(u, v), \\
 l_{uv21}(t) &= [u(t) \ v(t)]^T = \left[\frac{\pi t}{2} \ 0 \right]^T, & l_{uv21}(t) &\in r_2(u, v), \\
 & & 0 \leq t \leq 1; \\
 l_{uv12}(t) &= [u(t) \ v(t)]^T = \left[\frac{\pi}{2} \ \frac{\pi t}{2} \right]^T, & l_{uv12}(t) &\in r_1(u, v), \\
 l_{uv22}(t) &= [u(t) \ v(t)]^T = \left[\frac{\pi}{2} \ \frac{\pi t}{2} \right]^T, & l_{uv22}(t) &\in r_2(u, v), \\
 & & 0 \leq t \leq 1; \\
 l_{uv13}(t) &= [u(t) \ v(t)]^T = \left[\frac{\pi(1-t)}{2} \ \frac{\pi}{2} \right]^T, & l_{uv13}(t) &\in r_1(u, v), \\
 l_{uv23}(t) &= [u(t) \ v(t)]^T = \left[\frac{\pi(1-t)}{2} \ \frac{\pi}{2} \right]^T, & l_{uv23}(t) &\in r_2(u, v), \\
 & & 0 \leq t \leq 1; \\
 l_{uv14}(t) &= [u(t) \ v(t)]^T = \left[0 \ \frac{\pi(1-t)}{2} \right]^T, & l_{uv14}(t) &\in r_1(u, v), \\
 l_{uv24}(t) &= [u(t) \ v(t)]^T = \left[0 \ \frac{\pi(1-t)}{2} \right]^T, & l_{uv24}(t) &\in r_2(u, v), \\
 & & 0 \leq t \leq 1.
 \end{aligned}$$

Поверхности (2.12) и (2.13) для построения сферического тела приведены на рис. 2.4.

Для построения граней тела достаточно знать местную систему координат \mathbf{p} , $\mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z$ и радиус r сферы.

Тороидальное тело.

Начало местной системы координат \mathbf{p} для тороидального тела поместим в его центр, а орт \mathbf{i}_z направим по оси симметрии тела. Пусть больший радиус тора равен R , а малый радиус тора равен r .

Тороидальное тело имеет одну грань, описываемую тороидальной поверхностью

$$\mathbf{r}(u, v) = \mathbf{p} + (R + r \cos v) \cos u \mathbf{i}_x + (R + r \cos v) \sin u \mathbf{i}_y + r \sin v \mathbf{i}_z$$

$$0 \leq u \leq 2\pi, \quad -\pi \leq v \leq \pi, \quad (2.14)$$

два ребра

$$l_{uvr1}(t) = [\mathbf{u}(t) \quad \mathbf{v}(t)]^\top = [t \quad -\pi]^\top$$

$$l_{uvr2}(t) = [\mathbf{u}(t) \quad \mathbf{v}(t)]^\top = [t \quad \pi]^\top,$$

$$0 \leq t \leq 2\pi,$$

$$l_{uvr1}(t) = [\mathbf{u}(t) \quad \mathbf{v}(t)]^\top = [0 \quad t]^\top,$$

$$l_{uvr2}(t) = [\mathbf{u}(t) \quad \mathbf{v}(t)]^\top = [2\pi \quad t]^\top,$$

$$-\pi \leq t \leq \pi, \quad (2.15)-(2.16)$$

и одну вершину в точке пересечения ребер $\mathbf{p}+(R+r) \mathbf{i}_x$. Нормаль поверхности и грани тороидального тела совпадают по направлению. Грань тела имеет один цикл. Цикл грани состоит из списка ребер с соответствующими флагами:

ребро на базе кривой (2.15) — флаг положительный,

ребро на базе кривой (2.16) — флаг положительный,

ребро на базе кривой (2.15) — флаг отрицательный,

ребро на базе кривой (2.16) — флаг отрицательный.

Тороидальная поверхность грани является замкнутой по обоим параметрам, поэтому оба ребра оболочки тела замкнуты и являются швами. Тороидальное тело, его ребра и цикл грани показаны на рис. 2.5.

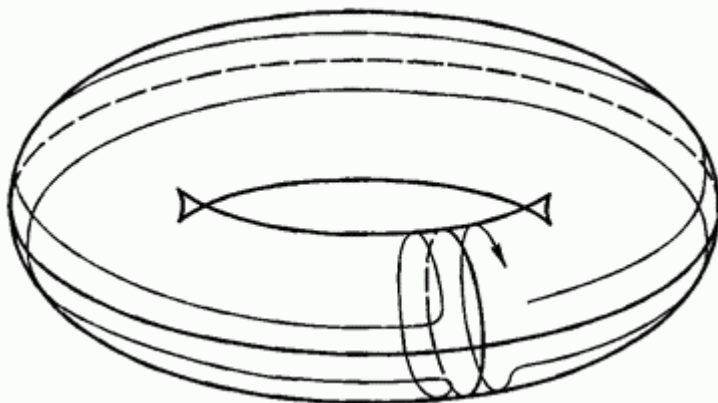


Рис. 2.5. Ориентация цикла грани тороидального тела

Для построения граней тела достаточно знать местную систему координат \mathbf{p} , $\mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z$ большой радиус R и меньший радиус r .

Принцип построения тел.

Рассмотренные тела иллюстрируют принцип построения математической модели тел. Эти тела называют твердотельными примитивами. К ним могут быть отнесены еще некоторые тела простой формы, например, треугольная призма (клин). Все тела построены по такому же принципу, что и твердотельные примитивы.

Может возникнуть вопрос: для чего требуется так усложнять модель тела, в частности, для чего требуется строить ребра на кривых пересечения поверхностей? Действительно, тела можно было бы описать набором поверхностей, не используя ни грани, ни ребра, ни вершины. У реальных деталей эти поверхности могут иметь очень сложную форму, как в смысле кривизны, так и в смысле границ, и эти поверхности каким-то образом нужно построить. Одним из удобных способов построения поверхностей, описывающих тело, является способ одновременного построения всех требуемых поверхностей с помощью операций над телами. Для этого берется одно из простых тел, и далее в определенных местах к нему добавляется объем или от него отнимается объем. Например, для того чтобы просверлить отверстие в некотором теле, выполняется булева операция вычитания из этого тела цилиндрического тела, играющего роль сверла. Аналогично

выполняются пазы и вырезы. Для того чтобы сварить модели двух деталей, выполняется булева операция объединения тел.

Использование топологических объектов необходимо для корректного выполнения этих операций. Пусть требуется отрезать от одного из описанных выше простейших тел некоторую часть и пусть резка производится плоскостью. Тогда мы вынуждены найти линии пересечения поверхностей тела с этой плоскостью и по этим линиям обрезать поверхности тела и саму плоскость. Кроме того, нужно найти место стыковки обрезанной плоскости с частью исходного тела. Для плоскости нужно определить, какая ее сторона будет смотреть наружу тела, а какая — внутрь. Все это приводит к тому, что нужно знать топологию исходного тела и строить его по общим правилам.

При проектировании приходится рассматривать несколько вариантов деталей и сборочных единиц. Различные варианты одной и той же детали можно получить путем изменения требуемых параметров ее исходного варианта. Для этого в математической модели детали необходимо иметь информацию о пути и способах ее построения. Таким образом, геометрическая модель детали или сборочной единицы должна быть дополнена еще некоторой информацией о последовательности ее построения.

4.3. Тела, полученные движением плоского контура

Многие тела или заготовки для них можно получить путем движения плоского контура по заданной траектории. Пусть траектория движения описывается кривой $g(v)$ которую мы будем называть направляющей. Плоский контур будем называть образующей кривой. Пусть дана ограниченная плоскость

$$\mathbf{r}(x, y) = \mathbf{p} + x\mathbf{i}_x + y\mathbf{i}_y, \quad x, y \in \Omega, \quad (3.1)$$

где Ω область на плоскости параметров, ограниченная двумерным контуром $c(t)=[x(t) \ y(t)]^T$. Контур $c(t)$ состоит из набора стыкующихся друг с другом двумерных кривых $c_i(u)=[x_i(u) \ y_i(u)]^T$, $u_{\min} \leq u \leq u_{\max}$, $i=1, 2, \dots, n$.

Каждая кривая контура в пространстве описывается радиус-вектором

$$\mathbf{c}_i(u) = \mathbf{p} + x_i(u)\mathbf{i}_x + y_i(u)\mathbf{i}_y, \quad u_{i \min} \leq u \leq u_{i \max}. \quad (3.2)$$

Контур не должен иметь точек самопересечения. Если направляющая кривая $g(v)$ движения контура не замкнута, то оболочка тела имеет торцевые грани. Торцевые грани тел движения представляют собой плоскости, ограниченные заданным контуром.

Они описываются радиус-вектором. Если направляющая кривая движения контура замкнута, то оболочка тела не имеет торцевых граней. Боковые грани оболочки тела базируются на поверхностях движения. Число боковых граней равно числу кривых в контуре. Каждая боковая поверхность в качестве образующей содержит пространственный аналог соответствующей кривой контура. В зависимости от типа траектории движения $g(v)$ можно построить несколько типов тел.

Тело выдавливания.

Если направляющей движения контура служит отрезок прямой

$g(v) = \mathbf{p} + v\mathbf{hd}$, $0 \leq v \leq 1$, то мы получим тело выдавливания. Одна из его торцевых граней будет описываться ограниченной плоскостью (3.1). Другая его торцевая грань будет описываться аналогичной плоскостью, только с началом в точке $\mathbf{p}' = \mathbf{p} + \mathbf{hd}$. Боковая грань тела выдавливания, соответствующая кривой контура $\mathbf{c}_i(u)$ описываются поверхностями

$$\mathbf{r}_i(u, v) = \mathbf{p} + x_i(u)\mathbf{i}_x + y_i(u)\mathbf{i}_y + v\mathbf{hd} = \mathbf{c}_i(u) + v\mathbf{hd}, \quad u_{i \min} \leq u \leq u_{i \max}, \quad 0 \leq v \leq 1. \quad (3.3)$$

Тело выдавливания, построенное по замкнутому контуру, приведено на рис. 3.1. Для контура можно построить эквидистантный контур. По заданному контуру и эквидистантному к нему контуру может быть построено тело выдавливания с уклоном.

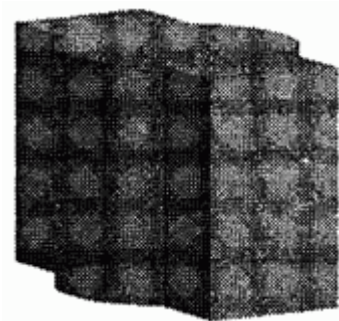


Рис. 3.1. Тело выдавливания



Рис. 3.2. Тело выдавливания с уклоном

Угол уклона определяется эквидистантой и длиной выдавливания. Тело выдавливания с уклоном приведено на рис. 3.2. Выдавливание выполнено в обе стороны от плоскости (3.1). Боковые грани тела с уклоном построены на линейчатых поверхностях.

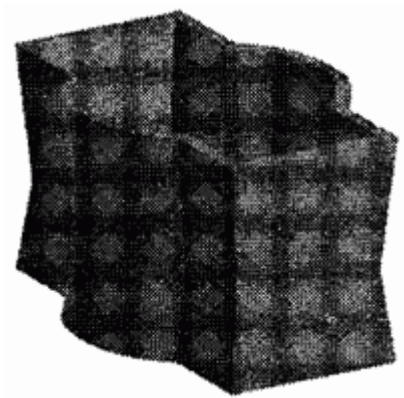


Рис. 3.3. Тонкостенное тело выдавливания

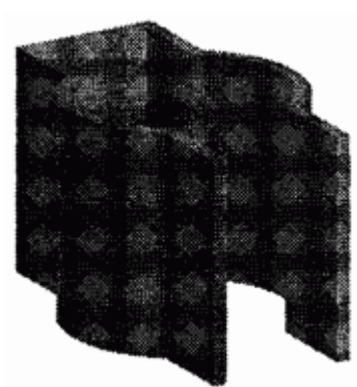


Рис. 3.4. Тонкостенное тело выдавливания

Тонкостенное тело выдавливания может быть построено по замкнутой или незамкнутой составной плоской кривой. Для его построения также используются эквидистантные составные кривые. Тонкостенные тела выдавливания приведены на рис. 3.3 и 3.4.

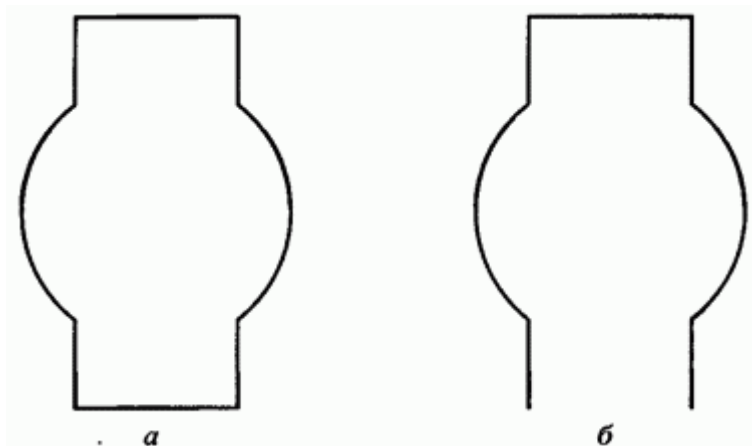


Рис. 3.5. Замкнутая (а) и незамкнутая (б) составная кривая

Замкнутая составная кривая и незамкнутая составная кривая, по которым построены приведенные выше тела, показаны на рис. 6.3.5.

Тело вращения.

Если направляющей кривой движения контура служит окружность или ее дуга, то мы получим тело вращения. Пусть ось вращения проходит через точку \mathbf{p}_0 , а ее направление характеризуется единичным вектором \mathbf{i} . Тогда боковая грань тела вращения, соответствующая кривой $\mathbf{c}_i(u)$, будет описываться поверхностью

$$\mathbf{r}_i(u, v) = \mathbf{p}_0 + \mathbf{r}_{3i}(u) + \cos v \mathbf{r}_{1i}(u) + \sin v \mathbf{r}_{2i}(u),$$

$$u_{i \min} \leq u \leq u_{i \max}, \quad 0 \leq v \leq \alpha, \quad (3.4)$$

где $\mathbf{r}_{3i}(u) = ((\mathbf{c}_i(u) - \mathbf{p}_0) \cdot \mathbf{i})\mathbf{i}$ — составляющая вектора $\mathbf{c}_i(u) - \mathbf{p}_0$, параллельная оси вращения, $\mathbf{r}_{1i}(u) = \mathbf{c}_i(u) - \mathbf{p}_0 - \mathbf{r}_{3i}(u)$ — составляющая вектора $\mathbf{c}_i(u) - \mathbf{p}_0$, перпендикулярная оси вращения, $\mathbf{r}_{2i}(u) = \mathbf{i} \times (\mathbf{c}_i(u) - \mathbf{p}_0)$ — ортогональный первым двум векторам вектор, длина которого равна длине вектора $\mathbf{r}_{1i}(u)$. Кривая $\mathbf{c}_i(u)$ описывается радиус-вектором (3.2).

Ось вращения не должна пересекать боковые грани. Определим положительное направление оси вращения. Пусть мы находимся с той стороны плоскости контура, откуда движение вдоль контура в положительном направлении выглядит против часовой стрелки. За положительное направление оси вращения примем направление, при взгляде вдоль которого контур находится слева. Если угол вращения α равен 2π , то оболочка тела имеет топологию тора, в противном случае — топологию призмы. В первом случае оболочка не имеет торцевых граней. Тело вращения с топологией призмы, построенное по замкнутому контуру, приведено на рис. 3.6. Тело вращения с топологией тора, построенное по замкнутому контуру, приведено на рис. 3.7.

По замкнутой или незамкнутой составной плоской кривой может быть построено тонкостенное тело вращения. Для его построения используются эквидистантные контуры или составные кривые. Тонкостенные тела вращения приведены на рис. 3.8 и 3.9.

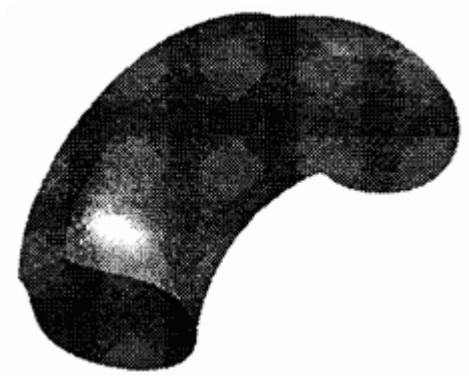


Рис. 3.6. Тело вращения

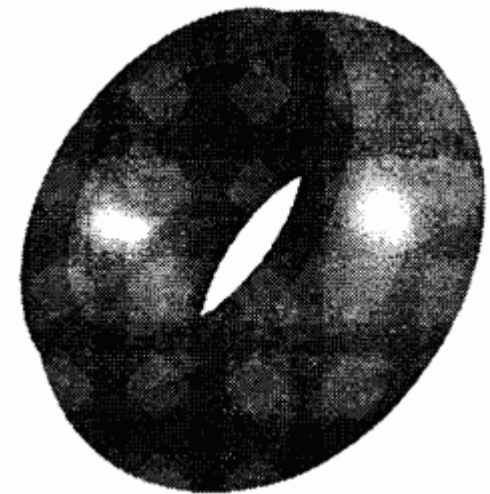


Рис. 3.7. Тело вращения с топологией тора

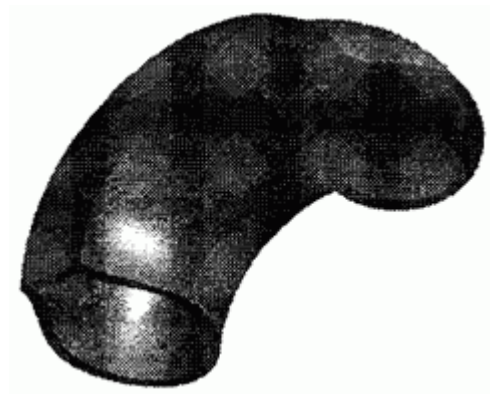


Рис. 3.8. Тонкостенное тело вращения

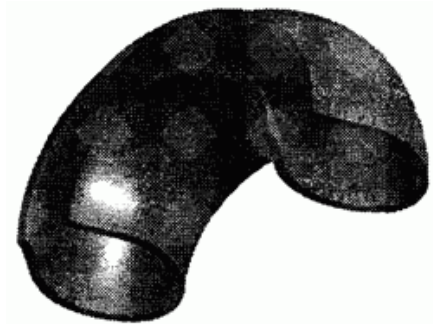


Рис. 3.9. Тонкостенное тело вращения

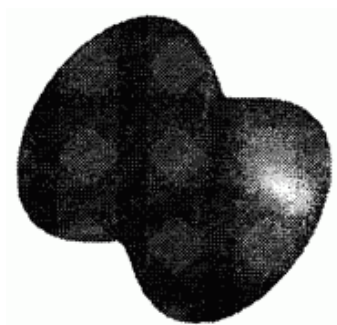


Рис. 3.10. Сфероид

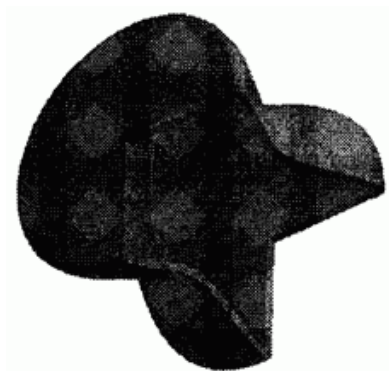


Рис. 3.11. Тонкостенный сфероид

По незамкнутой составной плоской кривой могут быть построены тела вращения, приведенные на рис. 3.10 и 3.11. Из концов образующей на ось вращения опущены перпендикуляры, по которым построены соответствующие грани. Для построения тонкостенного тела используется эквидистантная составная кривая. В отличие от предыдущих примеров тела, приведенные на рис. 3.10 и 3.11, касаются оси вращения. Такие тела будем называть сфероидами.

Кинематическое тело.

При всех других случаях формы направляющей кривой мы получим кинематическое тело. При движении плоского контура вдоль направляющей кривой ориентация контура относительно направляющей может меняться или может оставаться неизменной.

Если ориентация образующей в пространстве не меняется, то контур выполняет плоскопараллельное движение, оставаясь параллельным своему начальному положению, и мы получим тело сдвига. Боковые грани тела сдвига строятся для каждой кривой (3.2) и описываются поверхностью

$$\begin{aligned} \mathbf{r}_i(\mathbf{u}, v) &= \mathbf{g}(v) + (\mathbf{c}_i(\mathbf{u}) - \mathbf{g}(v_{\min}) - \mathbf{h}), \\ u_{i \min} &\leq u \leq u_{i \max}, \quad v_{\min} \leq v \leq v_{\max}, \end{aligned} \quad (3.5)$$

где \mathbf{h} — вектор привязки образующей к направляющей. Вектор \mathbf{h} смещает кривую $\mathbf{c}_i(\mathbf{u})$ на некоторую величину из ее начального положения относительно направляющей и сохраняет это смещение во время движения. Если $\mathbf{h}=\mathbf{0}$, то сохраняется исходное положение образующей кривой относительно начальной точки направляющей кривой. Тело сдвига всегда имеет торцевые грани. Если ориентация контура в пространстве при движении меняется, сохраняя некоторый заданный угол между плоскости контура и касательной к направляющей кривой, то мы получим тело заметания. Боковые грани тела заметания строятся для каждой кривой (3.2) и описываются поверхностью

$$\begin{aligned} \mathbf{r}_i(\mathbf{u}, v) &= \mathbf{g}(v) + \mathbf{M}(v) \cdot (\mathbf{c}_i(\mathbf{u}) - \mathbf{g}(v_{\min}) - \mathbf{h}), \\ u_{i \min} &\leq u \leq u_{i \max}, \quad v_{\min} \leq v \leq v_{\max}, \end{aligned} \quad (3.6)$$

где $\mathbf{M}(v)$ — матричная функция.

Если направляющая кривая замкнута, то получим кинематическое тело с топологией тора. Если направляющая кривая не замкнута, то получим кинематическое тело с топологией призмы.



Рис. 3.12. Тело сдвига



Рис. 3.13. Тело заметания

Боковые грани кинематического тела (тела сдвига или тела заметания) не должны пересекать сами себя. Это требование приводит к тому, что направляющая кривая тела сдвига не может быть замкнутой. Тело сдвига приведено на рис. 3.12. Тело заметания приведено на рис. 3.13.

На рис. 3.14 приведено тонкостенное кинематическое тело, построенное по незамкнутой образующей и замкнутой направляющей.



Рис. 3.14. Тонкостенное тело заметания

Для построения граней рассмотренных тел нужно знать двухмерный контур $c(t)$, местную его плоскость $\mathbf{p}, \mathbf{i}_x, \mathbf{i}_y$, направляющую кривую $g(v)$ и другие необходимые параметры (толщину стенки для тонкостенного тела, способ движения контура — плоскопараллельный или ортогональный).

4.4. Построение тела по плоским сечениям

Многие детали можно построить по их плоским сечениям в определенных местах. Пусть имеется несколько плоских контуров $g(v)$,

одинаково ориентированных, расположенных на некотором расстоянии друг от друга и состоящих из одинакового числа кривых. Построим тело, сечения которого ограничены данными контурами. Тело имеет две торцевые грани, построенные на крайних контурах. Торцевые грани тела представляют собой плоскости, ограниченные одним из крайних контуров. Боковыми гранями оболочки тела могут служить сглаживающие поверхности.



Рис. 4.1. Тело, построенное по плоским сечениям

Их количество равно количеству кривых в контуре. В каждой боковой грани участвует по одной кривой каждого контура. Эти кривые будем называть соответствующими. Между кривыми в контурах должно быть установлено соответствие, например, по порядковому номеру кривой в контуре. Тело, построенное по плоским сечениям, приведено на рис.4.1.

В общем случае все контуры, за исключением торцевых контуров, могут быть не плоскими. Тело, построенное по плоским сечениям, может быть замкнутым и иметь топологию тора.

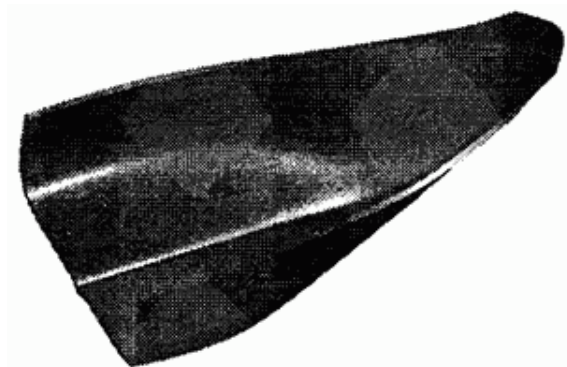


Рис. 4.2. Тело, построенное по плоским сечениям

Тело, построенное по сечениям, описываемым незамкнутыми линиями, приведено на рис. 4.2. По незамкнутым линиям в любом случае строятся замкнутые контуры, в которых участвуют эквидистантные кривые.

4.5. Тело в форме листа

На основе поверхности произвольной формы можно построить тело в форме листа конечной толщины. Пусть дана поверхность $\mathbf{b}(u, v)$. Выбрав толщину листа h , построим эквидистантную к ней поверхность

$$\mathbf{r}(u, v) = \mathbf{b}(u, v) + hm(u, v), \quad (5.1)$$

где $\mathbf{m}(u, v)$ — нормаль к поверхности $\mathbf{b}(u, v)$. На этих поверхностях построим две основные грани листового тела. Остальные (боковые) грани построим на линейчатых поверхностях, одной базовой линией которых является граничная линия на поверхности $\mathbf{b}(u, v)$, а второй — соответствующая ей линия на поверхности $\mathbf{r}(u, v)$.

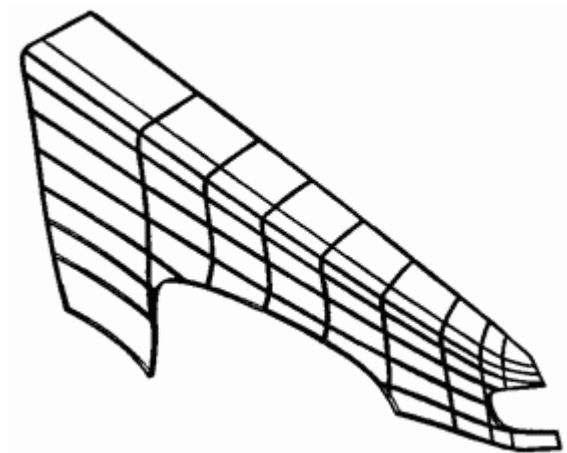


Рис. 5.1. Тело в форме листа

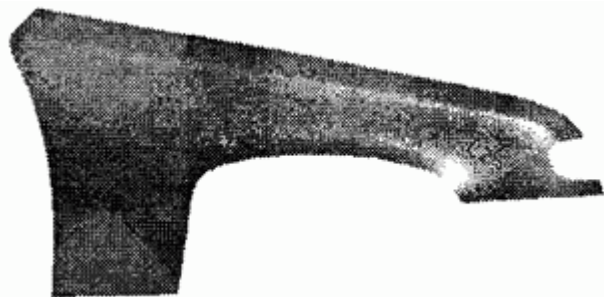


Рис. 5.2. Тело в форме листа

Если область определения параметров поверхности $\mathbf{b}(u, v)$ имеет прямоугольную форму, то листовое тело будет иметь четыре боковые грани. В общем случае листовое тело будет иметь столько боковых граней, сколько линий содержат двухмерные контуры, описывающие область определения параметров поверхности $\mathbf{b}(u, v)$.

Направление нормалей основных граней противоположно друг другу в соответствующих точках. Если поверхность $\mathbf{b}(u, v)$ является замкнутой по одному из параметров, то листовое тело будет иметь форму трубы и топологию тора. Если поверхность замкнута по обоим параметрическим направлениям, то мы получим тело с пустотой внутри. Такое тело будет иметь внешнюю оболочку на основе поверхности $\mathbf{r}(u, v)$ и внутреннюю оболочку на основе поверхности $\mathbf{b}(u, v)$,

Векторное изображение тела в форме листа, построенного по ограниченной контурами NURBS поверхности, приведено на рис. 6.5.1, а его тоновое изображение — на рис. 6.5.2.

С помощью тел в форме листа можно моделировать детали кузова автомобиля и планера самолета. В структуре данных тела достаточно иметь базовую поверхность $\mathbf{b}(u, v)$ и толщину тела h . Тело, построенное по сечениям, и тело в форме листа, так же как все рассмотренные выше тела, служат заготовками для моделей деталей. Дальнейшее моделирование деталей связано операциями над телами.

4.6. Булевы операции над телами

Над телами, как и над другими геометрическими объектами, можно выполнять операции — совокупность действий над одним или несколькими исходными телами, которая приводит к рождению нового тела. Одними из основных операций для двух тел являются булевы операции.

Булевыми операциями называют операции объединения, пересечения и вычитания тел, так как они выполняют одноименные операции над внутренними объемами тел (над множествами точек пространства, находящимися внутри тел). Булеву операцию объединения тел будем обозначать формулой $S=S_1 \cup S_2$, где S_1 , S_2 — исходные тела, S — результирующее тело. Булеву операцию пересечения тел будем обозначать формулой $S=S_1 \cap S_2$. Булеву операцию вычитания тел будем обозначать формулой $S=S_1 - S_2$. В порядке следования тел-операндов будем называть их первым телом и вторым телом.

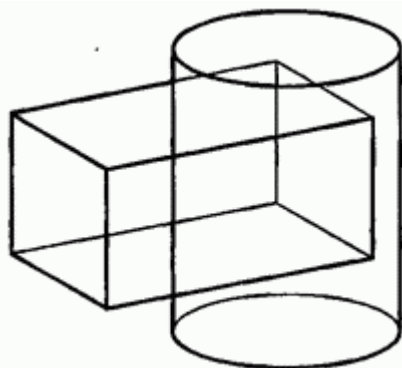


Рис. 6.1. Два исходных тела

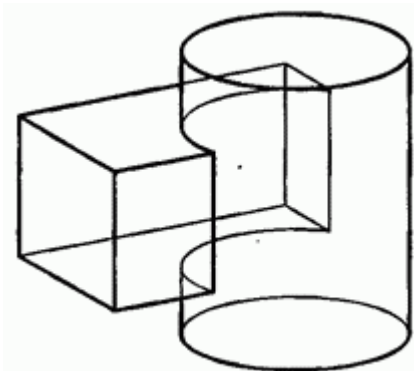


Рис. 6.2. Объединение тел

Результатом операции объединения двух тел является тело, которое содержит точки, принадлежащие внутреннему объему или первого, или второго тела. Результатом операции пересечения двух тел является тело, которое содержит точки, принадлежащие внутреннему объему как первого, так и второго тела.

Результатом операции вычитания двух тел является тело, которое содержит точки, принадлежащие внутреннему объему первого, но не принадлежащие внутреннему объему второго тела.

На рис. 6.1 приведены исходные для булевой операции тела. На рис. 6.2 приведен результат операции объединения тел, на рис. 6.3 приведен результат операции пересечения, на рис. 6.4 и 6.5 приведены результаты операции вычитания тел.

Операцию вычитания тел можно свести к операции пересечения тел; для этого нужно вывернуть второе тело наизнанку и найти точки его объема, одновременно принадлежащие и объему первого тела.

Вывернутое наизнанку тело S будем обозначать S^- .

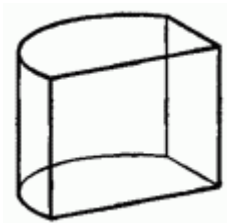


Рис. 6.3. Пересечение тел

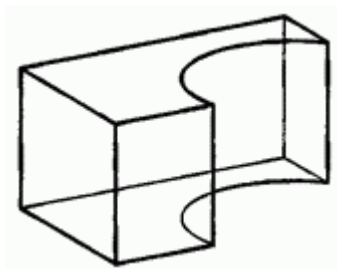


Рис. 6.4. Разность тел

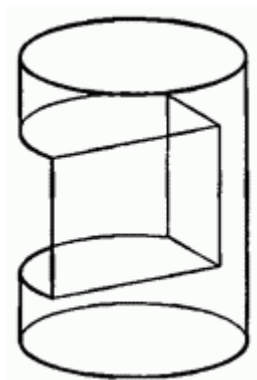


Рис. 6.5. Разность тел

При выворачивании тела наизнанку внутренние стороны граней становятся наружными сторонами, а наружные — внутренними и изменяются направления циклов на противоположные, в результате

чего внутренним объемом тела становится та часть пространства, которая до этого находилась снаружи тела. Математически операция вычитания сводится к операции пересечения тел $S = S_1 - S_2 = S_1 \cap S_2^c$. Конструктор при проектировании использует операции объединения и вычитания (они могут называться по-другому, например, операции сварки и сверления), а математический аппарат выполняет соответственно операции объединения и пересечения. Все булевы операции содержат много общего и выполняются по единому алгоритму.

Объединение тел.

Рассмотрим булеву операцию объединения тел. Кратко суть операции можно описать следующим образом: нужно найти линии пересечения граней тел, удалить ту часть первого тела, которая попала внутрь второго тела и ту часть второго тела, которая попала внутрь первого тела, а из всего остального построить новое тело. Операцию условно разобьем на три этапа. На первом этапе построим линии пересечения поверхностей граней и на их базе — новые ребра. Построенные новые ребра будем называть ребрами пересечения, а ребра тел будем называть старыми ребрами. На втором этапе определим точки пересечения новых ребер со старыми ребрами и в этих точках разрежем старые ребра на несколько ребер. На третьем этапе операции перестроим циклы пересекшихся граней. После этого добавим к пересекшимся граням тел грани, топологически связанные ними. Рассмотрим этапы построения тела более подробно.

Первый этап операции объединения тел начнем с того, что построим линии пересечения каждой грани первого тела с каждой гранью второго тела, если таковые имеются. Для этого используем алгоритм пересечения поверхностей граней.

Пусть грани первого тела описываются поверхностями

$$\mathbf{r}_i(u_i, v_i), \quad i = 1, 2, \dots, m, \quad (6.1)$$

а грани второго тела описываются поверхностями

$$\mathbf{s}_j(a_j, b_j), \quad j = 1, 2, \dots, n. \quad (6.2)$$

На базе линий пересечения граней первого и второго тел

$$\begin{aligned} \mathbf{l}_{uv}(t) &= [u_i(t) \ v_i(t)]^T, & \mathbf{l}_{uv}(t) &\in \mathbf{r}_i(u_i, v_i), \\ \mathbf{l}_{ab}(t) &= [a_j(t) \ b_j(t)]^T, & \mathbf{l}_{ab}(t) &\in \mathbf{s}_j(a_j, b_j), \\ t_{\min} &\leq t \leq t_{\max}, \end{aligned} \quad (6.3)$$

построим ребра пересечения. При этом ребрам пересечения дадим направление векторного произведения нормали грани первого тела с нормалью грани второго тела: $t_{\text{edge}} = \mathbf{m}_i \times \mathbf{m}_j$. Направление ребра определяется признаком совпадения направления производной линии пересечения с требуемым направлением ребра. За положительное направление нормали грани примем направление наружу тела. Нормаль грани может совпадать с нормалью ее поверхности или иметь противоположное направление в зависимости от признака их совпадения. На рис. 6.6 показаны направления ребер пересечения грани первого тела с двумя гранями второго тела.

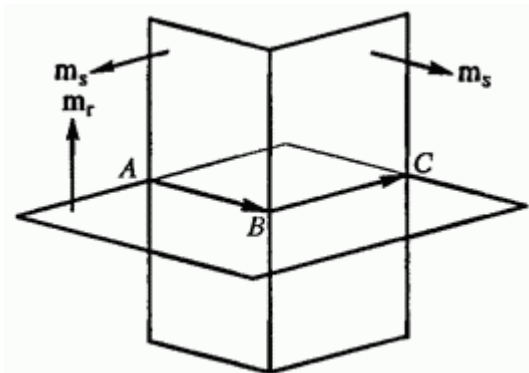


Рис. 6.6. Направление ребер пересечения граней тел

Ребра пересечения должны быть построены так, чтобы они полностью лежали внутри циклов граней исходных тел. Ребра пересечения могут подходить к границам грани только своими концами. В точках А, В и С

(рис. 6.6) старые ребра граней должны быть разбиты каждое на два ребра, так как в результирующее тело войдет только часть исходной грани.

На втором этапе разрежем старые ребра тела, к которым подходят ребра пересечения. Резка старого ребра осуществляется путем рассечения кривой, на которой базируется ребро. Из одной кривой получим две кривые, в совокупности заменяющие исходную кривую ребра. Одна из этих кривых останется геометрическим носителем разрезаемого ребра, а на базе второй построим новое ребро, которое получит от исходного ребра всю необходимую информацию. Как было сказано, каждое ребро строится на базе кривой пересечения поверхностей (1.1). Кривую пересечения поверхностей составляют две поверхности и две соответствующие им двумерные кривые. Рассечению подлежат именно двумерные кривые на двух разных поверхностях. Как до, так и после рассечения эти кривые должны иметь одинаковые области определения параметров и соответствие точек при всех значениях параметра.

Точки пересечения нового ребра со старым ребром грани ищутся как точки пересечения двумерных кривых, заданных на общей для них плоскости параметров. От каждого ребра в формуле точек пересечения линий участвует по одной двумерной кривой, входящей в линию пересечения.

Параметры точек пересечения ребер и сами координаты кривых, являющиеся параметрами поверхности, должны быть определены с заданной точностью.

Если разрезаемое ребро базируется на кривой пересечения, заданной отдельными точками (двумерные линии являются ломаными и точно совпадают в пространстве только в характеристических точках), то прежде чем разрезать такую кривую, нужно в обе линии вставить дополнительные точки, соответствующие точки пересечения трех поверхностей — двух поверхностей, лежащих по обе стороны разрезаемого ребра, и поверхности грани другого исходного тела. Например, если режется линия пересечения поверхностей $r_i(u_i, v_i)$, $r_n(u_n, v_n)$ поверхностью $s_j(a_j, b_j)$, то в линии $l_{iv}(q)=[u_i(q), v_i(q)]^T$ и $l_{nv}(q)=[u_n(q), v_n(q)]^T$ нужно вставить дополнительные двумерные характеристические точки, соответствующие пересечению поверхностей $r_i(u_i, v_i)$, $r_n(u_n, v_n)$, $s_j(a_j, b_j)$. Задача пересечения трех

поверхностей сводится к решению системы шести скалярных уравнений относительно шести параметров $u_i, v_i, u_n, v_n, a_j, b_j$. Начальное приближение решения известно достаточно точно.

Так как каждое ребро исходных тел входит в циклы двух смежных граней, то после резки ребер исходных тел необходимо произвести корректировку этих циклов с учетом разрезанных ребер.

После первых двух этапов мы получили совокупность ребер пересечения, ориентированных определенным образом, стыкующихся друг с другом и с ребрами исходных тел только в вершинах. Далее необходимо перестроить циклы пересекшихся граней.

Третий этап завершает булеву операцию. Для того чтобы каждую из показанных на рис. 6.6 граней разрезать на части, нужно перестроить ее циклы и в соответствии с циклами изменить контуры, описывающие область определения параметров поверхности грани. На рис. 6.7 показаны две пересекшиеся грани (тонкими линиями со стрелками показано направление циклов граней исходных тел) и ребро пересечения. На рис. 6.8 показаны те части граней, которые войдут в объединение тел.

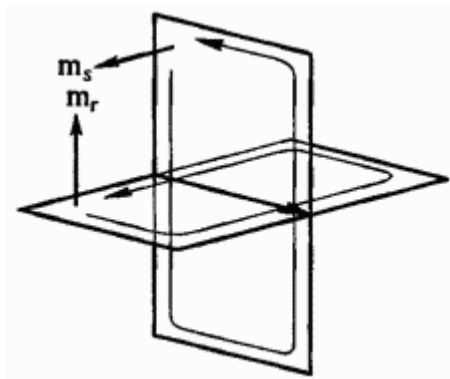


Рис. 6.7. Исходные грани тела

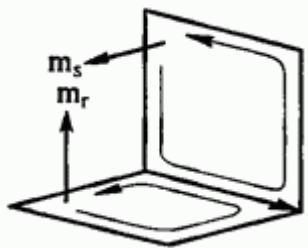


Рис. 6.8. Обрезанные грани

Стрелками показано направление перестроенных циклов граней тел. Каждый цикл представляет собой список ребер в порядке их следования и список флагов ориентации этих ребер в цикле.

Из рисунка видно, что при принятой ориентации ребер пересечения ($t_{\text{edge}} = \mathbf{m}_r \times \mathbf{m}_s$) в циклы граней первого тела они войдут с отрицательным флагом, а в циклы граней второго тела они войдут с положительным флагом.

Старые ребра исходных тел, которые сохранятся в результирующем теле, войдут в перестроенные циклы, сохранив свои флаги. При перестройке циклов пересекшихся граней будем использовать следующий алгоритм.

Рассмотрим одну из двух пересекшихся граней, принадлежащих первому телу. Берем любое ребро пересечения рассматриваемой грани и начинаем с него составлять список ребер цикла. Ребро пересечения должно войти в цикл грани первого тела с отрицательным флагом, следовательно, цикл будет иметь направление, противоположное первому ребру. Для продолжения цикла среди ребер пересечения и среди старых ребер грани ищем все ребра, стыкующиеся с данным ребром в его начальной вершине. Среди найденных ребер выберем то, которое лежит слева от остальных (заворачивает влево на больший угол по сравнению с другими найденными ребрами, если смотреть вдоль цикла с наружной стороны грани). Выбранное ребро ставим в список цикла. Если выбранное ребро является старым, то оно сохраняет свой флаг в цикле. Если выбранное ребро есть ребро пересечения, то оно получит отрицательный флаг. Ребра пересечения обладают преимущественным правом по отношению к старым ребрам быть выбранными. То есть, если левее других оказались два

совпадающих ребра, одно из которых является старым, а другое — ребром пересечения, то для продолжения цикла должно быть выбрано ребро пересечения.

Процесс перестроения цикла грани будем продолжать до тех пор, пока цикл не замкнется. На этом построение очередного цикла заканчивается. Если у рассматриваемой грани при построении цикла использованы не все ребра пересечения, то с любого из оставшихся ребер пересечения начинаем строить еще один цикл грани. Циклы перестраиваем до тех пор, пока не используем все ребра пересечения. Таким способом мы построим в общем случае несколько новых циклов рассматриваемой грани первого тела.

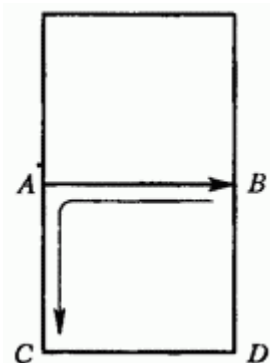


Рис. 6.9

Поиск стыкующихся ребер и определение угла поворота в точке стыка удобно выполнять по двумерным кривым ребер грани. При этом можно работать даже с такими ребрами, которые в пространстве стянуты в точку (например, ребро в вершине конуса или полюсе сферы). На рис. 6.9 показано, что с новым ребром В А в точке А стыкуются несколько ребер. В данном случае для продолжения цикла, начатого с ребра пересечения В А, следует выбрать ребро АС.

Вновь построенные циклы рассортируем по группам, каждая из которых состоит из внешнего цикла и входящих в него внутренних циклов. У рассматриваемой грани могут остаться нетронутыми один или несколько старых циклов. Нетронутыми мы будем называть старые циклы грани, ни одно ребро которых не вошло в перестроенные

циклы. Среди старых нетронутых циклов отберем те, которые необходимо включить в состав описания перестроенной грани. Ими являются старые внутренние циклы грани, лежащие внутри новых внешних циклов. Еще нужно определить, не потребуется ли включить в результат старый внешний цикл грани, если он остался нетронутым. Это необходимо сделать, если для некоторых новых внутренних циклов не найден новый внешний цикл и они лежат внутри старого внешнего цикла. Сортировку циклов удобно выполнять с помощью двумерных контуров, соответствующих каждому циклу.

Если в результате сортировки внешних циклов получилось больше одного, то это означает, что из исходной грани в результате операции образовалось несколько граней. На рис. 6.10 а - 6.13 а приведены варианты исходных граней первого тела.

На рис. 6.10 б-6.13 б. приведены грани с добавлением ребер пересечения (ребра пересечения выделены). На рис. 6.10 в-6.13 в приведены грани, которые получились в результате операции.

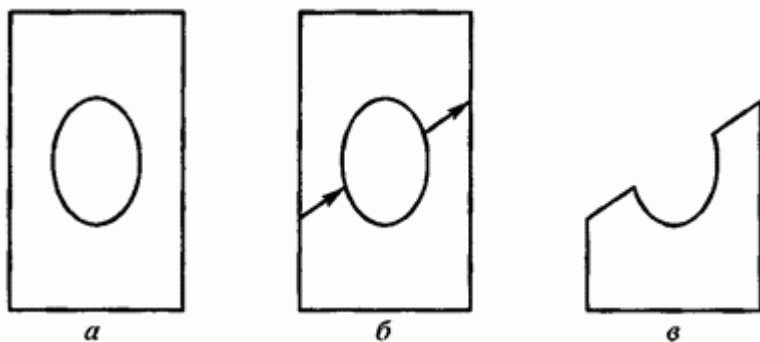


Рис. 6.10. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

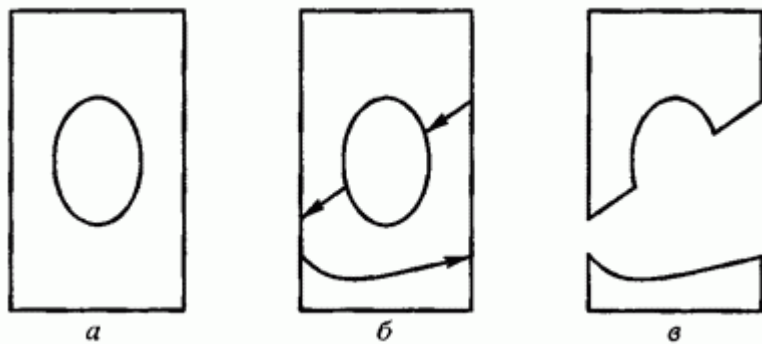


Рис. 6.11. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

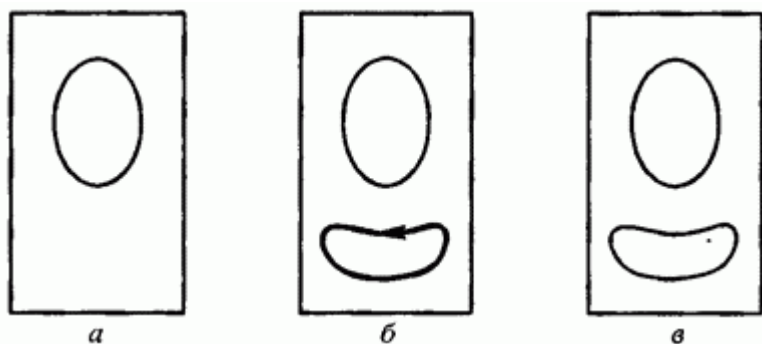


Рис. 6.12. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

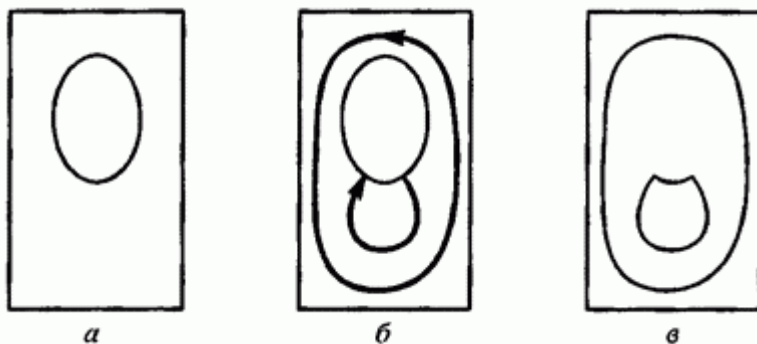


Рис. 6.13. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

На рис. 6.10 грань с двумя циклами была разрезана и получилась одна грань с одним циклом. На рис. 6.11 грань с двумя циклами дала две грани. В примере, приведенном на рис. 6.12, потребовалось использовать старый внешний и внутренний циклы. На рис. 6.13 из одной грани получено две грани, причем для одной из них потребовалось использовать старый внешний цикл.

Описанное перестроение циклов выполняется для каждой пересекшейся грани первого тела.

С пересекшимися гранями второго тела поступим аналогично тому, как мы поступили с гранями первого тела, но с одной небольшой разницей. Ребра пересечения должны войти в перестроенные циклы граней второго тела с положительным флагом (в перестроенные циклы граней первого тела они вошли с отрицательным флагом). В этом состоит отличие перестроения циклов граней второго тела. Все остальные действия над гранями первого и второго объединяемых тел одинаковы.

Мы перестроили пересекшиеся грани объединяемых тел. Все их включим в оболочку результирующего тела. Для получения результирующего тела остается к этим граням в оболочку тела добавить не пересекшиеся в операции грани, которые топологически связаны с пересекшимися гранями. Для этого будем брать последовательно ребра, входящие в оболочку нового тела, и включать в оболочку смежные грани ребер (если они там отсутствуют).

Продолжив эти действия для ребер всех добавленных граней, получим оболочку результирующего тела.

Пересечение тел.

Коротко суть булевой операции пересечения тел можно описать следующим образом: нужно найти линии пересечения тел, удалить ту часть первого тела, которая не попала внутрь второго, и ту часть второго тела, которая не попала внутрь первого, а из всего остального построить новое тело.

Эта операция имеет много общего с операцией объединения тел. Вернемся к рис. 6.6. На нем приведены пересекающиеся грани: одна грань первого тела и две грани второго тела. Грани режут друг друга, так что в результирующую оболочку войдут только части этих граней.

В пересечение тел войдет часть грани первого тела, лежащая внутри второго тела, и часть грани второго тела, лежащая внутри первого тела (рис. 6.14) (в объединение тел вошла часть грани первого тела, лежащая вне второго тела, и часть грани второго тела, лежащая вне первого тела).

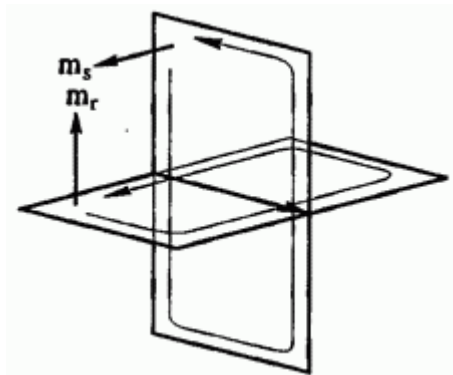


Рис. 6.14. Исходные грани

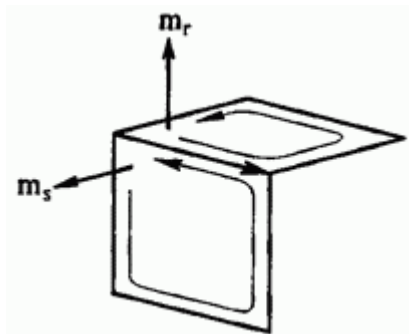


Рис. 6.15. Перестроенные грани пересечения тел

В этом и состоит основное отличие операций объединения и пересечения тел.

Операцию разобьем на три этапа. Первый и второй этапы операции пересечения тел полностью совпадают с соответствующими этапами операции объединения тел.

Третий этап операции пересечения тел выполняется аналогично третьему этапу операции объединения тел, но имеет одно отличие в том, с каким флагом входят в циклы ребра пересечения. Ребра пересечения входят в перестроенные циклы граней первого тела с положительным флагом, а в перестроенные циклы граней второго тела — с отрицательным флагом (в объединении тел флаги ребер пересечения в циклах имеют противоположное значение). Все остальные действия над гранями обоих тел в обеих операциях одинаковы.

Разность тел.

Коротко суть булевой операции вычитания тел можно описать следующим образом: нужно найти линии пересечения тел, удалить ту часть первого тела, которая попала внутрь второго, и ту часть второго тела, которая не попала внутрь первого, а из всего остального построить новое тело.

Булева операция вычитания тел сводится к булевой операции пересечения уменьшаемого тела и вывернутого наизнанку

вычитаемого тела. Вывернутое наизнанку тело мы получим из исходного тела путем переориентации направлений нормалей граней и направлений циклов граней. Переориентация направления нормали грани производится изменением признака совпадения нормали поверхности и нормали ее грани. Переориентация направления цикла грани производится перестроением списка ребер (изменением на обратный порядок следования ребер в списке) и заменой на противоположные флагов ребер в списках. Для вывернутого тела внутренним объемом является часть пространства, находящаяся вне его оболочки. Поэтому при пересечении уменьшаемого тела и вывернутого наизнанку вычитаемого тела результирующая оболочка будет содержать ту часть объема уменьшаемого тела, которая лежит вне вычитаемого тела.

Пересекающиеся ребра.

Наиболее трудоемким и требующим определенной точности в процессе выполнения булевых операций является построение ребер пересечения. Ребра пересечения не должны иметь выступающих за пределы грани частей, они должны обязательно стыковаться или друг с другом или со старыми ребрами граней. При корректном выполнении операции пересечения поверхностей эти условия обеспечиваются. Ребра пересечения не должны пересекать друг друга вне крайних точек.

В большинстве случаев новые ребра не пересекают друг друга, но в некоторых частных случаях это возможно. На рис. 6.16 показаны два цилиндрических тела одинакового диаметра, оси которых пересекаются — крест из цилиндров. При булевом объединении этих тел возможна ситуация, когда будут построены всего два замкнутых ребра пересечения. Такие ребра имеют две точки пересечения: А и В, по крайней мере, одна из которых не будет совпадать с вершинами ребер, а будет лежать где-то на ребре. Точки пересечения ребер лежат в точках касания цилиндров. Эти ребра должны быть разрезаны в точках А и В и у их частей должна быть уточнена ориентация, так как при прохождении точки касания поверхностей в данном случае векторное произведение нормалей к ним, по которому ориентируются новые ребра, меняет свое направление на противоположное. Обнаружить пересечение новых ребер можно по пересечению кривых на поверхностях, из которых состоит линия пересечения.

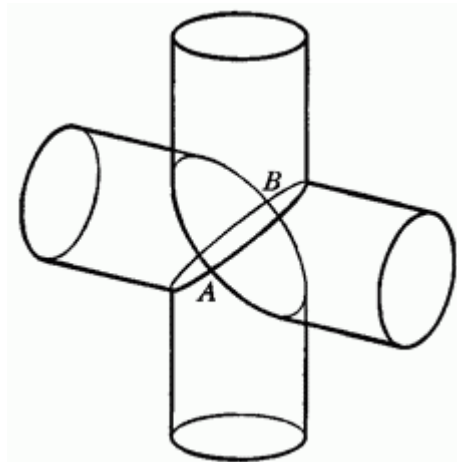


Рис. 6.16. В точках А и В нормали граней совпадают

Совпадающие ребра.

Описанный алгоритм выполнения булевых операций работает корректно, если ребра пересечения не совпадают с ребрами исходных тел. В противном случае он нуждается в уточнении. Рассмотрим примеры.

В булевых операциях возможен случай, когда какая-либо грань одного тела пересекает другое тело по его ребру. Возможен также случай, когда при построении ребер пересечения граней одного тела с гранями другого тела мы получим два новых ребра, совпадающих в пространстве друг с другом и с ребром одного из тел. Все ребра являются разными, так как в них стыкуются разные грани. На рис. 6.17 приведены два тела, при выполнении булевой операции над которыми, ребра пересечения совпадут с ребрами меньшего тела. Следуя общему алгоритму, мы в данном случае получим восемь ребер пересечения, половина из которых должна быть опущена (или не должна быть построена).

Правило для ребер пересечения.

При наличии ребер пересечения, совпадающих с ребрами граней исходных тел, и в некоторых других случаях будем выполнять

следующую проверку. Построим в плоскости каждой из двух пересекаемых граней по два вектора, ортогональные ребру пересечения. На рис. 6.18 это векторы \mathbf{a}_1 , \mathbf{b}_1 для грани с нормалью \mathbf{m}_1 и векторы \mathbf{a}_2 , \mathbf{b}_2 для грани с нормалью \mathbf{m}_2 . Векторы \mathbf{a}_1 , \mathbf{a}_2 поворачивают влево от ребра пересечения в соответствующих плоскостях, а векторы \mathbf{b}_1 , \mathbf{b}_2 поворачивают вправо от ребра пересечения в соответствующих плоскостях.

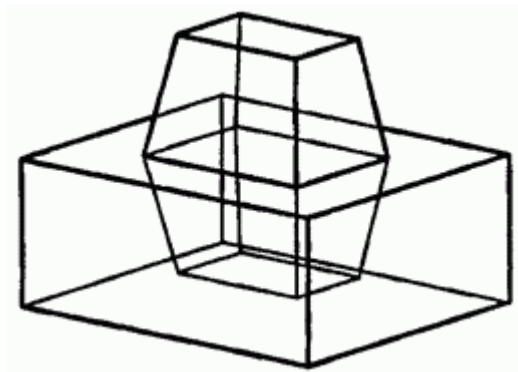


Рис. 6.17. Совпадение ребер пересечения с ребрами меньшего тела

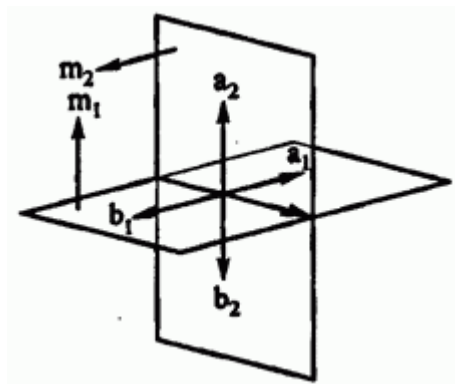


Рис. 6.18. Пересечение граней

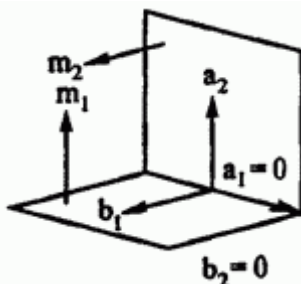


Рис. 6.19. Пересечение граней по ребру

Если грань не имеет продолжения за ребро пересечения (ребро пересечения частично или полностью совпадает с ребром тела), то соответствующий вектор положим равным нулю (рис. 6.19). Используя векторы $\mathbf{a}_1, \mathbf{a}_2, \mathbf{b}_1, \mathbf{b}_2$ и нормали $\mathbf{m}_1, \mathbf{m}_2$ мы можем определить, будет ли данное ребро пересечения использовано в операции или оно должно быть опущено. Обратим внимание на следующее обстоятельство.

В операции объединения тел грань первого тела мы сможем перестроить, если она имеет продолжение справа от ребра пересечения вне второго тела, а грань второго тела мы сможем перестроить, если она имеет продолжение слева от ребра пересечения вне первого тела.

Таким образом, в булевой операции объединения тел для ребра пересечения должны быть выполнены условия

$$\mathbf{m}_1 \cdot \mathbf{a}_2 > 0 \quad \text{и} \quad \mathbf{m}_2 \cdot \mathbf{b}_1 > 0. \quad (6.7)$$

В противном случае рассматриваемое ребро пересечения в булевой операции объединения тел строить не следует (если оно построено, то должно быть опущено).

В булевой операции пересечения тел грань первого тела мы сможем перестроить, если она имеет продолжение слева от ребра пересечения внутри второго тела, а грань второго тела мы сможем перестроить, если она имеет продолжение справа от ребра пересечения внутри первого тела. Таким образом, в булевой операции пересечения тел для ребра пересечения должны быть выполнены условия

$$\mathbf{m}_1 \cdot \mathbf{b}_2 < 0 \quad \text{и} \quad \mathbf{m}_2 \cdot \mathbf{a}_1 < 0. \quad (6.8)$$

В противном случае рассматриваемое ребро пересечения в булевой операции пересечения тел строить не следует (если оно построено, то должно быть опущено).

Как уже было сказано, ребра пересечения обладают преимущественным правом по отношению к старым ребрам тела быть включенными в перестраиваемый цикл.

Так как булева операция вычитания тел сводится к операции пересечения тел, то для нее должно выполняться правило (6.8) для вывернутой наизнанку оболочки тела.

Принадлежность точки пространству внутри тела.

Для ответа на вопрос, внутри или вне тела продолжается от ребра пересечения грань другого тела, нужно уметь определять, принадлежит ли некоторая точка \mathbf{p} пространству внутри тела — классифицировать точку относительно тела. Нам известно, что нормаль каждой грани направлена вне объема тела. Для точки \mathbf{p} найдем ближайшую точку \mathbf{p}_0 на ближайшей грани. Построим вектор \mathbf{n} из найденной точки \mathbf{p}_0 в точку \mathbf{p} . Вычислим нормаль \mathbf{m} тела в точке \mathbf{p}_0 . Если точка \mathbf{p}_0 лежит на границе граней (является точкой ребра или вершины), то в качестве нормали \mathbf{m} тела возьмем среднюю нормаль граничащих граней. Если $\mathbf{m} \cdot \mathbf{n} \geq 0$, то точка \mathbf{p} принадлежит внутреннему пространству тела. В противном случае — нет.

Перекрывающиеся грани.

В булевых операциях довольно часто можно встретить случаи частичного совпадения некоторых граней двух исходных тел. На рис. 6.20 приведены два тела, некоторые грани которых частично перекрывают друг друга.

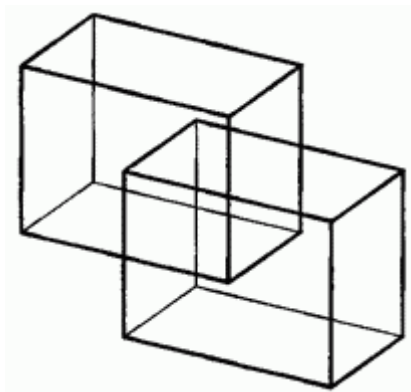


Рис. 6.20. Совпадающие грани тел

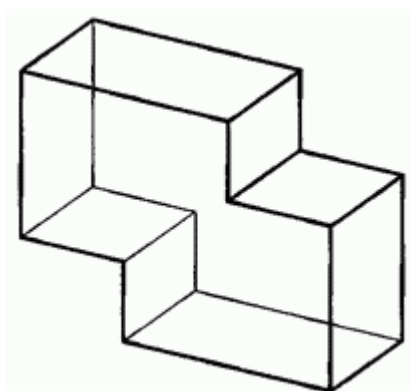


Рис. 6.21. Объединение тел

В данном случае не все ребра пересечения граней войдут в булев результат; некоторые ребра должны быть опущены (или не должны строиться). Результат объединения тел приведен на рис. 6.21.

Аналогичная ситуация возникает при выполнении булевой операции над двумя соосными цилиндрами одинакового радиуса и еще во многих случаях. Грани могут перекрываться полностью, частично или всего лишь по одной линии.. Нормали \mathbf{m}_1 , \mathbf{m}_2 частично перекрывающихся граней должны быть одинаково направлены в общих точках. Перекрывающимися являются грани, которые можно

перенести на одну общую поверхность. Таких граней может быть больше двух. При наличии у тел-операндов перекрывающихся граней все ребра пересечения этих граней должны быть проверены на предмет присутствия их в булевом результате. В операции объединения тел нам потребуются только те ребра пересечения перекрывающихся граней, смежные грани которых имеют продолжение вне одного из тел. В операции пересечения тел нам потребуются только те ребра пересечения перекрывающихся граней, смежные грани которых имеют продолжение внутрь одного из тел.

Циклы перекрывающихся граней тел должны быть перестроены по несколько иному алгоритму, чем для остальных граней. Прежде всего, все ребра этих граней должны быть приведены к одному общему носителю (к поверхности одной из перекрывающихся граней). В операции объединения тел старые ребра перекрывающихся граней первого тела, лежащие внутри второго тела, и старые ребра перекрывающихся граней второго тела, лежащие внутри первого тела, не войдут в результирующее тело. В операции пресечения тел старые ребра перекрывающихся граней первого тела, лежащие вне второго тела, и старые ребра перекрывающихся граней второго тела, лежащие вне первого тела, также не войдут в результирующее тело.

Построение некоторого цикла перекрывающихся граней начнем с ребра, которое точно должно войти в результат. Для продолжения цикла среди ребер пересечения и среди старых оставшихся ребер грани найдем все ребра, стыкующиеся с данным ребром в его соответствующей вершине. Среди найденных ребер выберем то, которое лежит справа от остальных (заворачивает вправо на больший угол по сравнению с другими найденными ребрами, если смотреть вдоль цикла с наружной стороны грани). Выбранное ребро ставим в список цикла. Отличие алгоритма состоит в том, что при продолжении цикла мы выбираем самое правое (а не самое левое) ребро.

Тела с несколькими оболочками.

Если какое-либо из тел-операндов имеет пустоты и, соответственно, описывается несколькими оболочками, то нетронутые операцией внутренние (внешние) оболочки должны быть проверены на вхождение в оболочку булева результата. В общем случае результирующее тело также может иметь несколько оболочек.

Дерево построения тел.

Булевы операции над телами показывают необходимость привлечения топологических понятий для построения тел.

Рассмотренные в предыдущих параграфах тела будем называть простыми, в отличие от тел, полученных в результате операций, которые будем называть сложными. Структура данных простых тел содержит минимум информации, по которой могут быть построены все грани тела. В структуре данных тела, полученного в результате булевой операции, положим структуры данных исходных тел и тип булевой операции. Оболочки тела всегда могут быть построены по этой информации. Таким образом, структура данных тела, полученного в результате многократного выполнения булевых операций, будет содержать дерево построения. Результирующее тело находится в корне дерева, а его ветви начинаются в простых телах.

Пример дерева приведен на рис. 6.22. В узлах дерева находятся тела. Дерево имеет несколько ярусов. Операции между телами обозначены соответствующими знаками. Операции выполняются между телами одного яруса.

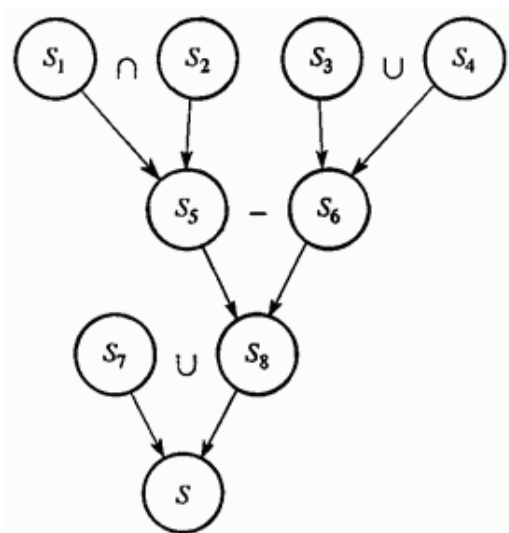


Рис. 6.22. Дерево построения тела

Структура данных тела в виде дерева построения может быть использована для всех тел — как для тел, полученных в результате некоторой операции, так и для простых тел.

4.7. Резка тела поверхностью

Если поверхность полностью пересекает тело, то тело можно разрезать этой поверхностью на две части. Результатом такой операции является новое тело, представляющее собой часть исходного тела, лежащую по одну или другую сторону режущей поверхности. Положительной стороной будем называть сторону поверхности, в которую направлены нормали к ней (если мы смотрим на положительную сторону поверхности, то нормаль направлена на нас). Другую сторону поверхности будем называть отрицательной.

Операцию резки тела поверхностью мы сведем к одной из булевых операций. На базе поверхности построим оболочку. Эта оболочка будет состоять из одной грани. Нормаль этой грани пусть совпадает с нормалью поверхности. Построенную незамкнутую оболочку мы будем рассматривать как некоторое незаконченное тело.

Пусть требуется построить ту часть тела, которая лежит с положительной стороны режущей поверхности. В этом случае выполним булеву операцию вычитания из этого тела незаконченного тела, построенного на основе режущей поверхности.

Пусть требуется построить ту часть тела, которая лежит с отрицательной стороны режущей поверхности. В этом случае выполним булеву операцию пересечения этого тела с незаконченным телом, построенным на основе режущей поверхности.

На рис. 7.1 показаны тело, режущая поверхность, ее нормаль и ребра пересечения граней тела с поверхностью. При резке тела поверхностью мы выполняем операцию пересечения тела с частью пространства, лежащей по ту или иную сторону поверхности.

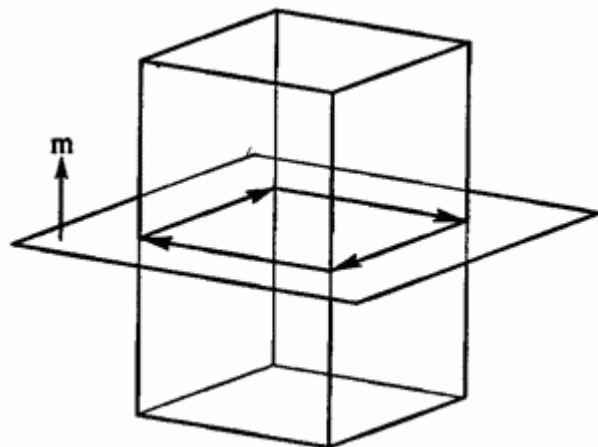


Рис. 7.1. Направление ребер пересечения

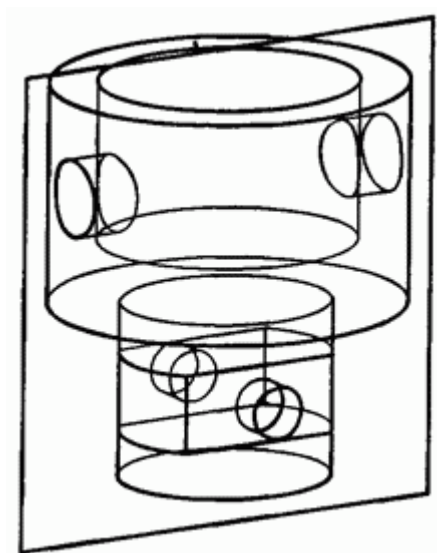


Рис. 7.2. Исходное тело и режущая поверхность

В частном случае тело можно разрезать плоскостью. На рис. 7.2 приведены исходное тело и режущая плоскость. На рис. 7.3 приведено разрезанное плоскостью тело.

С помощью рассматриваемой операции можно получить сложный разрез тела. Для этого оболочка незаконченного тела должна состоять из нескольких стыкующихся между собой граней.

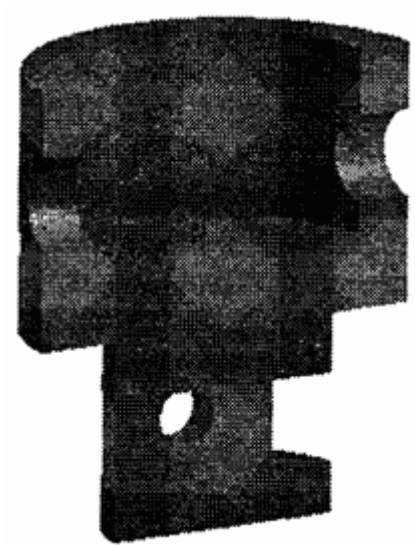


Рис. 7.3. Разрезанное тело

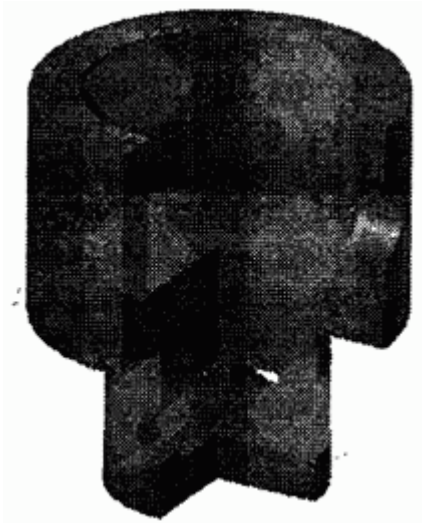


Рис. 7.4. Сложный разрез тела

Для получения сложного разреза возьмем составную кривую на плоскости и построим по ней незаконченное тело выдавливания так, чтобы оно пересекало заданное тело требуемым образом. Далее выполним булеву операцию над заданным телом и незаконченным телом. Результат сложного разреза тела показан на рис. 7.4.

7.8. Построение симметричного тела

Пусть имеется тело и плоскость. Построим тело, симметричное данному телу относительно данной плоскости. Исходное тело будем называть базовым. Симметричное тело будет зеркальной копией данного тела. Геометрия тела описывается точками, кривыми и поверхностями, которые в конечном итоге описываются точками векторами и скалярами. Поэтому построение симметричного тела в конечном итоге сводится к преобразованию симметрии радиус-векторов точек, свободных векторов и скаляров. Скалярные величины при преобразовании симметрии не изменяются, свободные векторы меняют свое направление, а точки — свое положение.

Пусть плоскость симметрии описывается формулой

$$\mathbf{p}(x, y) = \mathbf{p}_0 + x\mathbf{i}_1 + y\mathbf{i}_2,$$

где $\mathbf{i}_1, \mathbf{i}_2$ — ортогональные векторы единичной длины. Тогда матрица преобразования симметрии (3.8) тела относительно этой плоскости определяется формулой

$$\mathbf{A} = 2\mathbf{i}_1\mathbf{i}_1 + 2\mathbf{i}_2\mathbf{i}_2 - \mathbf{E}, \quad (8.1)$$

где $\mathbf{i}_1\mathbf{i}_1, \mathbf{i}_2\mathbf{i}_2$ — диадные произведения векторов. Свободный вектор \mathbf{v}_0 после преобразования симметрии относительно этой плоскости будет описываться радиус-вектором

$$\mathbf{v} = \mathbf{A} \cdot \mathbf{v}_0. \quad (8.2)$$

Произвольная точка \mathbf{r}_0 после преобразования симметрии относительно этой плоскости будет описываться радиус-вектором

$$\mathbf{r} = \mathbf{p}_0 + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{p}_0). \quad (8.3)$$

После преобразования симметрии все точки и линии на поверхностях останутся неизменными. А вот нормали поверхностей изменят свое направление на противоположное. Там, где нормаль грани и нормаль ее поверхности в исходном теле совпадали по направлению, в зеркальной копии будут иметь противоположное направление, и наоборот: там, где нормаль грани и нормаль ее поверхности в исходном теле не совпадали по направлению, в зеркальной копии будут иметь одинаковое направление. Кроме того, циклы граней зеркальной копии будут иметь направление, противоположное своим оригиналам. Поэтому в гранях зеркальной копии нужно произвести изменение признаков совпадения нормали поверхности и нормали ее грани на противоположные значения и переориентировать циклы. Переориентация направления цикла грани производится перестроением списка ребер (порядок следования ребер в списке обратный) и заменой флагов ребер в списках на противоположные флаги. Такая же переориентация производилась и при выворачивании тела наизнанку. Если выполнить только преобразование геометрических данных тела по матрице (8.1), то получим зеркальное

отражение тела, вывернутое наизнанку. Таким образом, построение симметричного тела сводится к преобразованию его копии по матрице (8.1) и выворачивании ее наизнанку (переориентации его граней).

Симметрия части тела.

Пусть плоскость симметрии пересекает тело. Представим, что оболочка тела разрезана плоскостью, одна из отрезанных частей удалена, по другой части выполнена зеркальная копия и склеена с ней. Мы получим тело, симметричное относительно плоскости, состоящее из двух половинок, одна из которых совпадает с частью исходного тела.

Плоскостью симметрии может служить плоская грань исходного тела. Построение такого симметричного тела имеет много общих моментов с резкой тела на части и операцией объединения тел.

Не теряя общности, будем строить симметричное тело по его части, находящейся с положительной стороны плоскости — с той стороны, в которую направлена нормаль плоскости. Операцию условно разобьем на три этапа.

На первом этапе построим линии пересечения поверхностей граней тела с плоскостью. Пусть пересеченные плоскостью грани тела описываются поверхностями $\mathbf{r}_i(u_i, v_i)$, $i=1, 2, \dots, n$. Тогда каждая линия пересечения будет состоять из пары двумерных кривых

$$\begin{aligned} l_{uv}(t) &= [\mathbf{u}_i(t) \ v_i(t)]^T, & l_{uv}(t) &\in \mathbf{r}_i(\mathbf{u}_i, \mathbf{v}_i), \\ l_{xy}(t) &= [x(t) \ y(t)]^T, & l_{xy}(t) &\in \mathbf{p}(x, y), \end{aligned} \quad (8.4)$$

имеющих общий параметр t . Первая кривая лежит на поверхности $\mathbf{r}_i(u_i, v_i)$, а вторая — на плоскости симметрии. На базе линий пересечения построим новые ребра (ребра пересечения). С помощью признака совпадения направления ребра и его кривой пересечения ребрам пересечения дадим направление векторного произведения нормали грани тела с нормалью к плоскости: $t_{\text{edge}} = \mathbf{m}_i \times \mathbf{m}_p$.

На втором этапе определим точки пересечения новых и старых ребер тела, в этих точках построим вершины и этими вершинами разрежем старые ребра на несколько ребер. Резка ребер описана в булевых операциях над телами.

Далее построим симметричную относительно плоскости копию части оболочки тела, лежащей над плоскостью. Для этого построим копию части оболочки, преобразуем ее по матрице (8.1) и вывернем ее наизнанку. При выворачивании оболочки нужно произвести изменение признаков совпадения нормали поверхности и нормали ее грани на противоположные значения и переориентировать циклы.

Остается сшить симметричные половинки тела по ребрам пересечения тела с плоскостью симметрии. Но прежде выполним замену линий на плоскости, входящих в кривые пересечения построенных ребер, на линии на симметричной оболочке. Для этого вместо кривых (8.4) в ребра положим кривые

$$\begin{aligned} l_{uv}(t) &= [u_i(t) \ v_i(t)]^\top, & l_{uv}(t) &\in \mathbf{r}_i(u_i, v_i), \\ l_{uv}'(t) &= [u_i(t) \ v_i(t)]^\top, & l_{uv}'(t) &\in \mathbf{r}_i'(u_i, v_i). \end{aligned} \quad (8.5)$$

Первую двумерную кривую $l_{uv}(t)$ кривой пересечения сохраним нетронутой, а вместо второй возьмем двумерную кривую $l_{uv}'(t)$ являющуюся точной копией $l_{uv}(t)$ и лежащую на симметричной копии поверхности $\mathbf{r}_i(u_i, v_i)$. Теперь каждое ребро пересечения будет базироваться на двух симметричных половинках искомого тела.

На третьем этапе перестроим циклы пересеченных плоскостью граней тела. Перестроение циклов грани подробно описано в булевой операции объединения тел. Каждое ребро пересечения должно войти в цикл исходной грани с отрицательным флагом, а цикл ее симметричной части — с положительным флагом.

К пересеченным граням тела добавим непересеченные грани, которые топологически связаны с первыми, и соответствующие симметричные копии. Так мы получим оболочку симметричного тела.

При построении симметричной оболочки следует учитывать все случаи совпадения ребер пересечения и старых ребер тела и частичное перекрытие граней, описанные в булевых операциях над телами. Кроме того, следует учесть наличие нескольких оболочек у исходного тела.

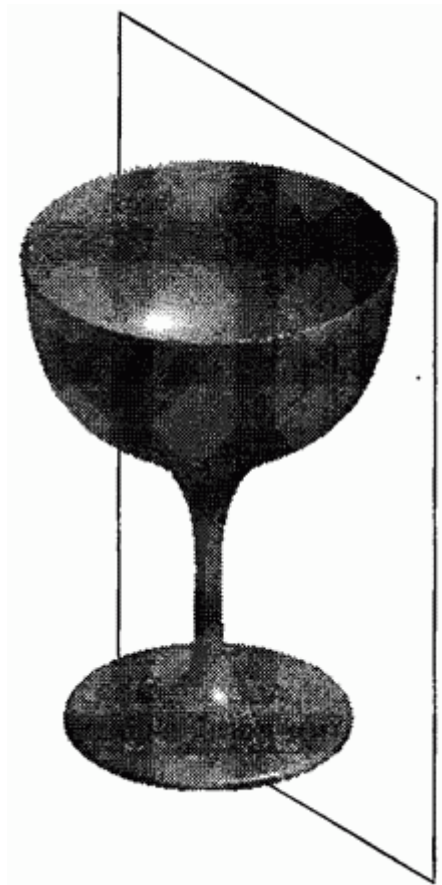


Рис. 8.1. Тело и плоскость симметрии

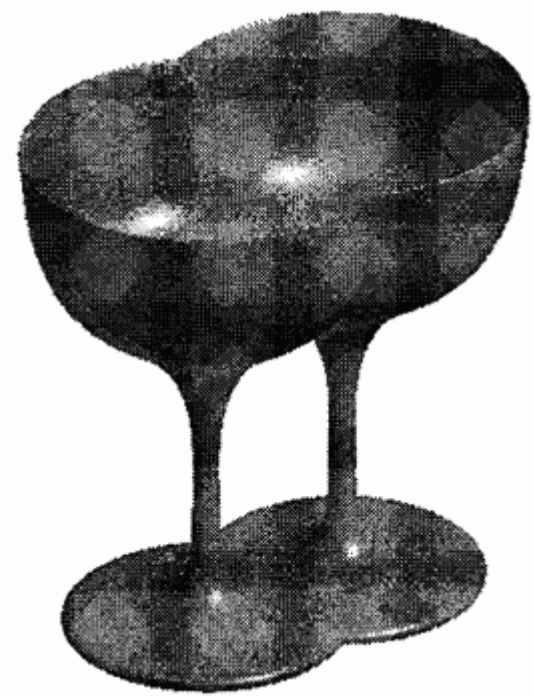


Рис. 6.8.2. Симметричное тело

На рис. 8.1 приведено исходное тело и плоскость симметрии. На рис. 8.2 приведен результат построения симметричного тела.

4.9. Построение эквидистантной оболочки тела

По данному телу можно построить тело с эквидистантной оболочкой. Оболочка нового тела расположена на заданном расстоянии по нормали от оболочки исходного тела. Это расстояние будем называть параметром эквидистанты и обозначим через h . Тело, по которому строится тело с эквидистантной оболочкой, будем называть базовым, а новое тело будем называть эквидистантным. Параметр эквидистанты этой операции может принимать как положительные, так и отрицательные значения. Если $h > 0$, то базовое тело располагается внутри эквидистантного тела. Если $h < 0$, то эквидистантное тело располагается внутри базового тела. Недопустимыми значениями

параметра являются такие, при которых оболочка нового тела получается самопересекающейся или вырожденной. Процесс построения оболочки эквидистантного тела условно разобьем на четыре этапа.

На первом этапе для каждой грани базового тела построим эквидистантную грань. Эквидистантная грань базируется на поверхности, эквидистантной к соответствующей поверхности базового тела. На рис. 9.1 приведены три грани базового тела, имеющие общую вершину *A*, и эквидистантные им поверхности.

Каждая эквидистантная поверхность должна быть продолжена до пересечения с соседними эквидистантными поверхностями. Радиус-вектор эквидистантной поверхности определяется известной формулой. На продолжении эквидистантной поверхности за пределы области определения параметров ее радиус-вектор будем вычислять по одной из известных формул в зависимости от замкнутости базовой поверхности.

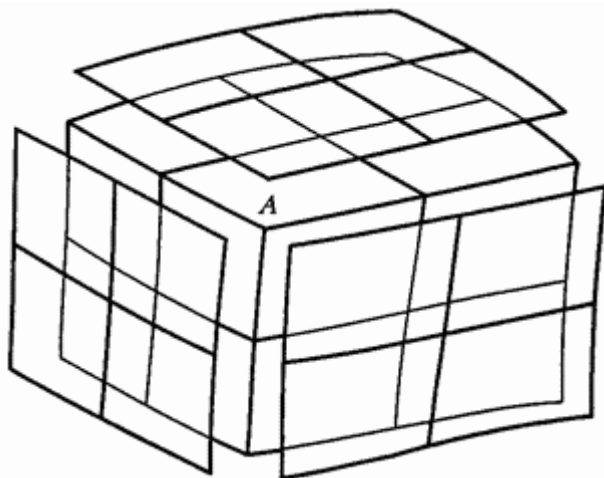


Рис. 9.1. Эквидистантные поверхности

Для построения эквидистантного тела нам остается построить его вершины и ребра. Для этого необходимо найти линии пересечения

эквидистантных поверхностей. Каждая линия пересечения должна начинаться и оканчиваться в вершине тела.

На втором этапе построим вершины эквидистантной оболочки. Рассмотрим последовательно вершины базового тела. В каждой вершине стыкуется несколько ребер. Нам нужно знать, какие ребра и какие грани базового тела стыкуются в данной вершине.

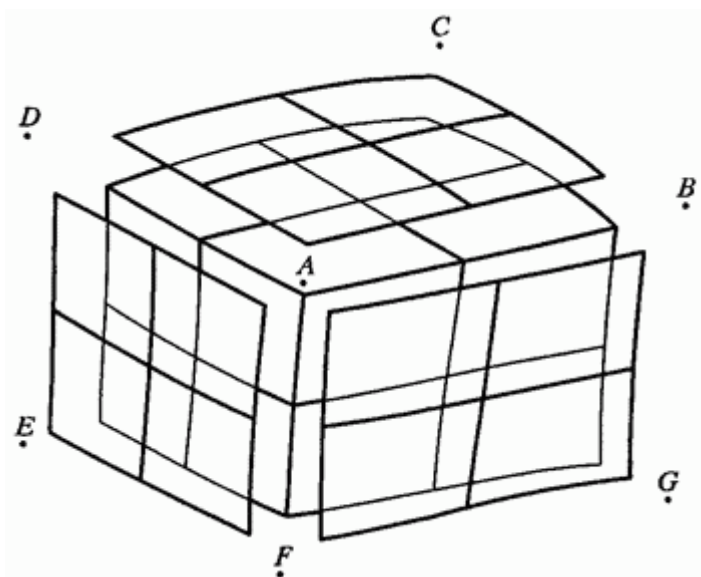


Рис. 9.2. Вершины эквидистантного тела

Вершине базового тела будут соответствовать одна или несколько вершин эквидистантного тела. В каждой вершине эквидистантного тела будут также стыковаться несколько ребер. Вершина эквидистантного тела базируется на точке пересечения эквидистантных поверхностей.

Вычислив эту точку, мы найдем параметры эквидистантных поверхностей, которые будут служить нам в качестве начальных и конечных точек линий пересечения поверхностей.

На рис. 9.2 приведены точки пересечения продолженных эквидистантных поверхностей, на которых будут базироваться вершины.

На третьем этапе построим ребра эквидистантной оболочки. Рассмотрим последовательно ребра базового тела. Для каждого ребра построим соответствующее ребро эквидистантной оболочки тела. Ребро будет базироваться на линии пересечения эквидистантных поверхностей.

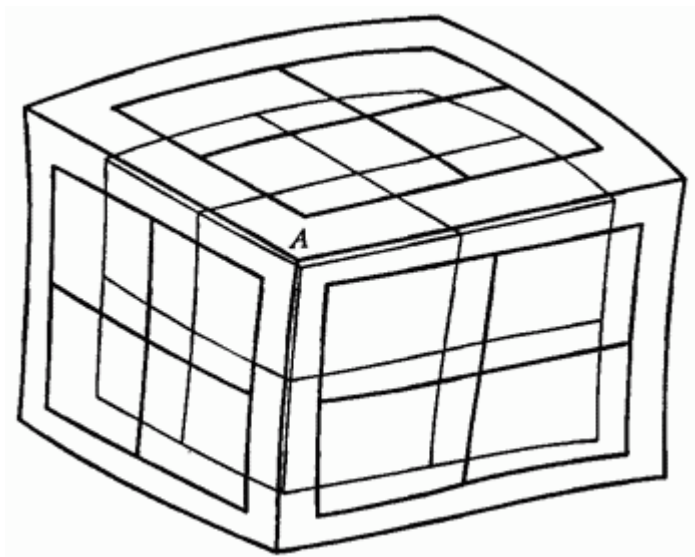


Рис. 9.3. Линии пересечения эквидистантных поверхностей

Начальные и конечные точки ребер нам известны из второго этапа. На рис. 9.3 приведены линии пересечения продолженных эквидистантных поверхностей, на которых будут базироваться ребра.

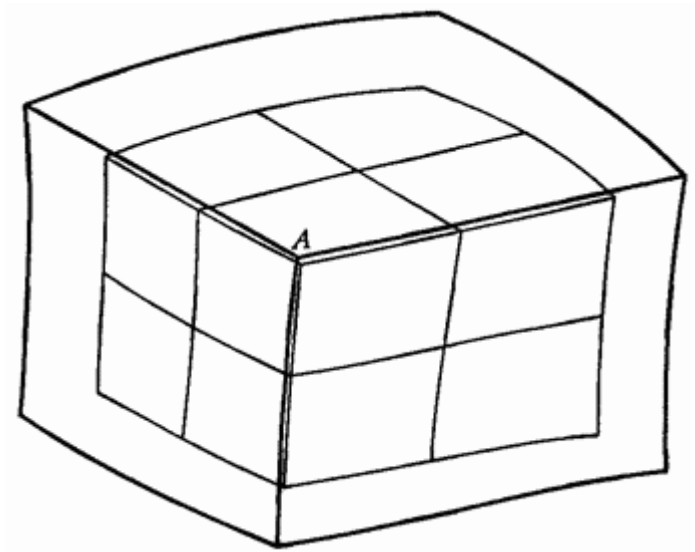


Рис. 9.4. Эквидистантные грани

На последнем, четвертом этапе построим циклы эквидистантных граней (рис. 9.4). Этот процесс аналогичен процессу перестроения циклов граней в булевых операциях.

Таким образом, мы получим оболочку эквидистантного тела.

Пример построения эквидистантного тела приведен на рис. 9.5 (исходное тело показано внутри тонкими линиями).

Заметим, что топология эквидистантного тела (количество вершин, ребер, граней и их взаимосвязь) не всегда совпадает с топологией базового тела.

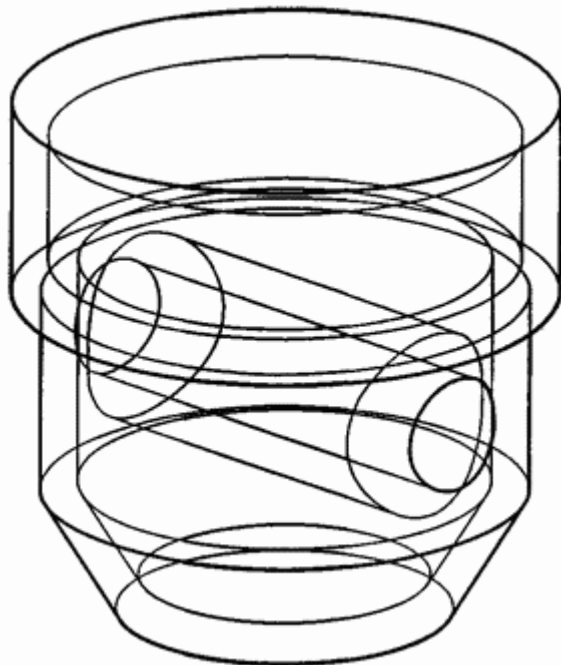


Рис. 9.5. Эквидистантное тело

Если в вершине базового тела стыкуется более трех ребер (не являющихся швами), то в эквидистантном теле этой вершине будет соответствовать несколько вершин и новых ребер. На рис. 9.6 приведена пирамида и эквидистантное пирамиде тело с отрицательным параметром эквидистанты, а на рис. 9.7 приведена аналогичная пирамида и эквидистантное пирамиде тело с положительным параметром эквидистанты.

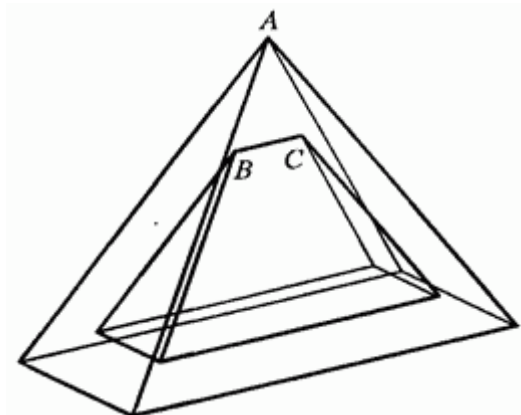


Рис. 9.6. Эквидистантная пирамида ($h < 0$)

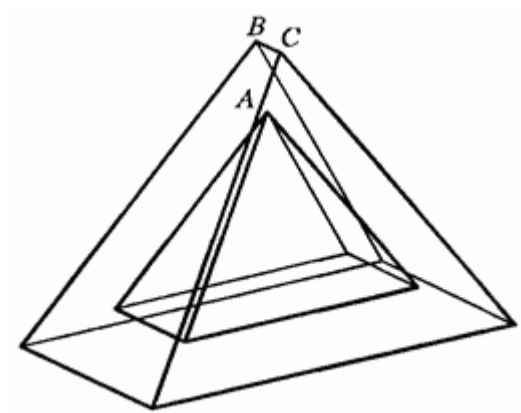


Рис. 9.7. Эквидистантная пирамида ($h > 0$)

В вершине А пирамиды стыкуется четыре ребра. В эквидистантном теле данной вершине соответствует две вершины В и С и одно дополнительное ребро ВС, что мы и наблюдаем на рис. 9.6 и 9.7.

Возможна и другая ситуация, когда некоторой вершине или некоторому ребру базового тела в эквидистантной оболочке не будет аналога. В таких случаях топология эквидистантного тела будет отличаться от топологии базового тела. Для обработки подобных

ситуаций следует проанализировать расположение соседних вершин и ребер.

4.10. Построение тонкостенного тела

Рассмотрим построение тонкостенных тел двух типов. Тела обоих типов будут строиться по некоторому базовому телу. Тело первого типа будет иметь замкнутую полость внутри. Оно имеет две оболочки, делящие пространство на три части: одна из них лежит вне тела, вторую занимает тело, а третья является внутренней полостью тела. Тело второго типа будет представлять собой открытое тонкостенное тело, имеющее одну оболочку. Тело первого типа будем называть закрытым тонкостенным телом, а тело второго типа — открытым тонкостенным телом. Закрытое тонкостенное тело можно считать частным случаем открытого тонкостенного тела.

Закрытое тело.

Закрытое тонкостенное тело с толщиной стенки h получим следующим образом. Построим эквидистантную оболочку базового тела оболочку. Процесс построения эквидистантной оболочки был описан при построении эквидистантного тела. Далее вывернем наизнанку одну из этих оболочек. Если $h > 0$, то вывернем наизнанку оболочку базового тела, если $h < 0$, то вывернем наизнанку эквидистантную оболочку. Эти две оболочки и создадут тонкостенное тело. Закрытое тонкостенное тело можно представить по рис. 9.5.

Данное тонкостенное тело является в отличие от других рассмотренных тел телом с пустотами. В общем случае тело с пустотами имеет несколько оболочек. Одна из них является внешней, а остальные — внутренними и лежат внутри внешней оболочки. Все оболочки не должны пересекать друг друга. Вектор нормали к внешней оболочке направлен вне объема оболочки, а векторы нормалей к внутренним оболочкам направлены внутрь объема, ограниченного ими.

Открытое тело.

Открытое тонкостенное тело строится на базе некоторого тела путем удаления одной или нескольких граней последнего и «придания

оставшимся граням конечной толщины». Конечно, грань не может иметь толщину, поэтому к оставшейся после удаления некоторых граней открытой оболочке строится эквидистантная открытая оболочка, а затем эти оболочки замыкаются частями удаляемых граней. В результате получается одна замкнутая оболочка.



Рис. 10.1. Исходное тело



Рис. 10.2. Тонкостенное тело

На рис. 10.2 приведен пример открытого тонкостенного тела ($h > 0$), построенного путем вскрытия одной грани тела, показанного на рис. 10.1.

Рассмотрим процесс построения оболочки открытого тонкостенного тела. Прежде всего, рассортируем грани базового тела на две группы: к первой группе отнесем вскрываемые грани базового тела, а ко второй

группе отнесем остальные грани базового тела, которые будем называть сохраняемыми. Сгруппируем ребра базового тела: к первой группе отнесем ребра, по которым пересекаются между собой вскрываемые грани базового тела, а ко второй группе отнесем ребра сохраняемых граней базового тела.

Процесс построения оболочки открытого тонкостенного тела имеет много общего с процессом построения эквидистантного тела.

Для каждой сохраняемой грани базового тела построим эквидистантную грань. Поверхность каждой эквидистантной грани и каждой вскрываемой грани должна иметь возможность быть продолженной до пересечения с поверхностями соседних граней. Радиус-вектор эквидистантной поверхности определяется известной формулой. На продолжении эквидистантной поверхности за пределы области определения параметров ее радиус-вектор будем вычислять по одной из известных формул в зависимости от замкнутости базовой поверхности.

Далее рассмотрим вершины сохраняемых граней базового тела. Каждой рассматриваемой вершине будут соответствовать одна или несколько вершин тонкостенного тела. В каждой вершине тонкостенного тела будут стыковаться несколько ребер. Вершина тонкостенного тела базируется на точке пересечения эквидистантных поверхностей или на точке пересечения эквидистантных поверхностей с поверхностями вскрываемых граней. Вычислив эти точки, мы найдем параметры пересекающихся поверхностей, которые будут служить нам в качестве начальных и конечных точек линий пересечения поверхностей.

Рассмотрим последовательно ребра сохраняемых граней базового тела (ребра второй группы). Для каждого ребра построим соответствующее ребро тонкостенного тела. Для этого найдем линии пересечения эквидистантных поверхностей между собой и с поверхностями вскрываемых граней. Начальные и конечные точки ребер нам известны.

По построенным вершинам перестроим ребра пересечения вскрываемых граней (ребра первой группы).

Далее вывернем наизнанку часть граней.

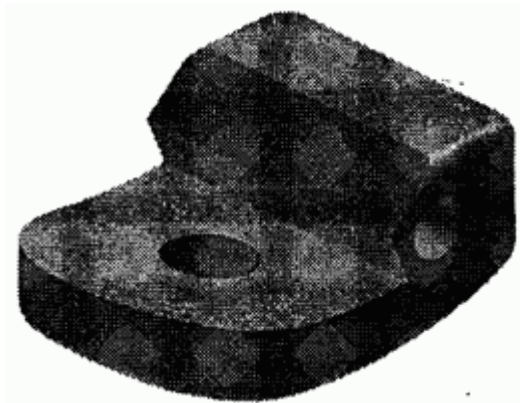


Рис. 10.3. Базовое тело



Рис. 10.4. Тонкостенное тело, построенное внутрь от базового тела

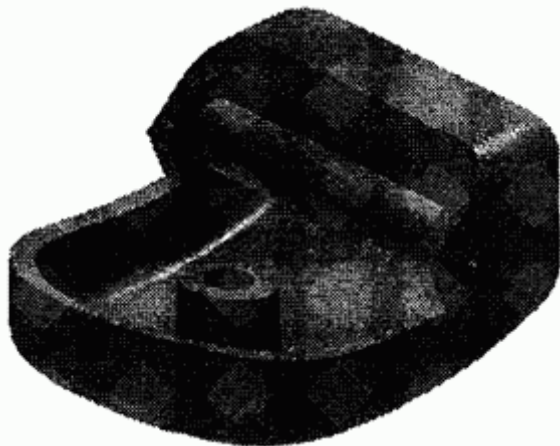


Рис. 10.5. Тонкостенное тело, построенное наружу от базового тела

Если $h > 0$, то изменим на противоположные направления циклов и нормалей сохраняемых граней базового тела и направления циклов вскрываемых граней базового тела. Если $h < 0$, то изменим на противоположные направления нормалей эквидистантных граней (циклов они еще не имеют).

На последнем этапе построим циклы эквидистантных граней и перестроим циклы вскрываемых граней. Таким образом, оболочка открытого тонкостенного тела будет составлена из сохраняемых граней, эквидистантных к ним граней и частей вскрываемых граней.

Открытое тонкостенное тело может быть построено как наружу ($h > 0$), так и внутрь ($h < 0$) от базового тела. На рис. 10.4 приведено открытое тонкостенное тело, построенное внутрь от базового тела путем вскрытия трех граней. Базовое тело приведено на рис. 10.3. На рис. 10.5 приведено открытое тонкостенное тело, построенное на основе этого же базового тела наружу от него тела путем вскрытия тех же трех граней (с большей толщиной стенки).

Операция построения тонкостенного тела наряду с булевыми операциями является мощным средством для построения тел сложной формы. На рис. 10.6 приведено открытое тонкостенное тело,

построенное на базе тела, показанного на рис. 10.5, путем вскрытия двух граней.

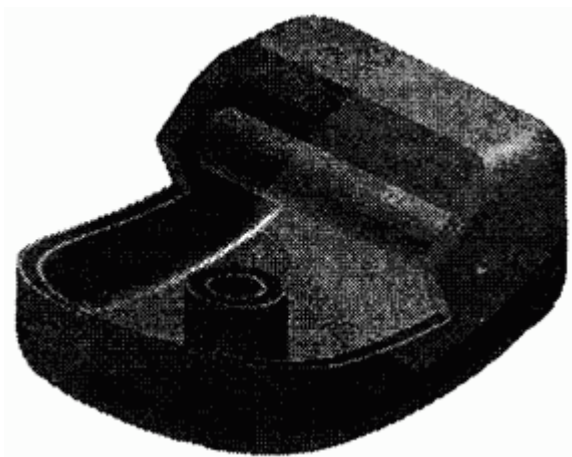


Рис. 10.6. Тонкостенное тело на базе тонкостенного тела

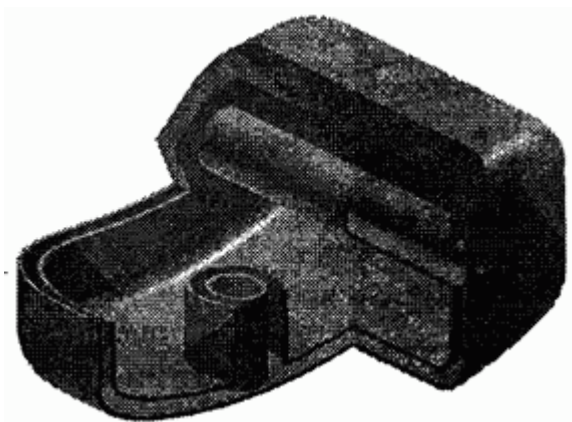


Рис. 10.7. Разрез тонкостенного тела

На рис. 10.7 приведен разрез двумя плоскостями этого тонкостенного тела, на котором видны внутренние полости тела.

Тонкостенное тело принадлежит к сложным телам. В его дерево построения положим структуру данных исходного тела, толщину стенки и список удаляемых граней.

4.11. Скругление ребер тела

Операция скругления ребер тела позволяет построить плавный переход от одной грани к другой. На рис. 11.1-11.4 приведены примеры скругления ребер тел. На рис. 11.1, 11.2 скруглены выпуклые ребра.

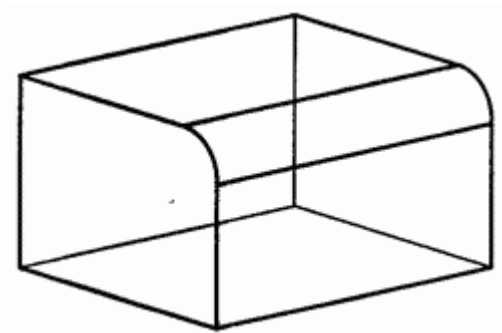


Рис. 11.1. Призма со скругленным ребром

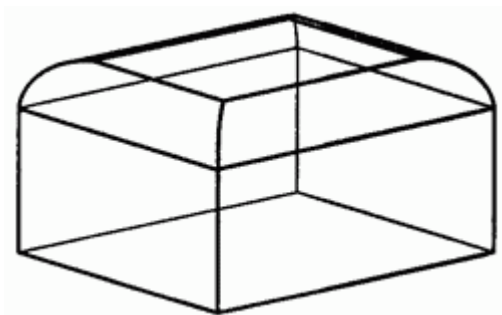


Рис. 11.2. Призма со скругленными ребрами

На рис. 11.3, 11.4 скруглены вогнутые ребра.

Рассмотрим общий случай скругления произвольного криволинейного ребра. Пусть грань, лежащая справа от скругляемого ребра (если смотреть снаружи тела вдоль ребра), базируется на поверхности $\mathbf{r}(u, v)$, а грань, лежащая слева от скругляемого ребра, базируется на поверхности $\mathbf{s}(a, b)$.

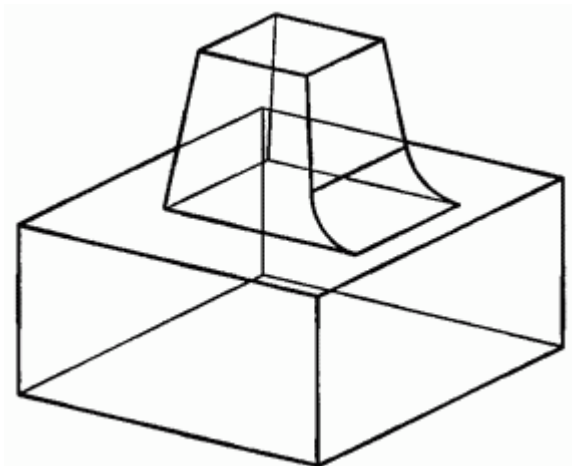


Рис. 11.3. Скругление ребра

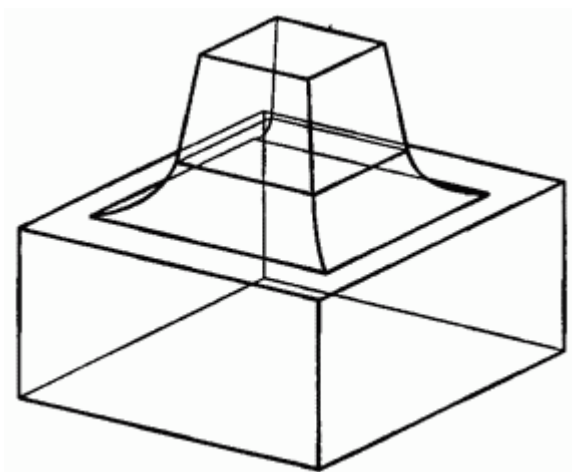


Рис. 11.4. Вариант скругления ребер

Эти грани будем называть сопрягаемыми. Пусть скругляемое ребро базируется на линии пересечения $\mathbf{c}(s)$ поверхностей $\mathbf{r}(u, v)$, $\mathbf{s}(a, b)$. По кривой $\mathbf{c}(s)$ построим поверхность скругления $\mathbf{q}(t, z)$, которая определяется известной формулой. Краями поверхности скругления являются линии, совпадающие с двухмерными линиями

$$\begin{aligned} l_{uv}(t) &= [u(t) \ v(t)]^T, & l_{uv}(t) &\in \mathbf{r}(u, v), \\ l_{ab}(t) &= [a(t) \ b(t)]^T, & l_{ab}(t) &\in \mathbf{s}(a, b) \end{aligned} \quad (11.1)$$

на сопрягаемых поверхностях. Построим два ребра вдоль краев поверхности скругления на базе линий пересечения

$$\begin{aligned} l_{uv}(t) &= [u(t) \ v(t)]^T, & l_{uv}(t) &\in \mathbf{r}(u, v), \\ l_{tz}(t) &= [t \ 0]^T, & l_{tz}(t) &\in \mathbf{q}(t, z), \\ l_{ab}(t) &= [a(t) \ b(t)]^T, & l_{ab}(t) &\in \mathbf{s}(a, b), \\ l_{tz}(t) &= [t \ 1]^T, & l_{tz}(t) &\in \mathbf{q}(t, z). \end{aligned} \quad (11.2) - (11.3)$$

Эти ребра будем называть продольными, так как они направлены вдоль скругляемого ребра и имеют ту же ориентацию.

Если скругляемое ребро замкнуто, то поверхность скругления также будет замкнутой по параметру t . Тогда на базе линии пересечения

$$\begin{aligned} l_{tz1}(w) &= [t_{\min} \ w]^T, & l_{tz1}(w) &\in \mathbf{q}(t, z), \\ l_{tz2}(w) &= [t_{\max} \ w]^T, & l_{tz2}(w) &\in \mathbf{q}(t, z), \\ & & 0 \leq w \leq 1, \end{aligned} \quad (11.4)$$

построим ребро, которое будет являться швом.

Если скругляемое ребро не замкнуто, то найдем все грани (за исключением сопрягаемых), пересекающиеся с поверхностью скругления (большая их часть стыкуется в начальной и конечной вершинах скругляемого ребра), и построим линии их пересечения с поверхностью скругления. На базе этих линий пересечения создадим ребра, которые будем называть поперечными.

На базе поверхности $\mathbf{q}(t, z)$ построим грань скругления. Цикл этой грани будет состоять из продольных и поперечных ребер (или шва). Для грани скругления определим признак совпадения ее нормали с направлением нормали поверхности. Нормаль грани скругления должна быть направлена наружу тела (нормаль поверхности скругления может совпадать с ней или быть ей противоположной).

После этого перестроим сопрягаемые поверхности $\mathbf{r}(u, v)$, $\mathbf{s}(a, b)$ и циклы граней на них. Для этого найдем пересечение линий (11.2) и (11.3) с ребрами сопрягаемых граней или их продолжениями и изменим эти ребра. В одних случаях упомянутые ребра нужно обрезать (рис. 11.1), в других случаях их нужно продлить (рис. 11.3). Вместо скругляемого ребра в цикл грани $\mathbf{r}(u, v)$ поставим ребро на базе линии пересечения (11.2), а в цикл грани $\mathbf{s}(a, b)$ поставим ребро на базе линии пересечения (11.3) (рис. 11.5).

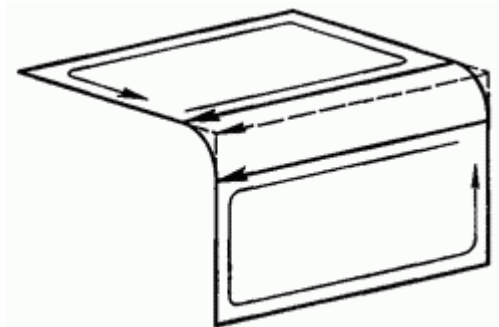


Рис. 11.5. Перестроение циклов граней при скруглении ребра

Далее перестроим циклы остальных граней, пересекшихся с поверхностью скругления $\mathbf{q}(t, z)$. В циклы этих граней войдут поперечные ребра. Перестроение циклов производится проверкой последовательности стыковки ребер между собой и составлением

списка ребер цикла в порядке их следования. Поперечные ребра определяют область изменения параметра t поверхности скругления $q(t, z)$.

Скругление сопряженных ребер.

Будем называть стыкующиеся ребра сопряженными, если в точках стыковки они имеют общую касательную. Если в точках стыковки ребра претерпевают излом, то такие ребра будем называть несопряженными.

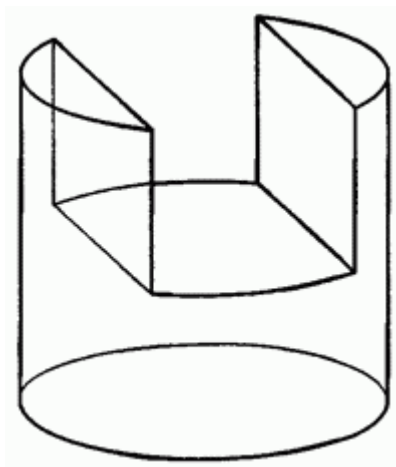


Рис. 11.6. Исходное тело

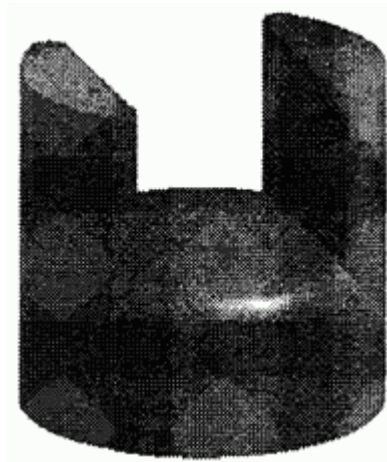


Рис. 11.7. Тело со скруглением нескольких ребер

Если скругляются сразу несколько несопряженных ребер тела, то выполним их скругление последовательно одно за другим. Скругление ребер тел, приведенных на рис. 11.2, 11.4, 11.7, выполнено последовательно.

В случаях скругления нескольких сопряженных ребер различные стадии операции скругления каждого ребра следует выполнять параллельно — одновременно для нескольких ребер. Перед началом операции скругления следует составить группы гладко стыкующихся ребер и далее работать с этими группами как с отдельным ребром описанным выше образом.

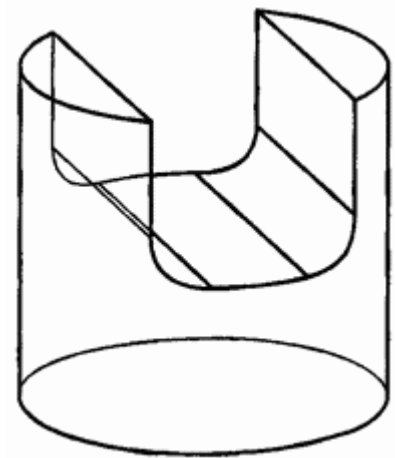


Рис. 11.8. Исходное тело

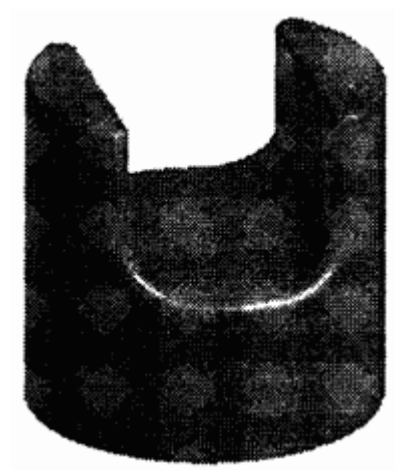


Рис. 11.9. Тело со скруглением цепочек ребер

Случай скругления нескольких сопряженных ребер приведен на рис. 11.9. Исходное тело, показанное на рис. 11.8, получено скруглением двух ребер тела, приведенного на рис. 11.6.

Для группы сопряженных ребер необходимо сначала построить все поверхности скругления, усечь этими поверхностями ребра тела, построить все продольные и поперечные ребра пересечения поверхностей скругления и граней тела, и только после этого построить грани скругления и произвести перестроение циклов граней тела.

На рис. 11.11 приведено тело со скруглением сопряженных ребер, у которых на концах продольные ребра грани скругления сходятся в одну точку (рис. 11.10).

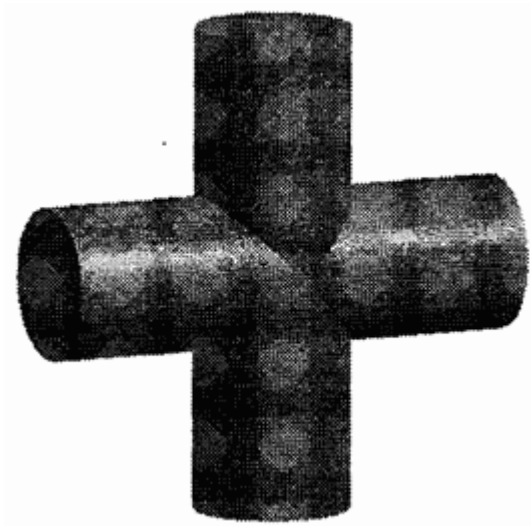


Рис. 11.10. Исходное тело

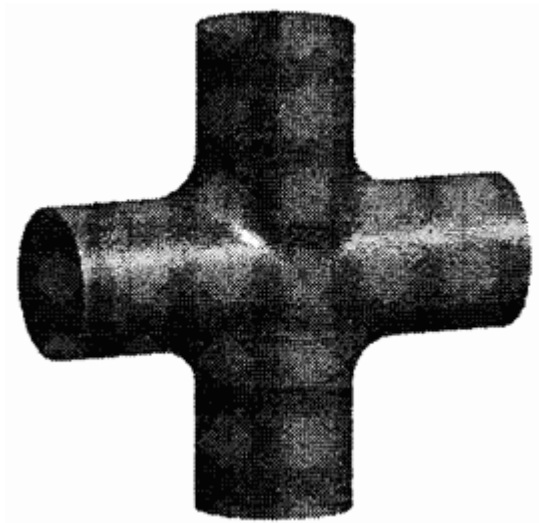


Рис. 11.11. Тело со скруглением цепочек ребер

В данном примере между сопряженными гранями поперечные ребра стянуты в точку, а поверхности скругления в них вырождаются.

Продольные ребра могут не полностью лежать в области сопрягаемых граней. В этом случае участки продольных ребер, выходящие за область определения сопрягаемых граней должны быть заменены на ребра пересечения грани скругления и соседних с сопрягаемой граней.

Скругление вершин.

Если скруглить три ребра, стыкующиеся в одной вершине, то в вершине получим картину, приведенную на рис. 11.12. Как правило, такую вершину скругляют. Скруглить вершину можно описанным выше способом: ребро пересечения двух граней скругления сопряжено с третьим ребром и может быть скруглено вместе с ним.

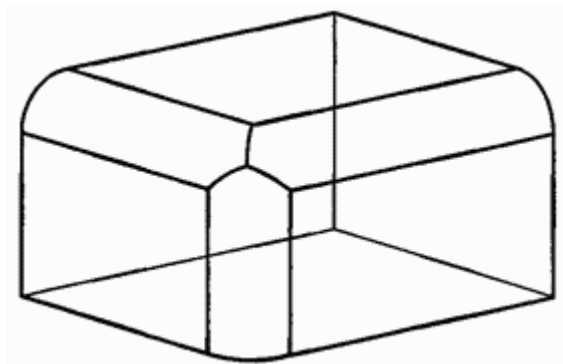


Рис. 11.12. Подлежащая скруглению вершина

Существует еще один способ скругления вершины. Для этого построим по одной линии на каждой поверхности скругления, используя то, что каждая из трех поверхностей скругления пересекается с двумя другими.

Рассмотрим построение упомянутой линии на одной из поверхностей скругления типа $\mathbf{q}(t, z)$. Пусть продольная линия $\mathbf{l}_r(t)$ этой поверхности скругления пересекается с продольной линией другой поверхности скругления в точке с параметром t_r , а продольная линия $\mathbf{l}_s(t)$ рассматриваемой поверхности скругления пересекается с продольной линией третьей поверхности скругления в точке с параметром t_s . Построим на рассматриваемой поверхности скругления двухмерный отрезок прямой из точки $[t_r, 1]^T$ в точку $[t_s, 1]^T$:

$$\mathbf{l}_{tz}(z) = [(1-z)t_r \quad zt_s]^T, \quad \mathbf{l}_{tz}(z) \in \mathbf{q}(t, z), \quad 0 \leq z \leq 1. \quad (11.5)$$

По отрезку и поверхности построим пространственную линию $\mathbf{l}_q(z)$. Линии (11.5) построим на каждой из трех поверхностей скругления, имеющих общую точку пересечения. Отрежем и опустим ту часть каждой грани скругления, которая лежит за построенной линией (11.5) (со стороны общей вершины). По трем линиям (11.5) построим поверхность, а на ее базе построим грань скругления вершины.

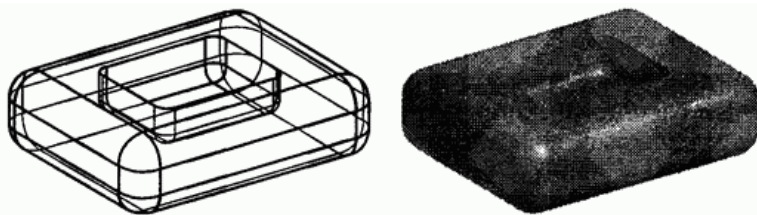


Рис. 11.13. Скругление ребер и вершин тела

На рис. 11.13 приведен пример скругления ребер и вершин призмы описанным способом.

Аналогично можно скруглить вершину, в которой стыкуются четыре скругляемых ребра. Для этого следует использовать поверхность, построенную по четырем кривым.

4.12. Построение фасок ребер тела

Фаски ребер тела строятся аналогично скруглению ребер с той лишь разницей, что поверхности скругления заменяются поверхностями фасок. Фаска описывается линейчатой поверхностью, построенной по двум линиям на пересекающихся по ребру поверхностях.

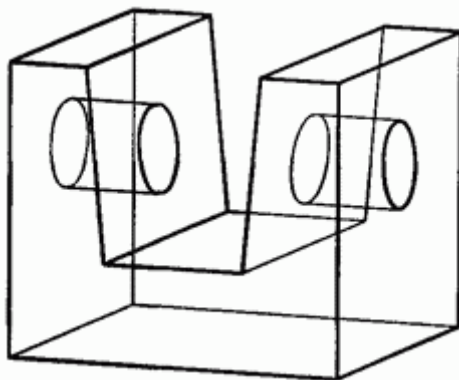


Рис. 12.1. Исходное тело

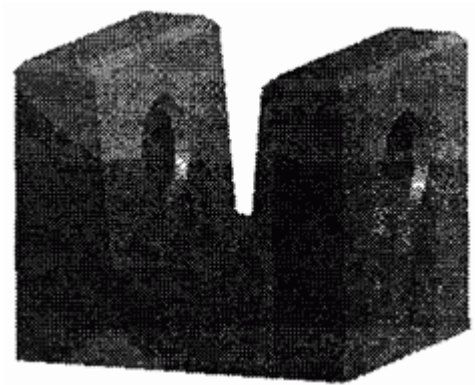


Рис. 12.2. Тело с фасками ребер

Результат построения фасок ребер тела, показанного на рис. 12.1, приведен на рис. 12.2.

Фаска вершин.

Если поверхности фасок строятся для трех ребер, стыкующихся в одной вершине, то общую вершину, как правило, срезают. Срез вершины выполним аналогично скруглению вершины.

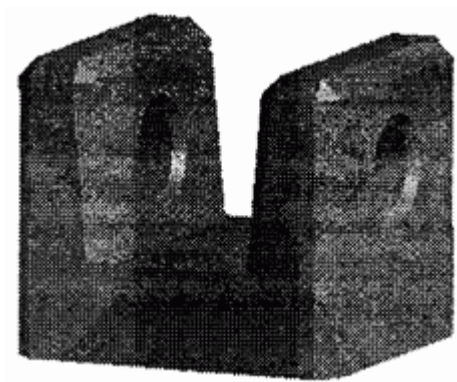


Рис. 12.3. Тело с фасками ребер и вершин

Для этого найдем три линии пересечения поверхностей фасок и построим по ним треугольную поверхность. На базе трех линий пересечения построим три ребра, а на базе треугольной поверхности построим грань. Результат среза вершины, в которой стыкуются три ребра, приведен на рис. 12.3.

4.13. Некоторые способы построения тел

Описанный выше метод выполнения булевых операций применим в случае полной определенности оболочек операндов. На практике часто требуется выполнить булеву операцию, когда оболочка одного из операндов определена не полностью и должна быть достроена в процессе операции. Например, к заданному телу нужно добавить часть тела, полученного выдавливанием заданного плоского контура, лежащую со стороны контура (рис. 13.1).

Результат такой операции приведен на рис. 13.2.

В данном случае требовалось «выдавить» заданный контур до ближайших к нему граней заданного тела. В разных местах контура ближайшими могут оказаться разные грани, поэтому точки контура должны быть «выдавлены» на различные расстояния.

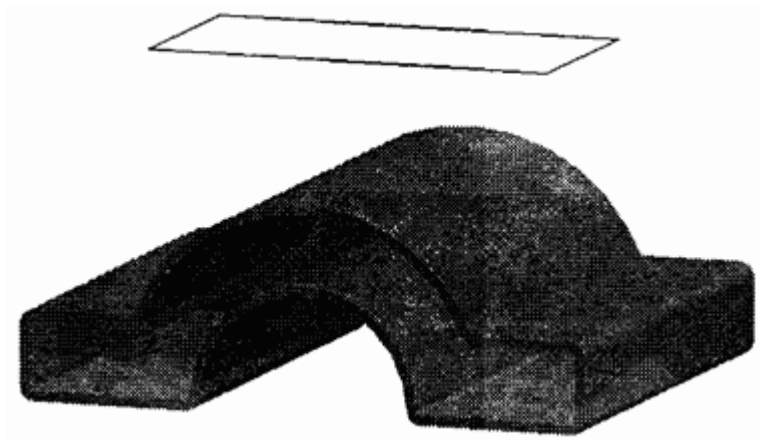


Рис. 13.1. Исходное тело и плоский контур

Выполнить данную операцию можно следующим образом. По заданному контуру построим тело выдавливания достаточной глубины для пересечения с заданным телом. Далее из тела выдавливания вычтем заданное тело. В результате операции мы в общем случае получим несколько оболочек. Выберем из них ближайшую к заданному контуру.

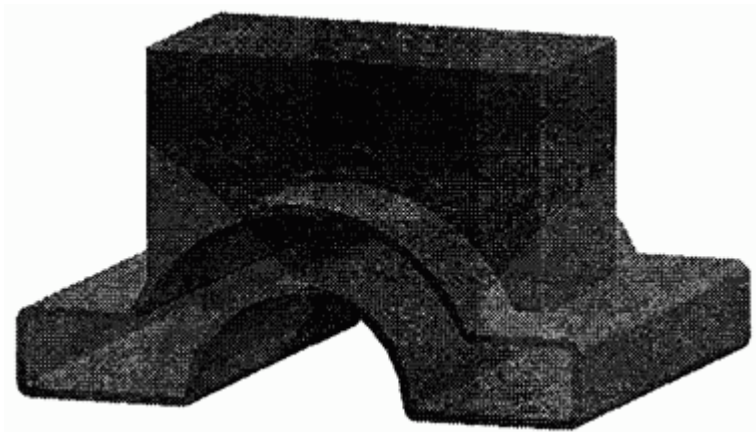


Рис. 13.2. Выдавливание контура до ближайшей поверхности тела

Тело с выбранной оболочкой объединим с заданным телом. При выполнении данных действий достаточно строить ребра пересечения только один раз. Во второй части операции (объединение тела с выбранной оболочкой) следует использовать только те ребра, которые принадлежат телу с выбранной оболочкой.

Другим примером может служить операция «вырезки» ближайших граней заданного тела заданным контуром (рис. 13.3).

В данном примере контуром вырезаются ближайшие стенки тела и не трогаются грани, лежащие за ними. В разных местах контура его точки «выдавливаются» на различные расстояния.

Вырезать ближайшие грани заданного тела заданным контуром можно следующим образом. По заданному контуру построим тело выдавливания достаточной глубины для пересечения с заданным телом.

Далее найдем пересечение тела выдавливания с заданным телом. В результате операции мы в общем случае получим несколько оболочек. Выберем из них ближайшую к заданному контуру. Тело с выбранной оболочкой вычтем из заданного тела. При выполнении данных действий достаточно строить ребра пересечения только один раз.

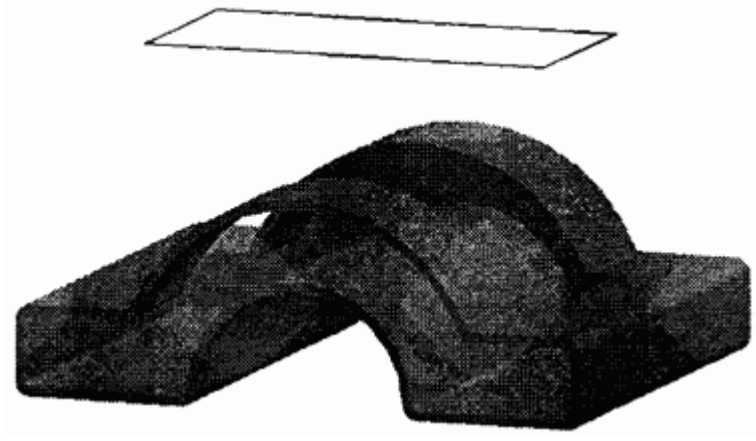


Рис. 13.3. Вырез контуром ближайших поверхностей тела

Во многих конструкциях используются ребра жесткости. Операция построения ребер жесткости выполняется по той же схеме, что и операция «выдавливания» контура до ближайших к нему граней тела. Тело с ребрами жесткости приведено на рис. 13.4.

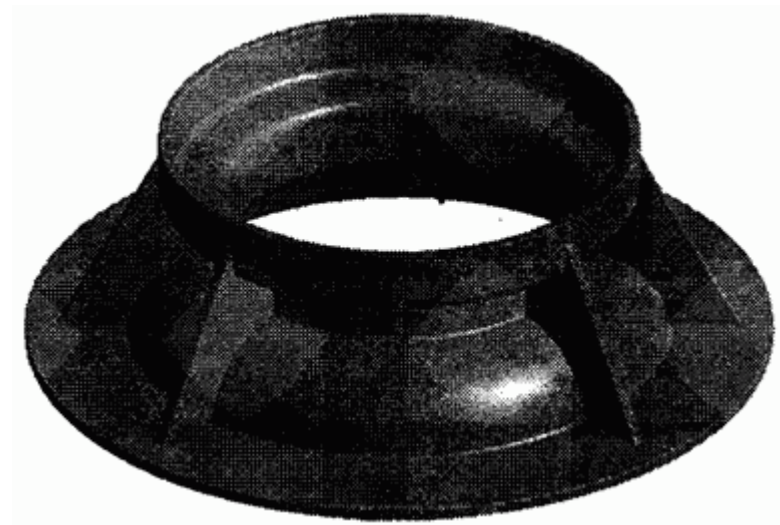


Рис. 13.4. Тело с ребрами жесткости

Когда оболочка одного из тел-операндов должна быть построена в процессе выполнения операции, после выполнения операций следует проверить оболочку результирующего тела на замкнутость.

4.14. Последовательность моделирования тел

Моделирование некоторого объекта может включать построение одного тела или построение нескольких тел — сборки тел.

Создание одиночного тела начинается с построения или одного из простых тел, или тела на базе линий, или тела на базе поверхности. Эти способы построения тел приведены в левой части рис. 14.1. Если исходное тело создается на базе плоских линий, то для их построения используются конструктивные плоскости.

Перед построением тела на базе поверхности нужно сначала создать исходную поверхность. Во многих случаях при построении поверхности также используются конструктивные плоскости. Далее путем выполнения операций, приведенных в правой части рис. 14.1, из исходного тела можно получить тело с более сложной оболочкой.

Операции над телом могут выполняться многократно и в произвольной последовательности, что отражено на рис. 14.1.



Рис. 14.1. Способы построения тел

Процесс построения оболочки сложного тела близок к процессу изготовления моделируемого объекта. С помощью булевой операции объединения к телу можно добавить требуемый объем. Для этого нужно построить еще одно тело и объединить исходное тело с ним. Аналогично с помощью булевых операций пересечения или вычитания из тела можно убрать требуемый объем. От тела может быть отрезана лишняя часть объема. Ребра тела могут быть скруглены или с них могут быть сняты фаски. Из тела можно получить тонкостенное тело путем «вскрытия» одной или нескольких граней и «придания оставшимся граням конечной толщины». Для симметричных тел можно построить только одну половину тела, а затем получить требуемое тело с помощью операции создания симметричного тела. К телу могут быть добавлены ребра жесткости.

Из нескольких тел можно получить сборку тел. В сборке все тела равноправны. Для взаимного расположения и ориентации тел сборки можно использовать преобразования сдвига, поворота, масштабирования или симметрии.

Построение отдельных тел и сборок должно сопровождаться протоколом построения, который называют деревом построения. Дерево построения позволяет выполнять редактирование тел и их сборок, создавать наборы однотипных моделей и управлять ими.

Наряду с деревом построения для редактирования кривых линий, поверхностей, тел, сборок тел и управления ими может использоваться механизм вариационных связей. Этому механизму посвящена следующая глава.

5. ВАРИАЦИОННЫЕ СВЯЗИ ГЕОМЕТРИЧЕСКИХ ОБЪЕКТОВ

5.1. Наложение вариационных связей

В данной главе мы рассмотрим один из эффективных способов управления информацией о геометрических объектах. Каждый геометрический объект имеет свою структуру данных. **Структура данных вместе с набором необходимых объекту функций представляет собой численную модель геометрического объекта. Скалярные величины, компоненты векторов, координаты точек, лежащие в структуре данных геометрического объекта, которые подлежат редактированию, будем называть параметрами этого объекта. Именно через эти параметры мы и будем осуществлять управление информацией о геометрических объектах.**

До сих пор геометрические объекты строились и существовали независимо друг от друга, т. е. редактирование одного из объектов не сказывалось на остальных объектах. **Редактирование объекта сводится к изменению численных значений его параметров.** Независимость геометрических объектов отражает тот факт, что значения параметров одного объекта не зависят от значений параметров других объектов. С практической точки зрения наложение зависимостей на параметры геометрических объектов является очень полезным.

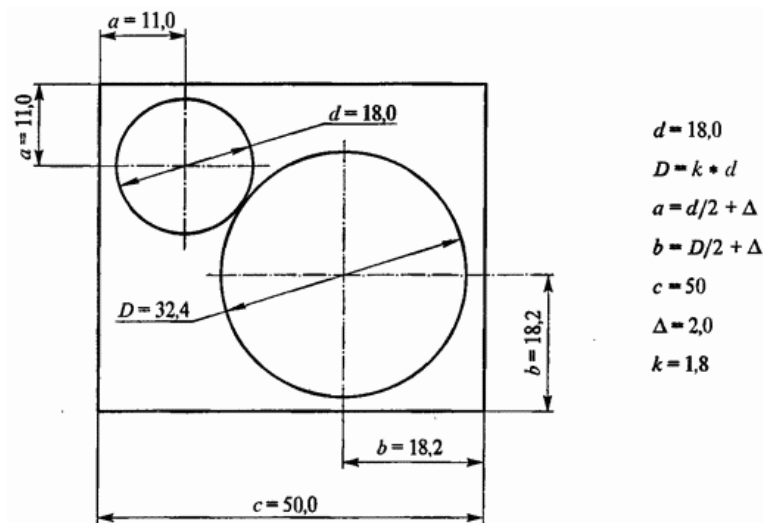


Рис. 1.1. Вариационные связи геометрических объектов

На рис. 1.1 приведен пример наложения зависимостей на две окружности и замкнутую ломаную линию на плоскости.

Зависимость геометрических объектов заключается в том, что окружности должны касаться друг друга, ломаная должна представлять собой прямоугольник, диаметры d и D окружностей связаны заданным коэффициентом k , центры окружностей должны отстоять на заданных расстояниях a и b от сторон прямоугольника, размер горизонтальной стороны прямоугольника должен быть равен c . Перечисленные связи представлены размерами и алгебраическими уравнениями, приведенными на рис. 1.1. Эскиз автоматически перестраивается при изменении одной или нескольких зависимостей. На рис. 1.2 приведены те же геометрические объекты при другом отношении диаметров окружностей.

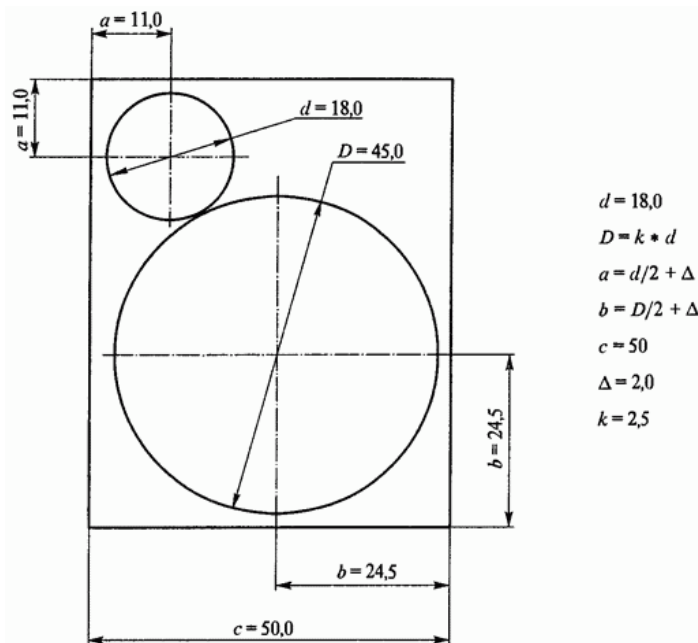


Рис. 1.2. Результат изменения одной из связей геометрических объектов

Изменение коэффициента отношения диаметров к привело к изменению размера вертикальной стороны прямоугольника, изменению диаметра и положения центра большей окружности. Все заданные уравнения при новом коэффициенте к также выполняются. Аналогично могут быть изменены и другие размеры и уравнения, что приведет к автоматическому перестроению геометрических объектов.

Наложение связей облегчает труд при проектировании нескольких однотипных деталей и при сборке различных деталей. Достигается это путем установления определенных зависимостей между параметрами геометрических объектов. Эти зависимости представляют собой некоторые уравнения

решение осуществляется итерационным методом. Объекты должны перестраиваться на каждой итерации решения, так как одни параметры объекта могут влиять на другие параметры этого же геометрического объекта. Уравнения будут изменять значения параметров, а объекты — перестраиваться в соответствии с новыми значениями параметров. Когда все уравнения системы будут удовлетворены с требуемой точностью, объекты будут перестроены соответствующим образом. Если систему уравнений удовлетворить нельзя, то всем параметрам присваиваются первоначальные значения. На практике часто приходится иметь дело с системой уравнений, содержащей большее число параметров, чем число уравнений. В последнем случае мы воспользуемся некоторым критерием, определяющим поведение всей системы параметров, и с помощью этого критерия сформируем систему уравнений для определения всех параметров.

Дополним вариационные связи информацией связываемых ими геометрических объектах, а также функциями общения с объектами и системой уравнений. В результате вариационные связи можно будет называть вариационными объектами. Они несут и обрабатывают геометрическую информацию. Как и геометрические объекты, вариационные объекты имеют свою структуру данных и свой набор функций. Состав структуры данных и функций определяется выполняемыми связями задачами. Каждый вариационный объект отвечает за то, чтобы заданные параметры удовлетворяли заданным уравнениям. В структуре данных вариационного объекта должны находиться: информация о связываемых им варьируемых параметрах, исходные значения параметров, информация об уравнениях связи (одном или нескольких).

В набор функций вариационных объектов должны войти функции, предоставляющие и изменяющие необходимую информацию о варьируемых параметрах, функции, предоставляющие информацию об уравнениях связей, функции, изменяющие параметры в процессе решения и после удовлетворения уравнений всех связей, функции восстановления исходного состояния параметров в случае неудачи в процессе решения.

Сначала мы рассмотрим отдельные вариационные связи и их уравнения. Далее введем критерий поведения геометрических объектов, позволяющий сформировать систему уравнений, в которой варьируемые параметры обладают равноправием при любом их числе.

Этот критерий в общем случае требует привлечение методов вариационного исчисления. На примере использования критерия поведения геометрических объектов мы рассмотрим вариационные связи двухмерных объектов.

5.2. Фиксирующие связи

Простейшей связью является фиксирующая связь. Она описывается одним уравнением, содержащим один параметр. Если параметр обозначить через q , то уравнение имеет вид

$$q = q^{(0)} = \text{const}, \quad (2.1)$$

где $q^{(0)}$ — заданное значение параметра, которое должно быть сохранено в дальнейшем. Фиксирующая связь используется тогда, когда требуется, чтобы некоторый параметр не изменялся в процессе решения системы уравнений связей. Если некоторый параметр неизменен, то он может быть удален из списка переменных системы уравнений связей. Иногда удобнее сохранить неизменный параметр, но при этом следует ввести для него фиксирующее уравнение (2.1).

Примером фиксирующей связи может служить вариационная зависимость, называемая закреплением точки. Она содержит три уравнения типа (2.1) и фиксирует три координаты точки $\mathbf{p}=[x \ y \ z]^T$. Закрепление точки описывается одним векторным уравнением

$$\mathbf{p} = \mathbf{p}^{(0)}, \quad (2.2)$$

состоящим из трех скалярных уравнений

$$x = x^{(0)}, \quad y = y^{(0)}, \quad z = z^{(0)}, \quad (2.3)$$

где $\mathbf{p}^{(0)}=[x^{(0)} \ y^{(0)} \ z^{(0)}]^T$ — заданное положение точки. Координаты закрепленных точек остаются неизменными, поэтому частные производные всех других уравнений системы по закрепленным параметрам равны нулю.

Уравнение (2.1) может использоваться для любого параметра геометрического объекта — скалярной величины, компоненты вектора или координаты точки.

Фиксирующие связи занимают особое место среди вариационных связей. Количество параметров в фиксирующих связях всегда равно количеству уравнений, сами параметры остаются неизменными, поэтому фиксирующие связи могут быть выделены из общей системы связей в самостоятельную группу. Система уравнений фиксирующих связей и система уравнений остальных связей не зависят друг от друга и могут быть решены отдельно.

5.3. Вариационные связи точек в пространстве

Наиболее простыми являются вариационные связи, накладываемые на координаты радиус-векторов точек. Эти связи могут быть наложены на две или несколько точек. При этом число уравнений связей может быть меньше числа участвующих в связях параметров. В этом случае можно составить дополнительные уравнения, необходимые для решения задачи. Мы рассмотрим пример, построения вариационной связи, фиксирующей расстояние между двумя точками в пространстве. В этом примере мы составим дополнительные уравнения исходя из симметрии поведения геометрических объектов. Дополнительные уравнения нам потребуются, чтобы проиллюстрировать процесс решения системы уравнений связи.

Линейный размер.

Пусть имеются две точки $\mathbf{p}_1=[x_1 \ y_1 \ z_1]^T$, $\mathbf{p}_2=[x_2 \ y_2 \ z_2]^T$. Наложим на точки вариационную связь, называемую линейным размером. Она описывается уравнением

$$|\mathbf{p}_2 - \mathbf{p}_1| - d = 0, \quad (3.1)$$

где d — требуемый размер. В координатном представлении уравнение связи имеет вид

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} - d = 0.$$

Линейный размер приведен на рис. 3.1. Точка p_1 лежит в структуре данных тора и определяет положение центра его местной системы координат. Точка p_2 лежит в структуре данных сферы и определяет положение ее центра.

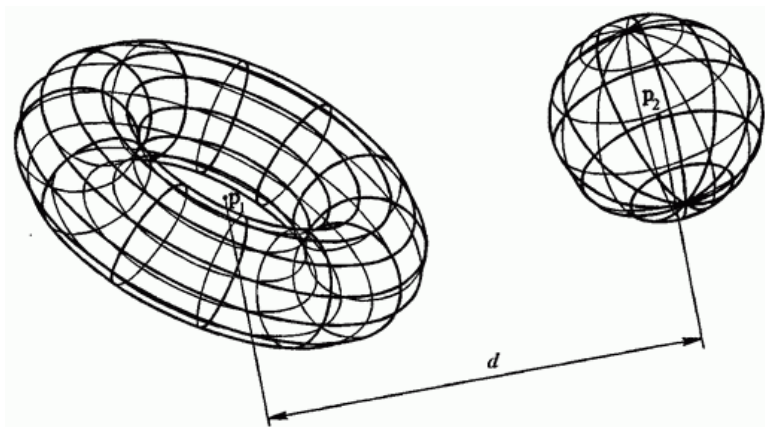


Рис. 3.1. Линейный размер между точками центров тора и сферы

В общем случае это могут быть любые точки из структуры данных геометрических объектов. Мы имеем одно уравнение (3.1), связывающее в общем случае шесть параметров: $x_1, y_1, z_1, x_2, y_2, z_2$. В уравнении (3.1) величину d будем считать константой. В общем случае величина d также может быть варьируемым параметром.

Рассмотрим принцип решения поставленной задачи. В данном случае для того, чтобы найти все параметры, необходимы шесть уравнений. Существует несколько способов получить систему уравнений с числом варьируемых параметров (неизвестных) равным числу уравнений. В данном случае мы составим некоторые дополнительные уравнения, связывающие те же параметры. Набор дополнительных уравнений может быть различным, но от этих дополнительных уравнений будет зависеть поведение точек. Потребуем, например, чтобы при изменении размера точки перемещались симметрично, т. е. точки перемещались бы вдоль соединяющей их линии, а центр тяжести точек оставался бы неподвижным. Эти условия описываются уравнениями

$$\begin{aligned}(x_2 - x_1)(y_2^{(0)} - y_1^{(0)}) &= (y_2 - y_1)(x_2^{(0)} - x_1^{(0)}), \\ (y_2 - y_1)(z_2^{(0)} - z_1^{(0)}) &= (z_2 - z_1)(y_2^{(0)} - y_1^{(0)}), \\ x_2 + x_1 &= x_2^{(0)} + x_1^{(0)}, \\ y_2 + y_1 &= y_2^{(0)} + y_1^{(0)}, \\ z_2 + z_1 &= z_2^{(0)} + z_1^{(0)},\end{aligned}\tag{3.2)-(3.6}$$

где $\mathbf{p}_1^{(0)}=[x_1^{(0)} y_1^{(0)} z_1^{(0)}]^T$, $\mathbf{p}_2^{(0)}=[x_2^{(0)} y_2^{(0)} z_2^{(0)}]^T$ — исходное положение точек, которое в общем случае не удовлетворяет уравнению (3.1). Добавив (3.2)-(3.6) к (3.1), получим системы шести уравнений относительно шести параметров. Перепишем все уравнения, придав им вид $f_i(x_1 y_1 z_1, x_2 y_2 z_2)=0$, $i=1, 2, \dots, 6$. Первое уравнение является нелинейным.

Решим систему уравнений методом Ньютона. Организуем итерационный процесс

$$\begin{aligned}x_i^{(r+1)} &= x_i^{(r)} + \Delta x_i, \\ y_i^{(r+1)} &= y_i^{(r)} + \Delta y_i, \\ z_i^{(r+1)} &= z_i^{(r)} + \Delta z_i, \\ i &= 1, 2,\end{aligned}\tag{3.7}$$

где r — номер итерации. Приращения координат $\Delta x_i, \Delta y_i, \Delta z_i$, $i=1,2$, на каждой итерации определим из системы линейных алгебраических уравнений

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & 0 & a_{24} & a_{25} & 0 \\ 0 & a_{32} & a_{33} & 0 & a_{35} & a_{36} \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \Delta x_1 \\ \Delta y_1 \\ \Delta z_1 \\ \Delta x_2 \\ \Delta y_2 \\ \Delta z_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} \quad (3.8)$$

где

$$\begin{aligned} a_{11} = -a_{14} &= \frac{x_1 - x_2}{d^{(r)}}, & a_{12} = -a_{15} &= \frac{y_1 - y_2}{d^{(r)}}, & a_{13} = -a_{16} &= \frac{z_1 - z_2}{d^{(r)}}, \\ a_{21} = -a_{24} &= y_1^{(0)} - y_2^{(0)}, & a_{22} = -a_{25} &= x_2^{(0)} - x_1^{(0)}, \\ a_{32} = -a_{35} &= z_1^{(0)} - z_2^{(0)}, & a_{33} = -a_{36} &= y_2^{(0)} - y_1^{(0)}, \\ b_1 &= d - d^{(r)}, \\ b_2 &= (y_2^{(r)} - y_1^{(r)})(x_2^{(0)} - x_1^{(0)}) - (x_2^{(r)} - x_1^{(r)})(y_2^{(0)} - y_1^{(0)}), \\ b_3 &= (z_2^{(r)} - z_1^{(r)})(y_2^{(0)} - y_1^{(0)}) - (y_2^{(r)} - y_1^{(r)})(z_2^{(0)} - z_1^{(0)}), \\ b_4 &= x_2^{(0)} + x_1^{(0)} - x_2^{(r)} - x_1^{(r)}, \\ b_5 &= y_2^{(0)} + y_1^{(0)} - y_2^{(r)} - y_1^{(r)}, \\ b_6 &= z_2^{(0)} + z_1^{(0)} - z_2^{(r)} - z_1^{(r)}, \\ d^{(r)} &= \sqrt{(x_2^{(r)} - x_1^{(r)})^2 + (y_2^{(r)} - y_1^{(r)})^2 + (z_2^{(r)} - z_1^{(r)})^2}. \end{aligned}$$

Система уравнений (3.8) в матричной записи имеет вид

$$\mathbf{A} \cdot \Delta \mathbf{x} = \mathbf{b}. \quad (3.9)$$

Эта система имеет единственное решение, если определитель $|\mathbf{A}| \neq 0$. На каждой итерации матрица \mathbf{A} вычисляется заново, так как точки перемещаются в пространстве, и может случиться так, что определитель матрицы \mathbf{A} на очередной итерации окажется равным нулю. Тогда точки следует вернуть в исходное положение, а размер положить равным расстоянию между ними. Каждая строка матрицы \mathbf{A} соответствует определенному уравнению. Для i -го уравнения $f_i(x_1, y_1, z_1, x_2, y_2, z_2) = 0$ в системе (3.8) коэффициенты i -й строки матрицы \mathbf{A} определяются по формулам

$$a_{i1} = \frac{\partial f_i}{\partial x_1}, \quad a_{i2} = \frac{\partial f_i}{\partial y_1}, \quad a_{i3} = \frac{\partial f_i}{\partial z_1}, \quad a_{i4} = \frac{\partial f_i}{\partial x_2}, \quad a_{i5} = \frac{\partial f_i}{\partial y_2}, \quad a_{i6} = \frac{\partial f_i}{\partial z_2},$$

Итерационный процесс (3.7) закончим, когда приращения координат на очередной итерации станут меньше заданной величины.

Возможны и другие варианты дополнительных уравнений. Например, если требуется сохранить точку \mathbf{p}_1 в заданном положении, то в качестве дополнительных уравнений можно использовать уравнения (3.2), (3.3) и

$$\mathbf{p}_1 = \mathbf{p}^{(0)},$$

где $\mathbf{p}^{(0)}$ — заданное положение точки. В дальнейшем для решения системы уравнений связей мы не будем использовать дополнительные уравнения, так как они нарушают равновесие варьируемых параметров.

Размер вдоль координаты.

Более простой, чем линейный размер, вариационной связью двух точек является размер вдоль одной координаты.

Она связывает две соответствующие координаты двух точек и описывается одним из уравнений

$$x_2 - x_1 = d_x,$$

$$y_2 - y_1 = d_y,$$

$$z_2 - z_1 = d_z,$$

(3.10.1)-(3.10.3)

где d_x , d_y , d_z — требуемые размеры. Уравнения (3.10), примененные к координатам точек отрезка $\mathbf{r}(t)=(1-t)\mathbf{p}_1+t\mathbf{p}_2$, могут сделать отрезок параллельным соответствующей координатной плоскости. Два уравнения (3.10) могут сделать отрезок параллельным одной из координатных осей.

Размер вдоль направления.

Еще одной вариационной связью двух точек является размер вдоль заданного направления. Она описывается уравнением

$$|(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{m}| = d, \quad (3.11)$$

где $\mathbf{m}=[m_x m_y m_z]^T$ — задающий направление вектор единичной длины, d — требуемый размер. Уравнение (3.11) устанавливает, что проекция вектора $\mathbf{p}_2 - \mathbf{p}_1$ построенного между заданными точками, на направление вектора \mathbf{m} по абсолютной величине равна d . В координатной записи уравнение (3.11) имеет вид

$$|(x_2 - x_1)m_x + (y_2 - y_1)m_y + (z_2 - z_1)m_z| = d. \quad (3.12)$$

Уравнение (3.12) связывает шесть параметров. Если $\mathbf{m}=[m_x m_y m_z]^T$, $\mathbf{l}=[l_x l_y l_z]^T$, $\mathbf{n}=[n_x n_y n_z]^T$ — три линейно независимых вектора единичной длины, то можно задать относительное положение двух точек \mathbf{p}_1 , \mathbf{p}_2 сразу тремя размерами вдоль трех некопланарных направлений \mathbf{m} , \mathbf{l} , \mathbf{n} :

$$\begin{aligned} |(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{m}| &= d_m, \\ |(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{l}| &= d_l, \\ |(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{n}| &= d_n. \end{aligned} \quad (3.13)$$

В координатной записи уравнения (3.13) имеют вид

$$\begin{aligned} |(x_2 - x_1)m_x + (y_2 - y_1)m_y + (z_2 - z_1)m_z| &= d_m, \\ |(x_2 - x_1)l_x + (y_2 - y_1)l_y + (z_2 - z_1)l_z| &= d_l, \\ |(x_2 - x_1)n_x + (y_2 - y_1)n_y + (z_2 - z_1)n_z| &= d_n. \end{aligned}$$

Совмещение точек.

Можно установить вариационную связь, которая сливает две точки в одну. Эта связь совмещения точек. Она приравнивает координаты двух заданных точек и описывается векторным уравнением

$$\mathbf{p}_2 = \mathbf{p}_1. \quad (3.14)$$

Данная вариационная связь содержит три скалярных уравнения

$$\begin{aligned} x_2 - x_1 &= 0, \\ y_2 - y_1 &= 0, \\ z_2 - z_1 &= 0. \end{aligned} \quad (3.15)$$

Симметрия точек относительно плоскости.

Рассмотрим вариационную связь симметрии точек \mathbf{p}_1 , \mathbf{p}_2 относительно плоскости

$$\mathbf{r}(x, y) = \mathbf{p} + x\mathbf{i}_1 + y\mathbf{i}_2. \quad (3.18)$$

Плоскость будем считать неподвижной, поэтому ее параметры варьироваться не будут.

Симметрия точек относительно плоскости описывается тремя уравнениями

$$\begin{aligned} (\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{i}_1 &= (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{i}_1, \\ (\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{i}_2 &= (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{i}_2, \\ (\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{i}_3 + (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{i}_3 &= 0, \end{aligned} \quad (3.19)-(3.21)$$

где $\mathbf{i}_3 = \mathbf{i}_1 \times \mathbf{i}_2$. Три уравнения (3.19)-(3.21) связывают шесть параметров. Симметрия точек показана на рис. 3.2.

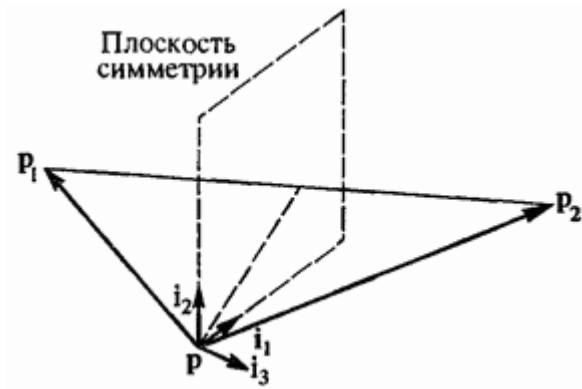


Рис. 3.2. Симметрия точек относительно плоскости

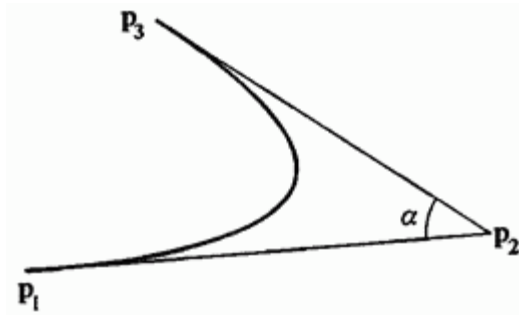


Рис. 3.3. Угловой размер между тремя точками NURBS кривой

Угловой размер.

Для трех точек можно установить угловую зависимость. Пусть имеется три точки $\mathbf{p}_1 = [x_1 \ y_1 \ z_1]^T$, $\mathbf{p}_2 = [x_2 \ y_2 \ z_2]^T$, $\mathbf{p}_3 = [x_3 \ y_3 \ z_3]^T$. На три точки можно наложить вариационную связь, определяющую угловой размер между отрезками $\mathbf{p}_1 - \mathbf{p}_2$, $\mathbf{p}_3 - \mathbf{p}_2$. На рис. 3.3 точками, на которые наложена вариационная связь, являются характеристические точки рациональной кривой. В общем случае это могут быть любые точки из структуры данных геометрических объектов.

Уравнение, описывающее угловой размер между тремя точками, имеет вид

$$\operatorname{arctg} \frac{s}{c} = \alpha, \quad (3.22)$$

где

$$\begin{aligned} s &= |(\mathbf{p}_1 - \mathbf{p}_2) \times (\mathbf{p}_3 - \mathbf{p}_2)|, \\ c &= (\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_2) = \\ &= (x_1 - x_2)(x_3 - x_2) + (y_1 - y_2)(y_3 - y_2) + (z_1 - z_2)(z_3 - z_2), \end{aligned} \quad (3.23)$$

α — заданный угол.

Величина s равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2, \mathbf{p}_3 - \mathbf{p}_2$ на $\sin \alpha$.

Величина c равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2, \mathbf{p}_3 - \mathbf{p}_2$ на $\cos \alpha$.

Квадрат s определяется формулой

$$\begin{aligned} s^2 &= ((y_1 - y_2)(z_3 - z_2) - (z_1 - z_2)(y_3 - y_2))^2 + \\ &+ ((z_1 - z_2)(x_3 - x_2) - (x_1 - x_2)(z_3 - z_2))^2 + \\ &+ ((x_1 - x_2)(y_3 - y_2) - (y_1 - y_2)(x_3 - x_2))^2. \end{aligned} \quad (3.24)$$

Уравнение (3.22) связывает девять параметров.

Другой вариационной связью трех точек $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ является угловой размер в плоскости. Эта связь устанавливает угловой размер между проекциями векторов $\mathbf{p}_1 - \mathbf{p}_2, \mathbf{p}_3 - \mathbf{p}_2$ на плоскость, ортогональную единичному вектору \mathbf{m} , и описывается уравнением вида (3.22), где

$$\begin{aligned} s &= \pm |\mathbf{q}_1 \times \mathbf{q}_3|, \\ c &= \mathbf{q}_1 \cdot \mathbf{q}_3, \\ \mathbf{q}_1 &= \mathbf{p}_1 - \mathbf{p}_2 - \mathbf{m}((\mathbf{p}_1 - \mathbf{p}_2) \cdot \mathbf{m}), \\ \mathbf{q}_2 &= \mathbf{p}_3 - \mathbf{p}_2 - \mathbf{m}((\mathbf{p}_3 - \mathbf{p}_2) \cdot \mathbf{m}). \end{aligned} \quad (3.25)-(3.27)$$

Векторы $\mathbf{q}_1, \mathbf{q}_3$ являются составляющими векторов $\mathbf{p}_1 - \mathbf{p}_2, \mathbf{p}_3 - \mathbf{p}_2$ ортогональными единичному вектору \mathbf{m} . Знак плюс в (3.25) выбирается в случае совпадения направления векторов $\mathbf{q}_1 \times \mathbf{q}_3$, знак минус — в противном случае.

Угол между векторами.

Пусть имеется четыре точки $\mathbf{p}_1 = [x_1 \ y_1 \ z_1]^T, \mathbf{p}_2 = [x_2 \ y_2 \ z_2]^T, \mathbf{p}_3 = [x_3 \ y_3 \ z_3]^T, \mathbf{p}_4 = [x_4 \ y_4 \ z_4]^T$. По четырем точкам можно построить два вектора. Построим векторы на точках $\mathbf{p}_1, \mathbf{p}_2$ и $\mathbf{p}_3, \mathbf{p}_4$. Будем считать, что эти векторы не коллинеарны. Два вектора определяют семейство плоскостей, параллельных одновременно им обоим. Каждая плоскость семейства будет ортогональна вектору $\mathbf{n} = (\mathbf{p}_1 - \mathbf{p}_2) \times (\mathbf{p}_3 - \mathbf{p}_4)$,

На четыре точки можно наложить вариационную связь, определяющую угол между векторами $\mathbf{p}_1 - \mathbf{p}_2, \mathbf{p}_3 - \mathbf{p}_4$. Эта связь аналогична угловому размеру (3.22). Уравнение, описывающее угол между четырьмя точками, имеет тот же вид, что и уравнение (3.22):

$$\arctg \frac{s}{c} = \alpha, \quad (3.28)$$

где

$$\begin{aligned} s &= |(\mathbf{p}_1 - \mathbf{p}_2) \times (\mathbf{p}_3 - \mathbf{p}_4)|, \\ c &= (\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_4) = \\ &= (x_1 - x_2)(x_3 - x_4) + (y_1 - y_2)(y_3 - y_4) + (z_1 - z_2)(z_3 - z_4). \end{aligned} \quad (3.29)-(3.30)$$

α — заданный угол. Величина s равна произведению длин векторов

$\mathbf{p}_1 - \mathbf{p}_2, \mathbf{p}_3 - \mathbf{p}_4$ на $\sin \alpha$. Величина c равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2, \mathbf{p}_3 - \mathbf{p}_4$ на $\cos \alpha$.

Квадрат s определяется формулой

$$\begin{aligned} s^2 &= ((y_1 - y_2)(z_3 - z_4) - (z_1 - z_2)(y_3 - y_4))^2 + \\ &+ ((z_1 - z_2)(x_3 - x_4) - (x_1 - x_2)(z_3 - z_4))^2 + \\ &+ ((x_1 - x_2)(y_3 - y_4) - (y_1 - y_2)(x_3 - x_4))^2. \end{aligned}$$

Уравнение (3.28) связывает двенадцать параметров.

Потребуем, чтобы при изменении угла между векторами точки перемещались ортогонально вектору \mathbf{n} . Для нахождения точек скрещения построим две линии

$$\begin{aligned}\mathbf{r}_1(t_1) &= (1 - t_1)\mathbf{p}_1 + t_1\mathbf{p}_2, \\ \mathbf{r}_3(t_3) &= (1 - t_3)\mathbf{p}_3 + t_3\mathbf{p}_4.\end{aligned}\quad (3.31)-(3.32)$$

Точками скрещения пространственных линий являются точки \mathbf{r}_1 и \mathbf{r}_3 на линиях $\mathbf{r}_1(t_1)$ и $\mathbf{r}_3(t_3)$, для которых выполняются равенства

$$\begin{aligned}(\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{r}_1 - \mathbf{r}_3) &= 0, \\ (\mathbf{p}_3 - \mathbf{p}_4) \cdot (\mathbf{r}_1 - \mathbf{r}_3) &= 0.\end{aligned}\quad (3.33)-(3.34)$$

Вектор, построенный по точкам скрещения, ортогонален обоим прямым. Параметры t_1 и t_3 прямых, соответствующие точкам скрещения, определим из системы уравнений

$$\begin{aligned}(\mathbf{p}_1 - \mathbf{p}_2) \cdot ((1 - t_1)\mathbf{p}_1 + t_1\mathbf{p}_2 - (1 - t_3)\mathbf{p}_3 - t_3\mathbf{p}_4) &= 0, \\ (\mathbf{p}_3 - \mathbf{p}_4) \cdot ((1 - t_1)\mathbf{p}_1 + t_1\mathbf{p}_2 - (1 - t_3)\mathbf{p}_3 - t_3\mathbf{p}_4) &= 0\end{aligned}\quad (3.35) - (3.36)$$

или

$$\begin{aligned}(\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_2 - \mathbf{p}_1)t_1 + (\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_4)t_3 &= (\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_1), \\ (\mathbf{p}_3 - \mathbf{p}_4) \cdot (\mathbf{p}_2 - \mathbf{p}_1)t_1 + (\mathbf{p}_3 - \mathbf{p}_4) \cdot (\mathbf{p}_3 - \mathbf{p}_4)t_3 &= (\mathbf{p}_3 - \mathbf{p}_4) \cdot (\mathbf{p}_3 - \mathbf{p}_1).\end{aligned}$$

По найденным параметрам найдем точки скрещения.

Ортогональность векторов.

Частным случаем вариационной связи (3.28) является ортогональность векторов, построенных по точкам \mathbf{p}_1 , \mathbf{p}_2 и \mathbf{p}_3 , \mathbf{p}_4 . Данная вариационная связь определяется уравнением

$$(\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_4) = 0. \quad (3.37)$$

Если на точках построены отрезки прямых (3.31) и (3.32), то уравнение (3.37) делает эти отрезки ортогональными.

Параллельность векторов.

Вариационная связь, делающая векторы $\mathbf{p}_1 - \mathbf{p}_2$, $\mathbf{p}_3 - \mathbf{p}_4$ параллельными, определяется векторным уравнением

$$(\mathbf{p}_1 - \mathbf{p}_2) \times (\mathbf{p}_3 - \mathbf{p}_4) = \mathbf{0}. \quad (3.38)$$

Связь (3.38) будем называть параллельностью векторов. Уравнение (3.38) содержит три скалярных уравнения

$$(y_1 - y_2)(z_3 - z_4) = 0, \quad (z_1 - z_2)(x_3 - x_4) = 0, \quad (x_1 - x_2)(y_3 - y_4) = 0.$$

5.4. Вариационные связи точек на кривых и поверхностях

Выше мы рассмотрели связи, накладываемые на существующие в структурах данных геометрических объектов параметры. Если требуется наложить вариационные связи на некоторые значения, которые отсутствуют в структуре данных геометрических объектов, то можно ввести в структуру данных дополнительные параметры. Например, для описания дуги окружности в структуре данных не нужны крайние ее точки, но их можно туда ввести, если необходимо накладывать на них вариационные связи. При этом остальные параметры должны пересчитываться в зависимости от положения крайних точек.

Вариационные связи можно накладывать на точки кривых и поверхностей, которые отсутствуют в их структурах данных. Например, мы хотим задать линейный размер между определенной точкой кривой $\mathbf{a}(t)$ и определенной точкой кривой $\mathbf{b}(w)$. Пусть для простоты кривыми являются отрезки

$$\mathbf{a}(t) = (1 - t)\mathbf{a}_1 + t\mathbf{a}_2,$$

$$\mathbf{b}(w) = (1 - w)\mathbf{b}_1 + w\mathbf{b}_2,$$

а в качестве точек на них мы возьмем середины отрезков. Середины отрезков определяются параметрами $t=t_0=1/2$, $w=w_0=1/2$ Будем считать, что t_0 , w_0 есть постоянные величины, если не будет объявлено, что они также являются варьируемыми параметрами уравнения. Линейный размер описывается уравнением (3.1), в котором теперь $\mathbf{p}_1=\mathbf{a}(t_0)$, $\mathbf{p}_2=\mathbf{b}(w_0)$. Координаты связываемых точек $\mathbf{p}_1=[x_1 \ y_1 \ z_1]^T$, $\mathbf{p}_2=[x_2 \ y_2 \ z_2]^T$ являются функциями координат точек кривых $\mathbf{a}_1=[x_{a1} \ y_{a1} \ z_{a1}]^T$, $\mathbf{a}_2=[x_{a2} \ y_{a2} \ z_{a2}]^T$, $\mathbf{b}_1=[x_{b1} \ y_{b1} \ z_{b1}]^T$, $\mathbf{b}_2=[x_{b2} \ y_{b2} \ z_{b2}]^T$. Число варьируемых параметров в данном случае будет равно 12. Ими являются координаты $x_{a1} \ y_{a1} \ z_{a1}$, $x_{a2} \ y_{a2} \ z_{a2}$, $x_{b1} \ y_{b1} \ z_{b1}$, $x_{b2} \ y_{b2} \ z_{b2}$. В общем случае варьируемыми параметрами будут являться параметры из структур данных связываемых кривых, которыми описываются точки $\mathbf{p}_1=\mathbf{a}(t_0)$, $\mathbf{p}_2=\mathbf{b}(w_0)$.

Аналогично можно установить размер по одной из координат (3.10) или размер вдоль заданного направления (3.12) между определенной точкой кривой $\mathbf{a}(t)$ и определенной точкой кривой $\mathbf{b}(w)$. Можно связать определенные точки кривых вариационной зависимостью (3.14). Можно установить угловой размер (3.28) между заданными точками трех геометрических объектов.

Все сказанное остается в силе, если вместо точек кривых связать определенные точки поверхностей. Например, описанным образом можно установить линейный размер между определенной точкой поверхности $\mathbf{a}(u, v)$ и определенной точкой поверхности $\mathbf{b}(t, w)$ или установить угловой размер между определенными точками поверхностей $\mathbf{a}(u, v)$, $\mathbf{b}(t, w)$ и $\mathbf{s}(x, y)$.

Между точкой \mathbf{p} и произвольным геометрическим объектом \mathbf{S} можно установить вариационную связь, «усаживающую» точку на геометрический объект. Эта связь описывается уравнением

$$\mathbf{p} = \mathbf{s}_p, \tag{4.1}$$

где s_p — проекция точки на геометрический объект. Векторное уравнение (4.1) содержит три скалярных уравнения относительно трех параметров — координат точки \mathbf{p} . Геометрический объект в процессе решения уравнений связей можно считать или неподвижным, или подвижным, но единым целым.

Если требуется, чтобы геометрический объект после установления вариационных связей не изменял свою форму, то его параметры нужно считать неизменными или к уравнениям связи следует добавить уравнения, сохраняющие неизменными параметры, определяющие его геометрию.

5.5. Алгебраические связи

В уравнения связей (3.1), (3.12), (3.22), (3.28) входят величины d и α , которые до сих пор читались константами. На рис. 1.1 размеры связаны друг с другом алгебраическими уравнениями. Для того, чтобы величины d и α , могли изменяться, они должны также являться варьируемыми параметрами. Эти параметры не принадлежат ни какому-либо геометрическому объекту, ни какой-либо вариационной связи. Будем называть их свободными параметрами.

Алгебраические уравнения, связывающие свободные параметры, будем называть алгебраическими связями. В частном случае алгебраические связи используются, чтобы сделать равными несколько параметров различных геометрических объектов. Для этого вводится свободный параметр и совокупность уравнений, приравнивающих требуемые параметры геометрических объектов введенному свободному параметру.

В качестве примера использования алгебраических связей рассмотрим два линейных размера, один из которых связывает точки $\mathbf{p}_1=[x_1 \ y_1 \ z_1]^T$, $\mathbf{p}_2=[x_2 \ y_2 \ z_2]^T$, а другой связывает точки $\mathbf{p}_3=[x_3 \ y_3 \ z_3]^T$, $\mathbf{p}_4=[x_4 \ y_4 \ z_4]^T$. Пусть линейные размеры описываются уравнениями

$$\begin{aligned} |\mathbf{p}_2 - \mathbf{p}_1| &= a, \\ |\mathbf{p}_3 - \mathbf{p}_4| &= b. \end{aligned} \quad (5.1) - (5.2)$$

Пусть значения размеров a и b связаны некоторыми алгебраическими уравнениями

$$\begin{aligned}g(a, b) &= 0, \\h(a, b) &= 0.\end{aligned}\tag{5.3)-(5.4}$$

Уравнения (5.1) и (5.2) отличаются от уравнения (3.1) тем, что в последнем размер d не варьируется, а в уравнениях (5.1) и (5.2) параметры a и b являются варьируемыми. На $(r+1)$ -й итерации решения системы уравнений изменяются как координаты точек, так и значения размеров:

$$\begin{aligned}x_i^{(r+1)} &= x_i^{(r)} + \Delta x_i, & y_i^{(r+1)} &= y_i^{(r)} + \Delta y_i, & z_i^{(r+1)} &= z_i^{(r)} + \Delta z_i, \\a^{(r+1)} &= a^{(r)} + \Delta a, & b^{(r+1)} &= b^{(r)} + \Delta b, \\i &= 1, 2, 3, 4.\end{aligned}$$

Свободными могут быть любые параметры уравнений связей, не являющиеся параметрами геометрических объектов. В общем случае алгебраические уравнения нельзя выделить из общей системы уравнений связей и решить отдельно, так как они могут быть связаны друг с другом через уравнения для параметров геометрических объектов.

Наряду с алгебраическими уравнениями вариационные связи могут строиться на неравенствах. Например, вместо уравнения (5.3) алгебраическая связь может определяться неравенством

$$a > b.\tag{5.5}$$

На каждой итерации решения системы уравнений неравенство заменяется соответствующим уравнением. Неравенство (5.5) на каждой итерации должно быть заменено уравнением

$$a - b = \epsilon,\tag{5.6}$$

где

$$\begin{aligned} \varepsilon &= a^{(r)} - b^{(r)}, & \text{если } a^{(r)} > b^{(r)}, \\ \varepsilon &= 0, & \text{если } a^{(r)} \leq b^{(r)}. \end{aligned}$$

Если на очередной итерации неравенство выполняется, то оно заменяется тождественным равенством, если неравенство не выполняется, то оно заменяется строгим равенством. Аналогичным образом в процессе численного решения системы уравнений и другие неравенства заменяются соответствующими равенствами. Связи, построенные на неравенствах, будем называть нестрогими алгебраическими связями.

5.6. Минимизация изменения параметров

Идеальным является случай, когда число уравнений связей равно числу варьируемых параметров. Этот идеальный случай реализуется, когда на заданную группу параметров наложено максимально возможное число связей. Если делается попытка наложить связей больше, чем это возможно, то система уравнений будет переопределена и отвергнет лишние связи. Если число уравнений связей меньше числа параметров, то необходимо найти способ дополнить систему связей уравнениями. Один из способов — составление дополнительных уравнений.

На примере линейного размера между двумя точками показано использование дополнительных уравнений. Возможны ситуации, когда одна и та же группа параметров участвует в нескольких связях (например, на одну и ту же группу точек наложено несколько размерных зависимостей). Набор дополнительных уравнений может быть различным и существенным образом влияет на поведение геометрических объектов. Использование дополнительных уравнений нарушает равноправие варьируемых параметров и может привести к нежелательным последствиям в поведении геометрических объектов.

Мы откажемся от использования дополнительных уравнений. Необходимую систему уравнений для удовлетворения всех вариационных связей мы получим, наложив некоторое общее для всей совокупности геометрических объектов условие поведения. Поведение системы параметров мы будем определять некоторым критерием, описываемым функцией или функционалом. Уравнения для

определения параметров мы получим из требования минимума или максимума функции или функционала данного критерия.

При наложении вариационных связей будем использовать метод минимизации суммы квадратов изменений параметров. Рассмотрим его применение на примере линейного размера между двумя точками, описываемого уравнением (3.1). Если ввести дополнительные уравнения (3.2)-(3.6), то поведение связанных точек будет симметричным, но сама система уравнений имеет несимметричный вид. Кроме того, если рассматриваемые точки участвуют еще в каких-нибудь связях, то возникает вопрос, какие из дополнительных уравнений следует оставить, а какие опустить. В большинстве случаев поведение такой системы связей теряет симметрию.

Обратим внимание на то, что в варианте дополнительных уравнений (3.2)-(3.6) суммарное перемещение связанных линейным размером точек является минимальным из возможных. Используем это свойство для составления системы уравнений. Квадрат суммарного перемещения связанных точек описывается функцией

$$\psi = \frac{1}{2} \left(|\mathbf{p}_1 - \mathbf{p}_1^{(0)}|^2 + |\mathbf{p}_2 - \mathbf{p}_2^{(0)}|^2 \right).$$

В координатном представлении эта функция имеет вид

$$\psi = \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (z_1 - z_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 + (z_2 - z_2^{(0)})^2 \right). \quad (6.1)$$

Эта функция пропорциональна сумме квадратов изменений всех связанных параметров. Аргументами этой функции являются те же шесть параметров: $x_1, y_1, z_1, x_2, y_2, z_2$. Потребуем, чтобы сумма квадратов изменений параметров была минимальной. При этом необходимо, чтобы координаты точек удовлетворяли уравнению связи (3.1). Используем метод неопределенных множителей Лагранжа для отыскания минимума функции (6.1) при условии (3.1). Необходимым условием минимума функции (6.1) при условии (3.1) является равенство нулю частных производных по параметрам функции

$$\begin{aligned}
 F = \psi + \lambda f &= \frac{1}{2} \left(|\mathbf{p}_1 - \mathbf{p}_1^{(0)}|^2 + |\mathbf{p}_2 - \mathbf{p}_2^{(0)}|^2 \right) + \lambda (|\mathbf{p}_2 - \mathbf{p}_1| - d) = \\
 &= \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (z_1 - z_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 + \right. \\
 &\quad \left. + (z_2 - z_2^{(0)})^2 \right) + \lambda \left(\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} - d \right), \quad (7.6.2) \quad (6.2)
 \end{aligned}$$

где $f = |\mathbf{p}_2 - \mathbf{p}_1| - d$, λ — подлежащий определению множитель. Искомые координаты точек и множитель λ найдем из системы уравнений

$$\begin{aligned}
 &\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} - d = 0, \\
 &x_1 - x_1^{(0)} + \lambda \frac{x_1 - x_2}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0, \\
 &y_1 - y_1^{(0)} + \lambda \frac{y_1 - y_2}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0, \\
 &z_1 - z_1^{(0)} + \lambda \frac{z_1 - z_2}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0, \\
 &x_2 - x_2^{(0)} + \lambda \frac{x_2 - x_1}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0, \\
 &y_2 - y_2^{(0)} + \lambda \frac{y_2 - y_1}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0, \\
 &z_2 - z_2^{(0)} + \lambda \frac{z_2 - z_1}{|\mathbf{p}_1 - \mathbf{p}_2|} = 0.
 \end{aligned} \tag{6.3}$$

Данная система уравнений имеет решение

$$\begin{aligned}
 x_1 &= x_1^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(x_1^{(0)} - x_2^{(0)}), \\
 y_1 &= y_1^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(y_1^{(0)} - y_2^{(0)}), \\
 z_1 &= z_1^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(z_1^{(0)} - z_2^{(0)}), \\
 x_2 &= x_2^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(x_2^{(0)} - x_1^{(0)}), \\
 y_2 &= y_2^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(y_2^{(0)} - y_1^{(0)}), \\
 z_2 &= z_2^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(z_2^{(0)} - z_1^{(0)}), \\
 \lambda &= -\frac{d - d^{(0)}}{2},
 \end{aligned} \tag{6.4}$$

где

$$d^{(0)} = \sqrt{(x_2^{(0)} - x_1^{(0)})^2 + (y_2^{(0)} - y_1^{(0)})^2 + (z_2^{(0)} - z_1^{(0)})^2}$$

Можно доказать, что найденное решение является точкой минимума функции (6.2). Точки переместятся вдоль прямой, проходящей через точки $\mathbf{p}_1^{(0)}$, $\mathbf{p}_2^{(0)}$ на одинаковое расстояние. В данной постановке задачи точки являются равноправными, а уравнения и решение системы уравнений являются симметричными.

Мы считали, что обе точки могут перемещаться. Если одна из точек закреплена, то на нее наложена связь (2.2) и ее координаты не варьируются. Пусть точка \mathbf{p}_1 закреплена. Тогда функция (6.2) будет иметь вид

$$\begin{aligned}
 F &= \frac{1}{2} |\mathbf{p}_2 - \mathbf{p}_2^{(0)}|^2 + \lambda (|\mathbf{p}_2 - \mathbf{p}_1| - d) = \\
 &= \frac{1}{2} \left((x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 + (z_2 - z_2^{(0)})^2 \right) + \\
 &\quad + \lambda \left(\sqrt{(x_2 - x_1^{(0)})^2 + (y_2 - y_1^{(0)})^2 + (z_2 - z_1^{(0)})^2} - d \right). \quad (6.5)
 \end{aligned}$$

Координаты закрепленной точки в функцию (6.5) не входят. Функция (6.5) достигает минимума при

$$\begin{aligned}
 x_2 &= x_2^{(0)} + \frac{d - d^{(0)}}{d^{(0)}} (x_2^{(0)} - x_1^{(0)}), \\
 y_2 &= y_2^{(0)} + \frac{d - d^{(0)}}{d^{(0)}} (y_2^{(0)} - y_1^{(0)}), \\
 z_2 &= z_2^{(0)} + \frac{d - d^{(0)}}{d^{(0)}} (z_2^{(0)} - z_1^{(0)}), \\
 \lambda &= -(d - d^{(0)}). \quad (6.6)
 \end{aligned}$$

И в решении (6.4), и в решении (6.6) свободные точки перемещаются вдоль прямой, проходящей через точки \mathbf{p}_1 , \mathbf{p}_2 . Требование минимума суммы квадратов варьируемых параметров автоматически создает симметрию в поведении системы. Симметрия присутствует и в системе уравнений. Это свойство особенно ценно для сложных связей, в которых много параметров и все они имеют различный геометрический смысл. На данном примере проиллюстрирован критерий, который мы будем использовать для удовлетворения произвольных вариационных связей.

Сумма квадратов изменений варьируемых параметров отражает реакцию геометрических объектов с наложенными вариационными связями, на изменение констант в уравнениях связей или на изменения самих объектов. Критерий изменения параметров сформулируем следующим образом. Реакция геометрических объектов с наложенными на них вариационными связями на любое возмущение должна быть такой, чтобы сумма квадратов изменений варьируемых

где $x_i^{(0)}$ — исходные значения варьируемых параметров.

На примере линейного размера выше было показано применение этого критерия совместно с методом неопределенных множителей Лагранжа.

Рассмотрим теоретические основы использованного выше метода неопределенных множителей Лагранжа для нахождения условного минимума или максимума действительной функции нескольких переменных параметров.

Экстремум функции.

Пусть некоторое требование, предъявленное к варьируемым параметрам x_i , $i=1,2,\dots,n$, определяется экстремумом функции

$$\psi = \psi(x_1, x_2, \dots, x_n). \quad (7.2)$$

По определению действительная функция (7.2) в точке $x_i=a_i$, $i=1,2,\dots,n$ (конкретную совокупность параметров будем называть точкой) имеет минимум (максимум), если существует такое положительное число S , что при любых $\Delta x_1, \Delta x_2, \dots, \Delta x_n$, удовлетворяющих неравенствам

$$0 < \sqrt{\Delta x_1^2 + \Delta x_2^2 + \dots + \Delta x_n^2} < \delta,$$

приращение функции

$$\Delta\psi = \psi(a_1 + \Delta x_1, a_2 + \Delta x_2, \dots, a_n + \Delta x_n) - \psi(a_1, a_2, \dots, a_n) \quad (7.3)$$

соответственно больше (меньше) нуля. Будем считать, что функция ψ в окрестности экстремума имеет непрерывные частные производные до второго порядка включительно. Разложим в ряд Тейлора в окрестности точки $x_i=a_i$, $i=1,2,\dots,n$ первое слагаемое правой части (7.3) и получим

$$\Delta\psi = \sum_{i=1}^n \frac{\partial\psi}{\partial x_i} \Delta x_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2\psi}{\partial x_i \partial x_j} \Delta x_i \Delta x_j + o(\Delta x^3), \quad (7.4)$$

где $o(\Delta x^3)$ — слагаемые третьего порядка относительно Δx_i . Слагаемые первого и второго порядка относительно Δx_i в (7.4) составляют первый $d\psi$ и второй $d^2\psi$ дифференциалы функции ψ .

Дифференцируемая функция $\psi(x_1, x_2, \dots, x_n)$ в точке $x_i = a_i$, $i = 1, 2, \dots, n$ имеет минимум или максимум лишь в том случае, когда ее первый дифференциал обращается в этой точке в нуль

$$d\psi = \frac{\partial\psi}{\partial x_1} dx_1 + \frac{\partial\psi}{\partial x_2} dx_2 + \dots + \frac{\partial\psi}{\partial x_n} dx_n = 0. \quad (7.5)$$

Для этого необходимо выполнение равенств

$$\frac{\partial\psi}{\partial x_1} = 0, \quad \frac{\partial\psi}{\partial x_2} = 0, \quad \dots, \quad \frac{\partial\psi}{\partial x_n} = 0, \quad (7.6)$$

при $x_1 = a_1$, $x_2 = a_2, \dots, x_n = a_n$. Выполнение равенств (7.6) является необходимым условием экстремума. Достаточным условием экстремума функции $\psi(x_1, x_2, \dots, x_n)$ является знакоопределенность второго дифференциала

$$d^2\psi = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2\psi}{\partial x_i \partial x_j} dx_i dx_j, \quad (7.7)$$

при $x_1 = a_1$, $x_2 = a_2, \dots, x_n = a_n$. Если второй дифференциал (7.7) представляет собой положительно определенную квадратичную форму, то функция в рассматриваемой точке имеет локальный минимум. Если второй дифференциал (7.7) есть отрицательно определенная квадратичная форма, то функция в рассматриваемой точке имеет локальный максимум.

Если второй дифференциал (7.7) не является знакоопределенной квадратичной формой, то функция в рассматриваемой точке не достигает экстремума.

относительно множителей $\lambda_1, \lambda_2, \dots, \lambda_m$. Как уже было сказано, определитель матрицы этой системы линейных алгебраических уравнений должен быть отличен от нуля. При выполнении равенств (7.15) для выполнения равенства (7.13) требуется, чтобы

$$\frac{\partial F}{\partial x_{m+1}} = 0, \quad \frac{\partial F}{\partial x_{m+2}} = 0, \quad \dots, \quad \frac{\partial F}{\partial x_n} = 0. \quad (7.17)$$

Объединив равенства (7.15), (7.17) и уравнения связи (7.1), приходим к тому, что в точке условного экстремума функции (7.2) должны выполняться $n+m$ равенств

$$\begin{aligned} \frac{\partial \psi}{\partial x_1} + \lambda_1 \frac{\partial f_1}{\partial x_1} + \lambda_2 \frac{\partial f_2}{\partial x_1} + \dots + \lambda_m \frac{\partial f_m}{\partial x_1} &= 0, \\ \frac{\partial \psi}{\partial x_2} + \lambda_1 \frac{\partial f_1}{\partial x_2} + \lambda_2 \frac{\partial f_2}{\partial x_2} + \dots + \lambda_m \frac{\partial f_m}{\partial x_2} &= 0, \\ \dots & \\ \frac{\partial \psi}{\partial x_n} + \lambda_1 \frac{\partial f_1}{\partial x_n} + \lambda_2 \frac{\partial f_2}{\partial x_n} + \dots + \lambda_m \frac{\partial f_m}{\partial x_n} &= 0, \\ f_1(x_1, x_2, \dots, x_n) &= 0, \\ f_2(x_1, x_2, \dots, x_n) &= 0, \\ \dots & \\ f_m(x_1, x_2, \dots, x_n) &= 0. \end{aligned} \quad (7.18)$$

Эта система уравнений позволяет определить параметры условного экстремума функции ψ и множители Лагранжа. Мы видим, что поиск условного экстремума функции ψ привел к уравнениям безусловного экстремума функции Лагранжа F . Практически метод Лагранжа реализуется следующим образом. Составляется функция Лагранжа (7.14) и для нее находят возможные точки безусловного экстремума. Для исключения множителей $\lambda_1, \lambda_2, \dots, \lambda_m$ привлекаются уравнения связей.

Найденное из системы уравнений (7.18) решение может не являться точкой экстремума. В точке условного минимума (условного максимума) функции ψ должны выполняться достаточные условия. При наличии связей экстремумы функций ψ и F совпадают. Тогда из вышеизложенного следует, что для получения достаточного условия условного экстремума функции ψ к уравнениям (7.18) нужно добавить требование знакоопределенности d^2F :

$$d^2F = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 F}{\partial x_i \partial x_j} dx_i dx_j. \quad (7.19)$$

В функции Лагранжа все параметры x_1, x_2, \dots, x_n можно рассматривать как независимые, если считать, что множители $\lambda_1, \lambda_2, \dots, \lambda_m$ определены из уравнений (7.16).

В методе Лагранжа все параметры равноправны, а система уравнений (7.18) является симметричной относительно параметров. Кроме того, все необходимые уравнения получены из единого требования к поведению параметров. Платой за эту симметрию является увеличение числа неизвестных из-за введения неопределенных множителей.

Функция $\psi(x_1, x_2, \dots, x_n)$ формулирует критерий поведения варьируемых параметров x_1, x_2, \dots, x_n при их переходе из начального состояния в конечное состояние, если изменены константы уравнений связей или сами уравнения. Принятый нами критерий в форме минимума функции

$$\psi(x_1, x_2, \dots, x_n) = \frac{1}{2} \sum_{i=1}^n (x_i - x_i^{(0)})^2, \quad (7.20)$$

где $x_i^{(0)}$ — начальные значения варьируемых параметров, предъявляет требование наименьшего суммарного изменения параметров, участвующих в связях. Конечные значения параметров x_1, x_2, \dots, x_n должны быть найдены из системы уравнений (7.18).

Система уравнений (7.18) при использовании критерия (7.20) примет вид

$$\begin{aligned}
 x_1 - x_1^{(0)} + \lambda_1 \frac{\partial f_1}{\partial x_1} + \lambda_2 \frac{\partial f_2}{\partial x_1} + \dots + \lambda_m \frac{\partial f_m}{\partial x_1} &= 0, \\
 x_2 - x_2^{(0)} + \lambda_1 \frac{\partial f_1}{\partial x_2} + \lambda_2 \frac{\partial f_2}{\partial x_2} + \dots + \lambda_m \frac{\partial f_m}{\partial x_2} &= 0, \\
 \dots & \\
 x_n - x_n^{(0)} + \lambda_1 \frac{\partial f_1}{\partial x_n} + \lambda_2 \frac{\partial f_2}{\partial x_n} + \dots + \lambda_m \frac{\partial f_m}{\partial x_n} &= 0, \\
 f_1(x_1, x_2, \dots, x_n) &= 0, \\
 f_2(x_1, x_2, \dots, x_n) &= 0, \\
 \dots & \\
 f_m(x_1, x_2, \dots, x_n) &= 0.
 \end{aligned}
 \tag{7.21}$$

Критерий (7.20) будем использовать при наложении вариационных связей на координаты точек и на любые другие параметры геометрических объектов.

5.8. Вариационный метод определения изменений параметров

Требование минимума суммарного изменения параметров связей приводит к задаче поиска условного экстремума некоторой функции. Покажем, что в более общей постановке эта задача является вариационной и сводится к поиску экстремума некоторого функционала.

Функционал.

Рассмотрим точку, координаты которой участвуют в некоторых уравнениях вариационных связей. Представим, что мы изменили или сами вариационные уравнения, или некоторые их параметры. В результате координаты точки в общем случае должны измениться. Пусть координаты рассматриваемой точки в начальном положении равны $x^{(0)}, y^{(0)}, z^{(0)}$, а в конечном положении равны x, y, z . В декартовой системе координат квадрат расстояния между начальным и конечным положением точки равен

$$2\psi = (x - x^{(0)})^2 + (y - y^{(0)})^2 + (z - z^{(0)})^2.$$

Для минимизации перемещения точки в декартовой системе координат можно потребовать условного минимума функции ψ . В результате мы приходим к системе уравнений (7.21), из которой определим координаты рассматриваемой точки в конечном положении.

В криволинейной системе координат квадрат расстояния между бесконечно близкими положениями рассматриваемой точки определяется формулой

$$(ds)^2 = g_{11}(dx)^2 + g_{22}(dy)^2 + g_{33}(dz)^2 + 2g_{12} dx dy + 2g_{23} dy dz + 2g_{13} dx dz.$$

где $g^{ij}=g_{ij}(x, y, z)$ — ковариантные компоненты метрического тензора. Пусть каждая координата является функцией некоторого параметра t . Тогда функции $x(t)$ будут описывать кривую, вдоль которой осуществляется перемещение точки. Величина

$$\left(\frac{ds}{dt}\right)^2 = g_{11}\left(\frac{dx}{dt}\right)^2 + g_{22}\left(\frac{dy}{dt}\right)^2 + g_{33}\left(\frac{dz}{dt}\right)^2 + 2g_{12}\frac{dx}{dt}\frac{dy}{dt} + 2g_{23}\frac{dy}{dt}\frac{dz}{dt} + 2g_{13}\frac{dx}{dt}\frac{dz}{dt}$$

равна квадрату первой производной этой кривой.

Рассмотрим интеграл

$$\Phi = \frac{1}{2} \int_{t_{\min}}^{t_{\max}} \left(g_{11}\left(\frac{dx}{dt}\right)^2 + g_{22}\left(\frac{dy}{dt}\right)^2 + g_{33}\left(\frac{dz}{dt}\right)^2 + 2g_{12}\frac{dx}{dt}\frac{dy}{dt} + 2g_{23}\frac{dy}{dt}\frac{dz}{dt} + 2g_{13}\frac{dx}{dt}\frac{dz}{dt} \right) dt,$$

где t — параметр траектории перемещения точки. Для минимизации перемещения точки в криволинейной системе координат потребуем минимума интеграла Φ . В этом интеграле координаты точки считаются функциями некоторого параметра t . Для начального значения параметра координаты принимают начальные значения:

$x(t_{\min})=x^{(0)}, y(t_{\min})=y^{(0)}, z(t_{\min})=z^{(0)}$ Конечные значения координат

$x(t_{\max})=x, y(t_{\max})=y, z(t_{\max})=z$ являются искомыми. Они должны удовлетворять уравнениям вариационных связей. Значение интеграла Φ зависит от функций $x(t), y(t), z(t)$. Подобные интегралы называются функционалами. Таким образом, требование минимума суммы

квадратов изменения параметров в более общей постановке сводится к

поиску минимума некоторого функционала. Если варьируемыми точкам приписать некоторую массу а параметр t считать временем, то принцип минимума суммы квадратов изменения параметров будет аналогичен принципу наименьшего действия в форме Гамильтона для движения механической системы.

В декартовой прямоугольной системе координат интеграл Φ для одной точки определяется формулой

$$\Phi = \frac{1}{2} \int_{t_{\min}}^{t_{\max}} \left(\left(\frac{dx}{dt} \right)^2 + \left(\frac{dy}{dt} \right)^2 + \left(\frac{dz}{dt} \right)^2 \right) dt.$$

Функционал представляет собой переменную величину, значение которой зависит от выбора одной или нескольких варьируемых функций. Параметры связей в этом функционале выступают в роли варьируемых функций. Начальные значения этих функций известны, а конечные значения должны удовлетворять уравнениям связей. Переход параметров из начального состояния в конечное состояние должен минимизировать некоторый функционал, выступающий в роли критерия такого перехода. Например, требование минимума суммы квадратов перемещений точек (6.1) заменим общим требованием минимума функционала

$$\Phi = \frac{1}{2} \int_{t_{\min}}^{t_{\max}} \left(\left(\frac{dx_1}{dt} \right)^2 + \left(\frac{dy_1}{dt} \right)^2 + \left(\frac{dz_1}{dt} \right)^2 + \left(\frac{dx_2}{dt} \right)^2 + \left(\frac{dy_2}{dt} \right)^2 + \left(\frac{dz_2}{dt} \right)^2 \right) dt, \quad (8.1)$$

где $x_1, y_1, z_1, x_2, y_2, z_2$ — координаты точек $\mathbf{p}_1, \mathbf{p}_2$ соответственно. Функционал записан в декартовой прямоугольной системе координат. Координаты точек мы представили как функции параметра t , которому можно приписать смысл параметра траекторий перехода точек из исходного состояния в новое состояние при изменении констант уравнений связей. Начальное t_{\min} и конечное t_{\max} значения параметра t будем считать известными. Для определенности положим $t_{\min}=0, t_{\max}=1$. Отличие вариационной постановки (8.1) данной задачи от задачи в постановке (6.1) заключается в следующем. Функция (6.1) предполагает, что точки перемещаются по прямой линии, а в функционале (8.1) такое предположение отсутствует, и перемещение

точек может быть произвольным. Функция (6.1) не зависит от пути изменения параметров, тогда как функционал (8.1) зависит от траектории перехода параметров из начального состояния в конечное состояние. Начальное положение точек нам известно.

В своем конечном положении координаты точек должны удовлетворять уравнениям связи. Вариационная постановка дает больше свободы в выборе критерия перехода параметров из начального состояния в конечное состояние. Мы увидим, что в декартовой системе координат условный экстремум функции (6.1) и условный экстремум функционала (8.1) приводят к одной и той же системе уравнений. В криволинейной системе координат это не так. Рассмотрим метод вариаций в общем случае.

Вариация функционала.

Пусть некоторое требование, предъявленное к варьируемым функциям $x_1(t), x_2(t), \dots, x_n(t)$ определяется экстремумом функционала вида

$$\Phi = \int_0^1 h \left(t, x_1, x_2, \dots, x_n, \frac{dx_1}{dt}, \frac{dx_2}{dt}, \dots, \frac{dx_n}{dt} \right) dt. \quad (8.2)$$

Функцию $h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')$ будем считать дифференцируемой необходимое число раз по t , по каждой функции x_1, x_2, \dots, x_n и по каждой первой производной $x_1'(t), x_2'(t), \dots, x_n'(t)$ каждой функции по параметру t . В функционале x_1, x_2, \dots, x_n выступают в роли варьируемых функций некоторого параметра t . Параметром t может служить общий параметр траекторий, описывающих изменение функций x_1, x_2, \dots, x_n . Областью изменения параметра t нами выбран отрезок $0 \leq t \leq 1$. Значение функционала Φ зависит от функций $x_1(t), x_2(t), \dots, x_n(t)$. Предположим, что экстремум функционала достигается на дважды дифференцируемых функциях $x_1(t), x_2(t), \dots, x_n(t)$.

Возьмем совокупность близких к ним функций $x_i(t) + \delta x_i(t)$, $i = 1, 2, \dots, n$ и представим их в виде

$$x_i(t, \alpha) = x_i(t) + \alpha \delta x_i(t), \quad 0 \leq \alpha \leq 1, \quad i = 1, 2, \dots, n. \quad (8.3)$$

Величины $\delta x_i(t)$ называются вариациями соответствующих функций. Эти функции можно дифференцировать по параметру t :

$$\frac{d(\delta x_i(t))}{dt} = \delta x_i' \equiv \delta \left(\frac{dx_i}{dt} \right), \quad \frac{d^2(\delta x_i(t))}{dt^2} = \delta x_i'' \equiv \delta \left(\frac{d^2 x_i}{dt^2} \right), \quad \dots$$

$$\dots, \quad \frac{d^k(\delta x_i(t))}{dt^k} = \delta x_i^{(k)} \equiv \delta \left(\frac{d^k x_i}{dt^k} \right), \quad i = 1, 2, \dots, n.$$

Подставим (8.3) в (8.2) и получим новое значение функционала $\Phi(\alpha)$. По предположению эта функция достигает экстремума при $\alpha=0$. Необходимым условием экстремума функции $\Phi(\alpha)$ является равенство нулю ее первой производной $\Phi'(0)$. Величина $\Phi'(0)$ называется вариацией функционала и обозначается через $\delta\Phi$. Найдем производную функции $\Phi(\alpha)$ по α как производную сложной функции и положим в ней $\alpha=0$. В результате получим

$$\left. \frac{d\Phi(\alpha)}{d\alpha} \right|_{\alpha=0} =$$

$$= \int_0^1 (h_1 \delta x_1 + h_2 \delta x_2 + \dots + h_n \delta x_n + h_{1(1)} \delta x_1' + h_{2(1)} \delta x_2' + \dots + h_{n(1)} \delta x_n') dt, \quad (8.4)$$

где

$$\begin{aligned}
 h_1 &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial x_1}, \\
 h_2 &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial x_2}, \\
 &\dots \\
 h_n &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial x_n}, \\
 h_{1(1)} &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial (dx_1/dt)}, \\
 h_{2(1)} &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial (dx_2/dt)}, \\
 &\dots \\
 h_{n(1)} &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial (dx_n/dt)}.
 \end{aligned}$$

Таким образом, необходимым условием экстремума функционала является равенство нулю его вариации, что выражается формулой

$$\int_0^1 (h_1 \delta x_1 + h_2 \delta x_2 + \dots + h_n \delta x_n + h_{1(1)} \delta x_1' + h_{2(1)} \delta x_2' + \dots + h_{n(1)} \delta x_n') dt = 0. \quad (8.5)$$

Проинтегрируем по частям вторую половину слагаемых в (8.5)

$$\begin{aligned}
 \int_0^1 h_{i(1)} \delta x_i' dt &= \int_0^1 h_{i(1)} \frac{d(\delta x_i)}{dt} dt = \int_0^1 \left(\frac{d(h_{i(1)} \delta x_i)}{dt} - \frac{dh_{i(1)}}{dt} \delta x_i \right) dt = \\
 &= h_{i(1)} \delta x_i \Big|_0^1 - \int_0^1 \frac{dh_{i(1)}}{dt} \delta x_i dt, \quad i = 1, 2, \dots, n.
 \end{aligned}$$

В результате получим, что необходимое условие экстремума функционала (8.2) выражается равенством

$$\sum_{i=1}^n h_{i(1)} \delta x_i \Big|_0^1 + \int_0^1 \sum_{i=1}^n \left(\left(h_i - \frac{dh_{i(1)}}{dt} \right) \delta x_i \right) dt = 0. \quad (8.6)$$

Уравнения Эйлера. Если функции x_1, x_2, \dots, x_n при вариации не изменяют свои значения в начальных и конечных точках, т. е. если $\delta x_i(t_{\min}) \equiv \delta x_i(0) = 0$ и $\delta x_i(t_{\max}) \equiv \delta x_i(1) = 0$, то все слагаемые

$$\sum_{i=1}^n h_{i(1)} \delta x_i \Big|_0^1 = 0, \quad i = 1, 2, \dots, n. \quad (8.7)$$

Это означает, что функции $x_1(t), x_2(t), \dots, x_n(t)$ и все их вариации проходят через фиксированные граничные точки. При этих условиях для выполнения равенства (8.6) необходимо выполнение равенств

$$\begin{aligned} h_1 - \frac{dh_{1(1)}}{dt} &= 0, \\ h_2 - \frac{dh_{2(1)}}{dt} &= 0, \\ \dots & \\ h_n - \frac{dh_{n(1)}}{dt} &= 0, \end{aligned} \quad (8.8)$$

так как вариации δx_i произвольны и в общем случае не равны нулю. Дифференциальные уравнения (8.8) называются уравнениями Эйлера. После дифференцирования по параметру t как сложной функции каждое из равенств (8.8) примет вид

$$h_i - \frac{\partial h_{i(1)}}{\partial t} - \sum_{j=1}^n \frac{\partial h_{i(1)}}{\partial x_j} \frac{dx_j}{dt} - \sum_{j=1}^n \frac{\partial h_{i(1)}}{\partial x_j'} \frac{d^2 x_j}{dt^2} = 0, \quad i = 1, 2, \dots, n,$$

или

$$\frac{\partial h}{\partial x_i} - \frac{\partial^2 h}{\partial x_i' \partial t} - \sum_{j=1}^n \frac{\partial^2 h}{\partial x_j \partial x_i'} \frac{dx_j}{dt} - \sum_{j=1}^n \frac{\partial^2 h}{\partial x_j' \partial x_i'} \frac{d^2 x_j}{dt^2} = 0,$$

$$i = 1, 2, \dots, n. \quad (8.9)$$

Проинтегрировав систему (8.8), получим общее решение для экстремалей, каждая из которых содержит по две константы. Константы могут быть найдены из граничных условий $\delta x_i|_0=0$. Заметим, что система (8.8) не всегда имеет решение, а если решение существует, то оно может быть не единственным или не удовлетворять достаточным условиям экстремума. Для функционала (8.1) уравнения (8.9) примут более простой вид:

$$\frac{d^2 x_i}{dt^2} = 0, \quad \frac{d^2 y_i}{dt^2} = 0, \quad \frac{d^2 z_i}{dt^2} = 0, \quad i = 1, 2. \quad (8.10)$$

Решением системы (8.10) являются линейные функции $x_i(t)=C_{0xi}+C_{1xi}t$, $y_i(t)=C_{0yi}+C_{1yi}t$, $z_i(t)=C_{0zi}+C_{1zi}t$. Константы $C_{0xi}, C_{1xi}, C_{0yi}, C_{1yi}, C_{0zi}, C_{1zi}$ должны быть найдены из краевых условий (8.7).

В функционале (8.2) функция h зависит от x_1, x_2, \dots, x_n и их первых производных x_1', x_2', \dots, x_n' по t . В общем случае функция h может зависеть от производных более высокого порядка. Рассмотрим функционал вида

$$\Phi(h) = \int_0^1 h(t, x_1, x_2, \dots, x_n, x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}, x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}, \dots, x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}) dt, \quad (8.11)$$

где $x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}$ — производные k -го порядка варьируемых функций x_1, x_2, \dots, x_n по параметру t .

Подставим вариации (8.3) функций и их производные в (8.11), найдем производную $\Phi'(a)$ при $a=0$ и приравняем ее нулю. В результате получим, что необходимое условие экстремума функционала (8.11) выражается равенством

$$\int_0^1 \sum_{i=1}^n (h_i \delta x_i + h_{i(1)} \delta x_i^{(1)} + h_{i(2)} \delta x_i^{(2)} + \dots + h_{i(k)} \delta x_i^{(k)}) dt = 0. \quad (8.12)$$

где

$$h_i = \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}, \dots, x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})}{\partial x_i},$$

$$h_{i(m)} = \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}, \dots, x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})}{\partial (d^m x_i / dt^m)},$$

$$i = 1, 2, \dots, n, \quad m = 1, 2, \dots, k.$$

Проинтегрируем k раз по частям слагаемые, в которые входят k -е производные вариаций $\delta x_i^{(k)}$

$$\int_0^1 h_{i(k)} \delta x_i^{(k)} dt = h_{i(k)} \delta x_i^{(k-1)} \Big|_0^1 - \frac{dh_{i(k)}}{dt} \delta x_i^{(k-2)} \Big|_0^1 + \frac{d^2 h_{i(k)}}{dt^2} \delta x_i^{(k-3)} \Big|_0^1 - \dots$$

$$\dots + (-1)^{k-1} \frac{d^{k-1} h_{i(k)}}{dt^{k-1}} \delta x_i \Big|_0^1 + (-1)^k \int_0^1 \frac{d^k h_{i(k)}}{dt^k} \delta x_i dt, \quad i = 1, 2, \dots, n. \quad (8.13)$$

В результате получим, что необходимое условие экстремума функционала (8.11) выражается равенством

$$\sum_{i=1}^n \left\{ h_{i(k)} \delta x_i^{(k-1)} \Big|_0^1 + \left(h_{i(k-1)} - \frac{dh_{i(k)}}{dt} \right) \delta x_i^{(k-2)} \Big|_0^1 + \right.$$

$$\left. + \left(h_{i(k-2)} - \frac{dh_{i(k-1)}}{dt} + \frac{d^2 h_{i(k)}}{dt^2} \right) \delta x_i^{(k-3)} \Big|_0^1 + \dots \right.$$

$$\left. + \left(h_{i(1)} - \frac{dh_{i(2)}}{dt} + \dots + (-1)^{k-1} \frac{d^{k-1} h_{i(k)}}{dt^{k-1}} \right) \delta x_i \Big|_0^1 \right\} +$$

$$+ \sum_{i=1}^n \int_0^1 \left(h_i - \frac{dh_{i(1)}}{dt} + \frac{d^2 h_{i(2)}}{dt^2} + \dots + (-1)^k \frac{d^k h_{i(k)}}{dt^k} \right) \delta x_i dt = 0,$$

$$i = 1, 2, \dots, n. \quad (8.14)$$

в конечном состоянии варьируемых функций. Эти равенства называются общими условиями трансверсальности. Совместно с уравнениями связей (8.17) уравнения (8.22) представляют систему $n+m$ уравнений относительно $\lambda_1, \lambda_2, \dots, \lambda_m$. Они позволяют найти вторую половину произвольных постоянных в общем решении уравнений Эйлера.

В результате мы получили, что необходимое условие экстремума функционала (8.2) с условиями (8.17) на границе, требует выполнения уравнений Эйлера (8.8), равенств нулю вариаций начального положения (8.18), выполнения условий трансверсальности (8.22) и уравнений связей (8.17) в конечном состоянии варьируемых функций (при $t=1$). Другими словами, при известных начальных условиях и известных уравнениях связей для конечных значений варьируемых функций необходимое условие экстремума функционала $\Phi(h)$ на функциях $x_1(t), x_2(t), \dots, x_n(t)$, удовлетворяющих уравнениям Эйлера, сводится к выполнению системы уравнений

$$\begin{aligned}
 & \left(\frac{\partial h}{\partial(dx_1/dt)} + \lambda_1 \frac{\partial f_1}{\partial x_1} + \lambda_2 \frac{\partial f_2}{\partial x_1} + \dots + \lambda_m \frac{\partial f_m}{\partial x_1} \right) \Big|_{t=1} = 0, \\
 & \left(\frac{\partial h}{\partial(dx_2/dt)} + \lambda_1 \frac{\partial f_1}{\partial x_2} + \lambda_2 \frac{\partial f_2}{\partial x_2} + \dots + \lambda_m \frac{\partial f_m}{\partial x_2} \right) \Big|_{t=1} = 0, \\
 & \dots \dots \dots \\
 & \left(\frac{\partial h}{\partial(dx_n/dt)} + \lambda_1 \frac{\partial f_1}{\partial x_n} + \lambda_2 \frac{\partial f_2}{\partial x_n} + \dots + \lambda_m \frac{\partial f_m}{\partial x_n} \right) \Big|_{t=1} = 0, \\
 & \quad f_1(x_1, x_2, \dots, x_n) \Big|_{t=1} = 0, \\
 & \quad f_2(x_1, x_2, \dots, x_n) \Big|_{t=1} = 0, \\
 & \quad \dots \dots \dots \\
 & \quad f_m(x_1, x_2, \dots, x_n) \Big|_{t=1} = 0.
 \end{aligned} \tag{8.23}$$

Эта система и ее решение симметричны относительно искомым функций. Функционал $\Phi(h)$ формулирует критерий поведения варьируемых функций $x_1(t), x_2(t), \dots, x_n(t)$ при их переходе из начального состояния в состояние, удовлетворяющее наложенным на них вариационным связям (8.17).

Связь с экстремумом функции.

Для принятого нами критерия поведения функций $x_1(t), x_2(t), \dots, x_n(t)$ потребуем наименьшей суммы квадратов изменений функций, которую мы опишем интегралом

$$\Phi = \frac{1}{2} \int_0^t \sum_{i=1}^n \left(\frac{dx_i}{dt} \right)^2 dt. \quad (8.24)$$

Подинтегральная функция в данном случае равна половине суммы квадратов производных

$$h = \frac{1}{2} \left(\left(\frac{dx_1}{dt} \right)^2 + \left(\frac{dx_2}{dt} \right)^2 + \dots + \left(\frac{dx_n}{dt} \right)^2 \right). \quad (8.25)$$

В соответствии с (8.9) и (8.18) общее решение системы уравнений Эйлера для функции (8.25) имеет линейный вид

$$x_i(t) = tx_i + (1-t)x_i^{(0)}, \quad i = 1, 2, \dots, n, \quad (8.26)$$

где $x_i^{(0)} = x_i(0)$ — начальные значения функций, $x_i = x_i(1)$ — искомые конечные значения варьируемых функций. Пусть функции $x_1(t), x_2(t), \dots, x_n(t)$ при $t=1$ должны удовлетворять уравнениям связей (8.17).

Конечные значения функций должны быть найдены из системы уравнений (8.23), которая после подстановки (8.26) в функцию (8.25) примет вид

получим уравнения геодезической линии из равенства нулю вариации функционала

$$\Phi = \frac{1}{2} \int_0^1 \left(\frac{ds}{dt} \right)^2 dt. \quad (9.1)$$

Пусть u^1, u^2, u^3 — криволинейные координаты в трехмерном пространстве. Криволинейные координаты в силу их поведения при преобразованиях координат должны иметь верхний индекс. Рассмотрим поведение точки, координаты которой являются варьируемыми параметрами связей, в криволинейной системе координат. Пусть в исходном состоянии рассматриваемая точка имела координаты $u^i = u^{i(0)}, i=1, 2, 3$. После изменения уравнений связей или констант в уравнениях связей она займет новое положение в пространстве. Проведем произвольную линию $u^i(t)$ из исходного положения рассматриваемой точки в ее новое положение.

Пусть при $t=0$ линия проходит через исходную точку с координатами $u^i = u^{i(0)}$, а при $t=1$ линия проходит через новое положение точки. Найдем уравнения линии, для которой функционал (9.1) принимает минимальное значение. В декартовой прямоугольной системе координат $u^1=x, u^2=y, u^3=z$, а функционал (9.1) имеет вид

$$\Phi = \frac{1}{2} \int_0^1 \left(\left(\frac{dx}{dt} \right)^2 + \left(\frac{dy}{dt} \right)^2 + \left(\frac{dz}{dt} \right)^2 \right) dt.$$

В криволинейной системе координат u^1, u^2, u^3 аналогичный функционал получим, подставив (1.10.5) в (9.1),

$$\Phi = \frac{1}{2} \int_0^1 g_{ij} \frac{du^i}{dt} \frac{du^j}{dt} dt = \frac{1}{2} \int_0^1 \left(g_{11} \left(\frac{du^1}{dt} \right)^2 + g_{22} \left(\frac{du^2}{dt} \right)^2 + g_{33} \left(\frac{du^3}{dt} \right)^2 + 2g_{12} \frac{du^1}{dt} \frac{du^2}{dt} + 2g_{23} \frac{du^2}{dt} \frac{du^3}{dt} + 2g_{13} \frac{du^1}{dt} \frac{du^3}{dt} \right) dt, \quad (9.2)$$

где g_{ij} — ковариантные компоненты метрического тензора. В (7.9.2) используется соглашение о суммировании по повторяющимся верхним и нижним индексам. Пусть функции $u^1(t)$, $u^2(t)$, $u^3(t)$ дифференцируемы требуемое число раз. Возьмем совокупность близких к ним функций $u^i(t) + \delta u^i(t)$, $i=1, 2, 3$, где величины $\delta u^i(t)$ называются вариациями соответствующих функций. Эти функции можно дифференцировать по параметру t

$$\frac{d(\delta u^i)}{dt} = \delta \left(\frac{du^i}{dt} \right), \quad \frac{d^2(\delta u^i)}{dt^2} = \delta \left(\frac{d^2 u^i}{dt^2} \right), \quad \frac{d^k(\delta u^i)}{dt^k} = \delta \left(\frac{d^k u^i}{dt^k} \right).$$

Символ δ обозначает переход из какой-нибудь точки искомой линии в точку некоторой другой кривой, которой соответствует то же значение параметра t . Вариация функционала (9.1) выражается формулой

$$\delta \Phi = \frac{1}{2} \int_0^1 \left(\frac{\partial g_{ij}}{\partial u^k} \frac{du^i}{dt} \frac{du^j}{dt} \delta u^k + 2g_{ij} \frac{du^i}{dt} \delta \left(\frac{du^j}{dt} \right) \right) dt, \quad (9.3)$$

где использовалась симметрия метрического тензора $g_{ij} = g_{ji}$. Проинтегрируем по частям вторую половину слагаемых в (9.3) и получим

$$\begin{aligned} \int_0^1 g_{ij} \frac{du^i}{dt} \delta \left(\frac{du^j}{dt} \right) dt &= g_{ij} \frac{du^i}{dt} \delta u^j \Big|_0^1 - \int_0^1 \frac{d(g_{ij} (du^i/dt))}{dt} \delta u^j dt = \\ &= g_{ij} \frac{du^i}{dt} \delta u^j \Big|_0^1 - \int_0^1 \left(\frac{\partial g_{ij}}{\partial u^m} \frac{du^i}{dt} \frac{du^m}{dt} + g_{ij} \frac{d^2 u^i}{dt^2} \right) \delta u^j dt. \end{aligned} \quad (9.4)$$

В силу равенства нулю вариаций $\delta u^i \Big|_0=0$ в конечных точках вариация функционала (9.1) выразится формулой

$$\delta \Phi = \frac{1}{2} \int_0^1 \left(\frac{\partial g_{ij}}{\partial u^k} \frac{du^i}{dt} \frac{du^j}{dt} - 2 \frac{\partial g_{ik}}{\partial u^m} \frac{du^i}{dt} \frac{du^m}{dt} - 2g_{ik} \frac{d^2 u^i}{dt^2} \right) \delta u^k dt. \quad (9.5)$$

Необходимое условие экстремума функционала выражается равенством нулю его вариации. Отсюда в силу произвольности выбора δu^k следует, что функционал (9.1) достигает экстремума на кривых $u^i(t)$, $i=1, 2, 3$, удовлетворяющих дифференциальным уравнениям

$$\frac{\partial g_{ij}}{\partial u^k} \frac{du^i}{dt} \frac{du^j}{dt} - 2 \frac{\partial g_{ik}}{\partial u^m} \frac{du^i}{dt} \frac{du^m}{dt} - 2g_{ik} \frac{d^2 u^i}{dt^2} = 0. \quad (9.6)$$

Изменим обозначения индексов (индексы, по которым выполняется суммирование, мы имеем право обозначать любыми буквами) и преобразуем первое и второе слагаемые в левой части равенства (9.6) следующим образом:

$$\begin{aligned} \frac{\partial g_{ij}}{\partial u^k} \frac{du^i}{dt} \frac{du^j}{dt} - 2 \frac{\partial g_{ik}}{\partial u^m} \frac{du^i}{dt} \frac{du^m}{dt} &\equiv \\ &\equiv \frac{\partial g_{ij}}{\partial u^k} \frac{du^i}{dt} \frac{du^j}{dt} - \frac{\partial g_{ik}}{\partial u^m} \frac{du^i}{dt} \frac{du^m}{dt} - \frac{\partial g_{ik}}{\partial u^m} \frac{du^i}{dt} \frac{du^m}{dt} \equiv \\ &\equiv \frac{\partial g_{ij}}{\partial u^k} \frac{du^i}{dt} \frac{du^j}{dt} - \frac{\partial g_{ik}}{\partial u^j} \frac{du^i}{dt} \frac{du^j}{dt} - \frac{\partial g_{jk}}{\partial u^i} \frac{du^j}{dt} \frac{du^i}{dt} = \\ &= \left(\frac{\partial g_{ij}}{\partial u^k} - \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{jk}}{\partial u^i} \right) \frac{du^j}{dt} \frac{du^i}{dt} = -2\Gamma_{ij,k} \frac{du^j}{dt} \frac{du^i}{dt}, \end{aligned}$$

где $\Gamma_{ij,k} = \frac{1}{2} \left(\frac{\partial g_{ik}}{\partial u^j} + \frac{\partial g_{jk}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^k} \right)$ — символы Кристоффеля 1-го рода (10.16).

С учетом полученного вариация функционала (9.1) будет равна нулю при выполнении равенства

$$g_{mk} \frac{d^2 u^m}{dt^2} + \Gamma_{ij,k} \frac{du^i}{dt} \frac{du^j}{dt} = 0. \quad (9.7)$$

Умножив равенства (9.7) на g^{rk} выполнив суммирование и используя свойство $g^{rk} g_{mk} = \delta_m^r$, где δ_k^r — символы Кронекера (10.7), получим

$$\frac{d^2 u^r}{dt^2} + \Gamma_{ij}^r \frac{du^i}{dt} \frac{du^j}{dt} = 0. \quad (9.8)$$

5.10. Вариационные связи двумерных точек

Применим метод минимизации изменений связанных параметров для вариационных связей двумерных геометрических объектов в декартовой прямоугольной системе координат. Применение метода минимизации изменений связанных параметров для вариационных связей трехмерных геометрических объектов аналогично тому, что будет изложено ниже. Уравнения связей для трехмерных геометрических объектов приведены выше и мы не будем повторяться. Варьируемые параметры трехмерных геометрических объектов посредством свободных параметров могут быть связаны с двумерными геометрическими объектами.

Двухмерные геометрические объекты используются для построения пространственных геометрических объектов. Многие тела строятся на основе плоских контуров путем выдавливания, вращения, движения вдоль заданной линии, плавного соединения нескольких плоских контуров. Плоские кривые и контуры строятся на двумерных линиях. Для того чтобы перестроить пространственный объект, требуется перестроить двумерный объект. Большие возможности в этом случае дают вариационные связи, делающие зависимыми параметры из структур данных геометрических объектов.

Для двумерных объектов многие уравнения вариационных связей имеют тот же векторный вид, что и для пространственных объектов. Отличие от вышеизложенного состоит в том, что двумерные точки и векторы имеют всего две координаты и, соответственно, каждая вариационная связь будет содержать меньшее число параметров. Рассмотрим некоторые двумерные вариационные связи.

Для фиксации скалярных параметров, например, радиуса окружности или угла дуги окружности, используются фиксирующие связи типа (2.1). Для закрепления двумерной точки применяется векторное уравнение (2.2), состоящее из двух скалярных уравнений (2.3).

Размер вдоль координаты.

Размер вдоль одной координаты между двумерными точками $p_1=[x_1 \ y_1]^T$ и $p_2=[x_2 \ y_2]^T$ описывается одним из уравнений

$$x_2 - x_1 - d = 0, \quad \text{или} \quad y_2 - y_1 - d = 0, \quad (10.1)$$

где d — требуемый размер. Пусть до постановки размера положение точек описывалось радиус-векторами $p_1^{(0)} = [x_1^{(0)} \ y_1^{(0)}]^T$ и $p_2^{(0)} = [x_2^{(0)} \ y_2^{(0)}]^T$. Для того, чтобы поведение точек при установке размера было симметричным, потребуем, чтобы сумма квадратов перемещений связанных точек

$$\begin{aligned} \psi &= \frac{1}{2} (|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2) = \\ &= \frac{1}{2} ((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2) \end{aligned} \quad (10.2)$$

была минимальной. Для определенности будем считать, что размер поставлен вдоль координат x . Необходимым условием минимума функции (10.2) при условии (10.1) является равенство нулю частных производных по параметрам $x_1, y_1, x_2, y_2, \lambda$ функции

$$F \doteq \psi + \lambda f = \frac{1}{2} ((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2) + \lambda(x_2 - x_1 - d). \quad (7.10.3) \quad (10.3)$$

Искомые координаты точек и параметр λ найдем из системы уравнений

$$\begin{aligned} x_2 - x_1 &= d, \\ x_1 - x_1^{(0)} - \lambda &= 0, & y_1 - y_1^{(0)} &= 0, \\ x_2 - x_2^{(0)} + \lambda &= 0, & y_2 - y_2^{(0)} &= 0. \end{aligned} \quad (10.4)$$

Данная система уравнений имеет решение

$$\begin{aligned}
 x_1 &= \frac{1}{2}(x_2^{(0)} + x_1^{(0)} - d), & y_1 &= y_1^{(0)}, \\
 x_2 &= \frac{1}{2}(x_2^{(0)} + x_1^{(0)} + d), & y_2 &= y_2^{(0)}, \\
 \lambda &= \frac{1}{2}(x_2^{(0)} - x_1^{(0)} - d).
 \end{aligned}
 \tag{10.5}$$

Можно доказать, что найденное решение является точкой минимума функции (10.2). Перемещение точек производится вдоль координаты x . В данной постановке задачи точки являются равноправными. Уравнения системы и ее решение являются симметричными без использования каких-либо дополнительных ограничений относительно поведения системы точек.

Если одна из точек закреплена, то на нее наложена связь (2.2) и ее координаты не варьируются. Пусть точка p_2 закреплена. Тогда система будет иметь решение

$$\begin{aligned}
 x_1 &= x_2^{(0)} - d, & y_1 &= y_1^{(0)}, & x_2 &= x_2^{(0)}, & y_2 &= y_2^{(0)}, \\
 \lambda &= x_2^{(0)} - x_1^{(0)} - d.
 \end{aligned}
 \tag{10.0}$$

Размер вдоль направления.

Размер вдоль заданного направления между двухмерными точками $p_1=[x_1 \ y_1]^T$ и $p_2=[x_2 \ y_2]^T$ описывается уравнением (3.12), которое в координатной записи имеет вид

$$(x_2 - x_1)m_x + (y_2 - y_1)m_y - d = 0, \tag{10.7}$$

где $\mathbf{m}=[m_x \ m_y]^T$ — задающий направление двухмерный вектор единичной длины. Найдем положение точек из условия минимума их суммарного перемещения (10.2). В соответствии с методом Лагранжа составим функцию

$$F = \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 \right) + \lambda \left((x_2 - x_1)m_x + (y_2 - y_1)m_y - d \right). \quad (10.8)$$

Искомые координаты точек и параметр λ найдем из равенство нулю частных производных этой функции по параметрам $x_1, y_1, x_2, y_2, \lambda$

$$\begin{aligned} (x_2 - x_1)m_x + (y_2 - y_1)m_y - d &= 0, \\ x_1 - x_1^{(0)} - \lambda m_x &= 0, & y_1 - y_1^{(0)} - \lambda m_y &= 0, \\ x_2 - x_2^{(0)} + \lambda m_x &= 0, & y_2 - y_2^{(0)} + \lambda m_y &= 0. \end{aligned} \quad (10.9)$$

Данная система уравнений имеет решение

$$\begin{aligned} x_1 &= x_1^{(0)} + \frac{m_x}{2} \left((x_2^{(0)} - x_1^{(0)})m_x + (y_2^{(0)} - y_1^{(0)})m_y - d \right), \\ y_1 &= y_1^{(0)} + \frac{m_y}{2} \left((x_2^{(0)} - x_1^{(0)})m_x + (y_2^{(0)} - y_1^{(0)})m_y - d \right), \\ x_2 &= x_2^{(0)} - \frac{m_x}{2} \left((x_2^{(0)} - x_1^{(0)})m_x + (y_2^{(0)} - y_1^{(0)})m_y - d \right), \\ y_2 &= y_2^{(0)} - \frac{m_y}{2} \left((x_2^{(0)} - x_1^{(0)})m_x + (y_2^{(0)} - y_1^{(0)})m_y - d \right), \\ \lambda &= \frac{1}{2} \left((x_2^{(0)} - x_1^{(0)})m_x + (y_2^{(0)} - y_1^{(0)})m_y - d \right). \end{aligned} \quad (10.10)$$

Перемещение точек производится вдоль вектора \mathbf{m} на одинаковое расстояние от исходного положения точек. Решение системы уравнений (7.10.9) является симметричным.

Линейный размер.

Линейный размер на плоскости между двумя точками $p_1 = [x_1 \ y_1]^T$ и $p_2 = [x_2 \ y_2]^T$ описывается уравнением (3.1), которое в координатной записи имеет вид

$$f = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d = 0. \quad (10.11)$$

Для симметричного поведения точек при установки размера потребуем, чтобы суммарное перемещение связанных точек было минимальным. Необходимым условием минимума функции (10.2) при условии (10.11) является равенство нулю частных производных по параметрам функции

$$\begin{aligned}
 F &= \frac{1}{2} (|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2) + \lambda(|p_2 - p_1| - d) = \\
 &= \frac{1}{2} ((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2) + \\
 &\quad + \lambda (\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d). \quad (7. (10.12))
 \end{aligned}$$

Система уравнений, из которой найдется искомое положение точек, аналогична системе (6.3). Решение такой системы уравнений имеет вид

$$\begin{aligned}
 x_1 &= x_1^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}} (x_1^{(0)} - x_2^{(0)}), \\
 y_1 &= y_1^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}} (y_1^{(0)} - y_2^{(0)}), \\
 x_2 &= x_2^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}} (x_2^{(0)} - x_1^{(0)}), \\
 y_2 &= y_2^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}} (y_2^{(0)} - y_1^{(0)}), \\
 \lambda &= -\frac{1}{2} (d - d^{(0)}).
 \end{aligned} \tag{10.13}$$

где $d^{(0)} = \sqrt{(x_2^{(0)} - x_1^{(0)})^2 + (y_2^{(0)} - y_1^{(0)})^2}$

Перемещение точек производится вдоль прямой линии, проходящей через точки $p_1^{(0)}$ и $p_2^{(0)}$, на одинаковое расстояние.

Если точки p_1 и p_2 входят в структуру данных отрезка, то с помощью связей

$$x_2 = x_1, \quad \text{или} \quad y_2 = y_1,$$

отрезок можно сделать параллельным одной из координатных осей. Пусть требуется, чтобы отрезок был параллельным оси y . Тогда из равенства нулю частных производных функции

$$F = \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 \right) + \lambda(x_2 - x_1) \quad (10.14)$$

получим систему уравнений

$$\begin{aligned} x_1 - x_1^{(0)} - \lambda &= 0, & x_2 - x_2^{(0)} + \lambda &= 0, \\ x_2 - x_1 &= 0. \end{aligned}$$

Ее решение равно

$$\begin{aligned} x_1 &= \frac{1}{2}(x_1^{(0)} + x_2^{(0)}), & x_2 &= \frac{1}{2}(x_2^{(0)} + x_1^{(0)}), \\ \lambda &= \frac{1}{2}(x_2^{(0)} - x_1^{(0)}). \end{aligned} \quad (10.15)$$

Симметрия двух точек.

Рассмотрим симметрию точек $p_1=[x_1 \ y_1]^T$ и $p_2=[x_2 \ y_2]^T$ относительно прямой линии

$$l(t) = p + ti, \quad (10.16)$$

что показано на рис. 10.1. Прямую будем считать неподвижной, поэтому ее параметры варьироваться не будут.

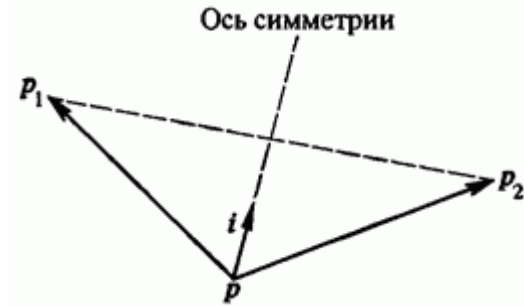


Рис. 10.1. Симметрия двумерных точек относительно прямой

Симметрия точек относительно прямой линии описывается двумя уравнениями

$$\begin{aligned} (\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{i} &= (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{i}, \\ (\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{j} + (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{j} &= 0, \end{aligned} \quad (10.17)$$

где вектор $\mathbf{j} = [-b \ a]^T$ ортогонален вектору $\mathbf{i} = [a \ b]^T$. Пусть точка p имеет координаты x и y . Составим функцию

$$\begin{aligned} F = \frac{1}{2} & \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 \right) + \\ & + \lambda_1 ((x_1 - x_2)a + (y_1 - y_2)b) + \lambda_2 ((y_1 + y_2 - 2y)a - (x_1 + x_2 - 2x)b). \end{aligned} \quad (10.18)$$

Систему уравнений для определения положения точек получим из равенства нулю частных производных функции (10.18)

$$\begin{aligned}x_1 - x_1^{(0)} + \lambda_1 a - \lambda_2 b &= 0, \\y_1 - y_1^{(0)} + \lambda_1 b + \lambda_2 a &= 0, \\x_2 - x_2^{(0)} - \lambda_1 a - \lambda_2 b &= 0, \\y_2 - y_2^{(0)} - \lambda_1 b + \lambda_2 a &= 0, \\(x_1 - x_2)a + (y_1 - y_2)b &= 0, \\(y_1 + y_2 - 2y)a - (x_1 + x_2 - 2x)b &= 0.\end{aligned}$$

Решение этой системы определит новое положение точек.

Если мы хотим, чтобы точка p_1 осталась неподвижной, то введем уравнение

$$p_1 = p_1^{(0)} \quad (10.19)$$

или в (10.18) ее координаты будем считать константами.

Если мы хотим, чтобы точка p_2 осталась неподвижной, а точка p_1 стала ей симметрична, то введем уравнение

$$p_2 = p_2^{(0)} \quad (10.20)$$

или в (7.10.18) ее координаты будем считать неварьируемыми константами. При закреплении одной из точек решение системы уравнений (7.10.18) определит положение другой точки.

Угловой размер между тремя точками на плоскости.

Угловой размер на плоскости между тремя точками $p_1=[x_1 \ y_1]^T$, $p_2=[x_2 \ y_2]^T$, $p_3=[x_3 \ y_3]^T$ описывается уравнением (3.28), где

$$s = (x_1 - x_2)(y_3 - y_2) - (y_1 - y_2)(x_3 - x_2),$$

$$c = (x_1 - x_2)(x_3 - x_2) + (y_1 - y_2)(y_3 - y_2). \quad (10.21)-(10.22)$$

Величина s равна произведению длин векторов $p_1 - p_2$, $p_3 - p_2$ на $\sin \alpha$.
 Величина c равна произведению длин векторов $p_1 - p_2$, $p_3 - p_2$ на $\cos \alpha$.
 Угловой размер на плоскости связывает шесть параметров. Составим функцию

$$F = \frac{1}{2} (|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 + |p_3 - p_3^{(0)}|^2) + \lambda \left(\arctg \frac{s}{c} - \alpha \right), \quad (10.23)$$

где $p_i^{(0)} = [x_i^{(0)} \ y_i^{(0)}]^T$, $p_2^{(0)} = [x_2^{(0)} \ y_2^{(0)}]^T$, $p_3^{(0)} = [x_3^{(0)} \ y_3^{(0)}]^T$ — исходные положения точек. Систему уравнений для определения параметров получим из равенства нулю частных производных функции (10.23)

$$x_1 - x_1^{(0)} + \lambda \frac{c(y_3 - y_2) - s(x_3 - x_2)}{s^2 + c^2} = 0,$$

$$y_1 - y_1^{(0)} + \lambda \frac{-c(x_3 - x_2) - s(y_3 - y_2)}{s^2 + c^2} = 0,$$

$$x_2 - x_2^{(0)} + \lambda \frac{c(y_1 - y_3) - s(2x_2 - x_1 - x_3)}{s^2 + c^2} = 0,$$

$$y_2 - y_2^{(0)} + \lambda \frac{c(x_3 - x_1) - s(2y_2 - y_1 - y_3)}{s^2 + c^2} = 0,$$

$$x_3 - x_3^{(0)} + \lambda \frac{-c(y_1 - y_2) - s(x_1 - x_2)}{s^2 + c^2} = 0,$$

$$y_3 - y_3^{(0)} + \lambda \frac{c(x_1 - x_2) - s(y_1 - y_2)}{s^2 + c^2} = 0,$$

$$\arctg \left(\frac{(x_1 - x_2)(y_3 - y_2) - (x_3 - x_2)(y_1 - y_2)}{(x_1 - x_2)(x_3 - x_2) + (y_1 - y_2)(y_3 - y_2)} \right) = \alpha. \quad (10.24)$$

Решение этой системы определит положение точек при заданном угле α .

5.11. Вариационные связи двумерных линий

Вариационные связи для нескольких двумерных точек рассмотрим в виде связей точек двумерных кривых. С помощью вариационных связей легко управлять кривыми. Например, отрезки можно сделать ортогональными, параллельными друг другу или осям координат, окружности и сплайны можно делать касательными отрезкам или друг другу.

Пусть даны два отрезка прямых

$$\begin{aligned} \mathbf{r}_1(t_1) &= (1 - t_1)\mathbf{p}_1 + t_1\mathbf{p}_2, \\ \mathbf{r}_3(t_3) &= (1 - t_3)\mathbf{p}_3 + t_3\mathbf{p}_4, \end{aligned} \quad (11.1)-(11.2)$$

построенных по точкам $\mathbf{p}_1=[x_1 \ y_1]^T$, $\mathbf{p}_2=[x_2 \ y_2]^T$, $\mathbf{p}_3=[x_3 \ y_3]^T$, $\mathbf{p}_4=[x_4 \ y_4]^T$. На данные четыре точки можно наложить вариационную связь, определяющую угол между отрезками $\mathbf{r}_1(t_1)$, $\mathbf{r}_3(t_3)$. Эта связь аналогична угловому размеру и описывается тем же уравнением

$$\operatorname{arctg} \frac{s}{c} = \alpha, \quad (11.3)$$

где α — заданный угол. Синус и косинус угла пропорциональны величинам:

$$\begin{aligned} s &= (x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3), \\ c &= (x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3). \end{aligned} \quad (11.4)-(11.5)$$

Уравнение (11.3) связывает восемь параметров.

Ортогональность отрезков.

Как частный случай связи (11.3) может рассматриваться вариационная связь устанавливающая ортогональность отрезков. Данная вариационная связь описывается уравнением

$$(p_2 - p_1) \cdot (p_4 - p_3) = (x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3) = 0. \quad (11.6)$$

Систему уравнений для определения положения точек получим из равенства нулю частных производных функции

$$F = \frac{1}{2} \left(|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 + |p_3 - p_3^{(0)}|^2 + |p_4 - p_4^{(0)}|^2 \right) + \lambda \left((x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3) \right), \quad (11.7)$$

где $p_1^{(0)} = [x_1^{(0)} \ y_1^{(0)}]^T$, $p_2^{(0)} = [x_2^{(0)} \ y_2^{(0)}]^T$, $p_3^{(0)} = [x_3^{(0)} \ y_3^{(0)}]^T$, $p_4^{(0)} = [x_4^{(0)} \ y_4^{(0)}]^T$ — исходные положения точек, в общем случае не удовлетворяющие уравнению (11.6).

Параллельность отрезков.

Другим частным случаем связи (11.3) является вариационная связь, устанавливающая параллельность отрезков. Она определяется уравнением

$$(x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3) = 0. \quad (11.8)$$

Пусть в исходном состоянии координаты точек $p_1^{(0)} = [x_1^{(0)} \ y_1^{(0)}]^T$, $p_2^{(0)} = [x_2^{(0)} \ y_2^{(0)}]^T$, $p_3^{(0)} = [x_3^{(0)} \ y_3^{(0)}]^T$, $p_4^{(0)} = [x_4^{(0)} \ y_4^{(0)}]^T$ не удовлетворяют уравнению (11.7). Применим метод минимизации суммарного изменения параметров. Систему уравнений для определения положения точек получим из равенства нулю частных производных функции

$$F = \frac{1}{2} \left(|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 + |p_3 - p_3^{(0)}|^2 + |p_4 - p_4^{(0)}|^2 \right) + \lambda \left((x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3) \right). \quad (11.9)$$

Для простоты предположим, что точки p_2, p_3, p_4 зафиксированы и, следовательно, не изменяют своего положения. Параллельность отрезков (11.1) и (11.2) будет достигнута перемещением точки p_1 . Система уравнений будет иметь вид

$$\begin{aligned} x_1 - x_1^{(0)} - \lambda(y_4 - y_3) &= 0, \\ y_1 - y_1^{(0)} + \lambda(x_4 - x_3) &= 0, \\ (x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3) &= 0. \end{aligned} \quad (11.10)$$

Эта система имеет решение

$$\begin{aligned} x_1 &= x_1^{(0)} + (y_4^{(0)} - y_3^{(0)}) \frac{h^{(0)}}{d^{(0)}}, \\ y_1 &= y_1^{(0)} - (x_4^{(0)} - x_3^{(0)}) \frac{h^{(0)}}{d^{(0)}}, \\ \lambda &= \frac{h^{(0)}}{d^{(0)}}, \end{aligned} \quad (11.11)$$

где $d^{(0)} = |p_4^{(0)} - p_3^{(0)}|^2$, $h^{(0)} = (x_2^{(0)} - x_1^{(0)}) (y_4^{(0)} - y_3^{(0)}) - (y_2^{(0)} - y_1^{(0)}) (x_4^{(0)} - x_3^{(0)})$. Перемещение точки p_1 происходит по нормали к отрезку, проходящему через точки $p_4^{(0)} - p_3^{(0)}$ (рис. 11.1).

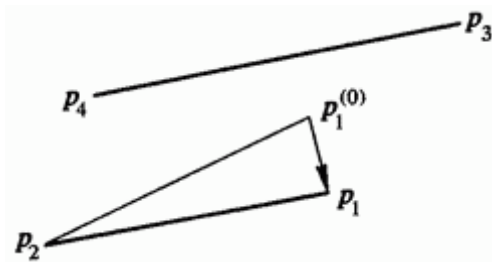


Рис. 11.1. Параллельность отрезков прямой линии

В описанном случае параллельности отрезков первый зависит от второго и подстраивается под него. Рассмотрим, как будет себя вести зависимый отрезок при вращении другого отрезка.

Пусть в исходном состоянии точки p_2 , p_4 закреплены, отрезки параллельны и их точки имеют координаты

$$\begin{aligned}x_1^{(0)} - x_2^{(0)} &= b, \\y_1^{(0)} - y_2^{(0)} &= 0, \\x_3^{(0)} - x_4^{(0)} &= a, \\y_4^{(0)} - y_3^{(0)} &= 0.\end{aligned}$$

Будем вращать точку p_3 вокруг точки p_4 так, чтобы их координаты изменялись по закону

$$\begin{aligned}x_3^{(0)} &= x_4^{(0)} + a \cos \varphi, \\y_3^{(0)} &= y_4^{(0)} + a \sin \varphi.\end{aligned}\quad (11.12)$$

В соответствии с решением (11.11) точка p_1 будет двигаться вокруг точки p_2 так, что ее координаты будут изменяться по закону

$$\begin{aligned}x_1 &= x_1^{(0)} - b \sin^2 \varphi = x_2^{(0)} + b \cos^2 \varphi = x_2^{(0)} + r \cos \varphi, \\y_1 &= y_1^{(0)} + b \sin \varphi \cos \varphi = y_2^{(0)} + b \cos \varphi \sin \varphi = y_2^{(0)} + r \sin \varphi,\end{aligned}$$

где $r = b \cos \alpha$. Из формул видно, что точка p_1 будет двигаться вокруг точки p_2 по окружности, но с удвоенной частотой по сравнению с частотой вращения точки p_3 вокруг точки p_4 .

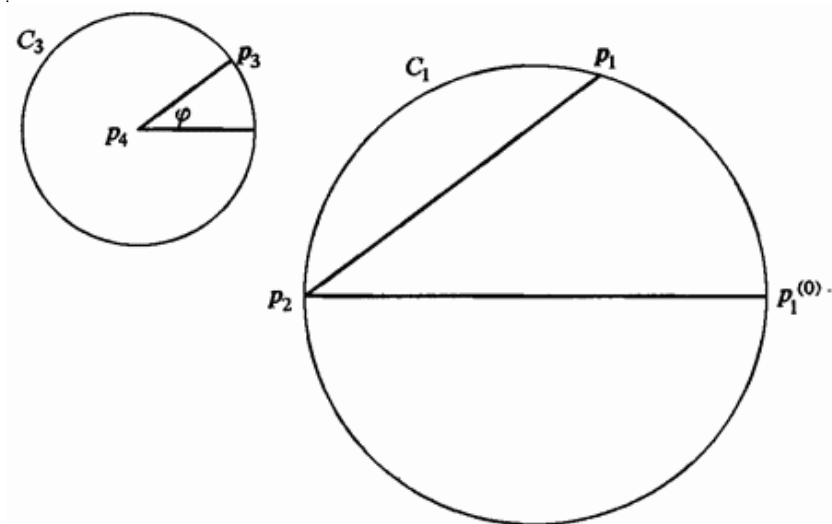


Рис. 11.2. Траектории движения отрезка $p_1 p_2$ параллельно зависимо от отрезка $p_3 p_4$ при закрепленных точках p_2, p_4

При изменении угла ψ от 0 до π точка p_3 сделает половину оборота по окружности вокруг точки p_4 , а точка p_1 сделает полный оборот по окружности вокруг точки p_2 . Траектории движения точек p_1, p_3 показаны на рис. 11.2. Точка p_3 движется по окружности C_3 , а точка p_1 движется по окружности C_1 .

Касание сплайнов.

Рассмотрим вариационные связи касания линий друг друга. Пусть даны две NURBS кривые. Одна из них описывается функцией

$$r_p(t) = \frac{\sum_{i=1}^n N_{i,m}(t) w_i p_i}{\sum_{i=1}^n N_{i,m}(t) w_i}, \quad t_{\min} \leq t \leq t_{\max}, \quad (11.13)$$

и построена на точках p_i , $i=1, 2, \dots, n$ с весами w_i , а другая описывается функцией

$$r_q(u) = \frac{\sum_{j=1}^k N_{j,l}(u) z_j q_j}{\sum_{j=1}^k N_{j,l}(u) z_j}, \quad u_{\min} \leq u \leq u_{\max}, \quad (11.14)$$

и построена на точках q_j , $j=1, 2, \dots, k$ с весами z_j . Найдем две наиболее близкие друг к другу точки $r_p(t_0)$ и $r_q(u_0)$ на этих кривых, одна — на первой, другая — на второй, касательные в которых параллельны.

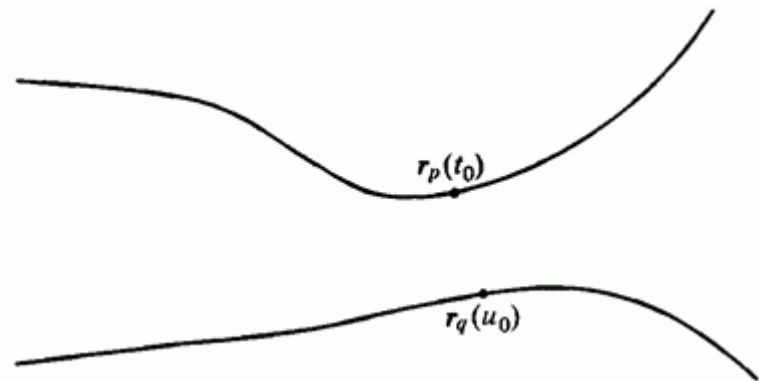


Рис. 11.3. NURBS кривые будут касаться ближайшими точками

Эти точки должны удовлетворять уравнениям имеющим вид

$$\begin{aligned} (r_p(t) - r_q(u)) \cdot \frac{dr_p}{dt} &= 0, \\ (r_q(u) - r_p(t)) \cdot \frac{dr_q}{du} &= 0. \end{aligned} \quad (11.15)$$

Пусть этим точкам соответствуют параметры t_0 и u_0 (рис. 11.3).

Точки на кривых выражаются в виде сумм

$$r_p(t_0) = P_1 p_1 + P_2 p_2 + \dots + P_n p_n = \sum_{i=1}^n P_i p_i,$$

$$r_q(u_0) = Q_1 q_1 + Q_2 q_2 + \dots + Q_k q_k = \sum_{j=1}^k Q_j q_j, \quad (11.16)-(11.17)$$

где

$$P_i = \frac{N_{i,m}(t_0) w_i}{\sum_{r=1}^n N_{r,m}(t_0) w_r}, \quad Q_j = \frac{N_{j,l}(u_0) z_j}{\sum_{r=1}^k N_{r,l}(u_0) z_r}.$$

Пусть точки имеют координаты $p_i = [x_i \ y_i]^T$, $i=1, 2, \dots, n$, $q_j = [a_j \ b_j]^T$, $j=1, 2, \dots, k$. Если бы точки $r_p(t_0)$ и $r_q(u_0)$ совпали, то являлись бы решением задачи касания кривых (11.13) и (11.14). Поэтому в качестве уравнения связи возьмем уравнение

$$|r_p(t_0) - r_q(u_0)| = |(P_1 p_1 + P_2 p_2 + \dots + P_n p_n) - (Q_1 q_1 + Q_2 q_2 + \dots + Q_k q_k)| =$$

$$= \sqrt{\left(\sum_{i=1}^n P_i x_i - \sum_{j=1}^k Q_j a_j\right)^2 + \left(\sum_{i=1}^n P_i y_i - \sum_{j=1}^k Q_j b_j\right)^2} = 0. \quad (7.11.18)$$

(11.18)

Составим функцию суммарного перемещения точек касающихся кривых совместно с уравнением связи (11.18). Эта функция имеет вид

$$F = \frac{1}{2} \left(|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 + \dots + |p_n - p_n^{(0)}|^2 + \right.$$

$$\left. + |q_1 - q_1^{(0)}|^2 + |q_2 - q_2^{(0)}|^2 + \dots + |q_k - q_k^{(0)}|^2 \right) +$$

$$+ \lambda |r_p(t_0) - r_q(u_0)|. \quad (11.19)$$

Систему уравнений для определения положения точек получим из равенства нулю частных производных функции (11.19)

$$\begin{aligned}
 x_i - x_i^{(0)} + \lambda P_i \frac{x_p(t_0) - a_q(u_0)}{\sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2}} &= 0, \quad i = 1, 2, \dots, n, \\
 y_i - y_i^{(0)} + \lambda P_i \frac{y_p(t_0) - b_q(u_0)}{\sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2}} &= 0, \quad i = 1, 2, \dots, n, \\
 a_j - a_j^{(0)} - \lambda Q_j \frac{x_p(t_0) - a_q(u_0)}{\sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2}} &= 0, \quad j = 1, 2, \dots, k, \\
 b_j - b_j^{(0)} - \lambda Q_j \frac{y_p(t_0) - b_q(u_0)}{\sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2}} &= 0, \quad j = 1, 2, \dots, k, \\
 \sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2} &= 0,
 \end{aligned} \tag{11.20}$$

где

$$\begin{aligned}
 x_p(t_0) &= P_1 x_1 + P_2 x_2 + \dots + P_n x_n = \sum_{i=1}^n P_i x_i, \\
 y_p(t_0) &= P_1 y_1 + P_2 y_2 + \dots + P_n y_n = \sum_{i=1}^n P_i y_i, \\
 a_q(u_0) &= Q_1 a_1 + Q_2 a_2 + \dots + Q_k a_k = \sum_{j=1}^k Q_j a_j, \\
 b_q(u_0) &= Q_1 b_1 + Q_2 b_2 + \dots + Q_k b_k = \sum_{j=1}^k Q_j b_j.
 \end{aligned}$$

Система уравнений (11.20) на $(r+1)$ -й итерации метода Ньютона дает приращения искомых функций, равные:

$$\begin{aligned}
 \Delta x_i^{(r+1)} &= -\frac{P_i}{s} \left(\sum_{m=1}^n P_m x_m^{(r)} - \sum_{m=1}^k Q_m a_m^{(r)} \right), & i = 1, 2, \dots, n, \\
 \Delta y_i^{(r+1)} &= -\frac{P_i}{s} \left(\sum_{m=1}^n P_m y_m^{(r)} - \sum_{m=1}^k Q_m b_m^{(r)} \right), & i = 1, 2, \dots, n, \\
 \Delta a_j^{(r+1)} &= \frac{Q_j}{s} \left(\sum_{m=1}^n P_m x_m^{(r)} - \sum_{m=1}^k Q_m a_m^{(r)} \right), & j = 1, 2, \dots, k, \\
 \Delta b_j^{(r+1)} &= \frac{Q_j}{s} \left(\sum_{m=1}^n P_m y_m^{(r)} - \sum_{m=1}^k Q_m b_m^{(r)} \right), & j = 1, 2, \dots, k.
 \end{aligned} \tag{11.21}$$

Значение коэффициента λ на текущей итерации равно

$$\begin{aligned}
 \lambda &= \frac{1}{s} \sqrt{\left(\sum_{i=1}^n P_i x_i - \sum_{j=1}^k Q_j a_j \right)^2 + \left(\sum_{i=1}^n P_i y_i - \sum_{j=1}^k Q_j b_j \right)^2} = \\
 &= \frac{1}{s} \sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2}, \tag{11.22}
 \end{aligned}$$

$$s = \sum_{i=1}^n P_i^2 + \sum_{j=1}^k Q_j^2$$

где . Пока решение не найдено, коэффициент λ остается не равным нулю.

Перемещение всех точек на каждой итерации происходит параллельно касательным к кривым, проходящим через точки $r_p(t_0)$ и $r_q(u_0)$ на расстояние, пропорциональное их вкладу (коэффициенту P_i или Q_j) в точки касания. Перед началом новой итерации необходимо заново вычислить параметры t_0 и u_0 для наиболее близких друг к другу точек на рассматриваемых кривых и коэффициенты P_i и Q_j для новых ближайших точек $r_p(t_0)$ и $r_q(u_0)$.

Следует заметить, что в искомой точке выполняется равенство (11.18), что обращает в нуль числители и знаменатели последних слагаемых уравнений системы (11.20).

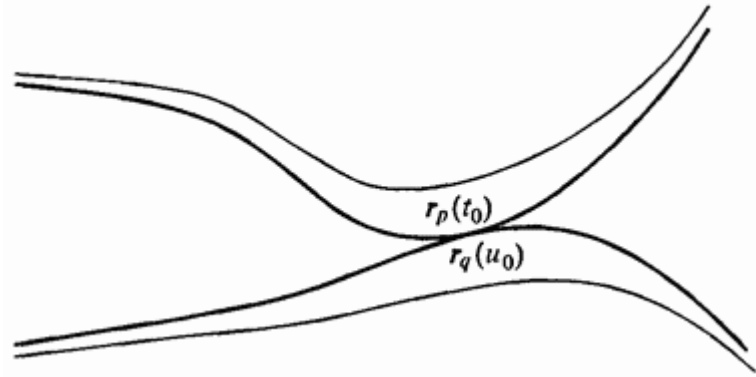


Рис. 11.4. Модификация NURBS кривых при наложении связи касания

Если на очередной r -й итерации выполнилось равенство (11.18), а решение системы нелинейных уравнений еще не закончено, то следует использовать $|r_p(t_0) - r_q(u_0)|^{(r-1)}$, $(x_p(t_0) - x_q(u_0))^{(r-1)}$, $(y_p(t_0) - y_q(u_0))^{(r-1)}$, вычисленные на предыдущей итерации. Поведение кривых приведено на рис. 11.4.

Касание отрезка и окружности.

Решим задачу касания в частных случаях. Касание отрезка $r_1(t)$ и окружности $r_0(u)$,

$$\begin{aligned} r_1(t) &= (1-t)p_1 + tp_2, \\ r_0(u) &= p_0 + i_x r \cos u + i_y r \sin u, \end{aligned} \quad (11.23)$$

описывается уравнением

$$|(1-t_0)p_1 + t_0p_2 - p_0| = r, \quad (11.24)$$

где $p_1=[x_1 \ y_1]^T$, $p_2=[x_2 \ y_2]^T$ — начальная и конечная точки отрезка, r — радиус окружности, $p_0=[x_0 \ y_0]^T$ — центр окружности,

$$t_0 = \frac{(p_0 - p_1) \bullet (p_2 - p_1)}{(p_2 - p_1) \bullet (p_2 - p_1)} \text{ — значение параметра отрезка,}$$

соответствующее проекции центра окружности на отрезок. В данном случае радиус окружности также является варьируемым параметром. Параметр t_0 , соответствующий проекции центра окружности на отрезок, мы будем вычислять на каждой итерации. Систему уравнений для определения параметров связи получим, приравняв нулю частные производные по параметрам функции

$$F = \frac{1}{2} \left(|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 + |p_0 - p_0^{(0)}|^2 + (r - r^{(0)})^2 \right) + \lambda (|(1 - t_0)p_1 + t_0p_2 - p_0| - r), \quad (11.25)$$

где $p_1^{(0)}, p_2^{(0)}, p_0^{(0)}$ — исходные положения точек p_1, p_2, p_0 , $r^{(0)}$ — исходное значение радиуса окружности.

Система уравнений для определения положения точек имеет вид

$$\begin{aligned} x_0 - x_0^{(0)} - \lambda \frac{x_0}{\sqrt{((1 - t_0)x_1 + t_0x_2 - x_0)^2 + ((1 - t_0)y_1 + t_0y_2 - y_0)^2}} &= 0, \\ y_0 - y_0^{(0)} - \lambda \frac{y_0}{\sqrt{((1 - t_0)x_1 + t_0x_2 - x_0)^2 + ((1 - t_0)y_1 + t_0y_2 - y_0)^2}} &= 0, \\ x_1 - x_1^{(0)} + \lambda \frac{(1 - t_0)x_1}{\sqrt{((1 - t_0)x_1 + t_0x_2 - x_0)^2 + ((1 - t_0)y_1 + t_0y_2 - y_0)^2}} &= 0, \\ y_1 - y_1^{(0)} + \lambda \frac{(1 - t_0)y_1}{\sqrt{((1 - t_0)x_1 + t_0x_2 - x_0)^2 + ((1 - t_0)y_1 + t_0y_2 - y_0)^2}} &= 0, \\ x_2 - x_2^{(0)} + \lambda \frac{t_0x_2}{\sqrt{((1 - t_0)x_1 + t_0x_2 - x_0)^2 + ((1 - t_0)y_1 + t_0y_2 - y_0)^2}} &= 0, \\ y_2 - y_2^{(0)} + \lambda \frac{t_0y_2}{\sqrt{((1 - t_0)x_1 + t_0x_2 - x_0)^2 + ((1 - t_0)y_1 + t_0y_2 - y_0)^2}} &= 0, \\ r - r^{(0)} - \lambda &= 0, \\ \sqrt{((1 - t_0)x_1 + t_0x_2 - x_0)^2 + ((1 - t_0)y_1 + t_0y_2 - y_0)^2} - r &= 0. \end{aligned} \quad (11.26)$$

Система уравнений (11.26) решается итерационно. Перед началом новой итерации необходимо заново вычислить параметр t_0 . Поведение отрезка и окружности приведено на рис. 11.5.

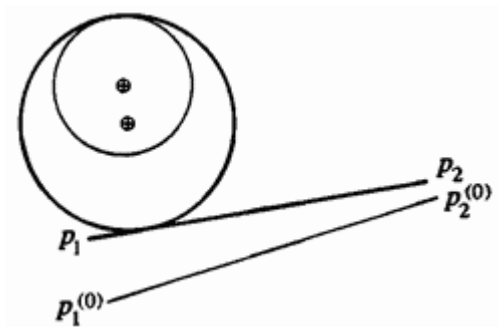


Рис. 11.5. Поведение окружности и отрезка при касании

Касание окружностей.

Касание двух окружностей

$$\begin{aligned} r_1(t_1) &= q_1 + i_x r_1 \cos t_1 + i_y r_1 \sin t_1, \\ r_2(t_2) &= q_2 + i_x r_2 \cos t_2 + i_y r_2 \sin t_2, \end{aligned} \quad (11.27)-(11.28)$$

описывается уравнением

$$|q_1 - q_2| = |r_1 \pm r_2|. \quad (11.29)$$

В данном случае радиусы окружностей также являются варьируемыми параметрами. Уравнение (11.29) связывает шесть параметров.

Систему уравнений для определения параметров связи получим, приравняв нулю частные производные по параметрам функции

$$F = \frac{1}{2} \left(|q_1 - q_1^{(0)}|^2 + |q_2 - q_2^{(0)}|^2 + (r_1 - r_1^{(0)})^2 + (r_2 - r_2^{(0)})^2 \right) + \lambda (|q_1 - q_2| - |r_1 \pm r_2|), \quad (11.30)$$

где $q_1^{(0)}$, $q_2^{(0)}$ — исходные положения центров, $r_1^{(0)}$, $r_2^{(0)}$ — исходное значение радиусов окружностей. Знак \pm должен быть раскрыт в зависимости от положения ближайших точек окружностей. Эти точки

лежат на линии, соединяющей их центры. Пусть окружности расположены вне друг друга. Тогда система уравнений для определения положения центров окружностей и их радиусов имеет вид

$$\begin{aligned} x_1 - x_1^{(0)} + \lambda \frac{x_1 - x_2}{d} &= 0, & y_1 - y_1^{(0)} + \lambda \frac{y_1 - y_2}{d} &= 0, \\ x_2 - x_2^{(0)} - \lambda \frac{x_1 - x_2}{d} &= 0, & y_2 - y_2^{(0)} - \lambda \frac{y_1 - y_2}{d} &= 0, \\ r_1 - r_1^{(0)} - \lambda &= 0, & r_2 - r_2^{(0)} - \lambda &= 0, & |q_1 - q_2| - (r_1 + r_2) &= 0, \end{aligned} \quad (11.31)$$

где $d = |q_1 - q_2|$. Эта система имеет решение

$$\begin{aligned} x_1 &= x_1^{(0)} - \frac{1}{4} a^{(0)} \frac{x_1^{(0)} - x_2^{(0)}}{d^{(0)}}, & y_1 &= y_1^{(0)} - \frac{1}{4} a^{(0)} \frac{y_1^{(0)} - y_2^{(0)}}{d^{(0)}}, \\ x_2 &= x_2^{(0)} + \frac{1}{4} a^{(0)} \frac{x_1^{(0)} - x_2^{(0)}}{d^{(0)}}, & y_2 &= y_2^{(0)} + \frac{1}{4} a^{(0)} \frac{y_1^{(0)} - y_2^{(0)}}{d^{(0)}}, \\ r_1 &= r_1^{(0)} + \frac{1}{4} a^{(0)}, & r_2 &= r_2^{(0)} + \frac{1}{4} a^{(0)}, & \lambda &= \frac{1}{4} a^{(0)}, \end{aligned} \quad (11.32)$$

где $d^{(0)} = |q_1^{(0)} - q_2^{(0)}|$, $a^{(0)} = |q_1^{(0)} - q_2^{(0)}| - (r_1^{(0)} + r_2^{(0)})$. Из решения следует, что радиусы изменяются на одинаковую величину, равную перемещению центров окружностей.

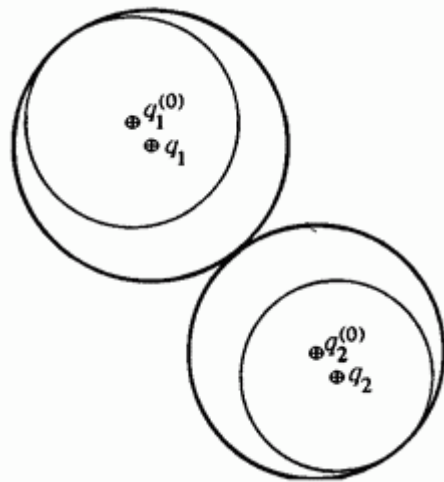


Рис. 11.6. Касание окружностей

Поведение окружностей приведено на рис. 11.6. При расположении окружностей одна внутри другой в уравнениях (11.31) изменяется знак перед одним из радиусов.

5.12. Формирование и решение системы уравнений связей

Вариационные связи являются мощным средством управления геометрическими объектами. Они позволяют редактировать геометрические объекты совместно, когда изменение параметра одного из объектов влечет за собой соответствующие изменения параметров других объектов. Параметром может служить любая величина из структуры данных геометрического объекта или специально введенная переменная. Специально введенные переменные используются для удобства редактирования геометрических объектов с помощью вариационных связей.

С помощью вариационных связей осуществляется управление сразу всеми связанными объектами. Как правило, на одни и те же параметры наложено несколько вариационных связей. Каждая связь представляет собой некоторый вариационный объект, который изменяет параметры геометрических объектов. Различные типы вариационных связей геометрических объектов приведены на рис. 12.1. Каждая связь имеет одно или несколько уравнений и перечень параметров, участвующих в каждом уравнении.

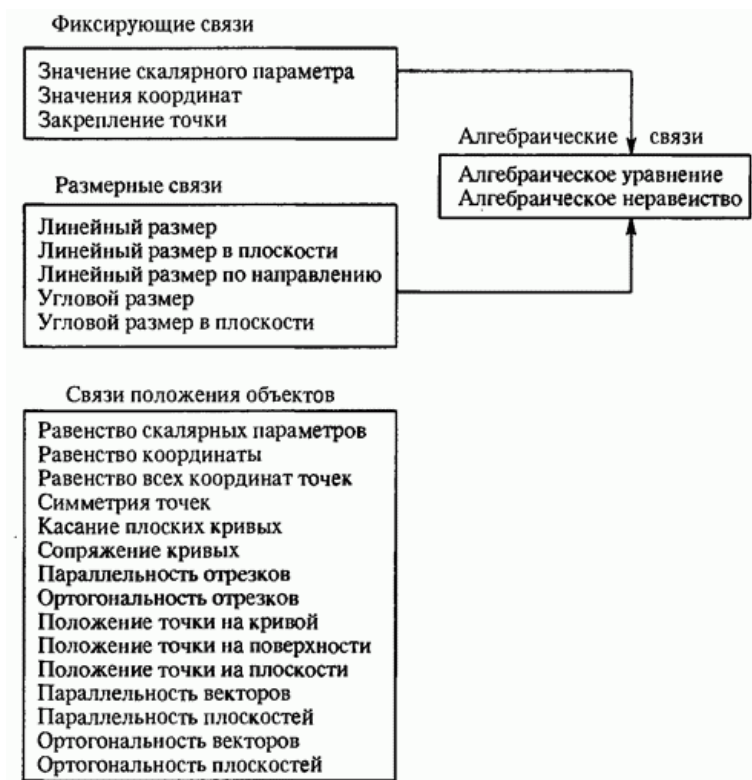


Рис. 12.1. Типы вариационных связей

Нормальным состоянием для вариационных связей является состояние, когда их уравнения выполняются. Будем называть это состояние равновесным. В процессе моделирования приходится модифицировать геометрические объекты или их взаимное положение. Таким образом, вариационные связи выводятся из состояния равновесия.

В результате все или часть уравнений вариационных связей перестает выполняться. Далее следует найти новые значения варьируемых параметров, чтобы уравнения связей опять выполнялись. Для того чтобы связанные параметры удовлетворяли уравнениям связей, требуется сформировать систему уравнений и решить ее. В процессе решения этой системы уравнений необходимо модифицировать все

На практике решение системы уравнений является самой сложной и тонкой проблемой при наложении вариационных связей. В разных случаях для решения системы уравнений используются разные подходы: аналитический, конструктивный, численный.

Аналитический подход используется для двумерных вариационных связей, когда искомое положение двумерных геометрических объектов можно найти с помощью «линейки и компаса». В некоторых простых случаях аналитический подход применим и для трехмерных объектов.

При наложении вариационных связей на положения твердых тел относительно друг друга эффективным является конструктивный подход. Конструктивный подход привлекает теорию графов. Для системы уравнений вариационных связей строится граф. Узлами графа являются геометрические объекты, а ребрами графа являются уравнения связей. На основе графа геометрические объекты и вариационные связи делятся на кластеры (группы). Кластер в модели ведет себя, как нечто жесткое целое. В результате задача сводится к задаче позиционирования геометрических объектов кластеров. Сначала определяется удовлетворяющее вариационным связям положение геометрических объектов для каждого кластера в отдельности, а затем кластеры последовательно стыкуются между собой. Размещение геометрических объектов внутри одного кластера осуществляется путем решения соответствующей системы уравнений. Стыковка кластеров между собой выполняется на основе анализа степеней свободы. Стыковка кластеров производится последовательно. Будем считать, что положение геометрических объектов некоторого кластера задано. Расположим геометрические объекты соседнего кластера так, чтобы удовлетворялись связывающие их уравнения. Тем самым мы состыкуем два кластера. Продолжим стыковку кластеров до тех пор, пока не будут выполняться все уравнения связей. В процессе стыковки кластеров возможно потребуются перебирать различные варианты и численно решать систему уравнений связей нескольких кластеров. Конструктивный подход уменьшает размерность системы уравнений, но его можно использовать для строго определенного набора геометрических объектов и вариационных связей.

Численный подход заключается в решении системы уравнений вариационных связей численными итерационными методами. При использовании этого подхода следует помнить, что вариационные

связи могут иметь несколько решений. Численный метод позволяет найти одно из этих решений, и оно может не совпадать с искомым. Найденное решение зависит от начального приближения, с которого начинается итерационный процесс. Чтобы начальное приближение находилось в области сходимости к искомому решению, следует не допускать сильных отклонений вариационных связей от их равновесных состояний. Например, если вариационные связи выведены из состояния равновесия путем перемещения некоторой точки геометрического объекта, то решение следует искать не для окончательного положения точки, а для нескольких промежуточных ее положений, постепенно сдвигая точку от исходного положения к конечному. Аналогично, при изменении значения некоторого размера следует последовательно искать решения для нескольких промежуточных значений этого размера, постепенно переходя от старого к новому значению. Численный подход чаще используется для решения двухмерных задач.

При использовании численного подхода желательно, чтобы ненулевые приращения параметров были одного порядка. Например, если одна часть варьируемых параметров является угловыми величинами, другая часть — компонентами векторов, а третья часть параметров — координатами точек, то желательно, чтобы их значения лежали в одних и тех же пределах. Для достижения этого используется нормирование варьируемых параметров.

Численное решение системы уравнений связей может производиться несколько иначе. Используемый критерий поведения геометрических объектов стремится сохранить параметры геометрических объектов ближе к их исходному состоянию. Воспользуемся этим свойством для уменьшения размерности системы уравнений. Попытаемся решить систему уравнений, зафиксировав еще некоторые из варьируемых параметров, кроме тех, которые зафиксированы уравнениями связей. Если нам это удастся, то мы будем придерживаться правила, которое заключается в сохранении объектов как можно ближе к их исходному состоянию, и уменьшим число неизвестных и уравнений в (12.4). Остается выяснить, какие из параметров можно дополнительно зафиксировать. Некоторые из уравнений связей могут быть удовлетворены при исходных значениях параметров. Определим эти уравнения и составим список варьируемых параметров, участвующих в них. Попробуем решить полную систему уравнений, считая параметры этого списка фиксированными. Если это удастся, попробуем расширить этот список, если нет, то будем сужать этот список до тех

пор, пока не удастся решить полную систему уравнений. Таким образом, нам, возможно, удастся найти новые значения параметров, сохранив некоторым из них исходные значения. Если даже при отсутствии дополнительной фиксации параметров систему уравнений (12.1) и (12.4) решить не удастся, то следует вернуться к исходному равновесному состоянию геометрических объектов и констант уравнений, при которых уравнения связей удовлетворяются.

В процессе решения уравнения связей изменяют значения параметров и тем самым перестраивают геометрические объекты в соответствии с новыми значениями параметров. Объекты будут перестроены в соответствии с наложенными связями, когда все уравнения полной системы будут удовлетворены с требуемой точностью.

6. КОМПЬЮТЕРНАЯ ГРАФИКА

6.1. Визуализация геометрических объектов

Одним из наиболее ярких применений геометрического моделирования является компьютерная графика. Исходная информация для получения изображения некоторого объекта на экране компьютера поставляется геометрической моделью этого объекта. На практике используются различные способы визуализации геометрических объектов. Наиболее простой из них — отображение линиями. Он позволяет получить общее представление об объекте, его размерах и форме. С помощью компьютера можно получить изображение моделируемого объекта, близкое к его фотографии. Геометрической модели или ее частям можно придать любой цвет и освещенность. Кроме того, объект на экране можно заставить двигаться в реальном времени и тем самым получить видеофильм.

Для того чтобы увидеть, как выглядит объект, нужно смоделировать поведение потока лучей света, идущего от источников к модели и от поверхности модели к наблюдателю. При этом граням модели можно придать необходимый цвет, зеркальность, прозрачность, излучение, фактуру и другие физические свойства взаимодействия со световыми потоками. Модель можно осветить с разных сторон светом различного цвета и интенсивности. Реалистические отображения объектов строятся из отдельных точек определенного цвета и яркости, причем точки должны быть сравнительно небольшой величины и

располагаться достаточно близко друг к другу (расстояние между точками не должно превосходить размеры точек).

Изображения геометрических объектов линиями называются векторными, а изображения геометрических объектов точками называются растровыми. Примеры векторного и растрового изображений приведены на рис. 1.1 и 1.2.

Растровая графика более информативна, чем векторная, так как позволяет получать тоновые изображения, но требует больше ресурсов и склонна к искажению при редактировании. В обоих случаях мы видим проекции геометрических объектов на выбранную плоскость.

Проекции могут быть построены с помощью линий, перпендикулярных проекционной плоскости, или с помощью линий, проходящих через общую точку. Первые называются параллельными проекциями (рис. 1.3), вторые — центральными проекциями или перспективными изображениями (рис. 1.4). Центральные проекции могут быть созданы на плоскости или криволинейной поверхности. Мы рассмотрим центральные проекции на плоскости.

Центральные проекции геометрических объектов ближе к тому изображению, которое возникает на сетчатке глаза, поэтому они дают более реалистические картины.

Хотя центральные проекции ближе к тому, что можно наблюдать в реальности, размеры отдельных частей объекта на центральных проекциях зависят от их расстояния до точки наблюдения, поэтому в качестве конструкторской документации чаще используются параллельные проекции.

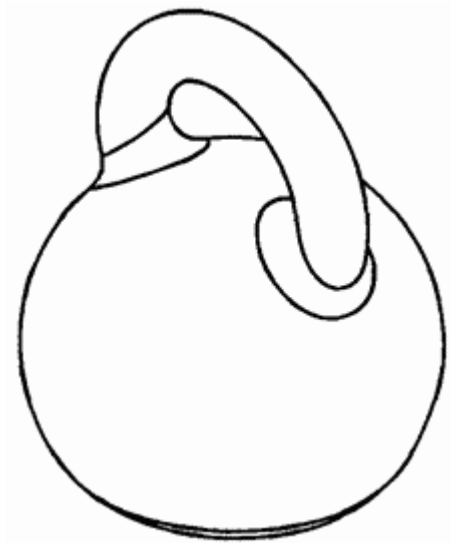


Рис. 1.1. Векторное изображение

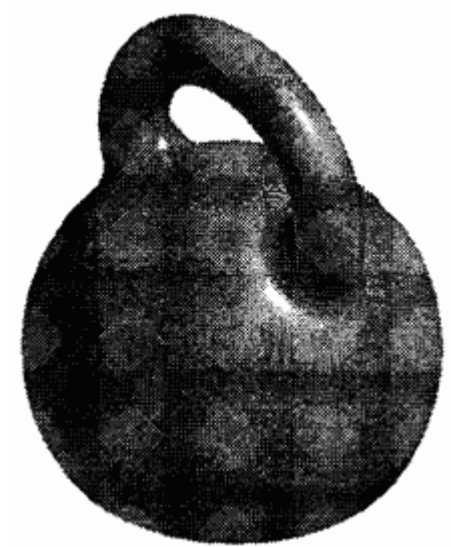


Рис. 1.2. Растровое изображение

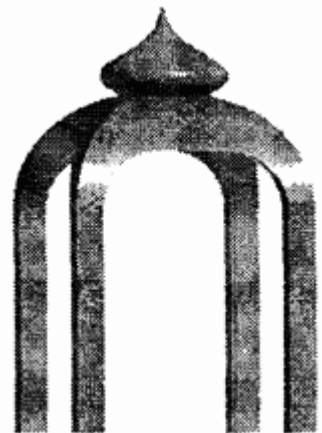


Рис. 1.3. Параллельная проекция

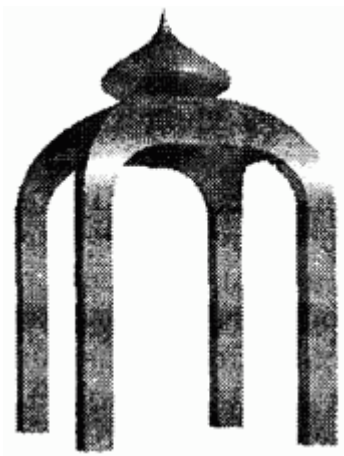


Рис. 1.4. Центральная проекция

Растровая и векторная графика, а также параллельные и центральные проекции не являются альтернативами. Они дополняют друг друга. В данной главе мы рассмотрим методы визуализации геометрических объектов с помощью компьютера или методы компьютерной графики.

6.2. Параллельные проекции на плоскость

Параллельные проекции геометрических объектов используются для получения чертежей, схем и других документов, где требуется сохранить определенную геометрическую точность относительных размеров и взаимного расположения отдельных частей объектов.

Пусть требуется построить параллельную проекцию геометрических объектов на плоскость, которая определяется точкой q и двумя взаимно ортогональными векторами единичной длины \mathbf{i}_x и \mathbf{i}_y . Проекционная плоскость может быть связана с чертежом, экраном компьютера или каким-либо другим устройством вывода.

Систему координат, в которой описаны отображаемые объекты, будем называть глобальной системой координат. Свяжем с проекционной плоскостью местную систему координат $Qxyz$. Пусть начальная точка q местной системы координат в глобальной системе координат описывается радиус-вектором $\mathbf{q}=[q_1 \ q_2 \ q_3]^T$, а орты \mathbf{i}_x , \mathbf{i}_y , $\mathbf{i}_z = \pm \mathbf{i}_x \times \mathbf{i}_y$ местной системы координат в глобальной системе координат описываются векторами $\mathbf{i}_x = [x_1 \ x_2 \ x_3]^T$, $\mathbf{i}_y = [y_1 \ y_2 \ y_3]^T$, $\mathbf{i}_z = [z_1 \ z_2 \ z_3]^T$. Местная система координат может быть как правой ($\mathbf{i}_z = \mathbf{i}_x \times \mathbf{i}_y$), так и левой ($\mathbf{i}_z = -\mathbf{i}_x \times \mathbf{i}_y$).

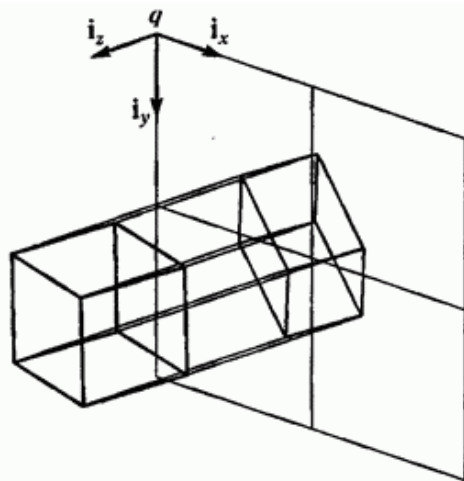


Рис..2.1. Проекционная плоскость

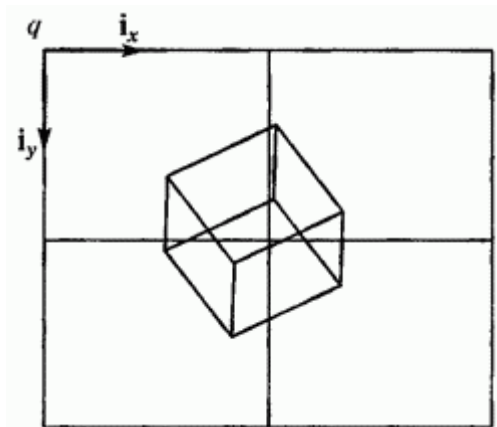


Рис. 2.2. Проекция объекта

На рис. 2.1 приведен пример построения местной системы координат и параллельных проекций ребер прямого параллелепипеда на плоскости, определяемой радиус-вектором точки q и ортами \mathbf{i}_x , \mathbf{i}_y , \mathbf{i}_z . На рис. 2.2 приведены параллельные проекции ребер прямого параллелепипеда на плоскости.

Проекция точки.

Рассмотрим произвольную точку $\mathbf{r} = [r_1 \ r_2 \ r_3]^T$. Определим двухмерную точку $\mathbf{p} = [x \ y]^T$ являющуюся ее проекцией на проекционную плоскость

$$\mathbf{p}(x, y) = \mathbf{q} + x\mathbf{i}_x + y\mathbf{i}_y. \quad (2.1)$$

Параллельную проекцию произвольной точки \mathbf{r} отображаемых объектов получим, опустив из нее перпендикуляр на проекционную плоскость. Пусть точка $\mathbf{r} = [r_1 \ r_2 \ r_3]^T$ в местной системе координат имеет координаты x , y , z . Координаты связаны соотношениями, которые в данном случае имеют вид

$$\begin{aligned} x &= x_1(r_1 - q_1) + x_2(r_2 - q_2) + x_3(r_3 - q_3) = x_1r_1 + x_2r_2 + x_3r_3 + o_x, \\ y &= y_1(r_1 - q_1) + y_2(r_2 - q_2) + y_3(r_3 - q_3) = y_1r_1 + y_2r_2 + y_3r_3 + o_y, \\ z &= z_1(r_1 - q_1) + z_2(r_2 - q_2) + z_3(r_3 - q_3) = z_1r_1 + z_2r_2 + z_3r_3 + o_z, \end{aligned} \quad (2.2)$$

где

$$\begin{aligned} o_x &= -x_1q_1 - x_2q_2 - x_3q_3, \\ o_y &= -y_1q_1 - y_2q_2 - y_3q_3, \\ o_z &= -z_1q_1 - z_2q_2 - z_3q_3. \end{aligned}$$

В однородных координатах преобразование (2.2) описывается равенством

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & o_x \\ y_1 & y_2 & y_3 & o_y \\ z_1 & z_2 & z_3 & o_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} = \mathbf{A}_0 \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix}. \quad (2.3)$$

Коэффициенты расширенной матрицы преобразования выражаются через компоненты векторов, определяющих положение проекционной плоскости.

Проекция точки на плоскость (2.1) описывается первыми двумя координатами вектора (2.3). Таким образом, первые две координаты расширенного радиус-вектора, полученного по формуле (2.3), являются координатами искомой параллельной проекцией $\mathbf{r} = [x \ y]^T$ рассматриваемой точки.

Масштаб проекции.

С помощью преобразования (2.3) может быть получена проекция точек линий в масштабе 1:1, если единица измерения длины для устройства вывода равны единице измерения пространства, в котором построена геометрическая модель. Вывод на экран компьютера производится в экранных единицах — пикселах. Пусть при построении геометрической модели использовалась единица длины — миллиметр. Если размер пиксела не равен миллиметру, то для получения на экране

изображения геометрического объекта в масштабе 1:1 радиус-вектор каждой точки проекции следует умножить на масштабную единицу устройства вывода, равную количеству пикселей в одном миллиметре. Масштабная единица может быть различной для координаты x и координаты y . В последнем случае радиус-вектор $\mathbf{r} = [x \ y]^T$ точки проекции необходимо умножить на матрицу

$$M = \begin{bmatrix} m_x & 0 \\ 0 & m_y \end{bmatrix}. \quad (2.4)$$

Если требуется получить изображение объекта в масштабе $m:1$, то радиус-вектор двухмерной точки проекции следует преобразовать по формуле (1.3.17). Таким образом, координаты x' и y' проекции некоторой точки \mathbf{r} геометрического объекта с учетом масштаба равны

$$\begin{aligned} \begin{bmatrix} x' \\ y' \end{bmatrix} &= \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + m \cdot \left(\begin{bmatrix} m_x & 0 \\ 0 & m_y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \right) = \\ &= \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + m \cdot \left(\begin{bmatrix} m_x & 0 \\ 0 & m_y \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 & x_3 & o_x \\ y_1 & y_2 & y_3 & o_y \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \right), \end{aligned} \quad (2.5)$$

где x_0, y_0 — координаты точки, которая при масштабировании должна остаться неподвижной. Вместо матрицы A_0 в (2.5) используется матрица, содержащая ее первые две строки.

Мы видим, что для получения координат x и y плоской проекции некоторой точки $\mathbf{r} = [r_1 \ r_2 \ r_3]^T$ достаточно знать положение проекционной плоскости в пространстве (орты плоскости проекции $\mathbf{i}_x = [x_1 \ x_2 \ x_3]^T$, $\mathbf{i}_y = [y_1 \ y_2 \ y_3]^T$, $\mathbf{i}_z = \pm \mathbf{i}_x \times \mathbf{i}_y$, положение центра плоскости $\mathbf{q} = [q_1 \ q_2 \ q_3]^T$, масштабные единицы m_x и m_y устройства вывода и масштаб отображения m).

6.3. Центральные проекции на плоскость

В параллельной проекции на плоскость каждая точка геометрического объекта проецируется на заданную плоскость вдоль одного и того же направления независимо от положения точки в пространстве. Если размеры объекта соизмеримы с расстоянием до объекта, то

параллельная проекция объекта будет отличаться его фотоизображения. **Вид также будет отличаться и от фотографии объекта. Это происходит потому, что изображение на сетчатке глаза и на фотографии получено не с помощью параллельной проекции.** Реальное изображение близко к центральной проекции объекта на плоскость. Центральные проекции называют также перспективными изображениями. При графическом отображении геометрической модели центральные проекции дают более реалистичную картину, чем параллельные проекции. Центральная проекция представляет собой выполненную по определенным правилам проекцию модели объекта на плоскость и также как и параллельная проекция является плоским объектом и строится с помощью двухмерных линий.

Пусть требуется построить центральную проекцию геометрических объектов на плоскость, которая определяется точкой q и двумя взаимно ортогональными векторами единичной длины $\mathbf{i}_x, \mathbf{i}_y$. Свяжем с проекционной плоскостью местную систему координат $Qxuz$. Пусть начальная точка q местной системы координат в глобальной системе координат описывается радиус-вектором $\mathbf{q}=[q_1 \ q_2 \ q_3]^T$, а орты $\mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z = \pm \mathbf{i}_x \times \mathbf{i}_y$ местной системы координат в глобальной системе координат описываются векторами $\mathbf{i}_x = [x_1 \ x_2 \ x_3]^T, \mathbf{i}_y = [y_1 \ y_2 \ y_3]^T, \mathbf{i}_z = [z_1 \ z_2 \ z_3]^T$.

Для построения центральной проекции на плоскость требуется знать еще точку пространства, из которой ведется наблюдение. Эту точку будем называть точкой наблюдения (точка наблюдения для параллельной проекции находится в бесконечности в положительном направлении нормали к проекционной плоскости). Пусть точкой наблюдения является точка W . Проекцию точки W на проекционную плоскость обозначим через F .

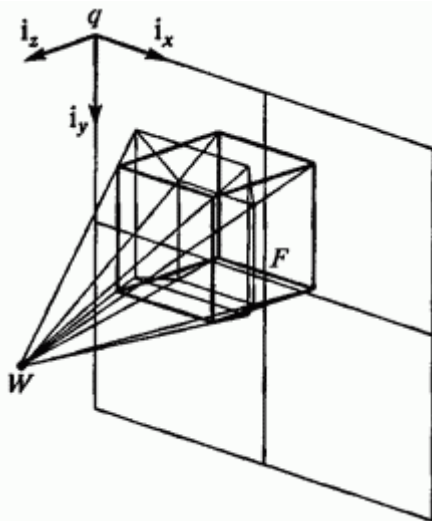


Рис. 3.1. Построение центральной проекции

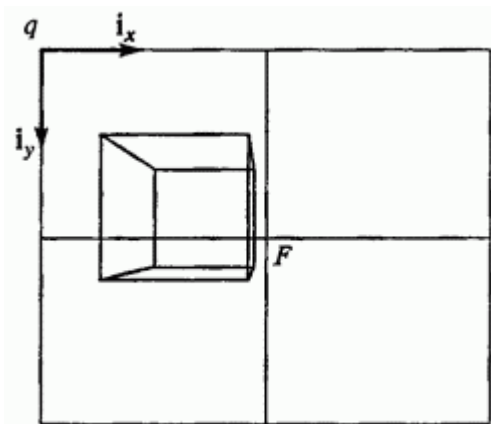


Рис. 3.2. Центральная проекция объекта

На рис. 3.1 приведены местная система координат, точка наблюдения и центральные проекции ребер прямого параллелепипеда. На рис. 3.2 приведены центральные проекции ребер прямого параллелепипеда на плоскости. Если провести плоскость через точку наблюдения,

параллельную плоскости проекции, то проецируемый объект должен полностью лежать в той же части пространства, что и плоскость проекции.

Проекция точки.

Рассмотрим построение центральной проекции одной из вершин прямого параллелепипеда на плоскость (2.1). На рис. 3.3 рассматриваемая точка обозначена через R , а ее центральная проекция обозначена через P . Центральная проекция произвольной точки R строится следующим образом. Построим прямую линию, проходящую через точку наблюдения W и проецируемую точку R . Точка P , в которой эта прямая или ее продолжение пересекают проекционную плоскость, и является центральной проекцией точки R ,

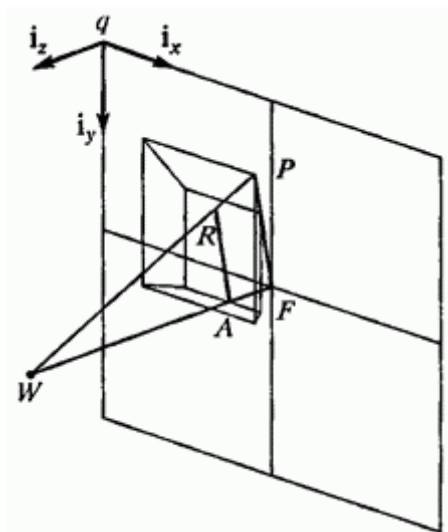


Рис. 3.3. Построение центральной проекции точки R

Радиус-вектор точки R обозначим через \mathbf{r} , а радиус-вектор ее центральной проекции — через \mathbf{p} . Пусть проецируемая точка в глобальной прямоугольной декартовой системе координат описывается радиус-вектором с координатами $\mathbf{r} = [r_1 \ r_2 \ r_3]^T$, а в системе координат $Qxyz$, связанной с проекционной плоскостью, эта точка имеет координаты x, y, z . Координаты точки \mathbf{r} в этих координатных системах

связаны соотношением (2.3). Обозначим радиус-вектор точки наблюдения через \mathbf{w} . Пусть в местной системе координат он имеет координаты x_w, y_w, z_w . Расстояние от точки наблюдения W до ее проекции на плоскость F равно z_w . Построим отрезок RA , параллельный отрезку PF . Длина отрезка WA равна $z_w - z$.

Рассмотрим на рис. 3.3 два подобных треугольника: WFP и WAR . Из подобия треугольников следует, что $PW/FW = RW/AW$. Запишем последнее равенство в векторном виде:

$$\frac{\mathbf{p} - \mathbf{w}}{z_w} = \frac{\mathbf{r} - \mathbf{w}}{z_w - z}, \quad (3.1)$$

Используя это равенство, найдем радиус-вектор \mathbf{r} центральной проекции точки

$$\mathbf{p} = \mathbf{w} + (\mathbf{r} - \mathbf{w}) \frac{z_w}{z_w - z} = \mathbf{r} \frac{z_w}{z_w - z} - \mathbf{w} \frac{z}{z_w - z}. \quad (3.2)$$

Чтобы получить координаты точки \mathbf{p} в проекционной плоскости, нужно с помощью равенства (2.3) перевести вектор \mathbf{r} из глобальной системы координат в местную систему координат. Первые две координаты вектора \mathbf{p} в проекционной системе координат и будут координатами центральной проекции точки \mathbf{r} на плоскости (2.1). Запишем равенство (9.3.2) в местной системе координат:

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = \frac{z_w}{z_w - z} \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \frac{z}{z_w - z} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \begin{bmatrix} \frac{xz_w - zx_w}{z_w - z} \\ \frac{yz_w - zy_w}{z_w - z} \\ 0 \end{bmatrix}. \quad (3.3)$$

Таким образом, декартовы координаты двухмерной точки $\mathbf{p} = [x_p, y_p]^T$ являющейся центральной проекцией рассматриваемой точки \mathbf{r} в пространстве, на плоскости (2.1) равны

$$x_p = \frac{xz_w - zx_w}{z_w - z}, \quad y_p = \frac{yz_w - zy_w}{z_w - z}. \quad (3.4)$$

Если удалить точку наблюдения на бесконечное расстояние от проекционной плоскости, то $z_w \rightarrow \infty$ и $x_p \rightarrow x, y_p \rightarrow y$.

В расширенном матричном виде с учетом (2.3) равенство (3.3) можно записать следующим образом:

$$\begin{aligned} \begin{bmatrix} x_p d \\ y_p d \\ z_p d \\ d \end{bmatrix} &= \begin{bmatrix} z_w & 0 & -x_w & 0 \\ 0 & z_w & -y_w & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & z_w \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \\ &= \begin{bmatrix} z_w & 0 & -x_w & 0 \\ 0 & z_w & -y_w & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & z_w \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 & x_3 & o_x \\ y_1 & y_2 & y_3 & o_y \\ z_1 & z_2 & z_3 & o_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} \end{aligned} \quad (3.5)$$

Напомним, что пространственные координаты вектора равны первым трем компонентам расширенного вектора, деленным на его четвертую компоненту. В силу предположения, что проекционная плоскость и рассматриваемая точка расположены с одной стороны от точки наблюдения, величина $d = z_w - z$ всегда положительна. Формула (3.5) описывает преобразование координат из глобальной системы координат в двухмерную систему координат на проекционной плоскости. Это является искусственным преобразованием, так как третья координата центральной проекции им не определяется (положена равной нулю). Оно не имеет обратного преобразования, так как определитель одной из матриц равен нулю.

Матрица преобразования.

Центральную проекцию объекта можно получить с помощью его преобразования по матрице

$$A_w = \begin{bmatrix} x_1 & y_1 & z_1 & q_x \\ x_2 & y_2 & z_2 & q_y \\ x_3 & y_3 & z_3 & q_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} z_w & 0 & -x_w & 0 \\ 0 & z_w & -y_w & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & z_w \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 & x_3 & o_x \\ y_1 & y_2 & y_3 & o_y \\ z_1 & z_2 & z_3 & o_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3.6)$$

После такого преобразования объект исказится так, что его параллельная проекция на плоскость (2.1) будет совпадать с его центральной проекцией на эту же плоскость. Определитель матрицы A_w не равен нулю, и преобразование (3.6) обратимо.

Для того, чтобы учесть масштабные единицы устройства вывода и масштаб проекции, двухмерный радиус-вектор $\mathbf{p}=[x_p, y_p]^T$ нужно умножить на матрицу (2.4) и масштаб изображения.

На рис. 3.4 и 3.5 приведены перспективные изображения геометрического объекта.

Для обоих рисунков использовалась одна и та же проекционная плоскость, но на рис. 3.4 проекция точки наблюдения лежит слева и выше центра рисунка, а на рис. 3.5 проекция точки наблюдения лежит правее центра рисунка.

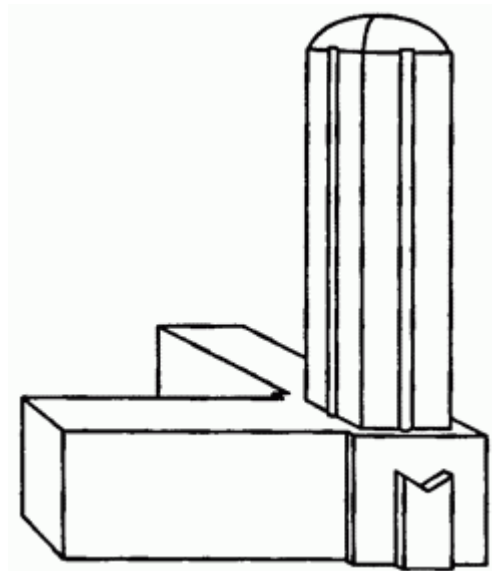


Рис. 3.4. Центральная проекция объекта

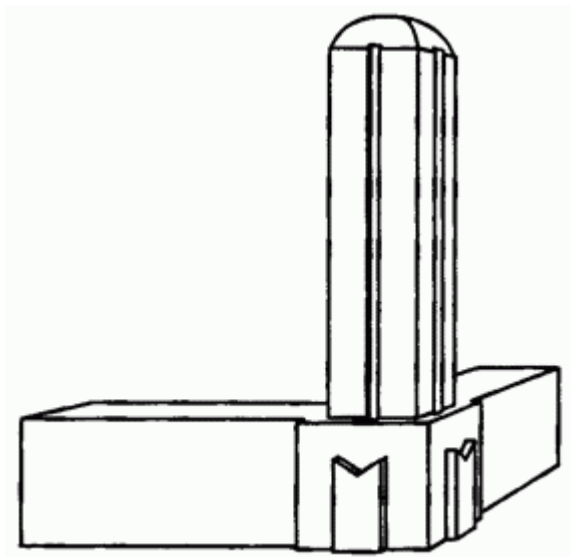


Рис. 3.5. Центральная проекция объекта

Чем дальше от точки наблюдения находится объект, тем меньше его перспективное изображение.

6.4. Полигоны кривых и поверхностей

Проекция на плоскость пространственного отрезка является отрезком на плоскости, а параллельная проекция на плоскость пространственной ломаной линии является ломаной линией на плоскости. Для того, чтобы получить изображение кривой линии, мы будем использовать проекцию полигона кривой.

Для каждой кривой линии мы построим набор точек p , принадлежащих кривой и следующих в порядке возрастания параметра. Соединив эти точки отрезками, получим ломаную линию, аппроксимирующую заданную кривую линию. Набор точек для построения аппроксимирующей ломаной будем называть полигоном кривой. Точки полигона будем вычислять достаточно близко друг от друга, чтобы линии изображения казались плавными.

Определение шага кривой.

Рассмотрим вычисление шага аппроксимации кривой линии. Пусть задана кривая $\mathbf{r}(t)$ и пусть мы находимся в некоторой ее точке t_0 являющейся точкой полигона. Требуется найти параметр t_1 следующей точки полигона, так чтобы отклонение кривой от ее полигона не превышало заданную величину δ . Предположим, что в ближайшей окрестности кривая близка к своей соприкасающейся в этой точке окружности (рис. 4.1).

Радиус ρ соприкасающейся окружности равен радиусу кривизны кривой и определяется формулой

$$\rho = \frac{|\mathbf{r}'|^3}{|\mathbf{r}' \times \mathbf{r}''|}. \quad (4.1)$$

По теореме Пифагора квадрат половины длины хорды равен $\rho^2 - (\rho - \delta)^2 = \delta(2\rho - \delta)$. Длина всей хорды будет равна

$$h = 2\sqrt{\delta(2\rho - \delta)}. \quad (4.2)$$

Если кривая близка к дуге соприкасающейся окружности и угол дуги достаточно мал, то можно считать, что длина хорды примерно равна длине дуги кривой и приращение параметра Δt на длине дуги кривой можно принять равным

$$\Delta t \approx \frac{h}{|\mathbf{r}'|} = 2 \frac{\sqrt{\delta(2\rho - \delta)}}{|\mathbf{r}'|}. \quad (4.3)$$

Таким образом, параметр следующей точки полигона может быть принят равным $t_1 = t_0 + \Delta t$. При резком изменении направления или длины второй производной в точке t_1 по сравнению с предыдущей точкой формула (4.3) дает ошибку, т. е. прогиб может оказаться значительно больше δ . В этом случае шаг должен быть уточнен путем

определения среднего радиуса кривизны кривой на рассматриваемом участке.

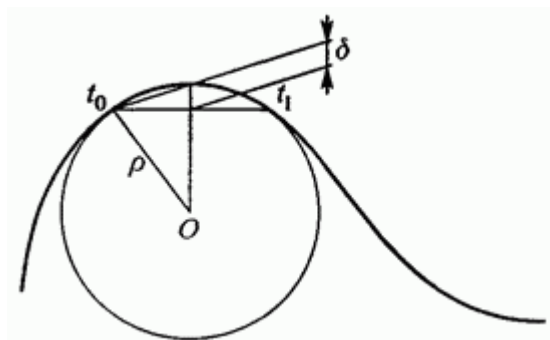


Рис. 4.1. Вычисление шага по параметру кривой по заданному прогибу δ

Формула вычисления параметрического шага между точками полигона (4.3) позволяет построить полигон кривой, отстоящий от самой кривой на величину, не превышающую δ . Проекция этого полигона будет отстоять от проекции кривой также на величину, не превышающую δ , при масштабе отображения 1:1 ($m=1$). Величину δ будем называть точностью отображения. Для произвольного масштаба отображения m :1 проекция полигона кривой будет отстоять от проекции кривой на величину, не превышающую $m\delta$. Таким образом, если мы хотим получить проекцию кривой в масштабе m :1 точностью отображения δ , то мы должны построить полигон, точки которого находятся на параметрической расстоянии, не превышающем

$$\Delta t \approx 2 \frac{\sqrt{(\delta/m)(2\rho - \delta/m)}}{|\mathbf{r}'|}. \quad (4.4)$$

Шаг по параметру кривой мы вычисляли при выполнении операций. Шаг определяется углом отклонения кривой и не зависит от ее размеров и масштаба отображения. Шаг (4.4) определяется точностью отображения и зависит от размеров кривой и масштаба отображения.

Определение шага кривой на поверхности.

Шаг для полигона кривой линии на поверхности вычислим аналогично вычислению шага полигона кривой. Пусть точка некоторого полигона поверхности $\mathbf{r}(u_0, v_0)$ определяется параметрами u_0, v_0 . Найдем параметры u_1, v_1 следующей точки поверхности так, чтобы пространственный отрезок из точки $\mathbf{r}(u_0, v_0)$ в точку $\mathbf{r}(u_1, v_1)$ отстоял от поверхности на расстоянии, не превышающем заданной величины δ . Для этого вычислим радиус кривизны линии на поверхности, имеющей направление $du:dv$. Кривизна линии на поверхности определяется формулой (1.8.5), а ее радиус кривизны определится равенством

$$\rho = \frac{g_{11} du^2 + 2g_{12} du dv + g_{22} dv^2}{b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2}. \quad (4.5)$$

Подставим (4.5) в (4.4) и получим приращения для параметров соседней точки полигона

$$\Delta u \approx 2 \frac{\sqrt{(\delta/m)(2\rho - \delta/m)}}{|\mathbf{r}_1 + \mathbf{r}_2(dv/du)|}, \quad \Delta v \approx 2 \frac{\sqrt{(\delta/m)(2\rho - \delta/m)}}{|\mathbf{r}_1(du/dv) + \mathbf{r}_2|}. \quad (4.6)$$

где $\mathbf{r}_1 = \partial \mathbf{r} / \partial u$ и $\mathbf{r}_2 = \partial \mathbf{r} / \partial v$, m — масштаб отображения. В частных случаях для u -полигонов получим $\rho_u = g_{11}/b_{11}$, а для v -полигонов — g_{22}/b_{22} . Соответственно ненулевые шаги вдоль координатных u -линий и v -линий поверхности равны

$$\Delta u \approx 2 \frac{\sqrt{(\delta/m)(2(g_{11}/b_{11}) - \delta/m)}}{|\partial \mathbf{r} / \partial u|}, \quad \Delta v = 0,$$

$$\Delta u = 0, \quad \Delta v \approx 2 \frac{\sqrt{(\delta/m)(2(g_{22}/b_{22}) - \delta/m)}}{|\partial \mathbf{r} / \partial v|}, \quad (4.7)$$

где $g_{11}, g_{12}, g_{21}, g_{22}, b_{11}, b_{12}, b_{21}, b_{22}$ — коэффициенты первой (7.8) и второй (7.21) основных квадратичных форм поверхности. Как можно заметить, шаги вычисляются приближенно. Это делается с целью

уменьшения вычислительных затрат. При необходимости шаги могут уточняться.

Величина прогиба δ выбирается достаточно малой, чтобы не было заметно изломов проекции в точках полигонов. Она составляет доли миллиметра. Расстояние между точками полигона зависит от кривизны линии и может быть достаточно большим.

Сетки полигонов.

С помощью полигонов можно отобразить кривые линии, поверхности или тела. Для удобства полигоны, принадлежащие одному геометрическому объекту, объединяют в сетки. Сетка представляет собой совокупность полигонов для векторного отображения некоторого геометрического объекта. Так сетка кривой состоит из одного полигона. Сетка поверхности состоит из полигонов ее границ и нескольких полигонов ее u -линий и v -линий. Сетка тела состоит из полигонов поверхностей его граней. Отображение геометрических объектов полигонами обладает универсальностью и простотой. Для получения параллельной проекции полигона достаточно вычислить проекции его точек и соединить их отрезками. На рис. 4.2 приведен пример отображения сетки поверхности, построенной по координатным u -линиям и v -линиям.

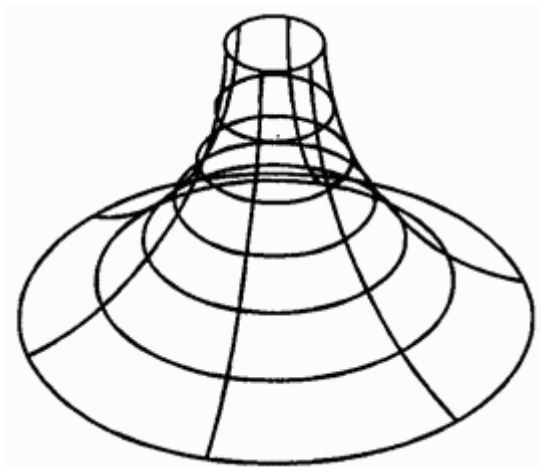


Рис. 4.2. Проекция сетки полигонов поверхности

В некоторых случаях для векторного отображения поверхности или тела граничных и координатных линий недостаточно. Для большей информативности отображения кривых поверхностей и граней к проекциям координатных линий добавляют проекции линий очерка.

6.5. Линии очерка

Каждая поверхность одной из своих сторон может быть направлена к наблюдателю и тогда эта сторона будет видимой. В противном случае сторона поверхности будет не видна из точки наблюдения. Может случиться так, что только часть стороны поверхности будет видимой. В этом случае на поверхности можно построить линию, разделяющую видимую и невидимую части поверхности. Линией очерка будем называть линию на поверхности, отделяющую видимую часть поверхности или грани от невидимой ее части.

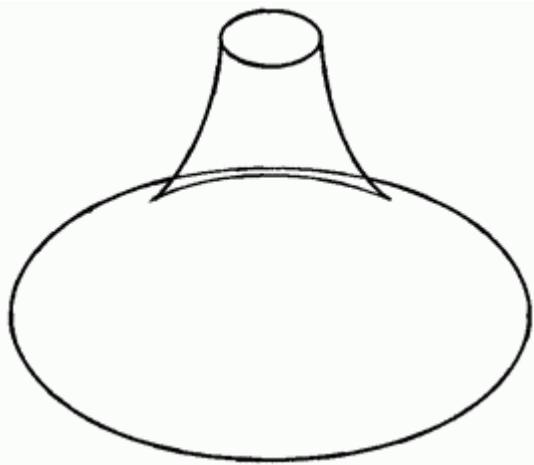


Рис. 5.1. Проекция линий очерка поверхности

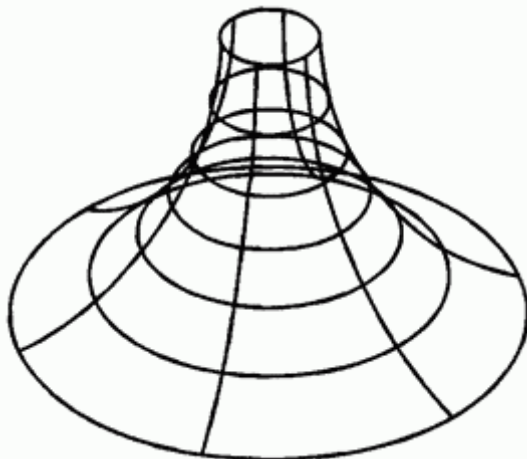


Рис. 5.2. Проекция сетки полигонов и линий очерка

На рис. 5.1 приведены линии очерка поверхности. На рис. 5.2 показаны линии очерка совместно с сеткой поверхности.

При переходе через линию очерка нормаль поверхности меняет направление по отношению к линии взгляда. В точках линии очерка нормаль поверхности ортогональна линии взгляда. В общем случае у поверхности линий очерка может быть несколько. Каждая линия очерка является пространственной кривой. Она или замкнута, или оканчивается на краях поверхности. Для разных направлений взгляда существует своя совокупность линий очерка, поэтому при повороте поверхности линии очерка необходимо строить заново.

Параллельные проекции.

Для некоторых поверхностей, например, сферы, цилиндра, конуса, линии очерка строятся достаточно просто. Рассмотрим общий случай построения линий очерка поверхности.

Пусть требуется найти линии очерка поверхности, описываемой радиус-вектором $\mathbf{r}(u, v)$. Каждая точка линии очерка для параллельной проекции на плоскость (2.1) должна удовлетворять уравнению

$$\mathbf{i}_z \cdot \mathbf{m} = 0, \quad (5.1)$$

где $\mathbf{m}=\mathbf{m}(u, v)$ — нормаль к поверхности, для которой строится линия очерка. Для поверхности, описываемой радиус-вектором $\mathbf{r}(u, v)$ нормаль $\mathbf{m}(u, v)$ также является функцией параметров u и v . Скалярное уравнение (5.1) содержит два искомых параметра u, v . Если задать один из параметров, то другой можно найти из уравнения (5.1), т. е. один из параметров является функцией от другого. Для равноправия параметров их можно представить в виде функций некоторого общего параметра t :

$$u = u(t), \quad v = v(t). \quad (5.2)$$

Результатом решения уравнения (5.1) является двумерная линия

$$l_{uv}(t) = [u(t) \quad v(t)]^T, \quad l_{uv}(t) \in \mathbf{r}(u, v) \quad (5.3)$$

на поверхности $\mathbf{r}(u, v)$. Эта линия и есть линия очерка поверхности.

Мы построим линию очерка по упорядоченной совокупности точек, удовлетворяющих уравнению (5.1). Точками мы называем пару параметров u_i, v_i поверхности, являющихся координатами двумерных точек на параметрической плоскости. Имея отдельные точки линии очерка, расположенные в порядке их следования и на определенном расстоянии друг от друга, всегда можно найти любую другую точку линии. Например, для нахождения точки, лежащей между двумя заданными соседними точками линии очерка, проведем плоскость перпендикулярно соединяющему соседние точки отрезку и найдем общую точку для поверхности и плоскости, решив три скалярных уравнения пересечения совместно с уравнением (5.1). Положение плоскости на отрезке можно задать параметром линии. По крайним точкам отрезка определяется нулевое приближение для искомой точки. Таким образом, совокупность отдельных двумерных точек линии очерка поверхности служит как бы нулевым приближением этой линии, по которому одним из численных методов всегда можно найти точное положение точки. Алгоритм построения линий очерка поверхности можно разбить на два этапа.

На первом этапе найдем хотя бы по одной точке на каждой линии очерка. Для этого, шагая по поверхности и исследуя знак скалярного произведения в соседних точках, найдем пары точек поверхности, в которых $\mathbf{m} \cdot \mathbf{i}_z$ меняет знак. Взяв в качестве нулевого приближения средние значения параметров этих точек, одним из численных методов найдем параметры u, v точки линии очерка. Пусть, например, при переходе из точки $[u_1, v_1]^T$ в близкую к ней точку $[u_2, v_2]^T$ $\mathbf{m} \cdot \mathbf{i}_z$ меняет знак. Тогда, положив $u^{(0)} = (u_1 + u_2)/2$, $v^{(0)} = (v_1 + v_2)/2$ с помощью итерационного процесса метода Ньютона

$$u^{(k+1)} = u^{(k)}, \quad v^{(k+1)} = v^{(k)} - \left(\frac{\mathbf{m} \cdot \mathbf{i}_z}{\mathbf{i}_z \cdot (\partial \mathbf{m} / \partial v)} \right)^{(k)} \quad (5.4)$$

или итерационного процесса

$$u^{(k+1)} = u^{(k)} - \left(\frac{\mathbf{m} \cdot \mathbf{i}_z}{\mathbf{i}_z \cdot (\partial \mathbf{m} / \partial u)} \right)^{(k)}, \quad v^{(k+1)} = v^{(k)} \quad (5.5)$$

найдем параметры одной из точек линии очерка. Производные нормали $\mathbf{m}_1, \mathbf{m}_2$ определяются формулами Вейнгартена (1.7.26), (1.7.28). Таким способом получим набор точек линий очерка. Точки из полученного на первом этапе набора никак не связаны друг с другом и могут принадлежать различным линиям очерка. Важно только, чтобы от каждой линии очерка в наборе присутствовала хотя бы одна точка.

На втором этапе берем любую точку из имеющегося набора и, двигаясь от нее с некоторым шагом сначала в одну сторону потом в другую, находим точку за точкой искомую совокупность точек линии очерка. Направление движения дает вектор

$$\mathbf{t} = \pm \frac{\mathbf{m}_1 \times \mathbf{m}}{|\mathbf{r}_1|} \pm \frac{\mathbf{m}_2 \times \mathbf{m}}{|\mathbf{r}_2|}, \quad (5.6)$$

где $\mathbf{m}_1, \mathbf{m}_2$ — частные производные нормали $\mathbf{m}(u, v)$, $\mathbf{r}_1, \mathbf{r}_2$ — частные производные радиус-вектора $\mathbf{r}(u, v)$ поверхности по параметрам u, v .

Знак перед слагаемым $(\mathbf{m}_i \times \mathbf{m})/|\mathbf{r}_i|$, $i=1, 2$ совпадает со знаком скалярного произведения $\mathbf{m}_i \cdot \mathbf{i}_z$. Шаг движения вычислим в соответствии с кривизнами поверхностей в текущей точке по формуле (4.7) или по формуле (4.8). Если

$$\frac{\mathbf{t} \cdot \mathbf{r}_1}{|\mathbf{r}_1|} > \frac{\mathbf{t} \cdot \mathbf{r}_2}{|\mathbf{r}_2|},$$

то по формуле (4.7) дадим приращение параметру u и по формуле (5.4) найдем соответствующий ему параметр v поверхности. В противном случае по формуле (4.8) дадим приращение параметру v и по формуле (5.5) найдем соответствующий ему параметр u поверхности. Движение по кривой закончим, когда дойдем до края одной из поверхностей или когда линия замкнется (новая точка окажется на расстоянии текущего шага от точки старта).

В процессе движения будем проверять, не лежат ли вблизи пути следования точки из набора, полученного на первом этапе. Для этого по пути следования будем вычислять расстояние от текущей точки кривой очерка до каждой точки из набора, полученного на первом этапе. Если вычисленное расстояние до какой-либо точки набора соизмеримо с текущим шагом движения, то эту точку удалим из набора как более ненужную. Так получим совокупность отдельных точек $[u_i, v_i]^T$, $i=1, 2, \dots, n$ одной линии очерка. При этом в наборе точек, полученном на первом этапе, не будет содержаться ни одной точки данной линии. Если в наборе останутся еще точки, то данная поверхность имеет, по крайней мере, еще одну линию очерка.



Рис. 5.3. Линии очерка тела

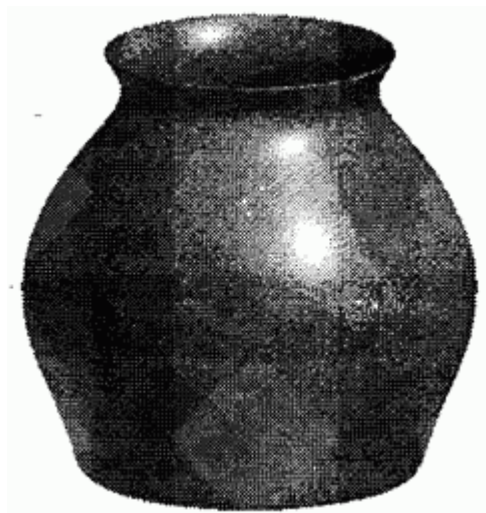


Рис. 5.4. Тело вращения

Совокупность ее точек найдем, взяв любую точку из набора и повторив второй этап построения. Построение линий закончим, когда в наборе

не останется ни одной точки. Описанным способом построим линии очерка всех граней модели.

Линии очерка граней являются линиями очерка их поверхностей. Линия очерка тела будет видимой, если она не закрыта гранью, лежащей ближе к точке наблюдения. На рис. 9.5.3 приведена линия очерка тела вращения, показанного на рис. 9.5.4. Проекция линии очерка может иметь изломы и точки возврата, но сама линия очерка является гладкой.

Точки излома у проекции возникают там, где касательная линия очерка коллинеарна вектору \mathbf{i}_z .

Для построения проекции линии очерка будем строить ее полигон, проекцию которого и возьмем в качестве проекции линии очерка.

Центральные проекции.

Линии очерка в центральных проекциях удовлетворяют уравнению

$$\mathbf{m} \cdot (\mathbf{r} - \mathbf{w}) = 0, \quad (5.7)$$

где $\mathbf{m} = \mathbf{m}(u, v)$ — нормаль поверхности $\mathbf{r} = \mathbf{r}(u, v)$, \mathbf{w} — радиус-вектор точки наблюдения. Линия очерка для центральной проекции отличается от линии очерка для параллельной проекции, хотя алгоритмы их построения аналогичны. Вместо постоянного вектора \mathbf{i}_z в (5.7) присутствует вектор $\mathbf{r} - \mathbf{w}$, направление которого зависит от проецируемой точки. Линия очерка для центральной проекции также представляет собой некоторую кривую на поверхности, описываемую зависимостями (5.3), и является пространственной кривой. Эта линия должна быть спроецирована на плоскость по правилам построения центральной проекции пространственной линии.

На рис. 5.5 приведена параллельная проекция линий очерка тора, а на рис. 5.6 для сравнения приведена центральная проекция линий очерка тора. Как можно видеть, эти проекции отличаются.

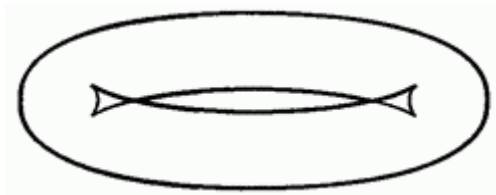


Рис. 5.5. Параллельная проекция линий очерка тора

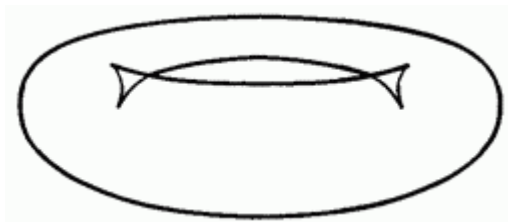


Рис. 5.6. Центральная проекция линий очерка тора

Алгоритм построения линий очерка для центральной проекции поверхности, описываемой радиус-вектором $\mathbf{r}(u, v)$ отличается от алгоритма построения линий очерка для параллельной проекции этой поверхности тем, что на первом этапе будем искать точки поверхности, в которых меняет знак скалярное произведение $\mathbf{m} \cdot (\mathbf{r} - \mathbf{w})$. Для определения этих точек вместо формул (5.4) и (5.5) следует использовать формулы

$$u^{(k+1)} = u^{(k)}, \quad v^{(k+1)} = v^{(k)} - \left(\frac{\mathbf{m} \cdot (\mathbf{r} - \mathbf{w})}{(\partial \mathbf{m} / \partial v) \cdot (\mathbf{r} - \mathbf{w}) + \mathbf{m} \cdot (\partial \mathbf{r} / \partial v)} \right)^{(k)} \quad (5.8)$$

и формулы

$$u^{(k+1)} = u^{(k)} - \left(\frac{\mathbf{m} \cdot (\mathbf{r} - \mathbf{w})}{(\partial \mathbf{m} / \partial u) \cdot (\mathbf{r} - \mathbf{w}) + \mathbf{m} \cdot (\partial \mathbf{r} / \partial u)} \right)^{(k)}, \quad v^{(k+1)} = v^{(k)} \quad (5.9)$$

соответственно. В остальном алгоритм построения линий очерка для центральной проекции поверхности не отличается от алгоритма построения линий очерка для параллельной проекции.

6.6. Определение видимой части геометрических объектов

Векторная графика.

При построении векторных изображений определение видимой части геометрических объектов сводится к задаче удаления невидимых линий. Сетки полигонов строятся для всех поверхностей тел. Часть полигонов может оказаться закрытой от взгляда наблюдателя. При удалении невидимых линий векторные проекции приобретают более реалистический вид. На рис..6.1 приведена проекция поверхности, показанной на рис. 4.2, без невидимых линий.

Линия очерка будет видимой, если она не закрыта другой гранью или частью своей же грани.

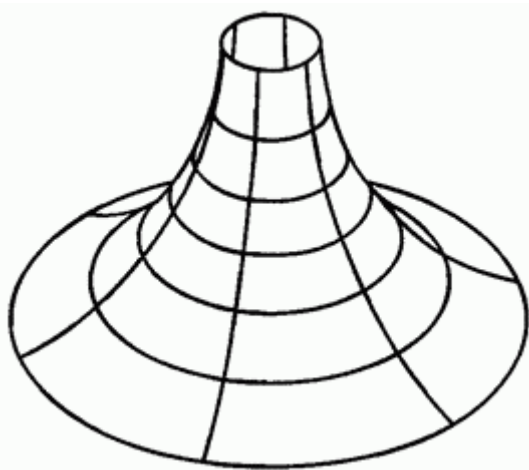


Рис. 6.1. Проекция поверхности с удалением невидимых линий

Рассмотрим один из алгоритмов удаления невидимых линий. Сначала построим пространственные полигоны геометрического объекта, включая полигоны линий очерка. Совокупность пространственных линий $\mathbf{c}_i(t_i)$, $i=1, 2, \dots, n$ соответствующих всем полигонам объекта, будем называть каркасной моделью. Далее построим проекцию (параллельную или центральную) каркасной модели на заданную

плоскость. Эта проекция будет состоять из двухмерных линий $\mathbf{c}_i(t_i)$, $i=1, 2, \dots, n$. Каждой пространственной кривой $\mathbf{c}_i(t_i)$ соответствует двухмерная кривая $\mathbf{c}_i(t_i)$. Каждой точке $\mathbf{r}=\mathbf{c}_i(t)$ пространственной кривой соответствует точка $\mathbf{r}=\mathbf{c}_i(t)$ двухмерной кривой. Среди линий проекции выделим те, которые являются проекциями линий очерка, ребер тел и границ поверхностей. Обозначим их через $l_j(t_j)$, $j=1, 2, \dots, m$. Определим, какие части проекции каркасной модели являются видимыми, а какие нет. Для этого найдем все точки пересечения и самопересечения двухмерных кривых $l_j(t_j)$, $j=1, 2, \dots, m$ между собой и со всеми кривыми $\mathbf{c}_i(t_i)$, $i=1, 2, \dots, n$ и разрежем этими точками линии на части. Соответствующим образом разрежем пространственные линии. Остается пометить и удалить невидимые части проекции. Теперь каждая часть линии будет или полностью видима или полностью невидима.

Рассмотрим алгоритм определения видимости линий проекции. Построим контур на плоскости проекции, охватывающий все двухмерные кривые. Все линии контура будут видимыми. Для определения видимости остальных линий возьмем произвольную точку \mathbf{r} на двухмерной кривой. Найдем соответствующую ей точку \mathbf{r} на соответствующей пространственной линии. Построим прямую линию $\mathbf{K}(t)=\mathbf{r}+t\mathbf{i}$, проходящую через точку \mathbf{r} и точку наблюдения и направленную от точки наблюдения. Для параллельной проекции вектор $\mathbf{i}=-\mathbf{i}_z$ для центральной проекции вектор $\mathbf{i}=(\mathbf{r}-\mathbf{w})/|\mathbf{r}-\mathbf{w}|$. Найдем все точки пересечения этой прямой с поверхностями геометрического объекта. Если значения параметра t прямой для всех точек пересечения неотрицательные, то кривая является видимой. Если среди значений параметра t для точек пересечения найдется хотя бы одно отрицательное, то кривая является невидимой, так как ее закрывает одна из поверхностей модели, находящаяся ближе к наблюдателю, чем точка \mathbf{r} . Проведем эту процедуру со всеми двухмерными кривыми и удалим те части проекции, которые окажутся невидимыми.

Для полигонов тел рассматриваемый алгоритм удаления невидимых линий проекции можно упростить. Проекция полигона (или его части) грани тела будет невидима, если нормаль грани в точках соответствующего трехмерного полигона направлена от точки наблюдения. Проекции ребер тела, нормали обеих смежных граней которых направлены от точки наблюдения, также являются невидимыми.

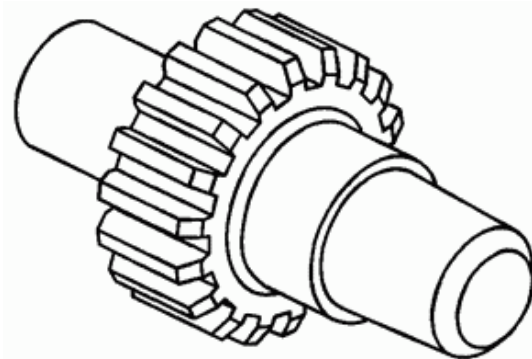


Рис. 6.2. Проекция детали с удалением невидимых линий

На рис. 6.2 приведено изображение детали с линиями очерка и удаленными невидимыми линиями. На рис. 9.6.3 приведено изображение детали, где невидимые линии имеют меньшую толщину, чем видимые.

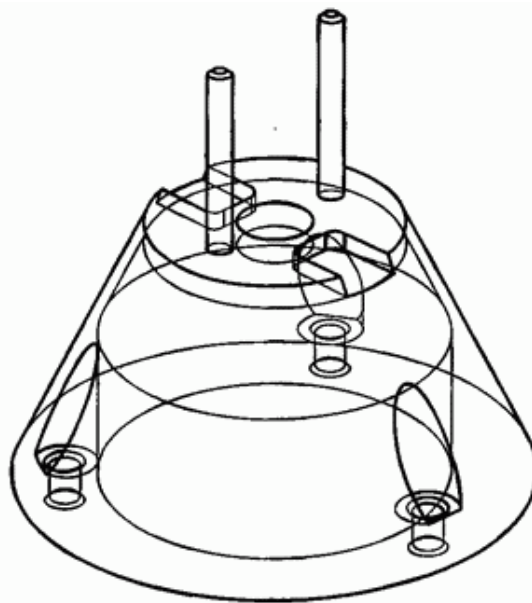


Рис. 6.3. Проекция детали с утонением невидимых линий

Описанный алгоритм является универсальным, но требует больших затрат времени. Для геометрических объектов, описываемых плоскими гранями, существует ряд быстрых алгоритмов удаления невидимых линий. В алгоритме Галимберти и Монтанари каждое ребро геометрического объекта сопоставляется с плоскостью каждой грани объекта. Каждая плоскость грани делит все пространство на два полупространства.

Если при сопоставлении ребро всегда располагается в одном полупространстве с точкой наблюдения или лежит на плоскости, частью которого является грань, то ребро видимо. Если для какой-то плоскости ребро располагается не в том полупространстве, где лежит точка наблюдения, то следует проверить, не закрыто ли ребро гранью, на которой построена плоскость. Для этого достаточно проверить положение проекции ребра относительно проекции контура грани и определить, располагается ли первая полностью вне второй, полностью внутри второй или частично внутри второй. Если при сопоставлении ребра и плоскости некоторой грани ребро попадает в оба полупространства, то его делят на две части и каждую часть рассматривают отдельно описанным способом. После каждой проверки от рассматриваемого ребра оставляют только видимую часть.

Растровая графика.

Растровое изображение формируется из отдельных точек. При построении растровых изображений возникает необходимость определять видимую часть геометрических объектов. Уоткинс предложил алгоритм построения видимой части изображения, имеющий сходство с построчной разверткой изображения. Проецируемый объект пересекают параллельными плоскостями, ортогональными плоскости проекции. Эти плоскости называют плоскостями развертки. Каждая плоскость пересекает геометрический объект по некоторым плоским кривым. Положение этих кривых анализируется на предмет удаленности от наблюдателя. На проекции должны быть построены только ближайшие к наблюдателю линии или их части. Таким образом, на линии развертки достаточно отобразить части отрезков, видимые из точки наблюдения. Изображение формируется из строк развертки. Алгоритм Уоткинса применим для построения параллельных проекций растровой графики.

Поверхности граней тел пересекаются только по своим краям. Если определить положение граней относительно друг друга и их удаленность от точки наблюдения, то для получения правильного изображения достаточно вывести на экран грани в порядке их удаления от точки наблюдения. Первыми следует вывести наиболее удаленные грани. Этот метод применим, если есть возможность изменять изображение в процессе его построения. Если позже выводимые грани проецируются на ранее выведенные грани, то последующие грани будут либо закрывать собой предыдущие (если грани непрозрачные), либо их цвет можно комбинировать с уже имеющимся цветом (если грани частично прозрачные). Этот алгоритм был предложен Ньюэллом и Санча. Основная трудность алгоритма заключается в определении порядка расположения граней.

При формировании растровых изображений требуется не только определять видимую часть геометрических объектов, но еще цвет и яркость каждой точки изображения. Если геометрические объекты не прозрачны, то цвет точки зависит от цвета ближайшей к наблюдателю грани в этой точке, а яркость точки зависит от освещенности объекта, оптических свойств его поверхности и от ориентации нормали поверхности по отношению к лучам падающего света и к линии визирования, соединяющей рассматриваемую точку и точку наблюдения. Если геометрические объекты прозрачные, то цвет точек изображения зависит от цвета всех граней, пересекаемых линией визирования, их оптических свойств и от ориентации их нормалей. Рассмотрим геометрические аспекты данной задачи.

В общем случае для каждой точки растрового изображения следует построить линию, проходящую через точку наблюдения для центральной проекции, или линию, перпендикулярную плоскости проекции для параллельной проекции, и найти все точки пересечения этой линии с отображаемыми объектами.

В точках пересечения потребуется вычислить нормаль к поверхности объекта для того, чтобы определить яркость точки растрового изображения. На рис. 6.4 приведено растровое изображение детали, построенное данным способом.

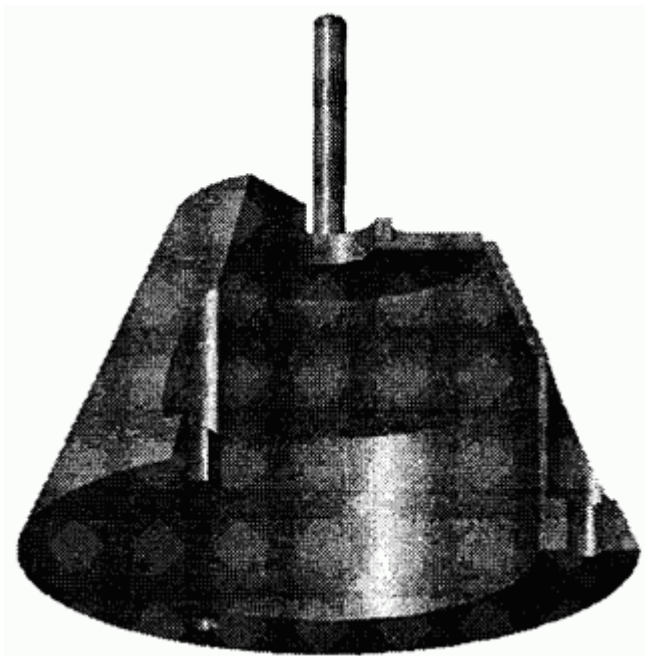


Рис. 6.4. Растровое изображение с определением яркости видимых точек

Для криволинейных поверхностей перечисленные действия требуют больших вычислительных затрат. Если поверхности геометрических объектов плоские, то задача существенно упрощается. Для уменьшения времени формирования растрового изображения все поверхности отображаемых объектов аппроксимируют треугольными пластинами.

6.7. Триангуляция поверхностей и тел

Триангуляция представляет собой аппроксимацию поверхности моделируемого объекта треугольными пластинами, отстоящими от нее на расстоянии, не превышающем некоторой заданной величины δ . Все треугольные пластины должны стыковаться между собой. Их вершины лежат на поверхности. С набором треугольных пластин легче работать, чем с поверхностью общего вида. Треугольные пластины будем называть треугольниками. Для треугольника достаточно быстро вычисляются расстояние до заданной точки или точка пересечения с

заданной прямой в пространстве. Триангуляция граней выполняется для визуального восприятия геометрической модели, поэтому стороны треугольников выбираются, такими, чтобы глаз не мог заметить изломы.

При отображении геометрических объектов по треугольникам на параметрических плоскостях поверхностей должна быть построена пространственная триангуляция граней тела путем вычисления массива точек в пространстве $\mathbf{p}_i(u_i, v_i)$ и массива нормалей $\mathbf{m}_i(u_i, v_i)$ к граням тела в этих точках по массиву двумерных точек $\mathbf{p}_i = [u_i, v_i]^T$. Для быстрого отображения тел их грани аппроксимируют треугольными пластинами, построенными на точках \mathbf{p}_i . Нормали требуются для определения поведения световых лучей, взаимодействующих с гранями тела. Тоновые рисунки в предыдущих главах и в данной главе выполнены с использованием триангуляции.

Результатом триангуляции поверхности мы хотим иметь массив двумерных точек $\mathbf{p}_i = [u_i, v_i]^T$ на параметрической плоскости и массив троек целых чисел, являющихся номерами точек в первом упомянутом массиве. Таким образом, каждый треугольник будет представлен тремя номерами его вершин в массиве параметров. По каждой двумерной точке параметрической области могут быть вычислены пространственная точка $\mathbf{p}_i(u_i, v_i)$ на поверхности и нормаль $\mathbf{m}_i(u_i, v_i)$ поверхности в ней. Пространственные точки и нормали могут храниться в массивах, аналогичных массиву двумерных точек.

Остановимся на некоторых способах триангуляции. Для плоских поверхностей существуют экономичные методы триангуляции, в которых треугольники строятся на граничных точках поверхности и не требуется искать точки внутри параметрической области.

Триангуляция Делоне.

Рассмотрим некоторую область на плоскости. Область будем называть выпуклой, если при движении вдоль ее границы приходится поворачивать только в одну сторону (только влево или только вправо). Для триангуляции выпуклых плоских областей можно использовать алгоритм Делоне. Мы не сможем напрямую применить этот алгоритм для триангуляции поверхностей произвольной формы, но мы будем использовать его метод построения треугольников.

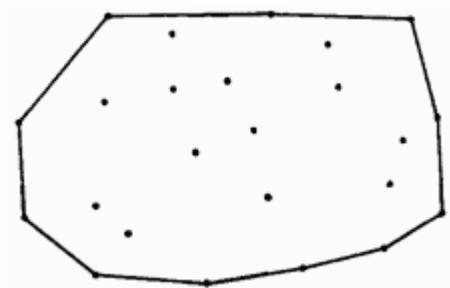


Рис. 7.1. Выпуклая область с заданными точками внутри

Пусть даны некоторая выпуклая двумерная область, ограниченная замкнутой ломаной линией, и набор точек внутри этой области (рис. 7.1).

Требуется разбить указанную область на треугольники, вершинами которых являются заданные точки внутри области и вершины ограничивающей ее ломаной линии. Треугольники не должны накрывать друг друга, а их стороны могут пересекаться только в вершинах. Можно построить несколько различных наборов треугольников, заполняющих указанную область. Во всех случаях число треугольников равно $K+I-2$, где K — число вершин ограничивающей ломаной, I — число заданных точек внутри области.

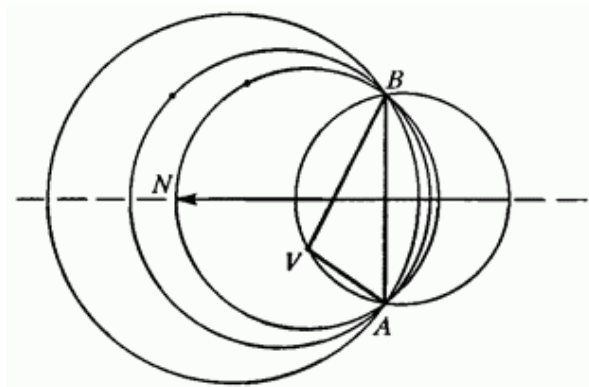


Рис. 7.2. Выбор третьей точки алгоритма Делоне

Триангуляция области будет триангуляцией Делоне, если внутри описанной вокруг каждого треугольника окружности отсутствуют вершины других треугольников. Триангуляция Делоне строит треугольники по возможности близкие к равноугольным (не допускает построение неоправданно вытянутых треугольников).

Ее можно назвать сбалансированной. Триангуляция Делоне будет уникальной, если никакие четыре вершины не лежат на одной окружности.

Рассмотрим триангуляцию Делоне. Вершины ограничивающей область ломаной и заданные точки внутри области будем называть вершинами триангуляции. Стороны треугольников будем называть ребрами. Среди ребер выделим отрезки ограничивающей ломаной, которые будем называть граничными ребрами. Сориентируем все граничные ребра так, чтобы выпуклая область лежала слева от каждого ребра. Пусть требуется построить треугольник, стороной которого является граничное ребро АВ, показанное на рис. 7.2.

Через вершины А, В и любую, не лежащую с ними на одной прямой, вершину можно провести окружность. В качестве третьей вершины треугольника выберем вершину V, соответствующая которой окружность, не содержит других вершин с той же стороны относительно отрезка АВ, с которой лежит точка V. Для граничного ребра в общем случае можно найти одну такую вершину. Будем называть ее ближайшей. Центр окружности, проходящей через точки А, В и V, лежит на пересечении перпендикуляров к серединам отрезков АВ, BV и VA. Положение центра окружности будем характеризовать параметром t отрезка MN, перпендикулярного ребру АВ, равного с ним по длине и проходящего через середину ребра АВ.

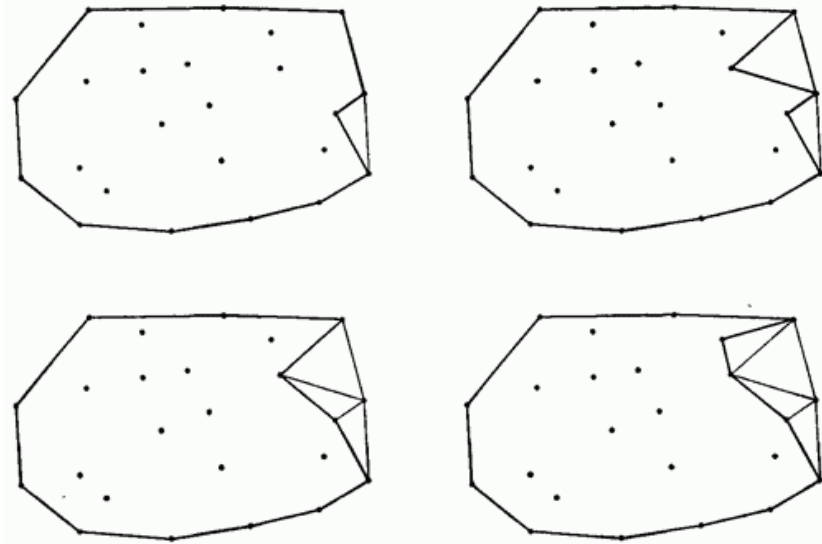


Рис. 7.3. Процесс триангуляции Делоне

Для всех вершин, лежащих слева от отрезка АВ, ближайшая вершина имеет наименьший параметр t . Соответствующая ближайшей вершине окружность не содержит других вершин слева от отрезка АВ. Пусть вершины А, В и V описываются двумерными радиус-векторами $a=[x_a, y_a]^T$, $b=[x_b, y_b]^T$, $v=[x_v, y_v]^T$ соответственно. Радиус-векторы середин отрезков АВ и BV будут равны

$$m = [x_m \ y_m]^T = \begin{bmatrix} \frac{1}{2}(x_a + x_b) \\ \frac{1}{2}(y_a + y_b) \end{bmatrix}, \quad q = [x_q \ y_q]^T = \begin{bmatrix} \frac{1}{2}(x_v + x_b) \\ \frac{1}{2}(y_v + y_b) \end{bmatrix}$$

Значение параметра t прямой $MN=(1-t)m+tn$, соответствующее положению на ней центра окружности, проходящей через точки А, В и V, равно

$$t = \frac{(x_v - x_b)(x_q - x_m) + (y_v - y_b)(y_q - y_m)}{(y_v - y_b)(x_b - x_a) - (x_v - x_b)(y_b - y_a)}. \quad (7.1)$$

Для ближайшей слева к отрезку АВ вершины параметр t имеет минимальное значение.

Сориентируем все граничные ребра так, чтобы подлежащая триангуляции область лежала слева от каждого из них. Построение треугольников начнем с любого граничного ребра. Найдем для него ближайшую вершину, соответствующая окружность которой не содержит других вершин. Пусть для граничного ребра АВ найдена ближайшая вершина V. Тогда построим треугольник АВV и переведем ребро АВ в разряд неактивных. Неактивными будем называть ребра и вершины, которые не участвуют в алгоритме триангуляции. Если среди граничных ребер отсутствует ребро BV, то на отрезке VB построим новое граничное ребро. Если же среди граничных ребер есть ребро BV, то переведем его и вершину V в разряд неактивных. Если среди граничных ребер отсутствует ребро VA, то на отрезке AV построим новое граничное ребро. Если же среди граничных ребер есть ребро VA, то переведем его и вершину A в разряд неактивных. Процесс триангуляции показан на рис. 7.3.

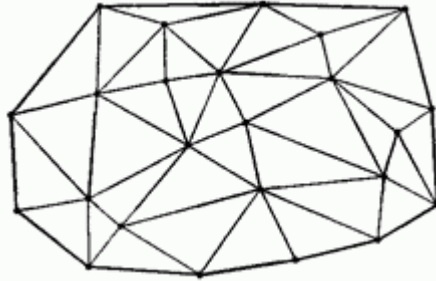


Рис. 7.4. Триангуляция Делоне

Триангуляцию закончим, когда все вершины и ребра станут неактивными. Результат триангуляции заданной области приведен на рис. 7.4.

Триангуляция методом коррекции.

Рассмотрим триангуляцию некоторой поверхности $\mathbf{r}(u, v)$ с прямоугольной областью определения параметров $u_{\min} \leq u \leq u_{\max}$, $v_{\min} \leq v \leq v_{\max}$. Разобьем область определения параметров поверхности на

прямоугольные ячейки двухмерными линиями $u_i = \text{const}$, $v_j = \text{const}$, $i=1, 2, \dots, n$, $j=1, 2, \dots, m$. Эти линии образуют прямоугольную сетку. Параметрические расстояния $\Delta u_i = u_{i+1} - u_i$ между соседними линиями $u_i = \text{const}$ в соответствии с формулой (4.7) возьмем равными

$$\Delta u_i = \min \left(2 \frac{\sqrt{(\delta/m)(2(g_{11}(u_i, v)/b_{11}(u_i, v)) - \delta/m)}}{\sqrt{g_{11}(u_i, v)}} \right) \quad (7.2)$$

для всех $v_{\min} \leq v \leq v_{\max}$.

Параметрические расстояния $\Delta v_j = v_{j+1} - v_j$ между соседними линиями $v_j = \text{const}$ в соответствии с формулой (4.8) возьмем равными

$$\Delta v_j = \min \left(2 \frac{\sqrt{(\delta/m)(2(g_{22}(u, v_j)/b_{22}(u, v_j)) - \delta/m)}}{\sqrt{g_{22}(u, v_j)}} \right) \quad (7.3)$$

для всех $u_{\min} \leq u \leq u_{\max}$

Построив диагонали во всех прямоугольных ячейках, мы получим триангуляцию поверхности (получим набор треугольников, удовлетворяющий предъявленным требованиям). На рис. 9.7.5 приведена триангуляция поверхности вращения описанным способом.

Рассмотрим триангуляцию поверхности $\mathbf{r}(u, v)$ с произвольной границей. Метод триангуляции построим на коррекции граничными контурами описанной выше триангуляции поверхности с прямоугольной областью определения параметров.

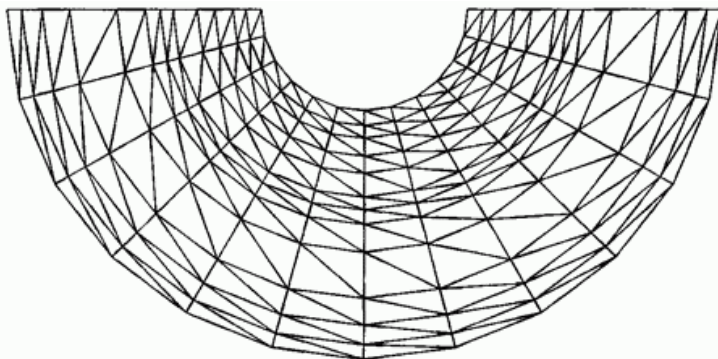


Рис. 7.5. Триангуляция поверхности с прямоугольной областью определения параметров

Пусть граница поверхности в области определения параметров описывается несколькими непересекающимися двухмерными контурами. Один из контуров является внешним и содержит остальные контуры. За положительное направление для каждого контура примем направление, при движении вдоль которого область определения поверхности находится всегда слева от контура, если смотреть навстречу нормали поверхности. Построим полигоны в положительном направлении граничных контуров области определения поверхности. Для построения граничных полигонов нужно пройти по граничным контурам поверхности с некоторым переменным шагом и заполнить массив двухмерных точек, координатами которых являются параметры поверхности. Полигон будем строить из точек на параметрической плоскости, но шаг при переходе от одной точки к другой будем определять из пространственной геометрии, а именно, из условия, чтобы прогиб дуги кривой между соседними точками был бы не более заданной величины δ . Параметрические шаги Δt построения полигона для кривой $\mathbf{r}(t) = \mathbf{r}(u(t), v(t))$ граничного контура поверхности вычислим по формуле (4.4).

Каждый полигон состоит из упорядоченного набора двухмерных точек $\mathbf{p}_i = [u_i, v_i]^T$. Каждый участок полигона можно рассматривать как отрезок двухмерной прямой линии, построенный на двух соседних точках. Будем использовать такие участки в качестве граничных ребер, а точки полигонов, на которых базируются ребра, будем использовать в качестве вершин триангуляции. Так как область определения

параметров поверхности лежит слева от граничных полигонов, то при построении треугольников для каждого граничного ребра триангуляции следует искать третью вершину треугольника слева от ребра.

Далее построим прямоугольную сетку для области $u_{\min} \leq u \leq u_{\max}$, $v_{\min} \leq v \leq v_{\max}$, где u_{\min}, u_{\max} , v_{\min}, v_{\max} определяют габаритный прямоугольник внешнего граничного контура. Узлы сетки мы также будем использовать в качестве вершин триангуляции.

Определим, какие узлы лежат внутри граничных полигонов, а какие лежат на границе или вне области определения поверхности. Используя эту информацию, рассортируем прямоугольные ячейки сетки на две группы. К первой группе отнесем ячейки, целиком лежащие внутри области определения параметров поверхности (ячейки не должны касаться граничных полигонов). Ко второй группе отнесем остальные ячейки (лежащие вне области определения поверхности или пересекаемые граничными полигонами).

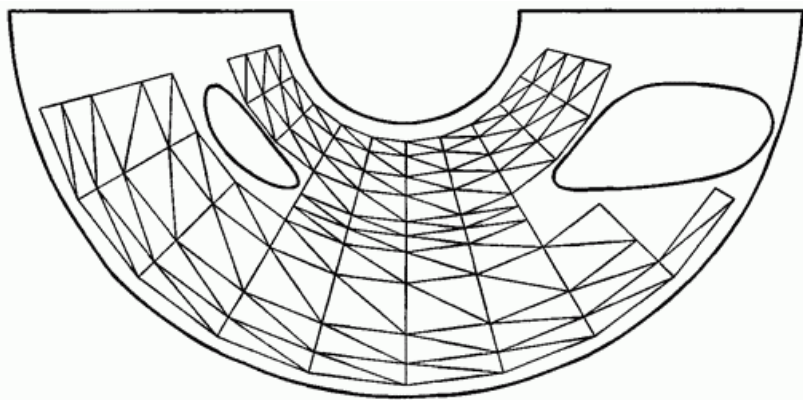


Рис. 7.6. Незаконченная триангуляция поверхности

Внутри каждой ячейки первой группы с помощью диагонали построим два треугольника. Тем самым мы получим незаконченную триангуляцию. Пример построения треугольников в ячейках первой группы для ограниченной контурами поверхности вращения приведен на рис. 7.6.

На непересеченных сторонах ячеек второй группы построим граничные ребра и направим их так, чтобы соответствующая ячейка находилась слева от ребра. Вокруг ячеек первой группы построим замкнутую ломаную линию (возможно несколько замкнутых линий) так, чтобы при движении по ней не разбитая на треугольники часть области лежала слева, если смотреть навстречу нормали поверхности. Прямолинейные участки этой ломаной линии также будем использовать в качестве граничных ребер. Мы будем считать все ребра равноправными. Для завершения триангуляции нам необходимо построить треугольники между граничными ребрами. Для каждого ребра будем искать вершину, которая лежит слева от него и может быть использована для построения треугольника. Поиск вершины будем осуществлять только среди тех вершин, которые лежат в одной ячейке с ребром. Для выбора вершины используем метод Делоне, описанный выше, и проиллюстрированный на рис. 9.7.2. Если такая вершина найдена, то следует проверить, не пересекаются ли два новых ребра треугольника с каким-либо граничным ребром. Пусть для граничного ребра АВ найдена ближайшая вершина V и проверено, что отрезки BV и VA не пересекают другие граничные ребра. Тогда построим треугольник ABV и переведем ребро АВ в разряд неактивных. Если среди граничных ребер отсутствует ребро BV, то на отрезке VB построим новое граничное ребро, если же среди граничных ребер есть ребро BV, то переведем его и вершину B в разряд неактивных. Если среди граничных ребер отсутствует ребро VA, то на отрезке AV построим новое граничное ребро, если же среди граничных ребер есть ребро VA, то переведем его и вершину A в разряд неактивных.

Если отрезок BV или VA пересекает другие граничные ребра, то перейдем к поиску ближайшей вершины для другого граничного ребра. Триангуляция будет закончена после перевода всех ребер и вершин в разряд неактивных.

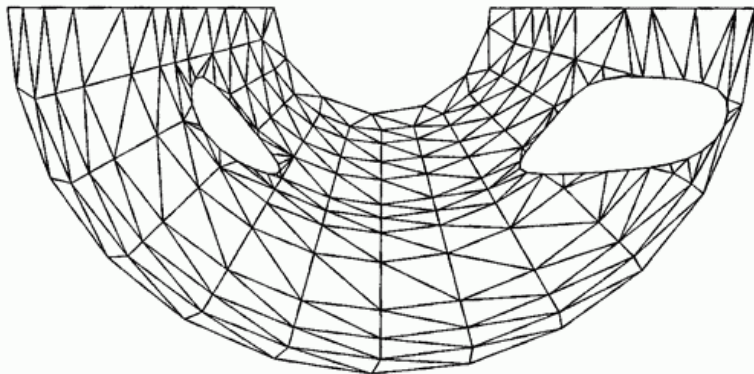


Рис. 7.7. Триангуляция методом коррекции

На рис. 7.7 приведена триангуляция поверхности методом коррекции треугольников в ячейках, пересеченных граничными контурами. На рис. 7.8 с помощью полученной триангуляции отображена сама поверхность.

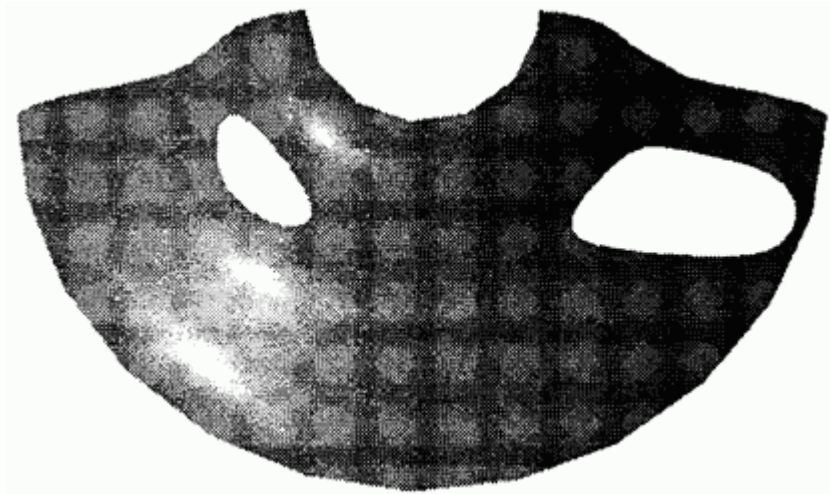


Рис. 7.8

Если граничные полигоны и поверхность обладают некоторой симметрией, то триангуляция методом коррекции будет обладать аналогичной симметрией.

Триангуляция методом поглощения.

Рассмотрим еще один метод триангуляции. По скорости он уступает триангуляции Делоне и ее модификациям. Для начала процедуры триангуляции необходимо представить границу поверхности в виде замкнутых полигонов. В процессе триангуляции нам потребуется определять шаги по параметрам поверхности Δu , Δv . При известном направлении движения эти шаги определяются формулами (4.6). Приблизительно шаги по параметрам поверхности Δu , Δv можно найти следующим образом. Определим область на плоскости параметров вокруг некоторой точки $[u_0, v_0]^T$ таким образом, чтобы любой пространственный отрезок из точки $[u_0, v_0]^T$ в точку $[u_1, v_1]^T$ этой области отстоял бы от поверхности не дальше заданной величины δ .

Для этого вычислим допустимые приращения параметров вдоль координатных линий

$$\Delta u \approx 2 \frac{\sqrt{(\delta/m)(2(g_{11}/b_{11}) - \delta/m)}}{|\partial \mathbf{r} / \partial u|}, \quad \Delta v \approx 2 \frac{\sqrt{(\delta/m)(2(g_{22}/b_{22}) - \delta/m)}}{|\partial \mathbf{r} / \partial v|}, \quad (7.4)$$

где $g_{11}, g_{22}, b_{11}, b_{22}$ — коэффициенты первой и второй квадратичных форм поверхности в точке $[u_0, v_0]^T$. За границу искомой области примем эллипс с центром в точке $[u_0, v_0]^T$ и полуосями Δu , Δv . Этот эллипс имеет уравнение

$$\left(\frac{u - u_0}{\Delta u} \right)^2 + \left(\frac{v - v_0}{\Delta v} \right)^2 = 1. \quad (7.5)$$

Если требуется на плоскости найти точку рядом с точкой $[u_0, v_0]^T$ в направлении, заданном углом φ с осью u , то ее параметрами будут

$$u_1 = u_0 + \Delta u \cos \varphi, \quad v_1 = v_0 + \Delta v \sin \varphi. \quad (7.6)$$

Сначала рассмотрим более простой случай, когда область параметров поверхности ограничена одним внешним контуром. Аппроксимируем границу поверхности замкнутым полигоном на параметрической области. При построении триангуляции будем использовать рабочий полигон, за который в данном случае примем полигон внешнего контура. Точки полигона занесем в результирующий массив двумерных точек. Треугольники будем строить от края рабочего полигона, сужая его до тех пор, пока в рабочем полигоне не останется всего три точки.

Найдем в рабочем полигоне вершину, в которой он поворачивает внутрь области. Такая точка всегда существует и угол поворота в ней меньше π . Обозначим эту точку через O , а ее параметры — через u_0, v_0 . Около этой точки построим один или два треугольника в зависимости от угла поворота. Если угол меньше $\pi/2$, то построим один треугольник на этих трех точках (рис. 7.9). В противном случае построим два треугольника на данной, двух соседних и одной новой точках (рис. 7.11). Новая точка обозначена через P . Точку P будем искать на диагонали параллелограмма BOC . Если вершина параллелограмма лежит внутри эллипса (рис. 7.10), то примем ее за точку P . В противном случае за точку P примем пересечение эллипса и диагонали параллелограмма. В последнем случае совсем не обязательно искать пересечение эллипса и отрезка.

Координаты u_p, v_p точки P определяются через координаты точек O BC

$$u_p = u_b + u_c - u_0, \quad v_p = v_b + v_c - v_0. \quad (7.7)$$

Угол отрезка OP с горизонталью определяется равенством

$$\operatorname{tg} \varphi = \frac{v_p - v_0}{u_p - u_0}. \quad (7.8)$$

Эти данные позволяют определить положение точки P относительно эллипса (7.5).

В случае, показанном на рис. 7.9, построим треугольник (запомним номера его вершин) и в рабочем полигоне удалим точку O . В случае,

показанном на рис. 7.11, построим два треугольника и в рабочем полигоне точку O заменим точкой P и поместим последнюю в результирующий массив точек. На рис. 7.12 приведен полигон, полученный после построения двух треугольников и ликвидации точки O . В обоих случаях точка O будет удалена из рабочего полигона и рабочий полигон сузится. Заметим, что треугольники можно строить только тогда, когда рабочий полигон после сужения не будет сам себя пересекать.

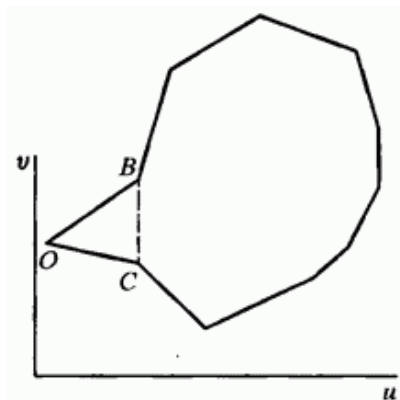


Рис. 7.9. Построение треугольника

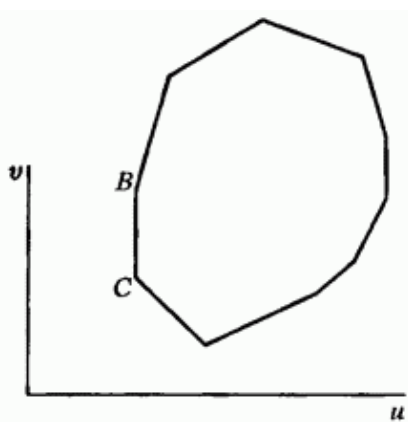


Рис. 7.10. Результирующий полигон

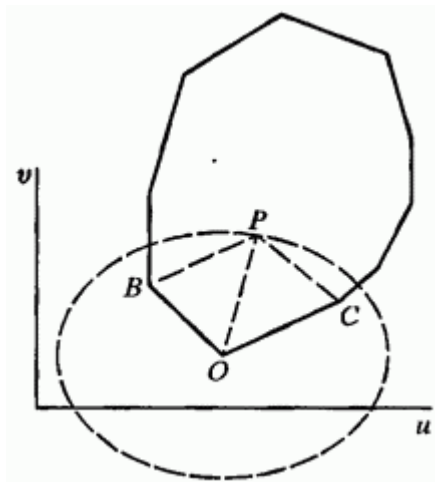


Рис. 7.11. Построение двух треугольников

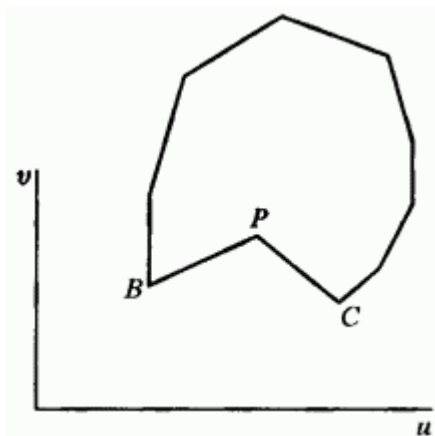


Рис. 7.12. Результирующий полигон

Такие ситуации показаны на рис. 7.13. Они могут возникнуть, когда стороны построенных треугольников пересекут несмежные с ними стороны рабочего полигона. Перед построением нового треугольника как в случае, показанном на рис. 7.9, так и в случае, показанном на рис.

7.11, должна быть выполнена проверка на отсутствие самопересечения результирующего полигона.

Более того, при определении положения точки Р важно, чтобы она находилась на достаточном расстоянии от других точек рабочего полигона и не подходила близко к отрезкам, соединяющим точки полигона. Иначе могут возникнуть трудности в дальнейшем при построении треугольников. Поэтому прежде, чем сузить рабочий полигон, следует проверить на самопересечение результирующий полигон. Если около точки О нельзя построить треугольник (треугольники), то вместо нее следует найти другую точку, в которой полигон более, чем в других, заворачивает внутрь контура, и выполнить в ней описанные действия.

Далее с измененным рабочим полигоном выполним те же действия, которые мы только что описали. Найдем в рабочем полигоне точку, в которой он более, чем в других, точках поворачивает внутрь области, выполним проверку на возможность сужения в ней полигона путем построения одного или двух треугольников и сузим полигон.

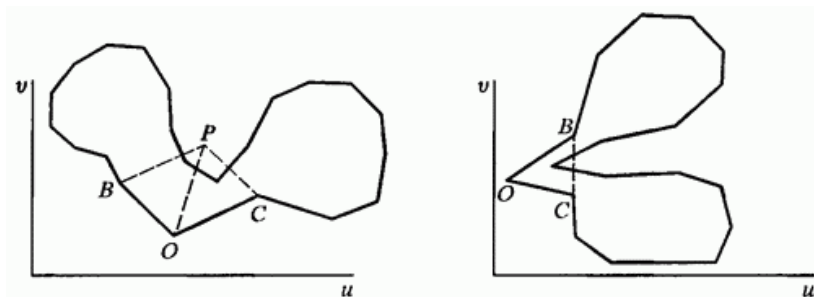


Рис. 7.13. В данном углу строить треугольники нельзя

Продолжая этот процесс, мы будем расширять массив двумерных точек и массив треугольников, и одновременно мы будем сужать рабочий полигон, уменьшая охватываемую им площадь и число его точек. На некотором этапе этих действий мы получим рабочий полигон, состоящий из трех точек. Построим на этих точках последний треугольник, ликвидируем рабочий полигон и закончим триангуляцию. В описываемом способе триангуляции область, ограниченная рабочим

полигоном, как бы ликвидируется путем отрезания от нее треугольников.

Рассмотрим общий случай, когда область параметров поверхности ограничена одним внешним контуром и несколькими внутренними контурами, целиком лежащими внутри внешнего контура. Аппроксимируем границу поверхности замкнутыми полигонами на параметрической области. Для каждого контура построим свой полигон. Так же как и для контуров, для полигонов, построенных на них, должно быть выполнено правило их взаимной ориентации. Ориентация внутренних полигонов должна быть противоположной ориентации внешнего полигона. Построение триангуляции начнем с полигона внешнего контура. Положим его точки в результирующий массив двумерных точек, а сам полигон сделаем рабочим.

Построение треугольников выполним так же, как и в случае односвязной области. Найдем в рабочем полигоне точку O , выполним проверку на возможность сужения в ней рабочего полигона и сузим полигон. При наличии внутренних контуров усложняется проверка возможности сужения рабочего полигона в выбранной точке. Кроме описанных проверок на пересечение сторон треугольников со сторонами рабочего полигона нужно выполнить проверку на пересечение сторон треугольников со сторонами всех внутренних полигонов.

Пусть мы проверяем возможность построения двух треугольников в точке O (рис. 7.11), и обнаружили, что новая точка P , будучи построенной, попадет внутрь одного из внутренних полигонов или окажется в недопустимой близости от его отрезков. В этом случае мы не будем строить точку P , а вместо этого включим в рабочий полигон данный внутренний полигон, построив два треугольника, показанных на рис. 7.14.

Для того чтобы точки одного из внутренних полигонов включить в рабочий полигон, найдем среди точек внутреннего полигона точку, ближайшую к точке C (смежную с точкой O) рабочего полигона.

Построим треугольники на точках OCF и CEP и между точками O и C рабочего полигона вставим точки внутреннего полигона, начиная с точки F и кончая точкой E . Тем самым мы разорвем рабочий полигон

на отрезке OC , разорвем внутренний полигон на отрезке EF и объединим их отрезками OF и EC .

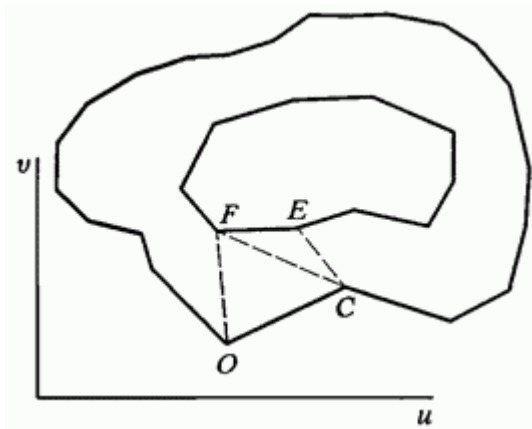


Рис. 7.14. Построение двух треугольников

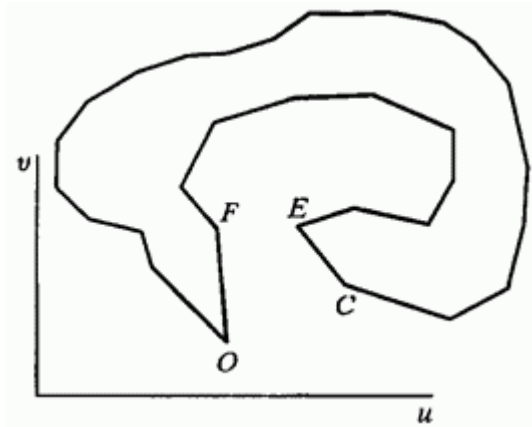


Рис. 7.15. Слияние внешнего и внутреннего полигонов

Результат слияния приведен на рис. 7.15. Конечно, перед объединением внешнего и внутреннего полигонов должны быть выполнены проверки на корректность этой операции.

Далее будем продолжать сужать рабочий полигон описанным способом до тех пор, пока не окажемся в непосредственной близости с другим внутренним полигоном и не включим его в рабочий полигон. В итоге, все внутренние полигоны будут включены в рабочий полигон, который должен быть сужен до последних трех точек. В результате, вся многосвязная область определения параметров поверхности будет покрыта треугольниками.

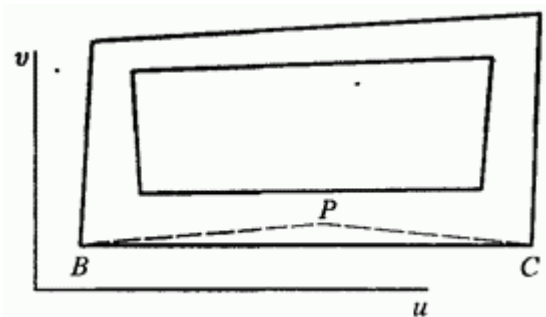


Рис. 7.16. В данном углу построить треугольники нельзя

Возможны ситуации, когда нельзя построить ни одного треугольника на заданных полигонах. На рис. 7.16 приведена область ограниченная двумя полигонами, каждый из которых состоит из четырех отрезков. Для внешнего полигона мы не можем продолжить триангуляцию, так как мешает внутренний полигон. В такой случае найдем две соседние точки В и С полигона, для которых можно построить треугольник ВСР. Точка Р проецируется на середину стороны ВС и находится на таком расстоянии от нее, чтобы новый треугольник не пересекал полигоны.

Далее продолжим триангуляцию описанным выше способом.

Другие способы триангуляции.

Существуют и другие способы триангуляции. Например, после построения полигонов внешнего и внутренних контуров области определения поверхности может быть выбрана иная стратегия построения треугольников. В другом варианте можно перед началом триангуляции объединить внешний и внутренние полигоны в один полигон. Можно внутри области определения параметров по

определенному алгоритму «набросать» точки и по ним и точкам полигонов граничных контуров выполнить триангуляцию Делоне. Существуют алгоритмы, строящие сначала крупные треугольники, а затем делящие их до приемлемых размеров.

Триангуляция тела.

Триангуляция тела представляет собой совокупность треугольников, полученных путем триангуляции поверхностей его граней.

Триангуляция отдельных поверхностей отличается от триангуляции граней тела тем, что в последнем случае должны быть согласованы граничные полигоны для смежных граней (рис. 7.17).

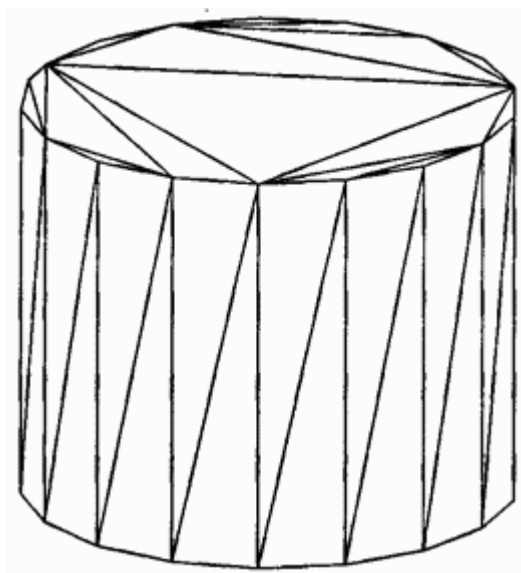


Рис. 7.17. Согласованность граничных полигонов граней тела

Участки полигонов смежных граней, проходящие по общим ребрам, будут согласованными, если их точки совпадают в пространстве.

Применение триангуляции.

Построенные в результате триангуляции треугольники используются для получения тоновых изображений. На рис. 9.7.18 и 9.7.19 приведены триангуляции грани листового тела, тоновое изображение которого показано на рис. 5.1.

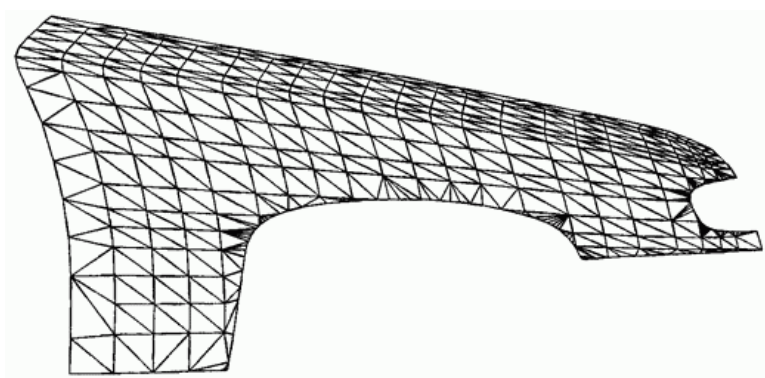


Рис. 7.18. Триангуляция грани тела методом коррекции

Разбиение области определения параметров поверхности на треугольники может быть использовано в интегралах (5.6.2), (5.6.3), (5.6.12), (6.7.17)-(5.7.22) при вычислении геометрических характеристик тел.

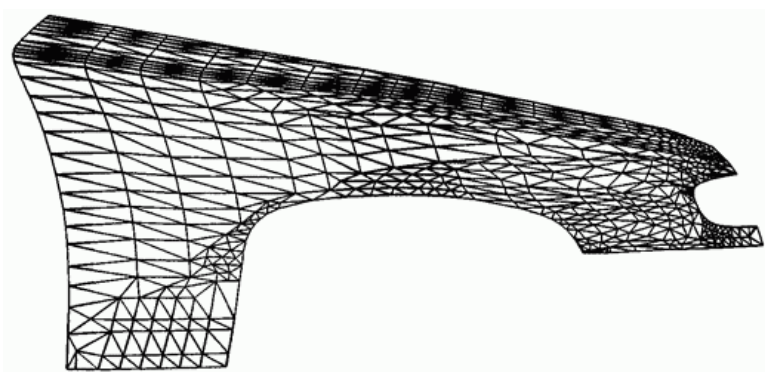


Рис. 7.19. Триангуляция грани тела методом поглощения

Приближенно интеграл для треугольной области поверхности вычисляется по кубатурной формуле (5.10.13).

6.8. Моделирование света

Изображение окружающих нас предметов формируется на сетчатке глаза потоками света, пришедшими от этих предметов. Пришедший свет является отраженным или излученным поверхностями окружающих предметов. Если некоторое тело полностью поглощает падающий на него свет, то оно называется абсолютно черным телом. Если тело полностью пропускает падающий на него свет, то оно называется абсолютно прозрачным телом. Реальные предметы некоторую часть падающего света поглощают, превращая его в тепло, некоторую часть света пропускают и некоторую часть — отражают. Будем считать, что для моделируемых объектов доля излученного и поглощенного света незначительна по сравнению с долей пропущенного и отраженного света.

Поглощенная часть, пропущенная часть и отраженная часть падающего света зависят от длины его волны. Цвет предмета зависит и от длины волны падающего света и от областей спектра, которые поверхность предмета отражает в большей степени. В этом параграфе мы не будем останавливаться на цвете тел, а рассмотрим взаимодействие тел с потоком света некоторой длины волны. **Световой поток мы будем характеризовать интенсивностью света — плотностью потока энергии световой волны.**

Тоновое изображение зависит от положения точки наблюдения, положения источников света относительно геометрической модели и от оптических свойств ее поверхностей. Поверхность объекта частично поглощает падающий свет, частично — пропускает, а частично — отражает. **При моделировании отраженного света его условно делят на две части: диффузно отраженный и зеркально отраженный.** Это деление связано с законами, по которым описывается направление лучей отраженного света. Соответственно поверхностям приписываются свойства диффузного и зеркального отражения. Если большую часть отраженного поверхностью света можно описать законом диффузного отражения, то поверхность называют **матовой**, если же большую часть отраженного поверхностью света можно описать законом зеркального отражения, то ее называют **зеркальной**. Объекты, поверхности которых способны

пропускать большую часть падающего на них света вглубь, называют прозрачными.

Пропущенный через поверхность свет также условно делят на две части: диффузно пропущенный и направленно пропущенный. Направленно пропущенный свет претерпевает преломление. Пропущенный свет может быть отражен от других поверхностей.

Падающий на поверхность свет можно условно разделить на **направленный и рассеянный**. Направленным является свет солнца или некоторого другого точечного источника. Рассеянный свет возникает в результате дифракции некоторого направленного света на мелких оптических неоднородностях окружающей среды (например, воздуха) из-за **флуктуации плотности (небольших отклонений плотности от ее среднего значения в пределах малых объемов)**. Рассеянный свет характеризуется равномерным распределением его интенсивности по всем направлениям. В отсутствие рассеянного света предметы выглядят контрастными. Так выглядят предметы, помещенные в темную комнату и освещенные направленным на них прожектором. При отсутствии воздуха части предметов, находящиеся в тени, могут быть совсем не видны. В большинстве реальных ситуаций рассеянный свет присутствует.

Приходящий в точку наблюдения свет разделим на четыре составляющие части:

- рассеянный свет,
- диффузно отраженный свет,
- зеркально отраженный свет,
- пропущенный свет.

Рассеянный свет.

Рассеянный свет, как и направленный, подвержен диффузному и зеркальному отражению, но в силу одинаковой интенсивности во всех направлениях в точку наблюдения от каждой точки поверхности приходит одинаковое количество рассеянного света. Обозначим через

Иа интенсивность рассеянного света. При наличии только рассеянного света в точку наблюдения от каждой точки объекта независимо от ориентации в ней поверхности придет свет интенсивности

$$I = k_a I_a, \quad (8.1)$$

где k_a — коэффициент, определяющий отраженную долю рассеянного света.

Диффузно отраженный свет.

Матовая поверхность большую часть падающего света отражает диффузно. Для диффузного отражения справедлив закон косинусов Ламберта, устанавливающий соответствие между количеством отраженного света и косинусом угла θ между направлением на точечный источник света интенсивности I_p и нормалью к поверхности. Интенсивность диффузно отраженного света определяется формулой

$$I_d = k_d I_p \cos \theta, \quad (8.2)$$

где k_d — коэффициент диффузного отражения, зависящий от материала поверхности. Количество диффузно отраженного света не зависит от положения наблюдателя. Матовые поверхности отражают одинаковое количество световой энергии во всех направлениях, но это количество пропорционально $\cos \theta$.

Зеркально отраженный свет.

Зеркальные поверхности отражают свет неодинаково по разным направлениям. Вследствие этого на гладких криволинейных поверхностях можно наблюдать световой блик. От зеркальной поверхности большая часть падающего света отражается в направлении, угол с нормалью которого равен углу падения и лежащем в одной плоскости с падающим светом и нормалью в точке падения.

Это направление называется направлением отраженного света (рис. 8.1). При отклонении на некоторый угол α от этого направления интенсивность отраженного света резко уменьшается. Изменение

интенсивности зеркально отраженного света около направления отраженного света Фонг предложил описывать функцией $\cos^n \alpha$, где n может зависеть от свойств поверхности и лежит в диапазоне от 1 до 200.

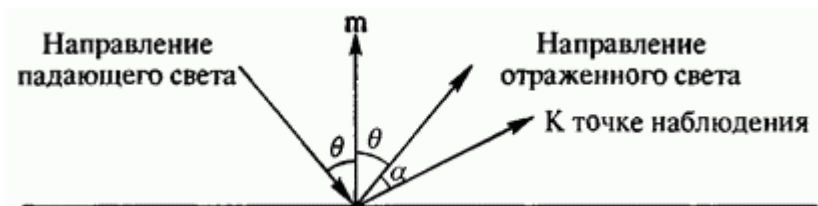


Рис. 8.1. Зеркальное отражение света

Количество отраженного света зависит также и от угла падения θ . При моделировании света зависимость отраженного света от угла падения заменяют константой $f(\theta)=k_s$, которую выбирают опытным путем так, чтобы результаты были приемлемы с эстетической точки зрения. Если интенсивность источника света равна I_p , то интенсивность зеркально отраженного света определим формулой Фонга

$$I_s = k_s I_p \cos^n \alpha, \quad (8.3)$$

где k_s будем называть коэффициентом зеркального отражения, а n будем называть коэффициентом блеска. Чем больше n , тем ярче и уже световой блик.

Пропущенный свет.

Если отображаемый объект является прозрачным, то от точки его поверхности в точку наблюдения придет не только отраженный свет, но и свет, пропущенный поверхностью из глубины объекта.

Пропущенный поверхностью свет может прийти в результате отражения от другой поверхности или от источника света через прозрачный объект. В обоих случаях свет пойдет через вещество и будет им частично поглощен. Интенсивность света при прохождении через вещество убывает по экспоненциальному закону и описывается законом Бугера

$$I = I_0 e^{-\mu l},$$

где I_0 — интенсивность света на входе в поглощающее вещество, I — пройденное расстояние, μ — коэффициент поглощения света данным веществом. При $l=1/\mu$ интенсивность пропущенного света уменьшается в e раз. Это равенство можно рассматривать, как определение коэффициента поглощения. Интенсивность I_0 является одной из составляющих частей падающего света. Поверхность как граница двух сред характеризуется коэффициентом отражения $k = k_s + k_d$ и коэффициентом пропускания k_t . Если пренебречь поглощением света, то интенсивность отраженного и пропущенного света в сумме должны давать интенсивность падающего света

$$I_p = I_d + I_t = k_d I_p + k_s I_p + k_t I_p. \quad (8.5)$$

Как было упомянуто выше, пропущенный свет условно делят на диффузно пропущенный и направленно пропущенный. Для простоты будем считать, что доля диффузно пропущенного света мала по сравнению с долей направленно пропущенного света, и опустим из рассмотрения диффузно пропущенный свет.

Направленно пропущенный свет претерпевает преломление на поверхности раздела двух сред. Преломленный луч лежит в одной плоскости с падающим лучом и нормалью в точке падения (рис. 8.2).



Рис. 8.2. Преломление света

Его направление составляет угол θ с нормалью и описывается законом преломления

$$n_0 \sin \theta = n' \sin \theta', \quad (8.6)$$

где n_0 — показатель преломления среды со стороны падающего света, n' — показатель преломления среды с другой стороны поверхности раздела. Для интенсивности направленно пропущенного света примем закон изменения от направления аналогичный закону (8.3) для направленно отраженного света

$$I_t = k_t I_p \cos^n \beta, \quad (8.7)$$

где β — угол отклонения направления пропущенного света от направления в точку наблюдения.

Пришедший к наблюдателю пропущенный свет проходит четное число границ сред и претерпевает столько же раз преломление.

Интенсивность наблюдаемого света.

Реальные поверхности обладают свойствами и диффузного и зеркального отражения, а также могут быть частично прозрачными. В реальных ситуациях присутствует и направленный и рассеянный свет. Рассмотрим вычисление интенсивности света, пришедшего в то точку наблюдения, от непрозрачной поверхности. Она определится суммой трех составляющих

$$I = k_a I_a + I_p k_d \cos \theta + I_p k_s \cos^n \alpha. \quad (8.8)$$

Результаты закраски, полученные с помощью формулы (8.8), обладают большой реалистичностью, но они имеют существенный недостаток. Если две плоские поверхности одного цвета параллельны друг другу и их проекции перекрываются, то поверхности будут закрасены одинаково и их изображения сольются независимо до точки наблюдения. Для устранения этого эффекта интенсивность пришедшего от поверхности света будем считать зависимой от

расстояния и запишем формулу вычисления интенсивности света в виде

$$I = \frac{k_a I_a}{1 + k_r(r - r_0)} + I_p(k_d \cos \theta + k_s \cos^n \alpha), \quad (8.9)$$

где r — расстояние от точки наблюдения до точки поверхности, r_0 — расстояние от точки наблюдения до ближайшей точки объекта, k_r — коэффициент влияния удаленности точки объекта на интенсивность пришедшего от нее света.

При $k_r = 0$ мы получим формулу (8.8). Интенсивность света от ближайшей точки также будет совпадать с (8.8). Расстояние r_0 введено из-за того, что для параллельных проекций точка наблюдения находится на бесконечности. Для центральных проекций можно положить $r_0 = 0$, тогда при удалении всего объекта от точки наблюдения будет наблюдаться уменьшение его яркости.

Формулу (9.8.9) стоит рассматривать как эмпирическую формулу, приемлемую с эстетической точки зрения и достаточно простую. В некоторых случаях вместо этой формулы используется формула

$$I = k_a I_a + I_p \frac{k_d \cos \theta + k_s \cos^n \alpha}{1 + k_r(r - r_0)}, \quad (8.10)$$

Она также позволяет получать достаточно реалистичные изображения.

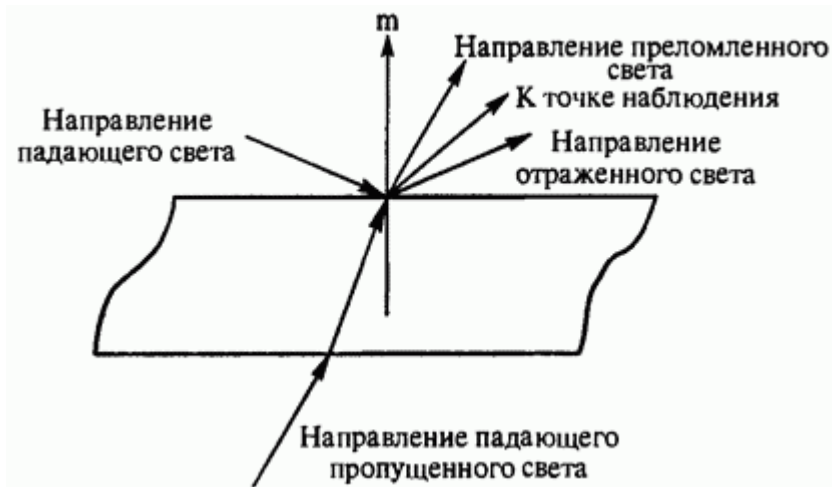


Рис. 8.3

Рассмотрим вычисление интенсивности света, пришедшего в точку наблюдения, от прозрачного объекта, показанного на рис. 8.3. Для построения реалистических изображений интенсивность пришедшего от некоторой точки объекта света опишем формулой

$$\begin{aligned}
 I &= \frac{k_a I_a}{1 + k_r(r - r_0)} + I_p(k_d \cos \theta + k_s \cos^n \alpha) + I_p k_t \cos^n \beta = \\
 &= \frac{k_a I_a}{1 + k_r(r - r_0)} + I_p(k_d \cos \theta + k_s \cos^n \alpha) + I_0 k_t^2 e^{-\mu l} \cos^n \beta.
 \end{aligned}
 \tag{8.11}$$

Последнее слагаемое получено следующим образом. После прохождения лучом света интенсивности I_0 первой поверхности интенсивность пропущенного света равна $k_t I_0$, по достижении второй поверхности его интенсивность будет равна $I_p = k_t I_0 e^{-\mu l}$ наконец, после прохождения второй поверхности его интенсивность станет равной $k_t^2 I_0 e^{-\mu l}$. Свет интенсивности I_0 может прийти от другого источника или может являться направленно отраженной частью света интенсивности I_p , но отраженной от некоторой другой поверхности.

При наличии нескольких источников света каждый из них даст свой вклад в интенсивность пришедшего в точку наблюдения света. Для

нескольких источников света, с интенсивностью каждого I_{pi} , формула (8.9) примет вид

$$I = \frac{k_a I_a}{1 + k_r(r - r_0)} + \sum_i I_{pi} (k_d \cos \theta_i + k_s \cos^n \alpha_i). \quad (8.12)$$

Аналогичным образом изменится формула (8.11) вычисления интенсивности света, пришедшего от прозрачной поверхности. Она примет вид

$$I = \frac{k_a I_a}{1 + k_r(r - r_0)} + \sum_i I_{pi} (k_d \cos \theta_i + k_s \cos^n \alpha_i) + \sum_j I_{pj'} k_t \cos^n \beta_j, \quad (8.13)$$

где $I_{pj'}$ — интенсивность j -го пропущенного луча света (проникающего с другой стороны поверхности).

Торрэнс и Спэрроу создали теоретически обоснованную модель отражающей поверхности. В этой модели предполагается, что поверхность представляет собой совокупность микроскопических идеально отражающих плоскостей. Ориентация граней задается функцией распределения вероятностей Гаусса. Свойства микроскопических плоскостей и их ориентация относительно направления падающего света определяют распределение интенсивности отраженного света по различным направлениям. Модель Торрэнса и Спэрроу хорошо согласуется с опытами, но является более сложной, чем рассмотренная выше модель.

6..9. Описание цвета

Свет и его восприятие человеком представляют собой сложные явления, которые изучены пока не полностью. Из экспериментальных исследований известно, что свет с длиной волны из определенной части спектра воспринимаются глазом в виде света, определенного цвета. Воспринимаемый цвет зависит от источника света, от свойств поверхности объекта и от зрительной системы человека.

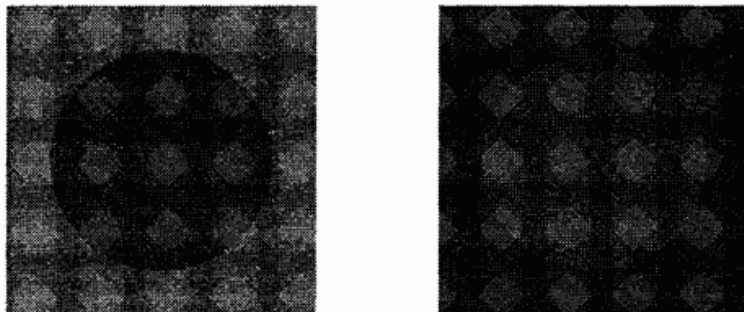


Рис. 9.1

Так, если объект освещен красным светом, а его поверхность отражает только синий свет, то объект будет казаться черным. Зависимость воспринимаемой глазом интенсивности света от зрительной системы человека показана на рис. 9.1.

Яркость круглых областей внутри двух квадратов одинакова, но круг на светлом фоне кажется более темным, чем такой же круг на темном фоне. Это явление называется одновременным контрастом.

Подобное явление существует и при восприятии цветных изображений. В зависимости от цвета и яркости окружения цветные области могут иметь разный оттенок (казаться меняющимися цвет). Кроме того, границы областей постоянной интенсивности кажутся более яркими, чем центральная часть областей. Это явление называется эффектом полос Маха.

Для получения цветных изображений используется представление света в виде трех компонент. Это представление основано на физиологической трехкомпонентной теории цвета, в соответствии с которой в сетчатке глаза имеется три типа колбочек и пик чувствительности каждого из этих типов приходится либо на красный, либо на зеленый, либо на синий цвета. Чувствительность глаза не одинакова к разным цветам: наибольшая она для зеленого цвета, наименьшая — для синего цвета. Кроме того, глаз воспринимает не абсолютное значение интенсивности того или иного цвета, а ее относительное значение. Следовательно, линейное распределение уровней интенсивности приходящего света в зрительной системе

человека становится логарифмическим. Все сказанное показывает, что в результате моделирования иногда можно получить не ту цветовую гамму, которая ожидалась.

Интенсивность падающего света будем представлять в виде красной, зеленой и синей компоненты. Положим, что в белом цвете все три компоненты имеют равные интенсивности (хотя области красного, зеленого и синего цвета видимой части спектра не являются равными). Такая цветовая модель является аддитивной системой смешивания цветов и называется RGB моделью (Red, Green, Blue). Перечисленные цвета являются основными цветами данной модели.

Соответствующими дополнительными цветами являются бирюзовый (Cyan), розовый (Magenta) и желтый (Yellow). Если основной цвет сложить в равной пропорции с его дополнительным цветом, то получим белый цвет. Таким образом, бирюзовый цвет можно представить в виде суммы зеленого и синего цветов, розовый цвет можно представить в виде суммы синего и красного цветов, желтый цвет можно представить в виде суммы красного и зеленого цветов. Остальные оттенки получаются в виде смешения основных цветов в различных (неравных) пропорциях. Эта модель не единственная, но она наиболее распространена, так как представление света в виде взвешенной суммы красного, зеленого и синего цветов согласуется с экспериментальными данными и легко реализуема.

В соответствии с RGB моделью интенсивность падающего и отраженного объектом света должна быть представлена в виде суммы интенсивностей красного, зеленого и синего цветов

$$I = I_R + I_G + I_B. \quad (9.1)$$

Интенсивность каждого цвета определяется по одной из приведенных формул, в которой каждая цветовая компонента описывается своими коэффициентами, например,

$$\begin{aligned} I_R &= \frac{k_{aR} I_a}{1 + k_r(r - r_0)} + I_p(k_{dR} \cos \theta + k_{sR} \cos^n \alpha) + I_p' k_{tR} \cos^n \beta, \\ I_G &= \frac{k_{aG} I_a}{1 + k_r(r - r_0)} + I_p(k_{dG} \cos \theta + k_{sG} \cos^n \alpha) + I_p' k_{tG} \cos^n \beta, \\ I_B &= \frac{k_{aB} I_a}{1 + k_r(r - r_0)} + I_p(k_{dB} \cos \theta + k_{sB} \cos^n \alpha) + I_p' k_{tB} \cos^n \beta. \end{aligned} \quad (9.2)$$

Значения интенсивностей должны быть нормированы при $r=r_0$. Варьируя направлением падающего света и значениями коэффициентов диффузного отражения, зеркального отражения и пропускания для разных компонент света, можно подобрать желаемое изображение объекта. Если источников света несколько, то каждый из них должен быть учтен в формулах (9.2) соответствующими слагаемыми.

Существуют другие физиологические теории зрения, например, теория дополнительных цветов, а также другие цветовые модели, например, модель СМУ, где основными цветами являются бирюзовый, розовый и желтый, а красный, зеленый и синий цвета являются дополнительными. Такая модель является субтрактивной системой смешивания основных цветов.

6.10. Формирование реалистических изображений

Реалистическое растровое изображение геометрического объекта формируется путем определения цвета и яркости его точек при некоторой модели света. Для ускорения этого процесса используется триангуляция поверхностей объекта. В результате задача сводится к закраске триангуляционной сетки объекта (аппроксимирующих треугольников).

В большинстве случаев выполняется построчное сканирование изображения. Для определения отражающей свет точки объекта выполняется пересечение его триангуляционной сетки со сканирующей линией визирования. Эта линия строится из точки изображения перпендикулярно проекционной плоскости (как для параллельной проекции, так и для центральной проекции). Поиск точек пересечения линии и треугольников упрощается, если его выполнять в местной системе координат, связанной с проекционной плоскостью. Для этого перед сканированием аппроксимирующие геометрический

объект треугольника переводятся в местную систему координат. Координаты вершин треугольников преобразуются по формуле (2.3). Компоненты векторов нормалей преобразуются по формуле

$$\begin{bmatrix} x_m \\ y_m \\ z_m \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & 0 \\ y_1 & y_2 & y_3 & 0 \\ z_1 & z_2 & z_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ 1 \end{bmatrix} = \mathbf{A}_0 \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ 1 \end{bmatrix} \quad (10.1)$$

где m_1, m_2, m_3 — компоненты нормали в глобальной системе координат, x_m, y_m, z_m — компоненты нормали в местной системе координат проекционной плоскости. Первые три строки матрицы \mathbf{A}_0 образованы компонентами ортов $\mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z$ местной системы координат. При построении перспективного изображения триангуляционная сетка подвергается дополнительному преобразованию: координаты x и y точек треугольников заменяются, соответственно, координатами x_p и y_p , вычисленными по формулам (3.5), а координата z остается без изменений. Аналогичным образом уточним и нормали \mathbf{m} в вершинах треугольников: они должны быть повернуты на угол

$$\eta = \arccos \left(-\frac{(\mathbf{r} - \mathbf{w}) \cdot \mathbf{i}_z}{|\mathbf{r} - \mathbf{w}|} \right) \quad (10.2)$$

вокруг вектора

$$\mathbf{v} = -\frac{(\mathbf{r} - \mathbf{w}) \times \mathbf{i}_z}{|(\mathbf{r} - \mathbf{w}) \times \mathbf{i}_z|}, \quad (10.3)$$

где \mathbf{w} — радиус-вектор точки наблюдения, \mathbf{r} радиус-вектор точки поверхности, в которой вычислена нормаль, \mathbf{i}_z — нормаль к проекционной плоскости.

Поворот вектора \mathbf{m} описывается формулой

$$\mathbf{m}_p = (\mathbf{v}\mathbf{v}) \cdot \mathbf{m} + \cos \eta (\mathbf{E} - \mathbf{v}\mathbf{v})\mathbf{m} + \sin \eta \mathbf{v} \times \mathbf{m}. \quad (10.4)$$

Сканирующая линия визирования ортогональна проекционной плоскости и в местной системе координат поиск ее пересечения с триангуляционной сеткой сводится к определению треугольника, в проекции которого лежит проекция точки. Последняя задача является двумерной, что уменьшает объем вычислений. Для уменьшения количества просмотров проекции треугольников можно сгруппировать по отдельным областям. На практике используются три основных способа закрашки: однотонная, основанная на интерполяции значений интенсивности (метод Гуро) и основанная на интерполяции нормалей (метод Фонга). Для каждой из них может применяться любая модель света. Тоновое изображение при любом способе закрашки формируется одинаковым образом: для каждой точки изображения определяется точка объекта (или несколько точек, если объект прозрачен), от которой приходит отраженный свет, и по интенсивности света и по оптическим свойствам отражающей поверхности определяются цвет точки изображения. Цвет вычисляется по интенсивностям цветовых компонент модели света.

Однотонная закрашка.

При однотонной закрашке для каждого треугольника вычисляется уровень интенсивности отраженного света, который используется для закрашки всего треугольника. При этом для простоты полагается, что v и a являются постоянными в пределах своего треугольника. Фактически получается, что триангуляционная сетка является реальной моделируемой поверхностью, а не аппроксимацией последней. Однотонная закрашка дает граненое изображение кривых поверхностей (рис. 10.1).

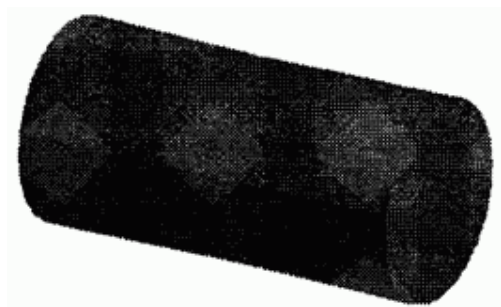


Рис. 10.1. Однотонная закрашка

Интенсивность отраженного света при переходе от одного треугольника к другому меняется скачком. Различие в цвете соседних треугольников хорошо заметно вследствие полос Маха.

Метод Гуро.

Способ закраски, основанный на интерполяции в пределах каждого треугольника значений интенсивности, был предложен Гуро. В методе Гуро по значениям нормалей в вершинах каждого треугольника определяются интенсивности отраженного света в его углах. По интенсивности света I_a, I_b и I_c в вершинах a, b и c некоторого треугольника интенсивность отраженного света от любой его внутренней точки определяется формулой

$$I(a, b, c) = aI_a + bI_b + cI_c, \quad (10.5)$$

где a, b, c — барицентрические координаты произвольной точки треугольника на параметрической плоскости, которые вычисляются по формулам (3.11.4). Метод Гуро хотя и сглаживает перепады интенсивности на границах треугольников, но не позволяет их полностью устранить. Методом Гуро трудно получить четкие блики, особенно, если они небольших размеров.

Метод Фонга.

Метод закраски, предложенный Фонгом, основан на интерполяции в пределах каждого треугольника значений нормалей к поверхности. В каждой вершине треугольника $\mathbf{p}_i(u_i, v_i)$ должна быть вычислена нормаль $\mathbf{m}_i(u_i, v_i)$ к поверхности отображаемого объекта. Нормали в вершинах имеют свое истинное направление, а в пределах каждого треугольника принимается линейный закон их изменения от вершины к вершине. Пусть в вершинах $\mathbf{p}_a(u_a, v_a), \mathbf{p}_b(u_b, v_b), \mathbf{p}_c(u_c, v_c)$ некоторого треугольника нормали равны $\mathbf{m}_a(u_a, v_a), \mathbf{m}_b(u_b, v_b), \mathbf{m}_c(u_c, v_c)$ тогда нормали на поверхности треугольника припишем направление вектора

$$\mathbf{m}(a, b, c) = a\mathbf{m}_a + b\mathbf{m}_b + c\mathbf{m}_c, \quad (10.6)$$

где a, b, c — барицентрические координаты произвольной точки $\mathbf{p}=[u \ v]^T$ треугольника на параметрической плоскости, которые вычисляются по формулам (3.11.4). Несмотря на то, что векторы $\mathbf{m}_a, \mathbf{m}_b, \mathbf{m}_c$ вектор $\mathbf{m}(a, b, c)$, в общем случае, не имеет единичной длины и его следует нормировать. На рис. 10.2 приведен треугольник и нормали в его вершинах.

Значение радиус-вектора точки треугольника определяется формулой

$$\mathbf{p}(a, b, c) = a\mathbf{p}_a + b\mathbf{p}_b + c\mathbf{p}_c. \quad (10.7)$$

Барицентрические координаты a, b и c связаны равенством (3.11.2) и в пределах треугольника принимают неотрицательные значения. Если хотя бы одна из барицентрических координат отрицательна, то это означает, что точка лежит за пределами треугольника. При аппроксимации (10.6) нормали плавно меняют свое направление при переходе от одного треугольника к другому, поэтому на тоновых изображениях поверхность выглядит гладкой.

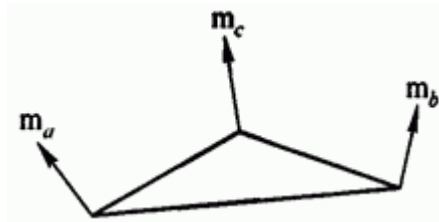


Рис. 10.2. Направление нормалей треугольника

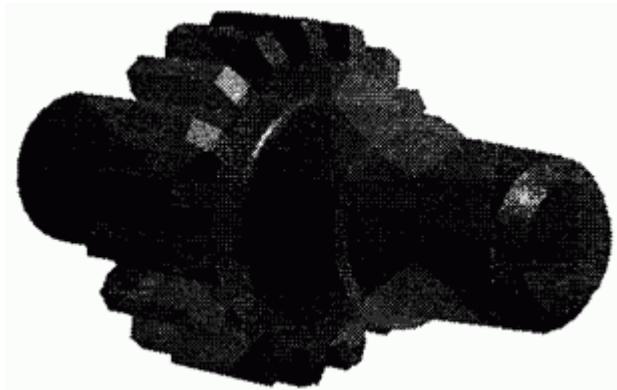


Рис. 10.3. Закраска методом Фонга

Метод Фонга полностью устраняет полосы Маха, однако, значительно увеличивает объем вычислений. На рис. 10.3 приведено изображение детали, полученное методом Фонга.

Детализация поверхностей.

Не все поверхности обладают достаточной гладкостью, например, поверхность апельсина. Некоторые поверхности имеют рисунок, например, поверхность деревянных досок.

Для создания регулярных неровностей на поверхностях в методе закраски Фонга могут использоваться регулярные отклонения нормали.

Для создания нерегулярных неровностей могут использоваться случайные отклонения нормали. На рис. 10.4 и 10.5 приведены тела, нормальям поверхностей которых сообщены случайные отклонения. На рис. 10.4 случайные отклонения нормали от исходного положения достигают 10° , а на рис. 10.5 эти отклонения достигают 60° .

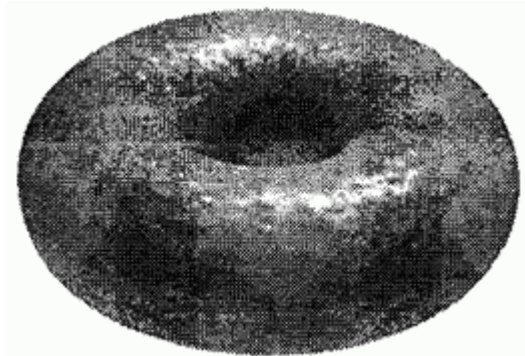


Рис. 10.4. Неровности на поверхности тора

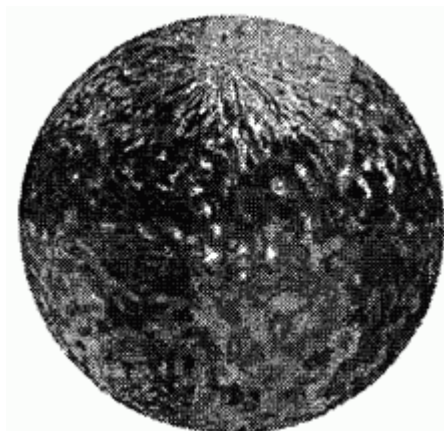


Рис. 10.5. Неровности на поверхности сферы

На поверхность моделируемого объекта может быть нанесен рисунок, подобный рисунку на стальных обоях. Рисунок выполняется цветом, описываемым на параметрической плоскости поверхности. Описание может быть аналитическим или может быть выполнено путем наложения некоторого рисунка на плоскость параметров поверхности. Рисунок в свою очередь может быть растровым изображением. Растровый рисунок будем рассматривать как набор цветных точек на целочисленной плоскости. Для нанесения рисунка на поверхность остается установить функциональную зависимость между

целочисленной плоскостью и плоскостью параметров поверхности. Детализация поверхности выполняется методом Фонга.

Тени.

В рассмотренных способах закраски мы считали, что на точки поверхности падает рассеянный свет интенсивности I_a и направленный свет интенсивности источника I_p .

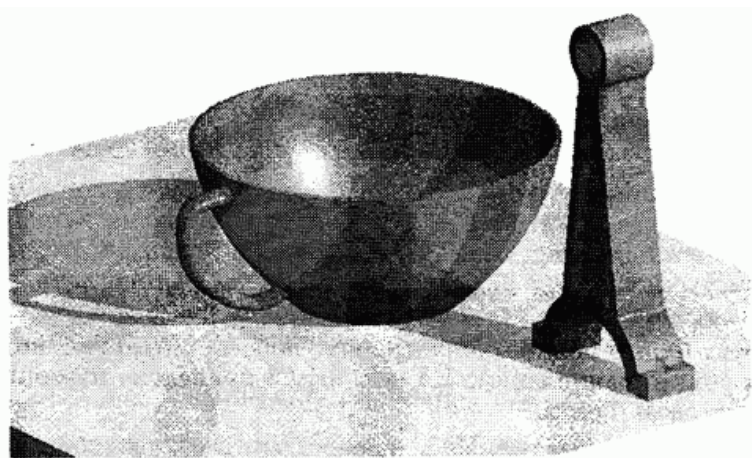


Рис. 10.6. Объекты и их тени

Это всегда справедливо, когда точка наблюдения совпадает с положением источника света.

В общем случае некоторая часть поверхности объекта будет закрыта от источника, т. е. будет находиться в тени. Алгоритм определения затенения точек аналогичен алгоритму определения видимости точек. Если точка поверхности «видна» из точки источника света, то она освещена. Те точки поверхности объекта, которые видны из точки наблюдения и из источника, освещены рассеянным и направленным светом. Точки поверхности объекта, которые видны из точки наблюдения, но не видны из источника света, находятся в тени и освещены только рассеянным светом. Это же справедливо и для нескольких источников света. Естественно, что при построении теней

объем вычислений увеличивается. На рис. 10.6 приведено изображение объектов с тенями.

Четкие тени дают точечные источники света. При освещении объекта распределенными источниками света возникают области полутеней. Полутени возникают в тех точках объекта, из которых видна только часть распределенного источника.

Прозрачность.

Для прозрачных моделей в точку наблюдения приходит свет, отраженный не только от ближайшей поверхности, но и от других поверхностей, которые видны сквозь нее. Если не учитывать преломление света, то в точку наблюдения придет свет от всех треугольников, с которыми пересекается линия визирования. На рис. 10.7 и 10.8 приведены изображения одного и того же объекта с различной прозрачностью материала.



Рис. 10.7. Непрозрачный объект

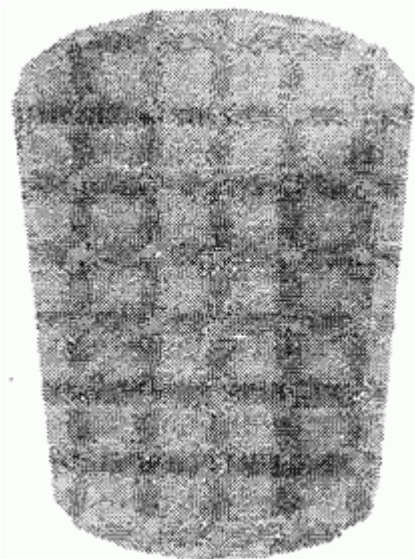


Рис. 10.8. Прозрачный объект

Материал объекта, приведенного на рис. 10.7, непрозрачный. Материал объекта, приведенного на рис. 10.8, отражает 40% падающего света, а пропускает 60% (преломление света не учитывалось).

При моделировании преломления света линия визирования для заслоненных треугольников должна быть смещена. Интенсивность света, пришедшего от каждой поверхности, будет ослаблена из-за поглощения света веществом. Чтобы определить, какие из световых лучей попадают в точку зрения, используют метод трассировки лучей.

Трассирование начинается из точки наблюдения, и лучи света отслеживаются в обратном направлении к каждому источнику света. Луч света, падающий на поверхность, разделяется на несколько основных частей: диффузно пропущенный свет, пропущенный преломленный свет, диффузно отраженный свет и зеркально отраженный свет.

Аналогично луч света, исходящий от поверхности объекта, является суммой двух отраженных составляющих света и двух составляющих света, пришедших из-за поверхности. Таким образом, каждый раз,

когда трассируемый из точки наблюдения луч подходит к поверхности отображаемого объекта, возможно появление нескольких новых лучей, т. е. трассируемый луч расщепляется на несколько составляющих. Диффузно отраженный свет и диффузно пропущенный свет могут прийти в общем случае из бесконечного числа направлений, поэтому обычно трассируют только лучи света, появляющиеся в результате зеркального отражения и преломления. В результате, при подходе трассируемого луча к поверхности прозрачного объекта он расщепляется на две части. Рассеянный свет и диффузно отраженный свет можно получить от ближайшей к точке наблюдения поверхности. Для каждого трассируемого луча нужно получить его пересечение со всеми поверхностями объекта. В точках пересечения по свойствам поверхности определяется, нужно ли расщеплять световой луч.

Как можно видеть, построение реалистических изображений геометрических моделей требует больших вычислительных затрат. Геометрическое моделирование имеет огромные возможности. Через компьютерную графику оно позволяет увидеть и исследовать то, что раньше могло существовать только в воображении.

Вместо заключения

Мы рассмотрели моделирование геометрических свойств окружающих нас и воображаемых объектов, связанные с моделированием математические проблемы и методы их решения. Чтобы создать систему моделирования, нужно обеспечить взаимодействие с создаваемой геометрической моделью. Это призвана обеспечить внутренняя организация модели и ее окружения — редактор системы. Проблема организации внутренней жизни геометрической модели и взаимодействия с ней человека представляет собой отдельную задачу, способы решения которой тесно связаны с программированием.

Программная реализация рассмотренных в книге математических проблем составляет математическое ядро системы моделирования. В объектном подходе к созданию математического ядра точки, кривые линии, поверхности, тела, топологические объекты, операции, вариационные связи, геометрические характеристики, полигоны, триангуляции и другие объекты реализуются в виде классов.

Геометрическое моделирование позволяет существенно сократить время и материальные затраты на производство проектируемых

объектов и повысить их качество. Геометрическое моделирование автоматизирует труд дизайнеров, конструкторов, архитекторов, технологов, позволяет им уйти от рутины и сосредоточиться на творчестве.

СПИСОК ЛИТЕРАТУРЫ

1. В.Ф.Асмус. Проблема интуиции в философии и математике. «Мысль», М.1965.
2. Дж. Барвайс. Введение в логику первого порядка. Справочная книга по математической логике. Ч.1. «Наука», М.1982. Пер. с англ.: Handbook of mathematical logic. J. Barwise (Ed). North-Holland P.C. 1977.
3. Дж.Булос, Р.Джеффри. Вычислимость и логика. М. «Мир» 1994. Пер. с английского: George S. Boolos, Richard C. Jeffrey. Computability and logic. Cambridge University press, 1989.
4. Е.А.Беляев, В.Я.Перминов. Философские и методологические проблемы математики. Изд-во Московского университета, 1981.
5. М.Бунге. Интуиция и наука. «Прогресс», М. 1967. Пер. с английского: M.Bunge. Intuition and Science. New York, 1962.
6. Н.Бурбаки. Начала математики. Ч.1, кн.1. Теория множеств. Мир. М. 1965. Пер. с французского.: Elements de Mathematique par N.Bourbaki. Livre 1. Theorie des ensembles. Troisieme edition, 1958.
7. Е.Вигнер. Непостижимая эффективность математики в естественных науках. УФН, т.94, вып.3, 1968, 535 – 546. Пер. с англ.: E.Wigner. The Unreasonable Effectiveness of Mathematics in the Natural Sciences, Comm. Pure and Appl. Math. 131, 1 (1960).
8. Р.Декарт. Правила для руководства ума. Избранные произведения. М.1950. Пер. с французского: Descartes R. Oeuvres, t. X. Paris, 1908.
9. М.Клайн. Математика. Утрата определённости. М.»Мир», 1984. Пер. с англ.: Morris Kline. MATHEMATICS. The Loss of Certainty. N-Y, Oxford University Press, 1980.
10. С.К.Клини. Введение в метаматематику. ИЛ М. 1957. Пер.с англ. : Introduction tu metamathematics by Stephen Cole Kleene. 1952. D.van Nostrand Company, inc. New York , Toronto.

11. М.Кац, С.Улам. Математика и логика. Ретроспектива и перспективы. «Мир», М. 1971. Пер. с англ.: Mathematics and Logic. Retrospect and Prospects. Mark Кас and Stanislaw M. Ulam. N.-Y. Washington. London. 1968.
12. А.Е. Кононюк. Общая теория познания и созидания. Кн.1. Киев: «Освіта України», 2013. 648 с. ecat.diit.edu.ua:81/ft/index_ru.html
13. А.Е. Кононюк. Общая теория познания и созидания. Кн.2, ч.1. Киев: «Освіта України», 2013. 544 с.
ecat.diit.edu.ua:81/ft/index_ru.html
14. А.Е. Кононюк. Общая теория познания и созидания. Кн.2, ч.2. Киев: «Освіта України», 2013. 644 с.
ecat.diit.edu.ua:81/ft/index_ru.html
15. А.Е. Кононюк. Информациология. Общая теория информации. Кн.1. Киев: «Освіта України», 2011. 476 с.
ecat.diit.edu.ua:81/ft/index_ru.html
16. А.Е. Кононюк. Информациология. Общая теория информации. Кн.2. Киев: «Освіта України», 2011. 476 с.
ecat.diit.edu.ua:81/ft/index_ru.html
17. А.Е. Кононюк. Информациология. Общая теория информации. Кн.3. Киев: «Освіта України», 2011. 412 с.
ecat.diit.edu.ua:81/ft/index_ru.html
18. А.Е. Кононюк. Информациология. Общая теория информации. Кн.4. Киев: «Освіта України», 2011. 488 с.
ecat.diit.edu.ua:81/ft/index_ru.html
19. А.Е. Кононюк. Общая теория понятий. Кн.1. Киев: «Освіта України», 2014. 514с.
20. А.Е. Кононюк. Общая теория понятий. Кн.2. Киев: «Освіта України», 2014. 544с.
21. А.Е. Кононюк. Общая теория понятий. Кн.3. Киев: «Освіта України», 2014. 614с.
22. А.Е. Кононюк. Системология. Общая теория систем. Кн.1. Киев: «Освіта України», 2012. 564с. ecat.diit.edu.ua:81/ft/index_ru.html

23. А.Е. Кононюк. Системология. Общая теория систем. Кн.2. Ч.1. Киев: «Освіта України», 2014. 558с.
ecat.diit.edu.ua:81/ft/index_ru.html
24. А.Е. Кононюк. Системология. Общая теория систем. Кн.2. Ч.2. Киев: «Освіта України», 2014. 658с.
ecat.diit.edu.ua:81/ft/index_ru.html
25. А.Е. Кононюк. Системология. Общая теория систем. Кн.2. Ч.1. Киев: «Освіта України», 2014. 558с.
ecat.diit.edu.ua:81/ft/index_ru.html
26. А.Е. Кононюк. Общая теория распознавания. Кн.1. Киев: «Освіта України», 2012. 584 с. *ecat.diit.edu.ua:81/ft/index_ru.html*
27. А.Е. Кононюк. Общая теория распознавания. Кн.2. Киев: «Освіта України», 2012. 588 с. *ecat.diit.edu.ua:81/ft/index_ru.html*
28. А.Е. Кононюк. Консалтология. Общая теория консалтинга. Кн.1. Киев: «Освіта України», 2013. 448 с.
ecat.diit.edu.ua:81/ft/index_ru.html
29. А.Е. Кононюк. Консалтология. Общая теория консалтинга. Кн.2. Киев: «Освіта України», 2013. 412 с.
ecat.diit.edu.ua:81/ft/index_ru.html
30. А.Е. Кононюк. Консалтология. Общая теория консалтинга. Кн.3. Киев: «Освіта України», 2013. 520 с.
ecat.diit.edu.ua:81/ft/index_ru.html
31. А.Е. Кононюк. Консалтология. Общая теория консалтинга. Кн.4. Киев: «Освіта України», 2013. 508 с.
ecat.diit.edu.ua:81/ft/index_ru.html
32. А.Е. Кононюк. Дискретно-непрерывная математика. Начала. Кн.1. Киев: «Освіта України», 2012. 652с.
ecat.diit.edu.ua:81/ft/index_ru.html
33. А.Е. Кононюк. Дискретно-непрерывная математика. Множества. Кн.2. Ч.1. Киев: «Освіта України», 2012. 452с.
ecat.diit.edu.ua:81/ft/index_ru.html
34. А.Е. Кононюк. Дискретно-непрерывная математика. Множества. Кн.2. Ч.2. Киев: «Освіта України», 2013. 536 с.
ecat.diit.edu.ua:81/ft/index_ru.html
35. А.Е. Кононюк. Дискретно-непрерывная математика. Отношения. Кн.3. Ч. 1. Киев: «Освіта України», 2013. 552с.
ecat.diit.edu.ua:81/ft/index_ru.html

36. А.Е. Кононюк. Дискретно-непрерывная математика. Отношения. Кн.3. Ч. 2. Киев: «Освіта України», 2013. 548 с.
ecat.diit.edu.ua:81/ft/index_ru.html
37. А.Е. Кононюк. Дискретно-непрерывная математика. Алгебры. Кн.4. Ч.1. Киев: «Освіта України», 2011. 452с.
ecat.diit.edu.ua:81/ft/index_ru.html
38. А.Е. Кононюк. Дискретно-непрерывная математика. Алгебры. Кн.4. Ч.2. Киев: «Освіта України», 2011. 668 с.
ecat.diit.edu.ua:81/ft/index_ru.html
39. А.Е. Кононюк. Дискретно-непрерывная математика. Алгебры. Кн.4. Ч.3. Киев: «Освіта України», 2015. 488 с.
http://lib.sumdu.edu.ua/library/DocDescription?doc_id=640902
40. А.Е. Кононюк. Дискретно-непрерывная математика. Алгебры. Кн.4. Ч.4. Киев: «Освіта України», 2015. 548 с.
41. А.Е. Кононюк. Дискретно-непрерывная математика. Алгебры. Кн.4. Ч.5. Киев: «Освіта України», 2015. 528 с.
42. А.Е. Кононюк. Дискретно-непрерывная математика. Алгебры. Кн.4. Ч.6. Киев: «Освіта України», 2015. 608 с.
43. А.Е. Кононюк. Дискретно-непрерывная математика. Матрицы. Кн.5. Ч.1. Киев: «Освіта України», 2013. 612 с.
ecat.diit.edu.ua:81/ft/index_ru.html
44. А.Е. Кононюк. Дискретно-непрерывная математика. Матрицы. Кн.5. Ч.2. Киев: «Освіта України», 2013. 500 с.
ecat.diit.edu.ua:81/ft/index_ru.html
45. А.Е. Кононюк. Дискретно-непрерывная математика. Матрицы. Кн.5. Ч.3. Киев: «Освіта України», 2013. 520 с.
ecat.diit.edu.ua:81/ft/index_ru.html
46. А.Е. Кононюк. Дискретно-непрерывная математика. Матрицы. Кн.5. Ч.4. Киев: «Освіта України», 2013. 508 с.
ecat.diit.edu.ua:81/ft/index_ru.html

47. А.Е. Кононюк. Дискретно-непрерывная математика. Матрицы. Кн.5. Ч.5. Киев: «Освіта України», 2013. 672 с.
ecat.diit.edu.ua:81/ft/index_ru.html
48. А.Е. Кононюк. Дискретно-непрерывная математика. Поверхности. Кн.6. Ч.1. Киев: «Освіта України», 2012. 652с.
ecat.diit.edu.ua:81/ft/index_ru.html
49. А.Е. Кононюк. Дискретно-непрерывная математика. Графы. Кн.7. Ч.1. Киев: «Освіта України», 2014. 652с.
ecat.diit.edu.ua:81/ft/index_ru.html
50. А.Е. Кононюк. Дискретно-непрерывная математика. Графы. Кн.7. Ч.2. Киев: «Освіта України», 2014. 552с.
ecat.diit.edu.ua:81/ft/index_ru.html
51. А.Е. Кононюк. Дискретно-непрерывная математика. Графы. Кн.7. Ч.3. Киев: «Освіта України», 2015. 512с.
ecat.diit.edu.ua:81/ft/index_ru.html
52. А.Е. Кононюк. Дискретно-непрерывная математика. Графы. Кн.7. Ч.4. Киев: «Освіта України», 2015. 552с.
ecat.diit.edu.ua:81/ft/index_ru.html
53. А.Е. Кононюк. Дискретно-непрерывная математика. Графы. Кн.7. Ч.5. Киев: «Освіта України», 2015. 660с.
54. А.Е. Кононюк. Обобщенная теория моделирования. Кн.1. Ч.1. Киев: «Освіта України», 2012. 602с.
ecat.diit.edu.ua:81/ft/index_ru.html
55. А.Е. Кононюк. Обобщенная теория моделирования. Кн.1. Ч.2. Киев: «Освіта України», 2012. 708с. *ecat.diit.edu.ua:81/ft/index_ru.html*
56. А.Е. Кононюк. Обобщенная теория моделирования. Кн.1. Ч.3. Киев: «Освіта України», 2012. 568с. *ecat.diit.edu.ua:81/ft/index_ru.html*
57. А.Е. Кононюк. Обобщенная теория моделирования. Кн.2. Киев: «Освіта України», 2012. 548с. *ecat.diit.edu.ua:81/ft/index_ru.html*
58. А.Е. Кононюк. Обобщенная теория моделирования. Кн.3. Ч.1. Киев: «Освіта України», 2012. 636с. *ecat.diit.edu.ua:81/ft/index_ru.html*
59. А.Е. Кононюк. Обобщенная теория моделирования. Кн.3. Ч.2. Киев: «Освіта України», 2012. 448с. *ecat.diit.edu.ua:81/ft/index_ru.html*
60. А.Е. Кононюк. Обобщенная теория моделирования. Кн.3. Ч.3. Киев: «Освіта України», 2013. 588с. *ecat.diit.edu.ua:81/ft/index_ru.html*
61. А.Е. Кононюк. Основы теории оптимизации. Кн.1. Киев: «Освіта України», 2011. 602с. *ecat.diit.edu.ua:81/ft/index_ru.html*

62. А.Е. Кононюк. Основы теории оптимизации. Кн.2. Ч.1. Киев: «Освіта України», 2011. 552с. ecat.diit.edu.ua:81/ft/index_ru.html
63. А.Е. Кононюк. Основы теории оптимизации. Кн.2. Ч.2. Киев: «Освіта України», 2011. 616с. ecat.diit.edu.ua:81/ft/index_ru.html
64. А.Е. Кононюк. Основы теории оптимизации. Кн.2. Ч.3. Киев: «Освіта України», 2012. 456с. ecat.diit.edu.ua:81/ft/index_ru.html
65. А.Е. Кононюк. Основы теории оптимизации. Кн.2. Ч.4. Киев: «Освіта України», 2012. 512с. ecat.diit.edu.ua:81/ft/index_ru.html
66. А.Е. Кононюк. Основы научных исследований. Кн.1. Киев: «Освіта України», 2011. 508с. ecat.diit.edu.ua:81/ft/index_ru.html
67. А.Е. Кононюк. Основы научных исследований. Кн.2. Киев: «Освіта України», 2011. 452с. ecat.diit.edu.ua:81/ft/index_ru.html
68. А.Е. Кононюк. Основы научных исследований. Кн.3. Киев: «Освіта України», 2011. 456с. ecat.diit.edu.ua:81/ft/index_ru.html
69. А.Е. Кононюк. Основы научных исследований. Кн.4. Киев: «Освіта України», 2011. 456с. ecat.diit.edu.ua:81/ft/index_ru.html
70. А.Е. Кононюк. Общая теория коммуникаций. Кн.1. Киев: «Освіта України», 2014. 488с.
71. А.Е. Кононюк. Нейроні мережі і генетичні алгоритми. Киев: «Корнійчук», 2010. 448с. ecat.diit.edu.ua:81/ft/index_ru.html
72. Кононюк А. Е. Обобщенная теория познания и созидания. [В 2 кн.] Кн. 1 : Начала / А. Е. Кононюк. — Киев : Освіта України, 2013. ecat.diit.edu.ua:81/ft/index_ru.html
73. Кононюк А. Е. Обобщенная теория познания и созидания. [В 2 кн.] Кн. 2 : Теория познания. Ч. 1 / А. Е. Кононюк. — Киев : Освіта України, 2013 ecat.diit.edu.ua:81/ft/index_ru.html
74. Кононюк А. Ю. Вища математика. (Модульна технологія навчання) : навчальний посібник : в 2 кн. / А. Ю. Кононюк. — Київ : КНТ, 2009 — Кн. 1. — 2009. — 702 с. ecat.diit.edu.ua:81/ft/index_ru.html
75. Кононюк А. Ю. Вища математика. (Модульна технологія навчання) : навчальний посібник : в 2 кн. / А. Ю. Кононюк. — Київ :

КНТ, 2009 Кн. 2. — 2009. — 790 с.
ecat.diit.edu.ua:81/ft/index_ru.html

76. А.Е. Кононюк. Дискретно-непрерывная математика. Поверхности. Кн.6. Ч.2. Киев: «Освіта України», 2012. 652с.
<http://www.dut.edu.ua/ua/lib/127/category/96/view/1297>

77. А.Е. Кононюк. Дискретно-непрерывная математика. Пространства. Кн.8. Ч.1. Киев: «Освіта України», 2016. 748 с.
<http://www.dut.edu.ua/ua/lib/1/category/96/view/1439>

78. А.Е. Кононюк. Дискретно-непрерывная математика. Пространства. Кн.8. Ч.2. Киев: «Освіта України», 2016. 480с.
http://lib.sumdu.edu.ua/library/DocDescription?doc_id=640775

79. А.Е. Кононюк. Истины и информация (фундаментальная теория представления истин и информации). К.1. Киев: «Освіта України», 2016. 568с.

80. А.Е. Кононюк. Истины и информация (фундаментальная теория представления истин и информации). К.2. Киев: «Освіта України», 2016. 558с.

81. А.Е. Кононюк. Истины и информация (фундаментальная теория представления истин и информации). К.3. Киев: «Освіта України», 2016. 588с.

82. А.Е. Кононюк. Истины и информация (фундаментальная теория представления истин и информации). К.4. Киев: «Освіта України», 2016. 552с

83. 82. А.Е. Кононюк. Истины и информация (фундаментальная теория представления истин и информации). К.5. Киев: «Освіта України», 2016. 836 с

84. Алберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения.—М.: Мир, 1972 (Ahlberg J.H., Nilson E.N., Walsh J.L. The Theory of Splines and their Applications.—New York: Academic Press, 1967).

85. Березин И.С., Жидков Н.П. Методы вычислений. Т. 1,2.—М.: Наука, 1962, 1966.

86. Будак Б. М., Фомин С.В. Кратные интегралы и ряды.—М.: Наука, 1967.
87. Гардан П., Люка М. Машинная графика и автоматизация конструирования.— М.: Мир, 1987 (*Techniques Graphiques Interactives et C.A.O./par Michel Lucas et Yvon Gar dan.—Prance: Hermes Publishing, 1983*).
88. Гильберт Д., Кон-Фоссен С. Наглядная геометрия.—М.: Наука, 1981 (*Hilbert D., Cohn-Vossen S. Anschauliche Geometrie.—Berlin: 1932*).
89. Де Бор К. Практическое руководство по сплайнам.—М.: Радио и связь, 1985 (*De Boor C. A Practical Guide to Splines.—Berlin: Springer, 1879*).
90. Демидович Б.П., Марон И.А. Основы вычислительной математики.—М.: Наука, 1970.
91. Дубровин Б.А., Новиков С.П., Фоменко А.Т. Современная геометрия.— М.: Наука, 1986.
92. Ильин В. А., Позняк Э.Г. Аналитическая геометрия.—М.: Наука, 1981.
93. Ильин В. А., Позняк Э.Г. Линейная алгебра.—М.: Наука, 1984.
94. Калиткин Н.Н. Численные методы.—М.: Наука, 1978.
95. Ласло М. Вычислительная геометрия и компьютерная графика на C++ К—М.: Бином, 1997 (*Laszlo M. J. Computational Geometry and Computer Graphics in C++.—Prentice Hall, 1996*).
96. Препарата Ф., Шеймос М. Вычислительная геометрия.—М.: Мир, 1989 (*Preparata F. P., Sham os M. Computational Geometry: An Introduction.— New York, Berlin, Tokyo: Springer-Verlag, 1985*).
97. Роджерс Д., Адамс Дж. Математические основы машинной графики.—М.: Машиностроение, 1980 (*Rogers D.F., Adams J. A. Mathematical Elements for Computer Graphics.—McGrow-Hill, 1976*).

□