

Н. Н. Головнов

ГЕОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Издание с компакт-дискom



Москва

Издательство «Физматлит»

2002

ББК 22.151.3

Г 61

УДК 513.7/681.3

ГОЛОВАНОВ Н. Н. Геометрическое моделирование.—М.: Издательство Физико-математической литературы, 2002.—472 с.—ISBN 5-94052-048-0.

Излагаются методы построения математической модели геометрических объектов с помощью компьютера. Даны основные сведения из дифференциальной геометрии, топологии, вариационного исчисления, численных методов, приведена теория *B*-сплайнов. Подробно рассмотрены методы моделирования различных кривых, поверхностей и тел, а также алгоритмы выполнения операций над ними и вычисления их геометрических характеристик. Описаны принципы установления вариационных зависимостей параметров геометрических объектов. Изложены методы компьютерной графики.

Для прикладных математиков, специалистов по системам автоматизированного проектирования и компьютерной графики. Может быть использована студентами соответствующих специальностей.

Научное издание

ГОЛОВАНОВ Николай Николаевич

ГЕОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Редактор *Л. А. Панюшкина*

Компьютерная графика *М. В. Ивановского*

Компьютерная верстка *Г. М. Красниковой*

ИД № 01389 от 30.03.2000

Гигиеническое заключение № 77.99.02.953.Д.003724.07.01

от 05.07.2001

Подписано в печать 18.03.2002. Формат 70x100/16.

Бумага офсетная № 1. Печать офсетная.

Усл. печ. л. 38,23. Усл. -изд. л. 42,05.

Тираж 1000 экз. Заказ № 5293.

Издательство Физико-математической литературы

117071 Москва В-71, Ленинский проспект, 15

Отпечатано с готовых диапозитивов в типографии

издательства "Экономика и жизнь - Гостиный ряд"

г. Калуга Тел. (0842)56-14-96

ISBN 5-94052-048-0



© Н. Н. Голованов, 2002

© Физматлит, оформление, 2002

ОГЛАВЛЕНИЕ

Предисловие	7
Благодарности	8
Глава 1. Основы геометрического моделирования	9
1.1. Описание геометрических объектов	9
1.2*. Преобразование декартовых прямоугольных координат	11
1.3. Модификации векторов и точек	15
1.4. Однородные координаты	20
1.5*. Геометрия кривых линий	22
1.6*. Геометрия двумерных кривых	29
1.7*. Геометрия поверхностей	32
1.8*. Кривизна линий на поверхности	39
1.9*. Тензоры поверхности	46
1.10*. Криволинейные координаты	54
1.11*. Тензоры в криволинейных координатах	62
1.12*. Ортогональные криволинейные координаты	67
1.13. Математическая модель геометрии объектов	70
Глава 2. Моделирование кривых линий	71
2.1. Математическая модель кривых линий	71
2.2. Аналитические линии	72
2.3*. Кривые второго порядка	77
2.4. Сплайны	81
2.5. Кривые Безье	88
2.6. Представление кривых второго порядка кривыми Безье	93
2.7. Рациональные кривые	100
2.8. <i>B</i> -сплайны	105
2.9. NURBS кривые	121
2.10. Линии, базирующиеся на линиях	135
2.11. Составные кривые	138
2.12. Двухмерные кривые	140
2.13. Способы построения линий	143

Глава 3. Моделирование поверхностей	145
3.1. Математическая модель поверхностей	145
3.2. Аналитические поверхности	146
3.3*. Поверхности второго порядка	153
3.4. Поверхности движения	158
3.5. Линейчатые поверхности	163
3.6. Поверхности Кунса	165
3.7. Сплайновые поверхности	169
3.8. Поверхности Безье	175
3.9. Рациональные поверхности	176
3.10. NURBS поверхности	179
3.11. Поверхности треугольной формы	184
3.12. Треугольные поверхности Безье	188
3.13. Треугольные сплайновые поверхности	194
3.14. Поверхности, базирующиеся на поверхностях	201
3.15. Ограниченные контурами поверхности	204
3.16. Способы построения поверхностей	207
 Глава 4. Операции над кривыми и поверхностями	 209
4.1. Выполнение операций	209
4.2. Движение по параметрической области	211
4.3*. Решение системы нелинейных уравнений	215
4.4*. Решение системы линейных уравнений	221
4.5. Проекция точки на линию	232
4.6. Проекция точки на поверхность	234
4.7. Определение точек пересечения линии и поверхности	237
4.8. Определение точек пересечения линий	241
4.9. Построение линий пересечения поверхностей	244
4.10. Поверхность скругления	253
4.11. Поверхность фаски	256
4.12. Определение точек пересечения трех поверхностей	259
4.13. Точность выполнения операций	260
 Глава 5. Топология оболочек	 262
5.1. Топологические объекты	262
5.2*. Эйлерова характеристика оболочек	262
5.3*. Связность оболочек	267
5.4*. Ориентируемость оболочек	269
5.5. Оболочки для моделирования тел	271
5.6. Поверхностное и твердотельное моделирование	275
 Глава 6. Моделирование тел	 276
6.1. Математическая модель тел	276
6.2. Простейшие тела	280

6.3.	Тела, полученные движением плоского контура	287
6.4.	Построение тела по плоским сечениям	292
6.5.	Тело в форме листа	293
6.6.	Булевы операции над телами	294
6.7.	Резка тела поверхностью	305
6.8.	Построение симметричного тела	307
6.9.	Построение эквидистантной оболочки тела	309
6.10.	Построение тонкостенного тела	313
6.11.	Скругление ребер тела	315
6.12.	Построение фасок ребер тела	320
6.13.	Некоторые способы построения тел	320
6.14.	Последовательность моделирования тел	322
Глава 7.	Вариационные связи геометрических объектов	324
7.1.	Наложение вариационных связей	324
7.2.	Фиксирующие связи	327
7.3.	Вариационные связи точек в пространстве	328
7.4.	Вариационные связи точек на кривых и поверхностях	335
7.5.	Алгебраические связи	336
7.6.	Минимизация изменения параметров	337
7.7*.	Условный экстремум функции изменения параметров	340
7.8*.	Вариационный метод определения изменений параметров	345
7.9*.	Геодезические линии	354
7.10.	Вариационные связи двумерных точек	357
7.11.	Вариационные связи двумерных линий	363
7.12.	Формирование и решение системы уравнений связей	372
Глава 8.	Вычисление геометрических характеристик	376
8.1.	Возможности геометрической модели	376
8.2*.	Криволинейные интегралы	377
8.3.	Геометрические характеристики плоских сечений	383
8.4.	Длина и центр масс кривой линии	388
8.5*.	Поверхностные интегралы	389
8.6.	Площадь поверхности, объем и центр масс тела	393
8.7.	Моменты инерции тела	397
8.8*.	Решение кубического уравнения	407
8.9*.	Квадратурные формулы	410
8.10*.	Кубатурные формулы	417
8.11.	Разбиение поверхностей при интегрировании	425
Глава 9.	Компьютерная графика	429
9.1.	Визуализация геометрических объектов	429
9.2.	Параллельные проекции на плоскость	430

9.3.	Центральные проекции на плоскость	433
9.4.	Полигоны кривых и поверхностей	436
9.5.	Линии очерка	439
9.6.	Определение видимой части геометрических объектов	443
9.7.	Триангуляция поверхностей и тел	446
9.8.	Моделирование света	458
9.9.	Описание цвета	463
9.10.	Формирование реалистических изображений	465
	Вместо заключения	470
	Список литературы	471

Посвящается Марии --
другу, любимой, жене

ПРЕДИСЛОВИЕ

Предлагаемая книга предназначена для тех, кто изучает, использует в работе или разрабатывает системы геометрического моделирования. Она посвящена математическому аппарату этих систем. В основу книги положен опыт работы автора над математическим ядром *CAD* системы КОМПАС.

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.

Глава 1 книги посвящена теоретическим основам геометрического моделирования. В гл. 2, 3 описаны методы моделирования различных кривых линий и поверхностей, начиная от элементарных и кончая рациональными кривыми и поверхностями на основе *B*-сплайнов (*NURBS* объектами). Глава 4 посвящена методам выполнения операций над кривыми линиями и поверхностями. В гл. 5, 6 рассмотрены принципы построения твердых тел. Для построения тел привлекаются топологические понятия. В гл. 7 описан механизм установления вариационных связей параметров геометрических объектов. Эти связи представляются в виде нелинейных уравнений относительно зависимых параметров. В гл. 8 рассмотрены методы расчета геометрических характеристик моделей. Последняя глава посвящена компьютерной графике.

Для чтения книги требуется знание математического анализа. Необходимые сведения из других разделов математики включены в главы книги. Соответствующие им параграфы отмечены звездочкой. Подготовленный читатель отмеченные звездочкой параграфы может пропустить. Книга будет полезна всем интересующимся геометрическим моделированием.

Список литературы не претендует на полноту перечня работ по рассматриваемой теме. В нем приведены работы, результаты которых использовались в данной книге.

К книге прилагается компакт-диск с системой КОМПАС-3D LT, которая использует программную реализацию описанных методов.

БЛАГОДАРНОСТИ

Хочу поблагодарить всех, кто помог мне написать эту книгу.

Благодарю своих учителей и преподавателей МГТУ им. Н.Э. Баумана.

Приношу глубокую благодарность В.И. Мяченкову, вдохновившему меня на написание книги.

Благодарю Н.Н. Астанина, Е.В. Бахина, О.Г. Волкова, А.В. Голикова, А.В. Горюхова, А.П. Духвалова, Д.М. Жамойцина, В.Н. Зорина, В.Н. Комяка, Ю.Г. Лысцева, А.О. Максименко, Э.О. Максименко, В.А. Матюшенкова, В.Б. Москалева, И.В. Николаеву, В.П. Панченко, Ю.В. Покидова, И.И. Радионюк, А.П. Столярова, С.Н. Талалу, А.Г. Тимошина, В.Г. Харю, В.И. Шпирько, Т.М. Янкину, всех своих коллег и друзей за помощь и поддержку во время работы над книгой.

Благодарю сотрудников АСКОН за практическое применение описанных в книге методов.

Благодарю Л.А. Панюшкину и А.С. Фурсова за полезные советы, позволившие улучшить книгу.

Благодарю свою семью и своих родителей за заботу и терпение.

Особую благодарность выражаю издательству, донесшему книгу до читателей.

Н. Н. Голованов

Глава 1

ОСНОВЫ ГЕОМЕТРИЧЕСКОГО МОДЕЛИРОВАНИЯ

1.1. Описание геометрических объектов

Геометрические объекты. Мы будем интересоваться формой окружающих предметов, их размерами и взаимным расположением, не вдаваясь в подробности физических свойств. Другими словами, мы будем изучать и моделировать геометрические свойства реальных или воображаемых объектов. Нашей конечной целью является построение математических моделей геометрии этих объектов. 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 позволят 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. Точки, линии, поверхности и тела будем называть *геометрическими объектами* (рис. 1.1.1).

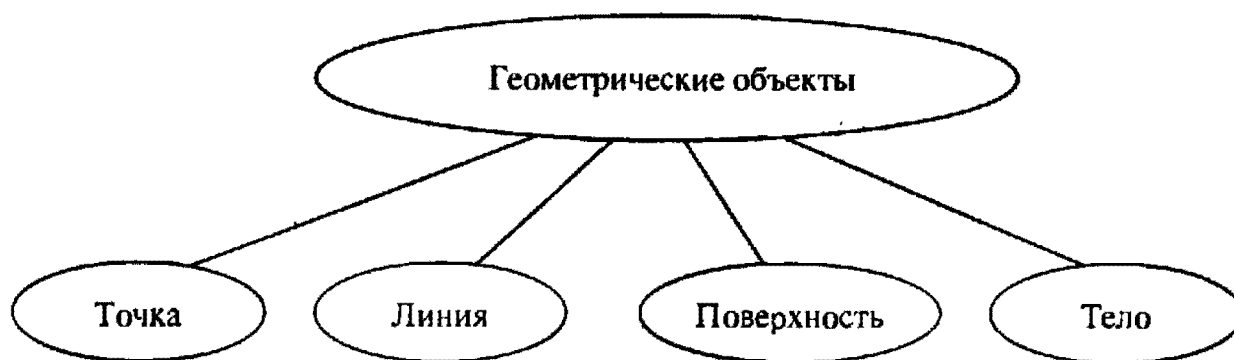


Рис. 1.1.1

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 $O\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$.

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,

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

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 $Qi_1i_2i_3$. Пусть положение некоторой точки R в первой

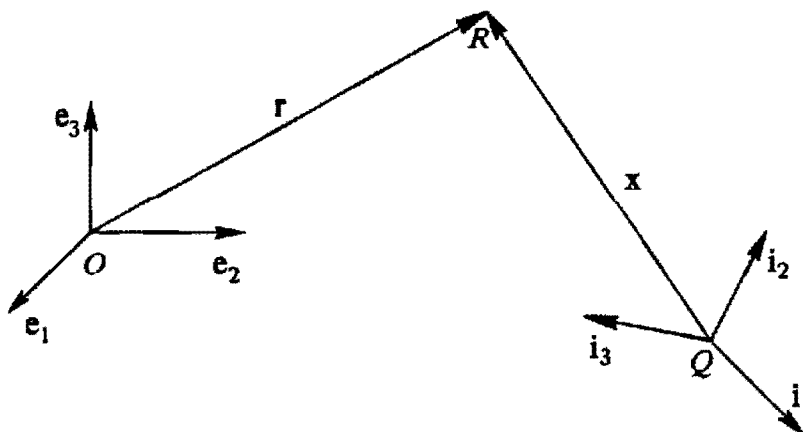


Рис. 1.2.1

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. 1.2.1).

Заметим, что равенствами

$$\begin{aligned} \mathbf{r} &= r_1 \mathbf{e}_1 + r_2 \mathbf{e}_2 + r_3 \mathbf{e}_3, \\ \mathbf{x} &= x_1 \mathbf{i}_1 + x_2 \mathbf{i}_2 + x_3 \mathbf{i}_3 \end{aligned} \quad (1.2.1)$$

точка R описывается в разных системах координат. Обозначим через

$$\mathbf{q} = q_1 \mathbf{e}_1 + q_2 \mathbf{e}_2 + q_3 \mathbf{e}_3 \quad (1.2.2)$$

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 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$:

$$\begin{aligned} \mathbf{i}_1 &= a_{11} \mathbf{e}_1 + a_{12} \mathbf{e}_2 + a_{13} \mathbf{e}_3, \\ \mathbf{i}_2 &= a_{21} \mathbf{e}_1 + a_{22} \mathbf{e}_2 + a_{23} \mathbf{e}_3, \\ \mathbf{i}_3 &= a_{31} \mathbf{e}_1 + a_{32} \mathbf{e}_2 + a_{33} \mathbf{e}_3, \end{aligned} \quad (1.2.3)$$

где a_{ij} , $i, j = 1, 2, 3$ — компоненты ортов $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$ в системе координат $O\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$. 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 $Q\mathbf{i}_1\mathbf{i}_2\mathbf{i}_3$ в систему координат $O\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$:

$$\begin{aligned} r_1 &= a_{11}x_1 + a_{21}x_2 + a_{31}x_3 + q_1, \\ r_2 &= a_{12}x_1 + a_{22}x_2 + a_{32}x_3 + q_2, \\ r_3 &= a_{13}x_1 + a_{23}x_2 + a_{33}x_3 + q_3. \end{aligned} \quad (1.2.4)$$

В матричном представлении (1.2.4) имеет вид

$$\mathbf{r} = \mathbf{A}^T \cdot \mathbf{x} + \mathbf{q}, \quad (1.2.5)$$

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 (1.2.3) скалярно на $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, получим еще одно выражение для компонент матрицы \mathbf{A} :

$$\begin{aligned} a_{11} &= \mathbf{i}_1 \cdot \mathbf{e}_1, & a_{12} &= \mathbf{i}_1 \cdot \mathbf{e}_2, & a_{13} &= \mathbf{i}_1 \cdot \mathbf{e}_3, \\ a_{21} &= \mathbf{i}_2 \cdot \mathbf{e}_1, & a_{22} &= \mathbf{i}_2 \cdot \mathbf{e}_2, & a_{23} &= \mathbf{i}_2 \cdot \mathbf{e}_3, \\ a_{31} &= \mathbf{i}_3 \cdot \mathbf{e}_1, & a_{32} &= \mathbf{i}_3 \cdot \mathbf{e}_2, & a_{33} &= \mathbf{i}_3 \cdot \mathbf{e}_3. \end{aligned} \quad (1.2.6)$$

Из (1.2.6) 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 (1.2.3):

$$\begin{aligned} \mathbf{e}_1 &= a_{11}\mathbf{i}_1 + a_{21}\mathbf{i}_2 + a_{31}\mathbf{i}_3, \\ \mathbf{e}_2 &= a_{12}\mathbf{i}_1 + a_{22}\mathbf{i}_2 + a_{32}\mathbf{i}_3, \\ \mathbf{e}_3 &= a_{13}\mathbf{i}_1 + a_{23}\mathbf{i}_2 + a_{33}\mathbf{i}_3. \end{aligned} \quad (1.2.7)$$

Действительно, если каждое из равенств (1.2.7) умножить скалярно на $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$, то получим равенства (1.2.6).

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 . Выполнив 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 $O\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$ в систему координат $Q\mathbf{i}_1\mathbf{i}_2\mathbf{i}_3$:

$$\begin{aligned} x_1 &= a_{11}r_1 + a_{12}r_2 + a_{13}r_3 + o_1, \\ x_2 &= a_{21}r_1 + a_{22}r_2 + a_{23}r_3 + o_2, \\ x_3 &= a_{31}r_1 + a_{32}r_2 + a_{33}r_3 + o_3. \end{aligned} \quad (1.2.8)$$

Если формулы (1.2.8) 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 (1.2.8), 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 (1.2.8) имеют вид

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{r} + \mathbf{o} = \mathbf{A} \cdot (\mathbf{r} - \mathbf{q}). \quad (1.2.9)$$

Умножив равенство (1.2.9) слева на \mathbf{A}^T и прибавив вектор \mathbf{q} , в соответствии с формулой (1.2.5) мы должны получить вектор \mathbf{r} . Отсюда следует, что

$$\mathbf{A}^T \cdot \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \equiv \mathbf{E}, \quad (1.2.10)$$

$$\mathbf{A}^T = \mathbf{A}^{-1}.$$

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:

$$|\mathbf{A}^T| = |\mathbf{A}| = 1. \quad (1.2.11)$$

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 (1.2.3) и (1.2.7) получим, что преобразования компонент векторов описываются формулами (1.2.4) и (1.2.8), 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 . Пусть положение неко-

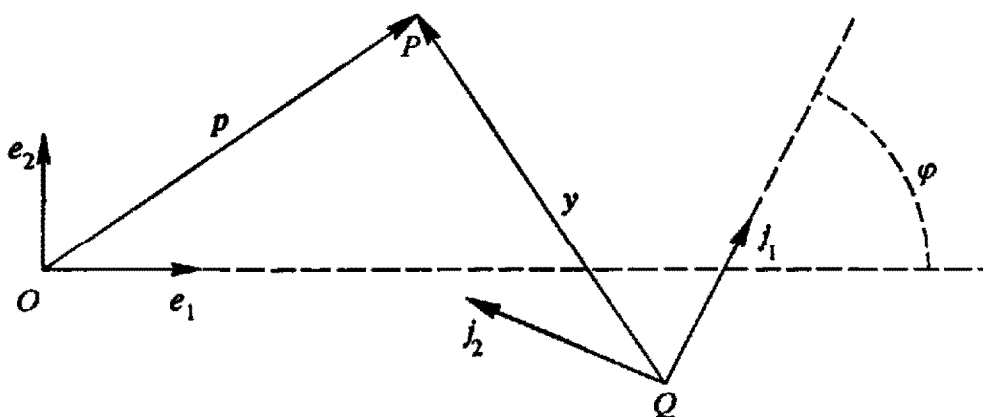


Рис. 1.2.2

торой точки P в первой системе описывается вектором $\mathbf{p} = p_1e_1 + p_2e_2$, а во второй системе положение этой же точки описывается вектором $\mathbf{y} = y_1j_1 + y_2j_2$ (рис. 1.2.2).

Разложим вектор \mathbf{q} , направленный из точки O в точку Q , а также орты j_1, j_2 по ортам e_1, e_2 :

$$\mathbf{q} = q_1e_1 + q_2e_2, \quad (1.2.12)$$

$$j_1 = a_{11}e_1 + a_{12}e_2, \quad (1.2.13)$$

$$j_2 = a_{21}e_1 + a_{22}e_2.$$

Сложив векторы q и y , получим

$$p_1 e_1 + p_2 e_2 = (a_{11}y_1 + a_{21}y_2 + q_1)e_1 + (a_{12}y_1 + a_{22}y_2 + q_2)e_2.$$

Отсюда следуют формулы преобразования координат точки:

$$\begin{aligned} p_1 &= a_{11}y_1 + a_{21}y_2 + q_1, \\ p_2 &= a_{12}y_1 + a_{22}y_2 + q_2. \end{aligned} \quad (1.2.14)$$

Матричное представление (1.2.14) совпадает с (1.2.5): $p = A^T \cdot y + q$. Так как базисные векторы j_1 и j_2 имеют единичную длину, равенствам (1.2.13) можно придать вид

$$\begin{aligned} j_1 &= \cos \varphi e_1 + \sin \varphi e_2, \\ j_2 &= -\sin \varphi e_1 + \cos \varphi e_2, \end{aligned} \quad (1.2.15)$$

если обе координатные системы правые или левые, и

$$\begin{aligned} j_1 &= \cos \varphi e_1 + \sin \varphi e_2, \\ j_2 &= \sin \varphi e_1 - \cos \varphi e_2, \end{aligned}$$

если одна из координатных систем правая, а другая — левая. Угол φ между векторами e_1 и j_1 отсчитывается от e_1 в сторону вектора 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 (1.2.14) при переходе от системы с базисными векторами e_1 и e_2 к системе координат с базисными векторами j_1 и j_2 примет вид

$$\begin{aligned} p_1 &= y_1 \cos \varphi - y_2 \sin \varphi + q_1, \\ p_2 &= y_1 \sin \varphi + y_2 \cos \varphi + q_2. \end{aligned} \quad (1.2.16)$$

Решив систему уравнений (1.2.16) относительно y_1 и y_2 , получим обратное преобразование

$$\begin{aligned} y_1 &= (p_1 - q_1) \cos \varphi + (p_2 - q_2) \sin \varphi, \\ y_2 &= -(p_1 - q_1) \sin \varphi + (p_2 - q_2) \cos \varphi. \end{aligned} \quad (1.2.17)$$

Если в (1.2.17) положить $p_1 = 0$, $p_2 = 0$, то получим координаты точки O в системе с базисными векторами j_1 , j_2 , выраженные через компоненты q_1 и q_2 :

$$\begin{aligned} o_1 &= -q_1 \cos \varphi - q_2 \sin \varphi, \\ o_2 &= q_1 \sin \varphi - q_2 \cos \varphi. \end{aligned} \quad (1.2.18)$$

В системе координат с базисными векторами j_1 , j_2 координаты o_1 и o_2 равны компонентам вектора o , построенного из точки Q в точку O . С учетом формул (1.2.18) преобразование координат точки (1.2.17) 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 (1.2.18) координаты q_1 и q_2 через координаты o_1 и o_2 и подставим их в (1.2.16). В результате получим

$$\begin{aligned} p_1 &= (y_1 - o_1) \cos \varphi - (y_2 - o_2) \sin \varphi, \\ p_2 &= (y_1 - o_1) \sin \varphi + (y_2 - o_2) \cos \varphi. \end{aligned} \quad (1.2.19)$$

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 $p_1e_1 + p_2e_2$, а в системе координат Qj_1j_2 этот же вектор описывается выражением $y_1j_1 + y_2j_2$. С учетом формул (1.2.15) получим, что преобразования компонент двумерных векторов описываются формулами (1.2.16) и (1.2.17), 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 t . Положение точки до модификации будем называть исходным и описывать радиус-вектором r_0 , положение точки после модификации будем называть новым и описывать радиус-вектором r . Положение точки после модификации будет описываться радиус-вектором, равным сумме радиус-вектора ее исходного положения r_0 и вектора сдвига t :

$$r = r_0 + t. \quad (1.3.1)$$

Компоненты вектора r равны сумме соответствующих компонент векторов r_0 и 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 r_0 , а ось вращения определяется

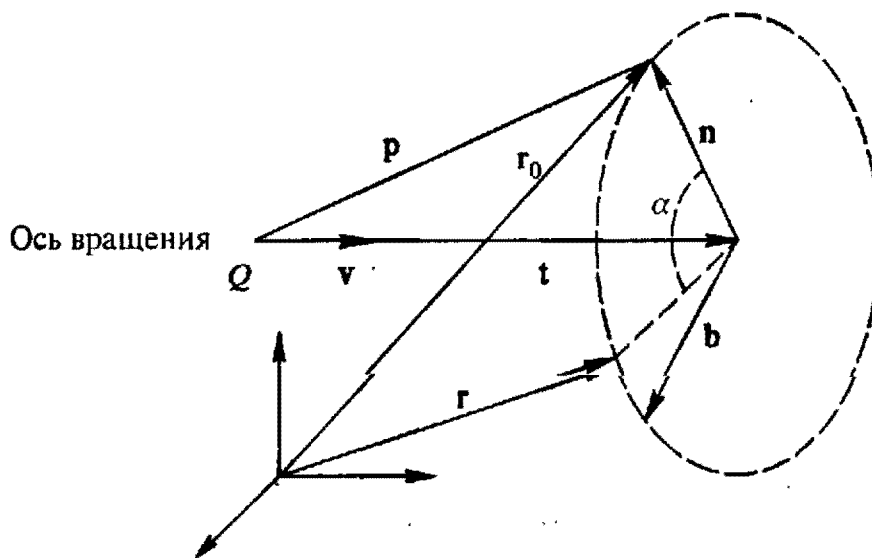


Рис. 1.3.1. Поворот точки вокруг оси

точкой Q и ортом v . Пусть q есть радиус-вектор точки Q . Выполним поворот точки вокруг оси на угол α против часовой стрелки, если взгляд направить навстречу вектору v (рис. 1.3.1).

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

$$\begin{aligned} \mathbf{p} &= \mathbf{t} + \mathbf{n}, \\ \mathbf{t} &= (\mathbf{p} \cdot \mathbf{v})\mathbf{v}, \quad \mathbf{n} = \mathbf{p} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}, \end{aligned} \quad (1.3.2)$$

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

$$\mathbf{b} = \mathbf{v} \times \mathbf{n} = \mathbf{v} \times (\mathbf{p} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}) = \mathbf{v} \times \mathbf{p}. \quad (1.3.3)$$

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

$$\begin{aligned} \mathbf{r} &= \mathbf{q} + \mathbf{t} + \mathbf{n} \cos \alpha + \mathbf{b} \sin \alpha = \\ &= \mathbf{q} + (\mathbf{p} \cdot \mathbf{v})\mathbf{v} + (\mathbf{p} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}) \cos \alpha + \mathbf{v} \times \mathbf{p} \sin \alpha, \end{aligned} \quad (1.3.4)$$

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

$$(\mathbf{p} \cdot \mathbf{v})\mathbf{v} = \begin{bmatrix} v_1 v_1 p_1 + v_1 v_2 p_2 + v_1 v_3 p_3 \\ v_2 v_1 p_1 + v_2 v_2 p_2 + v_2 v_3 p_3 \\ v_3 v_1 p_1 + v_3 v_2 p_2 + v_3 v_3 p_3 \end{bmatrix} = (\mathbf{v}\mathbf{v}) \cdot \mathbf{p}. \quad (1.3.5)$$

С учетом (1.3.5) выражение (1.3.4) примет вид

$$\mathbf{r} = \mathbf{q} + (\mathbf{v}\mathbf{v}) \cdot \mathbf{p} + \cos \alpha (\mathbf{E} - \mathbf{v}\mathbf{v}) \cdot \mathbf{p} + \sin \alpha \mathbf{v} \times \mathbf{p} = \mathbf{q} + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{q}). \quad (1.3.6)$$

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. 1.3.2).

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} , проекции на орт \mathbf{v}

и перпендикулярной плоскости составляющей \mathbf{n} :

$$\mathbf{p} = (\mathbf{p} \cdot \mathbf{u})\mathbf{u} + (\mathbf{p} \cdot \mathbf{v})\mathbf{v} + \mathbf{n}, \quad (1.3.7)$$

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.

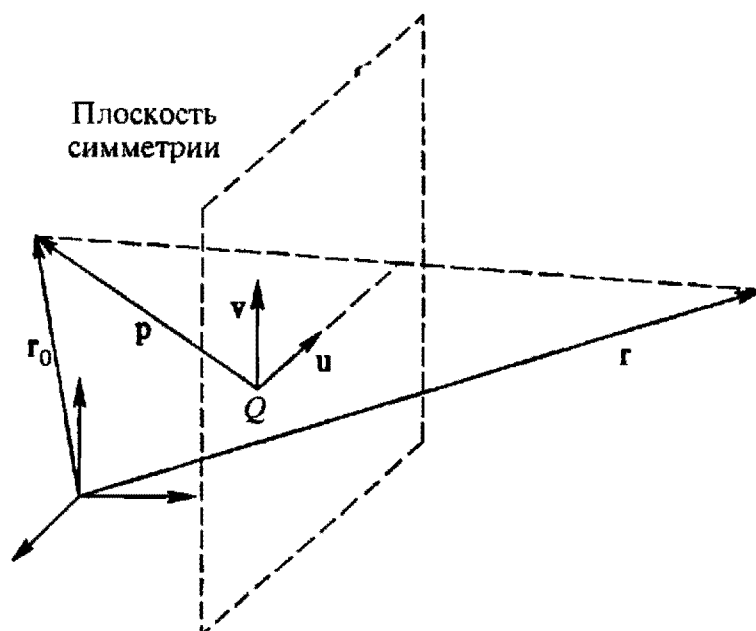


Рис. 1.3.2. Симметрия точки относительно плоскости

Положение симметричной точки будет описываться радиус-вектором

$$\begin{aligned} \mathbf{r} &= \mathbf{q} + (\mathbf{p} \cdot \mathbf{u})\mathbf{u} + (\mathbf{p} \cdot \mathbf{v})\mathbf{v} - (\mathbf{p} - (\mathbf{p} \cdot \mathbf{u})\mathbf{u} - (\mathbf{p} \cdot \mathbf{v})\mathbf{v}) = \\ &= \mathbf{q} + (2\mathbf{u}\mathbf{u} + 2\mathbf{v}\mathbf{v} - \mathbf{E}) \cdot \mathbf{p} = \mathbf{q} + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{q}), \end{aligned} \quad (1.3.8)$$

где матрица $\mathbf{A} = 2\mathbf{u}\mathbf{u} + 2\mathbf{v}\mathbf{v} - \mathbf{E}$ — матрица симметрии, $\mathbf{u}\mathbf{u}$ и $\mathbf{v}\mathbf{v}$ — диадные произведения векторов.

Масштабирование в пространстве. Рассмотрим масштабирование проекций на координатные оси расстояния до точки \mathbf{r}_0 относительно некоторой другой точки \mathbf{Q} , остающейся неподвижной после масштабирования. Пусть \mathbf{q} есть радиус-вектор точки \mathbf{Q} . В общем случае при масштабировании проекции на координатные оси вектора $\mathbf{p} = \mathbf{r}_0 - \mathbf{q}$ 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

$$\mathbf{r} = \mathbf{q} + \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \mathbf{q} + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{q}), \quad (1.3.9)$$

где \mathbf{A} — матрица масштабирования.

Модификация векторов в пространстве. Формулы модификации свободного вектора в пространстве получим из формул модификации радиус-вектора, положив в (1.3.1) $t = 0$, а в (1.3.6), (1.3.8), (1.3.9) — $q = 0$. 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 двумерной точки. Рассмотрим модификации двумерных точек. Векторная формула сдвига двумерной точки на вектор t совпадает с (1.3.1)

$$r = r_0 + t. \quad (1.3.10)$$

Поворот двумерной точки вокруг точки. Повороты двумерной точки выполняются вокруг оси, перпендикулярной плоскости, в которой лежит точка. Пусть начальное положение точки описывается радиус-вектором r_0 , а неподвижная

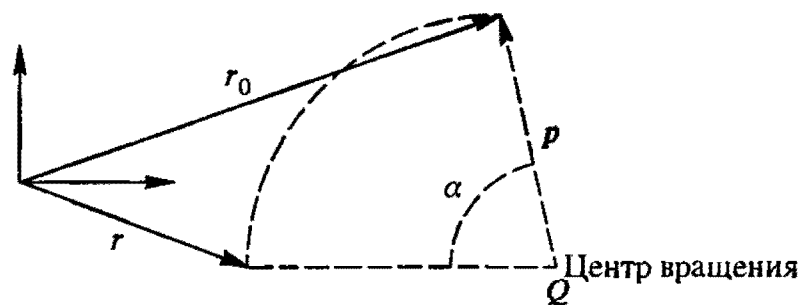


Рис. 1.3.3. Вращение двумерной точки

точка Q имеет радиус-вектор q . Выполним поворот точки в плоскости на угол α против часовой стрелки, если взгляд направить на плоскость (рис. 1.3.3).

Построим вектор $p = r_0 - q$ и вектор b , который имеет длину вектора p и повернут относительно него на прямой угол против часовой стрелки. Вектор b получен с помощью преобразования

$$b = N \cdot p, \quad (1.3.11)$$

где двумерная матрица N имеет вид

$$N = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \quad (1.3.12)$$

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

$$r = q + p \cos \alpha + b \sin \alpha = q + p \cos \alpha + N \cdot p \sin \alpha = q = A \cdot (r_0 - q), \quad (1.3.13)$$

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

Симметрия двумерной точки относительно линии. Определим координаты точки r , симметричной точки r_0 относительно линии. Пусть линия симметрии определяется точкой Q и ортом v (рис. 1.3.4).

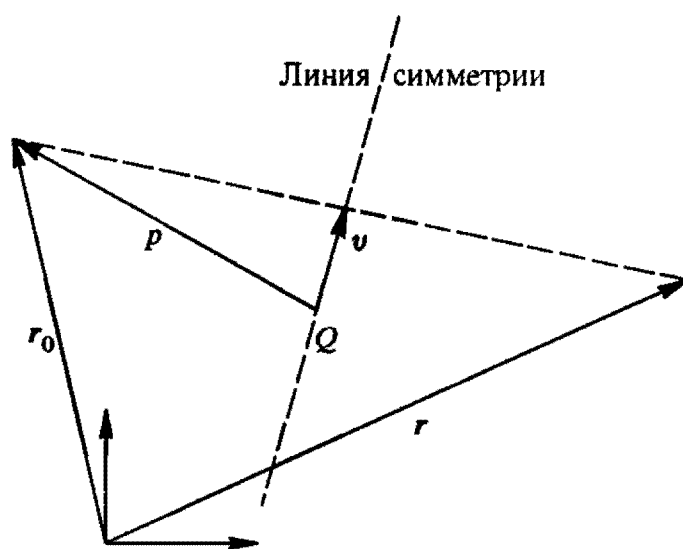


Рис. 1.3.4. Симметрия точки относительно линии

Пусть q есть радиус-вектор точки Q . Построим вектор $p = r_0 - q$ и представим его в виде суммы двух векторов — проекции на орт v и перпендикулярной ему составляющей n :

$$p = (p \cdot v)v + n, \quad (1.3.14)$$

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

$$(p \cdot v)v = \begin{bmatrix} v_1 v_1 p_1 + v_1 v_2 p_2 \\ v_2 v_1 p_1 + v_2 v_2 p_2 \end{bmatrix} = (vv) \cdot p, \quad (1.3.15)$$

где $vv = \begin{bmatrix} v_1 v_1 & v_1 v_2 \\ v_2 v_1 & v_2 v_2 \end{bmatrix}$ — диадное произведение векторов. После зеркального отражения вектора p его нормальная к линии составляющая изменит знак на противоположный. Положение симметричной точки будет описываться радиус-вектором

$$r = q + (p \cdot v)v - (p - (p \cdot v)v) = q + (2vv - E)p = q + A \cdot (r_0 - q), \quad (1.3.16)$$

где матрица $A = 2vv - E$ — матрица симметрии.

Масштабирование в двумерном пространстве. Пусть задана неподвижная точка q и требуется масштабировать относительно нее положения других точек. Положение точки с радиус-вектором r_0 после масштабирования по координатным осям относительно неподвижной точки q будет описываться радиус-вектором

$$r = q + \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = q + A \cdot (r_0 - q), \quad (1.3.17)$$

где p_1 и p_2 — компоненты вектора $p = r_0 - q$, m_1 — коэффициент увеличения компоненты p_1 , m_2 — коэффициент увеличения компоненты p_2 .

Модификация двумерных векторов. Модификации свободного двумерного вектора получим из модификаций радиус-вектора, положив в (1.3.10) $t = 0$, а в (1.3.14), (1.3.16), (1.3.17) — $q = 0$. Вектор в отличие от радиус-вектора не привязан ни к какой точке двумерного пространства и поэтому модификации вектора можно выполнить в местной системе координат, начало которой находится в точке Q , а координатные оси параллельны исходным координатным осям. Этим отличаются модификации вектора и радиус-вектора.

1.4. Однородные координаты

Рассмотренные выше модификации поворота (1.3.6), симметрии (1.3.8), масштабирования (1.3.9) радиус-вектора точки описываются формулами, имеющими одинаковый вид:

$$r = q + A \cdot (r_0 - q) = A \cdot r_0 + (q - A \cdot q) = A \cdot r_0 + t, \quad (1.4.1)$$

где $t = q - A \cdot q$ — преобразованный вектор сдвига. Сдвиг (1.3.1) точки описывается этой же формулой с единичной матрицей A и вектором t , равным вектору сдвига. Аналогичный вид имеет преобразование координат точки (1.2.5). Преобразования координат и модификациям точки можно придать единый простой вид

$$R = A_t \cdot R_0. \quad (1.4.2)$$

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

$$R = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} = [r_1 \quad r_2 \quad r_3 \quad 1]^T. \quad (1.4.3)$$

Матрица A_t представляет собой матрицу A , окаймленную снизу нулями, а справа — вектором сдвига t в расширенном виде

$$A_t = \begin{bmatrix} a_{11} & a_{12} & a_{13} & t_1 \\ a_{21} & a_{22} & a_{23} & t_2 \\ a_{31} & a_{32} & a_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (1.4.4)$$

Индекс матрицы A_t говорит о том, что она является расширенной, и включает вектор сдвига t . Для преобразований (1.4.1) будем использовать запись

$$R = Q + A_0 \cdot (R_0 - Q) = A_t \cdot R_0, \quad (1.4.5)$$

считая, что расширенная матрица A_t включает преобразование по матрице A и сдвиг по вектору $t = q - A \cdot q$. Расширенный вектор 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

вычисления 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

$$\mathbf{R} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{bmatrix} = \begin{bmatrix} wr_1 \\ wr_2 \\ wr_3 \\ w \end{bmatrix} \equiv \begin{bmatrix} w\mathbf{r} \\ w \end{bmatrix} \equiv [w\mathbf{r} \quad w]^\top. \quad (1.4.6)$$

При использовании однородных координат вычисления производятся для однородных компонент R_i , $i = 1, 2, 3, 4$ 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 (1.4.2).

С помощью однородных координат и расширенных матриц удобно описывать преобразования координат. Например, переход из системы координат $Q\mathbf{i}_1\mathbf{i}_2\mathbf{i}_3$ в систему $O\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$ вместо равенств (1.2.4) может быть описан матричным равенством

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & a_{31} & q_1 \\ a_{12} & a_{22} & a_{32} & q_2 \\ a_{13} & a_{23} & a_{33} & q_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} = \mathbf{A}_q^\top \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix}. \quad (1.4.7)$$

Обратное преобразование (1.2.8) может быть описано матричным равенством

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & o_1 \\ a_{21} & a_{22} & a_{23} & o_2 \\ a_{31} & a_{32} & a_{33} & o_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} = \mathbf{A}_o \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix}. \quad (1.4.8)$$

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

$$\begin{aligned} o_1 &= -a_{11}q_1 - a_{12}q_2 - a_{13}q_3, \\ o_2 &= -a_{21}q_1 - a_{22}q_2 - a_{23}q_3, \\ o_3 &= -a_{31}q_1 - a_{32}q_2 - a_{33}q_3. \end{aligned} \quad (1.4.9)$$

Матрицы преобразований (1.4.7) и (1.4.8) удовлетворяют равенству $\mathbf{A}_o \cdot \mathbf{A}_q^\top = \mathbf{A}_q^\top \cdot \mathbf{A}_o = \mathbf{E}$.

Двухмерное пространство. Двухмерные точки в однородных координатах имеют аналогичный вид

$$P = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} wp_1 \\ wp_2 \\ w \end{bmatrix}. \quad (1.4.10)$$

Преобразования двумерных расширенных векторов аналогичны преобразованиям пространственных расширенных векторов. Например, преобразование координат (1.2.16) может быть описано матричным равенством

$$\begin{bmatrix} p_1 \\ p_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \varphi & -\sin \varphi & q_1 \\ \sin \varphi & \cos \varphi & q_2 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ 1 \end{bmatrix}. \quad (1.4.11)$$

Обратное преобразование (1.2.17) может быть описано матричным равенством

$$\begin{bmatrix} y_1 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \varphi & \sin \varphi & o_1 \\ -\sin \varphi & \cos \varphi & o_2 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ 1 \end{bmatrix}, \quad (1.4.12)$$

где компоненты o_1, o_2 связаны с компонентами q_1, q_2 соотношениями (1.2.18).

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

$$\mathbf{r}(t) = \sum_{i=1}^3 r_i(t) \mathbf{e}_i, \quad t_{\min} \leq t \leq t_{\max}. \quad (1.5.1)$$

Представление кривой в виде (1.5.1) 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

$$t = t(v), \quad v_{\min} \leq v \leq v_{\max}, \quad (1.5.2)$$

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 (1.5.1), выраженная через параметр v , будет иметь вид

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

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 $c(v)$, $v_{\min} \leq v \leq v_{\max}$, 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

$$v(t) = v_{\min} \frac{t_{\max} - t}{t_{\max} - t_{\min}} + v_{\max} \frac{t - t_{\min}}{t_{\max} - t_{\min}}. \quad (1.5.3)$$

Вторая кривая теперь будет иметь вид $c = c(t) = c(v(t))$, $t_{\min} \leq t \leq t_{\max}$. 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 $a(t)$ и $c(t)$ и любой скалярной функции $\lambda(t)$ справедливы правила дифференцирования

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

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

$$r^{(n)} \equiv \frac{d^n r}{dt^n} = \sum_{i=1}^3 \frac{d^n r_i(t)}{dt^n} e_i. \quad (1.5.4)$$

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

$$r(t + \Delta t) = r(t) + r'(t)\Delta t + r''(t) \frac{(\Delta t)^2}{2} + \dots + r^{(n)}(t) \frac{(\Delta t)^n}{n!} + q_n \frac{(\Delta t)^{n+1}}{(n+1)!}, \quad (1.5.5)$$

где $q_n = \sum_{i=1}^3 \frac{d^{n+1} r_i}{dt^{n+1}} e_i$. Производные $\frac{d^{n+1} r_i}{dt^{n+1}}$ вычислены при некоторых значениях параметра $t \leq t_i \leq t + \Delta t$. Вектор q_n представляет собой остаточный член ряда Тейлора. Из свойств рядов Тейлора координатной функции $r_i(t)$ следует, что длина вектора 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

$$\lim_{n \rightarrow \infty} |q_n| = 0.$$

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

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

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. 1.5.1).

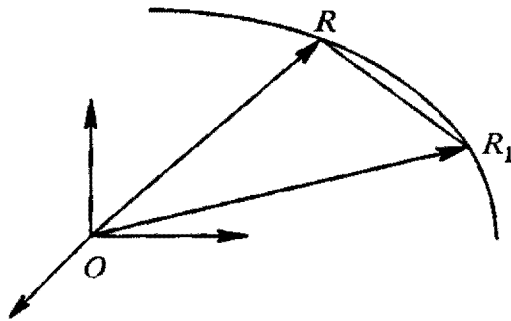


Рис. 1.5.1

Вектор

$$\mathbf{t} = \frac{\Delta\mathbf{r}}{\Delta t} \quad (1.5.6)$$

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 (1.5.6) будет стремиться к касательной к кривой в точке R . Предел отношения (1.5.6) при $\Delta t \rightarrow 0$ является первой производной векторной функции

$$\mathbf{r}' \equiv \sum_{i=1}^3 \frac{dr_i(t)}{dt} \mathbf{e}_i = \lim_{\Delta t \rightarrow 0} \sum_{i=1}^3 \frac{r_i(t + \Delta t) - r_i(t)}{\Delta t} \mathbf{e}_i. \quad (1.5.7)$$

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

$$s = \int_{t_{\min}}^{t_{\max}} \sqrt{\mathbf{r}' \cdot \mathbf{r}'} dt. \quad (1.5.8)$$

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

$$\mathbf{r} = \mathbf{r}(s), \quad (1.5.9)$$

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

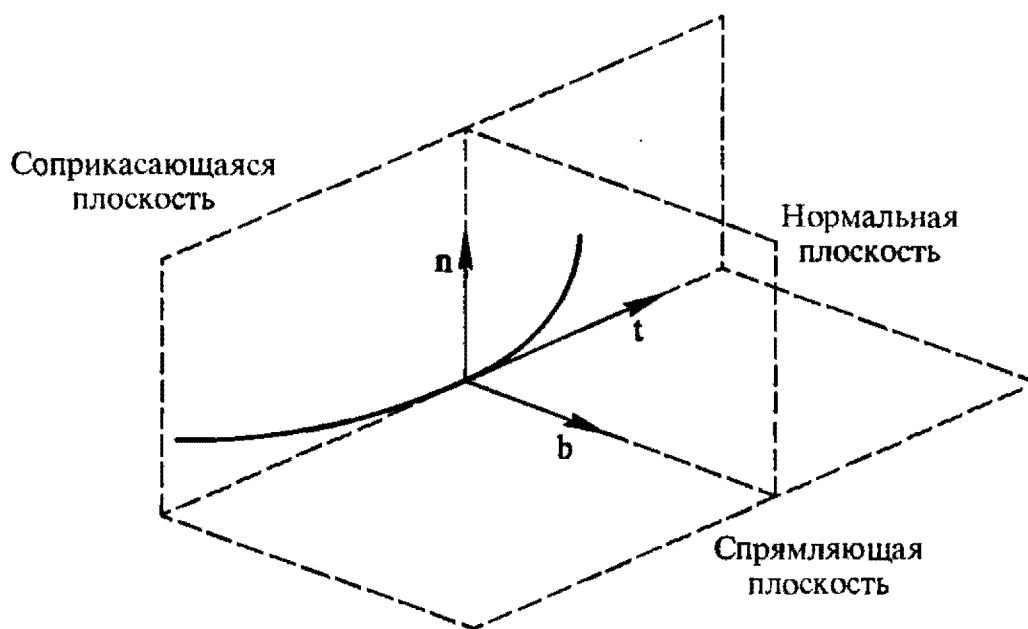


Рис. 1.5.2. Сопровождающий трехгранник кривой

по трем бесконечно близким точкам кривой. Плоскость, перпендикулярная нормальной и соприкасающейся плоскостям, называется *спрямляющей плоскостью* (рис. 1.5.2).

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

$$b = t \times n, \quad t = n \times b, \quad n = b \times t. \quad (1.5.10)$$

Формулы Френе–Серре. Предположим, что нам известно уравнение кривой с натуральной параметризацией (1.5.9). В этом случае

$$\mathbf{t} = \frac{d\mathbf{r}}{ds}. \quad (1.5.11)$$

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

$$\frac{d(\mathbf{t} \cdot \mathbf{t})}{ds} = 2\mathbf{t} \cdot \frac{d\mathbf{t}}{ds} = 2\mathbf{t} \cdot \frac{d^2\mathbf{r}}{ds^2} = 0, \quad (1.5.12)$$

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

$$\frac{d\mathbf{t}}{ds} = \frac{d^2\mathbf{r}}{ds^2} = k\mathbf{n}. \quad (1.5.13)$$

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

$$\begin{aligned} \frac{d(\mathbf{b} \cdot \mathbf{t})}{ds} &= \mathbf{t} \cdot \frac{d\mathbf{b}}{ds} + \mathbf{b} \cdot \frac{d\mathbf{t}}{ds} = \mathbf{t} \cdot \frac{d\mathbf{b}}{ds} + k\mathbf{b} \cdot \mathbf{n} = \mathbf{t} \cdot \frac{d\mathbf{b}}{ds} = 0, \\ \frac{d(\mathbf{b} \cdot \mathbf{b})}{ds} &= 2\mathbf{b} \cdot \frac{d\mathbf{b}}{ds} = 0. \end{aligned} \quad (1.5.14)$$

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

$$\frac{d\mathbf{b}}{ds} = -\chi\mathbf{n}. \quad (1.5.15)$$

Коэффициент χ называется *кручением* кривой линии. Равенства (1.5.13) и (1.5.15) определяют производные ортов \mathbf{t} и \mathbf{b} по длине дуги. Найдем производную нормали по длине дуги

$$\frac{d\mathbf{n}}{ds} = \frac{d(\mathbf{b} \times \mathbf{t})}{ds} = \frac{d\mathbf{b}}{ds} \times \mathbf{t} + \mathbf{b} \times \frac{d\mathbf{t}}{ds} = -\chi\mathbf{n} \times \mathbf{t} + k\mathbf{b} \times \mathbf{n} = \chi\mathbf{b} - k\mathbf{t}. \quad (1.5.16)$$

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

$$\frac{d\mathbf{t}}{ds} = k\mathbf{n}, \quad \frac{d\mathbf{n}}{ds} = \chi\mathbf{b} - k\mathbf{t}, \quad \frac{d\mathbf{b}}{ds} = -\chi\mathbf{n}, \quad (1.5.17)$$

которые известны как *формулы Френе–Серре*. Они выражают производные векторов сопровождающего трехгранника в виде разложения по самим этим векторам. Используя (1.5.10), (1.5.11) и формулы Френе–Серре, выразим векторы \mathbf{t} , \mathbf{n} , \mathbf{b} , кривизну и кручение кривой через производные радиус-вектора кривой по ее длине дуги следующим образом:

$$\begin{aligned} \mathbf{t} &= \frac{d\mathbf{r}}{ds}, \quad \mathbf{n} = \frac{1}{k} \frac{d^2\mathbf{r}}{ds^2}, \quad \mathbf{b} = \frac{d\mathbf{r}}{ds} \times \frac{1}{k} \frac{d^2\mathbf{r}}{ds^2}, \\ k &= \left| \frac{d^2\mathbf{r}}{ds^2} \right|, \quad \chi = \frac{1}{k^2} \left(\frac{d\mathbf{r}}{ds} \times \frac{d^2\mathbf{r}}{ds^2} \right) \cdot \frac{d^3\mathbf{r}}{ds^3}. \end{aligned} \quad (1.5.18)$$

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 i_1 , i_2 и 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

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

Очевидно, что коэффициент пропорциональности k для окружности в формуле (1.5.13) равен $1/\rho$. Таким образом, если мы построим для кривой *соприкасающуюся с ней окружность* (с таким же $d^2\mathbf{r}/ds^2$, как у кривой), то радиус этой окружности будет равен ρ (рис. 1.5.3).

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. Из (1.5.17) получим, что при увеличении параметра на

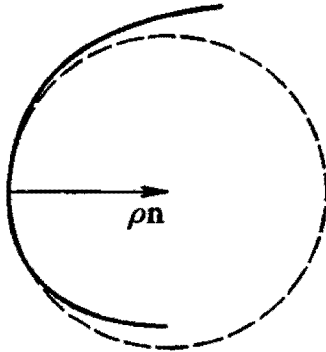


Рис. 1.5.3. Соприкасающаяся с кривой окружность

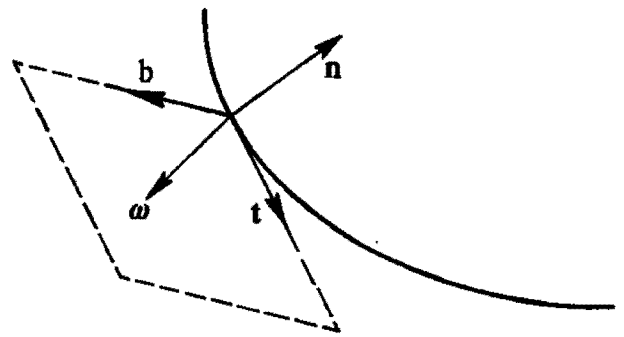


Рис. 1.5.4. Вектор Дарбу

небольшую величину Δt касательный вектор \mathbf{t} повернется в сторону главной нормали \mathbf{n} на угол k , а бинормаль \mathbf{b} повернется в сторону, противоположную главной нормали \mathbf{n} , на угол χ (рис. 1.5.4).

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

$$\omega = kb + \chi t. \quad (1.5.19)$$

Вектор Дарбу придает механический смысл формулам Френе–Серре (1.5.17), с использованием которого последние имеют вид

$$\frac{dt}{ds} = \omega \times t, \quad \frac{dn}{ds} = \omega \times n, \quad \frac{db}{ds} = \omega \times b.$$

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

$$k = k(s) > 0, \quad \chi = \chi(s). \quad (1.5.20)$$

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 (1.5.20) называются *натуральными уравнениями кривой*. Если две кривые имеют одинаковые натуральные уравнения, то они идентичны и отличаются только положением и ориентацией в пространстве. Функции $s(t)$, $k(t)$ и $\chi(t)$ однозначно определяют кривую с точностью до положения и ориентации в пространстве.

Как правило, параметризация кривой не является натуральной. Определим орты t , n , 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:

$$\mathbf{r}' \equiv \frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}}{ds} \frac{ds}{dt} = s' \mathbf{t}, \quad (1.5.21)$$

$$\mathbf{r}'' \equiv \frac{d^2\mathbf{r}}{dt^2} = \frac{d(s' \mathbf{t})}{dt} = \frac{d^2s}{dt^2} \mathbf{t} + \frac{dt}{ds} \left(\frac{ds}{dt} \right)^2 = s'' \mathbf{t} + s'^2 k \mathbf{n}, \quad (1.5.22)$$

$$\begin{aligned} \mathbf{r}''' \equiv \frac{d^3\mathbf{r}}{dt^3} &= \frac{d(s'' \mathbf{t} + s'^2 k \mathbf{n})}{dt} = \frac{d^3s}{dt^3} \mathbf{t} + 3k \mathbf{n} \frac{ds}{dt} \frac{d^2s}{dt^2} + \mathbf{n} \left(\frac{ds}{dt} \right)^3 \frac{dk}{ds} + (\chi \mathbf{b} - k \mathbf{t}) k \left(\frac{ds}{dt} \right)^3 = \\ &= (s''' - k^2 s'^3) \mathbf{t} + \left(3ks''s' + s'^3 \frac{dk}{dt} \right) \mathbf{n} + k\chi s'^3 \mathbf{b}, \end{aligned} \quad (1.5.23)$$

где s' , s'' и s''' — первая, вторая и третья производные длины дуги по параметру кривой. Из (1.5.21) получим формулу для вычисления производной длины дуги по параметру и формулу для вычисления касательного вектора t :

$$s' = |\mathbf{r}'| = \sqrt{\frac{d\mathbf{r}}{dt} \cdot \frac{d\mathbf{r}}{dt}}, \quad (1.5.24)$$

$$\mathbf{t} = \frac{d\mathbf{r}/dt}{|d\mathbf{r}/dt|} \equiv \frac{d\mathbf{r}/dt}{ds/dt} = \frac{\mathbf{r}'}{s'}. \quad (1.5.25)$$

Умножив (1.5.21) векторно на (1.5.22), получим формулу для определения кривизны кривой и направления бинормали

$$k\mathbf{b} = \frac{\mathbf{r}' \times \mathbf{r}''}{s'^3}. \quad (1.5.26)$$

Умножив (1.5.26) векторно на (1.5.25) 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

$$k\mathbf{n} = \frac{(\mathbf{r}' \times \mathbf{r}'') \times \mathbf{r}'}{s'^4} = \frac{\mathbf{r}''}{s'^2} - \frac{\mathbf{r}' \cdot \mathbf{r}''}{s'^4} \mathbf{r}'. \quad (1.5.27)$$

Из правой части равенства (1.5.27) 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 (1.5.26),

$$k = \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|^3}. \quad (1.5.28)$$

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

$$\rho = \frac{|\mathbf{r}'|^3}{|\mathbf{r}' \times \mathbf{r}''|}. \quad (1.5.28.1)$$

Умножив (1.5.26) скалярно на (1.5.23), получим формулу для определения кручения кривой

$$\chi = \frac{(\mathbf{r}' \times \mathbf{r}'') \cdot \mathbf{r}'''}{k^2 s'^6} = \frac{(\mathbf{r}' \times \mathbf{r}'') \cdot \mathbf{r}'''}{|\mathbf{r}' \times \mathbf{r}''|^2}. \quad (1.5.29)$$

Если кривизна кривой в данной точке не равна нулю, то, поделив обе части равенств (1.5.26) и (1.5.27) на кривизну (1.5.28), получим бинормаль и нормаль:

$$\mathbf{b} = \frac{\mathbf{r}' \times \mathbf{r}''}{|\mathbf{r}' \times \mathbf{r}''|^2}, \quad (1.5.30)$$

$$\mathbf{n} = \frac{\mathbf{r}''(\mathbf{r}' \cdot \mathbf{r}') - \mathbf{r}'(\mathbf{r}' \cdot \mathbf{r}'')}{|\mathbf{r}' \times \mathbf{r}''|^2 s'}. \quad (1.5.31)$$

Если известна векторная функция линии (1.5.1), то формулы (1.5.24)–(1.5.31) позволяют получить всю геометрическую информацию о кривой линии.

Из формул (1.5.28) и (1.5.29) 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}'' и \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 \mathbf{e}_1 и \mathbf{e}_2 . Компоненты двумерных векторов будем обозначать через x и y . Большинство

формулы 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

$$\mathbf{r}(t) = x(t)\mathbf{e}_1 + y(t)\mathbf{e}_2. \quad (1.6.1)$$

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

$$\frac{d\mathbf{t}}{ds} = k\mathbf{n}, \quad \frac{d\mathbf{n}}{ds} = -k\mathbf{t}, \quad (1.6.2)$$

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

$$k = \left| \frac{d^2\mathbf{r}}{ds^2} \right|. \quad (1.6.3)$$

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

$$\mathbf{r}_1 \times \mathbf{r}_2 = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = (x_1 y_2 - x_2 y_1). \quad (1.6.4)$$

Выразим через координаты x и y производную длины дуги и кривизну двумерной кривой, используя формулы (1.5.24) и (1.5.26),

$$s' = |\mathbf{r}'| = \sqrt{\mathbf{r}' \cdot \mathbf{r}'} = \sqrt{x'^2 + y'^2}, \quad (1.6.5)$$

$$k = \frac{\mathbf{r}' \times \mathbf{r}''}{|\mathbf{r}'|^3} = \frac{y''x' - x''y'}{(\sqrt{x'^2 + y'^2})^3}. \quad (1.6.6)$$

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:

$$\mathbf{a}(t) = \mathbf{r}(t) + \frac{1}{k}\mathbf{n}(t) = \mathbf{r} + \frac{1}{k^2} \frac{d^2\mathbf{r}}{ds^2}. \quad (1.6.7)$$

При произвольной параметризации кривой нормаль выразим с помощью равенства (1.5.27) и получим

$$\mathbf{a}(t) = \mathbf{r} + \frac{\mathbf{r}''(\mathbf{r}' \cdot \mathbf{r}') - \mathbf{r}'(\mathbf{r}' \cdot \mathbf{r}'')}{k^2 s'^4} = \mathbf{r} + \frac{\mathbf{r}''(\mathbf{r}' \cdot \mathbf{r}') - \mathbf{r}'(\mathbf{r}' \cdot \mathbf{r}'')}{|\mathbf{r}' \times \mathbf{r}''|^2} (\mathbf{r}' \cdot \mathbf{r}'). \quad (1.6.8)$$

Для двумерной кривой $r(t) = [x(t) \ y(t)]^T$ запишем векторное равенство (1.6.8) отдельно для каждой координаты эволюты $a(t) = [x_a(t) \ y_a(t)]^T$:

$$\begin{aligned} x_a(t) &= x + \frac{x''(x'^2 + y'^2) - x'(x'x'' + y'y'')}{(y''x' - x''y')^2}(x'^2 + y'^2) = x + y' \frac{x'^2 + y'^2}{x''y' - y''x'}, \\ y_a(t) &= y + \frac{y''(x'^2 + y'^2) - y'(x'x'' + y'y'')}{(y''x' - x''y')^2}(x'^2 + y'^2) = y + x' \frac{x'^2 + y'^2}{y''x' - x''y'}. \end{aligned} \quad (1.6.9)$$

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

$$r(s) = a - (s - s_0)t = a - (s - s_0)t, \quad (1.6.10)$$

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. 1.6.1 показана кривая $a(s)$ и ее эвольвента.

Эвольвента строится следующим образом. Положим, что длина дуги кривой $a(s)$ отсчитывается от точки C_0 . Для получения точки M эвольвенты, соответствующей некоторой точке C_i исходной кривой, построим в точке C_i касательную и отложим на

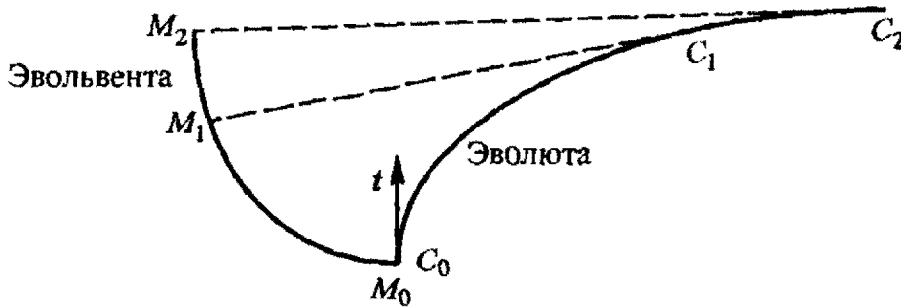


Рис. 1.6.1. Эволюта и эвольвента

ней с учетом знака отрезок длиной, равной длине дуги C_iC_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 (1.6.10) покажем, что эволюта кривой $r(s)$ 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 эволюты выразится правой частью формулы (1.6.8). Подставим в формулу эволюты (1.6.8) значения векторной функции r и ее производных, выраженные через векторную функцию a и ее производные по параметру s :

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

где s — длина дуги кривой $\alpha(s)$, k — ее кривизна, n — главная нормаль и t — касательная кривой $\alpha(s)$. В результате получим

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

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 $Oe_1e_2e_3$ поверхность можно описать радиус-вектором

$$r(u, v) = \sum_{i=1}^3 r_i(u, v)e_i, \quad u, v \in \Omega. \quad (1.7.1)$$

Представление поверхности в виде (1.7.1) 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

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

зависящим от параметров u_0 и v_0 . Новые параметры u_0 и 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 $r(u, v)$ 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

$$\begin{aligned} r_1 &= \frac{\partial r}{\partial u}, & r_2 &= \frac{\partial r}{\partial v}, \\ r_{11} &= \frac{\partial^2 r}{\partial u^2}, & r_{22} &= \frac{\partial^2 r}{\partial v^2}, \\ r_{12} &= \frac{\partial^2 r}{\partial u \partial v} = r_{21} = \frac{\partial^2 r}{\partial v \partial u}. \end{aligned} \quad (1.7.2)$$

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 r_1 и r_2 не коллинеарны и не обращаются в нуль. Такую точку поверхности $r(u, v)$ 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 $r(u, v)$, 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:

$$\frac{\partial r}{\partial u_0} = \frac{\partial r}{\partial u} \frac{\partial u}{\partial u_0} + \frac{\partial r}{\partial v} \frac{\partial v}{\partial u_0} = r_1 \frac{\partial u}{\partial u_0} + r_2 \frac{\partial v}{\partial u_0},$$

$$\frac{\partial r}{\partial v_0} = \frac{\partial r}{\partial u} \frac{\partial u}{\partial v_0} + \frac{\partial r}{\partial v} \frac{\partial v}{\partial v_0} = r_1 \frac{\partial u}{\partial v_0} + r_2 \frac{\partial v}{\partial v_0},$$

$$\frac{\partial^2 r}{\partial u_0^2} = r_{11} \left(\frac{\partial u}{\partial u_0} \right)^2 + 2r_{12} \frac{\partial u}{\partial u_0} \frac{\partial v}{\partial u_0} + r_{22} \left(\frac{\partial v}{\partial u_0} \right)^2 + r_1 \frac{\partial^2 u}{\partial u_0^2} + r_2 \frac{\partial^2 v}{\partial u_0^2},$$

$$\frac{\partial^2 r}{\partial v_0^2} = r_{11} \left(\frac{\partial u}{\partial v_0} \right)^2 + 2r_{12} \frac{\partial u}{\partial v_0} \frac{\partial v}{\partial v_0} + r_{22} \left(\frac{\partial v}{\partial v_0} \right)^2 + r_1 \frac{\partial^2 u}{\partial v_0^2} + r_2 \frac{\partial^2 v}{\partial v_0^2},$$

$$\frac{\partial^2 r}{\partial u_0 \partial v_0} = r_{11} \frac{\partial u}{\partial u_0} \frac{\partial u}{\partial v_0} + r_{12} \frac{\partial u}{\partial u_0} \frac{\partial v}{\partial v_0} +$$

$$+ r_{21} \frac{\partial u}{\partial v_0} \frac{\partial v}{\partial u_0} + r_{22} \frac{\partial v}{\partial u_0} \frac{\partial v}{\partial v_0} + r_1 \frac{\partial^2 u}{\partial u_0 \partial v_0} + r_2 \frac{\partial^2 v}{\partial u_0 \partial v_0}.$$

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 r_1 и r_2 радиус-вектора поверхности представляют собой векторы, касательные к соответствующим координатным линиям. Для поверхности можно построить семейство

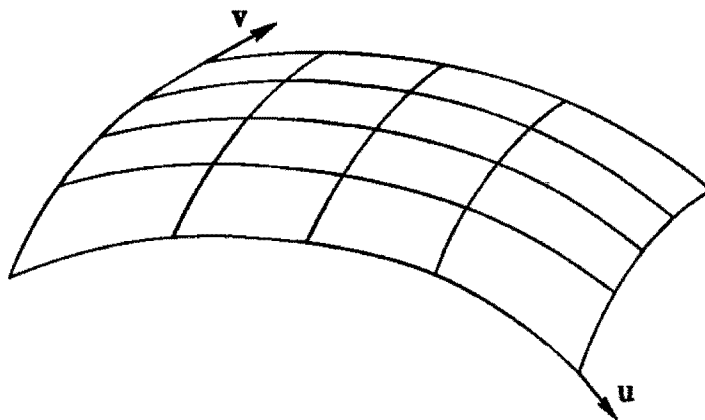


Рис. 1.7.1. Координатная сетка поверхности

u -кривых при различных значениях v , и семейство u -кривых при различных значениях u . Два таких семейства кривых образуют координатную сетку на поверхности (рис. 1.7.1). 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

$$u = u(t), \quad v = v(t). \quad (1.7.3)$$

Радиус-вектор линии на поверхности будет описываться зависимостью

$$\mathbf{r}(t) = \mathbf{r}(u(t), v(t)). \quad (1.7.4)$$

Через каждую точку поверхности можно провести множество различных кривых (1.7.4). Найдем дифференциал радиус-вектора (1.7.4) вдоль некоторой кривой на поверхности:

$$d\mathbf{r} = \mathbf{r}_1(u, v) du + \mathbf{r}_2(u, v) dv, \quad (1.7.5)$$

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 (1.7.4)

$$\frac{d\mathbf{r}}{dt} = \mathbf{r}_1 \frac{du}{dt} + \mathbf{r}_2 \frac{dv}{dt} \equiv \mathbf{r}_1 u' + \mathbf{r}_2 v' \quad (1.7.6)$$

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 (1.7.3), 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*

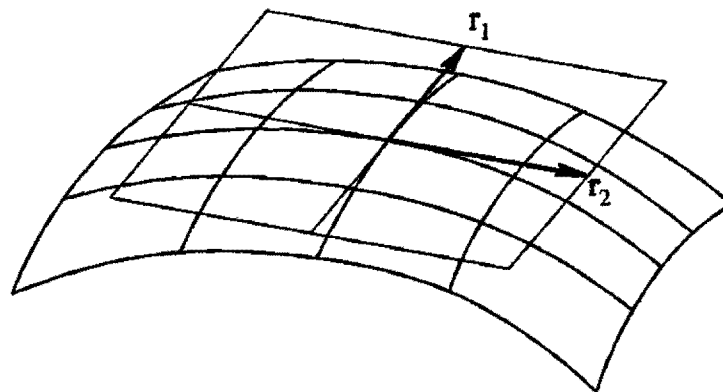


Рис. 1.7.2. Касательная плоскость и производные радиус-вектора поверхности

поверхности в данной точке (рис. 1.7.2). Дифференциал (1.7.5) имеет то же направление, что и производная (1.7.6). Из (1.7.5) видно, что направление дифференциала и производной (1.7.6) зависят от отношения $du:dv = u':v'$.

Пространственной кривой (1.7.4) соответствует двумерная 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 (1.7.3) 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 (1.7.5)

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

Квадрат длины бесконечно малой дуги равен

$$ds^2 = (\mathbf{r}_1 du + \mathbf{r}_2 dv) \cdot (\mathbf{r}_1 du + \mathbf{r}_2 dv) = \mathbf{r}_1 \cdot \mathbf{r}_1 du^2 + 2\mathbf{r}_1 \cdot \mathbf{r}_2 du dv + \mathbf{r}_2 \cdot \mathbf{r}_2 dv^2. \quad (1.7.7)$$

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

$$\begin{aligned} g_{11} &= \mathbf{r}_1 \cdot \mathbf{r}_1, \\ g_{12} &= g_{21} = \mathbf{r}_1 \cdot \mathbf{r}_2 = \mathbf{r}_2 \cdot \mathbf{r}_1, \\ g_{22} &= \mathbf{r}_2 \cdot \mathbf{r}_2. \end{aligned} \quad (1.7.8)$$

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

$$ds^2 = g_{11} du^2 + 2g_{12} du dv + g_{22} dv^2. \quad (1.7.9)$$

Выражение в правой части (1.7.9) является квадратичной формой дифференциалов du и 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 (1.7.9) можно записать в матричном виде

$$ds^2 = [du \ dv] \cdot \mathbf{G} \cdot [du \ dv]^T, \quad (1.7.10)$$

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

$$s = \int_{t_0}^{t_1} \sqrt{g_{11}u'u' + 2g_{12}u'v' + g_{22}v'v'} dt. \quad (1.7.11)$$

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 и dv дифференциалы параметров поверхности, соответствующие бесконечно малому смещению вдоль первой кривой на поверхности, а через δu и δv — 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:

$$\begin{aligned} \cos(\widehat{d\mathbf{r} \ \delta\mathbf{r}}) &= \frac{d\mathbf{r} \cdot \delta\mathbf{r}}{|d\mathbf{r}| |\delta\mathbf{r}|} = \\ &= \frac{g_{11} du \delta u + g_{12} du \delta v + g_{21} dv \delta u + g_{22} dv \delta v}{\sqrt{(g_{11} du du + 2g_{12} du dv + g_{22} dv dv)(g_{11} \delta u \delta u + 2g_{12} \delta u \delta v + g_{22} \delta v \delta v)}}. \end{aligned} \quad (1.7.12)$$

Выражение (1.7.12) позволяет найти угол φ между координатной u -линией и v -линией в некоторой точке поверхности, если мы положим в ней $du \neq 0$, $dv = 0$, $\delta u = 0$, $\delta v \neq 0$:

$$\cos \varphi = \frac{g_{12}}{\sqrt{g_{11}g_{22}}}. \quad (1.7.13)$$

Знак косинуса угла мы взяли совпадающим со знаком g_{12} , следовательно, исходя из обозначения (1.7.8), 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

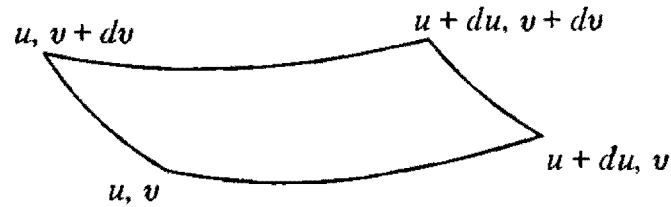


Рис. 1.7.3. Бесконечно малый элемент поверхности

координатных u -линий и v -линий. Рассматриваемая часть поверхности разобьется на конечное число так называемых криволинейных четырехугольников. Один из таких четырехугольников показан на рис. 1.7.3.

Пусть стороны этого четырехугольника равны du и 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:

$$\begin{aligned} |\mathbf{r}_1 \times \mathbf{r}_2|^2 &= |\mathbf{r}_1|^2 |\mathbf{r}_2|^2 \sin^2 \varphi = \\ &= |\mathbf{r}_1|^2 |\mathbf{r}_2|^2 (1 - \cos^2 \varphi) = g_{11}g_{22} \left(1 - \frac{g_{12}^2}{g_{11}g_{22}}\right) = g_{11}g_{22} - g_{12}^2. \end{aligned} \quad (1.7.14)$$

Таким образом, площадь четырехугольника в первом приближении определится формулой

$$dS = \sqrt{g_{11}g_{22} - g_{12}^2} du dv. \quad (1.7.15)$$

Будем теперь измельчать разбиение, увеличивая число координатных линий, таким образом, чтобы наибольшие из значений du и 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

$$S = \iint_{\Omega} \sqrt{g_{11}g_{22} - g_{12}^2} du dv. \quad (1.7.16)$$

Заметим, что так как первая квадратичная форма в (1.7.9) определяет квадрат длины дуги, который в обыкновенной точке больше нуля при всех du и dv , одновременно не равных нулю, то определитель матрицы \mathbf{G} положителен:

$$g = |\mathbf{G}| = g_{11}g_{22} - g_{12}^2 > 0. \quad (1.7.17)$$

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 (1.7.14), получим формулу для единичной нормали к поверхности в рассматриваемой точке

$$\mathbf{m} = \frac{\mathbf{r}_1 \times \mathbf{r}_2}{\sqrt{g_{11}g_{22} - g_{12}^2}}. \quad (1.7.18)$$

Продолжим исследование поверхности вблизи некоторой ее точки R путем исследования кривых на ней. Рассмотрим некоторую кривую на поверхности $u = u(t)$, $v = v(t)$. Вычислим приращение радиус-вектора кривой, которое он получит при бесконечно малом приращении ее параметра dt . Для этого используем разложение радиус-вектора в ряд Тейлора (1.5.5). Приращение радиус-вектора кривой $\Delta \mathbf{r}$ с точностью до второго порядка малости относительно dt равно

$$\Delta \mathbf{r} = \frac{d\mathbf{r}}{dt} dt + \frac{1}{2} \frac{d^2\mathbf{r}}{dt^2} dt^2. \quad (1.7.19)$$

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

$$\begin{aligned} \mathbf{r}' &= \mathbf{r}_1 u' + \mathbf{r}_2 v', \\ \mathbf{r}'' &= \mathbf{r}_{11} u'^2 + 2\mathbf{r}_{12} u'v' + \mathbf{r}_{22} v'^2 + \mathbf{r}_1 u'' + \mathbf{r}_2 v''. \end{aligned} \quad (1.7.20)$$

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 (1.7.19) на нормаль к поверхности, получим

$$\Delta \mathbf{r} \cdot \mathbf{m} = \frac{1}{2} \mathbf{r}'' \cdot \mathbf{m} dt^2 = \frac{1}{2} (\mathbf{r}_{11} \cdot \mathbf{m} u'^2 + 2\mathbf{r}_{12} \cdot \mathbf{m} u'v' + \mathbf{r}_{22} \cdot \mathbf{m} v'^2 + \mathbf{r}_1 \cdot \mathbf{m} u'' + \mathbf{r}_2 \cdot \mathbf{m} v'') dt^2.$$

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

$$\begin{aligned} b_{11} &= \mathbf{r}_{11} \cdot \mathbf{m}, \\ b_{12} &= b_{21} = \mathbf{r}_{12} \cdot \mathbf{m} = \mathbf{r}_{21} \cdot \mathbf{m}, \\ b_{22} &= \mathbf{r}_{22} \cdot \mathbf{m}. \end{aligned} \quad (1.7.21)$$

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

$$\Delta \mathbf{r} \cdot \mathbf{m} = \frac{1}{2} (b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2). \quad (1.7.22)$$

В (1.7.22) использовались обозначения (1.7.21). Выражение в скобках правой части (1.7.22) является квадратичной формой от дифференциалов du и 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 (1.7.22) можно записать в матричном виде

$$2\Delta\mathbf{r} \cdot \mathbf{m} = [du \ dv] \cdot \mathbf{B} \cdot [du \ dv]^T, \quad (1.7.23)$$

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 $\mathbf{m} \cdot \mathbf{r}_1 = 0$ и $\mathbf{m} \cdot \mathbf{r}_2 = 0$ 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 и \mathbf{m}_2 — частные производные вектора нормали по параметрам u и v . Подставим в эти равенства обозначения (1.7.21):

$$\begin{aligned} b_{11} &= -\mathbf{m}_1 \cdot \mathbf{r}_1, \\ b_{12} &= -\mathbf{m}_2 \cdot \mathbf{r}_1, \\ b_{21} &= -\mathbf{m}_1 \cdot \mathbf{r}_2, \\ b_{22} &= -\mathbf{m}_2 \cdot \mathbf{r}_2. \end{aligned} \quad (1.7.24)$$

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

$$\begin{aligned} d\mathbf{r} \cdot d\mathbf{m} &= \mathbf{m}_1 \cdot \mathbf{r}_1 du^2 + (\mathbf{m}_2 \cdot \mathbf{r}_1 + \mathbf{m}_1 \cdot \mathbf{r}_2) du dv + \mathbf{m}_2 \cdot \mathbf{r}_2 dv^2 = \\ &= -(b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2). \end{aligned} \quad (1.7.25)$$

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

$$\begin{aligned} \mathbf{m}_1 &= -b_1^1 \mathbf{r}_1 - b_1^2 \mathbf{r}_2, \\ \mathbf{m}_2 &= -b_2^1 \mathbf{r}_1 - b_2^2 \mathbf{r}_2, \end{aligned} \quad (1.7.26)$$

где $b_1^1, b_1^2, b_2^1, b_2^2$ — пока неизвестные коэффициенты разложения. Знак минус выбран из соображений удобства в дальнейшем. Для определения этих коэффициентов умножим каждое из равенств (1.7.26) скалярно на \mathbf{r}_1 и \mathbf{r}_2 . Используя (1.7.8), получим две системы уравнений

$$\begin{aligned} b_{11} &= b_1^1 g_{11} + b_1^2 g_{12}, \\ b_{12} &= b_1^1 g_{21} + b_1^2 g_{22}, \\ b_{21} &= b_2^1 g_{11} + b_2^2 g_{12}, \\ b_{22} &= b_2^1 g_{12} + b_2^2 g_{22}. \end{aligned} \quad (1.7.27)$$

Решая эти уравнения относительно искомых коэффициентов, получим

$$\begin{aligned} b_1^1 &= \frac{b_{11}g_{22} - b_{12}g_{12}}{g}, & b_1^2 &= \frac{b_{12}g_{11} - b_{11}g_{21}}{g}, \\ b_2^1 &= \frac{b_{21}g_{22} - b_{22}g_{12}}{g}, & b_2^2 &= \frac{b_{22}g_{11} - b_{21}g_{21}}{g}, \end{aligned} \quad (1.7.28)$$

где $g = g_{11}g_{22} - g_{12}g_{21} = \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} = |\mathbf{G}|$ — определитель матрицы \mathbf{G} . Формулы (1.7.26) с учетом (1.7.28) 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

$$\begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} = - \begin{bmatrix} b_1^1 & b_1^2 \\ b_2^1 & b_2^2 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} = -\mathbf{B}' \cdot \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix}, \quad (1.7.29)$$

где \mathbf{B}' — матрица коэффициентов (1.7.28).

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

$$\mathbf{G}^{-1} = \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix}, \quad (1.7.30)$$

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

$$\begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} = -\mathbf{B} \cdot \mathbf{G}^{-1} \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix}. \quad (1.7.31)$$

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

$$\mathbf{r}'' \cdot \mathbf{m} = \mathbf{r}_{11} \cdot \mathbf{m}u'^2 + 2\mathbf{r}_{12} \cdot \mathbf{m}u'v' + \mathbf{r}_{22} \cdot \mathbf{m}v'^2 = b_{11}u'^2 + 2b_{12}u'v' + b_{22}v'^2. \quad (1.8.1)$$

Вторая производная радиус-вектора кривой на поверхности согласно (1.5.22) равна

$$\mathbf{r}'' = \frac{d^2 \mathbf{r}}{dt^2} = \frac{d(s' \mathbf{t})}{ds} = s'' \mathbf{t} + s'^2 k \mathbf{n}, \quad (1.8.2)$$

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 (1.8.1) перепишем в следующем виде

$$ds^2 k \mathbf{n} \cdot \mathbf{m} = b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2. \quad (1.8.3)$$

Квадрат дифференциала длины дуги кривой определяется равенством (1.7.9). 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:

$$k \cos \omega = \frac{b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2}{g_{11} du^2 + 2g_{12} du dv + g_{22} dv^2}. \quad (1.8.4)$$

Для уяснения геометрического смысла последнего соотношения рассмотрим рис. 1.8.1, 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 $|\cos \omega| = 1$. Коэффициенты первой и второй квадратичных форм зависят только от положения точки M . А вот дифференциалы du и 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

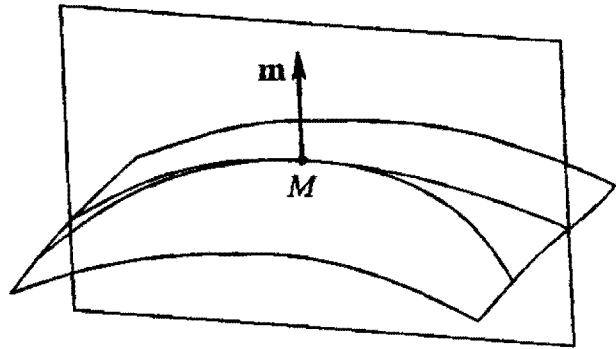


Рис. 1.8.1. Нормальное сечение поверхности

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

$$\mu = \frac{b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2}{g_{11} du^2 + 2g_{12} du dv + g_{22} dv^2}. \quad (1.8.5)$$

Угол ω есть 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:

$$\frac{1}{k} = \frac{1}{\mu} \cos \omega = \frac{1}{\mu} \mathbf{n} \cdot \mathbf{m}. \quad (1.8.6)$$

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 $k \sin \omega$ 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 случай, 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 (1.8.5) перепишем в другом виде (используя равенство (1.7.25))

$$\mu(d\mathbf{r} \cdot d\mathbf{r}) = -(d\mathbf{r} \cdot d\mathbf{m}), \quad (1.8.7)$$

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

$$dm = -\lambda dr$$

или

$$\begin{aligned} r_1 du + r_2 dv &= -\frac{1}{\lambda}(m_1 du + m_2 dv) = \frac{1}{\lambda}(b_1^1 r_1 + b_1^2 r_2) du + \frac{1}{\lambda}(b_2^1 r_1 + b_2^2 r_2) dv = \\ &= \frac{1}{\lambda}(b_1^1 du + b_2^1 dv)r_1 + \frac{1}{\lambda}(b_1^2 du + b_2^2 dv)r_2, \end{aligned} \quad (1.8.8)$$

где λ — неизвестный пока коэффициент. Мы можем считать r_1 и r_2 базисными векторами, по которым разложены векторы dr и dm . Для их коллинеарности нужно, чтобы коэффициенты при r_1 и r_2 в правой и левой частях (1.8.8) 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 $\mathbf{B}' = \mathbf{B} \cdot \mathbf{G}^{-1}$. Итак, для определения искомого направления мы пришли к системе линейных алгебраических уравнений для du и dv

$$(\mathbf{B} - \lambda \mathbf{G}) \begin{bmatrix} du \\ dv \end{bmatrix} = 0. \quad (1.8.9)$$

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 и λ_2 . Подставив каждый из корней в любое из уравнений (1.8.9), получим два направления на поверхности, определяемые отношениями $du_1:dv_1$ и $du_2:dv_2$.

Направления движения на поверхности, для которых векторы dr и dm коллинеарны, называются *главными направлениями поверхности*. Сравним соотношения (1.8.7) и (1.8.8) и увидим, что λ_1 и λ_2 равны кривизне нормальных сечений в главных направлениях, которые обозначим через μ_1 и 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

$$(g_{11}g_{22} - g_{12}^2)\lambda^2 - (b_{11}g_{22} + b_{22}g_{11} - 2b_{12}g_{12})\lambda + (b_{11}b_{22} - b_{12}^2) = 0. \quad (1.8.10)$$

Из (1.8.10) легко получить сумму и произведение корней уравнения, т. е. сумму и произведение главных кривизн:

$$\mu_1 + \mu_2 = \frac{b_{11}g_{22} + b_{22}g_{11} - 2b_{12}g_{12}}{g}, \quad (1.8.11)$$

$$\mu_1\mu_2 = \frac{b_{11}b_{22} - b_{12}b_{21}}{g}. \quad (1.8.12)$$

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

$$\mathbf{t}_1 = \mathbf{r}_1 du_1 + \mathbf{r}_2 dv_1,$$

$$\mathbf{t}_2 = \mathbf{r}_1 du_2 + \mathbf{r}_2 dv_2.$$

Их скалярное произведение равно

$$\mathbf{t}_1 \cdot \mathbf{t}_2 = [du_1 \ dv_1] \cdot \mathbf{G} \cdot [du_2 \ dv_2]^T. \quad (1.8.13)$$

Покажем, что оно в общем случае равно нулю. Для этого систему двух уравнений (1.8.9) 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:

$$[du_1 \ dv_1] \cdot ((\mu_2 - \mu_1) \mathbf{G}) \cdot [du_2 \ dv_2]^T = 0, \quad (1.8.14)$$

из которого следует, что если главные кривизны различны, то выражение (1.8.13) 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 и \mathbf{t}_2 главных направлений:

$$\frac{d\mathbf{r}}{ds} = \mathbf{t}_1 \cos \varphi + \mathbf{t}_2 \sin \varphi,$$

$$\frac{d\mathbf{m}}{ds} = -\mu_1 \mathbf{t}_1 \cos \varphi - \mu_2 \mathbf{t}_2 \sin \varphi,$$

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 (1.7.25) и (1.8.4) определится равенством

$$\mu = -\frac{d\mathbf{r}}{ds} \cdot \frac{d\mathbf{m}}{ds} = \mu_1 \cos^2 \varphi + \mu_2 \sin^2 \varphi. \quad (1.8.15)$$

Формула (1.8.15) 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 и 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 (1.8.12) может быть использована для определения поведения поверхности в некоторой ее точке M . Так как знаменатель в (1.8.12) 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. 1.8.2.

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

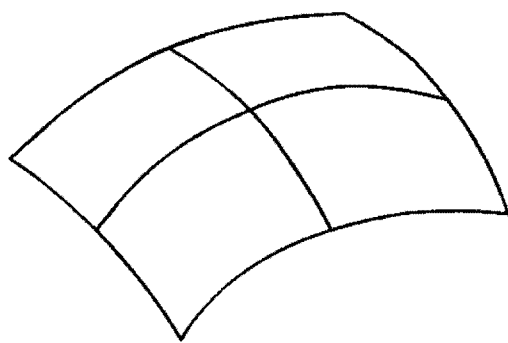


Рис. 1.8.2. Эллиптическая точка поверхности

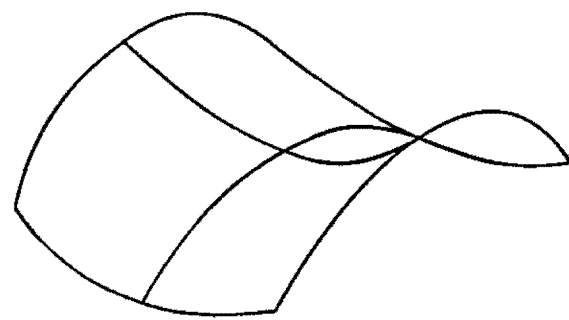


Рис. 1.8.3. Гиперболическая точка поверхности

кривизн. Если $|\mathbf{B}| < 0$, то точка M называется гиперболической. Поведение поверхности в гиперболической точке показано на рис. 1.8.3. Так как в такой точке главные кривизны имеют разные знаки, то согласно (1.8.15) существуют такие нормальные сечения, для которых выполняется равенство

$$\mu_1 \cos^2 \varphi + \mu_2 \sin^2 \varphi = 0. \quad (1.8.16)$$

Касательные к нормальным сечениям под углами

$$\varphi = \pm \operatorname{arctg} \left(\sqrt{-\frac{\mu_1}{\mu_2}} \right) \quad (1.8.17)$$

расположены в касательной плоскости симметрично относительно главных направлений и определяют асимптотические направления в точке M . Если в точке M $|\mathbf{B}| = 0$, то такая точка называется параболической. Поведение поверхности в параболической точке показано на рис. 1.8.4.

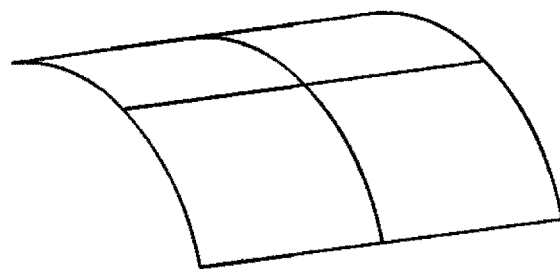


Рис. 1.8.3. Параболическая точка поверхности

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 M 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

$$g_{12} = g_{21} = b_{12} = b_{21} = 0 \quad (1.8.18)$$

в силу ортогональности главных направлений. Справедливо и обратное утверждение: если в каждой точке поверхности выполняются равенства (1.8.18), то координатные линии являются линиями кривизны. Действительно, в этом случае согласно (1.7.28) коэффициенты b_1^2 и b_2^1 в разложении (1.7.26) 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 и v . Модуль дифференциала нормали к поверхности равен углу между нормальными в двух бесконечно близких точках, связанных параметрическим смещением du и dv . Квадрат этого угла определяется равенством

$$d\vartheta^2 = d\mathbf{m} \cdot d\mathbf{m} = \mathbf{m}_1 \cdot \mathbf{m}_1 du^2 + 2\mathbf{m}_2 \cdot \mathbf{m}_1 du dv + \mathbf{m}_2 \cdot \mathbf{m}_2 dv^2. \quad (1.8.19)$$

Введем обозначения

$$\begin{aligned} f_{11} &= \mathbf{m}_1 \cdot \mathbf{m}_1, \\ f_{12} &= f_{21} = \mathbf{m}_1 \cdot \mathbf{m}_2 = \mathbf{m}_2 \cdot \mathbf{m}_1, \\ f_{22} &= \mathbf{m}_2 \cdot \mathbf{m}_2. \end{aligned} \quad (1.8.20)$$

Равенство (1.8.19) примет вид

$$d\mathbf{m} \cdot d\mathbf{m} = f_{11} du^2 + 2f_{12} du dv + f_{22} dv^2. \quad (1.8.21)$$

В правой части (1.8.21) мы получили квадратичную форму от du и 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 (1.8.21) можно записать в матричном виде

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

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

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

$$\begin{aligned} \mathbf{m} \cdot \mathbf{m}_{11} &= -\mathbf{m}_1 \cdot \mathbf{m}_1 = -f_{11}, \\ \mathbf{m} \cdot \mathbf{m}_{12} &= -\mathbf{m}_1 \cdot \mathbf{m}_2 = -f_{12}, \\ \mathbf{m} \cdot \mathbf{m}_{21} &= -\mathbf{m}_2 \cdot \mathbf{m}_1 = -f_{21}, \\ \mathbf{m} \cdot \mathbf{m}_{22} &= -\mathbf{m}_2 \cdot \mathbf{m}_2 = -f_{22}. \end{aligned} \quad (1.8.22)$$

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 и \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

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

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

$$[du \ dv] \cdot (\mathbf{F} + (\mu_1 + \mu_2)\mathbf{B} + \mu_1 \mu_2 \mathbf{G}) \cdot [du \ dv]^T = 0. \quad (1.8.23)$$

Так как du и dv в общем случае не равны нулю, то для выполнения соотношения (1.8.23) должно выполняться матричное равенство

$$\mathbf{F} + (\mu_1 + \mu_2)\mathbf{B} + \mu_1 \mu_2 \mathbf{G} = 0. \quad (1.8.24)$$

Это и есть уравнение, связывающее коэффициенты трех квадратичных форм поверхности. Из (1.8.24) следует, что коэффициенты третьей квадратичной формы поверхности выражаются через коэффициенты первой и второй квадратичных форм поверхности соотношением

$$\mathbf{F} = -(\mu_1 + \mu_2)\mathbf{B} - \mu_1 \mu_2 \mathbf{G}. \quad (1.8.25)$$

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 (1.7.9), (1.7.22), (1.8.21); верхние индексы использовались для обозначения компонент матрицы \mathbf{G}^{-1} в (1.7.30); смешанные индексы использовались в формулах (1.7.28).

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 *немыми*, 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 снизу 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 $u^{2'}$ производные радиус-вектора поверхности по новым параметрам будут связаны с производными радиус-вектора поверхности по старым параметрам соотношениями

$$\mathbf{r}_{k'} = \sum_{i=1}^2 \frac{\partial u^i}{\partial u^{k'}} \mathbf{r}_i, \quad k = 1, 2, \quad (1.9.1)$$

тогда как бесконечно малые приращения параметров будут связаны соотношениями

$$du^{k'} = \sum_{i=1}^2 \frac{\partial u^{k'}}{\partial u^i} du^i, \quad k = 1, 2. \quad (1.9.2)$$

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

$$A^i_{k'} = \frac{\partial u^i}{\partial u^{k'}}, \quad (1.9.3)$$

$$A^{k'}_i = \frac{\partial u^{k'}}{\partial u^i} \quad (1.9.4)$$

и запишем равенства (1.9.1) и (1.9.2), опустив знак суммы и используя вместо него соглашение о суммировании

$$\mathbf{r}_{k'} = A^i_{k'} \mathbf{r}_i, \quad (1.9.5)$$

$$du^{k'} = A^{k'}_i du^i, \quad (1.9.6)$$

где k' и i принимают значения 1 и 2.

Якобиан матриц преобразования (1.9.3) и (1.9.4) должен быть отличен от нуля. Матрицы с компонентами (1.9.3) и (1.9.4) описывают однородные линейные преобразования. При переходе к другой параметризации поверхности компоненты матриц G ,

B и G^{-1} изменяются по формулам

$$g_{k'n'} = A^i_{k'} A^j_{n'} g_{ij}, \quad (1.9.7)$$

$$b_{k'n'} = A^i_{k'} A^j_{n'} b_{ij}, \quad (1.9.8)$$

$$g^{k'n'} = A^{k'}_i A^{n'}_j g^{ij}. \quad (1.9.9)$$

Тензорами называются объекты, компоненты которых при изменении системы координат подвергаются линейным однородным преобразованиям, причем матрица преобразования для каждого нижнего индекса составлена из производных (1.9.3), а матрица преобразования для каждого верхнего индекса составлена из производных (1.9.4). 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 (1.7.10) или своими *контравариантными* компонентами в виде матрицы G^{-1} (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 (1.7.28) и (1.7.27), связывающим коэффициенты b_i^k и b_{ij} :

$$b_i^k = b_{ij} g^{jk}, \quad (1.9.10)$$

$$b_{ij} = b_i^k g_{jk}. \quad (1.9.11)$$

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^{jk} . Во втором равенстве мы получили ковариантные компоненты b_{ij} объекта через его смешанные компоненты b_i^k , а суммирование выполняется по индексу k .

Верхние и нижние индексы являются равноправными. Векторы r_1 , r_2 и m образуют местный базис в заданной точке поверхности, по которому можно разложить любой другой вектор в этой точке. Базис r_1 , r_2 , m называется *касательным*, так как его

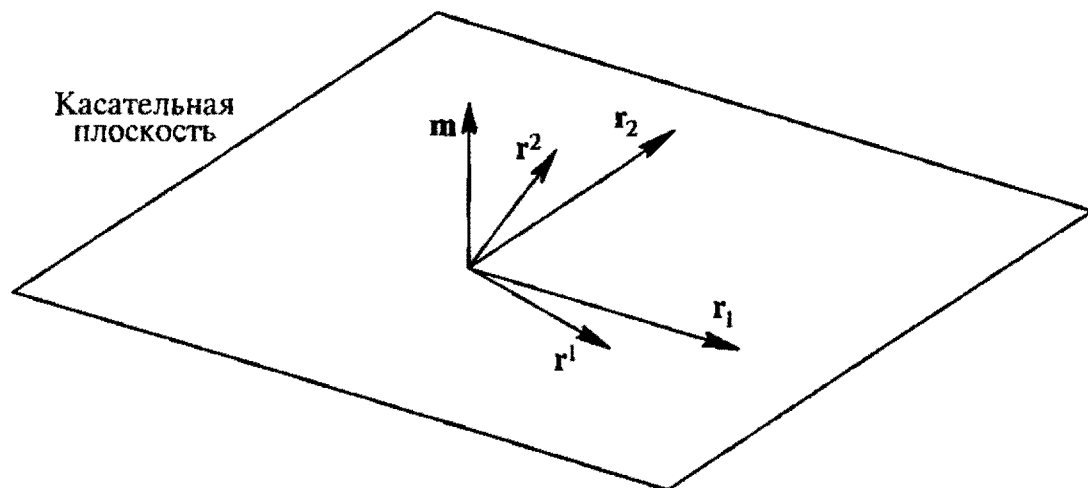


Рис. 1.9.1. Касательный и взаимный базисы поверхности в точке

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 r^1 , r^2 и 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 и \mathbf{r}^2 лежат в касательной плоскости, их ориентация показана на рис. 1.9.1.

Векторы \mathbf{r}^1 и \mathbf{r}^2 определим равенствами

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

Эти равенства можно получить другим способом: разложить векторы \mathbf{r}^1 и \mathbf{r}^2 по векторам \mathbf{r}_1 и \mathbf{r}_2 аналогично тому, как были разложены производные нормали \mathbf{m}_1 и \mathbf{m}_2 в (1.7.26), 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

$$g^{ki} = \mathbf{r}^k \cdot \mathbf{r}^i. \quad (1.9.13)$$

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 и \mathbf{m} . Для этого запишем их в виде

$$\mathbf{r}_{ij} = \Gamma_{ij}^k \mathbf{r}_k + b_{ij} \mathbf{m}, \quad (1.9.14)$$

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 (1.9.14) на \mathbf{m} мы должны получить равенства (1.7.21). Умножим (1.9.14) скалярно на \mathbf{r}_1 и \mathbf{r}_2 и получим систему уравнений для Γ_{ij}^k

$$\begin{aligned} \Gamma_{ij}^1 g_{11} + \Gamma_{ij}^2 g_{12} &= \Gamma_{ij,1}, \\ \Gamma_{ij}^1 g_{21} + \Gamma_{ij}^2 g_{22} &= \Gamma_{ij,2}, \end{aligned} \quad (1.9.15)$$

где введено обозначение $\Gamma_{ij,1} = \mathbf{r}_{ij} \cdot \mathbf{r}_1$ и $\Gamma_{ij,2} = \mathbf{r}_{ij} \cdot \mathbf{r}_2$. 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 $\Gamma_{ij,k}$

$$\Gamma_{ij,k} = \frac{1}{2} \left(\frac{\partial g_{kj}}{\partial u^i} + \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k} \right). \quad (1.9.16)$$

Теперь решим систему уравнений (1.9.15) относительно Γ_{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 (1.9.15) примут вид

$$\Gamma_{ij}^n = g^{nk} \Gamma_{ij,k}, \quad (1.9.17)$$

$$\Gamma_{ij,n} = g_{kn} \Gamma_{ij}^k. \quad (1.9.18)$$

Подставим в последние равенства значения (1.9.16) и получим окончательное выражение для коэффициентов в разложении (1.9.14)

$$\Gamma_{ij}^n = \frac{1}{2} g^{nk} \left(\frac{\partial g_{kj}}{\partial u^i} + \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k} \right). \quad (1.9.19)$$

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

В (1.9.14) мы выразили производные векторов \mathbf{r}_1 и \mathbf{r}_2 по параметрам. Формулы (1.9.14) называются *деривационными формулами Гаусса*. Совместно с деривационными формулами Вейнгартена (1.7.26) они представляют собой деривационные формулы для локального базиса $\mathbf{r}_1, \mathbf{r}_2, \mathbf{m}$ на поверхности. Перепишем их в виде

$$\begin{aligned} \frac{\partial \mathbf{r}_1}{\partial u^i} &= \Gamma_{1i}^1 \mathbf{r}_1 + \Gamma_{1i}^2 \mathbf{r}_2 + b_{1i} \mathbf{m}, \\ \frac{\partial \mathbf{r}_2}{\partial u^i} &= \Gamma_{2i}^1 \mathbf{r}_1 + \Gamma_{2i}^2 \mathbf{r}_2 + b_{2i} \mathbf{m}, \\ \frac{\partial \mathbf{m}}{\partial u^i} &= -b_i^1 \mathbf{r}_1 - b_i^2 \mathbf{r}_2. \end{aligned} \quad (1.9.20)$$

Эти формулы аналогичны формулам Френе–Серре (1.5.17) 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 (1.8.9) еще раз и пользуясь имеющимися разложе-

ниями первых производных, можно получить вторые производные локального базиса, а за ними — третьи производные и так далее. Например, вторые производные вектора нормали равны

$$\begin{aligned} m_{ij} &= -\frac{\partial b_i^k}{\partial u^j} \mathbf{r}_k - b_i^k \mathbf{r}_{kj} = -\frac{\partial b_i^k}{\partial u^j} \mathbf{r}_k - b_i^k (\Gamma_{kj}^n \mathbf{r}_n + b_{kj} \mathbf{m}) = \\ &= -\left(\frac{\partial b_i^k}{\partial u^j} - b_i^n \Gamma_{nj}^k \right) \mathbf{r}_k - b_i^k b_{kj} \mathbf{m}. \end{aligned} \quad (1.9.21)$$

Найдем разложение производных векторов \mathbf{r}^1 и \mathbf{r}^2 по взаимному базису поверхности в этой же точке, представленному векторами \mathbf{r}^1 , \mathbf{r}^2 и \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 дериационные формулы (1.9.20) и равенство

$$\frac{1}{2g} \frac{\partial g}{\partial u^i} = \Gamma_{1i}^1 + \Gamma_{2i}^2 = \Gamma_{ki}^k, \quad (1.9.24)$$

Равенство (1.9.24) следует из соотношения

$$\begin{aligned} \frac{\partial g}{\partial u^i} &= \frac{\partial (g_{11}g_{22} - g_{12}g_{21})}{\partial u^i} = 2\mathbf{r}_{1i} \cdot (g_{22}\mathbf{r}_1 - g_{12}\mathbf{r}_2) + 2\mathbf{r}_{2i} \cdot (g_{11}\mathbf{r}_2 - g_{21}\mathbf{r}_1) = \\ &= 2g\mathbf{r}_{1i} \cdot (g^{11}\mathbf{r}_1 + g^{12}\mathbf{r}_2) + 2g\mathbf{r}_{2i} \cdot (g^{22}\mathbf{r}_2 + g^{21}\mathbf{r}_1) = 2g\mathbf{r}_{ji} \cdot \mathbf{r}_k g^{kj} = 2g\Gamma_{ki}^k. \end{aligned} \quad (1.9.25)$$

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

$$\begin{aligned} \frac{\partial \mathbf{r}^1}{\partial u^i} &= -\Gamma_{1i}^1 \mathbf{r}^1 - \Gamma_{2i}^1 \mathbf{r}^2 + b_i^1 \mathbf{m}, \\ \frac{\partial \mathbf{r}^2}{\partial u^i} &= -\Gamma_{1i}^2 \mathbf{r}^1 - \Gamma_{2i}^2 \mathbf{r}^2 + b_i^2 \mathbf{m}, \\ \frac{\partial \mathbf{m}}{\partial u^i} &= -b_{i1} \mathbf{r}^1 - b_{i2} \mathbf{r}^2. \end{aligned} \quad (1.9.26)$$

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

они связаны между собой уравнениями, которые мы получим ниже. Дифференцируя равенство (1.9.14) по одному из параметров, получим вторую производную вектора \mathbf{r}_i

$$\begin{aligned} \frac{\partial^2 \mathbf{r}_i}{\partial u^n \partial u^j} &= \frac{\partial \Gamma_{ij}^k}{\partial u^n} \mathbf{r}_k + \Gamma_{ij}^k \mathbf{r}_{kn} + \frac{\partial b_{ij}}{\partial u^n} \mathbf{m} + b_{ij} \mathbf{m}_n = \\ &= \left(\frac{\partial \Gamma_{ij}^m}{\partial u^n} + \Gamma_{ij}^k \Gamma_{kn}^m - b_{ij} b_n^m \right) \mathbf{r}_m + \left(\frac{\partial b_{ij}}{\partial u^n} + \Gamma_{ij}^k b_{kn} \right) \mathbf{m}. \end{aligned} \quad (1.9.27)$$

Изменим последовательность дифференцирования вектора \mathbf{r}_i в (1.9.27), что приведет к смене мест индексов n и j :

$$\begin{aligned} \frac{\partial^2 \mathbf{r}_i}{\partial u^j \partial u^n} &= \frac{\partial \Gamma_{in}^k}{\partial u^j} \mathbf{r}_k + \Gamma_{in}^k \mathbf{r}_{kj} + \frac{\partial b_{in}}{\partial u^j} \mathbf{m} + b_{in} \mathbf{m}_j = \\ &= \left(\frac{\partial \Gamma_{in}^m}{\partial u^j} + \Gamma_{in}^k \Gamma_{kj}^m - b_{in} b_j^m \right) \mathbf{r}_m + \left(\frac{\partial b_{in}}{\partial u^j} + \Gamma_{in}^k b_{kj} \right) \mathbf{m}. \end{aligned} \quad (1.9.28)$$

Левые части равенств (1.9.27) и (1.9.28) равны, следовательно, должны быть равны и правые их части. Вычитая (1.9.28) из (1.9.27), получим равенство

$$\begin{aligned} \left(\frac{\partial \Gamma_{ij}^m}{\partial u^n} - \frac{\partial \Gamma_{in}^m}{\partial u^j} + \Gamma_{ij}^k \Gamma_{kn}^m - \Gamma_{in}^k \Gamma_{kj}^m + b_{in} b_j^m - b_{ij} b_n^m \right) \mathbf{r}_m + \\ + \left(\frac{\partial b_{ij}}{\partial u^n} - \frac{\partial b_{in}}{\partial u^j} + \Gamma_{ij}^k b_{kn} - \Gamma_{in}^k b_{kj} \right) \mathbf{m} = 0. \end{aligned} \quad (1.9.29)$$

Так как в обыкновенной точке векторы \mathbf{r}_1 , \mathbf{r}_2 , \mathbf{m} некопланарны, то для выполнения векторного равенства (1.9.29) должны выполняться следующие соотношения

$$\frac{\partial \Gamma_{ij}^m}{\partial u^n} - \frac{\partial \Gamma_{in}^m}{\partial u^j} + \Gamma_{ij}^k \Gamma_{kn}^m - \Gamma_{in}^k \Gamma_{kj}^m = b_{ij} b_n^m - b_{in} b_j^m, \quad (1.9.30)$$

$$\frac{\partial b_{ij}}{\partial u^n} - \Gamma_{in}^k b_{kj} = \frac{\partial b_{in}}{\partial u^j} - \Gamma_{ij}^k b_{kn}. \quad (1.9.31)$$

Эти уравнения связывают между собой коэффициенты первой и второй основных квадратичных форм поверхности. Уравнение (1.9.30) называется *уравнением Гаусса*. Уравнения (1.9.31) называются *уравнениями Петерсона-Кодацци*. Правая часть равенства (1.9.30) является тензором, следовательно, и его левая часть также представляет собой тензор. Этот тензор называется *тензором кривизны* или *тензором Римана* и обозначается

$$R_{jni}^m = \frac{\partial \Gamma_{ij}^m}{\partial u^n} - \frac{\partial \Gamma_{in}^m}{\partial u^j} + \Gamma_{kn}^m \Gamma_{ij}^k - \Gamma_{kj}^m \Gamma_{in}^k. \quad (1.9.32)$$

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

$$\frac{\partial \Gamma_{ij,l}}{\partial u^n} - \frac{\partial \Gamma_{in,l}}{\partial u^j} + \Gamma_{kn,l} \Gamma_{ij}^k - \Gamma_{kj,l} \Gamma_{in}^k = b_{ij} b_{nl} - b_{in} b_{jl}. \quad (1.9.33)$$

Аналогично получим ковариантные компоненты тензора кривизны

$$R_{jnli} = \frac{\partial \Gamma_{ij,l}}{\partial u^n} - \frac{\partial \Gamma_{in,l}}{\partial u^j} + \Gamma_{kn,l} \Gamma_{ij}^k - \Gamma_{kj,l} \Gamma_{in}^k. \quad (1.9.34)$$

В силу симметрии символов Кристоффеля по первым двум индексам тензор кривизны (1.9.34) обладает определенной симметрией, он кососимметричен по первой паре индексов и по второй паре индексов: $R_{jnli} = R_{iljn} = -R_{njil} = -R_{jnli}$. Для поверхностей равенство (1.9.34) эквивалентно одному равенству

$$\frac{\partial \Gamma_{11,2}}{\partial u^2} - \frac{\partial \Gamma_{12,2}}{\partial u^1} + \Gamma_{k2,2} \Gamma_{11}^k - \Gamma_{k1,2} \Gamma_{12}^k = b_{11} b_{22} - b_{12} b_{21} = \mu_1 \mu_2 g. \quad (1.9.35)$$

Мы использовали равенство (1.8.12) для гауссовой кривизны поверхности. Таким образом, гауссова кривизна поверхности может быть определена через коэффициенты первой квадратичной формы и их производные. Уравнения Петерсона–Кодацци (1.9.31) сводятся к двум уравнениям

$$\frac{\partial b_{11}}{\partial u^2} - \Gamma_{12}^k b_{k1} = \frac{\partial b_{12}}{\partial u^1} - \Gamma_{11}^k b_{k2}, \quad (1.9.36)$$

$$\frac{\partial b_{22}}{\partial u^1} - \Gamma_{21}^k b_{k2} = \frac{\partial b_{21}}{\partial u^2} - \Gamma_{22}^k b_{k1}. \quad (1.9.37)$$

Если над тензором кривизны (1.9.32) выполнить операцию свертки по индексам n и m , то мы получим ковариантные компоненты тензора Риччи

$$R_{ji} = R_{j1i}^1 + R_{j2i}^2 = R_{jni}^n = b_{ij} b_n^n - b_{in} b_j^n = (b_{ij} b_{nk} - b_{in} b_{jk}) g^{kn}. \quad (1.9.38)$$

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 kn на касательную плоскость к поверхности должна быть равна нулю. Найдем вектор кривизны произвольной кривой на поверхности

$$\begin{aligned} kn &= \frac{d^2 \mathbf{r}}{ds^2} = \mathbf{r}_{11} \left(\frac{du^1}{ds} \right)^2 + 2\mathbf{r}_{12} \frac{du^1}{ds} \frac{du^2}{ds} + \mathbf{r}_{22} \left(\frac{du^2}{ds} \right)^2 + \mathbf{r}_1 \frac{d^2 u^1}{ds^2} + \mathbf{r}_2 \frac{d^2 u^2}{ds^2} \equiv \\ &\equiv \mathbf{r}_{ij} \frac{du^i}{ds} \frac{du^j}{ds} + \mathbf{r}_k \frac{d^2 u^k}{ds^2} = \mathbf{r}_k \Gamma_{ij}^k \frac{du^i}{ds} \frac{du^j}{ds} + \mathbf{r}_k \frac{d^2 u^k}{ds^2} + m b_{ij} \frac{du^i}{ds} \frac{du^j}{ds}. \end{aligned} \quad (1.9.39)$$

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

$$\frac{d^2 u^k}{ds^2} + \Gamma_{ij}^k \frac{du^i}{ds} \frac{du^j}{ds} = 0. \quad (1.9.40)$$

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 и u^2 , принимающих значения 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:

$$\begin{aligned} \mathbf{r}_i &\equiv \frac{\partial \mathbf{r}}{\partial u^i}, & g_{ij} &= \mathbf{r}_i \cdot \mathbf{r}_j, & g &= |\mathbf{G}| = g_{11}g_{22} - g_{12}g_{21}, \\ \mathbf{m} &= \frac{\mathbf{r}_1 \times \mathbf{r}_2}{\sqrt{g}}, & \mathbf{r}^1 &= \frac{\mathbf{r}_2 \times \mathbf{m}}{\sqrt{g}}, & \mathbf{r}^2 &= \frac{\mathbf{m} \times \mathbf{r}_1}{\sqrt{g}}, \\ g^{ij} &= \mathbf{r}^i \cdot \mathbf{r}^j, & b_{ij} &= \mathbf{r}_{ij} \cdot \mathbf{m}, & b_i^j &= -\mathbf{m}_i \cdot \mathbf{r}^j. \end{aligned} \quad (1.9.41)$$

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

$$\begin{aligned} \mathbf{r}_{ik} &\equiv \frac{\partial^2 \mathbf{r}}{\partial u^i \partial u^k}, & \mathbf{r}_{ijk} &\equiv \frac{\partial^3 \mathbf{r}}{\partial u^i \partial u^j \partial u^k}, & \Gamma_{ij,k} &= \mathbf{r}_{ij} \cdot \mathbf{r}_k, & \Gamma_{ij}^k &= \mathbf{r}_{ij} \cdot \mathbf{r}^k, \\ \frac{\partial g_{ij}}{\partial u^k} &= \mathbf{r}_{ik} \cdot \mathbf{r}_j + \mathbf{r}_i \cdot \mathbf{r}_{jk}, & \frac{\partial b_{ij}}{\partial u^k} &= \mathbf{r}_{ijk} \cdot \mathbf{m} + \mathbf{r}_{ij} \cdot \mathbf{m}_k, \\ \frac{\partial \Gamma_{ij,k}}{\partial u^n} &= \Gamma_{ijn} \cdot \mathbf{r}_k + \mathbf{r}_{ij} \cdot \Gamma_{kn}, & \frac{\partial \Gamma_{ij}^k}{\partial u^n} &= \Gamma_{ijn} \cdot \mathbf{r}^k + \mathbf{r}_{ij} \cdot \Gamma_{kn}, \\ \mathbf{r}_{ik} &\equiv \frac{\partial \mathbf{r}_i}{\partial u^k} = \Gamma_{ik}^j \mathbf{r}_j + b_{ik} \mathbf{m}, & \mathbf{r}^i_k &\equiv \frac{\partial \mathbf{r}^i}{\partial u^k} = -\Gamma_{kj}^i \mathbf{r}^j + b_k^i \mathbf{m}, \\ \mathbf{m}_i &\equiv \frac{\partial \mathbf{m}}{\partial u^i} = -b_{ik} \mathbf{r}^k = -b_i^k \mathbf{r}_k, \\ \mathbf{m}_{ij} &\equiv \frac{\partial^2 \mathbf{m}}{\partial u^i \partial u^j} = -(\mathbf{r}_{ijk} \cdot \mathbf{m} + \mathbf{r}_{ik} \cdot \mathbf{m}_j + \mathbf{r}_{jk} \cdot \mathbf{m}_i) \mathbf{r}^k - b_{ik} b_j^k \mathbf{m}. \end{aligned} \quad (1.9.42)$$

В равенствах (1.9.41) и (1.9.42) верхние и нижние индексы принимают значения от 1 до 2. Вся необходимая информация о поверхности может быть получена на основе равенств (1.9.41) и (1.9.42) при известной векторной функции (1.7.1). 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 двумерной 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 двумерную 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

$$u^1 = u^1(x^1, x^2, x^3), \quad u^2 = u^2(x^1, x^2, x^3), \quad u^3 = u^3(x^1, x^2, x^3), \quad (1.10.1)$$

такие, что обратные им функции

$$x^1 = x^1(u^1, u^2, u^3), \quad x^2 = x^2(u^1, u^2, u^3), \quad x^3 = x^3(u^1, u^2, u^3) \quad (1.10.2)$$

также являются непрерывными дифференцируемыми и однозначными. Потребуем, чтобы определитель матрицы Якоби системы (1.10.2) был отличен от нуля:

$$\begin{vmatrix} \frac{\partial x^1}{\partial u^1} & \frac{\partial x^1}{\partial u^2} & \frac{\partial x^1}{\partial u^3} \\ \frac{\partial x^2}{\partial u^1} & \frac{\partial x^2}{\partial u^2} & \frac{\partial x^2}{\partial u^3} \\ \frac{\partial x^3}{\partial u^1} & \frac{\partial x^3}{\partial u^2} & \frac{\partial x^3}{\partial u^3} \end{vmatrix} \neq 0. \quad (1.10.3)$$

Если один из параметров в (1.10.2) зафиксировать ($u^i = \text{const}$), то мы получим некоторую поверхность в пространстве, которую назовем поверхностью i -й группы. Функции (1.10.2) 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 -й 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:

$$\begin{aligned} \mathbf{r}_1 &= \lim_{\delta u^1 \rightarrow 0} \frac{R(u^1 + \delta u^1, u^2, u^3) - R(u^1, u^2, u^3)}{\delta u^1}, \\ \mathbf{r}_2 &= \lim_{\delta u^2 \rightarrow 0} \frac{R(u^1, u^2 + \delta u^2, u^3) - R(u^1, u^2, u^3)}{\delta u^2}, \\ \mathbf{r}_3 &= \lim_{\delta u^3 \rightarrow 0} \frac{R(u^1, u^2, u^3 + \delta u^3) - R(u^1, u^2, u^3)}{\delta u^3}. \end{aligned} \quad (1.10.4)$$

В силу (1.10.3) 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

$$g_{ij} = g_{ij}(u^1, u^2, u^3) = \mathbf{r}_i \cdot \mathbf{r}_j, \quad i, j = 1, 2, 3 \quad (1.10.5)$$

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_{kn}' 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

$$\mathbf{r}_{k'} = A^i_{k'} \mathbf{r}_i. \quad (1.10.6)$$

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

$$\mathbf{r}_i \cdot \mathbf{r}^k = \delta_i^k, \quad (1.10.7)$$

где δ_i^k — символы Кронекера, они принимают значения $\delta_i^k = 1$ 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 \mathbf{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 g^{nj}' 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

$$\mathbf{r}^{i'} = A^{i'}_k \mathbf{r}^k. \quad (1.10.9)$$

Так как \mathbf{G}^{-1} является обратной матрицей \mathbf{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

$$\begin{aligned} \mathbf{r}^1 &= \frac{\mathbf{r}_2 \times \mathbf{r}_3}{\sqrt{g}}, & \mathbf{r}^2 &= \frac{\mathbf{r}_3 \times \mathbf{r}_1}{\sqrt{g}}, & \mathbf{r}^3 &= \frac{\mathbf{r}_1 \times \mathbf{r}_2}{\sqrt{g}}, \\ \mathbf{r}_1 &= \mathbf{r}^2 \times \mathbf{r}^3 \sqrt{g}, & \mathbf{r}_2 &= \mathbf{r}^3 \times \mathbf{r}^1 \sqrt{g}, & \mathbf{r}_3 &= \mathbf{r}^1 \times \mathbf{r}^2 \sqrt{g}, \end{aligned} \quad (1.10.10)$$

где g равен определителю матрицы \mathbf{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} и e_{ijk} . Величины e^{ijk} и 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

$$\mathbf{a} \cdot \mathbf{b} = a^i b^j g_{ij} = a_i b_j g^{ij} = a^i b_i = a_i b^i. \quad (1.10.11)$$

Их векторное произведение может быть представлено в виде

$$\mathbf{a} \times \mathbf{b} = a^i b^j e_{ijk} \sqrt{g} \mathbf{r}^k = a_i b_j e^{ijk} \frac{\mathbf{r}_k}{\sqrt{g}}, \quad (1.10.12)$$

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

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

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 $R_1(u^1 + \delta u^1, u^2 + \delta u^2, u^3 + \delta u^3)$. С точностью до линейных слагаемых относительно δu^i , $i = 1, 2, 3$ изменение базисных векторов можно представить в виде

$$\delta \mathbf{r}_i = \frac{\partial \mathbf{r}_i}{\partial u^j} \delta u^j = \mathbf{r}_{ij} \delta u^j = \Gamma_{ij}^k \mathbf{r}_k \delta u^j, \quad (1.10.13)$$

где Γ_{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

$$\mathbf{r}_{ij} \equiv \frac{\partial \mathbf{r}_i}{\partial u^j} = \Gamma_{ij}^k \mathbf{r}_k \quad (1.10.14)$$

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

$$\begin{aligned} \Gamma_{ij}^1 g_{11} + \Gamma_{ij}^2 g_{12} + \Gamma_{ij}^3 g_{13} &= \Gamma_{ij,1} \equiv \mathbf{r}_{ij} \cdot \mathbf{r}_1, \\ \Gamma_{ij}^1 g_{21} + \Gamma_{ij}^2 g_{22} + \Gamma_{ij}^3 g_{23} &= \Gamma_{ij,2} \equiv \mathbf{r}_{ij} \cdot \mathbf{r}_2, \\ \Gamma_{ij}^1 g_{31} + \Gamma_{ij}^2 g_{32} + \Gamma_{ij}^3 g_{33} &= \Gamma_{ij,3} \equiv \mathbf{r}_{ij} \cdot \mathbf{r}_3. \end{aligned} \quad (1.10.15)$$

Коэффициенты $\Gamma_{ij,k} \equiv \mathbf{r}_{ij} \cdot \mathbf{r}_k$ могут быть выражены через производные ковариантных компонент 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 $\Gamma_{ij,k}$

$$\Gamma_{ij,k} = \frac{1}{2} \left(\frac{\partial g_{kj}}{\partial u^i} + \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k} \right). \quad (1.10.16)$$

Теперь решим систему уравнений (1.10.15) относительно Γ_{ij}^k и получим равенства

$$\Gamma_{ij}^n = g^{nk} \Gamma_{ij,k}. \quad (1.10.17)$$

Саму систему (1.10.15) перепишем в виде

$$\Gamma_{ij,n} = g_{nk} \Gamma_{ij}^k. \quad (1.10.18)$$

Подставим в (1.10.17) равенства (1.10.16) и получим окончательное выражение для коэффициентов в разложении (1.10.13)

$$\Gamma_{ij}^n = \frac{1}{2} g^{nk} \left(\frac{\partial g_{kj}}{\partial u^i} + \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k} \right). \quad (1.10.19)$$

Коэффициенты $\Gamma_{ij,k}$ называются символами Кристоффеля 1-го рода, а коэффициенты Γ_{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 (1.10.15)–(1.10.19) идентичны соотношениям (1.9.15)–(1.9.19) 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 δt 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

$$\mathbf{a}(t) = \frac{du^i(t)}{dt} \mathbf{r}_i = a^i(t) \mathbf{r}_i. \quad (1.10.20)$$

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

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

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 δu^1 , δu^2 , δu^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

$$\frac{d\mathbf{a}}{dt} = \frac{da^i}{dt} \mathbf{r}_i + a^i \Gamma_{ij}^k \mathbf{r}_k \frac{du^j}{dt} = \frac{da^i}{dt} \mathbf{r}_i + a^n \Gamma_{nj}^i \mathbf{r}_i \frac{du^j}{dt} = \left(\frac{da^i}{dt} + a^n \Gamma_{nj}^i \frac{du^j}{dt} \right) \mathbf{r}_i. \quad (1.10.21)$$

Выражение

$$\frac{Da^i}{dt} = \frac{da^i}{dt} + a^n \Gamma_{nj}^i \frac{du^j}{dt} \quad (1.10.22)$$

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

$$\mathbf{a}(t) = a_i \mathbf{r}^i, \quad (1.10.23)$$

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

$$\mathbf{r}_i \cdot \frac{\partial \mathbf{r}^k}{\partial u^j} = -\frac{\partial \mathbf{r}_i}{\partial u^j} \cdot \mathbf{r}^k = -\Gamma_{ij}{}^n \mathbf{r}_n \cdot \mathbf{r}^k = -\Gamma_{ij}{}^k, \quad (1.10.24)$$

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

$$\mathbf{r}^k{}_j \equiv \frac{\partial \mathbf{r}^k}{\partial u^j} = -\Gamma_{ij}{}^k \mathbf{r}^i. \quad (1.10.25)$$

Действительно, если мы скалярно умножим равенство (1.10.25) на \mathbf{r}_i , то получим равенства (1.10.24). Таким образом, с точностью до линейных слагаемых относительно δu^j изменение векторов взаимного базиса равно

$$\delta \mathbf{r}^k = \frac{\partial \mathbf{r}^k}{\partial u^j} \delta u^j = -\Gamma_{ij}{}^k \mathbf{r}^i \delta u^j. \quad (1.10.26)$$

Используя покомпонентное представление (1.10.23), получим формулу для производной векторной функции кривой

$$\frac{d\mathbf{a}}{dt} = \frac{da_i}{dt} \mathbf{r}^i - a_i \Gamma_{kj}{}^i \mathbf{r}^k \frac{du^j}{dt} = \frac{da_i}{dt} \mathbf{r}^i - a_n \Gamma_{ij}{}^n \mathbf{r}^i \frac{du^j}{dt} = \left(\frac{da_i}{dt} - a_n \Gamma_{ij}{}^n \frac{du^j}{dt} \right) \mathbf{r}^i. \quad (1.10.27)$$

Выражение

$$\frac{Da_i}{dt} = \frac{da_i}{dt} - a_n \Gamma_{ij}{}^n \frac{du^j}{dt} \quad (1.10.28)$$

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 (1.10.21) и (1.10.27), 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^j на бесконечно

малую величину δu^j . 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^j 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^j к нулю, получим формулу для производной векторного поля в представлении через контравариантные компоненты

$$\frac{\partial \mathbf{a}}{\partial u^j} = \frac{\partial a^i}{\partial u^j} \mathbf{r}_i + a^i \Gamma_{ij}^k \mathbf{r}_k = \left(\frac{\partial a^i}{\partial u^j} + a^n \Gamma_{nj}^i \right) \mathbf{r}_i. \quad (1.10.29)$$

Выражение

$$\nabla_j a^i = \frac{\partial a^i}{\partial u^j} + a^n \Gamma_{nj}^i \quad (1.10.30)$$

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^j 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

$$\frac{\partial \mathbf{a}}{\partial u^j} = \frac{\partial a_i}{\partial u^j} \mathbf{r}^i - a_i \Gamma_{kj}^i \mathbf{r}^k = \left(\frac{\partial a_i}{\partial u^j} - a^n \Gamma_{ij}^n \right) \mathbf{r}^i. \quad (1.10.31)$$

Выражение

$$\nabla_j a_i = \frac{\partial a_i}{\partial u^j} - a_n \Gamma_{ij}^n \quad (1.10.32)$$

называют *ковариантной производной* ковариантных компонент векторного поля.

Итак, нами получены формулы (1.10.22) и (1.10.28) для вычисления производных вектора кривой $\mathbf{a}(t)$ по параметру t в произвольной криволинейной системе координат. Формулы (1.10.22), (1.10.28), (1.10.30), (1.10.32) 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 (1.10.14) и (1.10.25), получим выражения для производных компонент метрического тензора

$$\frac{\partial g_{ki}}{\partial u^j} = g_{kn} \Gamma_{ij}^n + g_{ni} \Gamma_{kj}^n, \quad (1.10.33)$$

$$\frac{\partial g^{ki}}{\partial u^j} = -g^{kn} \Gamma_{nj}^i - g^{ni} \Gamma_{nj}^k. \quad (1.10.34)$$

Подставив в равенство (1.10.30) соотношения $a^i = g^{ik} a_k$ и (1.10.34), получим

$$\begin{aligned} \nabla_j a^i &= \nabla_j (g^{ik} a_k) = g^{ik} \frac{\partial a_k}{\partial u^j} + a_k \frac{\partial g^{ki}}{\partial u^j} + a_k g^{nk} \Gamma_{nj}^i = \\ &= g^{ik} \frac{\partial a_k}{\partial u^j} - a_k g^{ni} \Gamma_{nj}^k = g^{ik} \left(\frac{\partial a_k}{\partial u^j} - a_n \Gamma_{kj}^n \right) = g^{ik} \nabla_j a_k. \end{aligned} \quad (1.10.35)$$

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 (1.10.32) соотношения $a_i = g_{ik} a^k$ и (1.10.33), получим

$$\begin{aligned} \nabla_j a_i &= \nabla_j (g_{ik} a^i) = g_{ik} \frac{\partial a^k}{\partial u^j} + a^k \frac{\partial g_{ik}}{\partial u^j} - a^k g_{kn} \Gamma_{ij}^n = \\ &= g_{ik} \frac{\partial a^k}{\partial u^j} + a^k g_{ni} \Gamma_{kj}^n = g_{ik} \left(\frac{\partial a^k}{\partial u^j} + a^n \Gamma_{nj}^k \right) = g_{ik} \nabla_j a^k. \end{aligned} \quad (1.10.36)$$

Из равенств (1.10.35) и (1.10.36) 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 свойствах 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

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

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

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

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

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

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

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)

$$\begin{aligned} D_w D_t \mathbf{a} - D_t D_w \mathbf{a} &= \left(-a_n \frac{\partial \Gamma_{ij}^n}{\partial w} \frac{\partial u^j}{\partial t} + a_m \Gamma_{nj}^m \frac{\partial u^j}{\partial t} \Gamma_{ir}^n \frac{\partial u^r}{\partial w} + \right. \\ &\quad \left. + a_n \frac{\partial \Gamma_{ij}^n}{\partial t} \frac{\partial u^j}{\partial w} - a_m \Gamma_{nj}^m \frac{\partial u^j}{\partial w} \Gamma_{ir}^n \frac{\partial u^r}{\partial t} \right) \mathbf{r}^i dw dt = \\ &= \left(\left(-\frac{\partial \Gamma_{ij}^n}{\partial u^r} + \Gamma_{qj}^n \Gamma_{ir}^q \right) \frac{\partial u^j}{\partial t} \frac{\partial u^r}{\partial w} + \left(\frac{\partial \Gamma_{ir}^n}{\partial u^j} - \Gamma_{qr}^n \Gamma_{ij}^q \right) \frac{\partial u^j}{\partial t} \frac{\partial u^r}{\partial w} \right) a_n \mathbf{r}^i dw dt = \\ &= \left(-\frac{\partial \Gamma_{ij}^n}{\partial u^r} + \Gamma_{kj}^n \Gamma_{ir}^k + \frac{\partial \Gamma_{ir}^n}{\partial u^j} - \Gamma_{kr}^n \Gamma_{ij}^k \right) \frac{\partial u^j}{\partial t} \frac{\partial u^r}{\partial w} a_n \mathbf{r}^i dw dt = \\ &= R_{rji}^n \frac{\partial u^j}{\partial t} \frac{\partial u^r}{\partial w} a_n dw dt \mathbf{r}^i, \quad (1.11.1) \end{aligned}$$

где введено обозначение

$$R_{rji}{}^n = \frac{\partial \Gamma_{ir}{}^n}{\partial u^j} - \frac{\partial \Gamma_{ij}{}^n}{\partial u^r} + \Gamma_{kj}{}^n \Gamma_{ir}{}^k - \Gamma_{kr}{}^n \Gamma_{ij}{}^k. \quad (1.11.2)$$

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 (1.11.1) примет вид

$$D_r D_j \mathbf{a} - D_j D_r \mathbf{a} = R_{rji}{}^n a_n \mathbf{r}^i du^r du^j, \quad (1.11.3)$$

так как $dw = du^r$ и $dt = du^j$. Левая часть (1.11.3) 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

$$D_r D_j \mathbf{a} - D_j D_r \mathbf{a} = -R_{rji}{}^n a^i \mathbf{r}_n du^r du^j. \quad (1.11.4)$$

Из формулы (1.11.3) видно, что тензор кривизны кососимметричен по первым двум индексам

$$R_{rji}{}^n = -R_{jri}{}^n. \quad (1.11.5)$$

Компоненты тензора кривизны (1.11.2) 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 и r , то мы получим *тензор Риччи*, ковариантные компоненты которого равны

$$R_{ji} = R_{jni}{}^n. \quad (1.11.6)$$

В евклидовом пространстве все компоненты тензора кривизны равны нулю:

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

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 $\mathbf{a} = a_i \mathbf{r}^i = a^i \mathbf{r}_i$ можно записать тензор второго ранга, например, метрический в виде суммы диад базисных векторов

$$\mathbf{G} = g_{ik} \mathbf{r}^i \mathbf{r}^k = g^{ik} \mathbf{r}_i \mathbf{r}_k. \quad (1.11.7)$$

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

$$\begin{aligned} \frac{\partial(g_{ik}r^i r^k)}{\partial u^j} &= \frac{\partial g_{ik}}{\partial u^j} r^i r^k + g_{ik} \frac{\partial r^i}{\partial u^j} r^k + g_{ik} r^i \frac{\partial r^k}{\partial u^j} = \\ &= \frac{\partial g_{ik}}{\partial u^j} r^i r^k - g_{ik} \Gamma_{nj}^i r^n r^k - g_{ik} r^i \Gamma_{nj}^k r^n = \\ &= \left(\frac{\partial g_{ik}}{\partial u^j} - g_{nk} \Gamma_{ij}^n - g_{in} \Gamma_{kj}^n \right) r^i r^k = \nabla_j g_{ik} r^i r^k, \quad (1.11.8) \end{aligned}$$

$$\begin{aligned} \frac{\partial(g^{ik}r_i r_k)}{\partial u^j} &= \frac{\partial g^{ik}}{\partial u^j} r_i r_k + g^{ik} \frac{\partial r_i}{\partial u^j} r_k + g^{ik} r_i \frac{\partial r_k}{\partial u^j} = \\ &= \frac{\partial g^{ik}}{\partial u^j} r_i r_k + g^{ik} \Gamma_{ij}^n r_n r_k + g^{ik} r_i \Gamma_{kj}^n r_n = \\ &= \left(\frac{\partial g^{ik}}{\partial u^j} + g^{nk} \Gamma_{nj}^i + g^{in} \Gamma_{nj}^k \right) r_i r_k = \nabla_j g^{ik} r_i r_k, \quad (1.11.9) \end{aligned}$$

где

$$\nabla_j g_{ki} = \frac{\partial g_{ki}}{\partial u^j} - g_{ni} \Gamma_{kj}^n - g_{kn} \Gamma_{ij}^n, \quad (1.11.10)$$

$$\nabla_j g^{ki} = \frac{\partial g^{ki}}{\partial u^j} + g^{ni} \Gamma_{nj}^k + g^{kn} \Gamma_{nj}^i \quad (1.11.11)$$

— ковариантные производные компонент метрического тензора. Как и следовало ожидать, из соотношений (1.10.33) и (1.10.34) ковариантные производные компонент метрического тензора равны нулю. Равенства (1.11.10) и (1.11.11) 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 $\mathbf{T} = t_m^{ki} \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m$ по правилам дифференцирования суммы и произведения функций

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

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

$$\nabla_j t_m^{ki} = \frac{\partial t_m^{ki}}{\partial u^j} + t_m^{ni} \Gamma_{nj}^k + t_m^{kn} \Gamma_{nj}^i - t_n^{ki} \Gamma_{mj}^n. \quad (1.11.12)$$

Из полученного результата можно сформулировать общее правило вычисления ковариантной производной компонент тензора произвольного ранга. Кроме частной производной данной компоненты тензора $\partial t_{****}^{****}/\partial u^j$ (индексы заменены точками) в ковариантной производной добавляются еще несколько сумм, каждая сумма соответствует одному из индексов. Так каждому нижнему индексу соответствует сумма вида $-t_{**n*}^{***} \Gamma_{*j}^n$, а каждому верхнему индексу соответствует сумма вида $+t_{***}^{*n*} \Gamma_{nj}^*$, где значение соответствующего индекса (обозначенного точкой) переходит от компоненты к символам Кристоффеля, а по индексу n производится суммирование. В символической записи это выглядит так:

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

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

$$\nabla = \frac{\partial}{\partial u^j} \mathbf{r}^j = \frac{\partial}{\partial u^1} \mathbf{r}^1 + \frac{\partial}{\partial u^2} \mathbf{r}^2 + \frac{\partial}{\partial u^3} \mathbf{r}^3. \quad (1.11.13)$$

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):

$$\nabla \mathbf{a} = \mathbf{r}^j \frac{\partial \mathbf{a}}{\partial u^j} = \mathbf{r}^j \mathbf{r}^i \left(\frac{\partial a_i}{\partial u^j} - a_n \Gamma_{ij}^n \right) = \nabla_j a_i \mathbf{r}^j \mathbf{r}^i = \text{grad } \mathbf{a}, \quad (1.11.14)$$

$$\nabla \cdot \mathbf{a} = \mathbf{r}^j \cdot \frac{\partial \mathbf{a}}{\partial u^j} = \mathbf{r}^j \cdot \mathbf{r}^i \left(\frac{\partial a_i}{\partial u^j} - a_n \Gamma_{ij}^n \right) = \nabla_j a^j = \text{div } \mathbf{a}, \quad (1.11.15)$$

$$\nabla \times \mathbf{a} = \mathbf{r}^j \times \frac{\partial \mathbf{a}}{\partial u^j} = \mathbf{r}^j \times \mathbf{r}^i \left(\frac{\partial a_i}{\partial u^j} - a_n \Gamma_{ij}^n \right) = \frac{1}{\sqrt{g}} e^{jik} \mathbf{r}_k \nabla_j a_i = \text{rot } \mathbf{a}. \quad (1.11.16)$$

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

$$\nabla \mathbf{T} = \mathbf{r}^j \frac{\partial \mathbf{T}}{\partial u^j} = \nabla_j t_m^{ki} \mathbf{r}^j \mathbf{r}_k \mathbf{r}_i \mathbf{r}^m = \text{grad } \mathbf{T},$$

$$\nabla \cdot \mathbf{T} = \mathbf{r}^j \cdot \frac{\partial \mathbf{T}}{\partial u^j} = \nabla_j t_m^{ji} \mathbf{r}_i \mathbf{r}^m = \text{div } \mathbf{T},$$

$$\nabla \times \mathbf{T} = \mathbf{r}^j \times \frac{\partial \mathbf{T}}{\partial u^j} = \frac{e^{jqn}}{\sqrt{g}} g_{kq} \nabla_j t_m^{ki} \mathbf{r}_n \mathbf{r}_i \mathbf{r}^m = \frac{e^{jkn}}{\sqrt{g}} \nabla_j t_{km}^i \mathbf{r}_n \mathbf{r}_i \mathbf{r}^m = \text{rot } \mathbf{T}.$$

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

$$\nabla^2 = \nabla \cdot \nabla = g^{ij} \frac{\partial^2}{\partial u^i \partial u^j} - g^{ik} \Gamma_{ik}^j \frac{\partial}{\partial u^j}. \quad (1.11.17)$$

Например, действие оператора Лапласа на скаляр записывается в виде

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

В декартовой прямоугольной системе координат оператор Лапласа имеет вид

$$\nabla^2 = \frac{\partial^2}{(\partial x^1)^2} + \frac{\partial^2}{(\partial x^2)^2} + \frac{\partial^2}{(\partial x^3)^2}. \quad (1.11.18)$$

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:

$$d\mathbf{a} = \frac{\partial \mathbf{a}}{\partial u^j} du^j = du^n \mathbf{r}_n \cdot \mathbf{r}^j \frac{\partial \mathbf{a}}{\partial u^j} = d\mathbf{r} \cdot \nabla \mathbf{a} = (\nabla \mathbf{a})^\top \cdot d\mathbf{r}, \quad (1.11.19)$$

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 (1.10.2), которые по формулам (1.10.4) 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

$$\begin{aligned} g_{ij} &= \mathbf{r}_i \cdot \mathbf{r}_j, \quad \mathbf{r}^k = g^{ki} \mathbf{r}_i, \quad \frac{\partial g_{ij}}{\partial u^k} = \mathbf{r}_{ik} \cdot \mathbf{r}_j + \mathbf{r}_i \cdot \mathbf{r}_{jk}, \\ \Gamma_{ij}^n &= \frac{1}{2} g^{nk} \left(\frac{\partial g_{kj}}{\partial u^i} + \frac{\partial g_{ik}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k} \right), \quad \mathbf{a} = a_k \mathbf{r}^k = a^k \mathbf{r}_k, \\ \mathbf{a}'(t) &= \frac{d\mathbf{a}}{dt} = \left(\frac{da^i}{dt} + a^n \Gamma_{nj}^i \frac{du^j}{dt} \right) \mathbf{r}_i = \left(\frac{da_i}{dt} - a_n \Gamma_{ij}^n \frac{du^j}{dt} \right) \mathbf{r}^i. \end{aligned} \quad (1.11.20)$$

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. $g_{ij} = 0$ и $g^{ij} = 0$ при $i \neq j$, а $g^{ii} = \frac{1}{g_{ii}}$. 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

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

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 (1.10.7). 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:

$$\mathbf{a} = a^r \mathbf{e}_r + a^\varphi \mathbf{e}_\varphi + a^z \mathbf{e}_z = a_r \mathbf{e}^r + a_\varphi \mathbf{e}^\varphi + a_z \mathbf{e}^z. \quad (1.12.1)$$

Частные производные векторов касательного базиса (1.10.14) по цилиндрическим координатам равны

$$\begin{aligned} \frac{\partial \mathbf{e}_r}{\partial r} &= 0, & \frac{\partial \mathbf{e}_r}{\partial \varphi} &= \frac{1}{r} \mathbf{e}_\varphi, & \frac{\partial \mathbf{e}_r}{\partial z} &= 0, \\ \frac{\partial \mathbf{e}_\varphi}{\partial r} &= -\frac{1}{r} \mathbf{e}_\varphi, & \frac{\partial \mathbf{e}_\varphi}{\partial \varphi} &= -r \mathbf{e}_r, & \frac{\partial \mathbf{e}_\varphi}{\partial z} &= 0, \\ \frac{\partial \mathbf{e}_z}{\partial r} &= 0, & \frac{\partial \mathbf{e}_z}{\partial \varphi} &= 0, & \frac{\partial \mathbf{e}_z}{\partial z} &= 0. \end{aligned} \quad (1.12.2)$$

Частные производные векторов взаимного базиса (1.10.25) по цилиндрическим координатам равны

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

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 (1.10.21) для производной векторной функции в цилиндрической системе координат имеет вид

$$\frac{d\mathbf{a}}{dt} = \left(\frac{da^r}{dt} - r a^\varphi \frac{d\varphi}{dt} \right) \mathbf{e}_r + \left(\frac{da^\varphi}{dt} + \frac{a^\varphi}{r} \frac{dr}{dt} + \frac{a^r}{r} \frac{d\varphi}{dt} \right) \mathbf{e}_\varphi + \frac{da^z}{dt} \mathbf{e}_z. \quad (1.12.3)$$

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 (1.10.20) описывается векторной функцией

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

Производная векторной функции спирали в соответствии с формулой (1.12.3) 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

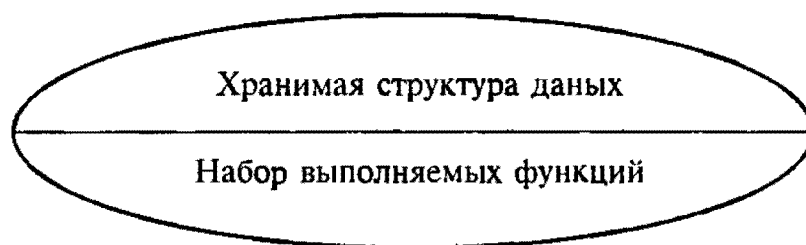


Рис. 1.13.1. Математическая модель

его модель требуемым образом (рис. 1.13.1). 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.

Глава 2

МОДЕЛИРОВАНИЕ КРИВЫХ ЛИНИЙ

2.1. Математическая модель кривых линий

В данной главе мы рассмотрим способы построения линий. Линиями можно описать отдельные геометрические свойства предметов, с помощью линий можно представить характерные черты воображаемых объектов. Кривые линии будут служить нам в качестве строительного материала для создания поверхностей и тел. В свою очередь, линии мы будем описывать с помощью скалярных величин, векторов, точек и других линий. Линии могут быть пространственными и двухмерными. Они имеют много общего. Сначала мы рассмотрим пространственные линии, а затем остановимся на отличиях двухмерных линий от них.

Пусть в пространстве задана некоторая глобальная декартова прямоугольная система координат с базисными векторами e_1, e_2, e_3 . Мы будем описывать пространственную линию векторной функцией

$$\mathbf{r}(t) = \sum_{i=1}^3 r_i(t) \mathbf{e}_i \quad (2.1.1)$$

скалярного аргумента t , изменяющегося в пределах $t_{\min} \leq t \leq t_{\max}$. Компоненты $r_1(t), r_2(t), r_3(t)$ радиус-вектора точки кривой $\mathbf{r}(t)$ являются однозначными непрерывными функциями параметра t . Такое описание линий называется *параметрическим*. Область изменения параметра кривой представляет собой отрезок $t_{\min} \leq t \leq t_{\max}$ в одномерном пространстве. Кривые могут быть замкнутыми или разомкнутыми. Для замкнутой кривой $\mathbf{r}(t \pm (t_{\max} - t_{\min})) = \mathbf{r}(t)$.

Двухмерную линию мы будем описывать векторной функцией

$$\mathbf{r}(t) = x(t) \mathbf{i}_1 + y(t) \mathbf{i}_2, \quad t_{\min} \leq t \leq t_{\max}. \quad (2.1.2)$$

Двухмерные векторы $\mathbf{i}_1, \mathbf{i}_2$ являются базисными векторами некоторой декартовой системы координат двухмерного пространства. Во многих случаях двухмерные линии аналогичны пространственным линиям.

В общем случае линию можно представить в виде системы двух уравнений, которым удовлетворяют координаты радиус-вектора точек линии. Каждое из этих уравнений можно считать уравнением поверхности, тогда система этих уравнений представит собой линию пересечения поверхностей. Если линия является плоской, то одним из уравнений может служить уравнение плоскости, а вторым — уравнение, связывающее двухмерные координаты радиус-вектора точек линии в этой плоскости. Если уравнение плоскости опустить, то мы придем к двухмерной линии, которая описывается одним уравнением. Такое описание двухмерной линии называется *нелинейным*. Описание линий уравнениями относительно ее координат в отличие от параметрического описания не всегда

однозначно. Оно также не является инвариантным относительно преобразования координат, т. е. при переходе в другую систему координат изменяются описывающие кривую линию уравнения. Неявное описание линий используется в теоретических исследованиях. При моделировании кривых линий мы будем использовать их параметрическое представление.

Для построения математической модели кривой линии нужно знать зависимость ее радиус-вектора $\mathbf{r}(t)$ от параметра и область изменения параметра t . Каждому типу кривой соответствует свой набор данных и свой алгоритм вычисления по ним радиус-вектора, а также своя область определения параметра кривой. При известной функциональной зависимости радиус-вектора от параметра кривой определяется вся геометрическая информация кривой. Для кривой всегда можно выполнить замену параметра и области его изменения.

Общение с математической моделью кривой происходит следующим образом. Мы обращаемся к функциям кривой с некоторым значением параметра и в качестве ответа получаем геометрическую информацию о кривой в точке, соответствующей заданному значению параметра. Различные функции кривой выдают различную информацию. Для одних функций значение параметра должно принадлежать области его определения, для других — может выходить за области определения. Если функция допускает выход значения параметра за область определения, то она должна выдавать геометрическую информацию для продолжения кривой.

2.2. Аналитические линии

Рассмотрим линии, координаты радиус-вектора которых могут быть представлены как аналитические функции некоторого параметра t . Простейшими из них являются конические сечения, спирали и некоторые другие линии, полученные как траектории движения точек механизмов. Многие кривые, описываемые аналитическими функциями, являются или замкнутыми или неограниченными. Для замкнутых кривых областью определения параметра будем считать отрезок параметрической оси, в пределах которого кривая проходит один цикл. Неограниченные кривые мы будем усекать, вводя минимальный и максимальный параметры.

Прямая линия. Одной из простейших линий является *прямая линия*. Она может быть описана точкой \mathbf{p} и вектором \mathbf{a} . Радиус-вектор прямой линии определим зависимостью

$$\mathbf{r}(t) = \mathbf{p} + t\mathbf{a}, \quad -\infty < t < \infty. \quad (2.2.1)$$

Длина вектора \mathbf{a} в общем случае может быть произвольной (но отличной от нуля). Теоретически областью изменения параметра прямой является вся числовая ось, но на практике удобно ввести ограничения для предельных значений параметра. В последнем случае мы получим отрезок прямой линии.

Отрезок. Другой простейшей линией является *отрезок прямой*. Он может быть представлен через начальную \mathbf{p}_1 и конечную \mathbf{p}_2 точки:

$$\mathbf{r}(t) = (1 - t)\mathbf{p}_1 + t\mathbf{p}_2, \quad 0 \leq t \leq 1. \quad (2.2.2)$$

Введем обозначение $v = (1 - t)$, тогда радиус-вектор отрезка будет описываться формулой

$$\mathbf{r}(t) = v\mathbf{p}_1 + t\mathbf{p}_2, \quad (2.2.3)$$

$$0 \leq t \leq 1, \quad 0 \leq v \leq 1, \quad t + v = 1.$$

Величины v и t , на которые умножаются координаты точек \mathbf{p}_1 и \mathbf{p}_2 , являются *барицентрическими координатами*. С использованием барицентрических координат точек \mathbf{p}_1 и \mathbf{p}_2 формула отрезка приобретает симметричный вид.

Плоские кривые. Описанию плоских кривых в пространстве можно придать следующий вид:

$$\mathbf{r}(t) = \mathbf{p} + x(t)\mathbf{i}_x + y(t)\mathbf{i}_y, \quad (2.2.4)$$

где \mathbf{p} — некоторая точка привязки характерной точки кривой, \mathbf{i}_x и \mathbf{i}_y — неколлинеарные векторы, $x(t)$ и $y(t)$ — скалярные функции. Аналогичный вид $\mathbf{r}(t) = \mathbf{p} + x(t)\mathbf{i}_x + y(t)\mathbf{i}_y$ имеет описание этих кривых в двухмерном пространстве. Разница заключается в том, что в двухмерном пространстве линии описываются точками и векторами, имеющими две компоненты, а в трехмерном пространстве — имеющими три компоненты. Точка \mathbf{p} и векторы \mathbf{i}_x и \mathbf{i}_y выбраны так, что скалярные функции $x(t)$ и $y(t)$ имеют канонический вид. Канонический вид примет и соотношение, связывающее эти скалярные функции. Векторы \mathbf{i}_x и \mathbf{i}_y вместе с вектором $\mathbf{i}_z = \mathbf{i}_x \times \mathbf{i}_y$ являются базисом местной декартовой системы координат, в которой кривая имеет канонический вид. Их всегда можно представить в виде разложения (1.2.3) по базису $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ глобальной декартовой системы координат:

$$\begin{aligned} \mathbf{i}_x &= a_{11}\mathbf{e}_1 + a_{12}\mathbf{e}_2 + a_{13}\mathbf{e}_3, \\ \mathbf{i}_y &= a_{21}\mathbf{e}_1 + a_{22}\mathbf{e}_2 + a_{23}\mathbf{e}_3, \\ \mathbf{i}_z &= a_{31}\mathbf{e}_1 + a_{32}\mathbf{e}_2 + a_{33}\mathbf{e}_3. \end{aligned} \quad (2.2.5)$$

Точка \mathbf{p} является началом этой местной декартовой системы координат. Она также может быть представлена в виде разложения по базису $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ глобальной декартовой системы координат:

$$\mathbf{p} = p_1\mathbf{e}_1 + p_2\mathbf{e}_2 + p_3\mathbf{e}_3. \quad (2.2.6)$$

Таким образом, структура данных аналитических линий содержит описание местной системы в виде начальной точки и трех ортов. При изменении положения и ориентации подобным образом описанной плоской кривой изменяются компоненты точки \mathbf{p} и векторов \mathbf{i}_x и \mathbf{i}_y , а скалярные функции $x(t)$ и $y(t)$ остаются неизменными, сохраняя канонический вид.

Окружность. *Окружность* можно описать, задав ее радиус r , положение центра \mathbf{p} и два взаимно ортогональных вектора единичной длины $\mathbf{i}_x, \mathbf{i}_y$, определяющих положение плоскости окружности:

$$\mathbf{r}(t) = \mathbf{p} + r \cos t \mathbf{i}_x + r \sin t \mathbf{i}_y, \quad 0 \leq t \leq 2\pi. \quad (2.2.7)$$

Окружность является плоской кривой. Окружность представляет собой геометрическое место точек на плоскости, равноудаленных от центра \mathbf{p} . Ее неявное описание с помощью координатных функций $x(t) = r \cdot \cos t$, $y(t) = r \cdot \sin t$ на плоскости имеет вид

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

Эллипс. *Эллипс* в изложенной концепции может быть описан радиус-вектором

$$\mathbf{r}(t) = \mathbf{p} + a \cos t \mathbf{i}_x + b \sin t \mathbf{i}_y, \quad 0 \leq t \leq 2\pi, \quad (2.2.8)$$

где a и b — полуоси эллипса. Скалярные функции $x(t) = a \cdot \cos t$, $y(t) = b \cdot \sin t$ эллипса связаны уравнением

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1.$$

Эллипс представляет собой геометрическое место точек на плоскости, для которых сумма расстояний до двух фиксированных точек f_1 и f_2 этой плоскости,

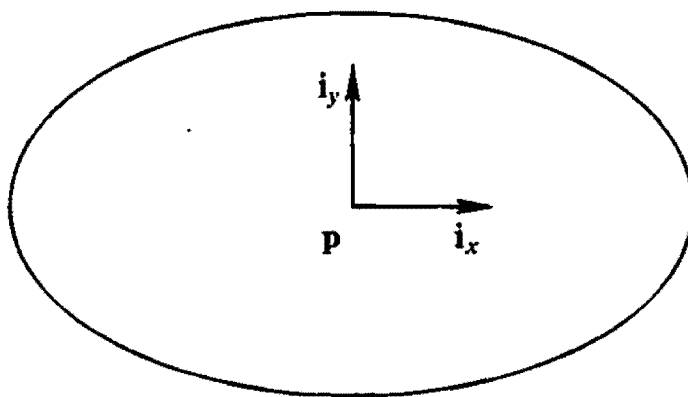


Рис. 2.2.1. Эллипс

называемых фокусами, есть постоянная величина, равная $2a$. Пусть $a > b$, тогда радиус-векторы фокусов равны

$$f_1 = p - \sqrt{a^2 - b^2} i_x, \quad f_2 = p + \sqrt{a^2 - b^2} i_x.$$

Эллипс показан на рис. 2.2.1. Окружность и эллипс являются замкнутыми линиями.

Гипербола. Гипербола может быть описана радиус-вектором

$$r(t) = p + a \operatorname{ch} t i_x + b \operatorname{sh} t i_y, \quad t_{\min} \leq t \leq t_{\max}. \quad (2.2.9)$$

Радиус-вектор (2.2.9) описывает только одну ветвь гиперболы. Вторая ветвь гиперболы может быть описана радиус-вектором

$$r(t) = p - a \operatorname{ch} t i_x + b \operatorname{sh} t i_y, \quad t_{\min} \leq t \leq t_{\max}.$$

Скалярные функции $x(t) = a \operatorname{ch} t$, $y(t) = b \operatorname{sh} t$ гиперболы связаны уравнением

$$\left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 = 1.$$

Данному уравнению удовлетворяют и функции $x(t) = \pm a \operatorname{ch} t$, $y(t) = \pm b \operatorname{sh} t$. У гиперболы мы будем использовать только одну ветвь. Гипербола представляет собой геометрическое место точек на плоскости, для которых абсолютная величина разности расстояний до двух фиксированных точек f_1 и f_2 этой плоскости, называемых фокусами, есть постоянная величина, равная $2a$. Радиус-векторы фокусов равны

$$f_1 = p - \sqrt{a^2 + b^2} i_x, \quad f_2 = p + \sqrt{a^2 + b^2} i_x.$$

Ветвь гиперболы, описываемая функцией (2.2.9), показана на рис. 2.2.2.

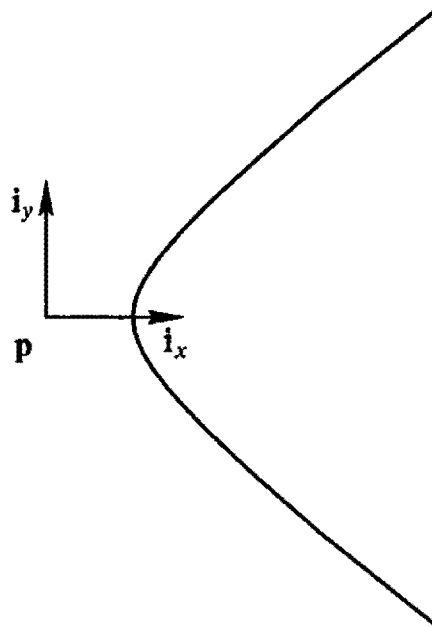


Рис. 2.2.2. Ветвь гиперболы

Парабола. Парабола может быть описана радиус-вектором

$$\mathbf{r}(t) = \mathbf{p} \frac{1}{2} p t^2 \mathbf{i}_x + p t \mathbf{i}_y, \quad t_{\min} \leq t \leq t_{\max}. \quad (2.2.10)$$

Скалярные функции $x(t) = \frac{1}{2} p t^2$, $y(t) = p t$ параболы связаны уравнением

$$2px - y^2 = 0.$$

Парабола представляет собой геометрическое место точек на плоскости, для которых расстояние до фиксированной точки \mathbf{f} этой плоскости, называемой *фоку-*

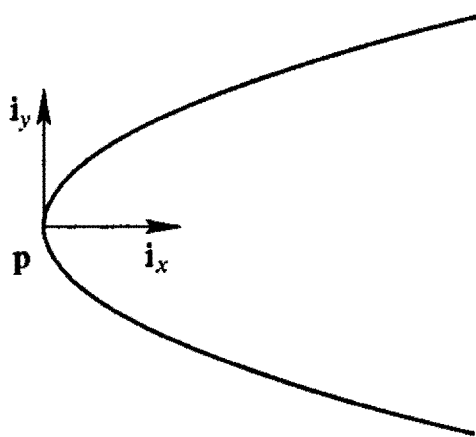


Рис. 2.2.3. Парабола

сом, равно расстоянию до некоторой фиксированной прямой $\mathbf{d}(w)$, называемой *директрисой*. Радиус-вектор фокуса и директрисы равны

$$\mathbf{f} = \mathbf{p} + \frac{p}{2} \mathbf{i}_x, \quad \mathbf{d}(w) = \mathbf{p} - \frac{p}{2} \mathbf{i}_x + w \mathbf{i}_y.$$

Величину $f = p/2$ называют фокусным расстоянием. Парабола показана на рис. 2.2.3.

Дуга. Дугу окружности можно описать, задав ее радиус r , положение центра \mathbf{p} , два взаимно ортогональных вектора единичной длины $\mathbf{i}_x, \mathbf{i}_y$, определяющих местную систему координат, начальный угол α_0 , угол дуги α и направление движения sign :

$$\mathbf{r}(t) = \mathbf{p} + r \cos(\alpha_0 + t \text{sign}) \mathbf{i}_x + r \sin(\alpha_0 + t \text{sign}) \mathbf{i}_y, \quad 0 \leq t \leq \alpha, \quad (2.2.11)$$

где $\text{sign} = +1$, если точка перемещается по дуге против часовой стрелки при взгляде навстречу вектору $\mathbf{i}_z = \mathbf{i}_x \times \mathbf{i}_y$, и $\text{sign} = -1$, если точка перемещается по дуге по часовой стрелке при взгляде навстречу вектору $\mathbf{i}_z = \mathbf{i}_x \times \mathbf{i}_y$ (в двухмерном случае при взгляде на плоскость).

Аналогичной формулой может быть описана дуга эллипса:

$$\mathbf{r}(t) = \mathbf{p} + a \cos(\alpha_0 + t \text{sign}) \mathbf{i}_x + b \sin(\alpha_0 + t \text{sign}) \mathbf{i}_y, \quad 0 \leq t \leq \alpha, \quad (2.2.12)$$

где a и b — полуоси эллипса.

Спираль. Цилиндрическая спираль также является аналитической кривой, но не является плоской. Она может быть описана положением начала оси спирали \mathbf{p} , тремя взаимно ортогональными векторами единичной длины $\mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z$, радиусом r , шагом h и параметрической длиной t_{\max} :

$$\mathbf{r}(t) = r \cos t \mathbf{i}_x + r \sin t \mathbf{i}_y + h \frac{t}{2\pi} \mathbf{i}_z, \quad 0 \leq t \leq t_{\max}. \quad (2.2.13)$$

Цилиндрическая спираль показана на рис. 2.2.4. Левую спираль можно получить путем присвоения отрицательного значения h .

Производные аналитических линий можно найти дифференцированием компонент радиус-вектора по параметру. Рассмотренные линии могут быть заданы

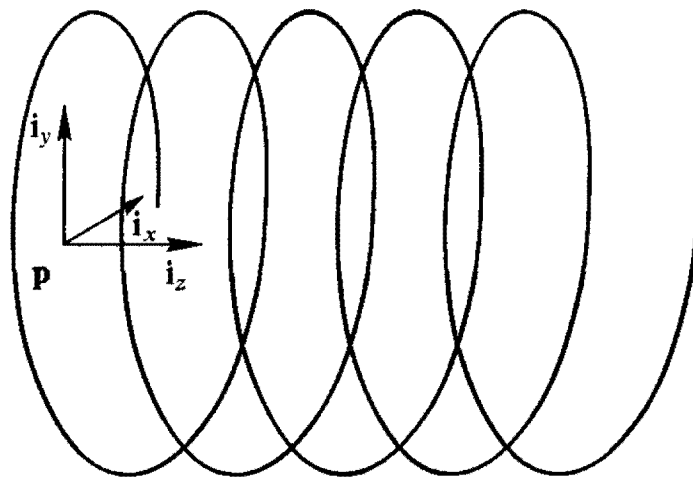


Рис. 2.2.4. Цилиндрическая спираль

и несколько иным образом. Например, радиус может быть заменен длиной вектора (уже не единичной), но в определенных случаях удобнее иметь данные о направлении и длине отдельно.

Можно построить большое количество других плоских и пространственных кривых, координаты радиус-вектора которых описываются аналитическими функциями. Мы не будем их рассматривать по той причине, что на практике они применяются редко.

2.3.* Кривые второго порядка

Эллипс, параболу и гиперболу называют коническими сечениями, так как они могут быть получены в результате пересечения конической поверхности с плоскостью, при условии, что плоскость не проходит через вершину конуса. Если угол плоскости с осью конуса больше угла образующей конуса с его осью, то получим эллипс. Если угол плоскости с осью конуса меньше угла образующей конуса с его осью, то получим гиперболу. Если указанные углы равны, то получим параболу. Конические сечения являются кривыми второго порядка. Выше мы рассмотрели параметрическое представление конических сечений. В некоторых случаях, например, при обмене данными, приходится иметь дело с координатными уравнениями для этих кривых. Поэтому требуется выполнять преобразование из неявного представления в параметрическое и обратно.

Декартовы координаты x и y на плоскости кривых второго порядка связаны уравнением второй степени

$$c_{11}x^2 + c_{22}y^2 + 2c_{12}xy + 2c_{13}x + 2c_{23}y + c_{33} = 0, \quad (2.3.1)$$

где c_{11} , c_{22} , c_{33} , c_{12} , c_{13} , c_{23} — коэффициенты, определяющие тип кривой, ее положение и ориентацию на плоскости (хотя бы один из коэффициентов c_{11} , c_{22} , c_{12} должен быть отличен от нуля). При перемещении или повороте кривой второго порядка коэффициенты c_{11} , c_{22} , c_{33} , c_{12} , c_{13} , c_{23} не остаются неизменными. Путем поворота и перемещения начала системы координат уравнение (2.3.1) можно преобразовать к наиболее простому виду, называемому каноническим. Величины

$$I_1 = c_{11} + c_{22}, \quad I_2 = c_{11}c_{22} - c_{12}^2, \quad I_3 = \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{vmatrix} \quad (2.3.2)$$

являются инвариантами линии второго порядка (2.3.1) относительно преобразований декартовой системы координат. Это можно доказать, рассмотрев преобразование переноса начала системы координат в точку x_0y_0 и поворота на угол φ

$$\begin{aligned} x &= x' \cos \varphi - y' \sin \varphi + x_0, \\ y &= x' \sin \varphi + y' \cos \varphi + y_0. \end{aligned} \quad (2.3.3)$$

Тип линии второго порядка и ее положение в пространстве определяются значениями инвариантов I_1 , I_2 , I_3 .

В расширенной матричной записи уравнение кривой второго порядка (2.3.1) имеет вид

$$\mathbf{R}^T \cdot \mathbf{C}_c \cdot \mathbf{R} = 0,$$

где использованы расширенный вектор (1.4.3)

$$\mathbf{R} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

и расширенная матрица

$$\mathbf{C}_c = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{bmatrix}.$$

В расширенной матричной записи преобразование координат (2.3.3) имеет вид

$$\mathbf{R} = \mathbf{A}_0^T \cdot \mathbf{R}',$$

где A_0 — расширенная матрица:

$$A_0 = \begin{bmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ x_0 & y_0 & 1 \end{bmatrix}.$$

Заметим, что определитель матрицы A_0 равен единице. После перехода к новым координатам x' , y' получим уравнение кривой второго порядка

$$R'^T \cdot A_0 \cdot C_c \cdot A_0^T \cdot R' = 0 \quad \text{или} \quad R'^T \cdot C_c' \cdot R' = 0, \quad \text{где} \quad C_c' = A_0 \cdot C_c \cdot A_0^T,$$

откуда следует, что I_3 есть инвариант относительно преобразования (2.3.3). Действительно,

$$I_3' \equiv |C_c'| = |A_0 \cdot C_c \cdot A_0^T| = |A_0| |C_c| |A_0^T| = |C_c| \equiv I_3.$$

Доказательство инвариантности величин I_1 , I_2 может быть выполнено путем замены координат (2.3.3) в (2.3.1) и вычисления величин (2.3.2) аналогично доказательству инвариантности I_3 .

Информация о типе линии и ее параметрах содержится в характеристической квадратичной форме, составленной из квадратичных членов относительно координат. Предположим, что найдена система координат $O'x'y'$, в которой уравнение кривой второго порядка имеет канонический вид

$$\lambda_1 x'^2 + \lambda_2 y'^2 + I = 0.$$

В силу свойства инвариантов будем иметь равенства $I_1 = \lambda_1 + \lambda_2$, $I_2 = \lambda_1 \lambda_2$, $I_3 = \lambda_1 \lambda_2 I$, откуда следует, что свободный член в каноническом уравнении равен $I = I_3 / (\lambda_1 \lambda_2) = I_3 / I_2$. Кривая второго порядка может быть приведена к указанному каноническому виду, если $I_2 \neq 0$. Кривые такого вида называются центральными. Для определения остальных канонических коэффициентов центральной кривой второго порядка составим квадратный многочлен

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1 \lambda_2 = \lambda^2 - I_1 \lambda + I_2,$$

корнями которого являются канонические коэффициенты. Так как I_1 и I_2 в любой системе координат имеют одни и те же значения, то следует вывод, что канонические коэффициенты λ_1 , λ_2 кривой второго порядка являются корнями уравнения

$$\lambda^2 - I_1 \lambda + I_2 = 0,$$

которое называется *характеристическим уравнением*. В матричной записи характеристическое уравнение имеет вид

$$\begin{vmatrix} c_{11} - \lambda & c_{12} \\ c_{12} & c_{22} - \lambda \end{vmatrix} = 0.$$

Таким образом, канонические коэффициенты являются собственными значениями матрицы характеристической квадратичной формы кривой второго порядка. Корни λ_1 , λ_2 характеристического уравнения являются действительными, так как матрица характеристической квадратичной формы симметричная.

Если $I_2 = 0$, то один из корней характеристического уравнения равен нулю. Пусть это будет первый по счету корень, и пусть $I_3 \neq 0$. Тогда кривая второго порядка может быть приведена к виду

$$2Jx' + \lambda_2 y'^2 = 0.$$

В силу свойства инвариантов для данного уравнения будем иметь равенства $I_1 = \lambda_2$, $I_2 = 0$, $I_3 = -\lambda_2 J^2$. Отсюда следует, что $J^2 = -I_3/\lambda_2 = -I_3/I_1$. Такая кривая не является центральной и имеет действительные точки при $I_3/I_1 < 0$.

Приведем классификацию линий второго порядка. Тип линии определяется знаком инварианта I_2 :

- эллиптический тип, если $I_2 > 0$,
- параболический тип, если $I_2 = 0$,
- гиперболический тип, если $I_2 < 0$.

Для того чтобы линии второго порядка не вырождались в прямые, необходимо, чтобы $I_3 \neq 0$. Если линия имеет эллиптический тип и $(I_3/I_1) > 0$, то уравнению (2.3.1) не удовлетворяет ни одна точка плоскости, а уравнение называют уравнением мнимого эллипса. Мы получим эллипс при $I_2 > 0$, $I_3 \neq 0$ и $(I_3/I_1) < 0$, гиперболу — при $I_2 < 0$ и $I_3 \neq 0$, параболу — при $I_2 = 0$ и $I_3 \neq 0$.

Каждая кривая второго порядка (2.3.1) в некоторой системе координат принимает канонический вид. Эту систему называют главной центральной (если линия имеет центр) системой координат. Кривая имеет центр, если $I_2 \neq 0$. Двухмерные координаты x_0 , y_0 центра системы координат, в которой кривая второго порядка примет канонический вид, находятся из системы уравнений

$$\begin{aligned} c_{11}x_0 + c_{12}y_0 + c_{13} &= 0, \\ c_{12}x_0 + c_{22}y_0 + c_{23} &= 0 \end{aligned} \quad (2.3.4)$$

и равны

$$x_0 = -\frac{1}{I_2} \begin{vmatrix} c_{13} & c_{12} \\ c_{23} & c_{22} \end{vmatrix}, \quad y_0 = -\frac{1}{I_2} \begin{vmatrix} c_{11} & c_{13} \\ c_{12} & c_{23} \end{vmatrix}. \quad (2.3.5)$$

Угол поворота φ главной системы координат относительно исходной удовлетворяет системе уравнений

$$\begin{aligned} (c_{11} - \lambda) \cos \varphi + c_{12} \sin \varphi &= 0, \\ c_{12} \cos \varphi + (c_{22} - \lambda) \sin \varphi &= 0, \end{aligned} \quad (2.3.6)$$

где λ — корни характеристического уравнения. Если корни характеристического уравнения различны, то система (2.3.6) имеет одно независимое уравнение. Исключим λ из уравнений (2.3.6) и получим

$$c_{12} \operatorname{tg} \varphi + c_{11} = c_{12} \operatorname{ctg} \varphi + c_{22}. \quad (2.3.7)$$

Если $c_{12} \neq 0$ или $c_{11} \neq c_{22}$, то мы можем найти угол поворота φ главной системы координат. Он определяется равенством

$$\operatorname{ctg} (2\varphi) = \frac{c_{11} - c_{22}}{2c_{12}}. \quad (2.3.8)$$

После преобразования координат в новой системе уравнение кривой (2.3.1) примет вид

$$\lambda_1 x^2 + \lambda_2 y^2 + \frac{I_3}{I_2} = 0,$$

если $I_2 \neq 0$ и, следовательно, оба корня характеристического уравнения отличны от нуля и

$$2x \sqrt{-\frac{I_3}{I_1}} + \lambda_2 y^2 = 0,$$

если $I_2 = 0$ и, следовательно, один из корней характеристического уравнения равен нулю (штрих около новых координат опущен). Если $I_2 \neq 0$, то каноническое уравнение кривой имеет симметричный вид и у кривой есть центральная точка (эллипс и гипербола). В противном случае кривая не имеет центра (парабола).

При переходе к главной центральной системе координат уравнение кривой второго порядка приобретает канонический вид:

$$\begin{aligned} \left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 &= 1 && \text{для эллипса,} \\ \left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 &= 1 && \text{для гиперболы,} \\ 2px - y^2 &= 0 && \text{для параболы.} \end{aligned}$$

Перенумеруем корни характеристического уравнения так, чтобы $\lambda_1 \geq \lambda_2$. Полуоси эллипса можно найти по формулам

$$a = \sqrt{-\frac{I_3}{\lambda_2 I_2}}, \quad b = \sqrt{-\frac{I_3}{\lambda_1 I_2}}, \quad \lambda_1 \geq \lambda_2 > 0, \quad (2.3.9)$$

действительную и мнимую полуоси гиперболы — по формулам

$$a = \sqrt{-\frac{I_3}{\lambda_1 I_2}}, \quad b = \sqrt{\frac{I_3}{\lambda_2 I_2}}, \quad \lambda_1 > 0 > \lambda_2, \quad (2.3.10)$$

параметр параболы — по формуле

$$p = -\frac{1}{\lambda_2} \sqrt{-\frac{I_3}{I_1}}, \quad \lambda_1 = 0 > \lambda_2. \quad (2.3.11)$$

Приведенные формулы позволяют построить параметрические зависимости для кривых второго порядка по их координатным уравнениям. Для перехода от параметрического представления к координатным уравнениям нужно знать каноническое уравнение $f(x, y) = 0$, которым связаны координатные функции $x(t)$, $y(t)$ кривой. Пусть на плоскости, в которой лежит кривая, заданы местная двухмерная декартова прямоугольная система координат и некоторая глобальная двухмерная декартова прямоугольная система координат. Пусть положение местной декартовой системы координат описывается центром в точке \mathbf{q} и ортами \mathbf{j}_1 и \mathbf{j}_2 , заданными в глобальной двухмерной системе координат выражениями

$$\begin{aligned} \mathbf{q} &= q_1 \mathbf{e}_1 + q_2 \mathbf{e}_2, \\ \mathbf{j}_1 &= \cos \varphi \mathbf{e}_1 + \sin \varphi \mathbf{e}_2, \\ \mathbf{j}_2 &= -\sin \varphi \mathbf{e}_1 + \cos \varphi \mathbf{e}_2, \end{aligned} \quad (2.3.12)$$

где $\mathbf{e}_1, \mathbf{e}_2$ — базис глобальной двухмерной системы координат. Местные координаты x, y связаны с глобальными двухмерными координатами p_1, p_2 уравнениями (1.2.17), которые в принятых здесь обозначениях имеют вид

$$\begin{aligned} x &= (p_1 - q_1) \cos \varphi + (p_2 - q_2) \sin \varphi, \\ y &= -(p_1 - q_1) \sin \varphi + (p_2 - q_2) \cos \varphi. \end{aligned} \quad (2.3.13)$$

Для получения уравнения $F(p_1, p_2) = 0$, связывающего глобальные координаты, подставим в уравнение $f(x, y) = 0$ выражения (2.3.13) для канонических координат.

2.4. Сплайны

Во многих практических задачах аналитическая формула линии не известна, но требуется, например, чтобы кривая проходила через заданные точки, имела определенную степень гладкости или в определенных точках имела заданные производные. Рассмотрим такие кривые.

Довольно большой класс линий можно построить по совокупности точек. Такие линии можно назвать *точечно-заданными*. Это ломаная линия и различные сплайны. Мы рассмотрим кубический сплайн, сплайн в форме Лагранжа, сплайн в форме Ньютона и сплайны Эрмита. Задача ставится следующим образом: имеется совокупность точек в пространстве, радиус-векторы которых равны \mathbf{p}_i , где $i = 0, 1, 2, \dots, n$ — номера точек; требуется построить линию, радиус-вектор которой при значениях параметра t_i , $i = 0, 1, 2, \dots, n$ был бы равен \mathbf{p}_i . Другими словами, необходимо построить интерполяционную кривую. Точки \mathbf{p}_i , которые определяют линию и являются ее внутренними данными, будем называть *характеристическими точками кривой*, точки t_i на числовой оси будем называть *узлами*, а значения параметра t_i будем называть *узловыми*.

Ломаная линия. Простейшей точечно-заданной линией является ломаная линия. Она состоит из отрезков, последовательно соединяющих заданные точки. Значение параметра в каждой последующей точке должно быть больше значения параметра в предыдущей точке $t_i < t_{i+1}$. Радиус-вектор ломаной определяется равенством

$$\begin{aligned} \mathbf{r}(t) &= \mathbf{p}_i(1 - w) + \mathbf{p}_{i+1}w, \\ w &= \frac{t - t_i}{t_{i+1} - t_i}, \quad t_0 \leq t \leq t_n, \end{aligned} \quad (2.4.1)$$

где $t_i \leq t \leq t_{i+1}$. Параметр w будем называть *местным параметром* на участке кривой между точками \mathbf{p}_i и \mathbf{p}_{i+1} . Первая производная ломаной линии в точках \mathbf{p}_i терпит разрыв по длине и по направлению. Параметр ломаной линии изменяется в одномерном пространстве. В этом пространстве для определения

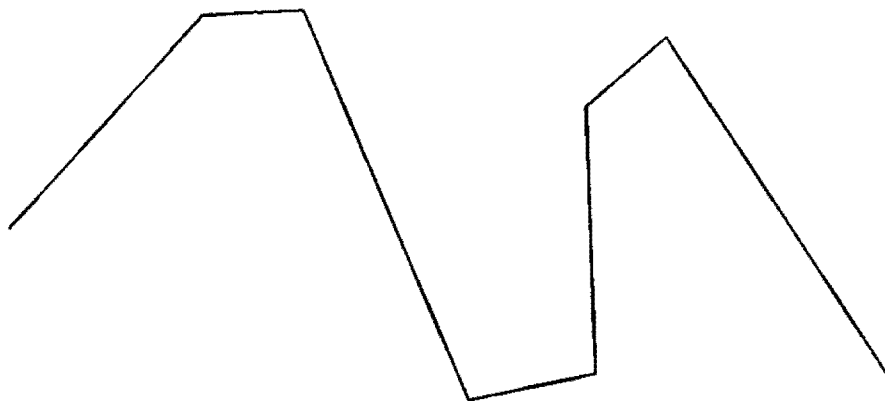


Рис. 2.4.1. Ломаная линия

параметра t мы вправе использовать любую систему координат. Для параметра можно использовать систему координат, где его значение в точке \mathbf{p}_i равно номеру точки: $t_i = i$. Такая параметризация называется *равномерной*, а параметрическая длина ломаной в таком случае равна числу точек минус один. Ломаная линия приведена на рис. 2.4.1.

Ломаная может быть замкнутой, в этом случае первая характеристическая точка одновременно является и последней. Параметрическая длина замкнутой ломаной линии равна числу точек, на которых она задана. Ломаная обладает

рядом полезных свойств: ее точки легко вычисляются, ее легко можно редактировать (вставить новую точку, удалить или сдвинуть имеющуюся), ее легко можно разрезать на части, каждая из которых также будет являться ломаной линией.

Сплайн Эрмита. Во многих практических задачах требуется построить плавную кривую линию, проходящую через заданные точки. Для этих целей строятся *сплайны*. Термин «сплайн» для кривых линий заимствован у названия чертежного инструмента — упругой гибкой линейки, которая может изгибаться так, чтобы проходить через заданные точки.

Если задана последовательность $m + 1$ точек, через которую должна пройти кривая, и производные ее радиус-вектора в этих точках, то по этим данным можно построить сплайн, описываемый полиномом степени $2m + 1$ и носящий имя Эрмита. Мы рассмотрим частный случай сплайна Эрмита для $m = 1$.

Ломаную линию можно рассматривать в качестве составной кривой, построенной из отрезков прямой линии. По аналогии можно построить составную кубическую кривую, состоящую из сплайнов Эрмита третьей степени, гладко стыкующихся между собой. Построим составной сплайн Эрмита, проходящий через заданную последовательность точек и имеющий в этих точках заданные производные. Пусть радиус-векторы этих точек равны \mathbf{p}_i , векторы производных кривой в этих точках равны \mathbf{q}_i , а значение параметра в этих точках равны t_i ($t_i < t_{i+1}$), где $i = 0, 1, 2, \dots, n$ — номера точек. На участке между точками \mathbf{p}_i и \mathbf{p}_{i+1} составной сплайн Эрмита является полиномом третьей степени местного параметра w

$$\begin{aligned} \mathbf{r}_i(w) &= \mathbf{a}_0 + \mathbf{a}_1 w + \mathbf{a}_2 w^2 + \mathbf{a}_3 w^3, \\ w &= \frac{t - t_i}{t_{i+1} - t_i}, \end{aligned} \quad (2.4.2)$$

Местный параметр w изменяется от 0 до 1. Векторы \mathbf{a}_j , $j = 0, 1, 2, 3$ найдем из условий на границе участка кривой

$$\begin{aligned} \mathbf{r}_i(0) &= \mathbf{p}_i, & \mathbf{r}_i(1) &= \mathbf{p}_{i+1}, \\ \mathbf{r}_i'(0) &= \mathbf{q}_i, & \mathbf{r}_i'(1) &= \mathbf{q}_{i+1}. \end{aligned}$$

После решения этой системы уравнений и подстановки искоемых значений в (2.4.2), получим зависимость радиус-вектора для сплайна Эрмита

$$\begin{aligned} \mathbf{r}(t) &= \mathbf{p}_i(1 - 3w^2 + 2w^3) + \\ &+ \mathbf{p}_{i+1}(3w^2 - 2w^3) + \mathbf{q}_i(w - 2w^2 + w^3) + \mathbf{q}_{i+1}(-w^2 + w^3) = \\ &= \alpha_0(w)\mathbf{p}_i + \alpha_1(w)\mathbf{p}_{i+1} + \beta_0(w)\mathbf{q}_i + \beta_1(w)\mathbf{q}_{i+1}, \end{aligned} \quad (2.4.3)$$

$$w = \frac{t - t_i}{t_{i+1} - t_i}, \quad t_0 \leq t \leq t_n,$$

где $t_i \leq t \leq t_{i+1}$. В (2.4.3) введены обозначения для функций

$$\begin{aligned} \alpha_0(w) &= 1 - 3w^2 + 2w^3, & \alpha_1(w) &= 3w^2 - 2w^3, \\ \beta_0(w) &= w - 2w^2 + w^3, & \beta_1(w) &= -w^2 + w^3, \end{aligned} \quad (2.4.4)$$

удовлетворяющих равенствам:

$$\begin{aligned} \alpha_0(0) &= 1, & \alpha_0(1) &= 0, & \alpha_1(0) &= 0, & \alpha_1(1) &= 1, \\ \alpha_0'(0) &= 0, & \alpha_0'(1) &= 0, & \alpha_1'(0) &= 0, & \alpha_1'(1) &= 0, \\ \beta_0(0) &= 0, & \beta_0(1) &= 0, & \beta_1(0) &= 0, & \beta_1(1) &= 0, \\ \beta_0'(0) &= 1, & \beta_0'(1) &= 0, & \beta_1'(0) &= 0, & \beta_1'(1) &= 1, \end{aligned}$$

где штрих означает дифференцирование по w . Если точки \mathbf{p}_i расположены равномерно, то можно принять значения параметра в точках \mathbf{p}_i равные номерам точек: $t_i = i$. При неравномерном расположении точек \mathbf{p}_i параметрическое расстояние $t_{i+1} - t_i$ можно положить пропорциональным расстоянию между соответствующими точками $|\mathbf{p}_{i+1} - \mathbf{p}_i|$. Составной сплайн Эрмита может быть замкнутым.

Мы рассмотрели случай, когда для кривой заданы производные в точках. Если производные \mathbf{q}_i неизвестны, то их можно вычислить по одной из схем. В первом случае их можно положить равными

$$\mathbf{q}_i = \frac{\mathbf{p}_{i+1} - \mathbf{p}_{i-1}}{t_{i+1} - t_{i-1}}.$$

При неравномерном расположении точек данный способ определения производных \mathbf{q}_i может привести к появлению нежелательных петель. Для предотвращения

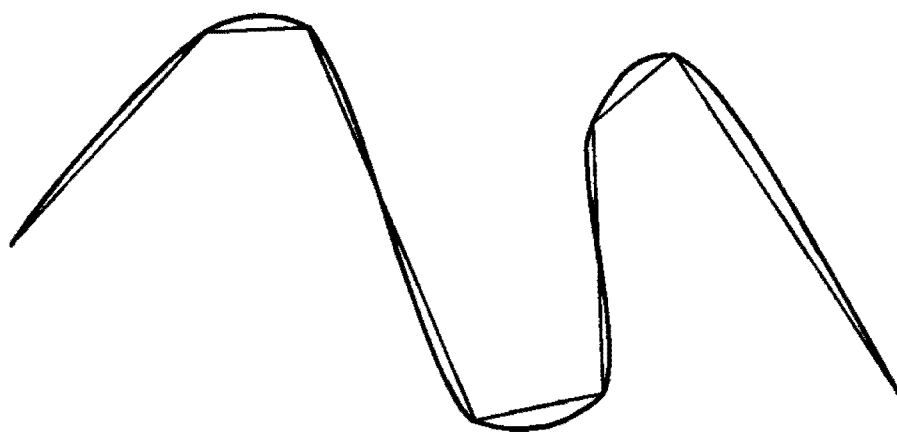


Рис. 2.4.2. Ломаная линия и составной сплайн Эрмита

появления петель нужно использовать другую схему определения производных \mathbf{q}_i . Например, их можно положить равными

$$\mathbf{q}_i = s_{i+1} \frac{\mathbf{p}_i - \mathbf{p}_{i-1}}{s_i + s_{i+1}} + s_i \frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{s_i + s_{i+1}},$$

где $s_i = |\mathbf{p}_i - \mathbf{p}_{i-1}|$, $s_{i+1} = |\mathbf{p}_{i+1} - \mathbf{p}_i|$ — расстояния между соседними точками. По третьей схеме меняются местами вклады расстояний между соседними точками:

$$\mathbf{q}_i = s_i \frac{\mathbf{p}_i - \mathbf{p}_{i-1}}{s_i + s_{i+1}} + s_{i+1} \frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{s_i + s_{i+1}}.$$

На рис. 2.4.2 приведены ломаная линия и составной сплайн Эрмита, построенный данным способом — по характеристическим точкам ломаной.

При неравномерном расположении точек данный способ определения производных q_i , как и первый способ, может привести к появлению петель.

Предложенные схемы не позволяют получить производные радиус-вектора кривой на ее краях, если она не замкнута. Производные на краях можно получить исходя из целей, которые преследуются при построении кривой. Найдем производные в крайних точках составной кривой из условия, что в этих точках обращаются в нуль третьи производные радиус-вектора. Для этого вычислим по (2.4.3) производные для соответствующих участков и подставим в них соответствующие значения параметра, в результате получим

$$\begin{aligned} q_0 &= 2(p_1 - p_0) - q_1, \\ q_n &= 2(p_n - p_{n-1}) - q_{n-1}. \end{aligned}$$

Аналогично можно найти производные в крайних точках составной кривой из других условий. Такими условиями могут служить: равенство нулю вторых производных на концах кривой; равенство производных радиус-вектора на концах заданным значениям. Составной сплайн Эрмита дает приемлемую аппроксимацию при большой плотности точек. Вторые производные в характеристических точках составном сплайне Эрмита не сохраняют непрерывность.

Кубический сплайн. Построим на заданной совокупности характеристических точек сплайн, который бы имел непрерывными первые и вторые производные радиус-вектора. На каждом участке между соседними характеристическими точками будем описывать радиус-вектор кривой кубическим полиномом типа (2.4.2). Введем обозначения для вторых производных в характеристических точках $r''(t_i) = s_i$. Вторая производная радиус-вектора на участке $t_i \leq t \leq t_{i+1}$ является линейной функцией параметра t

$$\frac{d^2 r}{dt^2} = s_i \frac{t_{i+1} - t}{t_{i+1} - t_i} + s_{i+1} \frac{t - t_i}{t_{i+1} - t_i}.$$

После двукратного интегрирования получим

$$r(t) = s_i \frac{(t_{i+1} - t)^3}{6(t_{i+1} - t_i)} + s_{i+1} \frac{(t - t_i)^3}{6(t_{i+1} - t_i)} + c_1 t + c_2. \quad (2.4.5)$$

Постоянные интегрирования c_1 и c_2 определим из условий на концах участка $r(t_i) = p_i$ и $r(t_{i+1}) = p_{i+1}$. После вычислений получим

$$\begin{aligned} r(t) &= s_i \frac{(t_{i+1} - t)^3}{6(t_{i+1} - t_i)} + s_{i+1} \frac{(t - t_i)^3}{6(t_{i+1} - t_i)} + \left(\frac{p_i}{t_{i+1} - t_i} - s_i \frac{t_{i+1} - t_i}{6} \right) (t_{i+1} - t) + \\ &+ \left(\frac{p_{i+1}}{t_{i+1} - t_i} - s_{i+1} \frac{t_{i+1} - t_i}{6} \right) (t - t_i). \quad (2.4.6) \end{aligned}$$

Выражение (2.4.6) описывает кубический полином на отрезке $t_i \leq t \leq t_{i+1}$ и содержит две неизвестные величины s_i и s_{i+1} . Для их определения приравняем первую производную сплайна на правом конце отрезка $t_{i-1} \leq t \leq t_i$ первой

производной сплайна на левом конце отрезка $t_i \leq t \leq t_{i+1}$. После дифференцирования (2.4.6) и подстановки $t = t_i$ получим

$$\left. \frac{dr}{dt} \right|_{t_i} = -\frac{(2s_i + s_{i+1})(t_{i+1} - t_i)}{6} + \frac{p_{i+1} - p_i}{t_{i+1} - t_i}. \quad (2.4.7)$$

После замены в (2.4.6) i на $i - 1$, дифференцирования и подстановки $t = t_i$ получим

$$\left. \frac{dr}{dt} \right|_{t_i} = \frac{(2s_i + s_{i-1})(t_i - t_{i-1})}{6} + \frac{p_i - p_{i-1}}{t_i - t_{i-1}}. \quad (2.4.8)$$

Приравняв правые части последних двух выражений, получим следующее уравнение:

$$\begin{aligned} s_{i-1}(t_i - t_{i-1}) + 2s_i(t_{i+1} - t_{i-1}) + s_{i+1}(t_{i+1} - t_i) = \\ = 6 \frac{p_{i+1} - p_i}{t_{i+1} - t_i} - 6 \frac{p_i - p_{i-1}}{t_i - t_{i-1}}. \end{aligned} \quad (2.4.9)$$

Таких уравнений можно составить для $n - 1$ характеристических точек (в случае незамкнутого сплайна). Подлежат определению $n + 1$ неизвестных векторов s_i . Два недостающих уравнения для определения всех неизвестных составим, исходя из условий на концах кривой. Например, если аналогично сплайну (линейке-инструменту) считать концы кривой свободными, то можно положить $s_0 = s_n = 0$. В некоторых случаях можно принять $s_0 = s_1$, $s_n = s_{n-1}$, тогда на концевых отрезках сплайн будет иметь постоянную кривизну. Дополнительные условия зависят от физических или геометрических условий в каждом конкретном случае. Если требуется построить замкнутую кривую, то мы будем иметь

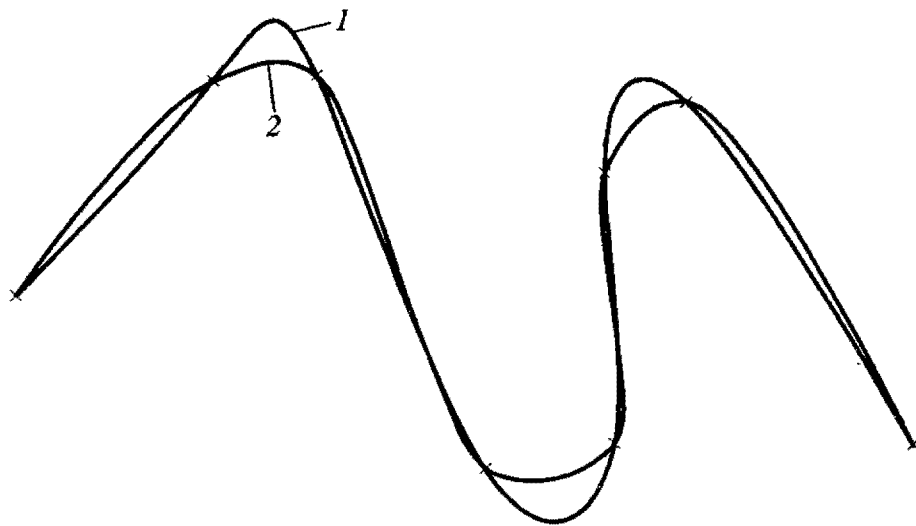


Рис. 2.4.3. Кубический сплайн (1), сплайн Эрмита (2)

равное число неизвестных и уравнений. Таким образом, неизвестные векторы вторых производных s_i в характеристических точках найдем из системы линейных уравнений. Матрица этой системы линейных уравнений трехдиагональная, что значительно упрощает решение системы. После вычисления

векторов s_i сплайн будет описываться формулой

$$\mathbf{r}(t) = (1 - w)\mathbf{p}_i + w\mathbf{p}_{i+1} + ((-2w + 3w^2 - w^3)s_i + (-w + w^3)s_{i+1}) \frac{(t_{i+1} - t_i)^2}{6}, \quad (2.4.10)$$

где $w = \frac{t - t_i}{t_{i+1} - t_i}$, а $t_i \leq t \leq t_{i+1}$.

Полученная кривая называется *кубическим сплайном*. В отличие от составного сплайна Эрмита для кубического сплайна изменение положения одной из характеристических точек приводит к перевычислению сплайна. На рис. 2.4.3 приведены кубический сплайн и составной сплайн Эрмита, построенные по одним и тем же восьми точкам.

Если положить, что значения параметра в характеристических точках равно их номерам ($t_i = i$), то уравнение (2.4.9) упростится и примет вид

$$s_{i-1} + 4s_i + s_{i+1} = 6(\mathbf{p}_{i+1} - \mathbf{p}_{i-1}). \quad (2.4.11)$$

Такая параметризация является равномерной. Кубические сплайны широко используются в задачах интерполяции. Равномерная параметризация при построении интерполяционного сплайна, проходящего через заданные точки, может привести к необоснованным изгибам кривой, если точки расположены неравномерно. Для избежания необоснованных изгибов кубического сплайна нужно использовать параметризацию, пропорциональную расстоянию между соседними точками. То есть параметризация должна соответствовать выражению

$$\frac{|\mathbf{p}_{i+1} - \mathbf{p}_i|}{t_{i+1} - t_i} \approx \frac{|\mathbf{p}_i - \mathbf{p}_{i-1}|}{t_i - t_{i-1}}.$$

По аналогии с кубическим сплайном могут быть построены сплайны более высокой степени, когда производные радиус-вектора третьего и более высокого порядка непрерывны в характеристических точках. Максимально возможный

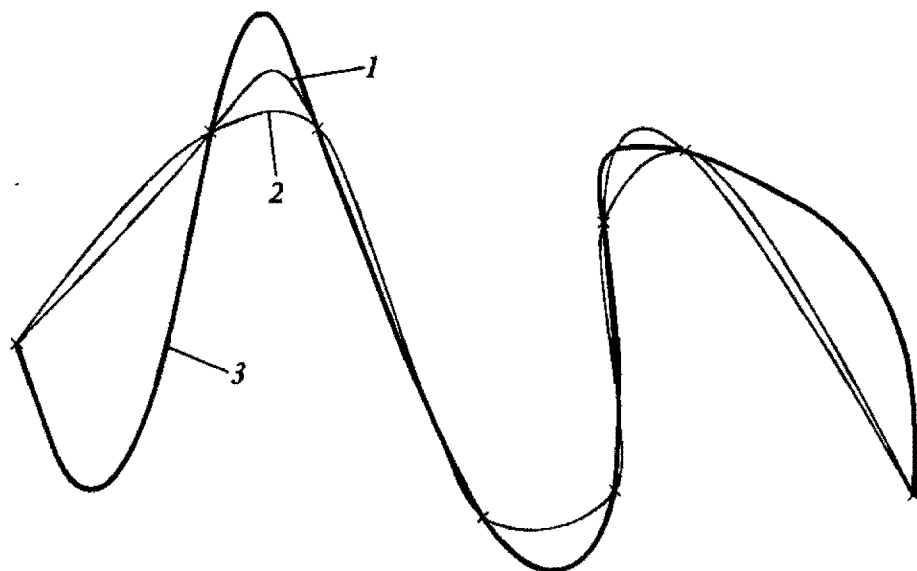


Рис. 2.4.4. Кубический сплайн (1), сплайн Эрмита (2), сплайн Лагранжа (3)

порядок производных радиус-вектора, для которых можно сохранить непрерывность, равен n при $n+1$ характеристических точках. Это можно сделать, построив интерполяционный полином Лагранжа или Ньютона. Интерполяционные

формулы Лагранжа и Ньютона дают один и тот же полином при разных способах его вычисления. В обоих случаях ищется векторный многочлен степени n вида

$$\mathbf{r}(t) = \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{a}_2 t^2 + \dots + \mathbf{a}_n t^n, \quad (2.4.12)$$

который при заданных значениях параметра t_i принимает заданные значения \mathbf{p}_i , $i = 0, 1, 2, \dots, n$. Для определения векторов \mathbf{a}_i получим систему линейных уравнений, определитель которой отличен от нуля, если среди точек нет совпадающих друг с другом. Сформулированная задача имеет единственное решение.

Сплайн Лагранжа. Интерполяционная формула Лагранжа представляет собой решение задачи построения кривой по заданным точкам в виде

$$\mathbf{r}(t) = \sum_{i=0}^n L_i(t) \mathbf{p}_i, \quad (2.4.13)$$

где функции $L_i(t)$ равны 1 при $t = t_i$ и равны 0 в остальных узлах t_j , $j \neq i$. Функции $L_i(t)$ будем искать в виде многочленов степени n , а именно:

$$L_i(t) = \lambda_i (t - t_0)(t - t_1) \dots (t - t_{i-1})(t - t_{i+1}) \dots (t - t_n).$$

Из условия $L_i(t_i) = 1$ находим $\frac{1}{\lambda_i} = (t_i - t_0)(t_i - t_1) \dots (t_i - t_{i-1})(t_i - t_{i+1}) \dots (t_i - t_n)$ и, соответственно, получим формулы для коэффициентов Лагранжа

$$L_i(t) = \frac{\prod_{j=0, j \neq i}^n (t - t_j)}{\prod_{j=0, j \neq i}^n (t_i - t_j)}. \quad (2.4.14)$$

Итак, интерполяционный полином Лагранжа имеет вид

$$\mathbf{r}(t) = \sum_{i=0}^n \frac{\prod_{j=0, j \neq i}^n (t - t_j)}{\prod_{j=0, j \neq i}^n (t_i - t_j)} \mathbf{p}_i, \quad t_0 \leq t \leq t_n. \quad (2.4.15)$$

Часто интерполяционный полином Лагранжа записывают в другом виде

$$\mathbf{r}(t) = \sum_{i=0}^n \frac{W_{0,n}(t)}{(t - t_i) W_{0,n}'(t_i)} \mathbf{p}_i, \quad (2.4.16)$$

где

$$W_{0,n}(t) = (t - t_0)(t - t_1)(t - t_2) \dots (t - t_{n-1})(t - t_n),$$

$$W_{0,n}'(t_i) = \left. \frac{dW_{0,n}}{dt} \right|_{t_i} = (t_i - t_0)(t_i - t_1) \dots (t_i - t_{i-1})(t_i - t_{i+1}) \dots (t_i - t_n).$$

На рис. 2.4.4 приведены кубический сплайн, составной сплайн Эрмита и сплайн Лагранжа, построенные по одним и тем же восьми точкам. На рисунке видно, что кривые ведут себя по-разному.

векторные коэффициенты \mathbf{p}_i являлись радиус-векторами характеристических точек кривой, а зависимость кривой от параметра сосредотачивалась в скалярных функциях $L_i(t)$, индивидуальных для каждой характеристической точки \mathbf{p}_i .

Де Кастелье и независимо от него Безье предложили аналогичный (2.5.2) подход для построения кривых. Эти кривые получили название *кривых Безье*. Кривая Безье описывается формулой

$$\mathbf{r}(t) = \sum_{i=0}^n B^n_i(t) \mathbf{p}_i = \sum_{i=0}^n \frac{n!}{i!(n-i)!} t^i (1-t)^{n-i} \mathbf{p}_i, \quad 0 \leq t \leq 1, \quad (2.5.3)$$

где \mathbf{p}_i — радиус-векторы точек, а n — число этих точек минус единица. Совокупность функций

$$B^n_i(t) = \frac{n!}{i!(n-i)!} t^i (1-t)^{n-i} = C^n_i t^i (1-t)^{n-i} \quad (2.5.4)$$

называется *базисом Бернштейна*. Можно сказать, что кривая Безье является полиномом Бернштейна с векторными коэффициентами. Сначала Де Кастелье и Безье использовали частные случаи кривой (2.5.3) и только позже Форрест установил связь коэффициентов кривой с полиномами Бернштейна. Коэффициенты при $t^i(1-t)^{n-i}$ в (2.5.4) равны коэффициентам $C^n_i = \frac{n!}{i!(n-i)!}$ бинома

Ньютона: $(a+b)^n = \sum_{i=0}^n C^n_i a^{n-i} b^i$. Из этого следует, что базис Бернштейна представляет собой разложение единицы (сумма всех коэффициентов при любом t равна единице)

$$\sum_{i=0}^n B^n_i(t) = (t + (1-t))^n = 1^n = 1. \quad (2.5.5)$$

Функции Бернштейна удовлетворяют рекуррентному соотношению

$$B^n_i = t B^{n-1}_{i-1} + (1-t) B^{n-1}_i.$$

Это соотношение доказывается непосредственной подстановкой:

$$t B^{n-1}_{i-1} + (1-t) B^{n-1}_i = \frac{i}{n} \frac{n!}{i!(n-i)!} t^i (1-t)^{n-i} + \frac{n-i}{n} \frac{n!}{i!(n-i)!} t^i (1-t)^{n-i} = B^n_i.$$

Используя это рекуррентное соотношение, можно вычислить все функции Бернштейна. Вычисление начинается с функции

$$B^0_0(t) = 1.$$

Далее получим $B^1_1(t) = t$, $B^1_0(t) = 1-t$. Все функции, один из индексов которых равен отрицательному числу, считаются равными нулю. Продолжив вычисления, получим все функции Бернштейна.

Алгоритм Де Кастелье. Подставим рекуррентное соотношение в выражение (2.5.3), выделив крайние точки и учитывая, что $B^n_0 = (1-t)^n$, $B^n_n = t^n$, получим

$$\begin{aligned} r(t) &= p_0(1-t)^n + p_n t^n + \sum_{i=1}^{n-1} (tB^{n-1}_{i-1} + (1-t)B^{n-1}_i) p_i = \\ &= (1-t) \left(p_0(1-t)^{n-1} + \sum_{i=1}^{n-1} B^{n-1}_i p_i \right) + t \left(p_n t^{n-1} + \sum_{i=1}^{n-1} B^{n-1}_{i-1} p_i \right) = \\ &= (1-t) \sum_{i=0}^{n-1} B^{n-1}_i p_i + t \sum_{i=0}^{n-1} B^{n-1}_i p_{i+1} = (1-t)r_0^{(n-1)}(t) + t r_1^{(n-1)}(t), \end{aligned}$$

где

$$r_0^{(n-1)}(t) = \sum_{i=0}^{n-1} B^{n-1}_i(t) p_i, \quad r_1^{(n-1)}(t) = \sum_{i=0}^{n-1} B^{n-1}_i(t) p_{i+1}.$$

Продолжив описанный процесс разложения для $r_0^{(n-1)}$ и $r_1^{(n-1)}$, в конце придем к равенствам

$$r_i^{(0)} = \sum_{j=0}^0 B^0_0(t) p_j = p_i, \quad i = 0, 1, \dots, n.$$

Обозначив $r(t)$ через $r_0^{(n)}(t)$, а p_i — через $r_i^{(0)}$, получим рекуррентное соотношение для вычисления точки кривой Безье

$$r_i^{(k)} = (1-t)r_i^{(k-1)} + t r_{i+1}^{(k-1)}, \quad i+k \leq n. \quad (2.5.6)$$

Алгоритм, описываемый соотношением (2.5.6), называется *алгоритмом Де Кастелье*. Алгоритм Де Кастелье позволяет вычислить любую точку кривой Безье по характеристическим точкам, ничего не зная о функциях Бернштейна. Величины $(1-t)$ и t , на которые умножаются координаты точек $r_i^{(k-1)}$ и $r_{i+1}^{(k-1)}$, являются барицентрическими координатами. Кривую Безье можно определить как линию, точки которой определяются рекуррентным соотношением (2.5.6).

Свойства кривых Безье. На рис. 2.5.1 и 2.5.2 приведены полиномы Бернштейна третьей и четвертой степени соответственно.

Из полиномов Бернштейна только нулевой и последний принимают максимально возможные значения $B^n_0(0) = 1$, $B^n_n(1) = 1$, поэтому кривая Безье проходит только через начальную p_0 и конечную p_n точки. Касательные к кривой Безье в начальной и конечной точках совпадают с направлением векторов $p_1 - p_0$ и $p_n - p_{n-1}$

$$\begin{aligned} r(0) &= p_0, \\ r(1) &= p_n, \\ r'(0) &= n(p_1 - p_0), \\ r'(1) &= n(p_n - p_{n-1}). \end{aligned}$$

По вершинам p_i строится *характеристическая ломаная* кривой Безье, которая касается кривой в крайних точках. Кривая Безье как бы сглаживает углы своей характеристической ломаной.

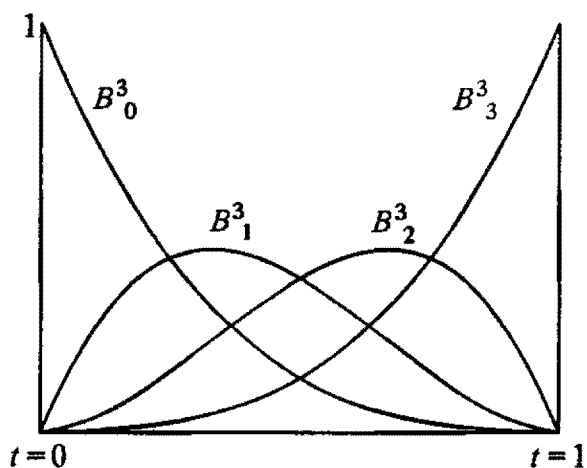


Рис. 2.5.1. Полиномы Бернштейна

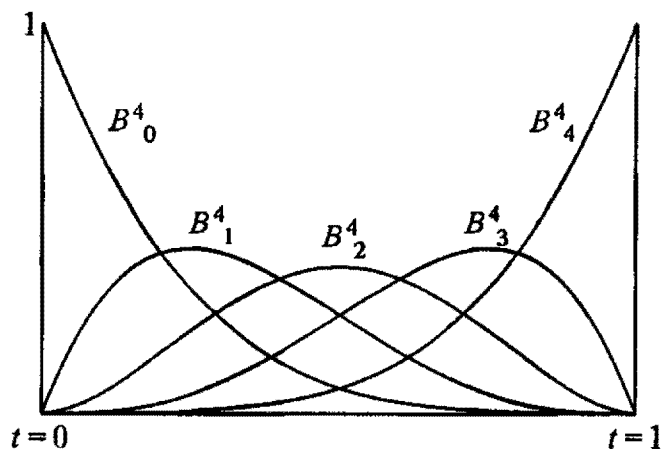


Рис. 2.5.2. Полиномы Бернштейна

Достаточно гибкой и в то же время достаточно простой для практического применения является кубическая кривая Безье, радиус-вектор которой описывается зависимостью

$$\mathbf{r}(t) = (1-t)^3 \mathbf{p}_0 + 3t(1-t)^2 \mathbf{p}_1 + 3t^2(1-t) \mathbf{p}_2 + t^3 \mathbf{p}_3.$$

Вид кубической кривой Безье и ее характеристической ломаной показаны на рис. 2.5.3 и 2.5.4.

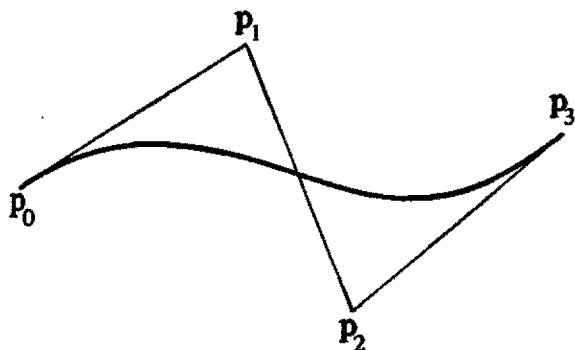


Рис. 2.5.3. Кривая Безье

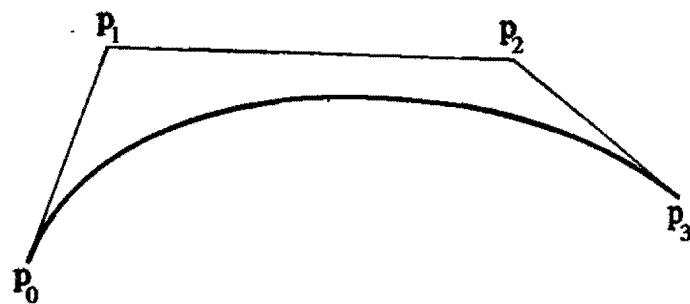


Рис. 2.5.4. Кривая Безье

Коэффициенты a_i этой же кривой в форме полинома (2.5.1) связаны с вершинами p_i соотношениями

$$\begin{aligned} a_0 &= p_0, & a_1 &= 3(p_1 - p_0), & a_2 &= 3(p_2 - 2p_1 + p_0), \\ & & a_3 &= p_3 - 3p_2 + 3p_1 - p_0. \end{aligned}$$

На примере кубической кривой Безье видно, что ее модификация путем перемещения точек p_i нагляднее, чем модификация путем изменения векторных коэффициентов a_i , так как в последнем случае трудно предугадать результат. Из-за того, что кривая не проходит через внутренние вершины своей характеристической ломаной, довольно сложно задать ей точное положение. Кривые в

форме Безье более удобны там, где важна форма линий, например, при проектировании конструкций типа кузова автомобиля. Чем больше точек участвуют

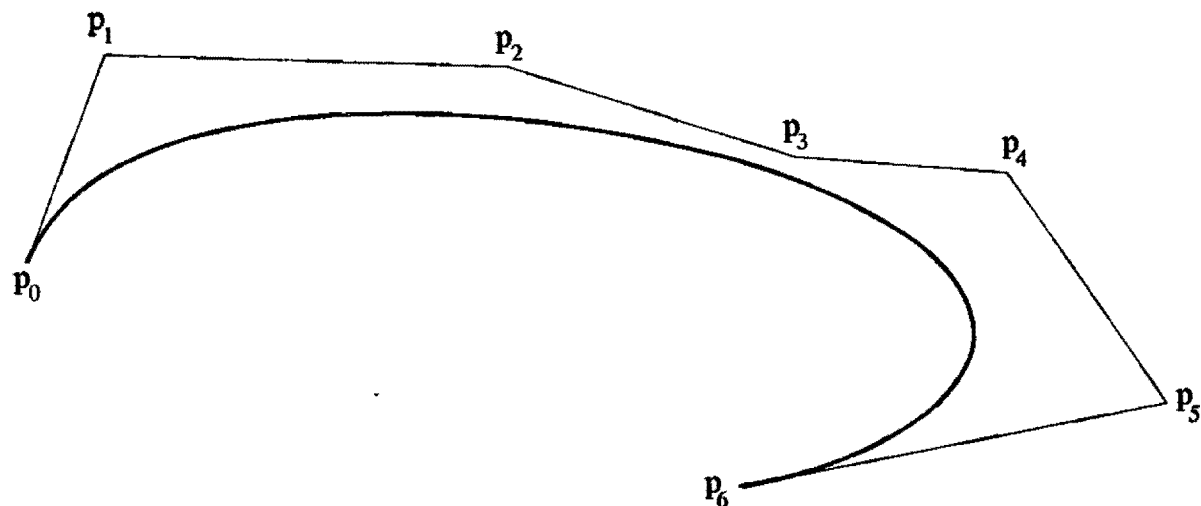


Рис. 2.5.5. Кривая Безье шестой степени

в построении кривой, тем она более гладкая. На рис. 2.5.5 приведена кривая Безье, построенная по 7 точкам.

Для построения кривой, проходящей через заданные точки, можно использовать составные кривые Безье. Кривая Безье всегда касается своей характеристической ломаной на концах. Если мы хотим, чтобы составная кривая была гладкой, необходимо следить за тем, чтобы в точках стыковки касательные к частям кривых совпадали. Но в общем случае в точках сочленения первая производная будет скачком менять свою длину, а все высшие производные будут претерпевать разрыв. Для непрерывности первой производной составной кривой Безье требуется касание и равенство длин крайних участков характеристических ломаных в месте стыковки. Параметрическая длина составной кривой Безье равна сумме параметрических длин частей, т. е. числу этих частей.

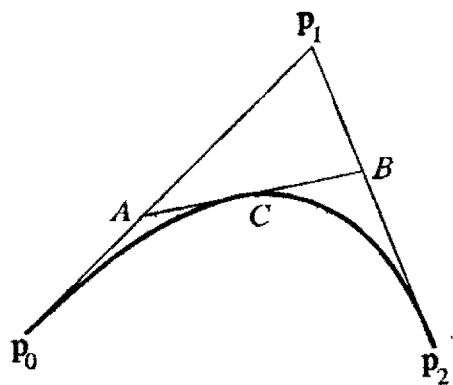


Рис. 2.5.6. Алгоритм Кастельяно для квадратичной кривой Безье

Если вершины кривой Безье лежат на одной прямой, кривая обращается в отрезок прямой. Отрезком прямой линии является и кривая Безье первой степени

$$\mathbf{r}(t) = (1 - t)\mathbf{p}_0 + t\mathbf{p}_1. \quad (2.5.7)$$

Радиус-вектор кривой Безье второй степени описывается зависимостью

$$\mathbf{r}(t) = (1 - t)^2\mathbf{p}_0 + 2t(1 - t)\mathbf{p}_1 + t^2\mathbf{p}_2. \quad (2.5.8)$$

Точка с произвольным параметром t квадратичной кривой Безье обладает тем свойством, что касательная в ней делит векторы $\mathbf{p}_1 - \mathbf{p}_0$ и $\mathbf{p}_2 - \mathbf{p}_1$ в отношении $t:(1 - t)$. Это можно заметить, если записать равенство (2.5.8) в виде

$$\mathbf{r}(t) = (1 - t)(\mathbf{p}_1 + (\mathbf{p}_0 - \mathbf{p}_1)(1 - t)) + t(\mathbf{p}_1 + (\mathbf{p}_2 - \mathbf{p}_1)t).$$

В этом же отношении точкой кривой делится и часть касательной, заключенная внутри характеристической ломаной (рис. 2.5.6).

Рисунок 2.5.6 иллюстрирует алгоритм Кастелье. Точка A имеет радиус-вектор $\mathbf{r}_0^{(1)}$, точка B имеет радиус-вектор $\mathbf{r}_1^{(1)}$, точка C имеет радиус-вектор $\mathbf{r}_0^{(2)} = \mathbf{r}(t)$.

Мы продолжим исследовать кривую Безье второй степени, и это приведет нас к новым линиям, обладающим весьма полезными свойствами.

2.6. Представление кривых второго порядка кривыми Безье

Квадратичная кривая Безье (2.5.8) является плоской линией и представляет собой полином второй степени параметра. Конические сечения (эллипс, парабола, гипербола) записываются в виде функции второй степени координат (но не параметра). Возникает вопрос: *можно ли некоторую часть конического сечения описать кривой Безье второй степени?* Попытаемся найти ответ на этот вопрос.

В декартовой системе координат на некоторой плоскости Oxy коническое сечение описывается уравнением второй степени (2.3.1)

$$c_{11}x^2 + 2c_{12}xy + c_{22}y^2 + 2c_{13}x + 2c_{23}y + c_{33} = 0.$$

Для нахождения коэффициентов c_{11} , c_{12} , c_{22} , c_{13} , c_{23} , c_{33} необходимо задать пять независимых условий. Воспользуемся методом построения конических сечений, который использовал Лайминг для проектирования поперечных сечений фюзеляжа самолета. Если обозначить через $l_1 = 0$ и $l_2 = 0$ уравнения двух прямых, записанных в виде $a_1x + b_1y + c_1 = 0$ и $a_2x + b_2y + c_2 = 0$, то уравнение $l_1l_2 = 0$ будет уравнением второго порядка и, следовательно, оно описывает некоторое коническое сечение. Возьмем еще одну пару прямых $a_3x + b_3y + c_3 = 0$ и $a_4x + b_4y + c_4 = 0$, записанных в виде уравнений $l_3 = 0$ и $l_4 = 0$, пересекающихся прямые первой пары, и составим уравнение

$$(1 - \lambda)l_1l_2 + \lambda l_3l_4 = 0, \quad (2.6.1)$$

где λ — некоторый параметр. Это уравнение также является уравнением второго порядка и при определенном значении параметра λ описывает некоторое коническое сечение. При любом значении параметра λ коническое сечение проходит через точки пересечения прямой l_1 с прямыми l_3 и l_4 и через точки пересечения прямой l_2 с прямыми l_3 и l_4 . В этих точках удовлетворяется уравнение (2.6.1) при любом λ . Уравнение (2.6.1) описывает семейство (или пучок) конических сечений, проходящих через четыре точки пересечения двух пар линий (рис. 2.6.1). Выбрать нужное коническое сечение из данного семейства можно, определив его пятую точку, не совпадающую с имеющимися точками, и тем самым определится параметр λ .

Используя уравнение (2.6.1), можно получить уравнение семейства конических сечений, имеющих в двух заданных точках заданные касательные.

На рис. 2.6.1 видно, что по мере приближения прямой l_3 к l_4 точка A приближается к точке B , точка C приближается к точке D , хорда AB и CD стремятся к касательным линиям конических сечений семейства. Когда прямые линии l_3 и l_4 совпадут, уравнение (2.6.1) примет вид

$$(1 - \lambda)l_1l_2 + \lambda l_3^2 = 0, \quad (2.6.2)$$

и будет представлять семейство конических сечений, касающихся линий l_1 и l_2 в точках их пересечения с линией l_3 (рис. 2.6.2). Если задать еще одну точку R , не совпадающую с точками A и C , то мы определим коническое сечение и параметр λ в (2.6.2). Коническое сечение в данном случае определяется четырьмя

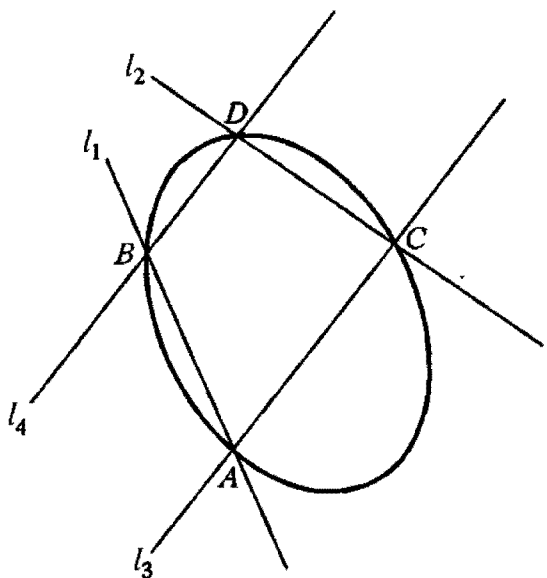


Рис. 2.6.1

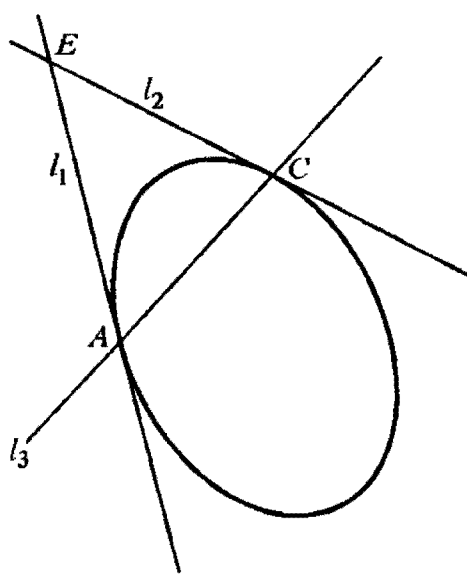


Рис. 2.6.2

точками: A , C , R и точкой E пересечения прямых линий l_1 и l_2 . Если точка R лежит внутри треугольника AEC , то коническое сечение будет иметь непрерывный участок, лежащий внутри этого треугольника. Можно показать, что если точка R делит пополам отрезок, соединяющий середины отрезков AE и CE , то коническое сечение является параболой, если точка R находится между упомянутой параболой и отрезком AC , то коническое сечение является эллипсом, в остальных случаях коническое сечение является гиперболой.

Уточним поставленный выше вопрос: можно ли описать часть конического сечения, лежащую внутри треугольника AEC , квадратичной кривой Безье, построенной по характеристической ломаной с вершинами в точках A , E и C . Пусть радиус-векторы точек A , E , C равны соответственно \mathbf{p}_0 , \mathbf{p}_1 , \mathbf{p}_2 . Введем косоугольную систему координат на плоскости, начало которой лежит в точке \mathbf{p}_1 , а координатным базисом являются векторы $\mathbf{p}_0 - \mathbf{p}_1$ и $\mathbf{p}_2 - \mathbf{p}_1$. Координаты этой системы обозначим через u и v . В косоугольной системе uv вектор \mathbf{p}_0 имеет координаты $u = 1$, $v = 0$, вектор \mathbf{p}_1 имеет координаты $u = 0$, $v = 0$, вектор \mathbf{p}_2 имеет координаты $u = 0$, $v = 1$. Произвольная точка R будет иметь радиус-вектор

$$\mathbf{r}(u, v) = \mathbf{p}_1 + u(\mathbf{p}_0 - \mathbf{p}_1) + v(\mathbf{p}_2 - \mathbf{p}_1). \quad (2.6.3)$$

В косоугольной системе прямые линии l_1 , l_2 , l_3 описываются уравнениями $v = 0$, $u = 0$, $u + v = 1$, соответственно. Подставим их в (2.6.2) и получим уравнение

$$S(u, v) = (1 - \lambda)uv + \lambda(u + v - 1)^2 = 0, \quad (2.6.4)$$

которое описывает семейство конических сечений в косоугольной системе координат.

Предположим, что мы выбрали из семейства (2.6.4) некоторое коническое сечение, т. е. нам известен в (2.6.4) параметр λ . Возьмем некоторую точку \mathbf{r} на коническом сечении внутри треугольника с вершинами в точках \mathbf{p}_0 , \mathbf{p}_1 , \mathbf{p}_2 (рис. 2.6.3).

Пусть ее координаты равны $0 < u_0 < 1$ и $0 < v_0 < 1$. Квадратичная кривая Безье описывается уравнением (2.5.8) как векторная функция параметра t . Попробуем записать в аналогичном виде коническое сечение. В общем случае

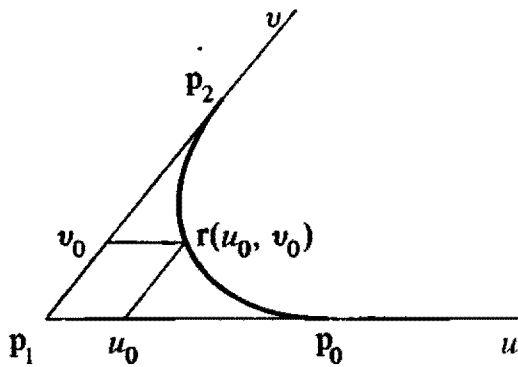


Рис. 2.6.3. Коническое сечение

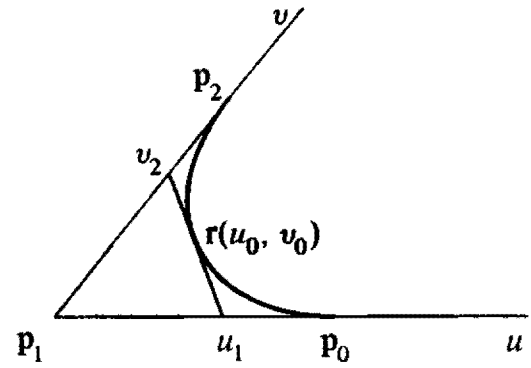


Рис. 2.6.4

можно считать, что оно описывается функцией (2.6.3), в которой параметры u и v связаны соотношением (2.6.4). Запишем радиус-вектор конического сечения в виде функции одного параметра, в свою очередь являющегося функцией u и v . Для этого нам потребуется касательная к коническому сечению в точке $r(u_0, v_0)$ и точки пересечения этой касательной с осями косоугольной системы координат.

Касательная линия к коническому сечению (2.6.4), записанному в виде $S(u, v) = 0$, в точке с координатами $u = u_0$, $v = v_0$ описывается уравнением

$$\left. \frac{\partial S}{\partial u} \right|_{u_0, v_0} (u - u_0) + \left. \frac{\partial S}{\partial v} \right|_{u_0, v_0} (v - v_0) = 0. \quad (2.6.5)$$

Производные $\partial S/\partial u$ и $\partial S/\partial v$ в точке $r(u_0, v_0)$ с учетом того, что u_0, v_0 удовлетворяют уравнению (2.6.4), равны

$$\begin{aligned} \left. \frac{\partial S}{\partial u} \right|_{u_0, v_0} &= (1 - \lambda)v_0 + 2\lambda(u_0 + v_0 - 1) = \lambda \frac{(u_0 + v_0 - 1)(u_0 - v_0 + 1)}{u_0}, \\ \left. \frac{\partial S}{\partial v} \right|_{u_0, v_0} &= (1 - \lambda)u_0 + 2\lambda(u_0 + v_0 - 1) = \lambda \frac{(u_0 + v_0 - 1)(v_0 - u_0 + 1)}{v_0}. \end{aligned}$$

Таким образом, уравнение касательной к коническому сечению в точке $r(u_0, v_0)$ имеет вид

$$(u_0 - v_0 + 1)v_0(u - u_0) + (v_0 - u_0 + 1)u_0(v - v_0) = 0. \quad (2.6.6)$$

Она пересекает координатные оси u и v в точках

$$\begin{aligned} u_1 = u_0 + u_0 \frac{v_0 - u_0 + 1}{u_0 - v_0 + 1} &= \frac{2u_0}{u_0 - v_0 + 1}, & v_1 &= 0, \\ v_2 = v_0 + v_0 \frac{u_0 - v_0 + 1}{v_0 - u_0 + 1} &= \frac{2v_0}{v_0 - u_0 + 1}, & u_2 &= 0, \end{aligned}$$

что показано на рис. 2.6.4.

Обозначим отношения длин отрезков, на которые касательная к коническому сечению делит векторы $\mathbf{p}_0 - \mathbf{p}_1$ и $\mathbf{p}_1 - \mathbf{p}_2$, через

$$z_1 = \frac{u_1}{1 - u_1} = \frac{2u_0}{1 - u_0 - v_0}, \quad z_2 = \frac{v_2}{1 - v_2} = \frac{2v_0}{1 - u_0 - v_0}. \quad (2.6.7)$$

Точку $\mathbf{r}(u_0, v_0)$ конического сечения можно характеризовать координатами u_0, v_0 , или координатами u_1, v_2 , или отношениями z_1, z_2 . Остановимся на последнем способе. Выразим u_0 и v_0 через z_1 и z_2 , используя соотношения (2.6.7)

$$u_0 = \frac{z_1}{2 + z_1 + z_2}, \quad v_0 = \frac{z_2}{2 + z_1 + z_2}. \quad (2.6.8)$$

Подставив (2.6.8) в (2.6.3), получим зависимость радиус-вектора конического сечения как функцию параметров z_1 и z_2

$$\mathbf{r}(z_1, z_2) = \frac{z_1 \mathbf{p}_0 + 2\mathbf{p}_1 + z_2 \mathbf{p}_2}{z_1 + 2 + z_2}. \quad (2.6.9)$$

Координаты u_0 и v_0 связаны уравнением (2.6.4), а параметры z_1 и z_2 — более простым соотношением

$$z_1 z_2 = \frac{4u_0 v_0}{(1 - u_0 - v_0)^2} = -\frac{4\lambda}{1 - \lambda}, \quad (2.6.10)$$

т. е. произведение $z_1 z_2 = \text{const}$ для заданного конического сечения. Фактически радиус-вектор конического сечения зависит от одного параметра.

Введем параметр t для конического сечения, от которого будут зависеть координаты u_0 и v_0 и, следовательно, параметры z_1 и z_2 . В соответствии с кривой Безье для точки \mathbf{p}_0 параметр $t = 0$, а для точки \mathbf{p}_2 параметр $t = 1$. При движении по коническому сечению от точки \mathbf{p}_0 к точке \mathbf{p}_2 координата u_0 изменяется от 1 до 0, координата v_0 изменяется от 0 до 1, а параметр t должен изменяться от 0 до 1. Обратим внимание на выражения (2.6.7). Заметим, что из (2.6.4) следует

$$1 - u_0 - v_0 = \sqrt{\frac{u_0 v_0 (\lambda - 1)}{\lambda}}.$$

Подставим это равенство в (2.6.7) и получим

$$z_1 = k \sqrt{\frac{u_0}{v_0}}, \quad z_2 = k \sqrt{\frac{v_0}{u_0}}, \quad (2.6.11)$$

где $k = \sqrt{\frac{4\lambda}{\lambda - 1}} = \frac{\sqrt{4u_0 v_0}}{1 - u_0 - v_0}$ — постоянный коэффициент. В качестве параметра t конического сечения примем величину

$$t = \frac{\sqrt{v_0}}{\sqrt{u_0} + \sqrt{v_0}}, \quad (2.6.12)$$

тогда равенства (2.6.11) примут вид

$$z_1 = k \frac{1 - t}{t}, \quad z_2 = k \frac{t}{1 - t}. \quad (2.6.13)$$

Подставим последние равенства в (2.6.9) и получим зависимость радиус-вектора конического сечения (2.6.3) как функцию одного параметра t

$$\mathbf{r}(t) = \frac{k(1-t)^2 \mathbf{p}_0 + 2t(1-t) \mathbf{p}_1 + kt^2 \mathbf{p}_2}{k(1-t)^2 + 2t(1-t) + kt^2}.$$

Более удобно для использования последнего равенства ввести параметр

$$w = \frac{1}{k} = \sqrt{\frac{\lambda - 1}{4\lambda}} = \frac{1 - u_0 - v_0}{\sqrt{4u_0v_0}} = \frac{1}{\sqrt{z_1 z_2}}, \quad (2.6.14)$$

который называется *весом* точки \mathbf{p}_1 . При использовании веса w формула (2.6.9) примет вид

$$\mathbf{r}(t) = \frac{(1-t)^2 \mathbf{p}_0 + 2t(1-t)w \mathbf{p}_1 + t^2 \mathbf{p}_2}{(1-t)^2 + 2t(1-t)w + t^2}. \quad (2.6.15)$$

Сравним выражение для радиус-вектора кривой Безье второй степени (2.5.8) и выражение для радиус-вектора конического сечения (2.6.15). Они совпадают в случае, когда $w = 1$, что справедливо для одной из кривых семейства конических сечений, которой соответствует $\lambda = -1/3$. Коническое сечение (2.6.4) при $\lambda = -1/3$ описывается уравнением

$$u^2 - 2uv + v^2 - 2u - 2v + 1 = 0,$$

которое является параболой, так как равен нулю его второй инвариант I_2 , составленный из коэффициентов при старших членах переменных u и v . Действительно, если ввести новые переменные $x = u - v$, $y = 2u + 2v - 1$, то они будут связаны параболическим уравнением $y = x^2$. Будем рассматривать выражение

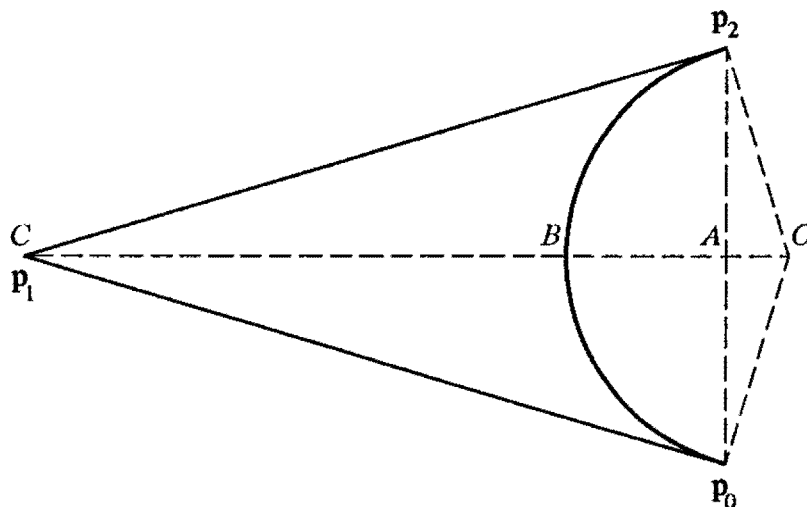


Рис. 2.6.5. Построение дуги окружности

(2.6.15) как обобщение квадратичной параметрической кривой, позволяющее в частных случаях получить как кривую Безье второй степени (при $w = 1$), так и коническое сечение.

Теперь мы можем ответить на поставленный выше вопрос: *кривой Безье второй степени в общем случае нельзя описать часть конического сечения, но его можно описать как частный случай кривой (2.6.15)*. Кривая Безье второй степени также может быть описана как частный случай кривой (2.6.15).

Построим с помощью формулы (2.6.15) дугу окружности радиуса ρ и углом раствора α , показанную на рис. 2.6.5. Через \mathbf{p}_0 и \mathbf{p}_2 обозначены радиус-векторы крайних точек дуги, а через \mathbf{p}_1 обозначена точка пересечения касательных линий к дуге, построенных в крайних ее точках.

Дуга симметрична относительно линии OC и имеет симметричную характеристическую ломаную. Воспользуемся симметрией и найдем коэффициент k по одной из формул (2.6.13) через отношения z_1 и z_2 , в которых касательная к дуге в точке B делит отрезки характеристической ломаной. В точке B параметр $t = 1/2$ и, следовательно, $k = z_1 = z_2$. Из подобия треугольников следует, что

$$k = z_1 = z_2 = \frac{BC}{AB}.$$

Длина отрезков BC и AB вычисляются через радиус ρ и угол раствора дуги α

$$AB = \rho \left(1 - \cos \frac{\alpha}{2}\right),$$

$$BC = CO - BO = \frac{\rho}{\cos(\alpha/2)} - \rho = \rho \frac{1 - \cos(\alpha/2)}{\cos(\alpha/2)}.$$

Тогда

$$w = \frac{1}{k} = \frac{AB}{BC} = \cos \frac{\alpha}{2}.$$

Таким образом, дуга окружности радиуса ρ с углом раствора α , может быть построена как квадратичная кривая (2.6.15), заданная точками p_0 , p_2 и p_1 с весом средней точки $w = \cos(\alpha/2)$. Радиус-вектор дуги описывается функцией

$$\mathbf{r}(t) = \frac{(1-t)^2 \mathbf{p}_0 + 2t(1-t) \cos(\alpha/2) \mathbf{p}_1 + t^2 \mathbf{p}_2}{(1-t)^2 + 2t(1-t) \cos(\alpha/2) + t^2}, \quad 0 \leq t \leq 1, \quad (2.6.16)$$

где точки связаны соотношениями

$$\begin{aligned} |\mathbf{p}_0 - \mathbf{p}_1| &= |\mathbf{p}_2 - \mathbf{p}_1| = \rho \operatorname{tg}(\alpha/2), \\ (\mathbf{p}_0 - \mathbf{p}_1) \cdot (\mathbf{p}_2 - \mathbf{p}_1) &= |\mathbf{p}_0 - \mathbf{p}_1| |\mathbf{p}_2 - \mathbf{p}_1| \cos \alpha. \end{aligned}$$

Параметрическая длина дуги равна единице. Формула (2.6.16) справедлива в диапазоне углов $0 < \alpha < 2\pi$. На рис. 2.6.6 приведена дуга окружности с углом 120° , а на рис. 2.6.7 приведена дуга окружности с углом 240° . Обе дуги имеют

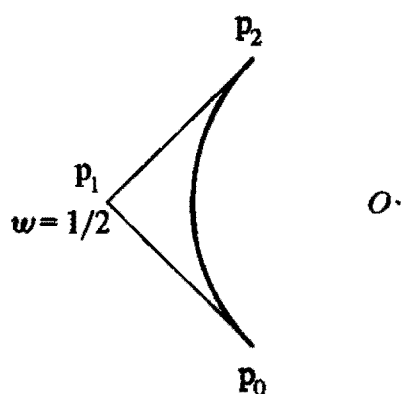


Рис. 2.6.6. Положительный вес точки p_1

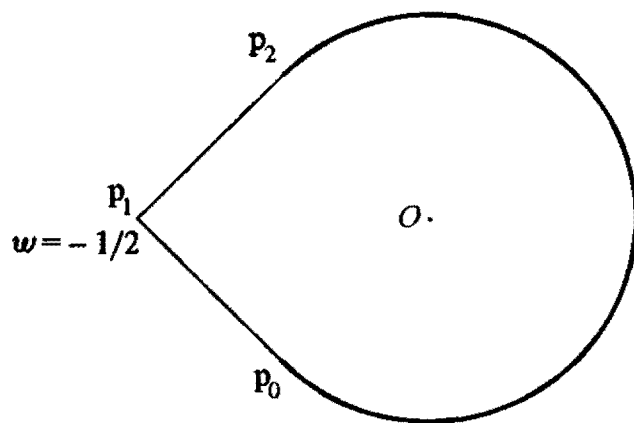


Рис. 2.6.7. Отрицательный вес точки p_1

одни и те же характеристические точки, но в первом случае вес средней точки равен $1/2$, а во втором случае вес средней точки равен $-1/2$. Можно разбить дугу на несколько частей и представить в виде составной кривой, тогда параметрическая длина дуги будет равна числу составных частей.

Построим с помощью формулы (2.6.15) четверть эллипса по трем вершинам \mathbf{p}_0 , \mathbf{p}_1 и \mathbf{p}_2 , образующим прямой угол: $(\mathbf{p}_0 - \mathbf{p}_1) \cdot (\mathbf{p}_2 - \mathbf{p}_1) = 0$, как показано на рис. 2.6.8.

Полуоси эллипса определяются длинами векторов: $a = |\mathbf{p}_0 - \mathbf{p}_1|$, $b = |\mathbf{p}_2 - \mathbf{p}_1|$. Для полного определения эллипса остается найти вес точки \mathbf{p}_1 . В косоугольных координатах u и v , приведенных на рис. 2.6.4, уравнение эллипса имеет вид $(u - 1)^2 + (v - 1)^2 = 1$, которому в (2.6.4) соответствует $\lambda = -1$. В соответствии с (2.6.14) вес средней вершины

$$w = \frac{\sqrt{2}}{2},$$

т. е. такой же, как и у дуги в четверть окружности. В косоугольных координатах уравнение эллипса имеет вид уравнения окружности, поэтому описание дуг эллипса похоже на описание дуг окружности.

Рассмотрим общий случай построения конического сечения, чем дуга окружности. Пусть известны три вершины характеристической ломаной кривой \mathbf{p}_0 ,

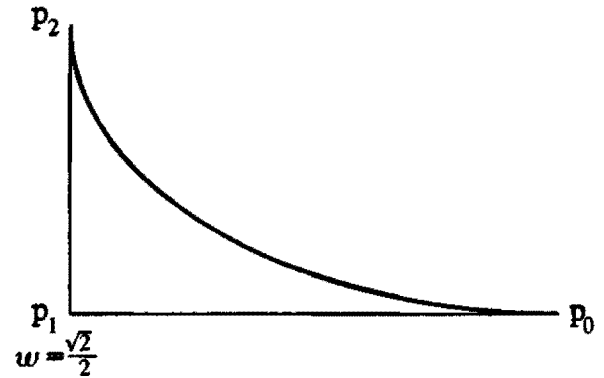


Рис. 2.6.8. Дуга эллипса

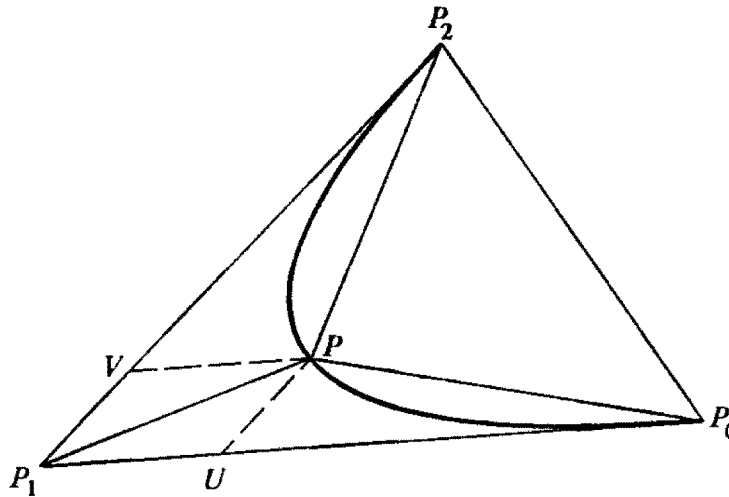


Рис. 2.6.9. Коническое сечение, построенное по четырем точкам

\mathbf{p}_1 , \mathbf{p}_2 и четвертая точка \mathbf{p} , лежащая на кривой. Точка \mathbf{p} должна лежать в одной плоскости с точками \mathbf{p}_0 , \mathbf{p}_1 , \mathbf{p}_2 строго внутри угла, построенного по этим точкам с вершиной в \mathbf{p}_1 . Построим коническое сечение в виде

$$\mathbf{r}(t) = \frac{(1-t)^2 \mathbf{p}_0 + 2t(1-t)w(\mathbf{p})\mathbf{p}_1 + t^2 \mathbf{p}_2}{(1-t)^2 + 2t(1-t)w(\mathbf{p}) + t^2}, \quad 0 \leq t \leq 1. \quad (2.6.17)$$

где $w = w(\mathbf{p})$ — вес вершины \mathbf{p}_1 , который определяется положением четвертой точки \mathbf{p} относительно первых трех. Точки \mathbf{p}_0 , \mathbf{p}_1 , \mathbf{p}_2 и \mathbf{p} на рис. 2.6.9 обозначены через P_0 , P_1 , P_2 и P .

Из (2.6.14) следует, что вес

$$w = \frac{1 - u_0 - v_0}{\sqrt{4u_0v_0}}. \quad (2.6.18)$$

Косоугольные координаты u_0, v_0 точки P , через которые вычисляется вес в (2.6.18), равны отношению длин отрезков VP/P_1P_0 и UP/P_1P_2 , соответственно. Обозначим через L_1 длину отрезка P_1P_0 , через L_2 — длину отрезка P_1P_2 , через l_1 — длину отрезка VP , через l_2 — длину отрезка UP , через S_1 — площадь треугольника P_1P_0P , через S_2 — площадь треугольника P_1P_2P , через S — площадь треугольника $P_1P_0P_2$, через α — угол $P_0P_1P_2$. Тогда

$$S = \frac{L_1 L_2 \sin \alpha}{2}, \quad S_1 = \frac{L_1 l_2 \sin \alpha}{2}, \quad S_2 = \frac{L_2 l_1 \sin \alpha}{2},$$

$$u_0 = \frac{VP}{P_1P_0} = \frac{l_1}{L_1} = \frac{S_2}{S}, \quad v_0 = \frac{UP}{P_1P_2} = \frac{l_2}{L_2} = \frac{S_1}{S}.$$

В принятых обозначениях вес средней точки выразится через площади треугольников

$$w = \frac{S - S_1 - S_2}{\sqrt{4S_1 S_2}}. \quad (2.6.19)$$

Вес w вершины \mathbf{p}_1 можно найти, если $S_1 S_2 \neq 0$, т. е. если точка \mathbf{p} не лежит на линиях P_1P_0 и P_1P_2 (точка \mathbf{p} должна лежать внутри угла $P_0P_1P_2$). В общем случае вес может быть как положительным, так и отрицательным. Как видно из рис. 2.6.6 и 2.6.7, если точка P лежит внутри треугольника $P_0P_1P_2$, то вес точки P_1 положителен, если же точка P лежит вне треугольника $P_0P_1P_2$, то вес точки P_1 отрицателен.

Если мы вычислили вес w средней вершины, то в соответствии с (2.6.14) вычислим параметр $\lambda = \frac{1}{1 - 4w^2}$ конического сечения и конкретный вид уравнения (2.6.4) сечения

$$u^2 + v^2 + 2(1 - 2w^2)uv - 2u - 2v + 1 = 0.$$

Второй инвариант данного конического сечения равен: $I_2 = 4w^2(1 - w^2)$. В зависимости от знака инварианта I_2 конические сечения делятся на три типа: эллиптический тип, если $I_2 > 0$; гиперболический, если $I_2 < 0$; параболический, если $I_2 = 0$. Значение инварианта и соответствующий тип конического не зависят от того, в какой системе координат вычислен инвариант. Таким образом, если $|w| < 1$, то кривая (2.6.17) описывает эллипс, если $|w| > 1$, то кривая описывает гиперболу, если $|w| = 1$, то кривая описывает параболу.

Возможности кривой (2.6.15) можно существенно расширить, если для каждой из вершин ввести вес.

2.7. Рациональные кривые

В формуле (2.6.15) средняя точка \mathbf{p}_1 имеет коэффициент w , называемый *весом*. Для полного равноправия точек $\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2$ в этой формуле припишем каждой из них свой вес w_0, w_1, w_2 . В результате правая часть равенства (2.6.15) примет вид

$$\mathbf{r}(t) = \frac{(1-t)^2 w_0 \mathbf{p}_0 + 2t(1-t) w_1 \mathbf{p}_1 + t^2 w_2 \mathbf{p}_2}{(1-t)^2 w_0 + 2t(1-t) w_1 + t^2 w_2} = \frac{\sum_{i=0}^2 B^2_i w_i \mathbf{p}_i}{\sum_{i=0}^2 B^2_i w_i}. \quad (2.7.1)$$

Три коэффициента w_0, w_1, w_2 придают правой части выражения (2.7.1) симметричный вид — каждой вершине \mathbf{p}_i соответствует свой коэффициент w_i . Будем рассматривать выражение (2.7.1) как обобщение квадратичной параметрической кривой, позволяющее в частных случаях получить как кривую Безье второй степени (при $w_0 = w_1 = w_2$), так и коническое сечение (при $w_0 = w_2 = kw_1$).

Коническое сечение можно описать с помощью кривой более высокой степени, чем это было сделано в (2.6.15). Пусть коническое сечение описывается равенством (2.7.1). Умножим числитель и знаменатель правой части равенства (2.7.1) на равное единице число $t + (1 - t)$, перегруппируем слагаемые по степеням произведений t и $(1 - t)$ и, в результате, придем к выражению

$$\mathbf{r}(t) = \frac{(1-t)^3 w_0 \mathbf{p}_0 + t(1-t)^2 (2w_1 \mathbf{p}_1 + w_0 \mathbf{p}_0) + t^2(1-t)(2w_1 \mathbf{p}_1 + w_2 \mathbf{p}_2) + t^3 w_2 \mathbf{p}_2}{(1-t)^3 w_0 + t(1-t)^2 (2w_1 + w_0) + t^2(1-t)(2w_1 + w_2) + t^3 w_2}.$$

Придадим правой части этого равенства форму кубической кривой Безье. Для этого введем обозначения для весов и вершин в правой части последнего выражения:

$$w_0' = w_0, \quad w_1' = \frac{2w_1 + w_0}{3}, \quad w_2' = \frac{2w_1 + w_2}{3}, \quad w_3' = w_2,$$

$$\mathbf{p}_0' = \mathbf{p}_0, \quad \mathbf{p}_1' = \frac{2w_1 \mathbf{p}_1 + w_0 \mathbf{p}_0}{3w_1'}, \quad \mathbf{p}_2' = \frac{2w_1 \mathbf{p}_1 + w_2 \mathbf{p}_2}{3w_2'}, \quad \mathbf{p}_3' = \mathbf{p}_2.$$

Теперь квадратичная кривая (2.7.1) примет форму кубической кривой (опустим штрих у вершин и их весов)

$$\mathbf{r}(t) = \frac{(1-t)^3 w_0 \mathbf{p}_0 + 3t(1-t)^2 w_1 \mathbf{p}_1 + 3t^2(1-t) w_2 \mathbf{p}_2 + t^3 w_3 \mathbf{p}_3}{(1-t)^3 w_0 + 3t(1-t)^2 w_1 + 3t^2(1-t) w_2 + t^3 w_3} = \frac{\sum_{i=0}^3 B^3_i w_i \mathbf{p}_i}{\sum_{i=0}^3 B^3_i w_i}.$$

Полученная кубическая кривая полностью совпадает с исходной квадратичной кривой, но внутренние вершины ее характеристической ломаной и их веса стали другими. Умножив числитель и знаменатель правой части последнего равенства на неравное нулю число $t + (1 - t)$, перейдем к кривой четвертой степени, которая описывает все то же коническое сечение. Этот процесс можно продолжать и дальше.

Рациональные кривые Безье. Используем полученный результат для обобщения параметрической кривой Безье произвольной степени. Аналогично зависимости (2.7.1) запишем формулу для вычисления радиус-вектора кривой Безье, построенной по произвольному числу n вершин, в виде

$$\mathbf{r}(t) = \frac{\sum_{i=0}^n B^n_i(t) w_i \mathbf{p}_i}{\sum_{i=0}^n B^n_i(t) w_i}, \quad 0 \leq t \leq 1, \quad (2.7.2)$$

где $B^n_i(t)$ — базисные функции Бернштейна. В общем случае каждая вершина \mathbf{p}_i кривой (2.7.2) имеет свой коэффициент w_i . Коэффициент w_i называется *весом* (значимостью) вершины \mathbf{p}_i . Кривые, характеристические точки которых

обладают весами, называются *рациональными*. Рациональные кривые широко применяются в проектировании благодаря Кунсу. Одним из важных свойств рациональных кривых является то, что описанные ими конические сечения сохраняют свою природу (остаются рациональными кривыми) при любом проективном преобразовании. Кривая (2.7.2), построенная на основе базиса Бернштейна, представляет собой *рациональную кривую Безье* степени n . Если веса всех вершин равны, то в силу свойства полиномов Бернштейна (2.5.5) рациональная кривая Безье (2.7.2.) становится равной обычной кривой Безье (2.5.3). В общем случае мы всегда можем перейти от рациональной кривой Безье степени n к рациональной кривой Безье степени $n + 1$, как это было продемонстрировано выше.

Рациональные кривые дают дополнительные возможности при моделировании линий. Изменяя веса точек, мы изменяем форму кривой. На рис. 2.7.1 приведены рациональные кривые Безье пятой степени, построенные по одним

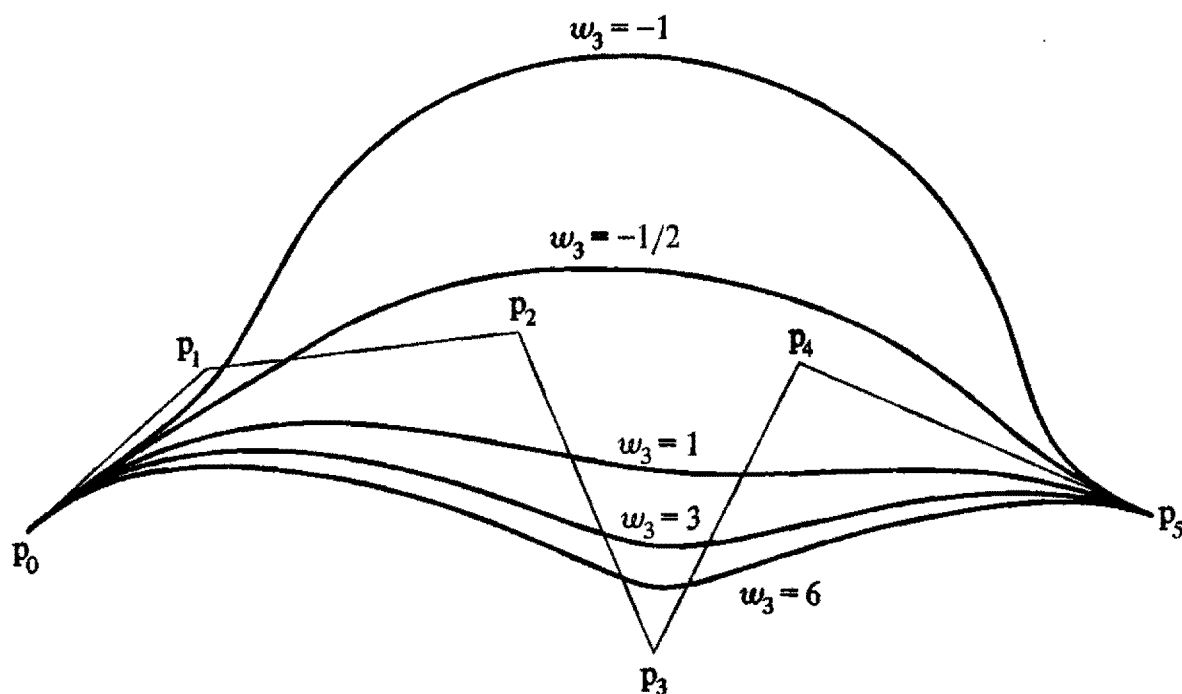


Рис. 2.7.1. Рациональные кривые Безье с различным весом точки P_3

и тем же характеристическим точкам. Веса всех точек, кроме P_3 , равны единице, а вес w_3 точки P_3 приведен на рисунке. Из рисунка видно, что чем больше вес точки, тем ближе она к этой точке. Если вес точки отрицательный, то точка как бы отталкивает кривую от себя.

Если считать, что в знаменателе правой части (2.7.2) стоит вес $w(t)$ радиус-вектора кривой $r(t)$, то его можно считать дополнительной координатой, так как формула его вычисления через веса вершин совпадает с формулой вычисления координат радиус-вектора через координаты вершин. При работе с точками, имеющими вес, удобно пользоваться однородными координатами (1.4.6)

$$R = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{bmatrix} = \begin{bmatrix} wr_1 \\ wr_2 \\ wr_3 \\ w \end{bmatrix} \equiv \begin{bmatrix} wr \\ w \end{bmatrix}, \quad (2.7.3)$$

где r_1, r_2, r_3 — координаты точки r , w — вес точки. В терминах однородных координат радиус-вектор (2.7.3) рациональной параметрической кривой Безье

описывается равенством, по форме совпадающим с равенством для кривой Безье (2.5.3)

$$\mathbf{R}(t) = \sum_{i=0}^n B_i^n(t) \mathbf{P}_i, \quad 0 \leq t \leq 1. \quad (2.7.4)$$

Мы видим, что вес точки подвергается тем же преобразованиям, что и ее декартовы координаты, поэтому с однородными компонентами R_i радиус-вектора можно работать без выделения декартовых координат. Декартовы координаты любой вершины и радиус-вектора точки рациональной кривой получим из однородных координат делением их на последнюю (дополнительную) координату

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

В однородных координатах рациональные кривые имеют более компактную запись, а ее точки могут быть подвержены расширенному матричному преобразованию (1.4.2), обобщающему все преобразования простой точки.

С помощью кривых Безье можно строить линии произвольной формы на любом числе точек. Но, если мы хотим построить кривую на n точках, так чтобы она была заданной степени $m < n - 1$, то нам не удастся обойтись одной кривой. Если к тому же n не кратно m , то не все составные части будут кривыми заданной степени. Если совместить первую и последнюю вершины характеристической ломаной, то получим замкнутую кривую, но в точке замыкания в общем случае кривая будет иметь излом. Кривые Безье не обладают достаточной гибкостью в том смысле, что на n точках можно построить кривую только степени $n - 1$.

Обобщение рациональных кривых. Рассмотрим формулу для рациональной кривой Безье (2.7.2) в более широком смысле, а именно, подставим в нее вместо функций $B_i^n(t)$ некоторые произвольные базисные функции, которые обозначим через $N_i(t)$, и запишем в общем виде формулу для радиус-вектора рациональной кривой, построенной на n вершинах,

$$\mathbf{r}(t) = \frac{\sum_{i=1}^n N_i(t) w_i \mathbf{p}_i}{\sum_{i=1}^n N_i(t) w_i}, \quad t_{\min} \leq t \leq t_{\max}, \quad (2.7.5)$$

где n — число вершин, \mathbf{p}_i — радиус-вектор i -й вершины характеристической ломаной кривой, w_i — вес i -й вершины, $N_i(t)$ — функция параметра кривой, дающая вклад i -й вершины в общую сумму при данном t .

Функции $N_i(t)$ должны позволять интерполировать произвольную функцию $f(t)$ на некотором заданном отрезке $t_1 \leq t \leq t_n$, т.е. для любой функции $f(t)$ должна существовать линейная комбинация

$$\varphi(t) = c_1 N_1(t) + c_2 N_2(t) + \dots + c_n N_n(t),$$

удовлетворяющая равенствам $\varphi(t_i) = f(t_i)$, $i = 1, 2, \dots, n$, где $t_1 < t_2 < \dots < t_n$ — последовательность узлов, c_1, c_2, \dots, c_n — числовые коэффициенты.

Потребуем, чтобы базисные функции $N_i(t)$ обладали свойством функций $B_i^n(t)$, а именно, чтобы при любом t они представляли бы собой разложение

единицы, т. е. удовлетворять равенству

$$\sum_{i=1}^n N_i(t) = 1. \quad (2.7.6)$$

Это требование выдвинуто для того, чтобы в случае равных весов всех вершин рациональная кривая (2.7.5) превращалась в обычную параметрическую кривую вида

$$\mathbf{r}(t) = \sum_{i=1}^n N_i(t) \mathbf{P}_i, \quad t_{\min} \leq t \leq t_{\max}, \quad (2.7.7)$$

которая в частном случае (в зависимости от функций $N_i(t)$) может быть равна или кривой Безье, или векторному полиному Лагранжа (2.4.13), или ломаной линии (2.4.1). Таким образом, рациональную параметрическую кривую (2.7.5) можно было бы считать наиболее общей формой записи параметрической линии, которая в качестве частных случаев содержит в себе почти все рассмотренные выше линии.

Для удобства в кривой (2.7.5) нумерация вершин начинается с 1, а не с 0, как это было у кривых Безье. Заметим, что формула (2.7.5) отличается от (2.7.2) еще и тем, что порядок функции $N_i(t)$ не связан с количества вершин n . Единственная косвенная связь порядка кривой и числа вершин заключается в том, что порядок кривой не может превосходить число вершин. Мы хотим получить более универсальную кривую, чем все рассмотренные выше. Область изменения параметра t кривой (2.7.5) в общем случае может быть выбрана удобной для нас.

В терминах однородных координат радиус-вектор *рациональной параметрической кривой* (2.7.5) описывается равенством

$$\mathbf{R}(t) = \sum_{i=1}^n N_i(t) \mathbf{P}_i, \quad t_{\min} \leq t \leq t_{\max}, \quad (2.7.8)$$

по форме совпадающим с равенством (2.7.7).

Радиус-вектор рациональной кривой вычисляется как частное от деления двух функций параметра кривой t , поэтому при вычислении производной рациональной кривой правую часть (2.7.5) следует рассматривать как сложную функцию. Если условно обозначим радиус-вектор рациональной кривой как $\mathbf{r} = \frac{w\mathbf{r}}{w}$, то производные радиус-вектора рациональной кривой определяются формулами

$$\frac{d\mathbf{r}}{dt} = \frac{1}{w} \frac{d(w\mathbf{r})}{dt} - \frac{w\mathbf{r}}{w^2} \frac{dw}{dt}, \quad (2.7.9)$$

$$\frac{d^2\mathbf{r}}{dt^2} = \frac{1}{w} \frac{d^2(w\mathbf{r})}{dt^2} - \frac{2}{w^2} \frac{d(w\mathbf{r})}{dt} \frac{dw}{dt} - \frac{w\mathbf{r}}{w^2} \frac{d^2w}{dt^2} + \frac{2w\mathbf{r}}{w^3} \left(\frac{dw}{dt} \right)^2, \quad (2.7.10)$$

$$\begin{aligned} \frac{d^3\mathbf{r}}{dt^3} = & \frac{1}{w} \frac{d^3(w\mathbf{r})}{dt^3} - \frac{3}{w^2} \frac{d^2(w\mathbf{r})}{dt^2} \frac{dw}{dt} - \frac{3}{w^2} \frac{d(w\mathbf{r})}{dt} \frac{d^2w}{dt^2} - \frac{w\mathbf{r}}{w^2} \frac{d^3w}{dt^3} + \\ & + \frac{6}{w^3} \frac{d(w\mathbf{r})}{dt} \left(\frac{dw}{dt} \right)^2 + \frac{6w\mathbf{r}}{w^3} \frac{d^2w}{dt^2} \frac{dw}{dt} - \frac{6w\mathbf{r}}{w^4} \left(\frac{dw}{dt} \right)^3. \end{aligned} \quad (2.7.11)$$

Ниже мы найдем непрерывные дифференцируемые функции $N_i(t)$, принимающие ненулевые значения на отрезке $t_i \leq t \leq t_{i+m}$, равные нулю на остальной части параметрической области кривой и удовлетворяющие равенству (2.7.6). Такими функциями являются фундаментальные сплайны или В-сплайны.

2.8. В-сплайны

Поставим перед собой цель найти некоторую непрерывную $m - 1$ раз дифференцируемую функцию $S_m(t)$, принимающую ненулевые значения только внутри отрезка $t_0 \leq t \leq t_m$ и имеющую равные нулю производные от 0-го до $(m - 2)$ -го порядка включительно на его концах. Производной нулевого порядка будем называть значение функции в заданной точке.

Покажем, что теоретически можно построить на отрезке $t_0 \leq t \leq t_m$ кусочно-полиномиальную функцию, удовлетворяющую поставленным условиям. Для этого разобьем отрезок $t_0 \leq t \leq t_m$ оси параметра t на m произвольных участков. Точки разбиения обозначим через $t_1, t_2, t_3, \dots, t_{m-1}$, и построим на каждом отрезке $t_{i-1} \leq t \leq t_i$, $i = 1, 2, \dots, m$, полиномы степени $m - 1$, гладко стыкуя их между собой по производным от 0-го до $(m - 2)$ -го порядка включительно. Вид искомой функции показан на рис. 2.8.1.

Каждый полином имеет m коэффициентов, всего полиномов m . Для их определения можно составить $m - 1$ уравнений в точках стыковки полиномов и $m - 1$ уравнений на каждом из краев. Всего мы получим $(m - 1)(m + 1) = m^2 - 1$ уравнений для m^2 коэффициентов. Недостающее уравнение может нормировать величину прогиба кусочно-полиномиальной линии, например, путем нормирования площади под ней, что дает возможность удовлетворить требованию

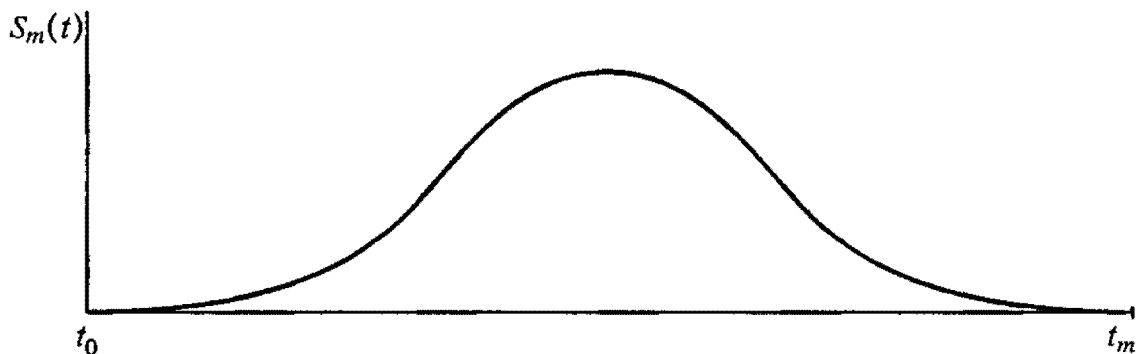


Рис. 2.8.1. Составной полиномиальный сплайн

(2.7.6). Если какие-либо два значения узла совпадают, например $t_i = t_{i+1}$, то полином на этом участке пропадает, а с ним пропадают m неизвестных и $m - 1$ уравнений стыковки. Для сохранения равенства числа неизвестных и числа уравнений откажемся от непрерывности наивысшей производной в кратном узле. Аналогично поступим при совпадении трех и более узлов. Таким образом, предъявленные к функции $S_m(t)$ требования не являются противоречивыми и могут быть выполнены.

Рассмотренные теоретически возможные кусочно-полиномиальные функции, а также базисные функции Бернштейна, являются частными случаями В-сплайнов. Основу теории В-сплайнов заложили Фергюсон, Шёнберг, Уитни и Карри. Гордон и Розенфельд установили связь между кривыми Безье и В-сплайнами и показали, что В-сплайны являются мощным обобщением функций Бернштейна. Теория В-сплайнов базируется на разделенных разностях.

Разделенные разности. Приведем необходимые сведения о разделенных разностях. Пусть имеется некоторая функция $f(t)$ и пусть нам известны значения функции для некоторых значений параметра $f_i = f(t_i)$, $i = 0, 1, 2, \dots, m$. Разделенными разностями первого порядка называются отношения

$$f[t_i, t_{i+1}] = \frac{f(t_{i+1}) - f(t_i)}{t_{i+1} - t_i}. \quad (2.8.1)$$

Точки t_i на параметрической оси называются *узлами*. Будем считать, что разделенные разности составлены по узлам с соседними номерами. Пока не оговорено противное, нумерация узлов может быть произвольной, и необязательно, чтобы узлы были пронумерованы в порядке возрастания значений. По разделенным разностям первого порядка вычисляются разделенные разности второго порядка

$$f[t_i, t_{i+1}, t_{i+2}] = \frac{f[t_{i+1}, t_{i+2}] - f[t_i, t_{i+1}]}{t_{i+2} - t_i}. \quad (2.8.2)$$

Следуя данному алгоритму, вычисляются разделенные разности более высоких порядков. Например, если известны разделенные разности $(k-1)$ -го порядка, то разделенная разность k -го порядка определяется как отношение

$$f[t_i, t_{i+1}, \dots, t_{i+k}] = \frac{f[t_{i+1}, t_{i+2}, \dots, t_{i+k}] - f[t_i, t_{i+1}, \dots, t_{i+k-1}]}{t_{i+k} - t_i}. \quad (2.8.3)$$

Разделенные разности будем обозначать символами функции, по которой они вычислены, с квадратными скобками, в которых перечислены значения узлов. Порядок разделенной разности на единицу меньше числа узлов в квадратных скобках. Разделенными разностями нулевого порядка являются значения функций в узлах $f[t_i] = f(t_i)$. Если функция $f(t)$ есть полином степени k , то как будет показано ниже, разделенная разность k -го порядка, построенная по этому полиному, равна коэффициенту при t^k полинома.

Если в разделенной разности первого порядка значения узлов совпадают, то она равна пределу отношения (2.8.1), т. е. производной функции в данной точке

$$f[t_i, t_i] = \lim_{t \rightarrow t_i} \frac{f(t_i) - f(t)}{t_i - t} = \left. \frac{df}{dt} \right|_{t_i} = f'(t_i). \quad (2.8.4)$$

Аналогично находятся разделенные разности более высокого порядка, построенные на совпадающих узлах. Так разделенная разность k -го порядка, составленная на $k+1$ равных узлах, равна k -й производной функции, вычисленной в данном узле, деленной на $k!$

$$f[t_i, t_i, \dots, t_i] = \frac{1}{k!} \left. \frac{d^k f}{dt^k} \right|_{t_i} = \frac{f^{(k)}(t_i)}{k!} \quad (2.8.5)$$

($k!$ в знаменателе правой части является следствием того, что в знаменателе разделенной разности стоит длина нескольких участков $t_i - t$). Узлы, значения которых совпадают, называются *кратными узлами*. Так, если $t_{i+1} = t_{i+2} = \dots = t_{i+k}$, то говорят, что эти узлы имеют кратность k .

Разделенная разность k -го порядка следующим образом выражается через значения функции $f(t)$ в узлах:

$$\begin{aligned} f[t_i, t_{i+1}, \dots, t_{i+k}] &= \\ &= \sum_{j=i}^{i+k} \frac{f(t_j)}{(t_j - t_i)(t_j - t_{i+1}) \dots (t_j - t_{j-1})(t_j - t_{j+1}) \dots (t_j - t_{j+k})} = \\ &= \sum_{j=i}^{i+k} \frac{f(t_j)}{W_{i,k}'(t_j)}, \quad (2.8.6) \end{aligned}$$

где $W_{i,k}'(t_j)$ — производная функции $W_{i,k}(t) = (t - t_i)(t - t_{i+1})(t - t_{i+2}) \dots (t - t_{i+k-1})(t - t_{i+k})$, вычисленная при $t = t_j$. Индексы i и k в обозначении функции $W_{i,k}(t)$ говорят об индексе начального узла и о количестве узлов. Равенство (2.8.6) докажем методом индукции. Для разделенной разности первого порядка формула (2.8.6) совпадает с (2.8.1), т. е. для $k = 1$ равенство (2.8.6) справедливо. Предположим, что равенство (2.8.6) справедливо для разделенных разностей $(k-1)$ -го порядка и убедимся, что из этого предположения следует его верность для разделенных разностей k -го порядка. Для этого вычтем равенства

$$f[t_{i+1}, t_{i+2}, \dots, t_{i+k}] = \sum_{j=i+1}^{i+k} \frac{(t_j - t_i)f(t_j)}{(t_j - t_i)W_{i+1,k-1}'(t_j)}$$

и

$$f[t_i, t_{i+1}, \dots, t_{i+k-1}] = \sum_{j=i}^{i+k-1} \frac{(t_j - t_{i+k})f(t_j)}{(t_j - t_{i+k})W_{i,k-1}'(t_j)},$$

аналогичные (2.8.6), для разделенных разностей на порядок ниже, у которых соответствующие слагаемые приведены к общему знаменателю. В результате получим равенство

$$\begin{aligned} f[t_{i+1}, t_{i+2}, \dots, t_{i+k}] - f[t_i, t_{i+1}, \dots, t_{i+k-1}] &= \\ &= \frac{(t_{i+k} - t_i)f(t_i)}{W_{i,k}'(t_i)} + \sum_{j=i+1}^{i+k-1} \frac{(t_j - t_i)f(t_j) - (t_j - t_{i+k})f(t_j)}{W_{i,k}'(t_j)} + \\ &\quad + \frac{(t_{i+k} - t_i)f(t_{i+k})}{W_{i,k}'(t_{i+k})} = (t_{i+k} - t_i) \sum_{j=i}^{i+k} \frac{f(t_j)}{W_{i,k}'(t_j)}, \end{aligned}$$

из которого после деления на $(t_{i+k} - t_i)$ на основании (2.8.3) следует (2.8.6). Из равенства (2.8.6) видно, что разделенная разность не зависит от порядка, в котором следуют узлы $t_i, t_{i+1}, \dots, t_{i+k}$ в списке аргументов.

Разделенные разности полиномов. Если $p(t)$ является полиномом m -й степени параметра t и нам известны значения полинома $p(t_i)$ в точках $t_i, i = 0, 1, 2, \dots, m$, то разделенная разность первого порядка

$$p[t, t_0] = \frac{p(t_0) - p(t)}{t_0 - t}$$

$= t_2 = \dots = t_{m-1} = t_m$, то в соответствии с (2.8.5) разделенные разности пропорциональны производным соответствующего порядка интерполируемой функции, и мы получим усеченный ряд Тейлора.

Вычислим с помощью (2.8.10) разделенную разность m -го порядка для m совпадающих узлов $t_0 = t_1 = \dots = t_{m-1}$ и одного отличного от остальных $t_{m-1} \neq t_m$. Все разделенные разности более низкого порядка в соответствии с (2.8.5) будут пропорциональны производным соответствующего порядка функции $f(t)$, а $p(t_m) = f(t_m)$. Искомая разделенная разность определяется формулой

$$\begin{aligned} f[t_0, t_1, \dots, t_m] &= f[t_0, t_0, t_0, \dots, t_0, t_m] = \\ &= \frac{f(t_m)}{(t_m - t_0)^m} - \sum_{j=0}^{m-1} \frac{f^{(j)}(t_0)}{j!(t_m - t_0)^{m-j}}, \end{aligned} \quad (2.8.11)$$

где $f^{(j)}(t_0)$ — производная j -го порядка в точке t_0 (производной нулевого порядка мы называем значение функции).

Свойства разделенных разностей. Известно, что по заданной совокупности значений функции $f(t_i)$ в узлах t_i , $i = 0, 1, 2, \dots, m$, можно построить единственный полином степени m , совпадающий в узлах t_i с заданной функцией. Из интерполяционной формулы Ньютона (2.8.9) видим, что разделенная разность $f[t_0, t_1, \dots, t_m]$ m -го порядка на заданной последовательности $m + 1$ узлов равна коэффициенту при наивысшей степени аргумента t^m полинома степени m , значения которого в заданных узлах согласуются с функцией $f(t)$. Это свойство может быть принято в качестве определения разделенной разности.

Свойство 1. Если $f(t) = a_0 + a_1t + a_2t^2 + \dots + a_mt^m$, то

$$\begin{aligned} f[t_0, t_1, \dots, t_m] &= a_m, \\ f[t_0, t_1, \dots, t_m, t_{m+1}] &= 0 \end{aligned} \quad (2.8.12)$$

для любого дополнительного узла t_{m+1} .

Свойство 2. В силу единственности такого полинома, а также в силу (2.8.6), любая разделенная разность является симметричной функцией своих аргументов, т.е. значение $f[t_0, t_1, \dots, t_m]$ не зависит от порядка, в котором следуют узлы t_0, t_1, \dots, t_m в списке аргументов

$$f[t_0, t_1, t_2, \dots, t_m] = f[t_1, t_0, t_2, \dots, t_m] = \dots = f[t_k, t_i, \dots, t_j], \quad (2.8.13)$$

где t_k, t_i, \dots, t_j есть последовательность узлов t_0, t_1, \dots, t_m , записанных в произвольном порядке. Разделенная разность выражается через значения функции $f(t)$ в узлах с помощью формулы (2.8.6). Если некоторые узлы являются кратными, то разделенная разность выражается через значения функции $f(t)$ в простых узлах и ее производные в кратных узлах.

Свойство 3. Если $f(t) = k_g g(t) + k_h h(t)$, то

$$f[t_0, t_1, \dots, t_m] = k_g g[t_0, t_1, \dots, t_m] + k_h h[t_0, t_1, \dots, t_m]. \quad (2.8.14)$$

Это равенство следует из единственности интерполяционного полинома.

Свойство 4. Если $f(t) = g(t)h(t)$, то

$$f[t_0, t_1, \dots, t_m] = \sum_{k=0}^m g[t_0, t_1, \dots, t_k]h[t_k, t_{k+1}, \dots, t_m]. \quad (2.8.15)$$

Для доказательства этой формулы аппроксимируем каждую из функций $g(t)$ и $h(t)$ полиномом (2.8.10) и перемножим их. Коэффициент при наивысшей степени t^m результирующего полинома будет равен правой части (2.8.15), что и требовалось доказать. Равенство (2.8.15) носит имя *формулы Лейбница*.

Свойство 5. Если для функции $f(t)$ построен полином $p_{m-1}(t)$ степени $m-1$ на последовательности узлов t_0, t_1, \dots, t_{m-1} , то

$$f(t) = p_{m-1}(t) + f[t_0, t_1, \dots, t_{m-1}, t](t-t_0)(t-t_1)\dots(t-t_{m-2})(t-t_{m-1}). \quad (2.8.16)$$

Это равенство получим, рассматривая t как дополнительный узел t_m в последовательности узлов и рассматривая равенство как необходимое условие в дополнительном узле для полинома степени m .

Усеченная степенная функция. Рассмотрим усеченную степенную функцию со смещенным началом, которую обозначим через

$$\sigma_m(z) = (z-t)_+^m \equiv (\max(0, z-t))^m. \quad (2.8.17)$$

Эта функция равна 0 при $z \leq t$ и равна $(z-t)^m$ в противном случае. Функция $\sigma_m(z)$ — кусочно-монотонная, непрерывная при $m > 0$; при $m = 0$ она имеет скачок в точке $z = t$, равный единице. Функция $\sigma_m(z)$ имеет непрерывные производные до $(m-1)$ -го порядка включительно. Производная m -го порядка имеет разрыв в точке $z = t$, равный $m!$. Функция (2.8.17) примечательна тем, что ее производные являются аналогичными функциями, только на порядок ниже. Нам понадобятся как производные по z , так и производные по t :

$$\begin{aligned} \frac{d\sigma_m}{dz} &= m(z-t)^{m-1}_+ = m\sigma_{m-1}, \\ \frac{d\sigma_m}{dt} &= -m(z-t)^{m-1}_+ = -m\sigma_{m-1}, \\ &\dots \end{aligned}$$

$$\frac{d^k \sigma_m}{dz^k} = \frac{m!}{(m-k)!} (z-t)^{m-k}_+ = \frac{m!}{(m-k)!} \sigma_{m-k}, \quad (2.8.18)$$

$$\frac{d^k \sigma_m}{dt^k} = \frac{(-1)^k m!}{(m-k)!} (z-t)^{m-k}_+ = \frac{(-1)^k m!}{(m-k)!} \sigma_{m-k}, \quad (2.8.19)$$

$$k = 0, 1, \dots, m.$$

Графики функций $\sigma_m(z)$ приведены на рис. 2.8.2. Чем выше порядок функции $\sigma_m(z)$, тем она более гладкая в точке $z = t$.

Рассмотрим свойства вычисленных по значениям функции $\sigma_{m-1}(z) = (z-t)^{m-1}_+$ в узлах t_0, t_1, \dots, t_m разделенных разностей m -го порядка $M(t) = \sigma_{m-1}[t_0, t_1, \dots, t_m]$. Разделенная разность вычисляется при фиксированном t

рованном параметре t , но является функцией параметра t . Если все $t_i < t$, $i = 0, 1, \dots, m$, то все $\sigma_{m-1}(t_i) = 0$ и разделенная разность $\sigma_{m-1}[t_0, t_1, \dots, t_m]$ также равна нулю. Если все $t_i > t$, $i = 0, 1, \dots, m$, то разделенная разность m -го порядка равна нулю, так как все узлы находятся на участке, где функция представляет собой полином степени, меньшей m .

Предположим, что последовательность узлов является возрастающей, т. е. $t_i < t_{i+1}$ для всех $i = 0, 1, \dots, m-1$. Тогда по указанным выше причинам разделенная разность $\sigma_{m-1}[t_0, t_1, \dots, t_m]$ и ее производные по t при $t = t_0$ и

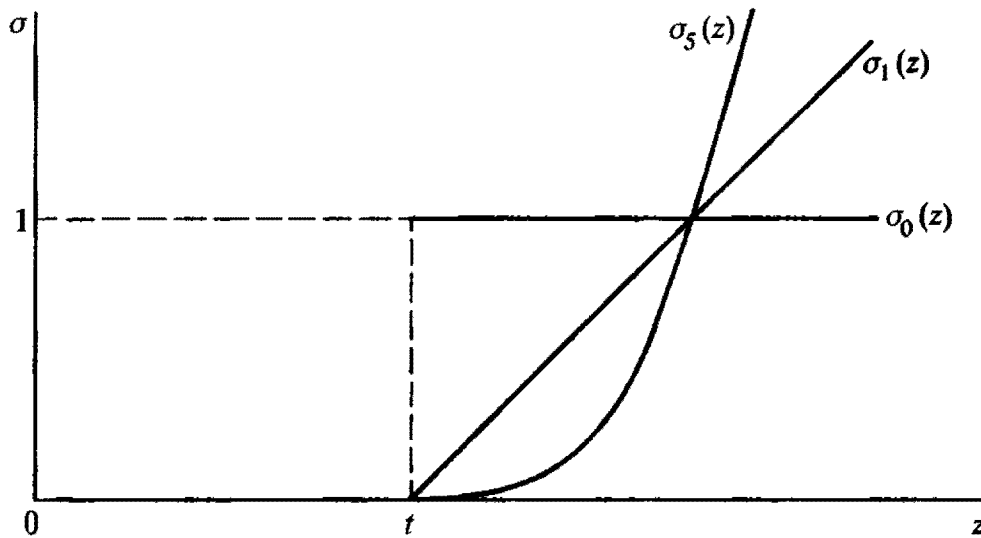


Рис. 2.8.2. Усеченная степенная функция со смещенным началом

$t = t_m$ равны нулю. Разделенная разность $\sigma_{m-1}[t_0, t_1, \dots, t_m]$ как функция параметра t является непрерывной функцией, отличной от нуля только на интервале $t_0 < t < t_m$. В соответствии с (2.8.6) она выражается через значения усеченной степенной функции в узлах

$$\begin{aligned} \sigma_{m-1}[t_0, t_1, \dots, t_m] &= \\ &= \sum_{j=0}^m \frac{(t_j - t)^{m-1} +}{(t_j - t_0)(t_j - t_1) \dots (t_j - t_{j-1})(t_j - t_{j+1}) \dots (t_j - t_m)} = \sum_{j=0}^m \frac{(t_j - t)^{m-1} +}{W_{0,m}'(t_j)}, \end{aligned}$$

где $W_{0,m}'(t_j)$ — производная функции $W_{0,m}(t) = (t - t_0)(t - t_1)(t - t_2) \dots (t - t_{m-1})(t - t_m)$, вычисленная при $t = t_j$. Производные разделенной разности по t до $(m-2)$ -го порядка

$$\begin{aligned} \frac{d^k \sigma_{m-1}[t_0, t_1, \dots, t_m]}{dt^k} &= \frac{(-1)^k (m-1)!}{(m-1-k)!} \sigma_{m-1-k}[t_0, t_1, \dots, t_m] = \\ &= \frac{(-1)^k (m-1)!}{(m-1-k)!} \sum_{j=0}^m \frac{(t_j - t)^{m-1-k} +}{W_{0,m}'(t_j)} \quad (2.8.20) \end{aligned}$$

также являются непрерывными функциями параметра t .

Таким образом, разделенная разность $M(t) = \sigma_{m-1}[t_0, t_1, \dots, t_m]$ как функция параметра t удовлетворяет всем требованиям кроме требования (2.7.6), предъявленным к базисным функциям $N_i(t)$ в формуле (2.7.5), и имеет вид,

приведенный на рис. 2.8.1. Набор функций $M(t)$ является претендентом на роль базисных функций в формуле (2.7.5). Для удовлетворения требованию (2.7.6) функции $M(t)$ можно нормировать, умножив их на некоторый коэффициент.

***B*-сплайн (фундаментальный сплайн).** В литературе используется несколько различных определений и обозначений *B*-сплайна. Вводятся определения ненормированного *B*-сплайна и нормированного *B*-сплайна. В общем случае эти определения звучат так. *Ненормированный B-сплайн для неубывающей последовательности m узлов есть разделенная разность m -го порядка усеченной степенной функции со смещенным началом $\sigma_{m-1}(z) = (z-1)^{m-1}_+$. Нормированный B-сплайн для неубывающей последовательности m узлов есть разделенная разность m -го порядка усеченной степенной функции со смещенным началом $\sigma_{m-1}(z) = (z-t)^{m-1}_+$, умноженная на разность значений последнего и первого узла.* Ненормированный и нормированный *B*-сплайны связаны между собой только коэффициентом и с точностью до множителя их определения одинаковы. Далее, если не будет указано конкретно, термин «*B*-сплайн» мы будем использовать для обоих указанных сплайнов.

В зависимости от того, к какому узлу заданной последовательности (первому или последнему) привязывается обозначение *B*-сплайна, и от того, какой порядок (равный порядку разделенной разности или равный степени усеченной степенной функции) приписывается *B*-сплайну, в литературе встречается четыре определения *B*-сплайна. Приведем их все. Начнем с нормированных *B*-сплайнов, так как именно они представляют для нас наибольший интерес.

Определение 1. Нормированный *i*-й *B*-сплайн m -го порядка $N_{i,m}(t)$ для неубывающей последовательности узлов $t_i, t_{i+1}, t_{i+2}, \dots, t_{i+m}$ есть разделенная разность m -го порядка усеченной степенной функции $\sigma_{m-1}(z) = (z-t)^{m-1}_+$, умноженная на $(t_{i+m} - t_i)$:

$$N_{i,m}(t) = (t_{i+m} - t_i)\sigma_{m-1}[t_i, t_{i+1}, \dots, t_{i+m}]. \quad (2.8.21.1)$$

Определение 2. Нормированный *i*-й *B*-сплайн m -го порядка $N_{m,i}(t)$ для неубывающей последовательности узлов $t_{i-m}, t_{i-m+1}, \dots, t_{i-1}, t_i$ есть разделенная разность m -го порядка усеченной степенной функции $\sigma_{m-1}(z) = (z-t)^{m-1}_+$, умноженная на $(t_i - t_{i-m})$:

$$N_{m,i}(t) = (t_i - t_{i-m})\sigma_{m-1}[t_{i-m}, t_{i-m+1}, \dots, t_i]. \quad (2.8.21.2)$$

Определение 3. Нормированный *i*-й *B*-сплайн $(m-1)$ -го порядка $N_i^{m-1}(t)$ для неубывающей последовательности узлов $t_i, t_{i+1}, t_{i+2}, \dots, t_{i+m}$ есть разделенная разность m -го порядка усеченной степенной функции $\sigma_{m-1}(z) = (z-t)^{m-1}_+$, умноженная на $(t_{i+m} - t_i)$:

$$N_i^{m-1}(t) = (t_{i+m} - t_i)\sigma_{m-1}[t_i, t_{i+1}, \dots, t_{i+m}]. \quad (2.8.21.3)$$

Определение 4. Нормированный *i*-й *B*-сплайн $(m-1)$ -го порядка $N^{m-1}_i(t)$ для неубывающей последовательности узлов $t_{i-m}, t_{i-m+1}, \dots, t_{i-1}, t_i$ есть разделенная разность m -го порядка усеченной степенной функции $\sigma_{m-1}(z) = (z-t)^{m-1}_+$, умноженная на $(t_i - t_{i-m})$:

$$N^{m-1}_i(t) = (t_i - t_{i-m})\sigma_{m-1}[t_{i-m}, t_{i-m+1}, \dots, t_i]. \quad (2.8.21.4)$$

Аналогично звучат определения для соответствующих ненормированных *B*-сплайнов, которые мы рассмотрим ниже. Для приведенных определений мы ввели различные обозначения *B*-сплайнов, чтобы не возникла путаница. В первом определении *B*-сплайн привязан к первому узлу (наследует его индекс) и этот индекс стоит перед индексом m порядка *B*-сплайна. Во втором определении *B*-сплайн привязан к последнему узлу (наследует его индекс) и этот индекс стоит после индекса m порядка *B*-сплайна. Третье определение аналогично первому, только *B*-сплайну приписывается порядок, равный степени усеченной функции, который оказывается на единицу меньше порядка разделенной разности. Для обращения внимания на это отличие мы обозначили порядок *B*-сплайна верхним индексом. Четвертое определение аналогично второму, только *B*-сплайну приписывается порядок, равный степени усеченной функции, который на единицу меньше порядка разделенной разности. Как и в третьем определении, мы обозначили порядок *B*-сплайна верхним индексом. В третьем определении *B*-сплайн привязан к первому узлу, а в четвертом определении *B*-сплайн привязан к последнему узлу (наследует его индекс). Соответственно, в третьем определении индекс узла привязки стоит перед индексом, обозначающим порядок *B*-сплайна, а в четвертом определении индекс узла привязки стоит после индекса порядка *B*-сплайна. Все четыре определения равноправны и приводят к одному и тому же рекуррентному соотношению для *B*-сплайнов, но форма записи этих рекуррентных соотношений различна.

Дадим пояснения к определениям *B*-сплайна на примере первого определения. Последовательность узлов $t_i, t_{i+1}, \dots, t_{i+m}$ должна быть неубывающей, т. е. должно выполняться неравенство $t_j \leq t_{j+1}$ для всех $j = i, i+1, \dots, i+m-1$. В остальном же значения узлов последовательности не оговариваются и в общем случае могут быть произвольными, в том числе и совпадающими. То, что *B*-сплайн $N_{i,m}(t)$ называется *i*-м, подчеркивает, что разделенная разность построена на последовательности узлов, начинающейся с узла t_i и кончающейся на узле t_{i+m} . Разделенная разность $\sigma_{m-1}[t_i, t_{i+1}, \dots, t_{i+m}]$ вычислена по функции $\sigma_{m-1}(z) = (z - t)_+^{m-1}$ при фиксированном значении t , при рассмотрении ее как функции только аргумента z . Значение разделенной разности, естественно, зависит от выбранного параметра t , о чем говорит аргумент $N_{i,m}(t)$. *B*-сплайн $N_{i,m}(t)$ нормирован коэффициентом $(t_{i+m} - t_i)$, на который умножена разделенная разность.

При втором определении *B*-сплайна последовательность узлов $t_{i-m}, t_{i-m+1}, \dots, t_i$ также должна быть неубывающей. То, что *B*-сплайн $N_{m,i}(t)$ называется *i*-м, подчеркивает, что разделенная разность построена на последовательности узлов, начинающейся с узла t_{i-m} и кончающейся на узле t_i . Разделенная разность $\sigma_{m-1}[t_{i-m}, t_{i-m+1}, \dots, t_i]$ вычислена по функции $\sigma_{m-1}(z) = (z - t)_+^{m-1}$ при фиксированном значении t , рассмотренная как функция только аргумента z . Значение разделенной разности зависит от выбранного параметра t , о чем говорит аргумент $N_{m,i}(t)$. *B*-сплайн $N_{m,i}(t)$ нормирован коэффициентом $(t_i - t_{i-m})$, на который умножена разделенная разность.

Далее мы будем строить кривые $\mathbf{r}(t)$ на базе точек \mathbf{p}_i , в которых индекс узла привязки *B*-сплайна будет соответствовать (равен) индексу точки \mathbf{p}_i . Так как индексация точек обычно начинается с нуля или единицы, то при использовании второго или четвертого определений *B*-сплайна индексацию узлов последовательности придется начинать с отрицательных значений так, чтобы индекс

m -го по порядку узла был равен нулю, тогда индекс первого по порядку узла будет равен $-m$.

Гладкость B -сплайнов характеризуется порядком, который в первом и втором определениях равен порядку разделенной разности. В первых двух определениях приходится различать степень полиномов и порядок сплайнов. Порядок сплайна, как и степень полинома, характеризует непрерывность ее производных. Для сплайна порядка m должна существовать и быть отличной от нуля (за исключением отдельных точек) производная $(m-1)$ -го порядка, а производная m -го порядка должна быть равна нулю. В первых двух определениях B -сплайны порядка m представляют собой кусочно-полиномиальные функции, построенные на полиномах степени $m-1$, поэтому их порядок на единицу больше наивысшей степени полинома. В третьем и четвертом определениях B -сплайны порядка m представляют собой кусочно-полиномиальные функции, построенные на полиномах той же степени m , поэтому их порядок равен наивысшей степени полинома.

Дальнейшие выкладки выполним одновременно для первого и второго определений нормированного B -сплайна. Для третьего и четвертого определений все формулы могут быть получены аналогично.

Пусть B -сплайн m -го порядка построен на неубывающей последовательности $m+1$ узлов $t_a, t_{a+1}, \dots, t_{b-1}, t_b$, где $t_b \equiv t_{a+m}$. B -сплайн отличен от нуля на небольшой области, т. е. $N_{a,m}(t) = 0$ и $N_{m,b}(t) = 0$ при $t < t_a$ и при $t \geq t_b$. Действительно, как только $t < t_a$, то между точками t_a и t_b усеченная функция примет вид $\sigma_{m-1}(z) = (z-t)^{m-1}$ и в силу того, что $(z-t)^{m-1}$ является полиномом степени меньшей m , любая разделенная разность m -го порядка, построенная по этому полиному, равна нулю. Как только $t \geq t_b$, то разделенная разность равна нулю в силу того, что между точками t_a и t_b $\sigma_{m-1}(z) = 0$. Только на участке $t_a \leq t < t_b$ функция $\sigma_{m-1}(z)$ представляет собой отрезок горизонтальной прямой, сопряженный с отрезком смещенной степенной функции, и поэтому ее разделенная разность $\sigma_{m-1}[t_a, t_{a+1}, \dots, t_b]$ отлична от нуля.

Если воспользоваться определением разделенной разности (2.8.4), то B -сплайн можно записать в виде

$$\begin{aligned} N_{a,m}(t) &= \sigma_{m-1}[t_{a+1}, t_{a+2}, \dots, t_b] - \sigma_{m-1}[t_a, t_{a+1}, \dots, t_{b-1}], \\ N_{m,b}(t) &= \sigma_{m-1}[t_{a+1}, t_{a+2}, \dots, t_b] - \sigma_{m-1}[t_a, t_{a+1}, \dots, t_{b-1}]. \end{aligned} \quad (2.8.22)$$

Для любого $t_j \leq t < t_{j+1}$, в силу локальности отличны от нуля только m B -сплайнов, а именно, $N_{j-m+1,m}(t), N_{j-m+2,m}(t), \dots, N_{j-1,m}(t), N_{j,m}(t)$ и $N_{m,j+1}(t), N_{m,j+2}(t), \dots, N_{m,j+m-1}(t), N_{m,j+m}(t)$. Используя равенство (2.8.22), найдем сумму всех B -сплайнов для $t_j \leq t < t_{j+1}$:

$$\begin{aligned} \sum_{i=j-m+1}^j N_{i,m}(t) &= \sum_{i=j-m+1}^j N_{m,i+m}(t) = \\ &= \sum_{i=j-m+1}^j \sigma_{m-1}[t_{i+1}, t_{i+2}, \dots, t_{i+m}] - \sum_{i=j-m+1}^j \sigma_{m-1}[t_i, t_{i+1}, \dots, t_{i+m-1}] = \\ &= \sigma_{m-1}[t_{j+1}, t_{j+2}, \dots, t_{j+m}] - \sigma_{m-1}[t_{j-m+1}, t_{j-m+2}, \dots, t_j] = 1 - 0 = 1. \end{aligned} \quad (2.8.23)$$

В (2.8.23) использовалось то, что $\sigma_{m-1}[t_{j+1}, t_{j+2}, \dots, t_{j+m}] = 1$ как коэффициент при наивысшей степени аргумента полинома $(z - t)^{m-1}$, которым описывается усеченная степенная функция на открытом интервале $t_{j+1} < z < t_{j+m}$. При этом на интервале $t_{j-m+1} < z < t_j$ усеченная степенная функция равна нулю, следовательно, $\sigma_{m-1}[t_{j-m+1}, t_{j-m+2}, \dots, t_j] = 0$. Так как участок $t_j \leq t < t_{j+1}$ был выбран произвольно, то аналогичное равенство выполняется для любого другого участка. Таким образом, при любом t для суммы всех *B*-сплайнов справедливо равенство

$$\sum_i N_{i,m}(t) = \sum_i N_{m,i}(t) = 1. \quad (2.8.24)$$

Именно ради свойства (2.8.24) нормируются разделенные разности в определении нормированного *B*-сплайна. Функции $N_{i,m}(t)$ и $N_{m,i}(t)$ представляют собой разложение единицы.

Вычисление нормированных *B*-сплайнов производится через ненормированные *B*-сплайны, являющиеся обычными разделенными разностями. Обратимся к ненормированным *B*-сплайнам. Введем четыре различных определения ненормированного *B*-сплайна. Они соответствуют приведенным выше четырем определениям нормированного *B*-сплайна.

Определение 1. Ненормированный i -й *B*-сплайн m -го порядка $M_{i,m}(t)$ для последовательности узлов $t_i, t_{i+1}, t_{i+2}, \dots, t_{i+m}$ есть разделенная разность m -го порядка усеченной степенной функции $\sigma_{m-1}(z) = (z - t)^{m-1}_+$:

$$M_{i,m}(t) = \sigma_{m-1}[t_i, t_{i+1}, \dots, t_{i+m}]. \quad (2.8.25.1)$$

Определение 2. Ненормированный i -й *B*-сплайн m -го порядка $M_{m,i}(t)$ для последовательности узлов $t_{i-m}, t_{i-m+1}, \dots, t_{i-1}, t_i$ есть разделенная разность m -го порядка усеченной степенной функции $\sigma_{m-1}(z) = (z - t)^{m-1}_+$:

$$M_{m,i}(t) = \sigma_{m-1}[t_{i-m}, t_{i-m+1}, \dots, t_i]. \quad (2.8.25.2)$$

Определение 3. Ненормированный i -й *B*-сплайн $(m - 1)$ -го порядка $M_i^{m-1}(t)$ для последовательности узлов $t_i, t_{i+1}, t_{i+2}, \dots, t_{i+m}$ есть разделенная разность m -го порядка усеченной степенной функции $\sigma_{m-1}(z) = (z - t)^{m-1}_+$:

$$M_i^{m-1}(t) = \sigma_{m-1}[t_i, t_{i+1}, \dots, t_{i+m}]. \quad (2.8.25.3)$$

Определение 4. Ненормированный i -й *B*-сплайн $(m - 1)$ -го порядка $M^{m-1}_i(t)$ для последовательности узлов $t_{i-m}, t_{i-m+1}, \dots, t_{i-1}, t_i$ есть разделенная разность m -го порядка усеченной степенной функции $\sigma_{m-1}(z) = (z - t)^{m-1}_+$:

$$M^{m-1}_i(t) = \sigma_{m-1}[t_{i-m}, t_{i-m+1}, \dots, t_i]. \quad (2.8.25.4)$$

Индексные обозначения ненормированного *B*-сплайна соответствуют индексным обозначениям нормированного *B*-сплайна. Все определения ненормированного *B*-сплайна равноценны. Рассмотрим свойства ненормированного *B*-сплайна.

Ненормированный B -сплайн первого порядка для первых двух определений и ненормированный B -сплайн нулевого порядка для последних двух определений равен разделенной разности первого порядка для функции $\sigma_0(z) = (z - t)_+^0$:

$$M_{i,1}(t) = \begin{cases} \frac{1}{t_{i+1} - t_i}, & \text{если } t_i \leq t < t_{i+1}, \\ 0, & \text{если } t < t_i \text{ или } t \geq t_{i+1}, \end{cases} \quad (2.8.26.1)$$

$$M_{1,i}(t) = \begin{cases} \frac{1}{t_i - t_{i-1}}, & \text{если } t_{i-1} \leq t < t_i, \\ 0, & \text{если } t < t_{i-1} \text{ или } t \geq t_i, \end{cases} \quad (2.8.26.2)$$

$$M_i^0(t) = \begin{cases} \frac{1}{t_{i+1} - t_i}, & \text{если } t_i \leq t < t_{i+1}, \\ 0, & \text{если } t < t_i \text{ или } t \geq t_{i+1}, \end{cases} \quad (2.8.26.3)$$

$$M_i^0(t) = \begin{cases} \frac{1}{t_i - t_{i-1}}, & \text{если } t_{i-1} \leq t < t_i, \\ 0, & \text{если } t < t_{i-1} \text{ или } t \geq t_i. \end{cases} \quad (2.8.26.4)$$

Формула Кокса-Де Бура. Пусть имеется неубывающая последовательность $m+1$ узлов $t_a \equiv t_{b-m}, t_{a+1}, \dots, t_{b-1}, t_b \equiv t_{a+m}$. Используем формулу Лейбница (2.8.15) для вычисления B -сплайна (2.8.21), представив функцию $\sigma_{m-1}(z) = (z - t)_+^{m-1}$ в виде произведения $\sigma_{m-1}(z) = g(z) \cdot \sigma_{m-2}(z)$, где $\sigma_{m-2}(z) = (z - t)_+^{m-2}$ и $g(z) = (z - t)$. Тогда при $t_a \leq t \leq t_b$ разделенная разность

$$\begin{aligned} \sigma_{m-1}[t_a, t_{a+1}, \dots, t_{b-1}, t_b] &= g[t_a] \sigma_{m-2}[t_a, t_{a+1}, \dots, t_b] + \\ &+ g[t_a, t_{a+1}] \sigma_{m-2}[t_{a+1}, t_{a+2}, \dots, t_b] + 0 + \dots + 0 = \\ &= (t_a - t) \sigma_{m-2}[t_a, t_{a+1}, \dots, t_b] + \sigma_{m-2}[t_{a+1}, t_{a+2}, \dots, t_b]. \end{aligned} \quad (2.8.27)$$

Подставим в (2.8.27) определение разделенной разности (2.8.3) для

$$\sigma_{m-2}[t_a, t_{a+1}, \dots, t_b] = \frac{\sigma_{m-2}[t_{a+1}, t_{a+2}, \dots, t_b] - \sigma_{m-2}[t_a, t_{a+1}, \dots, t_{b-1}]}{t_b - t_a}$$

и получим

$$\begin{aligned} \sigma_{m-1}[t_a, t_{a+1}, \dots, t_{b-1}, t_b] &= \\ &= \frac{t_a - t}{t_b - t_a} (\sigma_{m-2}[t_{a+1}, t_{a+2}, \dots, t_b] - \sigma_{m-2}[t_a, t_{a+1}, \dots, t_{b-1}]) + \\ &+ \frac{t_b - t_a}{t_b - t_a} \sigma_{m-2}[t_{a+1}, t_{a+2}, \dots, t_b] = \\ &= \frac{(t_b - t) \sigma_{m-2}[t_{a+1}, t_{a+2}, \dots, t_b] + (t - t_a) \sigma_{m-2}[t_a, t_{a+1}, \dots, t_{b-1}]}{t_b - t_a}. \end{aligned} \quad (2.8.28)$$

Перепишем равенство (2.8.28) с использованием определений ненормированного В-сплайна (2.8.25.1)–(2.8.25.4) и получим формулы

$$M_{a,m}(t) = \frac{(t_b - t)M_{a+1,m-1}(t) + (t - t_a)M_{a,m-1}(t)}{t_b - t_a}, \quad (2.8.29.1)$$

$$M_{m,b}(t) = \frac{(t_b - t)M_{m-1,b}(t) + (t - t_a)M_{m-1,b-1}(t)}{t_b - t_a}, \quad (2.8.29.2)$$

$$M_a^{m-1}(t) = \frac{(t_b - t)M_{a+1}^{m-2}(t) + (t - t_a)M_a^{m-2}(t)}{t_b - t_a}, \quad (2.8.29.3)$$

$$M_b^{m-1}(t) = \frac{(t_b - t)M_b^{m-2}(t) + (t - t_a)M_{b-1}^{m-2}(t)}{t_b - t_a}. \quad (2.8.29.4)$$

Рекуррентные соотношения (2.8.29.1)–(2.8.29.4) были получены независимо Мэнсфилдом, Коксом и Де Буrom и носят имя *формул Кокса–Де Бура*. Эти соотношения занимают центральное место в теории В-сплайнов. Они побуждают забыть о разделенных разностях и определить В-сплайн m -го или $(m - 1)$ -го порядка для неубывающей последовательности $m + 1$ узлов $t_a \equiv t_{b-1}, t_{a+1}, \dots, t_{b-1}, t_b \equiv t_{a+m}$ как функцию, вычисляемую по одному из рекуррентных соотношений (2.8.29.1)–(2.8.29.4) при соответствующем начальном значении (2.8.26.1)–(2.8.26.4). Все рекуррентные формулы равноценны.

Обратим внимание на то, что в формулах (2.8.29.1) и (2.8.29.2) В-сплайн m -го порядка для неубывающей последовательности $m + 1$ узлов вычисляется через два В-сплайна $(m - 1)$ -го порядка, определенных на последовательностях m узлов, лежащих внутри заданной последовательности. Один из двух В-сплайнов $(m - 1)$ -го порядка примыкает к началу вычисляемого В-сплайна, а другой — к концу вычисляемого В-сплайна (рис. 2.8.3). Каждый из двух

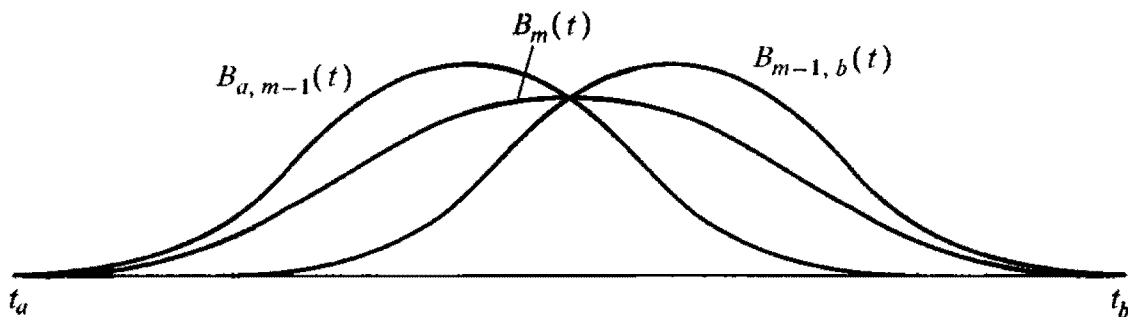


Рис. 2.8.3. Иллюстрация рекуррентного соотношения Кокса–Де Бура

В-сплайнов $(m - 1)$ -го порядка берется с коэффициентом, пропорциональным расстоянию параметра t от крайнего узла заданной последовательности, к которому В-сплайнов $(m - 1)$ -го порядка примыкает. Сумма этих двух коэффициентов равна единице.

Отвлечемся в формулах (2.8.29.1)–(2.8.29.4) от индексации и рассмотрим их с использованием барицентрических координат. Пусть дана произвольная последовательность $m + 1$ узлов t_0, t_1, \dots, t_m , хотя бы два из которых имели отличные друг от друга значения. Вспомним, что разделенная разность не зависит от порядка следования узлов в списке аргументов. Из последовательности $m + 1$ узлов выберем два произвольных несовпадающих узла t_a и t_b . На базе

этих двух узлов введем барицентрические координаты

$$a(t) = \frac{t_b - t}{t_b - t_a}, \quad b(t) = \frac{t_a - t}{t_a - t_b}.$$

Введем следующие обозначения. Последовательность $m + 1$ узлов t_0, t_1, \dots, t_m обозначим через T . Последовательность m узлов, полученную из T удалением узла t_a , обозначим через $T \setminus a$. Последовательность m узлов, полученную из T удалением узла t_b , обозначим через $T \setminus b$. Разделенную разность m -го порядка усеченной степенной функции $(z - t)^{m-1}_+$ на последовательности узлов T обозначим через $M^T(t)$. Разделенную разность $(m - 1)$ -го порядка усеченной степенной функции $(z - t)^{m-2}_+$ на последовательности узлов $T \setminus a$ обозначим через $M^{T \setminus a}(t)$. Разделенную разность $(m - 1)$ -го порядка усеченной степенной функции $(z - t)^{m-2}_+$ на последовательности узлов $T \setminus b$ обозначим через $M^{T \setminus b}(t)$. Тогда формулы Кокса-Де Бура (2.8.29.1)–(2.8.29.4) для вычисления ненормированного B -сплайна на последовательности узлов T можно заменить рекуррентным соотношением

$$M^T(t) = \frac{t_b - t}{t_b - t_a} M^{T \setminus a}(t) + \frac{t_a - t}{t_a - t_b} M^{T \setminus b}(t) = a(t) M^{T \setminus a}(t) + b(t) M^{T \setminus b}(t). \quad (2.8.30)$$

Так как в силу свойства 2 разделенная разность является симметричной функцией своих аргументов, то последнее рекуррентное соотношение не зависит от выбора узлов t_a и t_b . Рекуррентное соотношение (2.8.30) начинается с последовательности, состоящей из двух узлов t_a и t_b :

$$M(t) = \begin{cases} \frac{1}{|t_b - t_a|}, & \text{если } \min(t_a, t_b) \leq t < \max(t_a, t_b), \\ 0 & \text{в остальных случаях.} \end{cases}$$

Последовательность узлов T в общем случае может быть произвольной. Мы рассматриваем неубывающую последовательность узлов, так как нас интересует не отдельно взятый B -сплайн, а совокупность отличных от нуля B -сплайнов при заданном параметре t .

Рекуррентные соотношения (2.8.29.1)–(2.8.29.4) показывают, что если $M_{a+1, m-1}(t) \geq 0$ и $M_{a, m-1}(t) \geq 0$, то и $M_{a, m}(t) \geq 0$, и что если $M_{m-1, b}(t) \geq 0$ и $M_{m-1, b-1}(t) \geq 0$, то и $M_{m, b}(t) \geq 0$. Так как в (2.8.26.1) $M_{i, 1}(t) \geq 0$ и в (2.8.26.2) $M_{1, i}(t) \geq 0$, то можно сделать вывод, что B -сплайны при любых значениях параметра неотрицательные:

$$M_{a, m}(t) \geq 0, \quad M_{m, b}(t) \geq 0, \quad M_a^m(t) \geq 0, \quad M_b^m(t) \geq 0.$$

Нормированный B -сплайн связан с ненормированным B -сплайном соответствующим соотношением:

$$N_{a, m}(t) = (t_{a+m} - t_a) M_{a, m}(t), \quad (2.8.31.1)$$

$$N_{m, b}(t) = (t_b - t_{b-m}) M_{m, b}(t), \quad (2.8.31.2)$$

$$N_a^{m-1}(t) = (t_{a+m} - t_a) M_a^{m-1}(t), \quad (2.8.31.3)$$

$$N_b^{m-1}(t) = (t_b - t_{b-m}) M_b^{m-1}(t). \quad (2.8.31.4)$$

Тогда для нормированного В-сплайна формула Кокса-Де Бура примет вид

$$N_{a,m}(t) = \frac{t_b - t}{t_b - t_{a+1}} N_{a+1,m-1}(t) + \frac{t - t_a}{t_{b-1} - t_a} N_{a,m-1}(t), \quad (2.8.32.1)$$

$$N_{m,b}(t) = \frac{t_b - t}{t_b - t_{a+1}} N_{m-1,b}(t) + \frac{t - t_a}{t_{b-1} - t_a} N_{m-1,b-1}(t), \quad (2.8.32.2)$$

$$N_a^{m-1}(t) = \frac{t_b - t}{t_b - t_{a+1}} N_{a+1}^{m-2}(t) + \frac{t - t_a}{t_{b-1} - t_a} N_a^{m-2}(t), \quad (2.8.32.3)$$

$$N_b^{m-1}(t) = \frac{t_b - t}{t_b - t_{a+1}} N_b^{m-2}(t) + \frac{t - t_a}{t_{b-1} - t_a} N_{b-1}^{m-2}(t), \quad (2.8.32.4)$$

где индексы a и b связаны равенством $b = a + m$ или $a = b - m$. Вычисления начинаются с В-сплайна наименьшего порядка:

$$N_{i,1}(t) = \begin{cases} 1, & \text{если } t_i \leq t < t_{i+1}, \\ 0, & \text{если } t < t_i \text{ или } t \geq t_{i+1}, \end{cases} \quad (2.8.33.1)$$

$$N_{1,i}(t) = \begin{cases} 1, & \text{если } t_{i-1} \leq t < t_i, \\ 0, & \text{если } t < t_{i-1} \text{ или } t \geq t_i, \end{cases} \quad (2.8.33.2)$$

$$N_i^0(t) = \begin{cases} 1, & \text{если } t_i \leq t < t_{i+1}, \\ 0, & \text{если } t < t_i \text{ или } t \geq t_{i+1}, \end{cases} \quad (2.8.33.3)$$

$$N_i^0(t) = \begin{cases} 1, & \text{если } t_{i-1} \leq t < t_i, \\ 0, & \text{если } t < t_{i-1} \text{ или } t \geq t_i. \end{cases} \quad (2.8.33.4)$$

Если формально представить схему вычисления В-сплайна определения 1, то она выглядит следующим образом (порядок В-сплайна возрастает сверху вниз, а номера узлов возрастают слева направо):

$$\begin{array}{cccccc} N_{i,1}(t) & N_{i+1,1}(t) & \dots & N_{i+m-3,1}(t) & N_{i+m-2,1}(t) & N_{i+m-1,1}(t) \\ N_{i,2}(t) & N_{i+1,2}(t) & \dots & N_{i+m-3,2}(t) & N_{i+m-2,2}(t) & \\ N_{i,3}(t) & N_{i+1,3}(t) & \dots & N_{i+m-3,3}(t) & & \\ \dots & \dots & \dots & \dots & \dots & \dots \\ N_{i,m-1}(t) & N_{i+1,m-1}(t) & & & & \\ N_{i,m}(t) & & & & & \end{array}$$

Среди В-сплайнов первого порядка (2.8.26.1) только один отличен от нуля при данном параметре t , а для вычисления точки кривой (2.7.5) требуются все В-сплайны, отличные от нуля при данном параметре t . Поэтому для параметра $t_i \leq t < t_{i+1}$ вычисление всех m отличных от нуля В-сплайнов m -го порядка удобнее производить по схеме

$$\begin{array}{cccccc} & & & & & N_{i,1}(t) \\ & & & & N_{i-1,2}(t) & N_{i,2}(t) \\ & & & N_{i-2,3}(t) & N_{i-1,3}(t) & N_{i,3}(t) \\ & & \dots & \dots & \dots & \dots \\ & N_{i-m+2,m-1}(t) & \dots & N_{i-2,m-1}(t) & N_{i-1,m-1}(t) & N_{i,m-1}(t) \\ N_{i-m+1,m}(t) & N_{i-m+2,m}(t) & \dots & N_{i-2,m}(t) & N_{i-1,m}(t) & N_{i,m}(t) \end{array}$$

Эту треугольную таблицу можно вычислять построчно, полагая неупомянутые крайние соседние B -сплайны равными нулю. Вычислительные затраты могут быть меньше, если работать с ненормированными B -сплайнами $M_{i,m}(t)$, нормируя их в последней строке.

Схема вычисления B -сплайна определения 2 выглядит следующим образом (порядок B -сплайна возрастает сверху вниз, а номера узлов возрастают слева направо):

$$\begin{array}{cccccc}
 N_{1,i-m+1}(t) & N_{1,i-m+2}(t) & \dots & N_{1,i-2}(t) & N_{1,i-1}(t) & N_{1,i}(t) \\
 & N_{2,i-m+2}(t) & \dots & N_{2,i-2}(t) & N_{2,i-1}(t) & N_{2,i}(t) \\
 & & \dots & & & \dots \\
 & & & N_{m-2,i-2}(t) & N_{m-2,i-1}(t) & N_{m-2,i}(t) \\
 & & & & N_{m-1,i-1}(t) & N_{m-1,i}(t) \\
 & & & & & N_{m,i}(t)
 \end{array}$$

Среди B -сплайнов первого порядка (2.8.26.2) только один отличен от нуля при данном параметре t . Поэтому для параметра $t_{i-1} \leq t < t_i$ вычисление всех m отличных от нуля B -сплайнов m -го порядка при данном параметре t удобнее производить по схеме

$$\begin{array}{cccccc}
 N_{1,i}(t) & & & & & \\
 N_{2,i}(t) & N_{2,i+1}(t) & & & & \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 N_{m-2,i}(t) & N_{m-2,i+1}(t) & \dots & N_{m-2,i+m-3}(t) & & \\
 N_{m-1,i}(t) & N_{m-1,i+1}(t) & \dots & N_{m-1,i+m-3}(t) & N_{m-1,i+m-2}(t) & \\
 N_{m,i}(t) & N_{m,i+1}(t) & \dots & N_{m,i+m-3}(t) & N_{m,i+m-2}(t) & N_{m,i+m-1}(t)
 \end{array}$$

Эту треугольную таблицу можно также вычислять построчно, полагая неупомянутые крайние соседние B -сплайны равными нулю. Вычисления проще выполнять для ненормированных B -сплайнов $M_{m,i}(t)$, а нормированные B -сплайны вычислить в последней строке.

Таким образом, мы нашли непрерывные $m-1$ раз дифференцируемые функции $N_{a,m}(t)$ и $N_{m,b}(t)$, принимающие ненулевые значения только внутри отрезка $t_a \leq t \leq t_b$ и имеющие равные нулю значения функций и их производных до $(m-2)$ -го порядка включительно на его концах. Мы не имеем аналитического выражения этих функций, но имеем рекуррентные формулы для вычисления значений функций в любой заданной точке. Значения функций зависят от значений $(m-1)$ -го внутреннего узла, принадлежащих отрезку $t_a \leq t \leq t_b$. Действительно, значения последовательности узлов, на которой вычисляются B -сплайны, при выводе всех формул ничем не ограничивались. Единственное требование, предъявляемое к значениям узлов, заключается в том, чтобы они не убывали. Это дает огромные возможности при построении кривых и поверхностей с использованием B -сплайнов.

Дифференцируя выражение для B -сплайна, получим формулу для вычисления его производных требуемого порядка. Производные B -сплайна также вычисляются по рекуррентным соотношениям. Их вычисление может выполняться одновременно с вычислением значений самих B -сплайнов.

В дальнейшем мы будем пользоваться первым определением B -сплайна (если не оговорено противное), хотя все сказанное будет справедливо и при других определениях B -сплайна.

2.9. NURBS кривые

B -сплайны используют для построения рациональных кривых линий в форме (2.7.5). Формула для вычисления радиус-вектора рациональной параметрической кривой на основе B -сплайнов $N_{i,m}(t)$, построенной по вершинам \mathbf{p}_i , $i = 1, 2, \dots, n$ ($n \geq m$), обладающим весами w_i , имеет вид

$$\mathbf{r}(t) = \frac{\sum_{i=1}^n N_{i,m}(t)w_i\mathbf{p}_i}{\sum_{i=1}^n N_{i,m}(t)w_i}, \quad t_{\min} \leq t \leq t_{\max}. \quad (2.9.1)$$

Радиус-вектор кривой равен частному от деления некоторого вектора на вес кривой в данной точке. Каждый из B -сплайнов m -порядка $N_{i,m}(t)$ построен на последовательности из $m + 1$ узлов $t_i, t_{i+1}, t_{i+2}, \dots, t_{i+m}$. Значения узлов должны образовывать неубывающую последовательность, а в остальном на них не накладывається никаких ограничений. Кривые линии, построенные на множестве не равноотстоящих узлов, называются *неоднородными*. Параметрическое расстояние между соседними узлами у неоднородных кривых меняется при переходе от узла к узлу. Кривая (2.9.1) в общем случае является неоднородной и рациональной. Кривая (2.9.1) носит имя *NURBS* кривой, составленное из первых букв ее названия — *Non-Uniform Rational B-Spline* (неоднородный рациональный фундаментальный сплайн).

Когда все вершины NURBS кривой имеют равные веса, то формула (2.9.1) для вычисления радиус-вектора кривой на основе B -сплайнов примет вид

$$\mathbf{r}(t) = \sum_{i=1}^n N_{i,m}(t)\mathbf{p}_i, \quad t_{\min} \leq t \leq t_{\max}.$$

Эта кривая также может быть использована в моделировании, так как она обладает определенным порядком гладкости, достаточно легко вычисляется и легко управляется своими вершинами.

В формуле (2.9.1) для построения совокупности из n B -сплайнов m -го порядка требуется $n + m$ узлов в случае незамкнутой кривой и $n + 2m$ узлов в случае замкнутой кривой. Число узлов всегда больше числа вершин, поэтому множество узлов называется расширенным. Воспользуемся той свободой, которую предоставляет нам выбор значений узлов при вычислении B -сплайнов. Пусть вершины сплайна (2.9.1) пронумерованы от 1 до n . Пронумеруем узлы, на которых построены B -сплайны, от 1 до $n + m$. Чтобы незамкнутая NURBS кривая проходила через первую и последнюю вершину, первый B -сплайн должен иметь кратными первые m узлов (из $m + 1$ узлов, на которых он строится), а последний B -сплайн должен иметь кратными последние m узлов (из $m + 1$ узлов, на которых он строится). Для построения незамкнутой $(m - 2)$ -раз дифференцируемой NURBS кривой первые m узлов должны иметь равные значения: $t_1 = t_2 = \dots = t_m$, следующие $n - m$ узлов должны образовывать возрастающую последовательность: $t_{m+i} < t_{m+i+1}$ ($i = 1, 2, \dots, n - m$), оставшиеся m узлов должны принимать равные значения: $t_{n+1} = t_{n+2} = \dots = t_{n+m}$. Для построения замкнутой $(m - 2)$ -раз дифференцируемой NURBS кривой последовательность узлов должна отражать

замкнутость: значения первых $n + m$ узлов должны образовывать возрастающую последовательность: $t_i < t_{i+1}$ ($i = 1, 2, \dots, n + m$), следующие m узлов должны идти через интервалы, повторяющие первые m интервалов между узлами: $t_{i+n+m} = t_{n+m} + t_i$ ($i = 1, 2, \dots, m$). В общем случае параметр NURBS кривой изменяется от значения узла $t_{\min} = t_m$ до значения узла $t_{\max} = t_{n+m}$. Построенная на таких последовательностях узлов кривая будет иметь непрерывные производные до $(m - 2)$ -го порядка включительно на всей области определения. Если среди узлов будут кратные, кроме вышеуказанных, то непрерывность соответствующих производных кривой будет нарушена.

Вычисление радиус-вектора кривой. Для вычисления радиус-вектора NURBS кривой используется следующая схема. По значению параметра t из условия $t_i \leq t < t_{i+1}$ определяется номер i отличного от нуля B -сплайна первого порядка и вычисляется его значение, исходя из определения:

$$M_{i,1}(t) = \begin{cases} \frac{1}{t_{i+1} - t_i}, & \text{если } t_i \leq t < t_{i+1}, \\ 0, & \text{если } t < t_i \text{ или } t \geq t_{i+1}. \end{cases} \quad (2.9.2)$$

Далее, используя рекуррентное соотношение Кокса-Де Бура

$$M_{i,m}(t) = \frac{(t_{i+m} - t)M_{i+1,m-1}(t) + (t - t_i)M_{i,m-1}(t)}{t_{i+m} - t_i}, \quad (2.9.3)$$

последовательно вычисляются все отличные от нуля при данном параметре t B -сплайны до m -го порядка включительно: $M_{i-m+1,m}(t)$, $M_{i-m+2,m}(t)$, \dots , $M_{i,m}(t)$:

				$M_{i,1}(t)$	
			$M_{i-1,2}(t)$	$M_{i,2}(t)$	
		$M_{i-2,3}(t)$	$M_{i-1,3}(t)$	$M_{i,3}(t)$	
				
	$M_{i-m+2,m-1}(t)$	\dots	$M_{i-2,m-1}(t)$	$M_{i-1,m-1}(t)$	$M_{i,m-1}(t)$
$M_{i-m+1,m}(t)$	$M_{i-m+2,m}(t)$	\dots	$M_{i-2,m}(t)$	$M_{i-1,m}(t)$	$M_{i,m}(t)$

Данная треугольная таблица вычисляется построчно, поскольку из (2.9.2) известен элемент первой строки, а каждый элемент следующей строки можно построить по двум соседним элементам в предыдущей строке с помощью (2.9.3). При вычислении крайних элементов каждой строки используется тот факт, что один из элементов в предыдущей строке равен нулю. Далее B -сплайны m -го порядка нормируются

$$N_{j,m}(t) = (t_{j+m} - t_j)M_{j,m}(t), \quad j = i - m + 1, i - m + 2, \dots, i,$$

и подставляются в формулу (2.9.1), которая примет вид

$$\mathbf{r}(t) = \frac{\sum_{j=1-m+1}^i N_{j,m}(t)w_j \mathbf{P}_j}{\sum_{j=1-m+1}^i N_{j,m}(t)w_j}. \quad (2.9.4)$$

Одновременно с вычислением B -сплайнов может проводиться вычисление их производных. Формулы (2.9.2) и (2.9.3) можно считать определением B -сплайна. B -сплайны являются локальными функциями (имеют малый носитель).

Свойства кривой. Каждый B -сплайн отличен от нуля только на части области изменения параметра кривой, поэтому при изменении радиус-вектора одной из вершин подлежит пересчету не вся кривая, а только ее часть. B -сплайны являются неотрицательными функциями. Сумма значений всех нормированных B -сплайнов при любом параметре t NURBS кривой в соответствии со свойством (2.8.24) равна единице: $\sum_i N_{i,m}(t) = 1$, а площадь под любым ненормированным B -сплайном удовлетворяет равенству

$$\int_{-\infty}^{+\infty} M_{i,m}(t) dt = \int_{t_i}^{t_{i+m}} M_{i,m}(t) dt = \frac{1}{m}. \quad (2.9.5)$$

Это равенство получим, переставив операции интегрирования и вычисления разделенной разности. Проинтегрируем усеченную степенную функцию (2.8.17) в указанных пределах

$$\int_{t_i}^{t_{i+m}} (z-t)^{m-1}_+ dt = \frac{1}{m}(z-t_i)^m_+ - \frac{1}{m}(z-t_{i+m})^m_+.$$

Вычислим разделенную разность полученной функции на последовательности узлов B -сплайна. Второе слагаемое правой части на области $t_i \leq t \leq t_{i+m}$ равно нулю, а первое слагаемое равно полиному m -й степени $(z-t_i)^m/m$, разделенная разность m -го порядка которого в соответствии со свойством (2.8.12) равна коэффициенту при z^m , т. е. равна $1/m$, что доказывает равенство (2.9.5).

Из (2.9.5) следуют равенства:

$$\int_{-\infty}^{+\infty} N_{i,m}(t) dt = (t_{i+m} - t_i) \int_{t_i}^{t_{i+m}} M_{i,m}(t) dt = \frac{t_{i+m} - t_i}{m},$$

$$\int_{-\infty}^{+\infty} \left(\sum_{i=1}^n N_{i,m}(t) w_i \mathbf{p}_i \right) dt = \frac{1}{m} \sum_{i=1}^n (t_{i+m} - t_i) w_i \mathbf{p}_i,$$

$$\int_{-\infty}^{+\infty} \left(\sum_{i=1}^n N_{i,m}(t) w_i \right) dt = \frac{1}{m} \sum_{i=1}^n (t_{i+m} - t_i) w_i.$$

NURBS кривая (2.9.1) обладает достаточной гибкостью. Ее можно легко редактировать путем изменения положения вершин. Например, легко можно модифицировать имеющуюся NURBS кривую так, чтобы она проходила через указанную точку или чтобы она касалась указанной кривой в заданной точке.

Производные радиус-вектора кривой. Используя определение B -сплайна и тот факт, что первая производная усеченной степенной функции по параметру t равна

$$\frac{d\sigma_{m-1}}{dt} = -(m-1)(z-t)^{m-2}_+ = -(m-1)\sigma_{m-2},$$

вычислим первую производную B -сплайна, переставляя дифференцирование и вычисление разделенных разностей

$$\begin{aligned} \frac{dN_{i,m}}{dt} &= \frac{d}{dt} (\sigma_{m-1}[t_{i+1}, t_{i+2}, \dots, t_{i+m}] - \sigma_{m-1}[t_i, t_{i+1}, \dots, t_{i+m-1}]) = \\ &= -\frac{m-1}{t_{i+m} - t_{i+1}} N_{i+1, m-1} + \frac{m-1}{t_{i+m-1} - t_i} N_{i, m-1}. \end{aligned} \quad (2.9.6)$$

Продифференцируем числитель выражения (2.9.1) с учетом (2.9.6) и получим

$$\begin{aligned} \frac{d(w\mathbf{r})}{dt} &\equiv \frac{d}{dt} \left(\sum_{i=1}^n N_{i,m} w_i \mathbf{p}_i \right) = \frac{d}{dt} \left(\sum_{i=j}^{j+m-1} N_{i,m} w_i \mathbf{p}_i \right) = \sum_{i=j}^{j+m-1} \frac{dN_{i,m}}{dt} w_i \mathbf{p}_i = \\ &= (m-1) \sum_{i=j}^{j+m-1} N_{i, m-1} \frac{w_i \mathbf{p}_i}{t_{i+m-1} - t_i} - (m-1) \sum_{i=j}^{j+m-1} N_{i+1, m-1} \frac{w_i \mathbf{p}_i}{t_{i+m} - t_{i+1}} = \\ &= (m-1) N_{j, m-1} \frac{w_j \mathbf{p}_j}{t_{j+m-1} - t_j} + (m-1) \sum_{i=j+1}^{j+m-1} N_{i, m-1} \frac{w_i \mathbf{p}_i}{t_{i+m-1} - t_i} - \\ &- (m-1) \sum_{i=j+1}^{j+m-1} N_{i, m-1} \frac{w_{i-1} \mathbf{p}_{i-1}}{t_{i+m-1} - t_i} - (m-1) N_{j+m, m-1} \frac{w_{j+m-1} \mathbf{p}_{j+m-1}}{t_{j+2m-1} - t_{j+m}} = \\ &= (m-1) \sum_{i=j+1}^{j+m-1} N_{i, m-1} \frac{w_i \mathbf{p}_i - w_{i-1} \mathbf{p}_{i-1}}{t_{i+m-1} - t_i} = \sum_{i=j+1}^{j+m-1} N_{i, m-1} \mathbf{p}_i^{(1)}, \end{aligned} \quad (2.9.7)$$

где $j, j+1, j+2, \dots, j+m-1$ — индексы отличных от нуля B -сплайнов m -го порядка при заданном параметре $t_{j+m-1} \leq t < t_{j+m}$. В равенстве (2.9.7) использовались свойства $N_{j, m-1}(t) = 0$ и $N_{j+m, m-1}(t) = 0$. Действительно, для заданной последовательности узлов и при любом параметре t , принадлежащем области определения кривой, среди B -сплайнов m -го порядка отличны от нуля только m B -сплайнов (пусть их индексы равны $j, j+1, j+2, \dots, j+m-1$), а среди B -сплайнов $(m-1)$ -го порядка отличны от нуля $m-1$ B -сплайнов (их индексы, соответственно, равны $j+1, j+2, \dots, j+m-1$). Мы видим, что первая производная сплайн-функции m -го порядка представляет собой аналогичную сплайн-функцию, порядок и число вершин которой на единицу меньше. Вершины $\mathbf{p}_i^{(1)}$ вычисляются по формуле

$$\mathbf{p}_i^{(1)} = (m-1) \frac{w_i \mathbf{p}_i - w_{i-1} \mathbf{p}_{i-1}}{t_{i+m-1} - t_i}, \quad i = 2, 3, \dots, n. \quad (2.9.8)$$

Продолжая дифференцирование, можно найти производную требуемого порядка для числителя выражения (2.9.1):

$$\frac{d^k(w\mathbf{r})}{dt^k} \equiv \frac{d^k}{dt^k} \left(\sum_{i=1}^n N_{i,m} w_i \mathbf{p}_i \right) = \sum_{i=1+k}^n N_{i,m-k} \mathbf{p}_i^{(k)}, \quad (2.9.9)$$

где $\mathbf{p}_i^{(k)} = (m-k) \frac{\mathbf{p}_i^{(k-1)} - \mathbf{p}_{i-1}^{(k-1)}}{t_{i+m-k} - t_i}$, $k = 1, 2, \dots, m-1$, $i = 1+k, 2+k, \dots, n$,
 $\mathbf{p}_i^{(0)} = w_i \mathbf{p}_i$.

Для производных знаменателя (2.9.1) получим выражение, аналогичное (2.9.9),

$$\frac{d^k w}{dt^k} \equiv \frac{d^k}{dt^k} \left(\sum_{i=1}^n N_{i,m} w_i \right) = \sum_{i=1+k}^n N_{i,m-k} w_i^{(k)},$$

где $w_i^{(k)} = (m-k) \frac{w_i^{(k-1)} - w_{i-1}^{(k-1)}}{t_{i+m-k} - t_i}$, $k = 1, 2, \dots, m-1$, $i = 1+k, 2+k, \dots, n$,
 $w_i^{(0)} = w_i$.

Радиус-вектор NURBS кривой (2.9.1) вычисляется как частное от деления двух функций параметра кривой t , поэтому при вычислении производной NURBS кривой правую часть (2.9.1) следует рассматривать как сложную функцию. Первая производная радиус-вектора NURBS кривой равна

$$\frac{d\mathbf{r}}{dt} = \frac{d}{dt} \left(\frac{w\mathbf{r}(t)}{w(t)} \right) = \frac{1}{w} \frac{d(w\mathbf{r})}{dt} - \frac{w\mathbf{r}}{w^2} \frac{dw}{dt}, \quad (2.9.10)$$

где

$w\mathbf{r} = \sum_{i=1}^n N_{i,m} w_i \mathbf{p}_i$ — числитель правой части равенства (2.9.1),

$w = \sum_{i=1}^n N_{i,m} w_i$ — знаменатель правой части равенства (2.9.1),

$\frac{d(w\mathbf{r})}{dt} = (m-1) \sum_{i=2}^n N_{i,m-1} \frac{w_i \mathbf{p}_i - w_{i-1} \mathbf{p}_{i-1}}{t_{i+m-1} - t_i}$ — производная,

$\frac{dw}{dt} = (m-1) \sum_{i=2}^n N_{i,m-1} \frac{w_i - w_{i-1}}{t_{i+m-1} - t_i}$ — производная веса.

Аналогично вычисляются производные NURBS кривой более высокого порядка

$$\frac{d\mathbf{r}}{dt} = \frac{d^{(k)}}{dt^{(k)}} \left(\frac{w\mathbf{r}(t)}{w(t)} \right).$$

Производные второго и третьего порядка определены равенствами (2.7.10), (2.7.11). Можно заметить, что радиус-вектор точки и ее вес ($w\mathbf{r}$ или $w_i \mathbf{p}_i$) в формулах NURBS кривой выступают как единое целое.

Алгоритм Де Бура. Преобразуем числитель выражения (2.9.1), используя формулу (2.8.32.1), следующим образом:

$$\begin{aligned}
 wr &\equiv \sum_{i=1}^n N_{i,m}(t) w_i \mathbf{p}_i = \sum_{i=j}^{j+m-1} N_{i,m}(t) w_i \mathbf{p}_i = \\
 &= \sum_{i=j}^{j+m-1} \left(\frac{t_{i+m} - t}{t_{i+m} - t_{i+1}} N_{i+1,m-1}(t) + \frac{t - t_i}{t_{i+m-1} - t_i} N_{i,m-1}(t) \right) w_i \mathbf{p}_i = \\
 &= \sum_{i=j+1}^{j+m-1} \frac{t_{i+m-1} - t}{t_{i+m-1} - t_i} N_{i,m-1}(t) w_{i-1} \mathbf{p}_{i-1} + \\
 &\quad + \frac{t_{j+2m-1} - t}{t_{j+2m-1} - t_{j+m}} N_{j+m,m-1}(t) w_{j+m-1} \mathbf{p}_{j+m-1} + \\
 &\quad + \frac{t - t_j}{t_{j+m-1} - t_j} N_{j,m-1}(t) w_j \mathbf{p}_j + \sum_{i=j+1}^{j+m-1} \frac{t - t_i}{t_{i+m-1} - t_i} N_{i,m-1}(t) w_i \mathbf{p}_i = \\
 &= \sum_{i=j+1}^{j+m-1} N_{i,m-1} \left(\frac{t_{i+m-1} - t}{t_{i+m-1} - t_i} w_{i-1} \mathbf{p}_{i-1} + \frac{t - t_i}{t_{i+m-1} - t_i} w_i \mathbf{p}_i \right) = \\
 &= \sum_{i=j+1}^{j+m-1} N_{i,m-1}(t) \mathbf{r}_i^{(1)}, \quad (2.9.11)
 \end{aligned}$$

где $j, j+1, j+2, \dots, j+m-1$ — индексы отличных от нуля B -сплайнов m -го порядка при заданном параметре $t_{j+m-1} \leq t < t_{j+m}$. В преобразовании (2.9.11), как и при выводе равенства (2.9.7), использовались свойства $N_{j,m-1}(t) = 0$ и $N_{j+m,m-1}(t) = 0$ при заданном параметре t . Мы свели сумму m слагаемых к сумме $m-1$ слагаемых и понизили на единицу порядок B -сплайнов в этой сумме. Величины $\mathbf{r}_i^{(1)}$ вычисляются по формуле

$$\begin{aligned}
 \mathbf{r}_i^{(1)} &= \frac{t_{i+m-1} - t}{t_{i+m-1} - t_i} w_{i-1} \mathbf{p}_{i-1} + \frac{t - t_i}{t_{i+m-1} - t_i} w_i \mathbf{p}_i, \\
 &i = j+1, j+2, \dots, j+m-1.
 \end{aligned}$$

Продолжим упрощение числителя выражения (2.9.1) аналогичным образом и получим

$$\begin{aligned}
 wr &\equiv \sum_{i=1}^n N_{i,m}(t) w_i \mathbf{p}_i = \sum_{i=j}^{j+m-1} N_{i,m}(t) w_i \mathbf{p}_i = \sum_{i=j+1}^{j+m-1} N_{i,m-1}(t) \mathbf{r}_i^{(1)} = \\
 &= \sum_{i=j+2}^{j+m-1} N_{i,m-2} \left(\frac{t_{i+m-2} - t}{t_{i+m-2} - t_i} \mathbf{r}_{i-1}^{(1)} + \frac{t - t_i}{t_{i+m-2} - t_i} \mathbf{r}_i^{(1)} \right) = \\
 &= \sum_{i=j+2}^{j+m-1} N_{i,m-2}(t) \mathbf{r}_i^{(2)} = \dots = \sum_{i=j+m-1}^{j+m-1} N_{i,1}(t) \mathbf{r}_i^{(m-1)} = \mathbf{r}_{j+m-1}^{(m-1)}. \quad (2.9.12)
 \end{aligned}$$

Для знаменателя выражения (2.9.1) получим аналогичное выражение

$$\begin{aligned}
 w &\equiv \sum_{i=1}^n N_{i,m}(t)w_i = \sum_{i=j}^{j+m-1} N_{i,m}(t)w_i = \sum_{i=j+1}^{j+m-1} N_{i,m-1}(t)w_i^{(1)} = \\
 &= \sum_{i=j+2}^{j+m-1} N_{i,m-2}(t)w_i^{(2)} = \dots = \sum_{i=j+m-1}^{j+m-1} N_{i,1}(t)w_i^{(m-1)} = w_{j+m-1}^{(m-1)}. \quad (2.9.13)
 \end{aligned}$$

Мы пришли к выводу, что положение точки NURBS кривой (2.9.1) для заданного параметра $t_{j+m-1} \leq t < t_{j+m}$ может быть определено по формуле

$$\mathbf{r}(t) = \frac{w\mathbf{r}(t)}{w(t)} = \frac{\mathbf{r}_{j+m-1}^{(m-1)}}{w_{j+m-1}^{(m-1)}} \quad (2.9.14)$$

с помощью рекуррентных соотношений

$$\begin{aligned}
 \mathbf{r}_i^{(k)} &= \frac{t_{i+m-k} - t}{t_{i+m-k} - t_i} \mathbf{r}_{i-1}^{(k-1)} + \frac{t - t_i}{t_{i+m-k} - t_i} \mathbf{r}_i^{(k-1)}, \\
 w_i^{(k)} &= \frac{t_{i+m-k} - t}{t_{i+m-k} - t_i} w_{i-1}^{(k-1)} + \frac{t - t_i}{t_{i+m-k} - t_i} w_i^{(k-1)}, \quad (2.9.15)
 \end{aligned}$$

$$k = 1, 2, \dots, m-1, \quad i = j+k, j+k+1, \dots, j+m-1,$$

которые начинаются со значений $\mathbf{r}_i^{(0)} = w_i \mathbf{p}_i$, $w^{(0)} = w_i$, $i = j, j+1, j+2, \dots, j+m-1$. Эти соотношения являются обобщением алгоритма Де Кастелье и называются *алгоритмом Де Бура*. Алгоритм вычисления радиус-вектора точки NURBS кривой для параметра $t_{j+m-1} \leq t < t_{j+m}$ иллюстрирует рис. 2.9.1:

$$\begin{array}{ccccccc}
 \mathbf{r}_j^{(0)} & \mathbf{r}_{j+1}^{(0)} & \mathbf{r}_{j+2}^{(0)} & \dots & \mathbf{r}_{j+m-3}^{(0)} & \mathbf{r}_{j+m-2}^{(0)} & \mathbf{r}_{j+m-1}^{(0)} \\
 \mathbf{r}_{j+1}^{(1)} & \mathbf{r}_{j+2}^{(1)} & \dots & \mathbf{r}_{j+m-3}^{(1)} & \mathbf{r}_{j+m-2}^{(1)} & \mathbf{r}_{j+m-1}^{(1)} & \\
 \mathbf{r}_{j+2}^{(2)} & \mathbf{r}_{j+3}^{(2)} & \dots & \mathbf{r}_{j+m-2}^{(2)} & \mathbf{r}_{j+m-1}^{(2)} & & \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 & \mathbf{r}_{j+m-2}^{(m-2)} & \mathbf{r}_{j+m-1}^{(m-2)} & & & & \\
 & & \mathbf{r}_{j+m-1}^{(m-1)} & & & &
 \end{array}$$

Рис. 2.9.1. Алгоритм Де Бура

Вставка узлов. У заданной NURBS кривой m -го порядка можно увеличить число характеристических точек, сохранив неизменной ее форму и параметрическую длину. Для этого применяется вставка дополнительных узлов. Вставим дополнительный узел t^* в последовательность узлов t_1, t_2, \dots, t_n . Пусть значение t^* расположено между значениями узлов t_{j+m-1} и t_{j+m} , т.е. удовлетворяет неравенствам $t_{j+m-1} \leq t^* < t_{j+m}$. Используя алгоритм Де Бура, можно показать, что кривая не изменится, если величины $\mathbf{r}_{j+1}^{(0)}, \mathbf{r}_{j+2}^{(0)}, \dots, \mathbf{r}_{j+m-2}^{(0)}$ заменить на величины $\mathbf{r}_{j+1}^{(1)}, \mathbf{r}_{j+2}^{(1)}, \dots, \mathbf{r}_{j+m-1}^{(1)}$ (см. рис. 2.9.1). Таким образом,

вместе с вставкой узла $t_{j+m-1} \leq t^* < t_{j+m}$ мы должны заменить характеристические точки $\mathbf{P}_{j+1}, \mathbf{P}_{j+2}, \dots, \mathbf{P}_{j+m-2}$ с весами $w_{j+1}, w_{j+2}, \dots, w_{j+m-2}$ на вершины с радиус-векторами $\frac{\mathbf{r}_{j+1}^{(1)}}{w_{j+1}^{(1)}}, \frac{\mathbf{r}_{j+2}^{(1)}}{w_{j+2}^{(1)}}, \dots, \frac{\mathbf{r}_{j+m-1}^{(1)}}{w_{j+m-1}^{(1)}}$ и весами $w_{j+1}^{(1)}, w_{j+2}^{(1)}, \dots, w_{j+m-1}^{(1)}$. Новых вершин на одну больше, чем старых, поэтому номера узлов и вершин начиная с $j+m$ нужно увеличить на единицу.

Узел t^* может быть вставлен многократно (максимальная кратность равна $m-1$). Если узел t^* вставляется k раз, то величины $\mathbf{r}_{j+1}^{(0)}, \mathbf{r}_{j+2}^{(0)}, \dots, \mathbf{r}_{j+m-2}^{(0)}$ нужно заменить на величины $\mathbf{r}_{j+1}^{(1)}, \mathbf{r}_{j+2}^{(2)}, \dots, \mathbf{r}_{j+k}^{(k)}, \mathbf{r}_{j+k+1}^{(k)}, \dots, \mathbf{r}_{j+m-1}^{(k)}, \mathbf{r}_{j+m-1}^{(k-1)}, \dots, \mathbf{r}_{j+m-1}^{(2)}, \mathbf{r}_{j+m-1}^{(1)}$ (см. рис. 2.9.1). По их значениям можно определить новые вершины и их веса. Новых вершин будет на k больше, чем старых, поэтому номера узлов и вершин начиная с $j+m$ нужно увеличить на k .

Примеры. Для незамкнутой NURBS кривой мы будем использовать следующие значения узлов. Пусть первые m узлов имеют значения, равные нулю: $t_1 = t_2 = \dots = t_m = 0$, следующие $n-m$ узлов принимают целочисленные

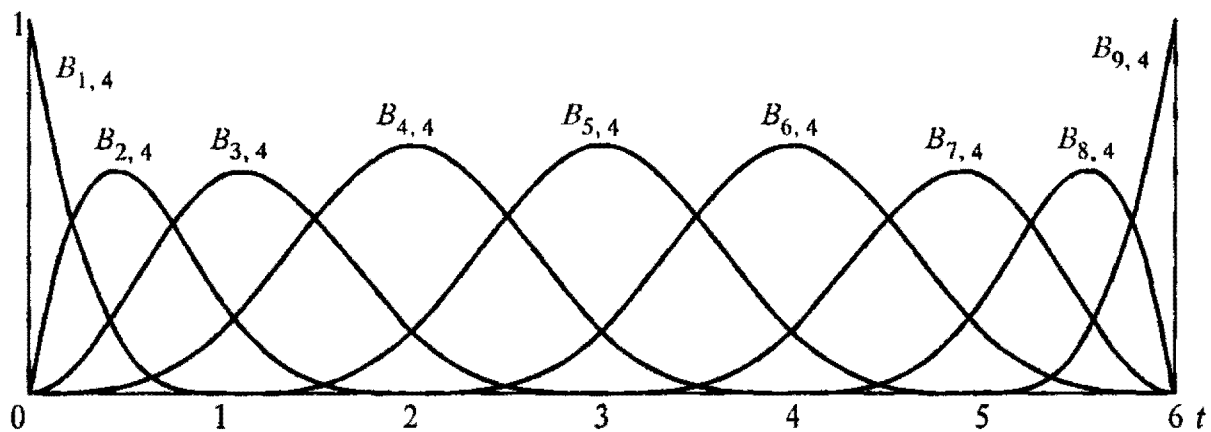


Рис. 2.9.2. Набор B -сплайнов 4-го порядка для незамкнутой NURBS кривой

значения от 1 до $n-m$: $t_{m+i} = i, i = 1, 2, \dots, n-m$, оставшиеся m узлов принимают значение $n-m+1$: $t_{n+1} = t_{n+2} = \dots = t_{n+m} = n-m+1$. Для замкнутой NURBS кривой мы будем использовать равноотстоящие значения

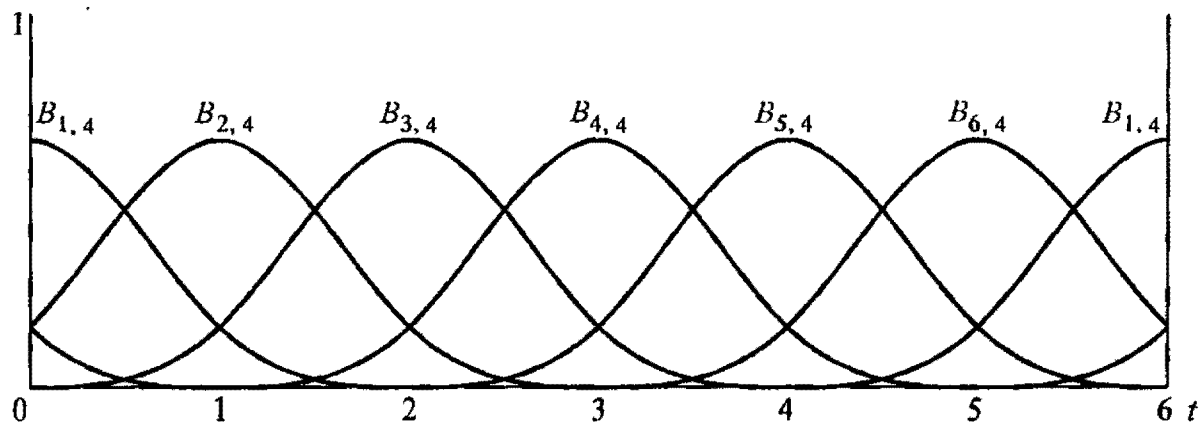


Рис. 2.9.3. Набор B -сплайнов 4-го порядка для замкнутой NURBS кривой

узлов: $t_i = i - m, i = 1, 2, \dots, n + 2m$. Параметр NURBS кривой изменяется в пределах: $t_m \leq t \leq t_{n+m}$, что для незамкнутой кривой есть $0 \leq t \leq n - m + 1$, а для замкнутой кривой есть $0 \leq t \leq n$. Параметризация с равноотстоящими

значениями узлов называется равномерной. Сплайны с равномерной параметризацией Шёнберг назвал *кардинальными*.

На рис. 2.9.2 показан полный набор B -сплайнов 4-го порядка (кубических) для 9 вершин, построенных на расширенном множестве узлов $0, 0, 0, 0, 1, 2, 3,$

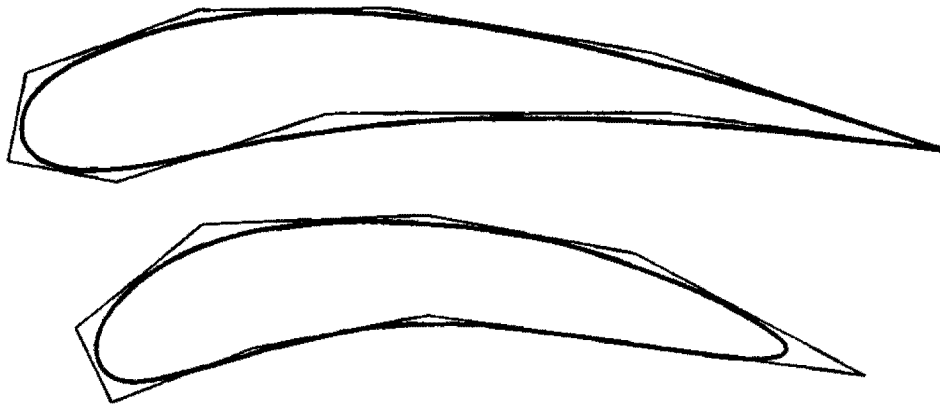


Рис. 2.9.4. NURBS кривые и их характеристические ломаные линии

4, 5, 6, 6, 6, 6. Параметр t сплайна (2.9.1), построенного на данном множестве узлов, принимает значения на отрезке $0 \leq t \leq 6$.

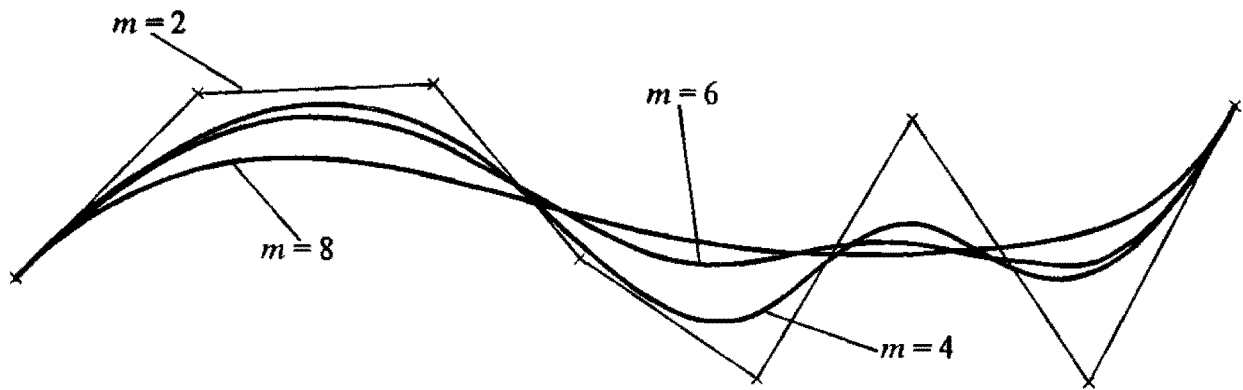


Рис. 2.9.5. Влияние порядка на форму NURBS кривой

Для замкнутой NURBS кривой используется последовательность равноотстоящих узлов: $t_i = i - m, i = 1, 2, \dots, n + 2m$. На рис. 2.9.3 показан полный

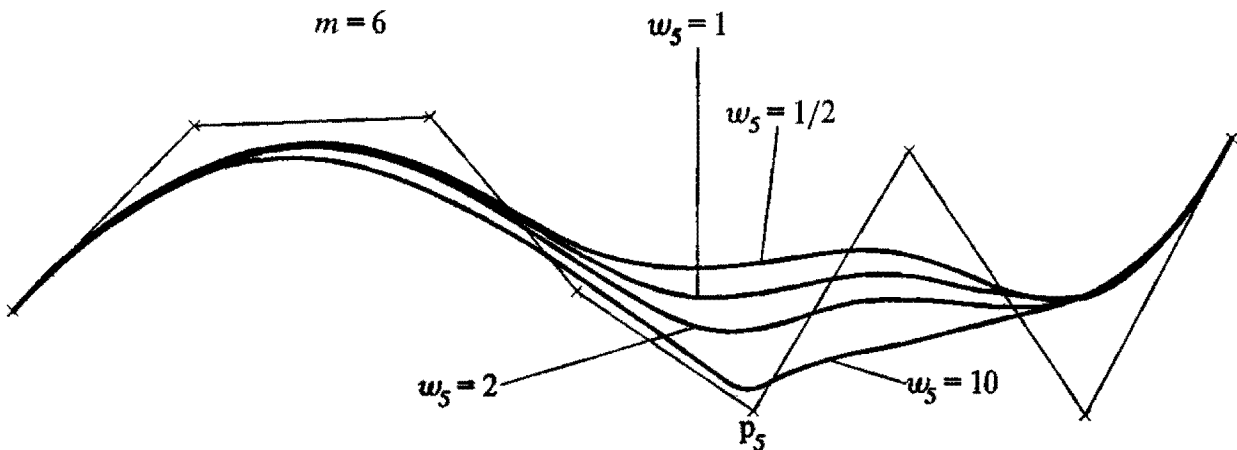


Рис. 2.9.6. Влияние веса точки p_5 на форму NURBS кривой

набор B -сплайнов 4-го порядка для 6 вершин, построенных на расширенном множестве узлов $-3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$.

Все B -сплайны для замкнутой кривой похожи друг на друга. Параметр t сплайна (2.9.1), построенного на данном множестве узлов, принимает значения на отрезке $0 \leq t \leq 6$.

На рис. 2.9.4 показаны незамкнутый и замкнутый сплайны и их характеристические ломаные, построенные по описанным выше последовательностям узлов. У незамкнутой кривой крайние точки совпадают. Как и кривая Безье, NURBS кривая не проходит через свои вершины за исключением крайних точек для незамкнутой кривой.

На рис. 2.9.5 по одним и тем же восьми вершинам построены NURBS кривые 2-го, 4-го, 6-го и 8-го порядка. На рисунке мы наблюдаем, что чем выше порядок кривой, тем она более гладкая. NURBS кривая второго порядка совпадает со своей характеристической ломаной линией.

На рис. 2.9.6 показано влияние веса вершины p_5 на форму кривой 6-го порядка. Чем больше вес вершины, тем ближе к ней проходит NURBS кривая. В общем случае вес вершины может быть нулевым и даже отрицательным.

NURBS представление ломаной линии. Если построить NURBS кривую (2.9.1) по n вершинам на последовательности узлов $t_1 = t_2 = 0$, $t_3 = 1$, $t_4 = 2, \dots, t_i = i - 2, \dots, t_n = n - 2$, $t_{n+1} = t_{n+2} = n - 1$ на базе B -сплайнов второго порядка $N_{i,2}(t)$, то она совпадет с ломаной линией (2.4.1). Область изменения параметра NURBS кривой в этом случае равна $0 \leq t \leq n - 1$. B -сплайны второго порядка $N_{i,2}(t)$ пропорциональны разделенным разностям второго порядка, вычисленным на последовательности узлов t_i, t_{i+1}, t_{i+2} для функции $\sigma_1(z) = (z - 1)_+$. B -сплайны второго порядка определяются равенством

$$N_{i,2}(t) = \begin{cases} \frac{t - t_i}{t_{i+1} - t_i}, & \text{если } t_i \leq t < t_{i+1}, \\ \frac{t_{i+2} - t}{t_{i+2} - t_{i+1}}, & \text{если } t_{i+1} \leq t < t_{i+2}, \\ 0, & \text{если } t < t_i \text{ или } t \geq t_{i+2}. \end{cases} \quad (2.9.16)$$

Мы видим, что $N_{i,2}(t)$ представляют собой кусочно-линейные функции, принимающие значения $N_{i,2}(t_{i+1}) = 1$ и $N_{j,2}(t_{i+1}) = 0$, $j \neq i$. B -сплайны (2.9.16) для шести вершин показаны на рис. 2.9.7.

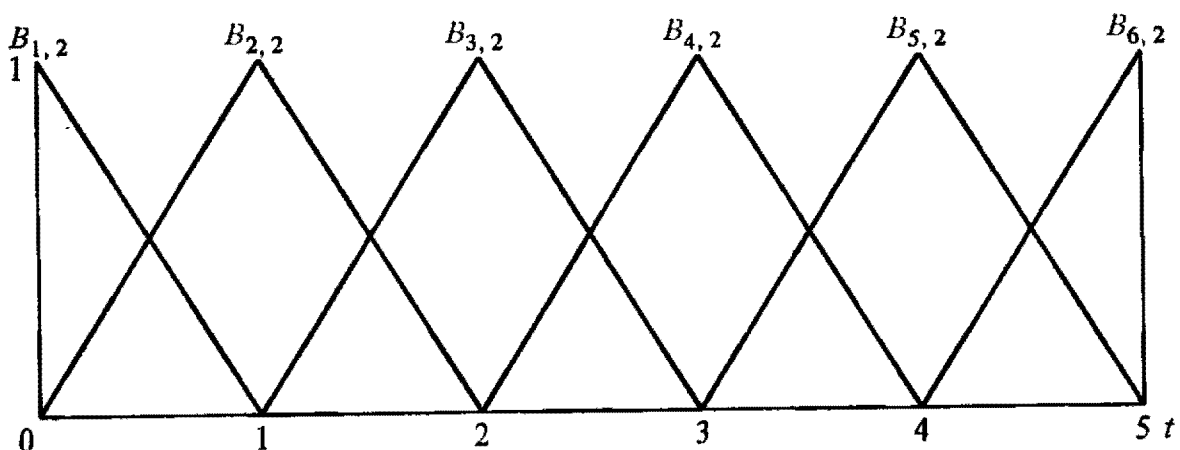


Рис. 2.9.7. B -сплайны второго порядка

Замкнутую ломаную линию как частный случай NURBS кривой (2.9.1) можно построить на последовательности узлов $t_1 = -1$, $t_2 = 0$, $t_3 = 1, \dots, t_i = i - 2, \dots, t_n = n - 2$, $t_{n+1} = n - 1$, $t_{n+2} = n$. Область изменения параметра NURBS кривой в этом случае равна $0 \leq t \leq n$.

NURBS представление отрезка прямой. Отрезок прямой может быть представлен в виде NURBS кривой (2.9.1)

$$\mathbf{r}(t) = \mathbf{p}_1 N_{1,2}(t) + \mathbf{p}_2 N_{2,2}(t),$$

по двум вершинам \mathbf{p}_1 и \mathbf{p}_2 на последовательности узлов $t_1 = t_2 = 0$, $t_3 = t_4 = 1$ и на базе B -сплайнов второго порядка $N_{i,2}(t)$ (2.9.16):

$$N_{1,2}(t) = \frac{t_3 - t}{t_3 - t_2} = 1 - t, \quad N_{2,2}(t) = \frac{t - t_2}{t_3 - t_2} = t, \quad 0 \leq t \leq 1.$$

Кривая Безье как частный случай NURBS кривой. Нормированные B -сплайны $N^n_i(t)$ определения 4 (см. § 2.8) в частном случае последовательности узлов $t_{-1-n} = t_{-n} = t_{1-n} = \dots = t_{-1} = 0$, $t_0 = t_1 = \dots = t_n = 1$ совпадают с базисными функциями Бернштейна $B^n_i(t)$, а NURBS кривая (2.9.1), записанная в виде

$$\mathbf{r}(t) = \frac{\sum_{i=0}^n N^n_i(t) w_i \mathbf{p}_i}{\sum_{i=0}^n N^n_i(t) w_i}, \quad t_{\min} \leq t \leq t_{\max}, \quad (2.9.17)$$

совпадает с рациональной кривой Безье (2.7.2). Это происходит, когда число вершин на единицу больше порядка B -сплайна $N^n_i(t)$. Область изменения параметра NURBS кривой в этом случае равна $0 \leq t \leq 1$.

Равенство B -сплайнов $N^n_i(t)$ определения (2.8.21.4), построенных на последовательности узлов $t_{-1-n} = t_{-n} = t_{1-n} = \dots = t_{-1} = 0$, $t_0 = t_1 = \dots = t_n = 1$,

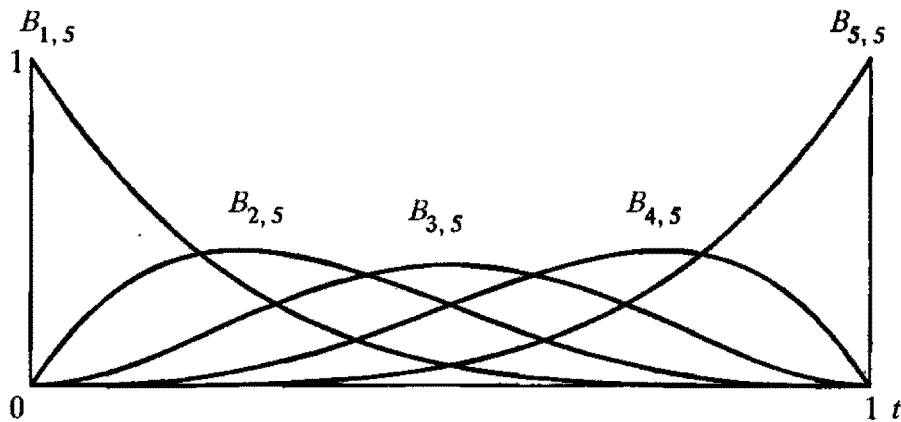


Рис. 2.9.8. Функции Бернштейна — частный случай B -сплайнов

и коэффициентов Бернштейна $B^n_i(t)$ можно доказать следующим образом. Вспомним, что коэффициенты Бернштейна связаны рекуррентным соотношением

$$B^n_i(t) = t B^{n-1}_{i-1}(t) + (1-t) B^{n-1}_i(t), \quad i = 0, 1, 2, \dots, n.$$

Мы получим все коэффициенты Бернштейна, начав вычисления с $B^0_0(t) = 1$. B -сплайны $N^n_i(t)$ связаны рекуррентным соотношением

$$\begin{aligned} N^n_i(t) &= \frac{t_i - t}{t_i - t_{i-n}} N^{n-1}_i(t) + \frac{t - t_{i-n-1}}{t_i - t_{i-n-1}} N^{n-1}_{i-1}(t) = \\ &= (1-t) N^{n-1}_i(t) + t N^{n-1}_{i-1}(t), \quad i = 0, 1, 2, \dots, n, \end{aligned} \quad (2.9.18)$$

где только $N^0_0(t) = 1$, а остальные B -сплайны нулевого порядка равны нулю. Из приведенных формул видно, что коэффициенты Бернштейна $B^n_i(t)$ и B -сплайны $N^n_i(t)$ вычисляются совершенно одинаково, что доказывает их равенство. Из сказанного следует, что нормированные B -сплайны $N_{i,n}(t)$ определения 1 в частном случае последовательности узлов $t_1 = t_2 = \dots = t_n = 0$, $t_{n+1} = t_{n+2} = \dots = t_{2n} = 1$ совпадают с базисными функциями Бернштейна $B^{n-1}_{i-1}(t)$. Этим доказывается, что кривые Безье являются частным случаем NURBS кривой. B -сплайны (2.9.17) 5-го порядка на последовательности узлов $t_1 = t_2 = \dots = t_5 = 0$, $t_6 = t_7 = \dots = t_{10} = 1$ показаны на рис. 2.9.8.

NURBS представление кривой второго порядка. Выше мы показали, что кривые второго порядка могут быть представлены в виде рациональных кривых Безье. NURBS кривые являются обобщением рациональных кривых Безье,

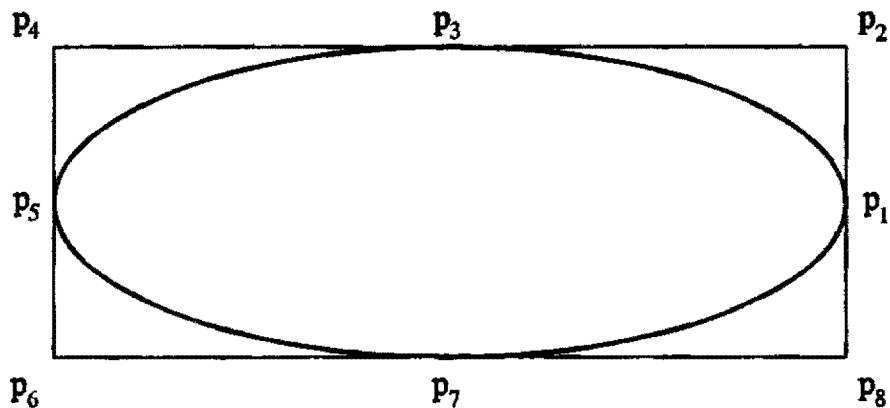


Рис. 2.9.9. Эллиптическая NURBS кривая

поэтому некоторая часть любой кривой второго порядка может быть представлена в виде NURBS кривой по трем вершинам на последовательности узлов $t_1 = t_2 = t_3 = 0$, $t_4 = t_5 = t_6 = 1$ на базе B -сплайнов третьего порядка $N_{i,3}(t)$.

Полностью кривую второго порядка можно представить в виде NURBS кривой, если использовать кратные узлы. На рис. 2.9.9 приведена эллиптическая NURBS кривая 3-го порядка и ее характеристическая ломаная.

Она построена по точкам $P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8$. Ее характеристическая ломаная представляет собой прямоугольник, описанный около эллиптической

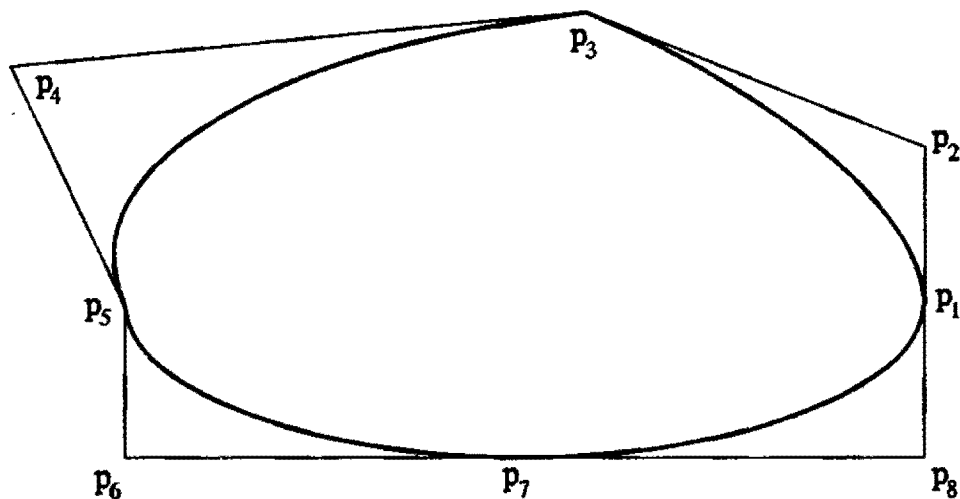


Рис. 2.9.10. NURBS кривая с кратными узлами

ской кривой. Веса характеристических точек равны: $w_1 = w_3 = w_5 = w_7 = 1$, $w_2 = w_4 = w_6 = w_8 = \cos(\pi/4)$. Последовательность узлов имеет вид $t_1 = t_2 = t_3 = 0$, $t_4 = t_5 = 1$, $t_6 = t_7 = 2$, $t_8 = t_9 = 3$, $t_{10} = t_{11} = t_{12} = 4$. Если мы хотим, чтобы параметрическая длина кривой, показанной на рис. 2.9.9, была

равна 2π , то вместо указанной выше последовательности узлов можно использовать последовательность: $t_1 = t_2 = t_3 = 0$, $t_4 = t_5 = \pi/2$, $t_6 = t_7 = \pi$, $t_8 = t_9 = 3\pi/2$, $t_{10} = t_{11} = t_{12} = 2\pi$. NURBS кривая может иметь изломы в точках с кратными узлами. Изломы возникают, когда соответствующие соседние вершины не расположены на одной прямой. На рис. 2.9.10 приведена та же эллиптическая NURBS кривая, но с измененными положениями точек \mathbf{p}_3 , \mathbf{p}_4 .

NURBS кривая (2.9.1) обобщает все рассмотренные выше кривые — отрезок прямой, ломаная линия, кривая Безье, конические сечения могут быть описаны в виде NURBS кривой. По своей природе NURBS кривая представляет собой линейную комбинацию кусочно-полиномиальных функций заданной степени, она позволяет строить кривые заданного порядка гладкости.

NURBS представление кубического сплайна. Пусть построен незамкнутый кубический сплайн (2.4.10) $\mathbf{a}(x)$, проходящий через точки \mathbf{a}_j при значениях

параметра x_j , $j = 1, 2, \dots, n$. Построим NURBS кривую $\mathbf{r}(t) = \sum_{i=1}^{n+2} N_{i,4}(t)\mathbf{p}_i$ четвертого порядка ($m = 4$) на последовательности узлов $t_1 = t_2 = t_3 = t_4 = x_1$, $t_5 = x_2$, $t_6 = x_3$, \dots , $t_{n+m-1} = t_{n+m} = t_{n+m+1} = t_{n+m+2} = x_n$. Для этого найдем $n + 2$ вершины \mathbf{p}_i , $i = 1, 2, \dots, n + 2$ из условий:

$$\begin{aligned} \mathbf{r}(x_1) &= \mathbf{a}_1, & \left. \frac{d\mathbf{a}}{dx} \right|_{x=x_1} &= \mathbf{a}'_1, \\ \mathbf{r}(x_j) &= \mathbf{a}_j, & j &= 2, 3, \dots, n-1, \\ \left. \frac{d\mathbf{a}}{dx} \right|_{x=x_n} &= \mathbf{a}'_n, & \mathbf{r}(x_n) &= \mathbf{a}_n. \end{aligned}$$

Веса вершины \mathbf{p}_i равны единице. Кубический сплайн $\mathbf{a}(x)$ и NURBS кривая $\mathbf{r}(t)$ будут полностью совпадать, так как они состоят из гладко стыкующихся (до второй производной включительно) участков кривых третьей степени. Таким образом, вершины \mathbf{p}_i , $i = 1, 2, \dots, n+2$ мы найдем из системы $n+2$ уравнений:

$$\begin{aligned} N_{1,4}(x_1)\mathbf{p}_1 + N_{2,4}(x_1)\mathbf{p}_2 + \dots + N_{n+1,4}(x_1)\mathbf{p}_{n+1} + N_{n+2,4}(x_1)\mathbf{p}_{n+2} &= \mathbf{a}_1, \\ 3N_{2,3}(x_1)\frac{\mathbf{p}_2 - \mathbf{p}_1}{t_5 - t_2} + 3N_{3,3}(x_1)\frac{\mathbf{p}_3 - \mathbf{p}_2}{t_6 - t_3} + \dots + 3N_{n+2,3}(x_1)\frac{\mathbf{p}_{n+2} - \mathbf{p}_{n+1}}{t_{n+5} - t_{n+2}} &= \mathbf{a}'_1, \\ N_{1,4}(x_2)\mathbf{p}_1 + N_{2,4}(x_2)\mathbf{p}_2 + \dots + N_{n+1,4}(x_2)\mathbf{p}_{n+1} + N_{n+2,4}(x_2)\mathbf{p}_{n+2} &= \mathbf{a}_2, \\ N_{1,4}(x_3)\mathbf{p}_1 + N_{2,4}(x_3)\mathbf{p}_2 + \dots + N_{n+1,4}(x_3)\mathbf{p}_{n+1} + N_{n+2,4}(x_3)\mathbf{p}_{n+2} &= \mathbf{a}_3, \\ \dots & \\ N_{1,4}(x_{n-1})\mathbf{p}_1 + N_{2,4}(x_{n-1})\mathbf{p}_2 + \dots + N_{n+1,4}(x_{n-1})\mathbf{p}_{n+1} + & \\ &+ N_{n+2,4}(x_{n-1})\mathbf{p}_{n+2} = \mathbf{a}_{n-1}, \\ 3N_{2,3}(x_n)\frac{\mathbf{p}_2 - \mathbf{p}_1}{t_5 - t_2} + 3N_{3,3}(x_n)\frac{\mathbf{p}_3 - \mathbf{p}_2}{t_6 - t_3} + \dots + 3N_{n+2,3}(x_n)\frac{\mathbf{p}_{n+2} - \mathbf{p}_{n+1}}{t_{n+5} - t_{n+2}} &= \mathbf{a}'_n, \\ N_{1,4}(x_n)\mathbf{p}_1 + N_{2,4}(x_n)\mathbf{p}_2 + \dots + N_{n+1,4}(x_n)\mathbf{p}_{n+1} + N_{n+2,4}(x_n)\mathbf{p}_{n+2} &= \mathbf{a}_n. \end{aligned} \tag{2.9.19}$$

Используя формулы (2.9.2) и (2.9.3), определим, что для каждого значения параметра $t = t_i$, $i = 4, 5, \dots, n + m - 1$, отличны от нуля только три B -сплайна четвертого порядка:

$$\begin{aligned}
 N_{i-3,4}(t_i) &= \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-2}} \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-1}}, \\
 N_{i-2,4}(t_i) &= \frac{t_i - t_{i-1}}{t_{i+2} - t_{i-1}} \frac{t_{i+2} - t_i}{t_{i+1} - t_{i-1}} + \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-2}} \frac{t_i - t_{i-2}}{t_{i+1} - t_{i-1}}, \\
 N_{i-1,4}(t_i) &= \frac{t_i - t_{i-1}}{t_{i+2} - t_{i-1}} \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}
 \end{aligned} \tag{2.9.20}$$

и два B -сплайна третьего порядка:

$$N_{i-2,3}(t_i) = \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-1}}, \quad N_{i-1,3}(t_i) = \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}. \tag{2.9.21}$$

После подстановки (2.9.20) и (2.9.21) в (2.9.19) система уравнений для определения вершин примет вид

$$\begin{bmatrix}
 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
 -b & b & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
 0 & a_{32} & a_{33} & a_{34} & 0 & \dots & 0 & 0 & 0 & 0 \\
 0 & 0 & a_{43} & a_{44} & a_{45} & \dots & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & a_{54} & a_{55} & \dots & 0 & 0 & 0 & 0 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 0 & 0 & 0 & 0 & 0 & \dots & a_{n-1n-1} & a_{n-1n} & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & \dots & a_{nn-1} & a_{nn} & a_{nn+1} & 0 \\
 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & -c & c \\
 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1
 \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ \dots \\ p_{n-1} \\ p_n \\ p_{n+1} \\ p_{n+2} \end{bmatrix} = \begin{bmatrix} a_1 \\ a'_1 \\ a_2 \\ a_3 \\ a_4 \\ \dots \\ a_{n-2} \\ a_{n-1} \\ a'_n \\ a_n \end{bmatrix}, \tag{2.9.22}$$

где

$$\begin{aligned}
 b &= \frac{3}{x_2 - x_1}, \quad c = \frac{3}{x_n - x_{n-1}}, \\
 a_{kk-1} &= \frac{x_k - x_{k-1}}{x_k - x_{k-3}} \frac{x_k - x_{k-1}}{x_k - x_{k-2}}, \\
 a_{kk} &= \frac{x_{k-1} - x_{k-2}}{x_{k+1} - x_{k-2}} \frac{x_{k+1} - x_{k-1}}{x_k - x_{k-2}} + \frac{x_k - x_{k-1}}{x_k - x_{k-3}} \frac{x_{k-1} - x_{k-3}}{x_k - x_{k-2}}, \\
 a_{kk+1} &= \frac{x_{k-1} - x_{k-2}}{x_{k+1} - x_{k-2}} \frac{x_{k-1} - x_{k-2}}{x_k - x_{k-2}}.
 \end{aligned}$$

Матрица системы уравнений (2.9.22) является трехдиагональной.

Для замкнутого кубического сплайна система уравнений (2.9.19) упрощается, так как в ней исчезнут второе и предпоследнее уравнения, определяющие производные кривой на краях. Число вершин p_i , $i = 1, 2, \dots, n$, NURBS кривой будет равно числу заданных точек a_i , $i = 1, 2, \dots, n$. Последовательность узлов будет иметь вид $t_i = x_{i-2}$, $i = 1, 2, \dots, n$ (плюс циклическое расширение).

кривой t изменяется в пределах $t_{\min} \leq t \leq t_{\max}$. Усеченную кривую определим как часть базовой кривой, начинающейся при параметре t_{begin} и оканчивающейся при параметре t_{end} , где $t_{\min} \leq t_{\text{begin}} \leq t_{\max}$ и $t_{\min} \leq t_{\text{end}} \leq t_{\max}$. Направление усеченной кривой может совпадать с направлением базовой кривой или быть ему противоположным, например, при $t_{\text{end}} < t_{\text{begin}}$. Если кривая замкнута, то движение от точки t_{begin} к точке t_{end} можно выполнить двумя способами: в положительном направлении базовой кривой и в противоположном направлении. Чтобы преодолеть эту неоднозначность для замкнутых кривых, в данные усеченной кривой вводится параметр sign , характеризующий совпадение ее направления с направлением базовой кривой и принимающий значения $+1$ или -1 . Параметру базовой кривой t_{begin} соответствует параметр усеченной кривой $w_{\min} = 0$, параметру базовой кривой t_{end} соответствует параметр усеченной кривой $w_{\max} = s$, где $s = s(t_{\text{begin}}, t_{\text{end}})$ — параметрическое расстояние между t_{end} и t_{begin} с учетом замкнутости кривой. Если кривая не замкнута, то $s = |t_{\text{end}} - t_{\text{begin}}|$. Радиус-вектор усеченной кривой описывается формулой

$$\mathbf{r}(w) = \mathbf{r}_b(t_{\text{begin}} + w \cdot \text{sign}), \quad 0 \leq w \leq s(t_{\text{begin}}, t_{\text{end}}), \quad (2.10.1)$$

где $\mathbf{r}_b(t)$, $t_{\min} \leq t \leq t_{\max}$, — базовая кривая, $s(t_{\text{begin}}, t_{\text{end}})$ — расстояние между параметрами усечения t_{begin} , t_{end} базовой кривой.

Эквидистантная кривая. Эквидистантная линия описывается радиус-вектором

$$\mathbf{r}(t) = \mathbf{r}_b(t) + \mathbf{a} \times \mathbf{t}_b(t), \quad t_{\min} \leq t \leq t_{\max}, \quad (2.10.2)$$

где $\mathbf{r}_b(t)$ — базовая кривая, $\mathbf{t}_b(t) = \frac{\mathbf{r}_b'}{\sqrt{\mathbf{r}_b' \cdot \mathbf{r}_b'}}$ — единичный касательный вектор к базовой кривой в данной точке, \mathbf{a} — заданный вектор. Область изменения параметра эквидистантной кривой совпадает с областью изменения параметра базовой кривой. Эквидистантная кривая оправдывает свое название, если $\mathbf{r}_b(t)$ — плоская кривая, а вектор \mathbf{a} ортогонален плоскости базовой кривой. В этом случае

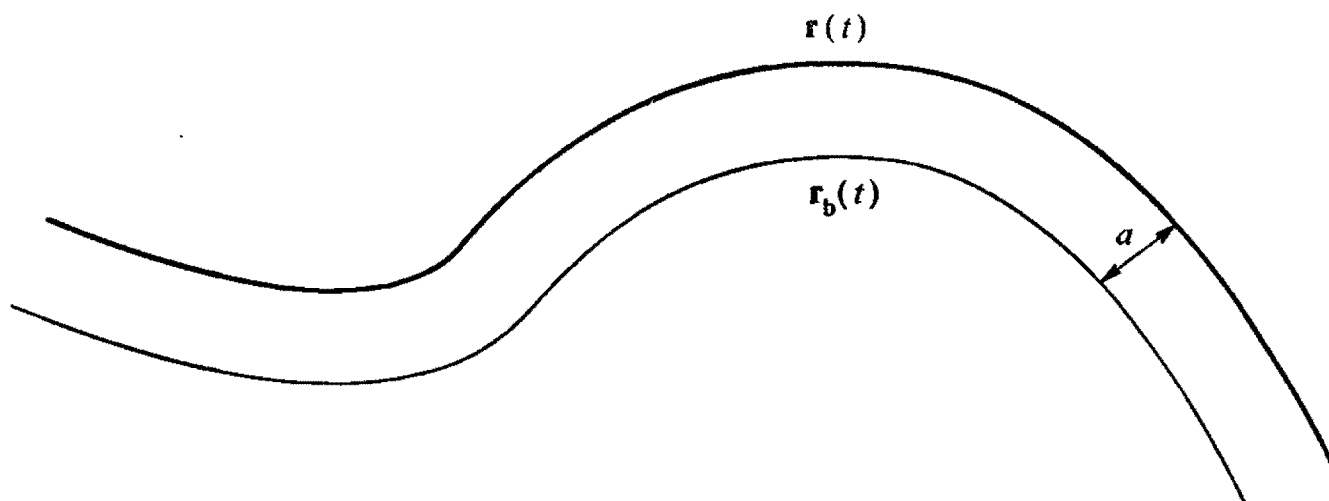


Рис. 2.10.1. Эквидистантная кривая

второе слагаемое в правой части (2.10.2) есть вектор, который лежит в плоскости базовой кривой, ортогонален ей и имеет длину вектора \mathbf{a} . В результате получим кривую, каждая точка которой отстоит по нормали от соответствующей точки базовой кривой на одинаковом расстоянии (рис. 2.10.1).

Производные эквидистантной линии описываются вектором

$$\mathbf{r}'(t) = \mathbf{r}_b'(t) + \mathbf{a} \times \mathbf{t}_b'(t),$$

где $\mathbf{t}_b' = \frac{\mathbf{r}_b''}{s'} - \mathbf{r}_b' \frac{\mathbf{r}' \cdot \mathbf{r}''}{s'^3}$ — производная касательного вектора базовой кривой, $s' = \sqrt{\mathbf{r}' \cdot \mathbf{r}'}$ — производная длины дуги.

Ссылочная кривая. Ссылочная кривая представляет собой линию, каждая точка которой получена путем некоторого преобразования соответствующей точки базовой кривой. Ссылочная кривая описывается радиус-вектором

$$\mathbf{r}(t) = \mathbf{M}_t \cdot \mathbf{r}_b(t), \quad t_{\min} \leq t \leq t_{\max}, \quad (2.10.3)$$

где $\mathbf{r}_b(t)$ — базовая кривая, \mathbf{M}_t — расширенная матрица преобразования (1.4.5) базовой кривой. Область изменения параметра ссылочной кривой совпадает с областью изменения параметра базовой кривой.

Репараметризованная кривая. К линиям, построенным на базе линий, можно отнести кривую с измененной областью параметра. Пусть требуется, чтобы кривая $\mathbf{r}_b(t)$, $t_{\min} \leq t \leq t_{\max}$, имела область определения параметра $w_{\min} \leq w \leq w_{\max}$. В этом случае можно построить репараметризованную кривую

$$\mathbf{r}(w) = \mathbf{r}_b(t(w)), \quad w_{\min} \leq w \leq w_{\max}, \quad (2.10.4)$$

где

$$t(w) = t_{\min} \frac{w_{\max} - w}{w_{\max} - w_{\min}} + t_{\max} \frac{w - w_{\min}}{w_{\max} - w_{\min}},$$

которая полностью совпадает с кривой $\mathbf{r}_b(t)$, но имеет другую область определения параметра. Линия с измененной длиной параметра применяется для согласования областей изменения параметра двух кривых, лежащих в основе кривой пересечения при построении ребер тел.

Продолженная кривая. Произвольную линию можно не только усекать, но и продлевать требуемым образом. Любая кривая может быть продолжена (или усечена) на заданное параметрическое расстояние. Пусть требуется продолжить кривую $\mathbf{r}_b(t)$, $t_{\min} \leq t \leq t_{\max}$ путем расширения области определения параметров до $a + t_{\min} \leq t \leq t_{\max} + b$. При $a < 0$ и $b > 0$ кривая продлевается за свои пределы, при $a > 0$ и $b < 0$ кривая усекается. Если кривая является замкнутой, то при выходе параметра за границу области определения выполним его циклический пересчет

$$\mathbf{r}(t) = \begin{cases} \mathbf{r}_b(t + t_{\Delta}), & \text{если } t < t_{\min}, \\ \mathbf{r}_b(t), & \text{если } t_{\min} \leq t \leq t_{\max}, \\ \mathbf{r}_b(t - t_{\Delta}), & \text{если } t_{\max} < t, \end{cases} \quad (2.10.5)$$

где $t_{\Delta} = t_{\max} - t_{\min}$. Если же кривая не является замкнутой, а ее параметр вышел за границу области определения, то продолжим кривую по касательной, которую она имела на соответствующем конце, и вычислим по продленной кривой необходимые геометрические характеристики. Радиус-вектор продолженной

кривой вычислим по формуле

$$\mathbf{r}(t) = \begin{cases} \mathbf{r}_b(t_{\min}) + (t - t_{\min}) \left. \frac{d\mathbf{r}_b}{dt} \right|_{t_{\min}}, & \text{если } t < t_{\min}, \\ \mathbf{r}_b(t), & \text{если } t_{\min} \leq t \leq t_{\max}, \\ \mathbf{r}_b(t_{\max}) + (t - t_{\max}) \left. \frac{d\mathbf{r}_b}{dt} \right|_{t_{\max}}, & \text{если } t_{\max} < t. \end{cases} \quad (2.10.6)$$

Дифференцируя формулы (2.10.5) и (2.10.6), получим производные радиус-вектора продолженной кривой. По формулам (2.10.5) или (2.10.6) может вычисляться геометрическая информация соответственно замкнутой или незамкнутой кривой при выходе ее параметра за область определения.

Общее правило. Все кривые, в данных которых лежит другая кривая, не должны допускать многократного наследования своего же типа. Например, в качестве базовой кривой для параметрически усеченной кривой не должна быть использована другая параметрически усеченная кривая, а должна быть использована базовая кривая последней с соответствующим пересчетом параметров усечения. Аналогичные правила должны действовать и для других базирующихся кривых. Если нужно построить эквидистантную линию на базе другой эквидистантной кривой $\mathbf{r}'(t) = \mathbf{r}_b(t) + \mathbf{a}' \times \mathbf{t}_b(t)$, то в качестве базовой линии должна быть использована базовая кривая последней $\mathbf{r}_b(t)$, а вектор эквидистанты должен быть равен $\mathbf{a} + \mathbf{a}'$. Если требуется построить ссылочную кривую на базе другой ссылочной кривой $\mathbf{r}'(t) = \mathbf{M}_t' \cdot \mathbf{r}_b(t)$, то в качестве базовой линии должна быть использована базовая кривая последней, а матрицу преобразования получим как произведение матриц $\mathbf{M}_t \cdot \mathbf{M}_t'$.

2.11. Составные кривые

Из нескольких кривых можно построить *составную кривую* — наиболее общий тип кривой. Составные кривые мы уже строили выше. Они характерны тем, что первая производная радиус-вектора в точках стыковки терпит разрыв или по длине или по длине и направлению. Рассмотренные выше сплайновые составные кривые формировались из кривых одного типа, и в них в общем случае первая производная радиус-вектора терпела разрыв по длине. Но составные кривые можно формировать и из кривых разной природы. Кривые, образующие составную кривую, будем называть *сегментами*. При построении составной кривой должны быть выполнены определенные условия: начало каждого последующего сегмента должно совпадать с концом предыдущего сегмента. Если сегменты составной кривой стыкуются не гладко, то составная кривая будет иметь изломы. В общем случае в местах стыка сегментов производные составной кривой терпят разрыв по длине и направлению. Составная кривая приведена на рис. 2.11.1.

Замкнутую составную кривую будем называть *контуром*. Для контура начало первого сегмента должно совпадать с концом последнего сегмента.

Пусть составная кривая содержит n сегментов

$$\mathbf{r}_i(w_i), \quad w_{i \min} \leq w_i \leq w_{i \max}, \quad i = 1, 2, \dots, n. \quad (2.11.1)$$

Начальное значение параметра t составной кривой положим равным нулю. Параметрическую длину составной кривой положим равной сумме параметрических длин составляющих его кривых

$$t_{\min} = 0, \quad t_{\max} = \sum_{i=1}^n (w_{i \max} - w_{i \min}).$$

При вычислении радиус-вектора составной кривой сначала необходимо определить тот сегмент, на который попадем, двигаясь по параметру. Затем нужно определить соответствующее значение собственного параметра этого сегмента

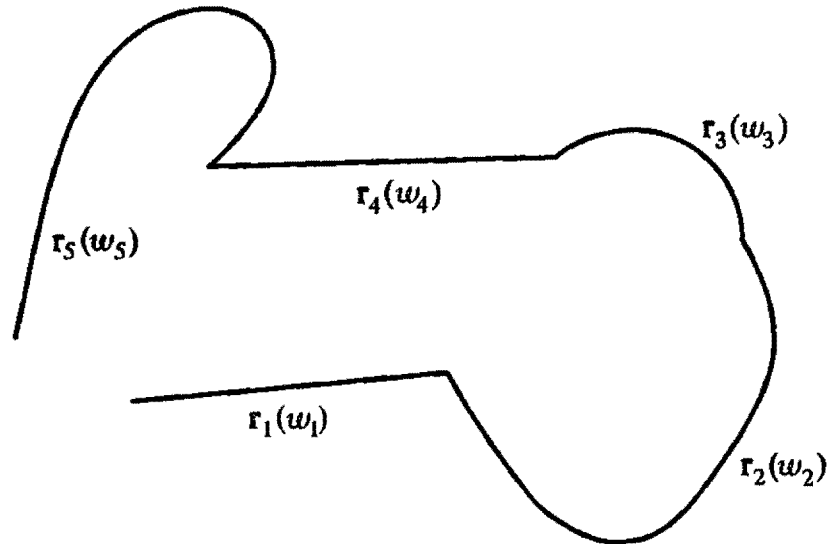


Рис. 2.11.1. Составная кривая

и с помощью его значения вычислить радиус-вектор сегмента или его производные. Пусть для параметра t составной кривой найден номер сегмента k , для которого выполняется соотношение

$$\sum_{i=1}^{k-1} (w_{i \max} - w_{i \min}) \leq t < \sum_{i=1}^k (w_{i \max} - w_{i \min}).$$

Тогда в соответствии со сказанным радиус-вектор составной кривой определяется равенством

$$\mathbf{r}(t) = \mathbf{r}_k \left(w_{k \min} + t - \sum_{i=1}^{k-1} (w_{i \max} - w_{i \min}) \right) = \mathbf{r}_k(w_k), \quad 0 \leq t \leq t_{\max}, \quad (2.11.2)$$

где параметр k -го сегмента равен

$$w_k = w_{k \min} + t - \sum_{i=1}^{k-1} (w_{i \max} - w_{i \min}). \quad (2.11.3)$$

Как можно заметить, подход к вычислению радиус-вектора составной кривой аналогичен подходу к вычислению радиус-вектора ломаной линии.

Общее правило. Составная кривая является линией, базирующейся на других линиях. Для нее должно выполняться требование о недопустимости многократного наследования своего же типа, предъявленное к кривым в предыдущем параграфе. В качестве сегментов составной кривой не должны использоваться другие составные кривые, но это не означает, что нельзя построить составную кривую из набора составных кривых. Если составную кривую нужно построить на основе других составных кривых, то последние должны рассматриваться

как совокупность кривых, а не как единые линии. Задача построения составной кривой из набора составных кривых должна решаться путем формирования кривой составляющими сегментами.

2.12. Двухмерные кривые

Все рассмотренные в этой главе кривые (за исключением цилиндрической спирали), могут строиться как в трехмерном пространстве, так и в двухмерном (на плоскости). В последнем случае их радиус-вектор имеет две координаты и преобразуется по двухмерным матрицам. Выражения для вычисления радиус-вектора двухмерных кривых полностью совпадают с соответствующими выражениями для трехмерных кривых. Для построения двухмерных кривых единственное, что нужно сделать, это заменить трехмерные векторы $\mathbf{r}(t) = [r_1(t) \ r_2(t) \ r_3(t)]^T$ на двухмерные векторы $\mathbf{r}(t) = [x(t) \ y(t)]^T$. Например, аналитические двухмерные кривые описываются выражениями вида

$$\mathbf{c}(t) = \mathbf{p} + x(t)\mathbf{i}_1 + y(t)\mathbf{i}_2, \quad (2.12.1)$$

где \mathbf{p} — начало двухмерной декартовой прямоугольной системы координат, \mathbf{i}_1 и \mathbf{i}_2 — орты этой системы.

Другой пример — двухмерные NURBS кривые описываются формулой, подобной (2.9.1):

$$\mathbf{r}(t) = \frac{\sum_{i=1}^n N_{i,m}(t)w_i\mathbf{p}_i}{\sum_{i=1}^n N_{i,m}(t)w_i}, \quad t_{\min} \leq t \leq t_{\max},$$

где пространственные точки характеристической ломаной заменены на двухмерные точки \mathbf{p}_i . Подобным образом строятся и другие двухмерные аналоги пространственных кривых вплоть до эквидистантных, усеченных, продолженных и составных кривых.

Двухмерные кривые будут использоваться для нескольких целей: для построения плоских пространственных кривых, для построения кривых на поверхностях, для построения линий (кривых) пересечения поверхностей и для описания области определения параметров поверхности.

Плоские пространственные кривые. Плоские пространственные кривые будем представлять в виде совокупности плоскости $\mathbf{r}(x, y) = \mathbf{p} + x\mathbf{i}_x + y\mathbf{i}_y$ и двухмерной кривой $\mathbf{c}(t) = [x(t) \ y(t)]^T$. Трехмерные векторы \mathbf{i}_x и \mathbf{i}_y являются базисными векторами некоторой местной системы координат. Плоская кривая является пространственной кривой и описывается радиус-вектором

$$\mathbf{c}(t) = \mathbf{p} + x(t)\mathbf{i}_x + y(t)\mathbf{i}_y. \quad (2.12.2)$$

Данный способ уже использовался для описания кривых второго порядка и других аналитических кривых. Теперь распространим его на сплайновые кривые, кривые Безье, NURBS кривые и другие. В качестве кривых $\mathbf{c}(t)$ могут использоваться и составные кривые.

Кривые на поверхности. Кривые на поверхности $\mathbf{r}(u, v)$ будем описывать совокупностью двухмерной кривой $\mathbf{c}_{uv}(t) = [u(t) \ v(t)]^T$ и поверхности. Двухмерная кривая и поверхность входят в структуру данных кривой на поверхности.

функцией

$$\mathbf{r}(u, v) = \mathbf{p} + av \operatorname{ch} u \mathbf{i}_x + bv \operatorname{sh} u \mathbf{i}_y + v^2 \mathbf{i}_z, \quad (3.2.14)$$

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

то она будет описывать только часть поверхности.

Скалярные функции $x(u, v)$, $y(u, v)$, $z(u, v)$ обоих представлений гиперболического параболоида связаны уравнением

$$\left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 = z. \quad (3.2.15)$$

Сечения параболоида плоскостями, ортогональными векторам \mathbf{i}_x и \mathbf{i}_y , являются параблами, сечения параболоида плоскостями, ортогональными вектору \mathbf{i}_z ,

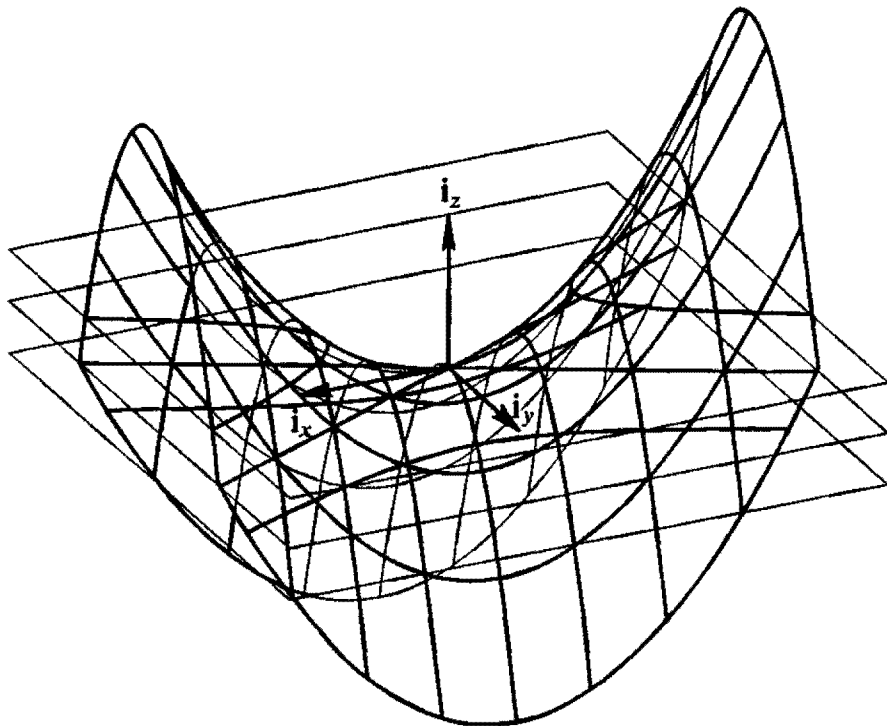


Рис. 3.2.6. Гиперболический параболоид

являются гиперболами. Сечение гиперболического параболоида плоскостью, ортогональной вектору \mathbf{i}_z и проходящей через точку \mathbf{p} , представляет собой две

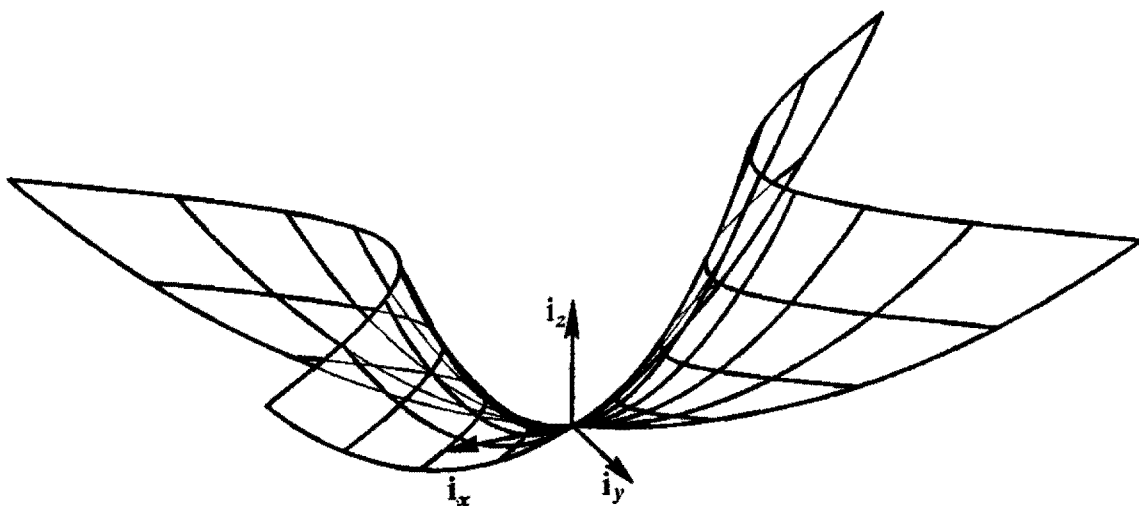


Рис. 3.2.7. Часть гиперболического параболоида

прямые линии. Гиперболический параболоид (3.2.13) и его сечения плоскостями, ортогональными вектору \mathbf{i}_z , показаны на рис. 3.2.6. Гиперболический параболоид (3.2.14) показан на рис. 3.2.7.

Формулы (3.2.6)–(3.2.13) описывают ограниченные гиперboloиды и параболоиды, усеченные параметрами v_{\min} и v_{\max} или u_{\min} и u_{\max} . Если в формулах (3.2.4), (3.2.6) и (3.2.8) $a = b$, то мы получим соответствующие поверхности вращения вокруг местной оси i_z . Частными случаями поверхности второго порядка являются цилиндрическая и коническая поверхности.

Цилиндр. *Эллиптический цилиндр* может быть описан формулой

$$\begin{aligned} \mathbf{r}(u, v) &= \mathbf{p} + a \cos u \mathbf{i}_x + b \sin u \mathbf{i}_y + hv \mathbf{i}_z, \\ 0 \leq u \leq 2\pi, \quad 0 \leq v \leq 1, \end{aligned} \quad (3.2.16)$$

где a и b — полуоси эллипса, являющегося поперечным сечением цилиндрической поверхности, h — длина цилиндра. Если $a = b$, то мы получим круговой цилиндр. Цилиндрическая поверхность (3.2.16) является замкнутой по параметру u и усеченной по параметру v .

Конус. *Круговая коническая поверхность* может быть описана формулой

$$\begin{aligned} \mathbf{r}(u, v) &= \mathbf{p} + (r + hv \operatorname{tg} \gamma)(\cos u \mathbf{i}_x + \sin u \mathbf{i}_y) + hv \mathbf{i}_z, \\ 0 \leq u \leq 2\pi, \quad 0 \leq v \leq 1, \end{aligned} \quad (3.2.17)$$

где r — радиус одного из оснований конуса, h — длина конуса, γ — угол между образующей и осью конуса. Коническая поверхность (3.2.17), как и цилиндрическая, является замкнутой по параметру u и усеченной по параметру v .

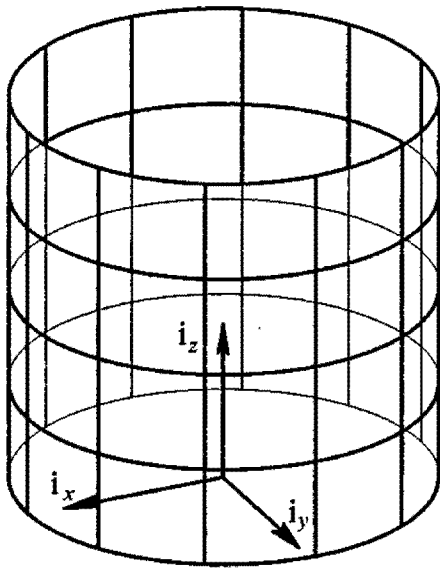


Рис. 3.2.8. Цилиндрическая поверхность

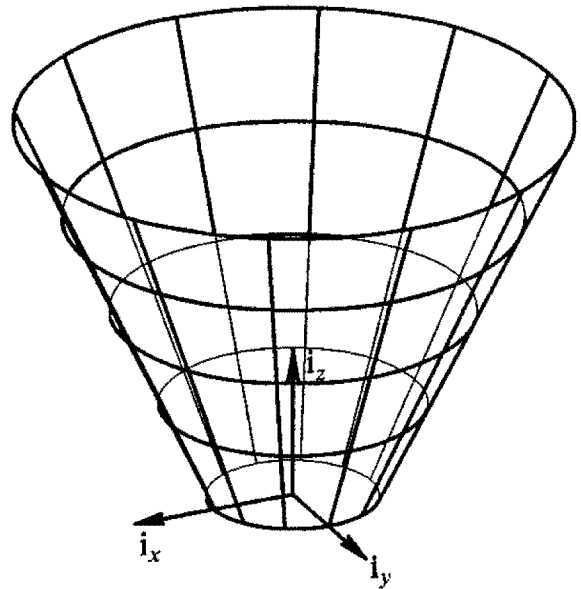


Рис. 3.2.9. Коническая поверхность

Круговой цилиндр и круговой конус показаны на рис. 3.2.8 и 3.2.9.

Тор. *Поверхность тора* может быть описана положением центра \mathbf{p} , тремя взаимно ортогональными векторами единичной длины i_x , i_y , i_z , определяющими положение и ориентацию местной декартовой системы координат, главным радиусом R и малым радиусом r

$$\begin{aligned} \mathbf{r}(u, v) &= \mathbf{p} + (R + r \cos v) \cos u \mathbf{i}_x + (R + r \cos v) \sin u \mathbf{i}_y + r \sin v \mathbf{i}_z, \\ 0 \leq u \leq 2\pi, \quad -\pi \leq v \leq \pi. \end{aligned} \quad (3.2.18)$$

Тороидальная поверхность (3.2.18) является замкнутой по параметру u и по параметру v . Скалярные функции $x(u, v) = (R + r \cos v) \cos u$, $y(u, v) = (R + r \cos v) \sin u$, $z(u, v) = r \sin v$ тора связаны уравнением

$$(\sqrt{x^2 + y^2} - R)^2 + z^2 = r^2.$$

Область изменения параметра v тороидальной поверхности (3.2.18) записана в предположении, что $r < R$. Если $r > R$, то для того, чтобы тор не пересекал сам себя, нужно уменьшить область изменения параметра v : $-\pi + v_0 \leq v \leq \pi - v_0$, где $v_0 = \arccos(R/r)$. При $R = 0$ получим $v_0 = \pi/2$ и тор превратится в

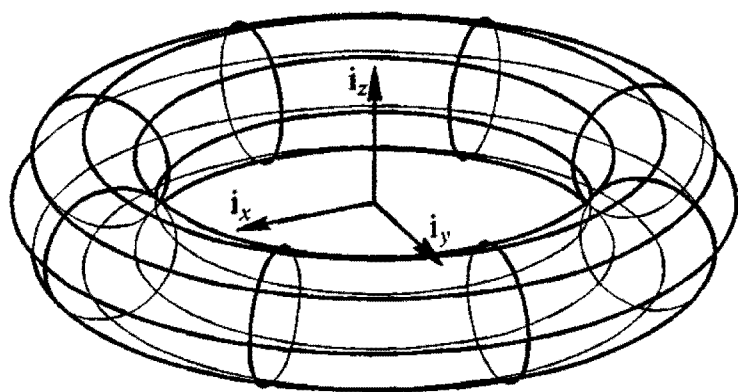


Рис. 3.2.10. Открытый тор

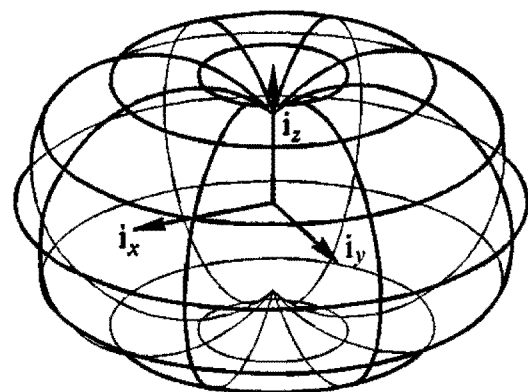


Рис. 3.2.11. Закрытый тор

сферу радиуса r . Тороидальная поверхность называется открытой, если $r < R$, и называется закрытой, если $r > R$. Открытый и закрытый торы показаны на рис. 3.2.10 и 3.2.11.

Существуют еще ряд поверхностей, которые можно описать аналитическими параметрическими зависимостями. Производные радиус-вектора аналитических поверхностей можно найти дифференцированием компонент радиус-вектора по параметрам.

3.3.* Поверхности второго порядка

Эллипсоид, параболоиды, гиперболоиды, цилиндр и конус являются представителями поверхностей второго порядка. Выше приведены параметрические зависимости для их радиус-вектора. В некоторых случаях, например, при обмене данными, приходится иметь дело с неявным описанием поверхностей. Возникает необходимость из неявного представления поверхности получить параметрическое представление и наоборот. Аналитические поверхности можно представить в виде уравнения, которому удовлетворяют координаты x, y, z радиус-вектора $\mathbf{r}(x, y, z) = x\mathbf{i}_x + y\mathbf{i}_y + z\mathbf{i}_z$ поверхности. Поверхности второго порядка в декартовой системе координат $Oxyz$ описываются уравнением второй степени

$$c_{11}x^2 + c_{22}y^2 + c_{33}z^2 + 2c_{12}xy + 2c_{23}yz + 2c_{13}xz + 2c_{14}x + 2c_{24}y + 2c_{34}z + c_{44} = 0, \quad (3.3.1)$$

где $c_{11}, c_{22}, c_{33}, c_{44}, c_{12}, c_{13}, c_{23}, c_{14}, c_{24}, c_{34}$ — коэффициенты, определяющие тип поверхности, ее положение и ориентацию в пространстве (считается, что хотя бы один из коэффициентов $c_{11}, c_{22}, c_{33}, c_{12}, c_{13}, c_{23}$ отличен от нуля). В отличие от параметрического такое описание поверхности не всегда однозначно, а также не является инвариантным относительно преобразования координат. При перемещении или повороте в пространстве поверхности второго порядка коэффициенты $c_{11}, c_{22}, c_{33}, c_{44}, c_{12}, c_{13}, c_{23}, c_{14}, c_{24}, c_{34}$ не остаются неизменными, причем трудно сразу сказать, изменена ли поверхность или только ее положение.

В расширенной матричной записи уравнение поверхности второго порядка (3.3.1) имеет вид

$$\mathbf{R}^T \cdot \mathbf{C}_c \cdot \mathbf{R} = 0, \quad (3.3.2)$$

где использованы расширенный вектор (1.4.3)

$$\mathbf{R} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

и расширенная матрица \mathbf{C}_c :

$$\mathbf{C}_c = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{12} & c_{22} & c_{23} & c_{24} \\ c_{13} & c_{23} & c_{33} & c_{34} \\ c_{14} & c_{24} & c_{34} & c_{44} \end{bmatrix}. \quad (3.3.3)$$

Для уравнения (3.3.1) поверхности второго порядка величины

$$I_1 = c_{11} + c_{22} + c_{33}, \quad I_2 = \begin{vmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{vmatrix} + \begin{vmatrix} c_{22} & c_{23} \\ c_{23} & c_{33} \end{vmatrix} + \begin{vmatrix} c_{33} & c_{13} \\ c_{13} & c_{11} \end{vmatrix}, \quad (3.3.4)$$

$$I_3 = \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{vmatrix}, \quad I_4 = \begin{vmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{12} & c_{22} & c_{23} & c_{24} \\ c_{13} & c_{23} & c_{33} & c_{34} \\ c_{14} & c_{24} & c_{34} & c_{44} \end{vmatrix}$$

являются инвариантами относительно преобразований (1.2.5) параллельного переноса и поворота осей декартовой системы координат. Эти преобразования координат описываются формулами

$$\begin{aligned} x &= a_{11}x' + a_{21}y' + a_{31}z' + x_0, \\ y &= a_{12}x' + a_{22}y' + a_{23}z' + y_0, \\ z &= a_{13}x' + a_{23}y' + a_{33}z' + z_0. \end{aligned} \quad (3.3.5)$$

В расширенной матричной записи преобразование координат (3.3.5) имеет вид

$$\mathbf{R} = \mathbf{A}_0^T \cdot \mathbf{R}',$$

где \mathbf{A}_0 — расширенная матрица:

$$\mathbf{A}_0 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ x_0 & y_0 & z_0 & 1 \end{bmatrix}. \quad (3.3.6)$$

Заметим, что определитель матрицы \mathbf{A}_0 равен единице. После перехода к координатам x', y', z' получим уравнение поверхности второго порядка

$$\mathbf{R}'^T \cdot \mathbf{A}_0 \cdot \mathbf{C}_c \cdot \mathbf{A}_0^T \cdot \mathbf{R}' = 0 \quad \text{или} \quad \mathbf{R}'^T \cdot \mathbf{C}_c' \cdot \mathbf{R}' = 0, \quad \text{где} \quad \mathbf{C}_c' = \mathbf{A}_0 \cdot \mathbf{C}_c \cdot \mathbf{A}_0^T,$$

откуда следует, что I_4 есть инвариант относительно преобразования (3.3.5). Действительно,

$$I_4' \equiv |\mathbf{C}_c'| = |\mathbf{A}_0 \cdot \mathbf{C}_c \cdot \mathbf{A}_0^T| = |\mathbf{A}_0| |\mathbf{C}_c| |\mathbf{A}_0^T| = |\mathbf{C}_c| \equiv I_4.$$

Доказательство инвариантности величин I_1, I_2, I_3 может быть выполнено путем замены координат (3.3.5) в (3.3.1) и вычисления величин (3.3.4) аналогично доказательству инвариантности I_4 . Тип поверхности второго порядка и ее положение в пространстве определяются значениями инвариантов I_1, I_2, I_3, I_4 .

Информация о типе поверхности и ее параметрах содержится в характеристической квадратичной форме $c_{11}x^2 + c_{22}y^2 + c_{33}z^2 + 2c_{12}xy + 2c_{23}yz + 2c_{13}xz$. Предположим, что найдена система координат $O'x'y'z'$, в которой уравнение поверхности второго порядка имеет канонический (простейший) вид

$$\lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2 + I = 0. \quad (3.3.7)$$

В силу свойства инвариантов будем иметь равенства

$$I_1 = \lambda_1 + \lambda_2 + \lambda_3, \quad I_2 = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1, \quad I_3 = \lambda_1\lambda_2\lambda_3, \quad I_4 = \lambda_1\lambda_2\lambda_3 I.$$

Отсюда следует, что свободный член в каноническом уравнении равен $I = \frac{I_4}{\lambda_1\lambda_2\lambda_3} = \frac{I_4}{I_3}$.

Поверхность второго порядка может быть приведена к указанному каноническому виду, если $I_3 \neq 0$. Поверхности такого вида называются центральными. Для определения остальных канонических коэффициентов центральной поверхности второго порядка составим кубический полином

$$\begin{aligned} (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) &= \\ &= \lambda^3 - (\lambda_1 + \lambda_2 + \lambda_3)\lambda^2 + (\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1)\lambda - \lambda_1\lambda_2\lambda_3 = \lambda^3 - I_1\lambda^2 + I_2\lambda - I_3, \end{aligned}$$

корнями которого являются канонические коэффициенты $\lambda_1, \lambda_2, \lambda_3$. Так I_1, I_2, I_3, I_4 в любой системе координат имеют одни и те же значения, то следует вывод, что канонические коэффициенты $\lambda_1, \lambda_2, \lambda_3$ поверхности второго порядка являются корнями уравнения

$$\lambda^3 - I_1\lambda^2 + I_2\lambda - I_3 = 0, \quad (3.3.8)$$

которое называется *характеристическим уравнением*. В матричной записи характеристическое уравнение (3.3.8) имеет вид

$$\begin{vmatrix} c_{11} - \lambda & c_{12} & c_{13} \\ c_{12} & c_{22} - \lambda & c_{23} \\ c_{13} & c_{23} & c_{33} - \lambda \end{vmatrix} = 0. \quad (3.3.9)$$

Таким образом, канонические коэффициенты являются собственными значениями матрицы характеристической квадратичной формы поверхности второго порядка. Корни $\lambda_1, \lambda_2, \lambda_3$ характеристического уравнения (3.3.8) являются действительными, так как матрица характеристической квадратичной формы симметричная. Метод решения кубического уравнения приведен в гл. 8.

Если $I_3 = 0$, то один из корней характеристического уравнения равен нулю. Пусть это будет третий по счету корень, и пусть $I_4 \neq 0$. Тогда поверхность второго порядка может быть приведена к виду

$$\lambda_1 x'^2 + \lambda_2 y'^2 - 2Jz' = 0.$$

В силу свойства инвариантов для данного уравнения будем иметь равенства

$$I_1 = \lambda_1 + \lambda_2, \quad I_2 = \lambda_1\lambda_2, \quad I_3 = 0, \quad I_4 = -\lambda_1\lambda_2 J^2.$$

Отсюда следует, что $J^2 = -\frac{I_4}{\lambda_1\lambda_2} = -\frac{I_4}{I_2}$. Такая поверхность не является центральной и имеет действительные точки, если $\frac{I_4}{I_2} < 0$.

Если $I_3 = 0$ и $I_4 = 0$, то поверхность второго порядка может быть приведена к одному из видов:

$$\begin{aligned} \lambda_1 x'^2 + \lambda_2 y'^2 &= 1, & \text{если } I_2 \neq 0, \\ \lambda_1 x'^2 + 2y' &= 0, & \text{если } I_2 = 0. \end{aligned}$$

Поверхность в этих случаях представляет собой цилиндрические поверхности, сечения $z = \text{const}$ которых представляют кривые второго порядка.

Приведем классификацию поверхностей второго порядка. Пусть $I_4 \neq 0$, тогда поверхность второго порядка не вырождена и представляет собой:

- эллипсоид, если $I_1 I_3 > 0$, $I_2 > 0$, $I_3 \neq 0$, $I_4 < 0$,
- однополостный гиперболоид, если $I_1 I_3 < 0$, $I_2 > 0$, $I_3 \neq 0$, $I_4 > 0$
или $I_1 I_3 > 0$, $I_2 < 0$, $I_3 \neq 0$, $I_4 > 0$,
- двуполостный гиперболоид, если $I_1 I_3 < 0$, $I_2 > 0$, $I_3 \neq 0$, $I_4 < 0$
или $I_1 I_3 > 0$, $I_2 < 0$, $I_3 \neq 0$, $I_4 < 0$,
- эллиптический параболоид, если $I_3 = 0$, $I_4 < 0$,
- гиперболический параболоид, если $I_3 = 0$, $I_4 > 0$.

При $I_1 I_3 > 0$, $I_2 > 0$, $I_3 \neq 0$, $I_4 > 0$ уравнению (3.3.1) не удовлетворяет ни одна точка пространства и оно описывает мнимый эллипсоид. Если $I_4 = 0$, то уравнение (3.3.1) описывает:

- коническую поверхность, если $I_1 I_3 < 0$, $I_2 > 0$, $I_3 \neq 0$
или $I_1 I_3 > 0$, $I_2 < 0$, $I_3 \neq 0$,
- эллиптический цилиндр, если $I_2 > 0$, $I_3 = 0$,
- гиперболический цилиндр, если $I_2 < 0$, $I_3 = 0$,
- параболический цилиндр, если $I_2 = 0$, $I_3 = 0$,

при условии, что

$$\begin{vmatrix} c_{22} & c_{23} & c_{24} \\ c_{23} & c_{33} & c_{34} \\ c_{24} & c_{34} & c_{44} \end{vmatrix} + \begin{vmatrix} c_{11} & c_{13} & c_{14} \\ c_{13} & c_{33} & c_{34} \\ c_{14} & c_{34} & c_{44} \end{vmatrix} + \begin{vmatrix} c_{11} & c_{12} & c_{14} \\ c_{12} & c_{22} & c_{24} \\ c_{14} & c_{24} & c_{44} \end{vmatrix} \neq 0.$$

Если последнее условие не выполняется, то цилиндры вырождаются в плоскости.

Если поверхность второго порядка привести к главной центральной системе координат, то ее уравнение в ней примет канонический вид. Уравнения, которым удовлетворяют координатные функции $x(u, v)$, $y(u, v)$, $z(u, v)$ поверхностей (3.2.4)–(3.2.13), имеют канонический вид. Координаты x_0 , y_0 , z_0 центральной точки p поверхности определяются из системы уравнений

$$\begin{aligned} c_{11}x_0 + c_{12}y_0 + c_{31}z_0 + c_{14} &= 0, \\ c_{12}x_0 + c_{22}y_0 + c_{23}z_0 + c_{24} &= 0, \\ c_{13}x_0 + c_{23}y_0 + c_{33}z_0 + c_{34} &= 0. \end{aligned} \tag{3.3.10}$$

Отсюда

$$x_0 = -\frac{1}{I_3} \begin{vmatrix} c_{14} & c_{12} & c_{13} \\ c_{24} & c_{22} & c_{23} \\ c_{34} & c_{23} & c_{33} \end{vmatrix}, \quad y_0 = -\frac{1}{I_3} \begin{vmatrix} c_{11} & c_{14} & c_{13} \\ c_{12} & c_{24} & c_{23} \\ c_{13} & c_{34} & c_{33} \end{vmatrix}, \quad z_0 = -\frac{1}{I_3} \begin{vmatrix} c_{11} & c_{12} & c_{14} \\ c_{12} & c_{22} & c_{24} \\ c_{13} & c_{23} & c_{34} \end{vmatrix}.$$

Каждая поверхность второго порядка имеет три главные плоскости, как минимум две из которых являются плоскостями симметрии. Все они проходят через центральную точку $p = [x_0, y_0, z_0]^T$. Коэффициенты a_{ij} , $i, j = 1, 2, 3$, преобразования (3.3.5) уравнения (3.3.1) в главную центральную систему координат являются компонентами векторов

$\mathbf{i}_1 = [a_{11} \ a_{12} \ a_{13}]^T$, $\mathbf{i}_2 = [a_{21} \ a_{22} \ a_{23}]^T$, $\mathbf{i}_3 = [a_{31} \ a_{32} \ a_{33}]^T$, ортогональных главным плоскостям. Векторы \mathbf{i}_1 , \mathbf{i}_2 , \mathbf{i}_3 называются *собственными векторами*. Каждый из них соответствует одному из корней характеристического уравнения.

Пусть $\lambda_1 \neq \lambda_2 \neq \lambda_3$ (корни характеристического уравнения различные). В этом случае базисные векторы главной центральной системы координат могут быть найдены следующим образом. Для каждого λ_i составим систему уравнений

$$\begin{aligned} (c_{11} - \lambda_i)a_{i1} + c_{12}a_{i2} + c_{13}a_{i3} &= 0, \\ c_{12}a_{i1} + (c_{22} - \lambda_i)a_{i2} + c_{23}a_{i3} &= 0, \\ c_{13}a_{i1} + c_{23}a_{i2} + (c_{33} - \lambda_i)a_{i3} &= 0. \end{aligned} \quad (3.3.11)$$

Среди уравнений (3.3.11) одно является линейной комбинацией двух других, так как определитель (3.3.9) системы равен нулю. Далее вычислим определители

$$A_1 = \begin{vmatrix} c_{22} - \lambda_i & c_{23} \\ c_{23} & c_{33} - \lambda_i \end{vmatrix}, \quad A_2 = \begin{vmatrix} c_{33} - \lambda_i & c_{13} \\ c_{13} & c_{11} - \lambda_i \end{vmatrix}, \quad A_3 = \begin{vmatrix} c_{11} - \lambda_i & c_{12} \\ c_{12} & c_{22} - \lambda_i \end{vmatrix}.$$

По крайней мере, один из них будет отличен от нуля. Пусть $A_3 \neq 0$. Тогда компоненты соответствующего главного вектора найдем из системы уравнений

$$\begin{aligned} (c_{11} - \lambda_i)a_{i1} + c_{12}a_{i2} + c_{13}a_{i3} &= 0, \\ c_{12}a_{i1} + (c_{22} - \lambda_i)a_{i2} + c_{23}a_{i3} &= 0, \\ a_{i1}^2 + a_{i2}^2 + a_{i3}^2 &= 1. \end{aligned}$$

Для ее решения из первых двух уравнений выразим a_{i1} и a_{i2} через a_{i3} и подставим их в третье уравнение. В результате получим a_{i3} . Далее из первых двух уравнений найдем a_{i1} и a_{i2} . Вектор $\mathbf{i}_i = [a_{i1} \ a_{i2} \ a_{i3}]^T$ соответствует λ_i .

Пусть два из трех корней характеристического уравнения равны между собой, например, $\lambda_1 = \lambda_2 \neq \lambda_3$. В этом случае базисный вектор $\mathbf{i}_3 = [a_{31} \ a_{32} \ a_{33}]^T$ может быть найден описанным выше способом. Другими базисными векторами могут служить любые два ортогональные друг другу и вектору \mathbf{i}_3 орта, так как поверхность второго порядка будет поверхностью вращения.

Если $\lambda_1 = \lambda_2 = \lambda_3$, то поверхность второго порядка будет являться сферой (или мнимой сферой). Базисными векторами ее главной центральной системы координат могут служить любые три взаимно ортогональные орта.

В главной центральной системе координат уравнение поверхности (3.3.1) примет вид

$$\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + \frac{I_4}{I_3} = 0, \quad (3.3.12)$$

если $I_3 \neq 0$ и, следовательно, все корни характеристического уравнения отличны от нуля, или

$$\lambda_1 x^2 + \lambda_2 y^2 - 2z\sqrt{-\frac{I_4}{I_2}} = 0, \quad (3.3.13)$$

если $I_3 = 0$, следовательно, хотя бы один из корней характеристического уравнения равен нулю (штрих около новых координат опущен). Если $I_3 \neq 0$, то каноническое уравнение поверхности имеет симметричный вид и у поверхности есть центральная точка (эллипсоид и гиперболоиды). В противном случае поверхность не имеет центра (параболоиды).

Перенумеруем корни характеристического уравнения так, чтобы $\lambda_1 \geq \lambda_2 \geq \lambda_3$. По корням характеристического уравнения определяются параметры поверхностей второго порядка. Если $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$, то, сравнив (3.3.12) с (3.2.5), найдем полуоси

эллипсоида (3.2.4) по формулам

$$a = \sqrt{-\frac{I_4}{\lambda_3 I_3}}, \quad b = \sqrt{-\frac{I_4}{\lambda_2 I_3}}, \quad c = \sqrt{-\frac{I_4}{\lambda_1 I_3}}. \quad (3.3.14)$$

Если $\lambda_1 \geq \lambda_2 > 0 > \lambda_3$, то, сравнив (3.3.12) с (3.2.7), найдем действительные и мнимую полуоси однополостного гиперболоида (3.2.6) по формулам

$$a = \sqrt{-\frac{I_4}{\lambda_3 I_3}}, \quad b = \sqrt{-\frac{I_4}{\lambda_1 I_3}}, \quad c = \sqrt{\frac{I_4}{\lambda_2 I_3}}. \quad (3.3.15)$$

Если $\lambda_1 > 0 > \lambda_2 \geq \lambda_3$, то, сравнив (3.3.12) с (3.2.9), найдем действительную и мнимые полуоси двуполостного гиперболоида (3.2.8) по формулам

$$a = \sqrt{\frac{I_4}{\lambda_3 I_3}}, \quad b = \sqrt{\frac{I_4}{\lambda_2 I_3}}, \quad c = \sqrt{-\frac{I_4}{\lambda_1 I_3}}. \quad (3.3.16)$$

Если $\lambda_1 \geq \lambda_2 > \lambda_3 = 0$, то, сравнив (3.3.13) с (3.2.11), найдем полуоси эллиптического параболоида (3.2.10) по формулам

$$a = \sqrt{\frac{2}{\lambda_2} \sqrt{-\frac{I_4}{I_2}}}, \quad b = \sqrt{\frac{2}{\lambda_1} \sqrt{-\frac{I_4}{I_2}}}. \quad (3.3.17)$$

Если $\lambda_1 > \lambda_2 = 0 > \lambda_3$, то, сравнив (3.3.13) с (3.2.15), найдем действительную и мнимую полуоси гиперболического параболоида (3.2.13) по формулам

$$a = \sqrt{\frac{2}{\lambda_1} \sqrt{-\frac{I_4}{I_2}}}, \quad b = \sqrt{-\frac{2}{\lambda_3} \sqrt{-\frac{I_4}{I_2}}}. \quad (3.3.18)$$

Если $0 > \lambda_1 \geq \lambda_2 \geq \lambda_3$, то уравнение (3.3.1) описывает мнимый эллипсоид.

Коническая поверхность является вырожденным случаем гиперболоида. Она представляет собой предельный случай, к которому стремятся однополостный и двуполостный гиперболоиды. Цилиндрическая поверхность является вырожденным случаем эллиптического и гиперболического параболоидов. Характеристики вырожденных случаев поверхностей можно найти по общим формулам.

Приведенные формулы позволяют строить параметрические зависимости для поверхностей второго порядка по их координатным уравнениям. Для перехода от параметрического описания к неявному описанию нужно знать каноническое уравнение $f(x, y, z) = 0$, которым связаны координатные функции $x(u, v)$, $y(u, v)$, $z(u, v)$ поверхности. Местные координаты x, y, z связаны с глобальными координатами r_1, r_2, r_3 уравнениями (1.2.8), которые в принятых здесь обозначениях имеют вид

$$\begin{aligned} x &= a_{11}(r_1 - p_1) + a_{12}(r_2 - p_2) + a_{13}(r_3 - p_3), \\ y &= a_{21}(r_1 - p_1) + a_{22}(r_2 - p_2) + a_{23}(r_3 - p_3), \\ z &= a_{31}(r_1 - p_1) + a_{32}(r_2 - p_2) + a_{33}(r_3 - p_3). \end{aligned} \quad (3.3.19)$$

Для получения уравнения $F(r_1, r_2, r_3) = 0$, связывающего глобальные координаты, подставим в уравнение $f(x, y, z) = 0$ выражения (3.3.19) для канонических координат.

3.4. Поверхности движения

Поверхность можно получить путем движения кривой по заданной траектории. Такие поверхности называются поверхностями движения. Область изменения параметров таких поверхностей представляет собой прямоугольник. Кривая, движением которой получается поверхность, называется *образующей*,

а траектория движения некоторой ее точки называется *направляющей*. Поверхность движения в качестве данных содержит образующую кривую и в том или ином виде направляющую линию. В общем случае направляющей может быть произвольная кривая. Среди всевозможных направляющих выделяют две наиболее простые — отрезок прямой линии и дугу окружности. Если направляющей служит отрезок прямой, то поверхность называется поверхностью выдавливания. Если направляющей служит дуга окружности или вся окружность, то поверхность называется поверхностью вращения. Во всех остальных случаях поверхность будем называть кинематической поверхностью. Для построения тел мы не будем использовать поверхности, пересекающие сами себя.

Поверхность выдавливания. Поверхность выдавливания, полученная движением кривой $c(t)$, $t_{\min} \leq t \leq t_{\max}$, вдоль вектора d , описывается радиус-вектором

$$r(u, v) = c(u) + vd, \quad t_{\min} \leq u \leq t_{\max}, \quad 0 \leq v \leq 1. \quad (3.4.1)$$

Длина поверхности выдавливания определяется длиной вектора d . В зависимости от замкнутости кривой $c(t)$ поверхность выдавливания может быть замкнутой или нет по параметру u . Направляющий вектор $d = id$ может быть представлен в виде произведения вектора единичной длины на скаляр. Поверхность выдавливания приведена на рис. 3.4.1.

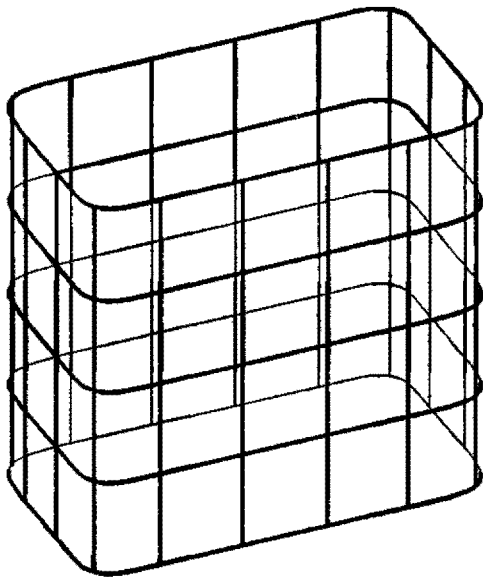


Рис. 3.4.1. Поверхность выдавливания

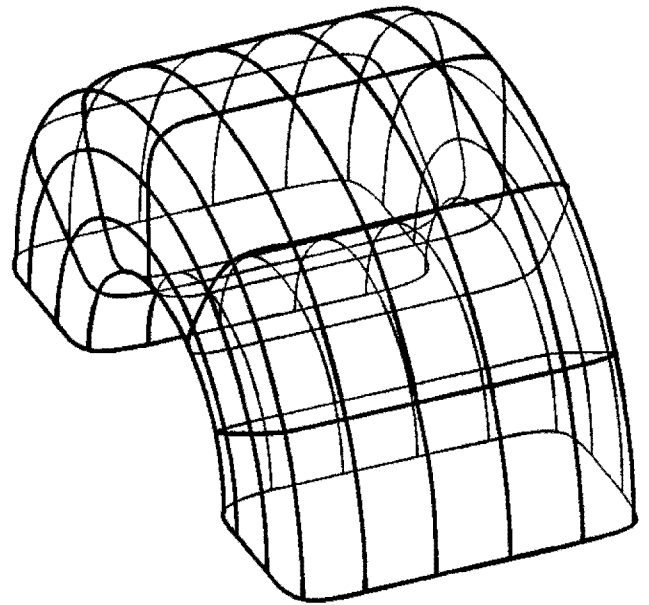


Рис. 3.4.2. Поверхность вращения

Поверхность вращения. Поверхность, полученная вращением кривой $c(t)$, $t_{\min} \leq t \leq t_{\max}$, на угол α вокруг оси, заданной единичным вектором i и точкой p , описывается радиус-вектором

$$r(u, v) = p + ((c(u) - p) \cdot i)i + (c(u) - p - ((c(u) - p) \cdot i)i) \cos v + \\ + i \times (c(u) - p) \sin v = p + r_3(u) + r_1(u) \cos v + r_2(u) \sin v, \quad (3.4.2) \\ t_{\min} \leq u \leq t_{\max}, \quad 0 \leq v \leq \alpha,$$

где $r_3(u) = ((c(u) - p) \cdot i)i$ — составляющая вектора $c(u) - p$, параллельная оси вращения, $r_1(u) = c(u) - p - r_3(u)$ — составляющая вектора $c(u) - p$, перпендикулярная оси вращения, $r_2(u) = i \times (c(u) - p)$ — ортогональный первым двум векторам вектор, длина которого равна длине вектора $r_1(u)$. Замкнутость поверхности вращения по параметру u совпадает с замкнутостью образующей

кривой $c(t)$. Если угол вращения $\alpha = 2\pi$, то поверхность (3.4.2) замкнута по параметру v , если $\alpha < 2\pi$, то поверхность вращения не замкнута по параметру v . Будем строить поверхности вращения, у которых угол вращения не превосходит 2π . Поверхность вращения приведена на рис. 3.4.2.

Кинематические поверхности. Рассмотрим кинематическую поверхность, которая является общим случаем поверхности движения. Пусть образующая описывается кривой $c(t)$, $t_{\min} \leq t \leq t_{\max}$, а направляющей является кривая $g(v)$, $v_{\min} \leq v \leq v_{\max}$. Пусть при построении кинематической поверхности параметр u совпадает с параметром образующей кривой $c(u)$, а параметр v совпадает с параметром направляющей кривой $g(v)$. При движении образующей кривой вдоль направляющей кривой ориентация первой относительно второй может меняться или может оставаться неизменной. Если образующая выполняет плоскопараллельное движение, т. е. остается «параллельной» своему начальному положению, то такую поверхность будем называть *поверхностью сдвига*, в противном случае поверхность будем называть *поверхностью заметания*.

Поверхность сдвига. Поверхность сдвига проще поверхности заметания. Радиус-вектор поверхности сдвига определяется формулой

$$\begin{aligned} \mathbf{r}(u, v) &= \mathbf{g}(v) + (\mathbf{c}(u) - \mathbf{g}(v_{\min}) - \mathbf{h}), \\ t_{\min} \leq u \leq t_{\max}, \quad v_{\min} \leq v \leq v_{\max}, \end{aligned} \quad (3.4.3)$$

где \mathbf{h} — вектор привязки образующей к направляющей. Вектор \mathbf{h} сдвигает кривую $c(u)$ на некоторую величину из ее начального положения относительно направляющей и сохраняет этот сдвиг во время движения. Если $\mathbf{h} = \mathbf{0}$, то сохраняется исходное положение образующей кривой относительно начальной точки направляющей кривой. Радиус-вектор поверхности сдвига строится как сумма двух векторов: вектора точки на направляющей кривой $g(v)$ и вектора положения точки образующей относительно начальной точки направляющей со сдвигом $c(u) - g(v_{\min}) - \mathbf{h}$. Поверхность сдвига приведена на рис. 3.4.3.

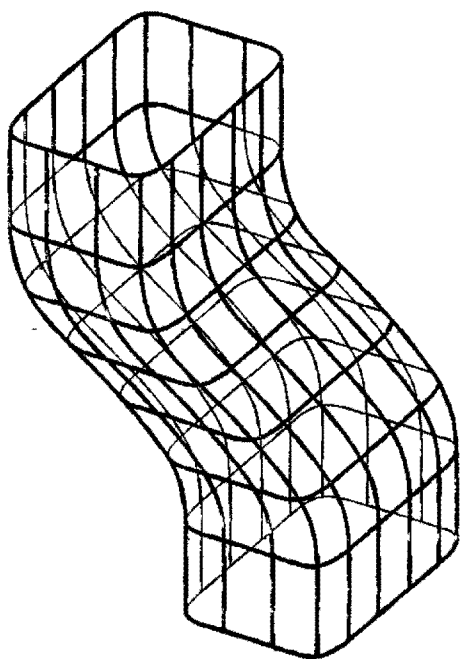


Рис. 3.4.3. Поверхность сдвига

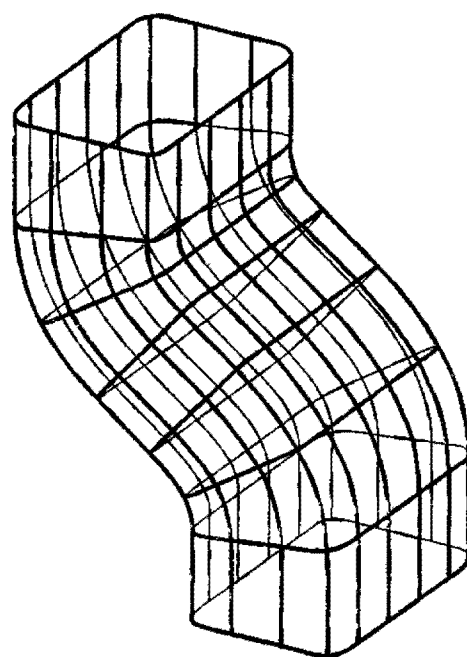


Рис. 3.4.4. Поверхность заметания

Поверхность заметания. Для построения кинематической поверхности заметания (рис. 3.4.4) будем использовать *подвижный касательный базис*, связанный с направляющей кривой $g(v)$. Подвижный касательный базис представляет

собой местную декартову прямоугольную систему координат. Ее начало находится в текущей точке направляющей кривой и определяется радиус-вектором $\mathbf{p}_0 = \mathbf{g}(v)$. Первый базисный орт $\mathbf{i}_1 = \mathbf{g}'/|\mathbf{g}'|$ местной системы координат совпадает с касательной к направляющей кривой, второй базисный орт \mathbf{i}_2 ортогонален первому, а третий $\mathbf{i}_3 = \mathbf{i}_1 \times \mathbf{i}_2$ образует с первым и вторым правую тройку векторов. Из сказанного не ясно, в какую сторону направлен вектор \mathbf{i}_2 . На первый взгляд, его можно было бы направить по вектору главной нормали к направляющей, но в точках распрямления кривой нормаль не определена, а в точках перегиба ее направление может резко поменяться на противоположное. Поэтому вектор \mathbf{i}_2 будем строить следующим образом: пусть \mathbf{d} есть некоторый не изменяющийся вектор, который не коллинеарен касательной к направляющей $\mathbf{g}(v)$ ни в одной ее точке. Тогда вектор \mathbf{i}_2 примем параллельным вектору

$$\mathbf{d}_2 = \mathbf{d} - (\mathbf{i}_1 \cdot \mathbf{d})\mathbf{i}_1. \quad (3.4.4)$$

Вектор \mathbf{d}_2 представляет собой составляющую вектора \mathbf{d} , перпендикулярную орту \mathbf{i}_1 (касательной к направляющей в данной точке). Составляющая \mathbf{d}_2 никогда не равна нулю, так как мы предполагаем, что \mathbf{d} никогда не коллинеарен \mathbf{i}_1 . Будем считать, что для любой кривой существует такой вектор \mathbf{d} , который никогда не коллинеарен касательной к кривой. Подвижный касательный базис связан с кривой $\mathbf{g}(v)$ и поэтому является функцией параметра v .

$$\mathbf{p}_0(v) = \mathbf{g}(v), \quad (3.4.5)$$

$$\mathbf{i}_1(v) = \frac{\mathbf{g}'}{|\mathbf{g}'|}, \quad \mathbf{i}_2(v) = \frac{\mathbf{d}_2}{|\mathbf{d}_2|}, \quad \mathbf{i}_3(v) = \mathbf{i}_1 \times \mathbf{i}_2,$$

где \mathbf{d}_2 определяется равенством (3.4.4), а $\mathbf{g}' = \frac{d\mathbf{g}}{dv}$.

Запомним положение образующей кривой $\mathbf{c}(u)$ в подвижном касательном базисе в начале направляющей (при начальном значении параметра $v = v_{\min}$) и будем сохранять его при движении вдоль направляющей. При движении вдоль направляющей подвижный касательный базис меняет свое положение и ориентацию в пространстве и увлекает за собой жестко связанную с ним образующую кривую. Таким образом, в данных кинематической поверхности находятся две кривые линии, вектор \mathbf{d} и положение касательного базиса (3.4.5) в начале направляющей: $\mathbf{g}(v_{\min})$, $\mathbf{i}_1(v_{\min})$, $\mathbf{i}_2(v_{\min})$, $\mathbf{i}_3(v_{\min})$. В общем случае можно запомнить положение образующей кривой относительно начальной точки направляющей, сдвинув первую на некоторый вектор \mathbf{h} (например, чтобы по направляющей двигался центр тяжести образующей), и сохранить его при движении вдоль направляющей. Если $\mathbf{h} = \mathbf{0}$, то сохраняется исходное положение образующей кривой относительно начальной точки направляющей кривой и касательной к ней. Радиус-вектор поверхности для произвольных параметров u и v получим следующим способом. Вычислим положение точки на образующей кривой в местной системе координат при $v = v_{\min}$ по формуле (1.2.9)

$$\mathbf{x}(u, v_{\min}) = \mathbf{A}(v_{\min}) \cdot (\mathbf{c}(u) - \mathbf{g}(v_{\min}) - \mathbf{h}),$$

где через $\mathbf{A}(v_{\min})$ обозначена матрица, вычисленная по матричной функции $\mathbf{A}(v)$ при начальном параметре v . Матричная функция $\mathbf{A}(v)$ является матрицей преобразования (1.2.6) координат радиус-вектора точки при переходе из глобальной в местную систему координат, ее строки составлены из компонент

базисных ортов местной системы координат

$$\mathbf{A}(v) = \begin{bmatrix} \mathbf{i}_1(v) \\ \mathbf{i}_2(v) \\ \mathbf{i}_3(v) \end{bmatrix}. \quad (3.4.6)$$

Вектор $\mathbf{x}(u, v_{\min})$ выражает положение точки образующей относительно точки на направляющей кривой в подвижном касательном базисе, которое сохраняется для произвольного параметра v . Переходя из местной системы координат в глобальную систему координат уже при текущем параметре v в соответствии с (1.2.5), получим радиус-вектор точки на поверхности

$$\mathbf{r}(u, v) = \mathbf{g}(v) + \mathbf{A}^T(v) \cdot \mathbf{x}(u, v_{\min}).$$

Таким образом, радиус-вектор кинематической поверхности заметания описывается формулой

$$\begin{aligned} \mathbf{r}(u, v) &= \mathbf{g}(v) + \mathbf{M}(v) \cdot (\mathbf{c}(u) - \mathbf{g}(v_{\min}) - \mathbf{h}), \\ t_{\min} \leq u \leq t_{\max}, \quad v_{\min} \leq v \leq v_{\max}, \end{aligned} \quad (3.4.7)$$

где $\mathbf{M}(v)$ — матричная функция, дающая матрицу поворота текущего подвижного касательного базиса относительно его начального положения. Эта матрица вычисляется по формуле

$$\mathbf{M}(v) = \mathbf{A}^T(v) \cdot \mathbf{A}(v_{\min}) = \begin{bmatrix} \mathbf{i}_1(v) \\ \mathbf{i}_2(v) \\ \mathbf{i}_3(v) \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{i}_1(v_{\min}) \\ \mathbf{i}_2(v_{\min}) \\ \mathbf{i}_3(v_{\min}) \end{bmatrix}. \quad (3.4.8)$$

В начале направляющей кривой матрица $\mathbf{M}(v_{\min})$ равна единичной матрице, так как $\mathbf{A}^T = \mathbf{A}^{-1}$.

Для вычисления производных радиус-вектора кинематической поверхности заметания нужно вычислять производные матрицы $\mathbf{M}(v)$, что в свою очередь сводится к вычислению производных ортов $\mathbf{i}_1(v)$, $\mathbf{i}_2(v)$, $\mathbf{i}_3(v)$. Рассмотрим вычисление производных векторов единичной длины на примере вычисления производных орта \mathbf{i}_1 . Этот орт получен нормированием вектора $\mathbf{g}(v)$. Вычислим первую, вторую и третью производные этого вектора — \mathbf{g}' , \mathbf{g}'' и \mathbf{g}''' . По этим производным сам орт \mathbf{i}_1 и его первая и вторая производные (\mathbf{i}_1' и \mathbf{i}_1'') определятся по формулам:

$$\begin{aligned} \mathbf{i}_1 &= \frac{\mathbf{g}'}{|\mathbf{g}'|}, \\ \frac{d\mathbf{i}_1}{dv} &= \frac{\mathbf{g}''}{|\mathbf{g}'|} - \left(\mathbf{i}_1 \cdot \frac{\mathbf{g}''}{|\mathbf{g}'|} \right) \mathbf{i}_1, \\ \frac{d^2\mathbf{i}_1}{dv^2} &= \frac{\mathbf{g}'''}{|\mathbf{g}'|} - 2 \left(\mathbf{i}_1 \cdot \frac{\mathbf{g}''}{|\mathbf{g}'|} \right) \frac{d\mathbf{i}_1}{dv} - \left(\mathbf{i}_1 \cdot \frac{\mathbf{g}'''}{|\mathbf{g}'|} + \frac{d\mathbf{i}_1}{dv} \cdot \frac{\mathbf{g}''}{|\mathbf{g}'|} \right) \mathbf{i}_1, \end{aligned} \quad (3.4.9)$$

где вторая и третья формулы получены путем дифференцирования первой с учетом того, что длина первой производной направляющей кривой $|\mathbf{g}'| \equiv \left| \frac{d\mathbf{g}}{dv} \right| = \sqrt{\mathbf{g}' \cdot \mathbf{g}'}$. Аналогично получим производные остальных ортов. Орт \mathbf{i}_2 получен

нормированием вектора $\mathbf{d}_2 = \mathbf{d} - (\mathbf{i}_1 \cdot \mathbf{d})\mathbf{i}_1$, а орт \mathbf{i}_3 получен нормированием вектора $\mathbf{d}_3 = \mathbf{g}' \times \mathbf{d}_2$. Вычислив первые, вторые и третьи производные \mathbf{d}_2 и \mathbf{d}_3 по (3.4.9), получим производные соответствующих ортов. Далее вычислим производные матричных функций $\mathbf{A}(v)$ и $\mathbf{M}(v)$, что позволит вычислить производные радиус-вектора поверхности (3.4.7).

Замкнутость кинематической поверхности по параметру u совпадает с замкнутостью образующей кривой $\mathbf{c}(v)$, а замкнутость поверхности по параметру

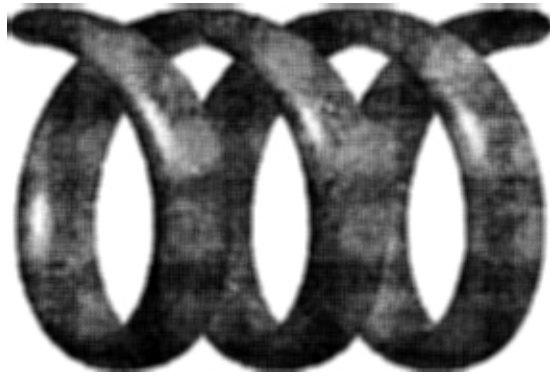


Рис. 3.4.5. Кинематическая поверхность

v совпадает с замкнутостью направляющей кривой $\mathbf{g}(v)$. На рис. 3.4.5 показана кинематическая поверхность, полученная движением эллипса вдоль спиральной линии.

При неудачном сочетании параметров кривых кинематическая поверхность может пересекать сама себя. Мы не будем рассматривать самопересекающиеся поверхности.

Из (3.4.9) следует, что для вычисления частных производных второго порядка радиус-вектора поверхности требуются производные радиус-вектора кривой третьего порядка. Для получения всей геометрической информации о кривой достаточно знать векторную функцию ее радиус-вектора и ее первую и вторую производные, но для построения некоторых поверхностей требуется третья производная кривой. Кинематическая поверхность является одной из них.

3.5. Линейчатые поверхности

Линейчатая поверхность представляет собой геометрическое место отрезков прямых, соединяющих соответствующие точки заданных двух линий. Ее можно получить путем движения прямой линии по двум направляющим кривым линиям, при этом каждой точке одной кривой должна соответствовать вполне определенная точка другой кривой. Пусть заданы две направляющие кривые $\mathbf{a}(t)$ и $\mathbf{d}(w)$, где параметры кривых t и w изменяются в пределах $t_{\min} \leq t \leq t_{\max}$ и $w_{\min} \leq w \leq w_{\max}$. Радиус-вектор линейчатой поверхности определяется формулой

$$\begin{aligned} \mathbf{r}(u, v) &= (1 - v)\mathbf{a}(t) + v\mathbf{d}(w), \\ t &= t_{\min}(1 - u) + t_{\max}u, \quad w = w_{\min}(1 - u) + w_{\max}u, \\ 0 &\leq u \leq 1, \quad 0 \leq v \leq 1. \end{aligned} \quad (3.5.1)$$

Для направляющих кривых неявно производится репараметризация — приведение области изменения параметров к отрезку от 0 до 1. Можно было бы репараметризовать только одну кривую — привести область изменения одной из них к другой. Если параметризация кривых $\mathbf{a}(u)$ и $\mathbf{d}(u)$ совпадает, то радиус-вектор

линейчатой поверхности может быть описан функцией

$$\mathbf{r}(u, v) = (1 - v)\mathbf{a}(u) + v\mathbf{d}(u), \quad u_{\min} \leq u \leq u_{\max}, \quad 0 \leq v \leq 1.$$

На рис. 3.5.1 приведен пример линейчатой поверхности.

Если обе направляющие замкнуты, то линейчатая поверхность замкнута по параметру u . По другому параметру линейчатая поверхность всегда не замкнута.

Если направляющими линиями являются отрезки прямых линий, то поверхность является линейчатой по обоим параметрам и для ее построения

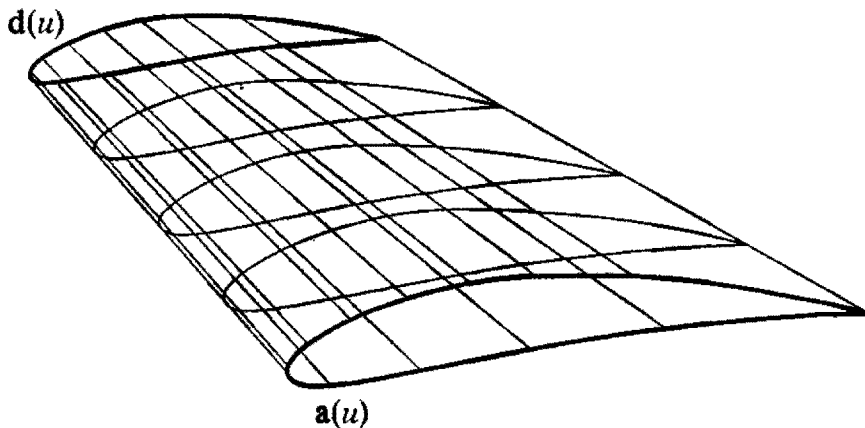


Рис. 3.5.1. Линейчатая поверхность

достаточно знать радиус-векторы концевых точек отрезков. Пусть один направляющий отрезок в форме (2.2.3) проведен из точки \mathbf{p}_1 в точку \mathbf{p}_2 , а другой направляющий отрезок проведен из точки \mathbf{p}_3 в точку \mathbf{p}_4 . Тогда линейчатая по двум параметрам поверхность определится векторной функцией

$$\begin{aligned} \mathbf{r}(u, v) &= (1 - v)(\mathbf{p}_1(1 - u) + \mathbf{p}_2u) + v(\mathbf{p}_3(1 - u) + \mathbf{p}_4u) = \\ &= (1 - u)(1 - v)\mathbf{p}_1 + u(1 - v)\mathbf{p}_2 + (1 - u)v\mathbf{p}_3 + uv\mathbf{p}_4, \quad (3.5.2) \\ &0 \leq u \leq 1, \quad 0 \leq v \leq 1. \end{aligned}$$

Такая поверхность называется *билинейной*. Она всегда не замкнута. Если все четыре точки лежат в одной плоскости, то поверхность представляет собой часть плоскости. В некоторых случаях удобно описывать плоскость по четырем точкам, лежащим в углах квадрата, в виде зависимости (3.5.2). Описанная таким образом плоскость отличается от (3.2.1) тем, что ее параметрические размеры не зависят от размеров самой плоскости. При выполнении пересечений и других операций над поверхностями желательно, чтобы у поверхностей области определения параметров не зависели от размеров поверхностей. Если точки \mathbf{p}_1 и \mathbf{p}_3 или \mathbf{p}_2 и \mathbf{p}_4 совпадают, то мы получим треугольную поверхность, которая всегда плоская.

Частным случаем линейчатой поверхности является секториальная поверхность. Она получается из линейчатой поверхности (3.5.1), если одну из кривых, например $\mathbf{d}(u)$, заменить точкой \mathbf{p} . Радиус-вектор секториальной поверхности описывается формулой

$$\mathbf{r}(u, v) = (1 - v)\mathbf{a}(u) + v\mathbf{p}, \quad t_{\min} \leq u \leq t_{\max}, \quad 0 \leq v \leq 1. \quad (3.5.3)$$

Если направляющая $\mathbf{a}(t)$ замкнута, то секториальная поверхность также замкнута по параметру u . По другому параметру секториальная поверхность всегда не замкнута.

Треугольная и секториальная поверхности имеют точку, в которой одна из частных производных ее радиус-вектора равна нулю. Аналогичные точки имеют сферическая и коническая поверхности. Эти точки не являются особыми, так как другая частная производная радиус-вектора поверхности не равна нулю. Если рассматривать кривые на поверхностях, то в таких точках находятся отрезки двумерных координатных линий.

С помощью линейчатой поверхности можно моделировать детали с уклоном и фаски деталей. Направляющими кривыми поверхности-фаски служат кривые на поверхностях, отстоящие от ребра на заданном расстоянии.

3.6. Поверхности Кунса

Возьмем векторную функцию линейчатой поверхности (3.5.1), прибавим к ней и вычтем из нее векторную функцию билинейной поверхности (3.5.2), где \mathbf{p}_1 и \mathbf{p}_2 — концевые точки направляющей кривой $\mathbf{a}(t)$, а \mathbf{p}_3 и \mathbf{p}_4 — концевые точки направляющей кривой $\mathbf{d}(w)$. В результате этих действий векторная функция не изменится, но будет иметь вид

$$\begin{aligned} \mathbf{r}(u, v) = & (1 - v)\mathbf{a}(u) + v\mathbf{d}(u) + (1 - u)(\mathbf{p}_1(1 - v) + \mathbf{p}_3v) + \\ & + u(\mathbf{p}_2(1 - v) + \mathbf{p}_4v) - (1 - u)(1 - v)\mathbf{p}_1 - u(1 - v)\mathbf{p}_2 - \\ & - (1 - u)v\mathbf{p}_3 - uv\mathbf{p}_4 = (1 - v)\mathbf{a}(u) + v\mathbf{d}(u) + (1 - u)\mathbf{b}(v) + u\mathbf{c}(v) - \\ & - (1 - u)(1 - v)\mathbf{p}_1 - u(1 - v)\mathbf{p}_2 - (1 - u)v\mathbf{p}_3 - uv\mathbf{p}_4, \quad (3.6.1) \\ & 0 \leq u \leq 1, \quad 0 \leq v \leq 1. \end{aligned}$$

В обозначениях функций $\mathbf{a}(u)$ и $\mathbf{d}(u)$ мы использовали их реальный аргумент u , а также ввели обозначения для отрезков прямых $\mathbf{b}(v) = \mathbf{p}_1(1 - v) + \mathbf{p}_3v$, $\mathbf{c}(v) = \mathbf{p}_2(1 - v) + \mathbf{p}_4v$, соединяющих концы кривых $\mathbf{a}(u)$ и $\mathbf{d}(u)$. Если допустить, что в качестве $\mathbf{b}(v)$ и $\mathbf{c}(v)$ могут использоваться произвольные кривые,

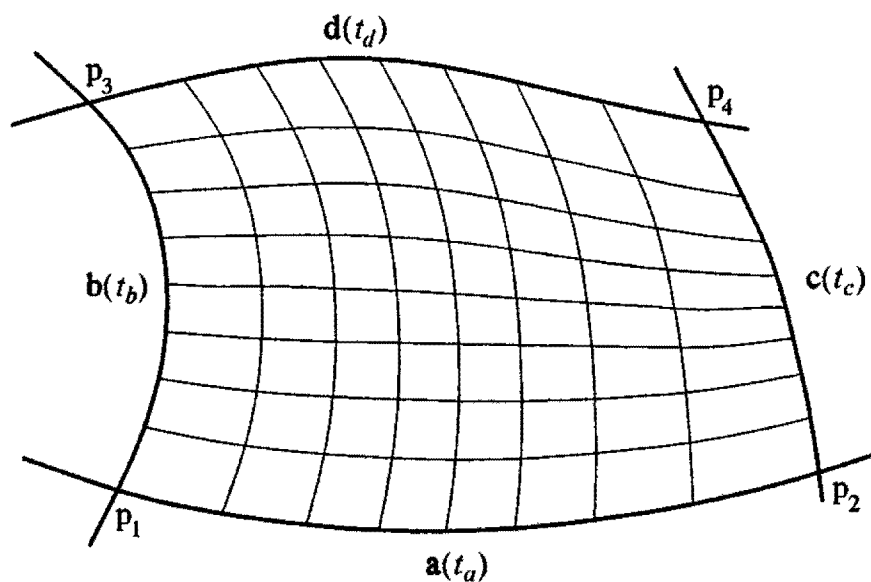


Рис. 3.6.1. Поверхность Кунса

начинающиеся и оканчивающиеся в тех же точках, то получим вариант поверхности Кунса. Поверхность Кунса строится по четырем равноправным кривым.

Рассмотрим четыре кривые $\mathbf{a}(t_a)$, $\mathbf{d}(t_d)$, $\mathbf{b}(t_b)$, $\mathbf{c}(t_c)$, попарно пересекающиеся в точках \mathbf{p}_1 , \mathbf{p}_3 , \mathbf{p}_2 , \mathbf{p}_4 , как показано на рис. 3.6.1.

Построим поверхность внутри «четыреугольника», образованного кривыми. Пусть точкам \mathbf{p}_1 и \mathbf{p}_2 на кривой $\mathbf{a}(t_a)$ соответствуют параметры $t_{a \min}$ и $t_{a \max}$, точкам \mathbf{p}_1 и \mathbf{p}_3 на кривой $\mathbf{b}(t_b)$ соответствуют параметры $t_{b \min}$ и $t_{b \max}$, точкам \mathbf{p}_3 и \mathbf{p}_4 на кривой $\mathbf{d}(t_d)$ соответствуют параметры $t_{d \min}$ и $t_{d \max}$, точкам \mathbf{p}_2 и \mathbf{p}_4 на кривой $\mathbf{c}(t_c)$ соответствуют параметры $t_{c \min}$ и $t_{c \max}$. Поверхность, построенную по данным четырем кривым, можно описать радиус-вектором

$$\begin{aligned} \mathbf{r}(u, v) &= (1 - v)(\mathbf{a}(t_a) - (1 - u)\mathbf{p}_1) + \\ &+ (1 - u)(\mathbf{b}(t_b) - v\mathbf{p}_3) + v(\mathbf{d}(t_d) - u\mathbf{p}_4) + u(\mathbf{c}(t_c) - (1 - v)\mathbf{p}_2) = \\ &= (1 - v)\mathbf{a}(t_a) + (1 - u)\mathbf{b}(t_b) + v\mathbf{d}(t_d) + u\mathbf{c}(t_c) - (1 - u)(1 - v)\mathbf{p}_1 - \\ &\quad - u(1 - v)\mathbf{p}_2 - (1 - u)v\mathbf{p}_3 - uv\mathbf{p}_4, \quad (3.6.2) \end{aligned}$$

$$\begin{aligned} t_a &= t_{a \min}(1 - u) + t_{a \max}u, & t_b &= t_{b \min}(1 - v) + t_{b \max}v, \\ t_d &= t_{d \min}(1 - u) + t_{d \max}u, & t_c &= t_{c \min}(1 - v) + t_{c \max}v, \\ &0 \leq u \leq 1, & &0 \leq v \leq 1. \end{aligned}$$

Легко проверить, что вектор $\mathbf{r}(u, 0)$ описывает участок кривой $\mathbf{a}(t_a)$, вектор $\mathbf{r}(u, 1)$ описывает участок кривой $\mathbf{d}(t_d)$, вектор $\mathbf{r}(0, v)$ описывает участок кривой $\mathbf{b}(t_b)$, вектор $\mathbf{r}(1, v)$ описывает участок кривой $\mathbf{c}(t_c)$. Совсем не обязательно, чтобы $t_{k \min}$ было меньше $t_{k \max}$, $k = a, b, c, d$.

Для того чтобы формула (3.6.2) легче воспринималась, переобозначим кривые и точки, входящие в структуру данных описываемой ею поверхности: кривую $\mathbf{a}(t_a)$ обозначим через $\mathbf{r}(u, 0)$, кривую $\mathbf{d}(t_d)$ обозначим через $\mathbf{r}(u, 1)$, кривую $\mathbf{b}(t_b)$ обозначим через $\mathbf{r}(0, v)$, кривую $\mathbf{c}(t_c)$ обозначим через $\mathbf{r}(1, v)$, точку \mathbf{p}_1 обозначим через $\mathbf{r}(0, 0)$, точку \mathbf{p}_2 обозначим через $\mathbf{r}(1, 0)$, точку \mathbf{p}_3 обозначим через $\mathbf{r}(0, 1)$, точку \mathbf{p}_4 обозначим через $\mathbf{r}(1, 1)$, и будем считать, что для кривых выполнен переход к параметрам u и v в соответствии с (3.6.2). Для коэффициентов, с которыми граничные кривые входят в (3.6.2), введем обозначение

$$\begin{aligned} \alpha_0(u) &= 1 - u, & \alpha_1(u) &= u, \\ \alpha_0(v) &= 1 - v, & \alpha_1(v) &= v. \end{aligned} \quad (3.6.3)$$

В результате векторная функция (3.6.1) примет вид

$$\begin{aligned} \mathbf{r}(u, v) &= \alpha_0(v)\mathbf{r}(u, 0) + \alpha_0(u)\mathbf{r}(0, v) + \\ &+ \alpha_1(v)\mathbf{r}(u, 1) + \alpha_1(u)\mathbf{r}(1, v) - \alpha_0(u)\alpha_0(v)\mathbf{r}(0, 0) - \\ &- \alpha_1(u)\alpha_0(v)\mathbf{r}(1, 0) - \alpha_0(u)\alpha_1(v)\mathbf{r}(0, 1) - \alpha_1(u)\alpha_1(v)\mathbf{r}(1, 1) = \\ &= [\alpha_0(u) \quad \alpha_1(u)] \cdot \begin{bmatrix} \mathbf{r}(0, v) \\ \mathbf{r}(1, v) \end{bmatrix} + [\alpha_0(v) \quad \alpha_1(v)] \cdot \begin{bmatrix} \mathbf{r}(u, 0) \\ \mathbf{r}(u, 1) \end{bmatrix} - \\ &\quad - [\alpha_0(u) \quad \alpha_1(u)] \cdot \begin{bmatrix} \mathbf{r}(0, 0) & \mathbf{r}(0, 1) \\ \mathbf{r}(1, 0) & \mathbf{r}(1, 1) \end{bmatrix} \cdot \begin{bmatrix} \alpha_0(v) \\ \alpha_1(v) \end{bmatrix} = \\ &= \sum_{i=0}^1 \alpha_i(u)\mathbf{r}(i, v) + \sum_{j=0}^1 \alpha_j(v)\mathbf{r}(u, j) + \sum_{i=0}^1 \sum_{j=0}^1 \alpha_i(u)\alpha_j(v)\mathbf{r}(i, j), \quad (3.6.4) \end{aligned}$$

$$0 \leq u \leq 1, \quad 0 \leq v \leq 1.$$

Поверхность (3.6.4) построена на четырех граничных кривых, коэффициенты при которых являются линейными функциями параметров u и v . Такое линейное объединение четырех кривых называется *линейной поверхностью Кунса*. Функции $\alpha_i(u)$ и $\alpha_j(v)$ называются *функциями смещения*. Если граничные линии являются отрезками прямых линий, то формула (3.6.4) даст билинейную поверхность (3.5.2).

Подобной поверхностью можно описать произвольную часть плоскости, границу которой можно представить в виде четырех попарно пересекающихся линий на ней. Например, часть плоскости, лежащей внутри эллипса, можно описать поверхностью

$$\begin{aligned} \mathbf{r}(u, v) = & \mathbf{p} + \\ & + a \left(\cos u + \cos v - 1 + \frac{2u(1 - \cos v - \sin v)}{\pi} + \frac{2v(1 - \cos u - \sin u)}{\pi} \right) \mathbf{i}_1 + \\ & + b \left(\sin u - \sin v - \frac{2u(1 - \cos v - \sin v)}{\pi} + \frac{2v(1 - \cos u - \sin u)}{\pi} \right) \mathbf{i}_2, \quad (3.6.5) \\ & 0 \leq u \leq \frac{\pi}{2}, \quad 0 \leq v \leq \frac{\pi}{2}, \end{aligned}$$

где \mathbf{p} — радиус-вектор центра, a и b — полуоси эллипса, $\mathbf{i}_1, \mathbf{i}_2$ — орты, лежащие в плоскости эллипса (рис. 3.6.2).

Формула (3.6.5) получена как частный случай формулы (3.6.4) путем подстановки в нее конкретных линий и точек. Поверхность (3.6.5) построена по четырем дугам эллипса. Данная поверхность отличается от своего аналога в

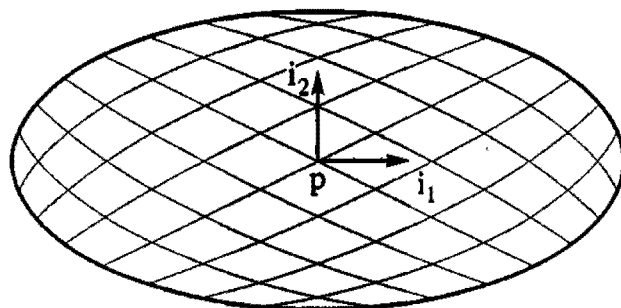


Рис. 3.6.2. Частный случай поверхности Кунса

виде плоскости с областью изменения параметров, ограниченной эллипсом, тем, что имеет прямоугольную область изменения параметров. Следует заметить, что поверхность (3.6.5) имеет особенности в угловых точках: производная по первому параметру коллинеарна производной по второму параметру. Дуги эллипса в (3.6.5) представлены в явном виде. Для удобства верхние границы области изменения параметров заменены с 1 на $\pi/2$. Поверхность (3.6.5) имеет четыре особые точки, в которых частные производные радиус-вектора по параметрам u и v коллинеарны. Это точки стыковки дуг эллипса.

Из поверхностей (3.6.4) можно сконструировать составную поверхность, стыкуя их по граничным кривым (делая граничные кривые общими). Если поверхность (3.6.4) стыкуется по своим краям с другой такой же поверхностью, то на границе производная в трансверсальном направлении будет претерпевать излом. Для того чтобы на граничных кривых производные радиус-вектора поверхности в трансверсальном к граничным кривым направлении являлись заданными

функциями, добавим их к ее описанию. То есть добавим к описанию поверхности ее производные вдоль граничных кривых $\mathbf{r}_u(i, v)$, $\mathbf{r}_v(u, j)$ и производные в углах поверхности $\mathbf{r}_u(i, j)$, $\mathbf{r}_v(i, j)$, $\mathbf{r}_{uv}(i, j)$, с коэффициентами $\beta_i(u)$, $\beta_j(v)$, где $i = 0, 1$, $j = 0, 1$. Линейные функции смещения (3.6.3) заменим кубическими функциями (2.5.4). Построенная таким образом поверхность описывается векторной функцией

$$\begin{aligned} \mathbf{r}(u, v) = & [\alpha_0(u) \quad \alpha_1(u) \quad \beta_0(u) \quad \beta_1(u)] \cdot \begin{bmatrix} \mathbf{r}(0, v) \\ \mathbf{r}(1, v) \\ \mathbf{r}_u(0, v) \\ \mathbf{r}_u(1, v) \end{bmatrix} + \\ & + [\alpha_0(v) \quad \alpha_1(v) \quad \beta_0(v) \quad \beta_1(v)] \cdot \begin{bmatrix} \mathbf{r}(u, 0) \\ \mathbf{r}(u, 1) \\ \mathbf{r}_v(u, 0) \\ \mathbf{r}_v(u, 1) \end{bmatrix} - [\alpha_0(u) \quad \alpha_1(u) \quad \beta_0(u) \quad \beta_1(u)] \times \\ & \times \begin{bmatrix} \mathbf{r}(0, 0) & \mathbf{r}(0, 1) & \mathbf{r}_v(0, 0) & \mathbf{r}_v(0, 1) \\ \mathbf{r}(1, 0) & \mathbf{r}(1, 1) & \mathbf{r}_v(1, 0) & \mathbf{r}_v(1, 1) \\ \mathbf{r}_u(0, 0) & \mathbf{r}_u(0, 1) & \mathbf{r}_{uv}(0, 0) & \mathbf{r}_{uv}(0, 1) \\ \mathbf{r}_u(1, 0) & \mathbf{r}_u(1, 1) & \mathbf{r}_{uv}(1, 0) & \mathbf{r}_{uv}(1, 1) \end{bmatrix} \cdot \begin{bmatrix} \alpha_0(v) \\ \alpha_1(v) \\ \beta_0(v) \\ \beta_1(v) \end{bmatrix}, \quad (3.6.6) \end{aligned}$$

$$0 \leq u \leq 1, \quad 0 \leq v \leq 1,$$

где кубические функции смещения определяются формулами

$$\begin{aligned} \alpha_0(w) &= 1 - 3w^2 + 2w^3, & \alpha_1(w) &= 3w^2 - 2w^3, \\ \beta_0(w) &= w - 2w^2 + w^3, & \beta_1(w) &= -w^2 + w^3. \end{aligned} \quad (3.6.7)$$

Поверхность (3.6.6) называется *кубической поверхностью Кунса*. Формула (3.6.6) описывает поверхность Кунса, построенную на тех же четырех кривых, показанных на рис. 3.6.1, что и поверхность (3.6.4). Но в отличие от последней она имеет заданные производные в трансверсальном направлении вдоль граничных кривых.

Форрест предложил обобщить поверхности (3.6.4) и (3.6.6) и использовать вместо линейных или кубических функций смещения некоторые *обобщенные функции смещения* $\alpha_{k,i}(u)$. Индекс k указывает на то, что обобщенная функция смещения умножается на векторную функцию, представляющую собой производную k -го порядка от граничной кривой. Производной нулевого порядка функции будем называть саму функцию. Для граничных кривых $k = 0$, для касательных векторов $k = 1$ или, другими словами, $\alpha_{0,i}(u) = \alpha_i(u)$, $\alpha_{1,i}(u) = \beta_i(u)$, $i = 0, 1$. Обобщенные функции смещения должны удовлетворять равенствам

$$\alpha_{k,i}^{(n)}(j) \equiv \left. \frac{d^n \alpha_{k,i}}{du} \right|_{u=j} = \delta_i^j \delta_k^n = \begin{cases} 1, & \text{если } i = j \text{ и } k = n, \\ 0, & \text{если } i \neq j \text{ или } k \neq n, \end{cases} \quad (3.6.8)$$

которые являются обобщением условий, которым удовлетворяют линейные и кубические функции смещения. Радиус-вектор обобщенной поверхности Кунса определяется формулой

$$\begin{aligned} \mathbf{r}(u, v) = & \sum_{i=0}^1 \sum_{k=0}^n \alpha_{k,i}(u) \mathbf{r}_u^{(k)}(i, v) + \sum_{j=0}^1 \sum_{l=0}^m \alpha_{l,j}(v) \mathbf{r}_v^{(l)}(u, j) - \\ & - \sum_{i=0}^1 \sum_{j=0}^1 \sum_{k=0}^n \sum_{l=0}^m \alpha_{k,i}(u) \alpha_{l,j}(v) \mathbf{r}_{uv}^{(k,l)}(i, j), \quad (3.6.9) \\ & 0 \leq u \leq 1, \quad 0 \leq v \leq 1, \end{aligned}$$

где используются заданные значения производных на краях поверхности

$$\begin{aligned} \mathbf{r}_u^{(k)}(i, v) &= \left. \frac{\partial^k \mathbf{r}(u, v)}{\partial u^k} \right|_{u=i}, \quad i = 0, 1, \\ \mathbf{r}_v^{(l)}(u, j) &= \left. \frac{\partial^l \mathbf{r}(u, v)}{\partial v^l} \right|_{v=j}, \quad j = 0, 1, \end{aligned}$$

и заданные значения производных в угловых точках параметрической области поверхности

$$\mathbf{r}_{uv}^{(k,l)}(i, j) = \left. \frac{\partial^{k+l} \mathbf{r}(u, v)}{\partial u^k \partial v^l} \right|_{\substack{u=i \\ v=j}}, \quad i, j = 0, 1.$$

В качестве обобщенных функций смещения могут быть использованы полиномы. Обобщенные поверхности Кунса имеют заданные производные k -го порядка радиус-вектора по параметру u вдоль граничных кривых $u = \text{const}$ и заданные производные l -го порядка радиус-вектора по параметру v вдоль граничных кривых $v = \text{const}$.

3.7. Сплайновые поверхности

Корпуса судов, фюзеляжи и крылья самолетов часто проектируются с помощью поперечных или продольных сечений, проходящих в заданных местах. Эти сечения соединяются плавными кривыми, в результате чего получается единая трехмерная форма. Это выполнялось в натуральную величину, для чего использовались плазы. По традиции процесс построения поверхности по ее сечениям называется плазированием, а поверхности, построенные таким способом, называются *сплайновыми* поверхностями. Рассмотрим некоторые сплайновые поверхности.

Поверхность Эрмита. Пусть имеется семейство кривых $\mathbf{r}_j(t_j)$, $j = 0, 1, 2, \dots, n$, показанных на рис. 3.7.1. Кривые семейства не должны пересекать друг друга. На этом семействе можно построить гладкую поверхность, имеющую непрерывные производные по всем направлениям. Приведем все кривые к одной параметрической длине, так чтобы область изменения параметра была

$u_{\min} \leq u \leq u_{\max}$. Кривые переобозначим в соответствии с их местом на поверхности: $\mathbf{r}(u, j) \equiv \mathbf{r}_j(t_j)$, $j = 0, 1, 2, \dots, n$. Пусть параметр поверхности v между кривыми $\mathbf{r}(u, i)$ и $\mathbf{r}(u, k)$ изменяется от i до k .

Радиус-вектор сглаживающей поверхности, построенной по семейству сечений, являющейся аналогом составного сплайна Эрмита (2.4.3), описывается формулой

$$\begin{aligned} \mathbf{r}(u, v) &= [\alpha_0(w) \quad \alpha_1(w) \quad \beta_0(w) \quad \beta_1(w)] \cdot \begin{bmatrix} \mathbf{r}(u, j) \\ \mathbf{r}(u, j+1) \\ \mathbf{r}_v(u, j) \\ \mathbf{r}_v(u, j+1) \end{bmatrix} = \\ &= \mathbf{r}(u, j)(1 - 3w^2 + 2w^3) + \mathbf{r}(u, j+1)(3w^2 - 2w^3) + \\ &\quad + \mathbf{r}_v(u, j)(w - 2w^2 + w^3) + \mathbf{r}_v(u, j+1)(-w^2 + w^3), \quad (3.7.1) \\ &w = w(v) = v - j, \quad u_{\min} \leq u \leq u_{\max}, \quad 0 \leq v \leq n, \end{aligned}$$

где j равно целой части параметра v и номеру кривой, $n + 1$ — число кривых. Данная поверхность является кубической по v направлению. Параметр w является местным параметром данной части поверхности в v направлении, он

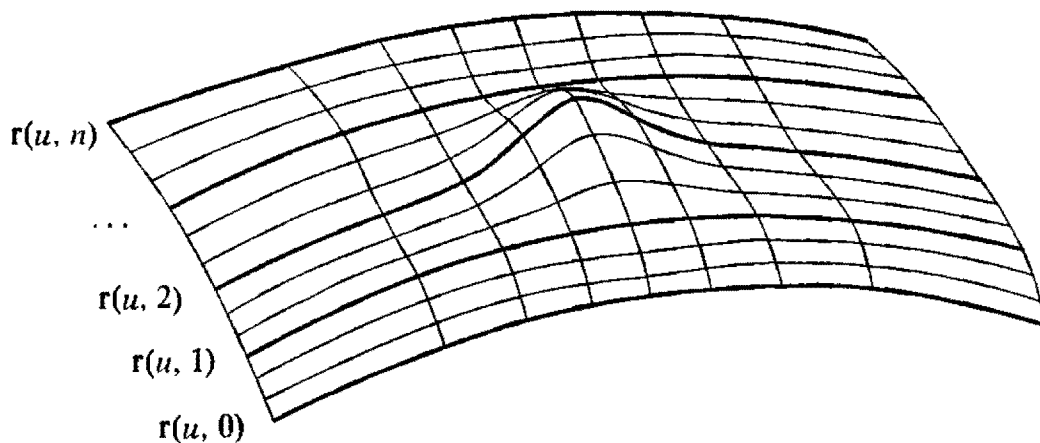


Рис. 3.7.1. Поверхность на семействе кривых

изменяется от 0 до 1. Производные на линиях сетки $\mathbf{r}_v(u, j)$ могут быть вычислены по соседним линиям, аналогично тому, как вычисляются производные в характеристических точках составного сплайна Эрмита, например,

$$\mathbf{r}_v(u, j) = \frac{\mathbf{r}(u, j+1) - \mathbf{r}(u, j-1)}{2}$$

для внутренних линий и

$$\begin{aligned} \mathbf{r}_v(u, 0) &= 2\mathbf{r}(u, 1) - 2\mathbf{r}(u, 0) - \mathbf{r}_v(u, 1), \\ \mathbf{r}_v(u, n) &= 2\mathbf{r}(u, n) - 2\mathbf{r}(u, n-1) - \mathbf{r}_v(u, n-1) \end{aligned}$$

для крайних линий.

Эти зависимости обеспечивают равенство нулю третьих производных на крайних кривых поверхности в трансверсальных к краю направлениях. Если все кривые семейства являются замкнутыми, то поверхность (3.7.1) получается

замкнутой в соответствующем направлении. Поверхность может быть замкнутой и в другом параметрическом направлении. Производные $\mathbf{r}_v(u, j)$ могут

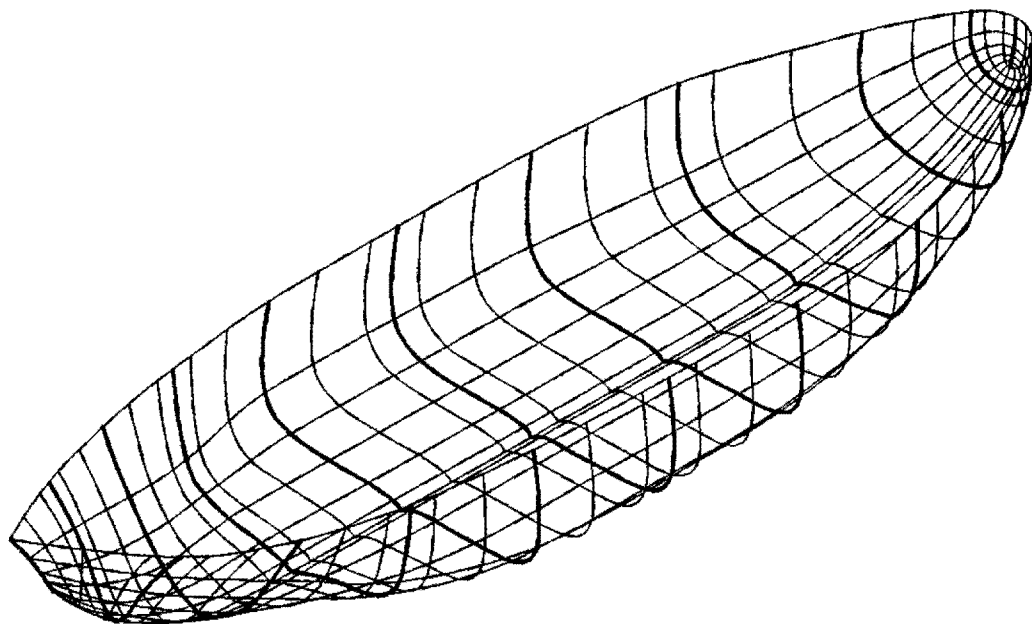


Рис. 3.7.2. Поверхность Эрмита на семействе кривых

быть вычислены и по иным формулам, приведенным для сплайна Эрмита. Поверхность Эрмита, построенная на семействе NURBS кривых, приведена на рис. 3.7.2.

Поверхность Лагранжа. По семейству кривых, показанных на рис. 3.7.1, можно построить поверхность, являющуюся аналогом кривой Лагранжа (2.4.13). Радиус-вектор такой поверхности, построенной по семейству сечений $\mathbf{r}(u, i) \equiv \mathbf{r}_i(t_i)$, $i = 0, 1, 2, \dots, n$, описывается формулой

$$\mathbf{r}(u, v) = \sum_{i=0}^n L_i(v) \mathbf{r}(u, i), \quad u_{\min} \leq u \leq u_{\max}, \quad 0 \leq v \leq n, \quad (3.7.2)$$

где $L_i(v)$ — коэффициенты Лагранжа (2.4.14) при $t_j = j$:

$$L_i(v) = \frac{\prod_{j=0, j \neq i}^n (v - j)}{\prod_{j=0, j \neq i}^n (i - j)}. \quad (3.7.3)$$

Каждая v -линия такой поверхности представляет собой кривую Лагранжа, построенную по точкам $\mathbf{r}(u, i)$, $i = 0, 1, \dots, n$, $u = \text{const}$. Существуют и другие способы построения сглаживающей поверхности по семейству кривых.

Поверхность перехода. Поверхность типа (3.7.1) может служить *поверхностью перехода* от одной поверхности к другой. Пусть, например, требуется плавно сопрячь край одной поверхности с краем другой. Обозначим сопрягаемый край первой поверхности через $\mathbf{a}(t)$, а край второй поверхности — через $\mathbf{b}(t)$, где t — некоторый общий параметр. Через $\mathbf{a}_w(t)$ обозначим частную производную первой из сопрягаемых поверхностей в трансверсальном к ее краю направлении, а через $\mathbf{b}_w(t)$ обозначим частную производную второй поверхности в трансверсальном к ее краю направлении. Под параметрически трансверсальным направлением будем иметь в виду следующее. Если первая поверхность

описывается радиус-вектором $\mathbf{p}(u, v)$, то $\mathbf{a}(t) = \mathbf{p}(u_{\max}, t)$ и $\mathbf{a}_w(t) = \mathbf{p}_v(u_{\max}, t)$ описывают край поверхности. Аналогично, если вторая поверхность описывается радиус-вектором $\mathbf{q}(u, v)$, то $\mathbf{b}(t) = \mathbf{q}(t, v_{\min})$ и $\mathbf{b}_w(t) = \mathbf{q}_u(t, v_{\min})$ описывают край этой поверхности. Поверхность

$$\begin{aligned} \mathbf{r}(t, w) = & \mathbf{a}(t)(1 - 3w^2 + 2w^3) + \mathbf{b}(t)(3w^2 - 2w^3) + \\ & + \mathbf{a}_w(t)(w - 2w^2 + w^3) + \mathbf{b}_w(t)(-w^2 + w^3), \quad (3.7.4) \\ & t_{\min} \leq t \leq t_{\max}, \quad 0 \leq w \leq 1, \end{aligned}$$

гладко сопрягает две заданные поверхности по их краям. Мы рассмотрели случай, когда сопрягаемые поверхности имеют прямоугольные области определения параметров. На двух сопрягаемых поверхностях и поверхности перехода может быть построена единая гладкая поверхность.

Поверхность Гордона. Перейдем к построению поверхностей на двух семействах кривых. Пусть имеется сетка кривых, показанных на рис. 3.7.3.

Сетка образована двумя семействами кривых: $\mathbf{a}_j(t_j)$, $j = 0, 1, 2, \dots, n$, $\mathbf{b}_i(w_i)$, $i = 0, 1, 2, \dots, m$. Каждая кривая одного семейства пересекает все кривые другого семейства и не пересекает ни одной кривой своего семейства.

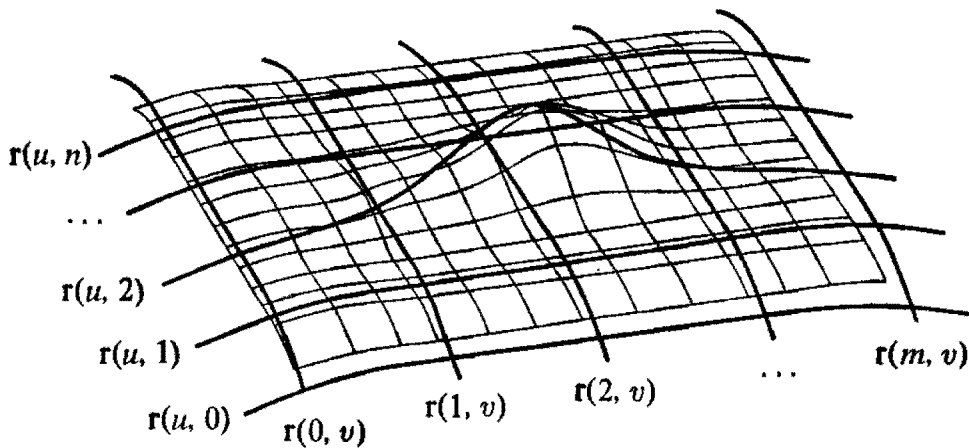


Рис. 3.7.3. Поверхность Гордона на сетке из кривых

На этой сетке можно построить составную поверхность, каждая ячейка которой представляет собой кубическую поверхность Кунса. Для этого кривые переобозначим в соответствии с их местом на поверхности: $\mathbf{r}(u, j) \equiv \mathbf{a}_j(t_j)$, $j = 0, 1, 2, \dots, n$, $\mathbf{r}(i, v) \equiv \mathbf{b}_i(w_i)$, $i = 0, 1, 2, \dots, m$. Точки пересечения кривых $\mathbf{r}(t_i, j)$ и $\mathbf{r}(i, t_j)$ обозначим через $\mathbf{r}(i, j)$, $j = 0, 1, 2, \dots, n$, $i = 0, 1, 2, \dots, m$. На каждом участке кривых введем параметризацию аналогичную (3.6.2) так, чтобы между точками $\mathbf{r}(i, j)$ и $\mathbf{r}(k, l)$ параметр u изменялся от i до k , а параметр v изменялся от j до l . Радиус-вектор поверхности, интерполирующей всю сетку и удовлетворяющей определенным условиям гладкости, описывается формулой

$$\begin{aligned} \mathbf{r}(u, v) = & \sum_{j=0}^m L_j(v) \mathbf{r}(u, j) + \sum_{i=0}^n L_i(u) \mathbf{r}(i, v) - \sum_{i=0}^n \sum_{j=0}^m L_i(u) L_j(v) \mathbf{r}(i, j), \quad (3.7.5) \\ & 0 \leq u \leq m, \quad 0 \leq v \leq n, \end{aligned}$$

где $L_j(v)$ и $L_i(u)$ — коэффициенты Лагранжа (3.7.3). Эта формула аналогична (3.7.2), только интерполяция выполняется по двум направлениям. В формулах (3.7.2) и (3.7.5) вместо коэффициентов Лагранжа могут использоваться произвольные дифференцируемые функции $a_k(w)$, удовлетворяющие равенству

$$\alpha_k(w) = \begin{cases} 1, & \text{если } w = k, \\ 0, & \text{если } w \neq k. \end{cases}$$

В качестве функций смещения $a_k(w)$ могут использоваться кубические сплайны. Поверхность (3.7.5) называется *поверхностью Гордона*. Гладкость такой поверхности зависит от гладкости функций смещения.

Для построения поверхности на сетке, образованной двумя семействами кривых, могут быть использованы поверхности Кунса (3.6.6) в виде ее порций. Так как в этом случае образуется поверхность, составленная из нескольких поверхностей, то она называется составной. Наличие в составе данных поверхности (3.6.6) производных на стыках отдельных порций позволяет получить гладкую составную поверхность. При использовании в качестве каждой порции поверхность (3.6.6) составная поверхность будет гладкой на заданной сетке кривых. Ее радиус-вектор описывается формулой

$$\begin{aligned} \mathbf{r}(u, v) = & [\alpha_0(w) \quad \alpha_1(w) \quad \beta_0(w) \quad \beta_1(w)] \cdot \begin{bmatrix} \mathbf{r}(u, j) \\ \mathbf{r}(u, j+1) \\ \mathbf{r}_v(u, j) \\ \mathbf{r}_v(u, j+1) \end{bmatrix} + \\ & + [\alpha_0(t) \quad \alpha_1(t) \quad \beta_0(t) \quad \beta_1(t)] \cdot \begin{bmatrix} \mathbf{r}(i, v) \\ \mathbf{r}(i+1, v) \\ \mathbf{r}_v(i, v) \\ \mathbf{r}_v(i+1, v) \end{bmatrix} - [\alpha_0(t) \quad \alpha_1(t) \quad \beta_0(t) \quad \beta_1(t)] \times \\ & \times \begin{bmatrix} \mathbf{r}(i, j) & \mathbf{r}(i, j+1) & \mathbf{r}_v(i, j) & \mathbf{r}_v(i, j+1) \\ \mathbf{r}(i+1, j) & \mathbf{r}(i+1, j+1) & \mathbf{r}_v(i+1, j) & \mathbf{r}_v(i+1, j+1) \\ \mathbf{r}_u(i, j) & \mathbf{r}_u(i, j+1) & \mathbf{r}_{uv}(i, j) & \mathbf{r}_{uv}(i, j+1) \\ \mathbf{r}_u(i+1, j) & \mathbf{r}_u(i+1, j+1) & \mathbf{r}_{uv}(i+1, j) & \mathbf{r}_{uv}(i+1, j+1) \end{bmatrix} \cdot \begin{bmatrix} \alpha_0(w) \\ \alpha_1(w) \\ \beta_0(w) \\ \beta_1(w) \end{bmatrix}, \\ & t = u - i, \quad w = v - j, \quad 0 \leq u \leq m, \quad 0 \leq v \leq n, \end{aligned} \quad (3.7.6)$$

где i равно целой части параметра u , j равно целой части параметра v , $m+1$ — число кривых v -направления, $n+1$ — число кривых u -направления. Кубические функции смещения определяются формулами (3.6.7). Фактически формула (3.7.6) описывает одну порцию составной поверхности, находящуюся между кривыми $\mathbf{r}(u, j)$, $\mathbf{r}(u, j+1)$, $\mathbf{r}(i, v)$, $\mathbf{r}(i+1, v)$, $i = 0, 1, \dots, m-1$, $j = 0, 1, \dots, n-1$. Параметры t и w являются местными параметрами данной порции составной поверхности, они изменяются от 0 до 1. Если $m = 1$ и $n = 1$, то формула (3.7.5) описывает одну порцию кубической поверхности Кунса (3.6.6), построенной на четырех кривых, показанных на рис. 3.6.1. Производные на линиях сетки могут быть вычислены по соседним линиям аналогично

тому, как по соседним точкам вычисляются производные составного сплайна Эрмита (2.4.3) в опорных точках

$$\begin{aligned} \mathbf{r}_u(i, v) &= \frac{\mathbf{r}(i+1, v) - \mathbf{r}(i-1, v)}{2}, \\ \mathbf{r}_v(u, j) &= \frac{\mathbf{r}(u, j+1) - \mathbf{r}(u, j-1)}{2} \end{aligned} \quad (3.7.7)$$

для внутренних линий и

$$\begin{aligned} \mathbf{r}_u(0, v) &= 2\mathbf{r}(1, v) - 2\mathbf{r}(0, v) - \mathbf{r}_u(1, v), \\ \mathbf{r}_u(m, v) &= 2\mathbf{r}(m, v) - 2\mathbf{r}(m-1, v) - \mathbf{r}_u(u, n-1), \\ \mathbf{r}_v(u, 0) &= 2\mathbf{r}(u, 1) - 2\mathbf{r}(u, 0) - \mathbf{r}_v(u, 1), \\ \mathbf{r}_v(u, n) &= 2\mathbf{r}(u, n) - 2\mathbf{r}(u, n-1) - \mathbf{r}_v(u, n-1) \end{aligned} \quad (3.7.8)$$

для крайних линий. Равенства (3.7.8) обеспечивают равенство нулю третьих производных на краях поверхности в трансверсальных к краю направлениях. Если все кривые какого-либо семейства являются замкнутыми, то поверхность (3.7.6) получается замкнутой в соответствующем направлении. Заметим, что поверхность (3.7.6) в точках пересечения линий $\mathbf{r}(i, j)$ может иметь уплощения.

Предположим, что в формуле (3.7.5) сегменты кривых сетки между точками их пересечения $\mathbf{r}(i, j)$ представляют собой кубические сплайны Эрмита (2.4.3), т. е. определяются зависимостями:

$$\begin{aligned} \mathbf{r}(i, v) &= \alpha_0(w)\mathbf{r}(i, j) + \alpha_1(w)\mathbf{r}(i, j+1) + \beta_0(w)\mathbf{r}_v(i, j) + \beta_1(w)\mathbf{r}_v(i, j+1), \\ \mathbf{r}(u, j) &= \alpha_0(t)\mathbf{r}(i, j) + \alpha_1(t)\mathbf{r}(i+1, j) + \beta_0(t)\mathbf{r}_u(i, j) + \beta_1(t)\mathbf{r}_u(i+1, j), \\ \mathbf{r}_u(i, v) &= \alpha_0(w)\mathbf{r}_u(i, j) + \alpha_1(w)\mathbf{r}_u(i, j+1) + \beta_0(w)\mathbf{r}_{vu}(i, j) + \beta_1(w)\mathbf{r}_{vu}(i, j+1), \\ \mathbf{r}_v(u, j) &= \alpha_0(t)\mathbf{r}_v(i, j) + \alpha_1(t)\mathbf{r}_v(i+1, j) + \beta_0(t)\mathbf{r}_{uv}(i, j) + \beta_1(t)\mathbf{r}_{uv}(i+1, j), \end{aligned}$$

где i равно целой части параметра u , j равно целой части параметра v , $t = u - i$, $w = v - j$. После подстановки этих зависимостей в (3.7.5) получим три одинаковых матричных члена, два из них со знаком плюс и один со знаком минус. В результате формула (3.7.5) примет вид

$$\begin{aligned} \mathbf{r}(u, v) &= [\alpha_0(t) \quad \alpha_1(t) \quad \beta_0(t) \quad \beta_1(t)] \times \\ &\times \begin{bmatrix} \mathbf{r}(i, j) & \mathbf{r}(i, j+1) & \mathbf{r}_v(i, j) & \mathbf{r}_v(i, j+1) \\ \mathbf{r}(i+1, j) & \mathbf{r}(i+1, j+1) & \mathbf{r}_v(i+1, j) & \mathbf{r}_v(i+1, j+1) \\ \mathbf{r}_u(i, j) & \mathbf{r}_u(i, j+1) & \mathbf{r}_{uv}(i, j) & \mathbf{r}_{uv}(i, j+1) \\ \mathbf{r}_u(i+1, j) & \mathbf{r}_u(i+1, j+1) & \mathbf{r}_{uv}(i+1, j) & \mathbf{r}_{uv}(i+1, j+1) \end{bmatrix} \cdot \begin{bmatrix} \alpha_0(w) \\ \alpha_1(w) \\ \beta_0(w) \\ \beta_1(w) \end{bmatrix}, \\ &t = u - i, \quad w = v - j, \quad 0 \leq u \leq m, \quad 0 \leq v \leq n, \end{aligned} \quad (3.7.9)$$

где i равно целой части параметра u , j равно целой части параметра v . Поверхность (3.7.9) полностью определяется точками $\mathbf{r}(i, j)$ и значением в них производных поверхности $\mathbf{r}_u(i, j)$, $\mathbf{r}_v(i, j)$ и $\mathbf{r}_{uv}(i, j)$. Данную поверхность можно считать составной, так как она состоит из порций кубических поверхностей

Эрмита, которые гладко стыкуются между собой, так как имеют общие производные на границах. Поверхность (3.7.9) еще называют поверхностью тензорного произведения. Фактически данная поверхность строится на сетке точек с заданными производными, а не на сетке кривых.

3.8. Поверхности Безье

Рассмотренные выше кривые Безье (2.5.3) характерны тем, что их радиус-вектор представлен в виде суммы радиус-векторов точек в пространстве, умноженных на некоторые коэффициенты, являющиеся скалярными функциями параметра кривой. Скалярными функциями являются функции Бернштейна. Аналогичным образом строятся поверхности Безье, радиус-вектор которых представляется в виде суммы радиус-векторов точек в пространстве, умноженных на коэффициенты, являющиеся функциями Бернштейна параметров поверхности. Рассмотрим поверхности Безье подробнее.

Пусть имеется совокупность точек \mathbf{p}_{ij} , образующих сетку, т. е. условно расположенных в виде $n + 1$ рядов по $m + 1$ точек в каждом ряду. Индексы точки \mathbf{p}_{ij} означают, что данная точка расположена j -й по счету в i -м ряду (первый индекс равен номеру ряда, второй — номеру точки в ряду). Нумеровать ряды будем от 0 до n , а точки в них будем нумеровать от 0 до m . Поверхность Безье, построенная на совокупности точек \mathbf{p}_{ij} , определяется радиус-вектором

$$\mathbf{r}(u, v) = \sum_{i=0}^n \sum_{j=0}^m B^m_j(u) B^n_i(v) \mathbf{p}_{ij} = \sum_{i=0}^n \sum_{j=0}^m B^{nm}_{ij}(v, u) \mathbf{p}_{ij}, \quad (3.8.1)$$

$$0 \leq u \leq 1, \quad 0 \leq v \leq 1,$$

где $B^{nm}_{ij}(v, u) = B^n_i(v) B^m_j(u)$ — функции базиса Бернштейна:

$$B^n_i(v) = \frac{n!}{(n-i)!i!} v^i (1-v)^{n-i} = C_n^i v^i (1-v)^{n-i}, \quad (3.8.2)$$

$$B^m_j(u) = \frac{m!}{(m-j)!j!} u^j (1-u)^{m-j} = C_m^j u^j (1-u)^{m-j}. \quad (3.8.3)$$

Напомним, что коэффициенты Бернштейна удовлетворяют равенству (2.5.5), в соответствии с которым функции $B^{nm}_{ij}(v, u)$ удовлетворяют равенству

$$\sum_{i=0}^n \sum_{j=0}^m B^{nm}_{ij}(v, u) = 1. \quad (3.8.4)$$

Будем говорить, что вдоль u -направления поверхность имеет степень m , а вдоль v -направления поверхность имеет степень n .

Если представить, что на каждых четырех соседних точках \mathbf{p}_{ij} , \mathbf{p}_{ij+1} , \mathbf{p}_{i+1j} , \mathbf{p}_{i+1j+1} построена билинейная поверхность, то мы получим многогранник, который будем называть *характеристическим*. Точки \mathbf{p}_{ij} будем называть *вершинами характеристического многогранника*. Поверхность Безье вместе со своим характеристическим многогранником показана на рис. 3.8.1. Так как из коэффициентов Бернштейна только нулевой и последний принимают максимально возможные значения, то поверхность Безье проходит только через

угловые вершины сетки: P_{00} , P_{0m} , P_{n0} , P_{nm} . Сечения поверхности (3.8.1) по линиям $u = \text{const}$ или $v = \text{const}$ представляют собой кривые Безье. Характеристический многогранник дает общее представление о поверхности Безье, а сама поверхность как бы сглаживает углы этого многогранника. Чем выше порядок поверхности (3.8.1), тем она более гладкая. Порядок поверхности в u -направлении жестко связан с количеством вершин в ряду, а порядок поверхности в v -направлении жестко связан с количеством рядов вершин. Область изменения параметров поверхности представляет собой квадрат со стороной, равной 1.

Степень поверхности Безье, а соответственно и ее гладкость, жестко связаны с количеством вершин, по которым она построена. Это может создать некоторые неудобства при использовании поверхностей Безье. Преодолеть эти неудобства

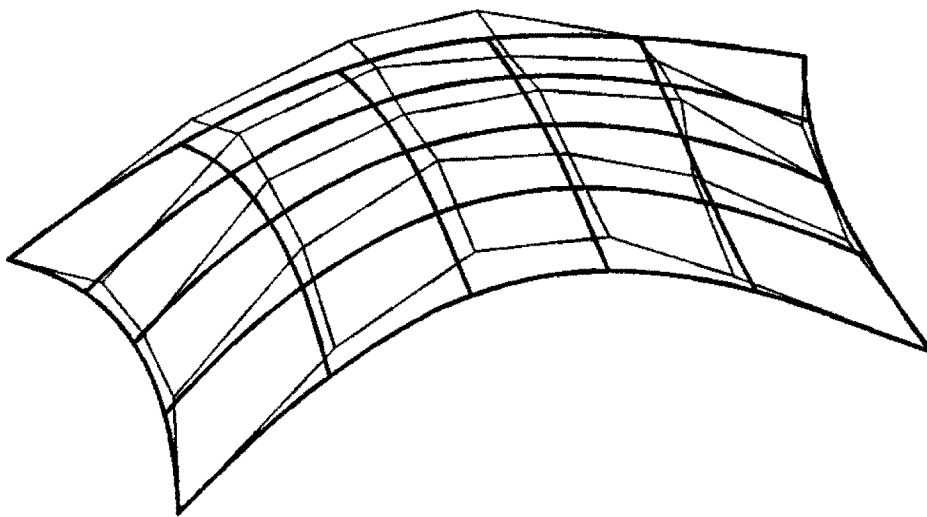


Рис. 3.8.1. Поверхность Безье и ее характеристический многогранник

могут помочь составные поверхности. Составная поверхность Безье получается в результате стыковки отдельных поверхностей Безье, имеющих вдоль стыкуемых краев одинаковые степени. В общем случае для того, чтобы порции поверхности Безье гладко стыковались друг с другом, нужно чтобы ребра, прилегающие к углам стыкуемых порций поверхности, лежали в одной плоскости.

3.9. Рациональные поверхности

В предыдущей главе было показано, что кривую Безье можно обобщить так, чтобы она могла описывать заданное коническое сечение. Рассуждая аналогичным образом, приходим к тому, что поверхность Безье можно обобщить так, чтобы она могла описать требуемую часть поверхности второго порядка — эллипсоида, гиперboloида, параболоида или их частных случаев.

Наделим каждую вершину кривой (3.8.1) P_{ij} весом w_{ij} и будем описывать поверхность в однородных координатах. Поверхности, характеристические точки которых обладают весами, называются *рациональными*. Сам радиус-вектор такой поверхности тоже приобретает вес. Вес радиус-вектора вычисляется по точно таким же формулам, что и его координаты, поэтому вес принято считать дополнительной координатой точки. Рациональные поверхности Безье можно получить подобно тому, как строились рациональные кривые Безье. Расширенный радиус-вектор (1.4.6) рациональной поверхности Безье описывается формулой, по виду совпадающей с (3.8.1), в которой обычные векторы $r(u, v)$ и P_{ij} заменены соответствующими расширенными векторами $R(u, v)$ и P_{ij} .

Обычный трехмерный радиус-вектор рациональной поверхности Безье определяется формулой

$$\mathbf{r}(u, v) = \frac{\sum_{i=0}^n \sum_{j=0}^m B^n_i(v) B^m_j(u) w_{ij} \mathbf{P}_{ij}}{\sum_{i=0}^n \sum_{j=0}^m B^n_i(v) B^m_j(u) w_{ij}}, \quad 0 \leq u \leq 1, \quad 0 \leq v \leq 1. \quad (3.9.1)$$

Как и у рациональных кривых, у рациональной поверхности играют роль не абсолютные значения весов вершин, а их относительные значения. Чем больше вес некоторой вершины по отношению к весам других вершин, тем ближе к этой вершине проходит поверхность.

С помощью поверхности (3.9.1) может быть построена часть поверхности второго порядка или часть поверхности тора так же, как с помощью рациональной кривой Безье (2.7.2) может быть построена некоторая часть конического сечения. Для этого проще всего использовать рациональную поверхность Безье второй степени, построенную по девяти вершинам:

$$\mathbf{r}(u, v) = \frac{\begin{bmatrix} (1-v)^2 & 2v(1-v) & v^2 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{P}_{00} & w_u \mathbf{P}_{01} & \mathbf{P}_{02} \\ w_v \mathbf{P}_{10} & w_v w_v \mathbf{P}_{11} & w_v \mathbf{P}_{12} \\ \mathbf{P}_{20} & w_u \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix} \cdot \begin{bmatrix} (1-u)^2 \\ 2u(1-u) \\ u^2 \end{bmatrix}}{((1-u)^2 + 2u(1-u)w_u + u^2)((1-v)^2 + 2v(1-v)w_v + v^2)}, \quad (3.9.2)$$

$$0 \leq u \leq 1, \quad 0 \leq v \leq 1.$$

В (3.9.2) веса угловых вершин положены равными единице, а для остальных вершин введены веса w_u и w_v , имеющие индексы параметрических направлений поверхности. Тип поверхности, получаемой по формуле (3.9.2), зависит от положения вершин и от весов w_u и w_v . Так, например, для того чтобы получить часть поверхности тора (3.2.18) с углами раствора α_u и α_v , нужно, чтобы вершины \mathbf{P}_{0i} , \mathbf{P}_{1i} , \mathbf{P}_{2i} и \mathbf{P}_{i0} , \mathbf{P}_{i1} , \mathbf{P}_{i2} , где $i = 0, 1, 2$, располагались друг относительно друга аналогично тому, как было показано на рис. 2.6.5, а веса вершин имели значения $w_u = \cos(\alpha_u/2)$ и $w_v = \cos(\alpha_v/2)$.

С помощью составной рациональной поверхности Безье может быть построена полностью любая поверхность второго порядка. Для этого нужно использовать несколько поверхностей (3.9.2), гладко стыкующихся между собой, как было описано выше.

Поверхность скругления. Рациональной поверхностью можно описать *поверхность скругления* ребра детали. Сечение поверхности скругления в одном из параметрических направлений должно представлять собой дугу окружности. Эту дугу мы опишем с помощью рациональной кривой Безье (2.6.16), в которой вершины \mathbf{P}_0 , \mathbf{P}_1 , \mathbf{P}_2 и угол дуги α будем считать функциями одного из параметров поверхности скругления. Радиус-вектор поверхности скругления выразится формулой

$$\mathbf{r}(u, t) = \frac{(1-t)^2 \mathbf{P}_0(u) + 2t(1-t)w(u) \mathbf{P}_1(u) + t^2 \mathbf{P}_2(u)}{(1-t)^2 + 2t(1-t)w(u) + t^2}, \quad (3.9.3)$$

$$u_{\min} \leq u \leq u_{\max}, \quad 0 \leq t \leq 1,$$

где $w(u) = \cos(\alpha(u)/2)$ — вес линии $\mathbf{p}_1(u)$, вычисленный по углу дуги в данном месте. Метод построения кривых $\mathbf{p}_0(u)$, $\mathbf{p}_1(u)$, $\mathbf{p}_2(u)$ по радиусу скругления

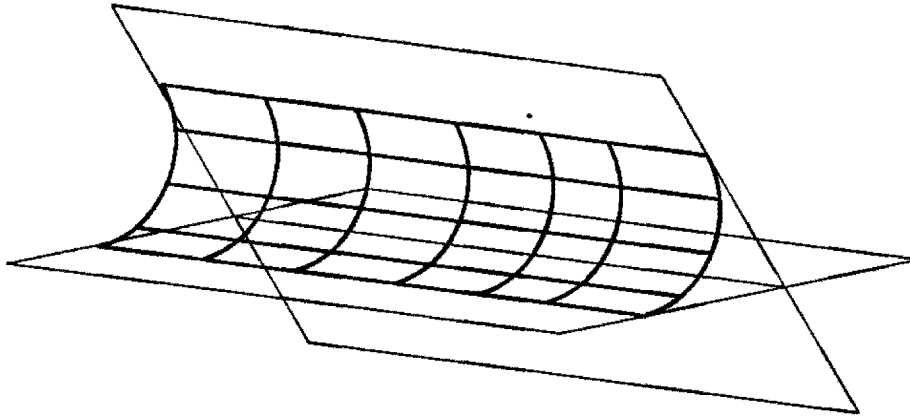


Рис. 3.9.1. Поверхность скругления — пример рациональной поверхности

(возможно переменному) будет приведен в следующей главе. Простейший вариант поверхности скругления (3.9.3), построенной между двумя плоскостями, приведен на рис. 3.9.1.

Будем рассматривать формулу (3.9.1) в более широком смысле, а именно, будем считать, что в них вместо функций B^{nm}_{ij} могут стоять некоторые произвольные базисные функции, которые обозначим через F_{ij} . В общем виде радиус-вектор рациональной параметрической поверхности, построенной на mn вершинах, определяется формулой

$$\mathbf{r}(u, v) = \frac{\sum_{i=1}^n \sum_{j=1}^m F_{ij}(v, u) w_{ij} \mathbf{p}_{ij}}{\sum_{i=1}^n \sum_{j=1}^m F_{ij}(v, u) w_{ij}}, \quad u_{\min} \leq u \leq u_{\max}, \quad v_{\min} \leq v \leq v_{\max}, \quad (3.9.4)$$

где \mathbf{p}_{ij} — радиус-векторы j -й вершины, расположенной в i -м ряду, характеристического многогранника поверхности, w_{ij} — вес соответствующей вершины, $F_{ij}(v, u)$ — функция j -й вершины в i -м ряду, n — число рядов вершин, m — число вершин в каждом ряду. В отличие от формулы (3.9.1) суммирование в формуле (3.9.4) начинается с 1, а не с 0. Это не имеет принципиального значения, просто для поверхностей Безье удобно было нумеровать, начиная с 0, а в других случаях удобнее проводить нумерацию, начиная с 1. Параметр u изменяется вдоль вершин, стоящих в некотором ряду, а параметр v изменяется вдоль рядов вершин, имеющих одинаковый номер в ряду.

В терминах однородных координат расширенный радиус-вектор (1.4.6) рациональной поверхности определяется формулой

$$\mathbf{R}(u, v) = \sum_{i=1}^n \sum_{j=1}^m F_{ij}(v, u) \mathbf{P}_{ij}, \quad u_{\min} \leq u \leq u_{\max}, \quad v_{\min} \leq v \leq v_{\max}, \quad (3.9.5)$$

где $\mathbf{P}_{ij} = [w_{ij} \mathbf{p}_{ij} \ w_{ij}]^T$ — расширенные векторы вершин поверхности.

Радиус-вектор рациональной поверхности вычисляется как частное от деления двух функций параметров u и v , поэтому при вычислении производных рациональной поверхности правую часть (3.9.5) следует рассматривать как

сложную функцию. Если условно обозначим радиус-вектор рациональной поверхности как $\mathbf{r} = \frac{w\mathbf{r}}{w}$, то производные радиус-вектора рациональной поверхности определяются формулами

$$\frac{\partial \mathbf{r}}{\partial u} = \frac{1}{w} \frac{\partial(w\mathbf{r})}{\partial u} - \frac{w\mathbf{r}}{w^2} \frac{\partial w}{\partial u}, \quad (3.9.6)$$

$$\frac{\partial \mathbf{r}}{\partial v} = \frac{1}{w} \frac{\partial(w\mathbf{r})}{\partial v} - \frac{w\mathbf{r}}{w^2} \frac{\partial w}{\partial v}, \quad (3.9.7)$$

$$\frac{\partial^2 \mathbf{r}}{\partial u^2} = \frac{1}{w} \frac{\partial^2(w\mathbf{r})}{\partial u^2} - \frac{2}{w^2} \frac{\partial(w\mathbf{r})}{\partial u} \frac{\partial w}{\partial u} - \frac{w\mathbf{r}}{w^2} \frac{\partial^2 w}{\partial u^2} + \frac{2w\mathbf{r}}{w^3} \left(\frac{\partial w}{\partial u} \right)^2, \quad (3.9.8)$$

$$\frac{\partial^2 \mathbf{r}}{\partial v^2} = \frac{1}{w} \frac{\partial^2(w\mathbf{r})}{\partial v^2} - \frac{2}{w^2} \frac{\partial(w\mathbf{r})}{\partial v} \frac{\partial w}{\partial v} - \frac{w\mathbf{r}}{w^2} \frac{\partial^2 w}{\partial v^2} + \frac{2w\mathbf{r}}{w^3} \left(\frac{\partial w}{\partial v} \right)^2, \quad (3.9.9)$$

$$\begin{aligned} \frac{\partial^2 \mathbf{r}}{\partial u \partial v} = & \frac{1}{w} \frac{\partial^2(w\mathbf{r})}{\partial u \partial v} - \frac{1}{w^2} \frac{\partial(w\mathbf{r})}{\partial u} \frac{\partial w}{\partial v} - \frac{1}{w^2} \frac{\partial(w\mathbf{r})}{\partial v} \frac{\partial w}{\partial u} - \\ & - \frac{w\mathbf{r}}{w^2} \frac{\partial^2 w}{\partial u \partial v} + \frac{2w\mathbf{r}}{w^3} \frac{\partial w}{\partial u} \frac{\partial w}{\partial v}. \end{aligned} \quad (3.9.10)$$

Производные более высокого порядка вычисляются аналогично.

3.10. NURBS поверхности

В качестве функций $F_{ij}(v, u)$ в формуле (3.9.4) будем использовать произведения B -сплайнов, один из которых зависит от параметра u , а другой зависит от параметра v . В результате получим неоднородную рациональную поверхность, определяемую B -сплайнами (2.8.21), которая приобрела название *NURBS поверхности* (*Non-Uniform Rational B-Spline surface*). NURBS поверхность, как и поверхность Безье, строится на сетке вершин — совокупности точек \mathbf{p}_{ij} , условно расположенных в виде n рядов по m точек в каждом ряду. Индексы вершины \mathbf{p}_{ij} означают то, что данная точка расположена j -й по счету в i -м ряду (первый индекс равен номеру ряда, второй — номеру точки в ряду). В данном случае нумеровать ряды удобнее от 1 до n , а точки в них — от 1 до m . Пусть вес вершины \mathbf{p}_{ij} равен w_{ij} . Пусть B -сплайны $N_{j,k}(u)$ u -направления имеют порядок k , а B -сплайны $N_{i,q}(v)$ v -направления имеют порядок q . Радиус-вектор NURBS поверхности определяется равенством

$$\mathbf{r}(u, v) = \frac{\sum_{i=1}^n \sum_{j=1}^m N_{i,q}(v) N_{j,k}(u) w_{ij} \mathbf{p}_{ij}}{\sum_{i=1}^n \sum_{j=1}^m N_{i,q}(v) N_{j,k}(u) w_{ij}} = \frac{\sum_{i=1}^n \sum_{j=1}^m N_{ij,qk}(v, u) w_{ij} \mathbf{p}_{ij}}{\sum_{i=1}^n \sum_{j=1}^m N_{ij,qk}(v, u) w_{ij}}, \quad (3.10.1)$$

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

где $N_{ij,qk}(v, u) = N_{i,q}(v) N_{j,k}(u)$ — базисные функции, n — число рядов вершин, m — число вершин в каждом ряду. Производные радиус-вектора NURBS поверхности определяются равенствами (3.9.6)–(3.9.10).

Каждый из B -сплайнов k -го порядка $N_{j,k}(u)$ построен на последовательности из $k+1$ узлов $u_j, u_{j+1}, u_{j+2}, \dots, u_{j+k}$; всего для построения совокупности m B -сплайнов k -го порядка требуется $m+k$ узлов u_j в случае незамкнутой кривой и $m+2k$ узлов в случае замкнутой кривой. Каждый из B -сплайнов q -го порядка $N_{i,q}(v)$ построен на последовательности из $q+1$ узлов $v_i, v_{i+1}, v_{i+2}, \dots, v_{i+q}$, всего для построения совокупности n B -сплайнов q -го порядка требуется $n+q$ узлов v_i в случае незамкнутой кривой и $n+2q$ узлов в случае замкнутой кривой. Как следует из (3.10.1), NURBS поверхность имеет две последовательности узлов, одну — в u -направлении, а другую — в v -направлении. Значения узлов должны образовывать неубывающие последовательности. Будем называть их u -последовательностью узлов и v -последовательностью узлов. Пронумеруем узлы каждой последовательности, начав нумерацию с 1.

Для незамкнутой NURBS поверхности используются следующие последовательности узлов. Первые k узлов u -последовательности имеют значения, равные нулю: $u_1 = u_2 = \dots = u_k = 0$; следующие $m-k$ узлов принимают целочисленные значения от 1 до $m-k$: $t_{k+j} = j$ ($j = 1, 2, \dots, m-k$); оставшиеся k узлов принимают значение $m-k+1$: $t_{m+1} = t_{m+2} = \dots = t_{m+k} = m-k+1$. Первые q узлов v -последовательности имеют значения, равные нулю: $v_1 = v_2 = \dots = v_q = 0$; следующие $n-q$ узлов принимают целочисленные значения от 1 до $n-q$: $t_{q+i} = i$ ($i = 1, 2, \dots, n-q$); оставшиеся q узлов принимают значение $n-q+1$: $t_{n+1} = t_{n+2} = \dots = t_{n+q} = n-q+1$.

Для замкнутой NURBS поверхности используются равномерные последовательности узлов с единичным шагом: $u_j = j-m$, $j = 1, 2, \dots, m+2k$, $v_i = i-n$, $i = 1, 2, \dots, n+2q$.

Приведенные примеры узловых последовательностей не являются единственно возможными. Как и для любой поверхности, внутренняя параметризация NURBS поверхности может быть произвольной.

Область изменения параметров NURBS поверхности представляет собой прямоугольник: $u_k \leq u \leq u_{m+k}$, $v_q \leq v \leq v_{n+q}$. Для указанных последовательностей узлов параметры NURBS поверхности изменяются в пределах: $0 \leq u \leq (m-k+1)$, $0 \leq v \leq (n-q+1)$ для незамкнутой поверхности и $0 \leq u \leq m$, $0 \leq v \leq n$ для замкнутой поверхности.

Для вычисления радиус-вектора NURBS поверхности используем формулы (2.9.2)–(2.9.4). По значению параметра u из условия $u_j \leq u < u_{j+1}$ определим номер j единственного отличного от нуля B -сплайна первого порядка

$$M_{j,1}(u) = \frac{1}{u_{j+1} - u_j}.$$

Далее, используя рекуррентное соотношение

$$M_{r,k}(u) = \frac{(u_{r+k} - u)M_{r+1,k-1}(u) + (u - u_r)M_{r,k-1}(u)}{u_{r+k} - u_r},$$

$$r = j - k + 1, j - k + 2, \dots, j,$$

последовательно вычислим все отличные от нуля при данном u B -сплайны до k -го порядка включительно: $M_{j-k+1,k}(u)$, $M_{j-k+2,k}(u)$, \dots , $M_{j,k}(u)$. Далее B -сплайны k -го порядка нормируем в соответствии с (2.9.4)

$$N_{r,k}(u) = (u_{r+k} - u_r)M_{r,k}(u), \quad r = j - k + 1, j - k + 2, \dots, j.$$

Аналогично по значению параметра v из условия $v_i \leq v < v_{i+1}$ определим номер i отличного от нуля B -сплайна первого порядка

$$M_{i,1}(v) = \frac{1}{v_{i+1} - v_i}.$$

Используя рекуррентное соотношение

$$M_{r,q}(v) = \frac{(v_{r+q} - v)M_{r+1,q-1}(v) + (v - v_r)M_{r,q-1}(v)}{v_{r+q} - v_r},$$

$$r = i - q + 1, i - q + 2, \dots, i,$$

последовательно вычислим все отличные от нуля при данном v B -сплайны до q -го порядка включительно: $M_{i-q+1,q}(v), M_{i-q+2,q}(v), \dots, M_{i,q}(v)$, и затем нормируем их

$$N_{r,q}(v) = (v_{r+q} - v_r)M_{r,q}(v), \quad r = i - q + 1, i - q + 2, \dots, i.$$

Напомним, что порядок NURBS поверхности в том или ином параметрическом направлении равен порядку соответствующих B -сплайнов, а порядок B -сплайна на единицу больше степени полиномов, из которых он построен.

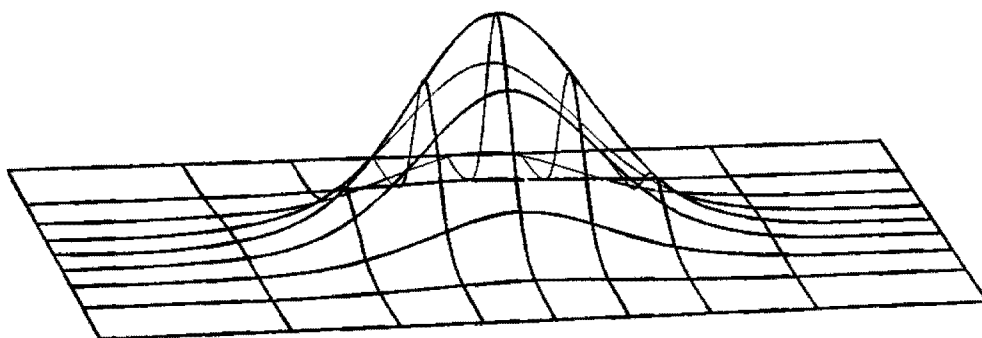


Рис. 3.10.1. NURBS поверхность

На рис. 3.10.1 показана NURBS поверхность 4-го порядка по обоим параметрическим направлениям, построенная на сетке, состоящей из 7×7 вершин. Все вершины, кроме центральной, лежат в одной плоскости. Веса всех вершин равны 1.

Если на каждых четырех соседних характеристических точках построить билинейные поверхности, то мы получим некоторый многогранник, который называется *характеристическим*. Характеристический многогранник поверхности, приведенной на рис. 3.10.1, показан на рис. 3.10.2.

В терминах однородных координат расширенный радиус-вектор (1.4.6) NURBS поверхности определяется формулой

$$\mathbf{R}(u, v) = \sum_{i=1}^n \sum_{j=1}^m N_{i,q}(v) N_{j,k}(u) \mathbf{P}_{ij}, \quad (3.10.2)$$

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

где $\mathbf{P}_{ij} = [w_{ij} \mathbf{p}_{ij} \ w_{ij}]^T$ — расширенные векторы вершин поверхности.

Когда все вершины NURBS поверхности имеют равные веса, то формула (3.10.1) в силу свойства B -сплайнов (2.8.24) примет вид

$$\mathbf{r}(u, v) = \sum_{i=1}^n \sum_{j=1}^m N_{i,q}(v) N_{j,k}(u) \mathbf{P}_{ij}, \quad (3.10.3)$$

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

Эта поверхность называется *B -сплайн поверхностью*. Она также может быть использована в проектировании, так как обладает определенной гладкостью, и для вычисления ее производных не требуется пользоваться формулами (3.9.6)–(3.9.10).

Если построить NURBS поверхность (3.10.3) на базе B -сплайнов второго порядка, то она совпадает со своим характеристическим многогранником и представляет собой совокупность билинейных поверхностей.

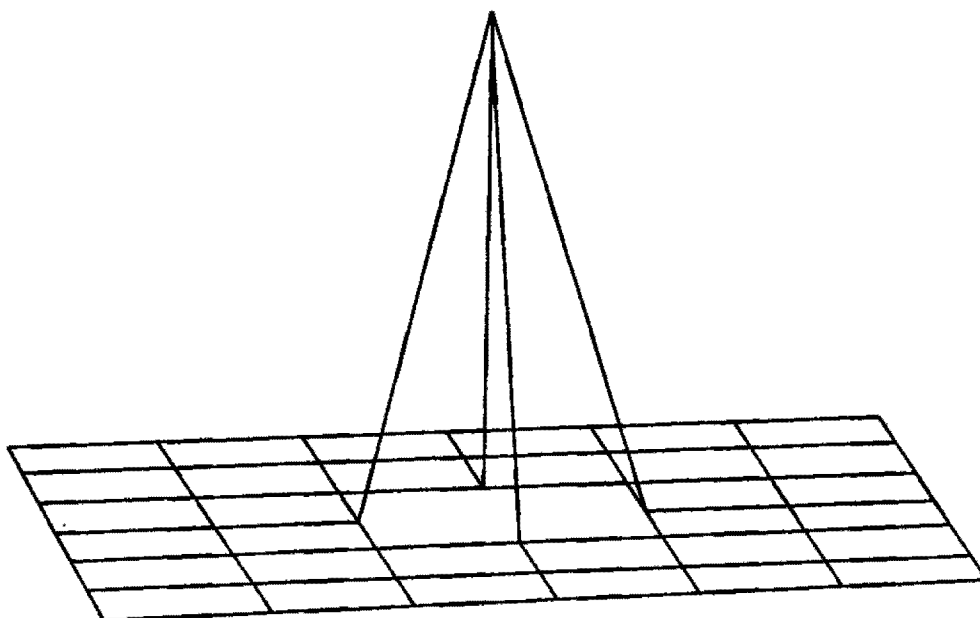


Рис. 3.10.2. Характеристический многогранник NURBS поверхности

Если $n = q$ и $m = k$, т.е. если число рядов вершин n равно порядку q B -сплайнов v -направления, число вершин в ряду m равно порядку k B -сплайнов u -направления, узлы u -последовательности имеют значения $u_1 = u_2 = \dots = u_m = 0$,

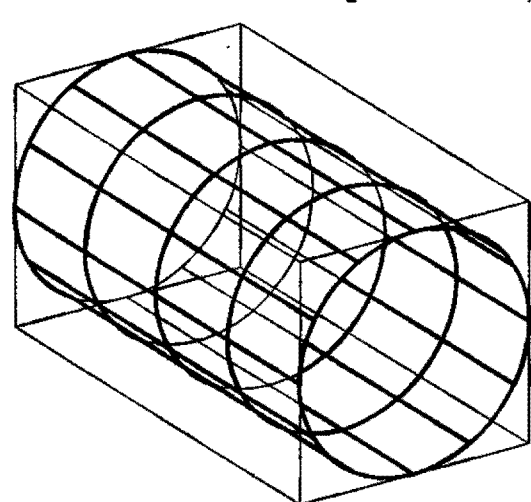


Рис. 3.10.3. Цилиндрическая NURBS поверхность

$u_{m+1} = u_{m+2} = \dots = u_{2m} = 1$, узлы v -последовательности имеют значения $v_1 = v_2 = \dots = v_n = 0$, $v_{n+1} = v_{n+2} = \dots = v_{2n} = 1$, то NURBS поверхность (3.10.1) совпадает с рациональной поверхностью Безье (3.9.1). Это следует из того, что функции Бернштейна являются частными случаями B -сплайнов. Область изменения параметров NURBS поверхности в этом случае представляет собой квадрат со стороной, равной 1.

С помощью поверхности (3.10.1) может быть построена любая поверхность второго порядка. На рис. 3.10.3 приведена цилиндрическая NURBS поверхность и ее характеристический многогранник.

На рис. 3.10.4 приведена тороидальная NURBS поверхность и ее характеристический многогранник. Обе поверхности имеют кратные узлы.

По своему внешнему виду NURBS поверхности и поверхности Безье похожи. Но в отличие от последних порядок NURBS поверхности не связан жестко с

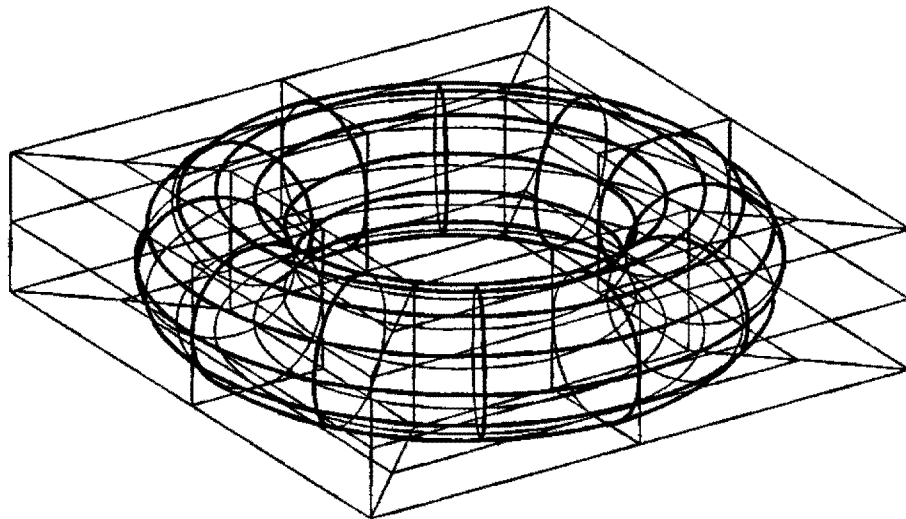


Рис. 3.10.4. Тороидальная NURBS поверхность

количеством вершин и предоставляет возможность строить поверхности невысокого порядка (3-го или 4-го порядка) на большом числе вершин. Эта возможность придает поверхности большую гибкость, так как поверхности Безье, построенные на большом числе вершин, имеют высокий порядок и являются слишком гладкими.

NURBS поверхность является обобщением большинства рассмотренных выше поверхностей: многогранной билинейной поверхности, поверхности Безье, поверхностей второго порядка, рациональных поверхностей.

Каждая линия $u = \text{const}$ или $v = \text{const}$ на поверхности (3.10.1) является NURBS кривой (2.9.1). Но можно построить рациональные неоднородные поверхности на основе B -сплайнов, которые необязательно давали бы NURBS кривые в качестве всех своих параметрических u -линий или v -линий. Например,

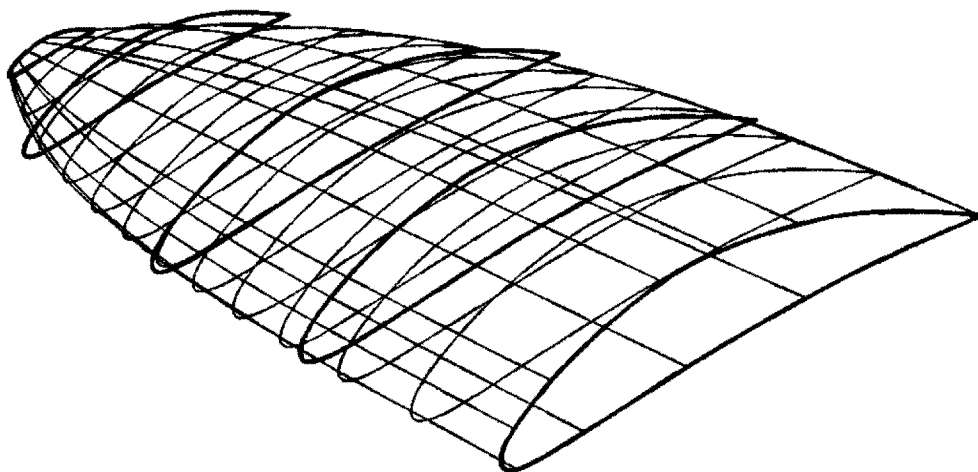


Рис. 3.10.5. NURBS поверхности на семействе кривых

можно получить поверхность, если в формуле (2.9.1) вместо вершин характеристической ломаной подставить кривые $\mathbf{p}_i = \mathbf{p}_i(u)$, $i = 1, 2, \dots, n$, обладающие весами $w_i = w_i(u)$, и имеющие одинаковые параметрические длины

$u_{\min} \leq u \leq u_{\max}$. В результате, получим поверхность, описываемую формулой

$$\mathbf{r}(u, v) = \frac{\sum_{i=1}^n N_{i,m}(v) w_i(u) \mathbf{p}_i(u)}{\sum_{i=1}^n N_{i,m}(v) w_i(u)}, \quad (3.10.4)$$

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

Только параметрические линии $u = \text{const}$ на поверхности (3.10.4) являются NURBS кривыми. Такую поверхность можно назвать плазовой NURBS поверхностью. На рис. 3.10.5. показана плазовая NURBS поверхность и пять кривых, по которым она построена. Частным случаем этой поверхности является поверхность скругления (3.9.3).

3.11. Поверхности треугольной формы

Существует много способов построения поверхностей треугольной формы. Сейчас мы остановимся на поверхностях, которые при треугольной форме имеют треугольную область определения параметров. Для треугольных областей в двумерном пространстве удобно использовать барицентрические координаты.

Барицентрические координаты на плоскости. Пусть в области определения параметров u и v некоторой поверхности задана двумерная декартова система координат. Пусть в этой области заданы три точки A, B, C , не лежащие на

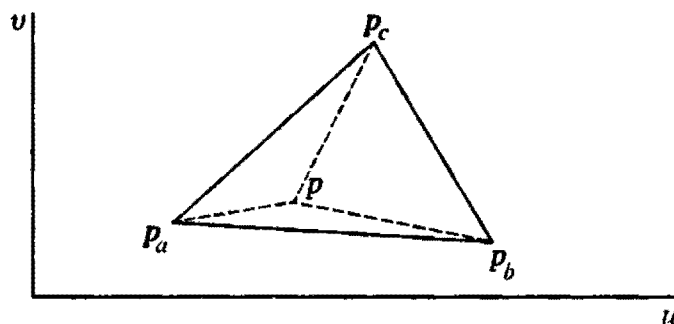


Рис. 3.11.1. Точки на параметрической плоскости

одной прямой. Радиус-векторы точек A, B и C обозначим соответственно через $\mathbf{p}_a = [u_a \ v_a]^T$, $\mathbf{p}_b = [u_b \ v_b]^T$ и $\mathbf{p}_c = [u_c \ v_c]^T$ (рис. 3.11.1).

Положение любой другой точки $\mathbf{p} = [u \ v]^T$ можно описать с помощью точек $\mathbf{p}_a, \mathbf{p}_b, \mathbf{p}_c$ равенством

$$\mathbf{p} = a\mathbf{p}_a + b\mathbf{p}_b + c\mathbf{p}_c, \quad (3.11.1)$$

где коэффициенты a, b, c определены с точностью до множителя. Для полной определенности потребуем, чтобы их сумма была равна единице:

$$a + b + c = 1. \quad (3.11.2)$$

Значения коэффициентов a, b, c , соответствующие декартовым координатам u и v , мы найдем из системы уравнений

$$\begin{aligned} a + b + c &= 1, \\ au_a + bv_b + cv_c &= u, \\ av_a + bv_b + cv_c &= v. \end{aligned} \quad (3.11.3)$$

Коэффициенты a, b, c определяются равенствами

$$\begin{aligned}
 a &= \frac{1}{\Delta_{abc}} \begin{vmatrix} 1 & 1 & 1 \\ u & u_b & u_c \\ v & v_b & v_c \end{vmatrix} = \frac{1}{\Delta_{abc}} (uv_b + u_bv_c + u_cv - vu_b - v_bu_c - v_cu), \\
 b &= \frac{1}{\Delta_{abc}} \begin{vmatrix} 1 & 1 & 1 \\ u_a & u & u_c \\ v_a & v & v_c \end{vmatrix} = \frac{1}{\Delta_{abc}} (u_av + uv_c + u_cv_a - v_a u - vu_c - v_cu_a), \\
 c &= \frac{1}{\Delta_{abc}} \begin{vmatrix} 1 & 1 & 1 \\ u_a & u_b & u \\ v_a & v_b & v \end{vmatrix} = \frac{1}{\Delta_{abc}} (u_av_b + u_bv + uv_a - v_a u_b - v_b u - vu_a),
 \end{aligned} \tag{3.11.4}$$

где определитель системы уравнений (3.11.3)

$$\Delta_{abc} \begin{vmatrix} 1 & 1 & 1 \\ u_a & u_b & u_c \\ v_a & v_b & v_c \end{vmatrix} = u_av_b + u_bv_c + u_cv_a - v_a u_b - v_b u_c - v_c u_a \tag{3.11.5}$$

равен удвоенной площади треугольника ABC ($\Delta_{abc} = 2S_{abc}$).

Во многих случаях для описания положения точки в двухмерном пространстве удобно перейти от координат u и v к трем *барицентрическим координатам*, которыми являются коэффициенты a, b, c . Барицентрические координаты произвольной точки $p = [u \ v]^T$ равны отношению площади треугольника, полученного из треугольника ABC заменой соответствующей вершины точкой p , к площади треугольника ABC . Барицентрические координаты a, b, c точки $p = [u \ v]^T$ вычислены относительно точек p_a, p_b, p_c и являются функциями точки p .

Отметим свойства барицентрических координат. Они подчинены равенству (3.11.2). Если точка p лежит внутри треугольника ABC , то ее барицентрические координаты неотрицательные. Если точка p совпадает с вершиной треугольника ABC , то две ее барицентрические координаты равны нулю. Если точка p принадлежит стороне треугольника ABC или лежит на ее продолжении, то одна барицентрическая координата точки равна нулю. Если точка p лежит вне треугольника ABC , то, по крайней мере, одна барицентрическая координата точки отрицательна.

Любой двухмерный вектор $d = [u \ v]^T$ можно описать с помощью точек p_a, p_b, p_c равенством

$$d = \alpha p_a + \beta p_b + \gamma p_c. \tag{3.11.6}$$

Коэффициенты α, β, γ называют барицентрическими компонентами вектора. Двухмерный вектор w можно представить в виде разности радиус-векторов двух точек, например, $p_1 = a_1 p_a + b_1 p_b + c_1 p_c$ и $p_0 = a_0 p_a + b_0 p_b + c_0 p_c$. Так как $a_1 + b_1 + c_1 = 1$ и $a_0 + b_0 + c_0 = 1$, то барицентрические компоненты α, β, γ двухмерного вектора связаны равенством

$$\alpha + \beta + \gamma = 0. \tag{3.11.7}$$

Значения барицентрических компонент α , β , γ , соответствующие декартовым компонентам u и v вектора \mathbf{d} , мы найдем из системы уравнений

$$\begin{aligned}\alpha + \beta + \gamma &= 0, \\ \alpha u_a + \beta u_b + \gamma u_c &= u, \\ \alpha v_a + \beta v_b + \gamma v_c &= v.\end{aligned}\tag{3.11.8}$$

Компоненты α , β , γ определяются равенствами

$$\begin{aligned}\alpha &= \frac{1}{\Delta_{abc}} \begin{vmatrix} 0 & 1 & 1 \\ u & u_b & u_c \\ v & v_b & v_c \end{vmatrix} = \frac{1}{\Delta_{abc}} (uv_b + u_c v - vu_b - v_c u), \\ \beta &= \frac{1}{\Delta_{abc}} \begin{vmatrix} 1 & 0 & 1 \\ u_a & u & u_c \\ v_a & v & v_c \end{vmatrix} = \frac{1}{\Delta_{abc}} (u_a v + uv_c - v_a u - vu_c), \\ \gamma &= \frac{1}{\Delta_{abc}} \begin{vmatrix} 1 & 1 & 0 \\ u_a & u_b & u \\ v_a & v_b & v \end{vmatrix} = \frac{1}{\Delta_{abc}} (u_b v + uv_a - v_b u - vu_a),\end{aligned}\tag{3.11.9}$$

где определитель Δ_{abc} вычисляется по формуле (3.11.5).

Билинейная треугольная поверхность. Одной из простейших треугольных поверхностей является плоская треугольная поверхность, построенная по трем точкам \mathbf{p}_1 , \mathbf{p}_2 , \mathbf{p}_3 . Эта поверхность представляет собой треугольный аналог билинейной поверхности (3.5.2). Пусть область определения параметров поверхности ограничена треугольником с вершинами в точках \mathbf{p}_a , \mathbf{p}_b , \mathbf{p}_c (рис. 3.11.1).

Введем барицентрические координаты a , b , c (3.11.4). Радиус-вектор \mathbf{r} плоской треугольной поверхности, построенной по трем вершинам \mathbf{p}_1 , \mathbf{p}_2 , \mathbf{p}_3 , опишем векторной функцией трех параметров a , b , c

$$\begin{aligned}\mathbf{r}(a, b, c) &= a\mathbf{p}_1 + b\mathbf{p}_2 + c\mathbf{p}_3, \\ 0 \leq a \leq 1, \quad 0 \leq b \leq 1, \quad 0 \leq c \leq 1.\end{aligned}\tag{3.11.10}$$

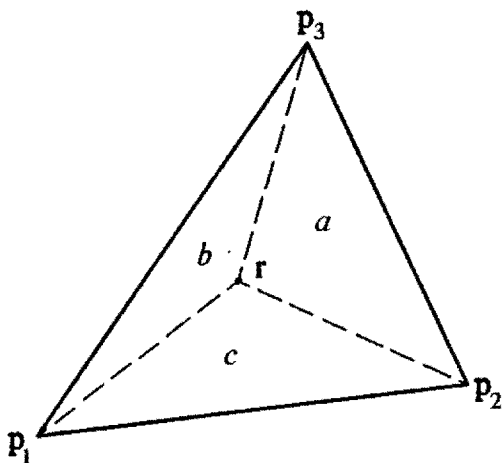


Рис. 3.11.2. Плоская треугольная поверхность

Если мы соединим произвольную точку \mathbf{r} поверхности (3.11.10) отрезками прямых линий с вершинами \mathbf{p}_1 , \mathbf{p}_2 , \mathbf{p}_3 , то этим мы разобьем площадь треугольной поверхности на три части (рис. 3.11.2). Отношение площади части треугольной поверхности к полной площади поверхности равно соответствующей барицентрической координате точки \mathbf{r} .

Фактически радиус-вектор (3.11.10) зависит от двух параметров, так как три параметра a , b , c связаны равенством (3.11.2). Можно переписать зависимость (3.11.10) как функцию двух параметров x и y , например: $a(x, y) = x$, $b(x, y) = y$, $c(x, y) = 1 - x - y$, где x и y принимают значения на треугольной области. Можно переписать зависимость (3.11.10) как функцию тех же двух параметров

x и y , принимающих значения на обычной квадратной области $0 \leq x \leq 1$, $0 \leq y \leq 1$, например: $a = x$, $b = y(1 - x)$, $c = 1 - x - y + xy$ или: $a = x(1 - y)$, $b = y$, $c = 1 - x - y + xy$.

Треугольная поверхность на трех кривых. Еще одну треугольную поверхность построим на трех попарно пересекающихся кривых. Ее можно использовать для скругления пространственного угла, в котором сходятся три скругленных ребра моделируемой детали. Рассмотрим три кривые $q_1(t_1)$, $q_2(t_2)$, $q_3(t_3)$, попарно пересекающиеся в точках p_1 , p_2 , p_3 , как показано на рис. 3.11.3.

Построим поверхность внутри «треугольника», образованного кривыми. Преобразуем с помощью (1.5.3) параметризацию кривых так, чтобы точкам p_3 и p_2 на кривой $q_1(t_1)$ соответствовали параметры 0 и 1, точкам p_1 и p_3 на кривой

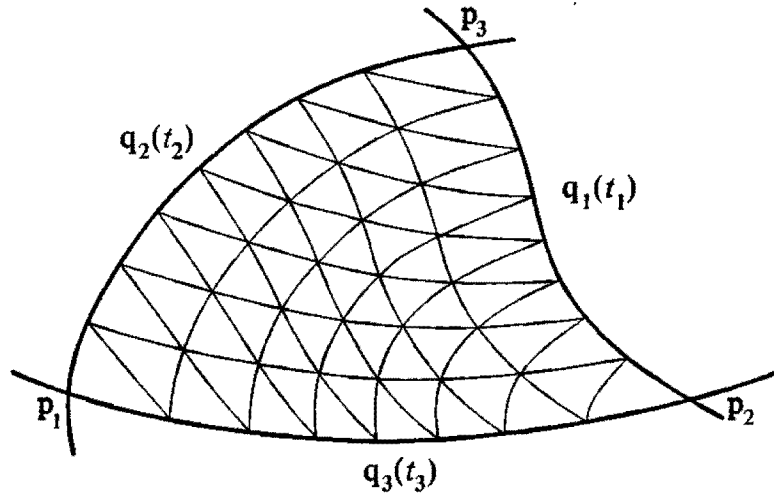


Рис. 3.11.3. Треугольная поверхность на трех кривых

$q_2(t_2)$ соответствовали параметры 0 и 1, точкам p_2 и p_1 на кривой $q_3(t_3)$ соответствовали параметры 0 и 1. Радиус-вектор поверхности, построенной по данным трем кривым, опишем векторной функцией трех барицентрических координат a , b , c

$$\begin{aligned} \mathbf{r}(a, b, c) = & a(\mathbf{q}_3(1 - b) + \mathbf{q}_2(c) - \mathbf{p}_1) + \\ & + b(\mathbf{q}_1(1 - c) + \mathbf{q}_3(a) - \mathbf{p}_2) + c(\mathbf{q}_2(1 - a) + \mathbf{q}_1(b) - \mathbf{p}_3), \quad (3.11.11) \\ & 0 \leq a \leq 1, \quad 0 \leq b \leq 1, \quad 0 \leq c \leq 1, \quad a + b + c = 1. \end{aligned}$$

Фактически радиус-вектор (3.11.11) зависит от двух параметров, так как три параметра a , b , c связаны уравнением (3.11.2). Края поверхности (3.11.11) совпадают с кривыми, по которым она построена. При $c = 0$, $b = 1 - a$ вектор

$$\mathbf{r}(a, 1 - a, 0) = a(\mathbf{q}_3(a) + \mathbf{q}_2(0) - \mathbf{p}_1) + (1 - a)(\mathbf{q}_1(1) + \mathbf{q}_3(a) - \mathbf{p}_2) = \mathbf{q}_3(a)$$

описывает участок кривой $q_3(t_3)$ между точками p_2 и p_1 . При $a = 0$, $c = 1 - b$ вектор

$$\mathbf{r}(0, b, 1 - b) = b(\mathbf{q}_1(b) + \mathbf{q}_3(0) - \mathbf{p}_2) + (1 - b)(\mathbf{q}_2(1) + \mathbf{q}_1(b) - \mathbf{p}_3) = \mathbf{q}_1(b)$$

описывает участок кривой $q_1(t_1)$ между точками p_3 и p_2 . При $b = 0$, $a = 1 - c$ вектор

$$\mathbf{r}(1 - c, 0, c) = c(\mathbf{q}_2(c) + \mathbf{q}_1(0) - \mathbf{p}_3) + (1 - c)(\mathbf{q}_3(1) + \mathbf{q}_2(c) - \mathbf{p}_1) = \mathbf{q}_2(c)$$

описывает участок кривой $q_2(t_2)$ между точками p_1 и p_3 . Приведенные формулы можно получить путем циклической перестановки индексов 1, 2, 3 или 3, 2, 1. Правая часть (3.11.11) симметрична относительно циклической перестановки индексов. Поверхность (3.11.11) является треугольным аналогом линейной поверхности Кунса (3.6.4).

На рис. 3.11.4 приведена треугольная поверхность, построенная на трех одинаковых дугах окружностей. Плоскости окружностей ортогональны плоскости,

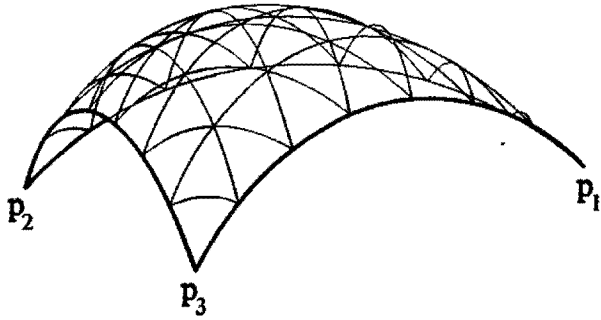


Рис. 3.11.4. Треугольная поверхность на трех дугах окружности

Рис. 3.11.5. Треугольная поверхность на трех дугах окружности

в которой лежат точки P_1, P_2, P_3 . Каждая дуга содержит четверть окружности. На рис. 3.11.5 приведена та же треугольная поверхность под другим ракурсом.

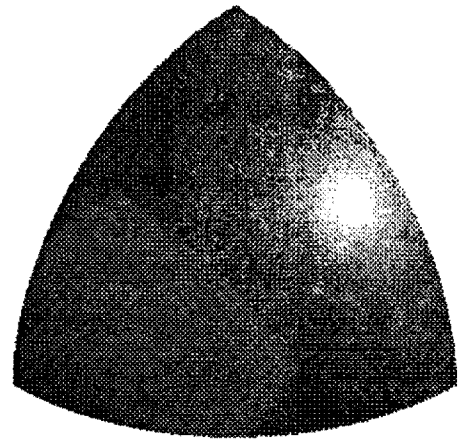
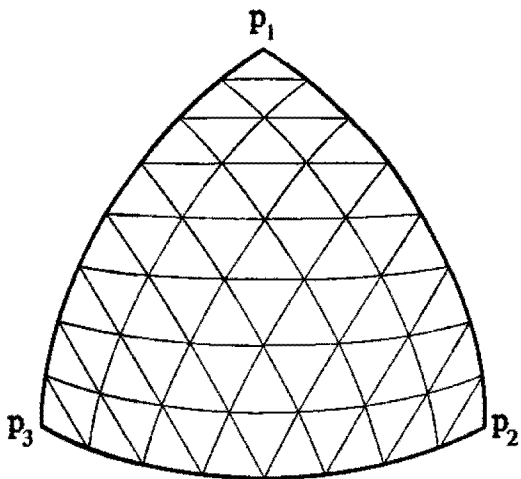


Рис. 3.11.6. Поверхность на трех ортогональных дугах в четверть окружности

Рис. 3.11.7. Поверхность на трех ортогональных дугах в четверть окружности

Если плоскости дуг окружностей сделать взаимно ортогональными, то треугольная поверхность будет описывать часть сферической поверхности. Такая поверхность приведена на рис. 3.11.6 и 3.11.7.

3.12. Треугольные поверхности Безье

Рассмотрим треугольные аналоги поверхности Безье. Пусть имеется совокупность некоторых характеристических точек p_{ijk} , образующих треугольную сетку. На рис. 3.12.1 показана треугольная сетка, вдоль каждой стороны которой расположены 6 характеристических точек. Точки расположены в виде треугольника, причем вдоль каждой стороны треугольника расположено одинаковое число точек. Пусть это число равно $n + 1$. В нашем случае $n = 5$. Общее число характеристических точек равно $(n + 1)(n + 2)/2$. Каждый из трех

индексов точки соответствует своей вершине треугольника. Точка, лежащая в вершине треугольника, имеет значение индекса, соответствующего этой вершине, равное n (максимальное значение), а значения других индексов равно нулю. Точки, принадлежащие ряду вдоль одной из трех сторон треугольника,

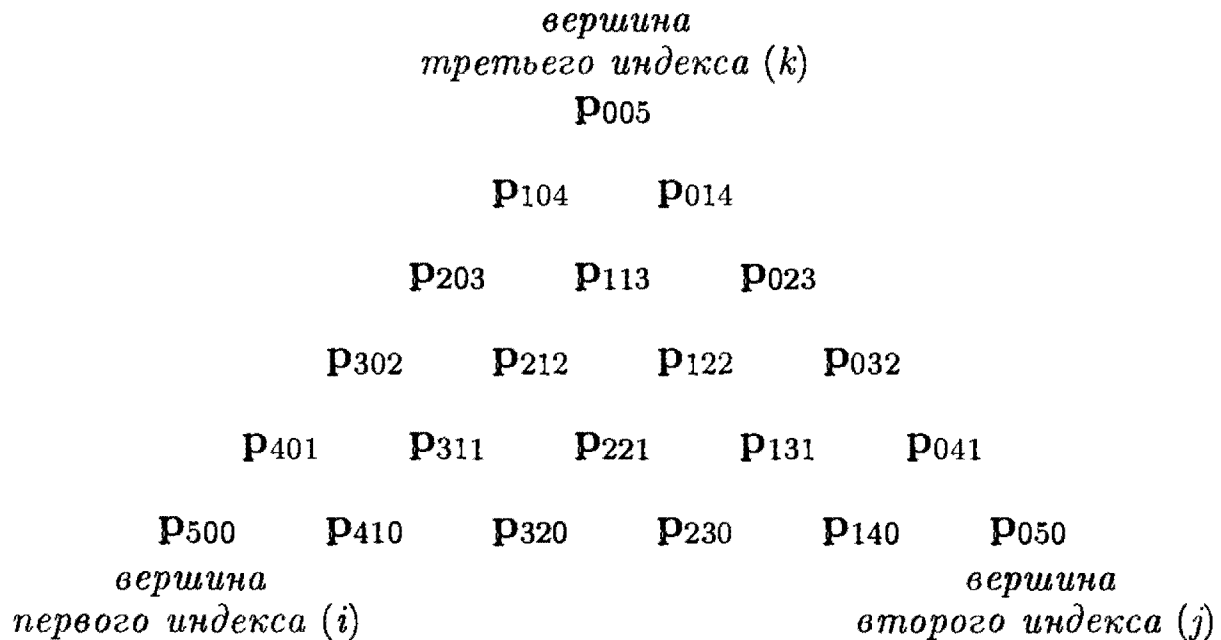


Рис. 3.12.1. Нумерация точек

имеют нулевой индекс, соответствующий противоположной вершине. Чем ближе ряд, в котором лежат точки, к вершине, тем большее значение имеет индекс этой вершины.

Индексы точки P_{ijk} означают целочисленное расстояние от нее до соответствующей стороны треугольника, измеренное количеством отделяющих рядов. Сумма индексов характеристической точки P_{ijk} имеет постоянное значение

$$i + j + k = n. \quad (3.12.1)$$

Треугольную поверхность Безье опишем с помощью барицентрических координат a, b, c . Будем считать, что поверхность Безье имеет ту же область определения параметров, что и поверхность (3.11.10). Радиус-вектор треугольной поверхности Безье, построенной по характеристическим точкам P_{ijk} , описывается формулой

$$\mathbf{r}(a, b, c) = \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n B^n_{ijk}(a, b, c) P_{ijk} = \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n \frac{n!}{i!j!k!} a^i b^j c^k P_{ijk}, \quad (3.12.2)$$

$$0 \leq a \leq 1, \quad 0 \leq b \leq 1, \quad 0 \leq c \leq 1, \quad a + b + c = 1. \quad (3.12.3)$$

Суммирование в (3.12.2) выполняется по всем характеристическим точкам, а трехиндексные функции Бернштейна имеют вид

$$B^n_{ijk}(a, b, c) = \frac{n!}{i!j!k!} a^i b^j c^k, \quad (3.12.4)$$

где должно выполняться равенство (3.12.1).

Треугольная поверхность Безье, построенная по 7 точкам вдоль каждой стороны, показана на рис. 3.12.2. Характеристический многогранник треугольной поверхности Безье приведен на рис. 3.12.3.

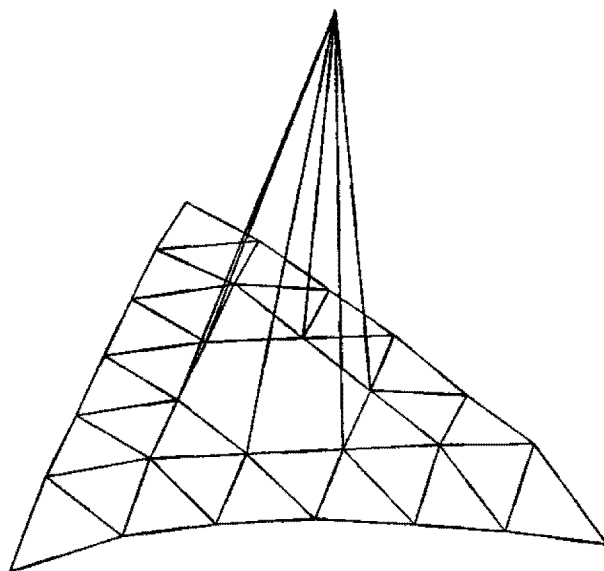
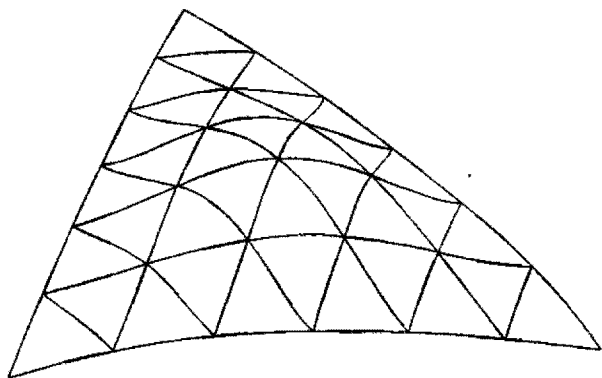


Рис. 3.12.2. Треугольная поверхность Безье

Рис. 3.12.3. Характеристический многогранник треугольной поверхности Безье

Крайние линии поверхности (3.12.2) описываются векторами $\mathbf{r}(a, 1-a, 0)$, $\mathbf{r}(0, b, 1-b)$, $\mathbf{r}(1-c, 0, c)$ и представляют собой кривые Безье (2.5.3). Покажем это на примере крайней кривой $\mathbf{r}(a, 1-a, 0)$. Для нее $c = 0$, $k = 0$ и функции (3.12.4) примут вид

$$B^n_{ij0}(a) = \frac{n!}{i!j!} a^i b^j = \frac{n!}{i!(n-i)!} a^i (1-a)^{n-i}, \quad (3.12.5)$$

что совпадает с (2.5.4).

Покажем, что трехиндексные функции Бернштейна удовлетворяют равенству

$$\sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n B^n_{ijk}(a, b, c) = 1. \quad (3.12.6)$$

Суммирование (3.12.6) выполним по строчкам третьего индекса $k = 0, 1, 2, \dots, n$. В k -й строке третьего индекса содержится $m = n - k$ характеристических точек.

$$\begin{aligned} \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n B^n_{ijk}(a, b, c) &= \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n \frac{n!}{i!j!k!} a^i b^j c^k = \\ &= \sum_{k=0}^n \frac{n!}{k!(n-k)!} c^k \sum_{i=0}^{n-k} \frac{(n-k)!}{i!(n-k-i)!} a^i b^{n-k-i} = \\ &= \sum_{k=0}^n C^k_n c^k \sum_{i=0}^{n-k} C^i_{n-k} a^i b^{n-k-i} = \sum_{k=0}^n C^k_n c^k (a+b)^{n-k} = (a+b+c)^n = 1, \end{aligned}$$

что и требовалось доказать.

Алгоритм Де Кастелье. Трехиндексные функции Бернштейна удовлетворяют рекуррентному соотношению

$$B^n_{ijk} = aB^{n-1}_{i-1jk} + bB^{n-1}_{ij-1k} + cB^{n-1}_{ijk-1}. \quad (3.12.7)$$

Это соотношение доказывается непосредственной подстановкой:

$$\begin{aligned} aB^{n-1}_{i-1jk}(a, b, c) + bB^{n-1}_{ij-1k}(a, b, c) + cB^{n-1}_{ijk-1}(a, b, c) &= \\ &= a \frac{(n-1)!}{(i-1)!j!k!} a^{i-1} b^j c^k + b \frac{(n-1)!}{i!(j-1)!k!} a^i b^{j-1} c^k + c \frac{(n-1)!}{i!j!(k-1)!} a^i b^j c^{k-1} = \\ &= \frac{i}{n} \frac{n!}{i!j!k!} a^i b^j c^k + \frac{j}{n} \frac{n!}{i!j!k!} a^i b^j c^k + \frac{k}{n} \frac{n!}{i!j!k!} a^i b^j c^k = \\ &= \frac{i+j+k}{n} \frac{n!}{i!j!k!} a^i b^j c^k = B^n_{ijk}(a, b, c), \end{aligned}$$

что и требовалось доказать.

Начнем вычисление с функции $B^0_{000}(a, b, c) = 1$.

Далее получим $B^1_{100}(a, b, c) = a$, $B^1_{010}(a, b, c) = b$, $B^1_{001}(a, b, c) = c$. Функция Бернштейна, один из индексов которой равен отрицательному числу, считается равной нулю. Подставим рекуррентное соотношение (3.12.7) в выражение (3.12.2), выделив крайние точки и учитывая, что $B^n_{n00} = a^n$, $B^n_{0n0} = b^n$, $B^n_{00n} = c^n$, $B^n_{0jk} = b^j c^k$, $B^n_{i0k} = a^i c^k$, $B^n_{ij0} = a^i b^j$, получим аналогично выводу (2.5.6)

$$\begin{aligned} r(a, b, c) &= \sum B^n_{ijk} \mathbf{P}_{ijk} = \sum (aB^{n-1}_{i-1jk} + bB^{n-1}_{ij-1k} + cB^{n-1}_{ijk-1}) \mathbf{P}_{ijk} = \\ &= a \sum_{\substack{i,j,k=0, \\ i+j+k=n-1}}^{n-1} B^{n-1}_{ijk} \mathbf{P}_{i+1jk} + b \sum_{\substack{i,j,k=0, \\ i+j+k=n-1}}^{n-1} B^{n-1}_{ijk} \mathbf{P}_{ij+1k} + \\ &+ c \sum_{\substack{i,j,k=0, \\ i+j+k=n-1}}^{n-1} B^{n-1}_{ijk} \mathbf{P}_{ijk+1} = ar_{i+1jk}^{(n-1)} + br_{ij+1k}^{(n-1)} + cr_{ijk+1}^{(n-1)}, \end{aligned}$$

где введены обозначения:

$$r_{i+1jk}^{(n-1)} = \sum_{\substack{i,j,k=0, \\ i+j+k=n-1}}^n B^{n-1}_{ijk} \mathbf{P}_{i+1jk},$$

$$r_{ij+1k}^{(n-1)} = \sum_{\substack{i,j,k=0, \\ i+j+k=n-1}}^n B^{n-1}_{ijk} \mathbf{P}_{ij+1k},$$

$$r_{ijk+1}^{(n-1)} = \sum_{\substack{i,j,k=0, \\ i+j+k=n-1}}^n B^{n-1}_{ijk} \mathbf{P}_{ijk+1}.$$

Выполнив описанные действия для $\mathbf{r}_{i+1jk}^{(n-1)}$, $\mathbf{r}_{ij+1k}^{(n-1)}$ и $\mathbf{r}_{ijk+1}^{(n-1)}$, в конце придем к равенствам

$$\mathbf{r}_{ijk}^{(0)} = B^0_{000}(a, b, c)\mathbf{P}_{ijk} = \mathbf{P}_{ijk}.$$

Обозначим $\mathbf{r}(a, b, c)$ через $\mathbf{r}_{000}^{(n)}$. В результате получим рекуррентное соотношение для вычисления точки треугольной поверхности Безье

$$\mathbf{r}_{ijk}^{(m)} = a\mathbf{r}_{i+1jk}^{(m-1)} + b\mathbf{r}_{ij+1k}^{(m-1)} + c\mathbf{r}_{ijk+1}^{(m-1)}, \quad (3.12.8)$$

где $i + j + k = n - m$ ($i, j, k \geq 0$), $\mathbf{r}_{ijk}^{(0)} = \mathbf{P}_{ijk}$, $\mathbf{r}_{000}^{(n)} = \mathbf{r}(a, b, c)$. Алгоритм, описываемый соотношением (3.12.8), называется *алгоритмом Де Кастелье*. Точки $\mathbf{r}_{ijk}^{(m)}$, получаемые в процессе вычисления, называются точками Де Кастелье. Алгоритм Де Кастелье позволяет вычислить любую точку треугольной поверхности Безье по характеристическим точкам, ничего не зная о функциях Бернштейна. Коэффициенты a, b и c при координатах точек $\mathbf{r}_{i+1jk}^{(m-1)}$, $\mathbf{r}_{ij+1k}^{(m-1)}$ и $\mathbf{r}_{ijk+1}^{(m-1)}$ являются барицентрическими координатами. Треугольную поверхность Безье можно определить как поверхность, точки которой определяются рекуррентным соотношением (3.12.8) по совокупности характеристических точек.

Увеличение числа характеристических точек. От треугольной поверхности Безье n -й степени можно перейти к треугольной поверхности $(n + 1)$ -й степени аналогично тому, как это было сделано для кривых Безье. Для этого умножим правую часть равенства (3.12.2) на сумму $a + b + c$, равную 1, и получим

$$\begin{aligned} \mathbf{r}(a, b, c) &= \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n B^n_{ijk}\mathbf{P}_{ijk} = \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n \frac{n!}{i!j!k!} a^i b^j c^k \mathbf{P}_{ijk} (a + b + c) = \\ &= \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n \frac{n!}{i!j!k!} a^{i+1} b^j c^k \mathbf{P}_{ijk} + \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n \frac{n!}{i!j!k!} a^i b^{j+1} c^k \mathbf{P}_{ijk} + \\ &+ \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n \frac{n!}{i!j!k!} a^i b^j c^{k+1} \mathbf{P}_{ijk} = \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n \frac{(n+1)!}{(i+1)!j!k!} a^{i+1} b^j c^k \mathbf{P}_{ijk} \frac{i+1}{n+1} + \\ &+ \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n \frac{(n+1)!}{i!(j+1)!k!} a^i b^{j+1} c^k \mathbf{P}_{ijk} \frac{j+1}{n+1} + \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n \frac{(n+1)!}{i!j!(k+1)!} a^i b^j c^{k+1} \mathbf{P}_{ijk} \frac{k+1}{n+1} = \\ &= \sum_{\substack{i,j,k=0, \\ i+j+k=n+1}}^{n+1} \frac{(n+1)!}{i!j!k!} a^i b^j c^k \frac{i\mathbf{P}_{i-1jk}}{n+1} + \sum_{\substack{i,j,k=0, \\ i+j+k=n+1}}^{n+1} \frac{(n+1)!}{i!j!k!} a^i b^j c^k \frac{j\mathbf{P}_{ij-1k}}{n+1} + \\ &+ \sum_{\substack{i,j,k=0, \\ i+j+k=n+1}}^{n+1} \frac{(n+1)!}{i!j!k!} a^i b^j c^k \frac{k\mathbf{P}_{ijk-1}}{n+1} = \sum_{\substack{i,j,k=0, \\ i+j+k=n+1}}^{n+1} B^{n+1}_{ijk} \mathbf{P}^*_{ijk}, \quad (3.12.9) \end{aligned}$$

где $\mathbf{P}^*_{ijk} = \frac{i\mathbf{P}_{i-1jk} + j\mathbf{P}_{ij-1k} + k\mathbf{P}_{ijk-1}}{n+1}$ — новые характеристические точки.

Новых характеристических точек на один ряд вдоль каждой из трех сторон треугольника стало больше. Точки в вершинах треугольника остались прежние: $\mathbf{P}^*_{n+100} = \mathbf{P}_{n00}$, $\mathbf{P}^*_{0n+10} = \mathbf{P}_{0n0}$, $\mathbf{P}^*_{00n+1} = \mathbf{P}_{00n}$. Выражение (3.12.9) описывает ту же поверхность, что и выражение (3.12.2), только через другие характеристические точки.

Треугольная рациональная поверхность Безье. Пусть треугольная поверхность Безье построена на совокупности характеристических точек \mathbf{P}_{ijk} , образующих треугольную сетку, показанную на рис. 3.12.1. Припишем каждой точке \mathbf{P}_{ijk} вес w_{ijk} . Построим треугольную рациональную поверхность Безье. Ее радиус-вектор описывается формулой

$$\mathbf{r}(a, b, c) = \frac{\sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n B^n_{ijk}(a, b, c) w_{ijk} \mathbf{P}_{ijk}}{\sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n B^n_{ijk}(a, b, c) w_{ijk}}, \quad (3.12.10)$$

$$0 \leq a \leq 1, \quad 0 \leq b \leq 1, \quad 0 \leq c \leq 1, \quad a + b + c = 1,$$

где суммирование выполняется по всем характеристическим точкам. В силу свойства (3.12.6), поверхность Безье (3.12.2) является частным случаем рациональной поверхности (3.12.10) при одинаковых весах точек. Чем больше вес точки, тем ближе к ней проходит поверхность. Напомним, что роль играет не абсолютное значение веса каждой точки, а отношения весов точек друг к другу. Если веса всех характерных точек одинаковы, то мы приходим к обычной поверхности Безье (3.12.2).

В терминах однородных координат расширенный радиус-вектор (1.4.6) треугольной поверхности Безье определяется формулой

$$\mathbf{R}(a, b, c) = \sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n B^n_{ijk}(a, b, c) w_{ijk} \mathbf{P}_{ijk}, \quad (3.12.11)$$

$$0 \leq a \leq 1, \quad 0 \leq b \leq 1, \quad 0 \leq c \leq 1, \quad a + b + c = 1,$$

где $\mathbf{P}_{ijk} = [w_{ijk} \mathbf{P}_{ijk} \ w_{ijk}]^T$ — расширенные радиус-векторы вершин поверхности. На рис. 3.12.4 показана треугольная поверхность Безье, аналогичная приведенной на рис. 3.12.2 поверхности, с весом средней точки, равным 5.

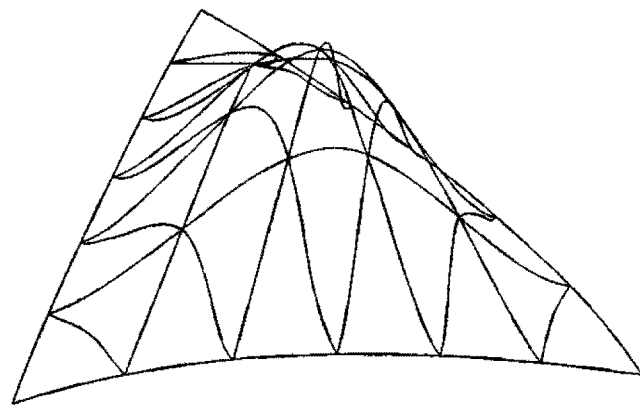


Рис. 3.12.4. Треугольная рациональная поверхность Безье.

3.13. Треугольные сплайновые поверхности

Рассмотрим поверхности, которые можно считать обобщением треугольной поверхности Безье. Они строятся в виде взвешенной суммы некоторых характеристических точек. Каждая характеристическая точка входит в сумму с коэффициентом, который является функцией координат двухмерной точки на области определения параметров поверхности. Каждая такая функция отлична от нуля на некоторой ограниченной двухмерной области. Базисные функции координат u и v двухмерных точек, отличные от нуля на ограниченных двухмерных областях простой формы, называют *двухмерными B -сплайнами* или *симплексными сплайнами*. Они представляют собой аналоги рассмотренных выше B -сплайнов и имеют с ними много общего.

Рассмотрим аналогию между B -сплайнами и симплексными сплайнами. Каждый B -сплайн строится на последовательности узлов. Узлами последовательности можно разделить область определения B -сплайна на полуоткрытые подобласти. При этом каждое значение параметра B -сплайна может принадлежать только одной подобласти. Область определения симплексного сплайна можно разбить на несколько подобластей. Каждая точка $p = [u \ v]^T$ области определения может принадлежать только одной подобласти. Нужно условиться, какой из двух подобластей принадлежат точки разделяющей их линии. Для этого вводится понятие полуоткрытой выпуклой области.

Полуоткрытая выпуклая область. Через $[V]$ обозначим выпуклую область, которую покрывает некоторый выпуклый многоугольник V , а через $[V)$ обозна-

чим *полуоткрытую выпуклую область* многоугольника V . Не все граничные точки выпуклой области $[V]$ принадлежат полуоткрытой выпуклой области $[V)$. Точка $p = [u \ v]^T$ принадлежит полуоткрытой выпуклой области $[V)$, если существуют такие $\varepsilon > 0$ и $\eta > 0$, при которых точки $p = [u \ v]^T$, $p_1 = [u + \varepsilon \ v]^T$ и $p_2 = [u + \varepsilon \ v + \eta]^T$ принадлежат $[V]$.

На рис. 3.13.1 приведена полуоткрытая выпуклая область. Для области, показанной на рис. 3.13.1, внутренние точки многоугольника и точки сплошных линий принадлежат $[V)$, а точки пунктирных линий и точки их стыковки со сплошными линиями не принадлежат $[V)$.

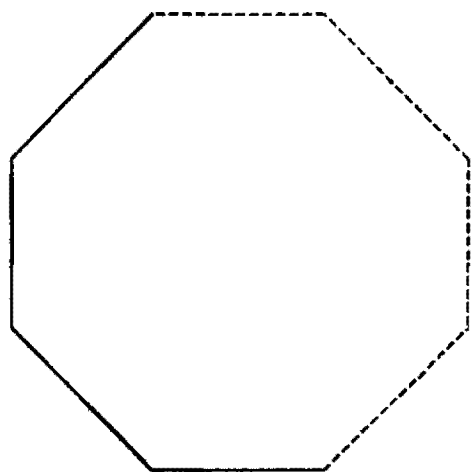


Рис. 3.13.1. Полуоткрытая выпуклая область

Симплексный сплайн. Симплексные сплайны являются аналогами ненормированных B -сплайнов. Они представляют собой кусочно-полиномиальные функции точек двухмерного пространства. Каждый симплексный сплайн больше нуля внутри своей полуоткрытой выпуклой области и равен нулю вне области.

Пусть задана последовательность двухмерных точек p_1, p_2, \dots, p_m , хотя бы три из которых не совпадают друг с другом и не лежат на одной прямой. Для такой последовательности можно построить выпуклый многоугольник $V = \{p_1, p_2, \dots, p_m\}$ с вершинами в точках последовательности, так чтобы все точки заданной последовательности принадлежали $[V]$. В общем случае часть точек заданной последовательности будет лежать в вершинах многоугольника V , часть — на его сторонах и остальная часть — внутри многоугольника V . Некоторые точки последовательности могут совпадать. Точки p_1, p_2, \dots, p_m после-

довательности называют *узлами*. На базе узлов p_1, p_2, \dots, p_m можно построить симплексный сплайн $M^V(p)$ n -го порядка, где $n = m - 3$. Полуоткрытую выпуклую область $[V]$ многоугольника V называют *носителем* симплексного сплайна или его областью определения. Все узлы симплексного сплайна равноправны. Симплексный сплайн является функцией точки $p = [u \ v]^T$. Введем еще одно обозначение: через $M^{V \setminus k}(p)$ обозначим симплексный сплайн $(n-1)$ -го порядка, построенный на последовательности узлов, полученной из заданной последовательности путем исключения из нее вершины p_k .

Симплексный сплайн n -го порядка определяется рекурсивно как взвешенная сумма трех симплексных сплайнов $(n-1)$ -го порядка. Используя равноправие узлов последовательности p_1, p_2, \dots, p_m , выберем из нее произвольную тройку не лежащих на одной прямой узлов p_a, p_b, p_c . Будем считать, что при обходе точек p_a, p_b, p_c мы будем двигаться против часовой стрелки. По формулам (3.11.4) вычислим барицентрические координаты a, b, c точки p относительно точек p_a, p_b, p_c . Симплексный сплайн n -го порядка определяется формулой

$$M^V(p) = aM^{V \setminus a}(p) + bM^{V \setminus b}(p) + cM^{V \setminus c}(p), \quad (3.13.1)$$

где $M^{V \setminus a}(p), M^{V \setminus b}(p), M^{V \setminus c}(p)$ — симплексные сплайны $(n-1)$ -го порядка, построенные на последовательностях узлов, полученных из последовательности p_1, p_2, \dots, p_m путем исключения из нее соответственно вершины p_a, p_b, p_c . Результат формулы (3.13.1) не зависит от выбранной тройки узлов p_a, p_b, p_c . Вычисление значения симплексного сплайна n -го порядка начинается с вычисления значений симплексных сплайнов нулевого порядка на треугольных областях. Пусть вершинами некоторого треугольника T являются точки $p_i = [u_i \ v_i]^T, p_j = [u_j \ v_j]^T, p_k = [u_k \ v_k]^T$. Значение симплексного сплайна нулевого порядка на невырожденной треугольной области T определяется формулой

$$M^T(p) = \begin{cases} \frac{1}{|\Delta_{ijk}|}, & \text{если } p \in [T], \\ 0, & \text{если } p \notin [T], \end{cases} \quad (3.13.2)$$

где $\Delta_{ijk} = \begin{vmatrix} 1 & 1 & 1 \\ u_i & u_j & u_k \\ v_i & v_j & v_k \end{vmatrix}$. Симплексный сплайн нулевого порядка является разрывной константной функцией положения точки p . Симплексный сплайн первого порядка является непрерывной функцией с разрывными первыми производными.

Симплексный сплайн $M^V(p)$ всегда неотрицателен. Он отличен от нуля внутри своей полуоткрытой выпуклой области $[V]$ и равен нулю вне ее.

В общем случае симплексный сплайн n -го порядка является кусочно-полиномиальной функцией степени n , имеющей непрерывные производные до $(n-1)$ -го порядка включительно. В используемых нами обозначениях порядок симплексного сплайна совпадает с наивысшей степенью полиномов, описывающих его.

Обратим внимание на аналогию формулы Кокса–Де Бура (2.8.30) и формулы (3.13.1). Обе формулы описывают рекуррентное соотношение для сплайна $(m-\lambda)$ -го порядка (порядок равен степени соответствующей кусочно-полиноми-

альной функции), построенного на последовательности m узлов, где λ — число барицентрических координат. В обоих случаях вычисления начинаются со сплайнов нулевого порядка для λ несовпадающих узлов. Эту аналогию можно продолжить для пространств большей размерности.

Для симплексного сплайна можно вычислить производную в направлении любого вектора на плоскости параметров. Пусть дан двухмерный вектор $t = [u \ v]^T$. По формулам (3.11.9) этот вектор можно описать с помощью барицентрических компонент α, β, γ относительно точек p_a, p_b, p_c равенством

$$t = \alpha p_a + \beta p_b + \gamma p_c.$$

Производная симплексного сплайна (3.13.1) в направлении вектора t вычисляется по формуле

$$\frac{\partial M^v(p)}{\partial t} = n(\alpha M^{v \setminus a}(p) + \beta M^{v \setminus b}(p) + \gamma M^{v \setminus c}(p)). \quad (3.13.3)$$

Производные симплексных сплайнов n -го порядка определяются через симплексные сплайны $(n-1)$ -го порядка. В качестве точек p_a, p_b, p_c могут быть взяты любые три не лежащие на одной прямой узла дифференцируемого симплексного сплайна.

На рис. 3.13.2 приведен симплексный сплайн первого порядка, построенный на прямоугольной области (значение симплексного сплайна откладывается в

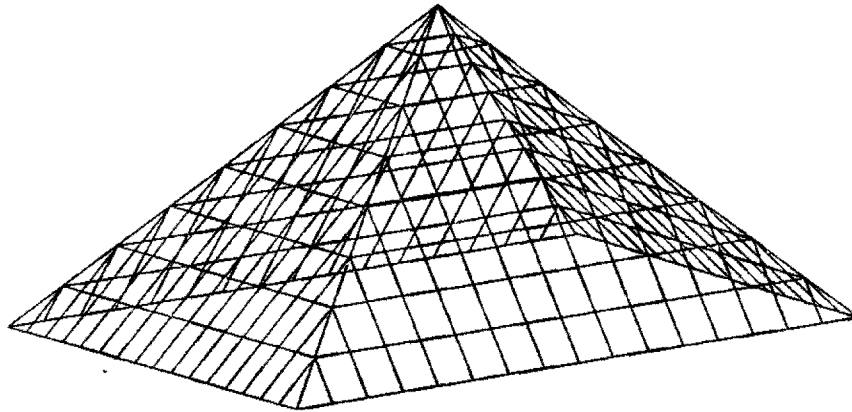


Рис. 3.13.2. Симплексный сплайн первого порядка

направлении, перпендикулярном плоскости параметров сплайна).

На рис. 3.13.3 приведен симплексный сплайн третьего порядка, построенный на шести несовпадающих узлах.

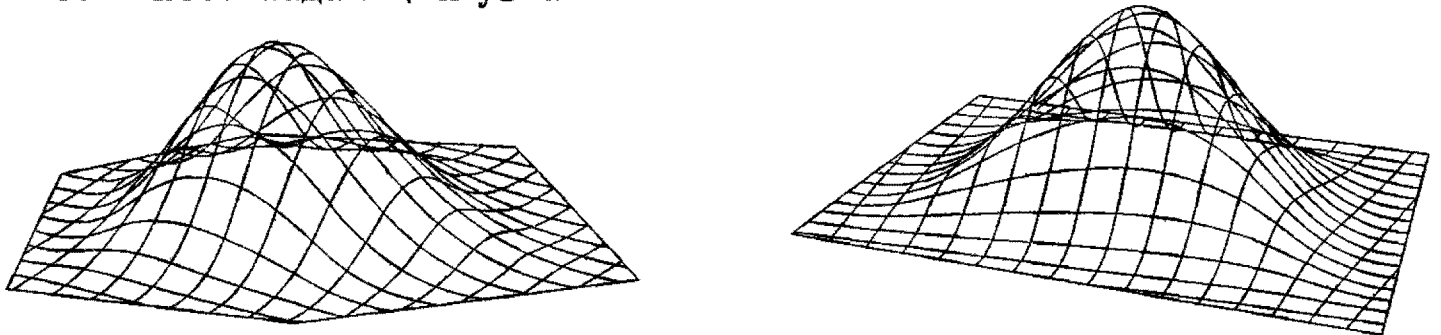


Рис. 3.13.3. Симплексный сплайн третьего порядка

Рис. 3.13.4. Симплексный сплайн пятого порядка с кратными узлами

Последовательность узлов B -сплайна может содержать кратные узлы. Необходимо только, чтобы в последовательности нашлись хотя бы два не лежащих

на одной прямой узла. Узлы p_1, p_2, \dots, p_m симплексного сплайна также могут быть кратными. Для построения симплексного сплайна необходимо только, чтобы среди узлов p_1, p_2, \dots, p_m нашлись хотя бы три не лежащих на одной прямой узла. В кратных узлах высшие производные симплексного сплайна теряют гладкость. На рис. 3.13.4 приведен симплексный сплайн пятого порядка, построенный на восьми узлах, образующих четырехугольную область. В каждой вершине четырехугольной области находится по два совпадающих узла.

Некоторые узлы симплексного сплайна могут лежать внутри его выпуклого многоугольника. На рис. 3.13.5 приведен симплексный сплайн третьего порядка, построенный на шести узлах, три из которых лежат в вершинах треугольной области, а остальные три узла совпадают и лежат в центре этой треугольной области. Все производные в центре такого сплайна теряют непрерывность.

Некоторые узлы симплексного сплайна могут лежать на границе его выпуклого многоугольника. На рис. 3.13.6 приведен симплексный сплайн второго

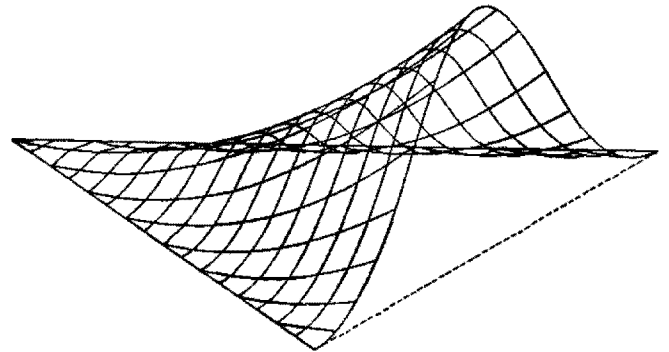
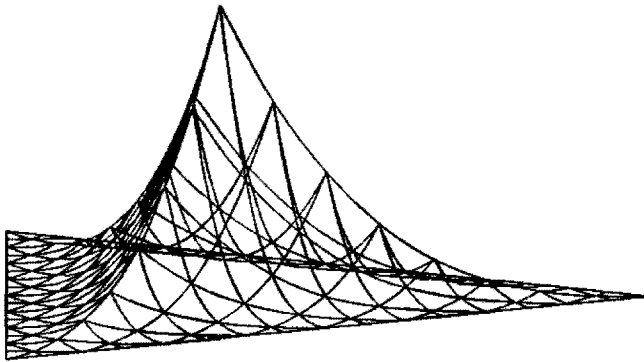


Рис. 3.13.5. Симплексный сплайн третьего порядка с тремя кратными узлами в центре

Рис. 3.13.6. Симплексный сплайн второго

порядка, построенный на пяти узлах, четыре из которых лежат вдоль стороны треугольника. На этой стороне производные сплайна и сам симплексный сплайн терпят разрыв.

В местах концентрации узлов значение симплексного сплайна заметно больше, чем в остальных местах. Если все узлы симплексного сплайна расположены

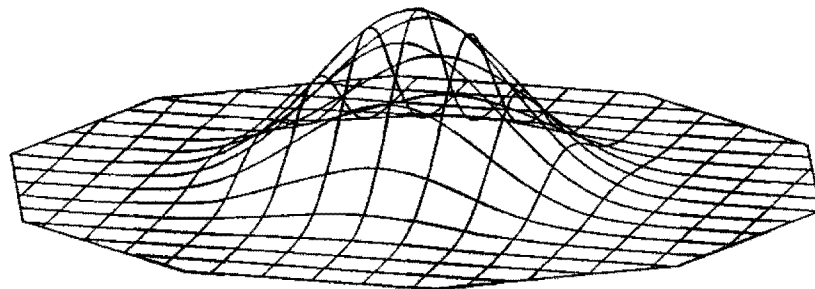


Рис. 3.13.7. Симплексный сплайн седьмого порядка

в вершинах его выпуклого многоугольника, то чем больше порядок сплайна, тем меньше центральная существенно отличная от нуля часть сплайна. На рис. 3.13.7 приведен симплексный сплайн, узлы которого лежат в вершинах правильного десятиугольника.

Трехиндексные функции Бернштейна (3.12.4) являются частными случаями симплексных сплайнов. Функция Бернштейна

$$B^n_{ijk}(a, b, c) = \frac{n!}{i!j!k!} a^i b^j c^k$$

с точностью до множителя равна симплексному сплайну n -го порядка, построенному на последовательности $(n+3)$ узлов, из которых $(i+1)$ узлов совпадают с точкой p_a , $(j+1)$ узлов совпадают с точкой p_b и $(k+1)$ узлов совпадают с точкой p_c . Пусть точки p_a, p_b, p_c являются вершинами треугольника T . На треугольной области T можно построить $C_{n+2}^2 = \frac{(n+2)!}{2!n!} = \frac{(n+1)(n+2)}{2}$ симплексных сплайнов n -го порядка. Эти симплексные сплайны и соответствующие им функции Бернштейна связаны равенством

$$M_{ijk}(p) = \frac{1}{\Delta_{abc}} \frac{n!}{i!j!k!} a^i b^j c^k, \quad (3.13.4)$$

где $p = ap_a + bp_b + cp_c$, Δ_{abc} — удвоенная площадь треугольника T . Симплексный сплайн $M_{ijk}(p)$ построен на последовательности узлов: p_a — кратности $(i+1)$, p_b — кратности $(j+1)$, p_c — кратности $(k+1)$.

DMS поверхности. NURBS кривые мы строим на семействе B -сплайнов, которые имеют общие узлы. Носители (области определения) B -сплайнов перекрываются. Это позволяет строить гладкие кривые. B -сплайны NURBS кривых нормированы, их сумма для каждого значения параметра равна единице, что позволяет вводить для характеристических точек веса. При построении поверхностей на семействе симплексных сплайнов нужно, чтобы сплайны семейства имели общие узлы, а их полуоткрытые выпуклые области перекрывались. В противном случае сплайновая поверхность не будет гладкой. Кроме того, если сумма симплексных сплайнов для каждой двумерной точки параметрической области поверхности будет равна единице, то это позволит нам строить рациональные поверхности, в которых характеристические точки могут иметь веса.

Дамен (Dahmen), Микели (Micchelli) и Зайдель (Seidel) предложили один из способов построения поверхностей на базе симплексных сплайнов. Этот способ имеет много общего со способом построения треугольных поверхностей

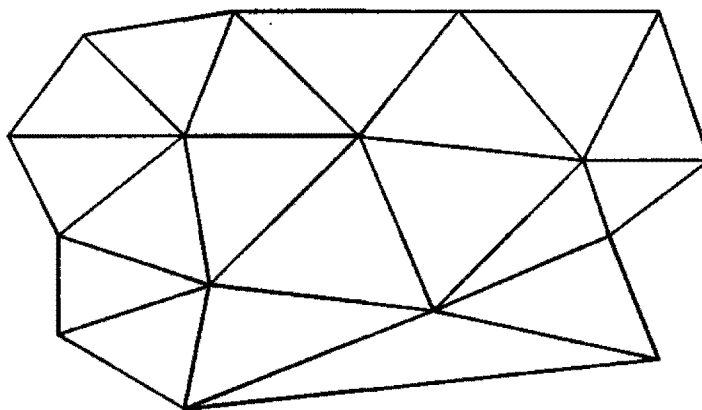


Рис. 3.13.8. Триангуляция в качестве области определения DMS поверхности

Безье. Поверхности, построенные этим способом, называют *DMS сплайнами* или *DMS поверхностями*. Рассмотрим построение DMS поверхности на базе симплексных сплайнов порядка n .

Областью определения DMS поверхности служит правильная триангуляция Ω некоторой двумерной области. Триангуляция считается правильной, если она состоит из невырожденных не перекрывающихся друг друга треугольников, стыкующихся по общим ребрам. Пример правильной триангуляции приведен на рис. 3.13.8.

Пусть вершины триангуляции описываются двумерными радиус-векторами p_i . С каждой вершиной некоторым образом свяжем $n + 1$ узлов: $p_{i,0}, p_{i,1}, \dots, p_{i,n}$. За узел $p_{i,0}$ примем вершину p_i : $p_{i,0} = p_i$, остальные узлы могут не совпадать с вершиной. Совокупность узлов $p_{i,0}, p_{i,1}, \dots, p_{i,n}$ называют *узловым облаком*. Узловые облака нам понадобятся в качестве узлов симплексных сплайнов.

Рассмотрим некоторый треугольник T^S из области определения DMS поверхности Ω (рис. 3.13.9). Пусть его вершинами являются точки p_a, p_b, p_c .

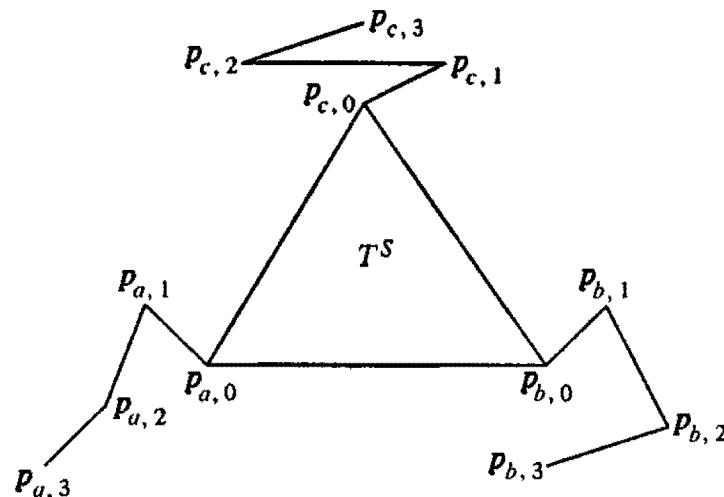


Рис. 3.13.9. Узловые облака вершин треугольника для $n = 3$

Совокупность всех узлов в вершинах рассматриваемого треугольника $p_{a,0}, p_{a,1}, \dots, p_{a,n}, p_{b,0}, p_{b,1}, \dots, p_{b,n}, p_{c,0}, p_{c,1}, \dots, p_{c,n}$ называют *сеткой узлов*. Для треугольника T^S мы можем получить $(n + 1)(n + 2)/2$ различных последовательностей узлов: $p_{a,0}, p_{a,1}, \dots, p_{a,i}, p_{b,0}, p_{b,1}, \dots, p_{b,j}, p_{c,0}, p_{c,1}, \dots, p_{c,k}$, где $i \geq 0, j \geq 0, k \geq 0$ и $i + j + k = n$. С каждым треугольником T^S свяжем $(n + 1)(n + 2)/2$ характеристических пространственных точек r^s_{ijk} , где $i \geq 0, j \geq 0, k \geq 0$ и $i + j + k = n$. Количество точек r^s_{ijk} равно количеству возможных последовательностей узлов для рассматриваемого треугольника. На характеристические точки никаких ограничений не накладывается. Их положение в пространстве определяет форму DMS поверхности. Пусть треугольники, образованные узлами $p_{a,i}, p_{b,j}, p_{c,k}$, $0 \leq i \leq n, 0 \leq j \leq n, 0 \leq k \leq n$, являются невырожденными, и при обходе их вершин мы будем двигаться против часовой стрелки.

На каждой последовательности узлов найдем выпуклую область $S_{ijk} = \{p_{a,0}, p_{a,1}, \dots, p_{a,i}, p_{b,0}, p_{b,1}, \dots, p_{b,j}, p_{c,0}, p_{c,1}, \dots, p_{c,k}\}$, построим симплексный сплайн $M^s_{ijk}(p)$. Вычислим *нормированный симплексный сплайн*

$$N^s_{ijk}(p) = |\Delta^s_{ijk}| M^s_{ijk}(p), \quad (3.13.5)$$

$$i \geq 0, \quad j \geq 0, \quad k \geq 0, \quad i + j + k = n,$$

где $|\Delta^s_{ijk}|$ — удвоенная площадь треугольника, образованного узлами $p_{a,i}, p_{b,j}, p_{c,k}$. С каждым нормированным симплексным сплайном $N^s_{ijk}(p)$ свяжем соответствующую характеристическую точку r^s_{ijk} . Радиус-вектор $r(p)$ DMS поверхности определяется как сумма по всем треугольникам области определения поверхности всех характеристических точек r^s_{ijk} , умноженных на значения

соответствующих нормированных симплексных сплайнов $N^s_{ijk}(\mathbf{p})$:

$$\begin{aligned} \mathbf{r}(\mathbf{p}) &= \sum_{\Omega} \left(\sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n N^s_{ijk}(\mathbf{p}) \mathbf{r}^s_{ijk} \right) = \\ &= \sum_{\Omega} \left(\sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n |\Delta^s_{ijk}| M^s_{ijk}(\mathbf{p}) \mathbf{r}^s_{ijk} \right), \quad (3.13.6) \\ \mathbf{p} &= [u \ v]^T \in \Omega. \end{aligned}$$

Внешнее суммирование выполняется по всем треугольникам области определения поверхности Ω . Внутреннее суммирование в формуле (3.13.6) выполняется по всем возможным последовательностям узлов треугольника T^s . Таких последовательностей $(n+1)(n+2)/2$, где n — порядок симплексных сплайнов.

В общем случае DMS поверхность является кусочно-полиномиальной функцией степени n . Она обладает большой гибкостью. Для одной и той же области определения Ω можно построить множество различных DMS поверхностей, варьируя узлами $\mathbf{p}_{i,1}, \mathbf{p}_{i,2}, \dots, \mathbf{p}_{i,n}$ в каждой вершине $\mathbf{p}_i = \mathbf{p}_{i,0}$ триангуляции и пространственным положением характеристических точек \mathbf{r}^s_{ijk} треугольников.

В работе [39] показано, что сумма всех отличных от нуля в заданной точке $\mathbf{p} = [u \ v]^T$ нормированных симплексных сплайнов (3.13.5) равна единице. То есть для правильной триангуляции выполняется равенство

$$\sum_{\Omega} \left(\sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n |\Delta^s_{ijk}| M^s_{ijk}(\mathbf{p}) \right) = 1. \quad (3.13.7)$$

Это свойство позволяет строить рациональные поверхности. Каждой характеристической точке \mathbf{r}^s_{ijk} треугольников припишем вес w^s_{ijk} . Мы можем построить рациональную DMS поверхность с радиус-вектором

$$\mathbf{r}(\mathbf{p}) = \frac{\sum_{\Omega} \left(\sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n |\Delta^s_{ijk}| M^s_{ijk}(\mathbf{p}) w^s_{ijk} \mathbf{r}^s_{ijk} \right)}{\sum_{\Omega} \left(\sum_{\substack{i,j,k=0, \\ i+j+k=n}}^n |\Delta^s_{ijk}| M^s_{ijk}(\mathbf{p}) w^s_{ijk} \right)}, \quad \mathbf{p} = [u \ v]^T \in \Omega. \quad (3.13.8)$$

Никаких ограничений на расположение узлов в узловых облаках мы не накладывали, за исключением того, что треугольники, образованные узлами $\mathbf{p}_{a,i}, \mathbf{p}_{b,j}, \mathbf{p}_{c,k}$, должны быть невырожденными и одинаково ориентированными. Если все узлы узловых облаков положить равными соответствующей вершине ($\mathbf{p}_{i,0} = \mathbf{p}_{i,1} = \dots = \mathbf{p}_{i,n} = \mathbf{p}_i$), то нормированные симплексные сплайны (3.13.5) будут равны соответствующим трехиндексным функциям Бернштейна, а каждая внутренняя сумма в (3.13.6) будет описывать отдельную треугольную поверхность Безье.

3.14. Поверхности, базирующиеся на поверхностях

На базе поверхности можно построить другую поверхность. Рассмотрим эквидистантную, ссылочную и продолженную поверхности. Поверхность, на основе которой строится новая поверхность, будем называть базовой поверхностью. Базовая поверхность будет лежать в структуре данных рассмотренных ниже поверхностей.

Эквидистантная поверхность. Для любой поверхности может быть построена эквидистантная поверхность. Каждая точка эквидистантной поверхности отстоит от соответствующей точки базовой поверхности на заданном расстоянии h по нормали к ней. Радиус-вектор эквидистантной поверхности описывается формулой

$$\mathbf{r}(u, v) = \mathbf{b}(u, v) + h\mathbf{m}(u, v), \quad u, v \in \Omega, \quad (3.14.1)$$

где $\mathbf{b}(u, v)$, $u, v \in \Omega$ — базовая поверхность, Ω — область определения параметров базовой поверхности, $\mathbf{m}(u, v)$ — нормаль к базовой поверхности.

Нормаль имеет единичную длину и направление вектора $\mathbf{b}_1 \times \mathbf{b}_2$. Область изменения параметров эквидистантной поверхности совпадает с областью

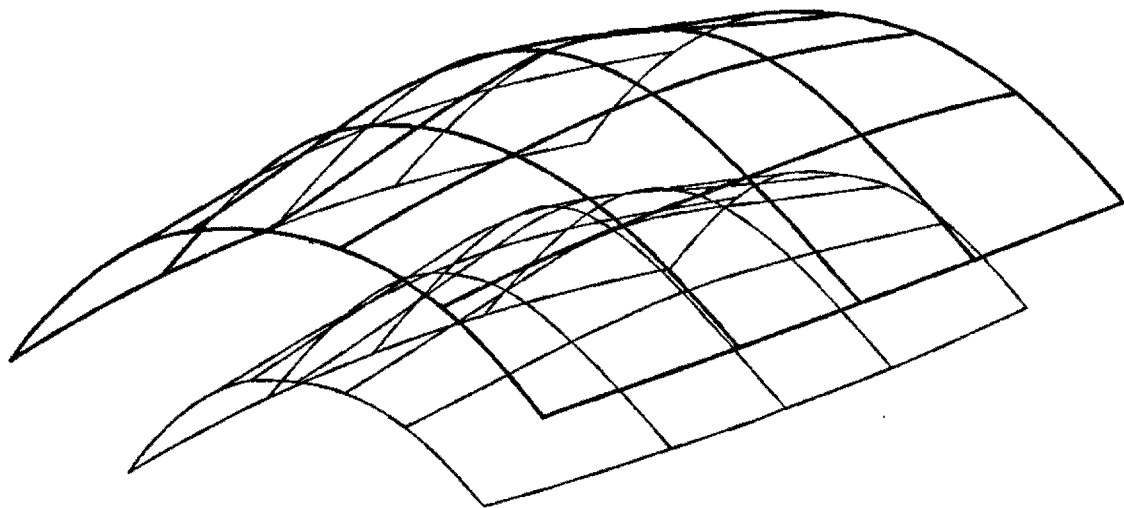


Рис. 3.14.1. Эквидистантная поверхность

изменения параметров базовой поверхности. Эквидистантная и базовая поверхности показаны на рис. 3.14.1. Производные $\mathbf{r}_1(u, v)$ и $\mathbf{r}_2(u, v)$ эквидистантной поверхности по параметрам u и v описываются векторами

$$\mathbf{r}_1(u, v) = \mathbf{b}_1(u, v) + h\mathbf{m}_1(u, v), \quad (3.14.2)$$

$$\mathbf{r}_2(u, v) = \mathbf{b}_2(u, v) + h\mathbf{m}_2(u, v), \quad (3.14.3)$$

где производные нормали $\mathbf{m}_1(u, v)$ и $\mathbf{m}_2(u, v)$ по параметрам определяются с помощью дериwационных формул Вейнгартена (1.7.26).

В общем случае можно построить эквидистантную поверхность, у которой расстояние $h = h(u, v)$ зависит от параметров поверхности.

Ссылочная поверхность. Ссылочная поверхность представляет собой поверхность, каждая точка которой получена путем некоторого преобразования соответствующей точки базовой поверхности. Ссылочная поверхность описывается радиус-вектором

$$\mathbf{r}(u, v) = \mathbf{M}_t \cdot \mathbf{b}(u, v), \quad u, v \in \Omega, \quad (3.14.4)$$

где $\mathbf{b}(u, v)$, $u, v \in \Omega$ — базовая поверхность, Ω — область определения параметров базовой поверхности, \mathbf{M}_t — расширенная матрица преобразования (1.4.5)

базовой поверхности. Область изменения параметра ссылочной поверхности совпадает с областью изменения параметра базовой поверхности. Ссылочная поверхность может быть использована при проективных преобразованиях.

Репараметризованная поверхность. Для любой поверхности может быть произведена линейная репараметризация. Пусть область определения параметров поверхности $\mathbf{r}(u, v)$ описывается неравенствами $u_{\min} \leq u \leq u_{\max}$, $v_{\min} \leq v \leq v_{\max}$. Пусть требуется, чтобы каждый из параметров изменялся бы в пределах от x до y . Будем считать, что u, v зависят от новых параметров p и q :

$$u(p) = u_{\min} \frac{y-p}{y-x} + u_{\max} \frac{p-x}{y-x}, \quad (3.14.5)$$

$$v(q) = v_{\min} \frac{y-q}{y-x} + v_{\max} \frac{q-x}{y-x}. \quad (3.14.6)$$

Тогда радиус-вектор поверхности будет описываться функцией

$$\mathbf{r}(p, q) = \mathbf{r}(u(p), v(q)), \quad x \leq p \leq y, \quad x \leq q \leq y. \quad (3.14.7)$$

После перехода к новым параметрам сама поверхность не изменится, но области определения ее параметров будет представлять собой квадрат со стороной $y-x$. Изменение области определения параметров можно считать одним из способов построения поверхности на базе другой поверхности.

Продолженная поверхность. Любая поверхность может быть продолжена на заданное параметрическое расстояние. Пусть требуется продолжить поверхность $\mathbf{b}(u, v)$, $u_1 \leq u \leq u_2$, $v_1 \leq v \leq v_2$ путем расширения области определения параметров до $a_u + u_1 \leq u \leq u_2 + b_u$, $a_v + v_1 \leq v \leq v_2 + b_v$. Если $a_u \leq 0$, $a_v \leq 0$, $b_u \geq 0$, $b_v \geq 0$, то поверхность будет продолжена за свои пределы. Если $a_u > 0$, $a_v > 0$, $b_u < 0$, $b_v < 0$, то поверхность будет усечена. Если поверхность замкнута по какому-либо из параметров, то при выходе замкнутого параметра за границу области определения выполним его циклический пересчет. Если поверхность $\mathbf{b}(u, v)$ замкнута по параметру u , тогда

$$\mathbf{r}(u, v) = \begin{cases} \mathbf{b}(u + u_{\Delta}, v), & \text{если } u < u_1, v_1 \leq v \leq v_2, \\ \mathbf{b}(u, v), & \text{если } u_1 \leq u \leq u_2, v_1 \leq v \leq v_2, \\ \mathbf{b}(u - u_{\Delta}, v), & \text{если } u > u_2, v_1 \leq v \leq v_2, \\ \mathbf{b}(u, v_1) + (v - v_1)\mathbf{b}_2(u, v_1), & \text{если } u_1 \leq u \leq u_2, v < v_1, \\ \mathbf{b}(u, v_2) + (v - v_2)\mathbf{b}_2(u, v_2), & \text{если } u_1 \leq u \leq u_2, v > v_2, \end{cases} \quad (3.14.8)$$

где $u_{\Delta} = u_2 - u_1$. Если поверхность $\mathbf{b}(u, v)$ замкнута по параметру v , тогда

$$\mathbf{r}(u, v) = \begin{cases} \mathbf{b}(u, v + v_{\Delta}), & \text{если } u_1 \leq u \leq u_2, v < v_1, \\ \mathbf{b}(u, v), & \text{если } u_1 \leq u \leq u_2, v_1 \leq v \leq v_2, \\ \mathbf{b}(u, v - v_{\Delta}), & \text{если } u_1 \leq u \leq u_2, v > v_2, \\ \mathbf{b}(u_1, v) + (u - u_1)\mathbf{b}_1(u_1, v), & \text{если } u < u_1, v_1 \leq v \leq v_2, \\ \mathbf{b}(u_2, v) + (u - u_2)\mathbf{b}_1(u_2, v), & \text{если } u > u_2, v_1 \leq v \leq v_2, \end{cases} \quad (3.14.9)$$

где $v_{\Delta} = v_2 - v_1$. Если же поверхность не является замкнутой, а один из параметров или оба вышли за границу области определения, то продолжим по-

верхность по касательной на ближайшей границе, и вычислим по продленной поверхности необходимые геометрические характеристики. Радиус-вектор продленной поверхности вычислим по формуле

$$r(u, v) = \begin{cases} b(u, v), & \text{если } u_1 \leq u \leq u_2, v_1 \leq v \leq v_2, \\ b(u_1, v) + (u - u_1)b_1(u_1, v), & \text{если } u < u_1, v_1 \leq v \leq v_2, \\ b(u_2, v) + (u - u_2)b_1(u_2, v), & \text{если } u > u_2, v_1 \leq v \leq v_2, \\ b(u, v_1) + (v - v_1)b_2(u, v_1), & \text{если } u_1 \leq u \leq u_2, v < v_1, \\ b(u, v_2) + (v - v_2)b_2(u, v_2), & \text{если } u_1 \leq u \leq u_2, v > v_2, \\ b(u_1, v_1) + (u - u_1)b_1(u_1, v_1) + (v - v_1)b_2(u_1, v_1) + \\ \quad + (u - u_1)(v - v_1)b_{12}(u_1, v_1), & \text{если } u < u_1, v < v_1, \\ b(u_2, v_1) + (u - u_2)b_1(u_2, v_1) + (v - v_1)b_2(u_2, v_1) + \\ \quad + (u - u_2)(v - v_1)b_{12}(u_2, v_1), & \text{если } u > u_2, v < v_1, \\ b(u_1, v_2) + (u - u_1)b_1(u_1, v_2) + (v - v_2)b_2(u_1, v_2) + \\ \quad + (u - u_1)(v - v_2)b_{12}(u_1, v_2), & \text{если } u < u_1, v > v_2, \\ b(u_2, v_2) + (u - u_2)b_1(u_2, v_2) + (v - v_2)b_2(u_2, v_2) + \\ \quad + (u - u_2)(v - v_2)b_{12}(u_2, v_2), & \text{если } u > u_2, v > v_2, \end{cases} \quad (3.14.10)$$

где $b_1(u, v)$ и $b_2(u, v)$ — частные производные радиус-вектора $b(u, v)$ по параметрам u и v , соответственно, $b_{12}(u, v)$ — смешанная производная радиус-вектора $b(u, v)$ по параметрам u и v . Дифференцируя формулы (3.14.8)–(3.14.10), получим производные радиус-вектора продленной поверхности. На рис. 3.14.2 приведена поверхность, на основе которой построена продленная поверхность. Продленная поверхность приведена на рис. 3.14.3.

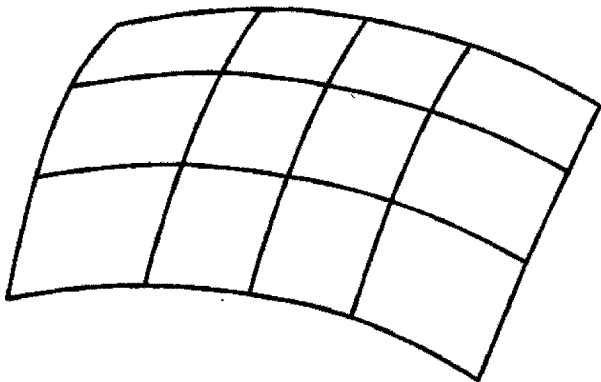


Рис. 3.14.2. Исходная поверхность

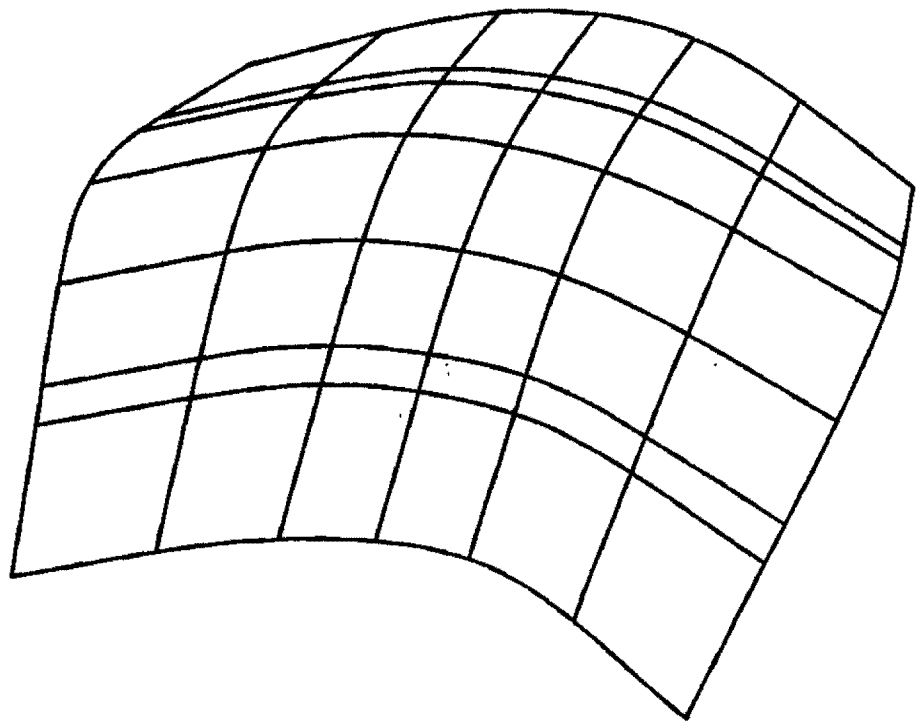


Рис. 3.14.3. Продленная поверхность

Формулы (3.14.8)–(3.14.10) могут быть использованы для вычисления радиус-вектора поверхности и его производных при выходе параметров поверхности за область их определения.

Общее правило. Все поверхности, в структуре данных которых лежит другая поверхность, не должны допускать многократного наследования своего же типа. Например, в качестве базовой поверхности для эквидистантной поверхности не должна быть использована другая эквидистантная поверхность, а должна быть использована базовая поверхность последней с соответствующим пересчетом длины эквидистанты. То есть, если нужно построить эквидистантную поверхность на расстоянии h_2 от другой эквидистантной поверхности $r_1(u, v) = \mathbf{b}(u, v) + h_1 \mathbf{m}(t)$, то в качестве базовой поверхности должна быть использована поверхность $\mathbf{b}(u, v)$, а длина эквидистанты должна быть равна $h = h_2 + h_1$. Если требуется построить ссылочную поверхность с матрицей преобразования \mathbf{M}_t на базе другой ссылочной поверхности $r_1(u, v) = \mathbf{N}_t \cdot \mathbf{b}(u, v)$, то в качестве базовой поверхности должна быть использована поверхность $\mathbf{b}(u, v)$, а матрицу преобразования получим как произведение матриц $\mathbf{M}_t \cdot \mathbf{N}_t$. Аналогичные правила должны действовать и для других базирующихся поверхностей.

3.15. Ограниченные контурами поверхности

Рассмотрим поверхности, которые выделяются среди всех остальных поверхностей и создаются «над ними» (на их основе).

Все приведенные выше поверхности имеют «природные» прямоугольные или треугольные области определения параметров. Прямоугольные или треугольные области удобны в использовании из-за их простоты. Прямоугольная область параметров определяется неравенствами $u_{\min} \leq u \leq u_{\max}$, $v_{\min} \leq v \leq v_{\max}$, и для любой пары параметров можно сразу сказать, лежит ли она вне области, внутри области или на границе области. Рассмотренные выше поверхности дают широкие возможности для моделирования. Неограниченные возможности для моделирования мы получим, если границы поверхностей смогут иметь произвольный вид. Это можно сделать с помощью двумерных кривых линий $r_i(t_i) = [u(t_i) \ v(t_i)]^T$, $t_{i \min} \leq t_i \leq t_{i \max}$, $i = 1, 2, \dots, n$, на поверхности, которые описываются в пространстве параметров u и v поверхности. Эти двумерные кривые собраны в контуры и разделяют используемую часть поверхности от остальных частей.

Такая поверхность с произвольной границей представляет собой наиболее общий тип поверхности. Будем называть ее *ограниченной контурами поверхностью*. Ограниченную контурами поверхность можно построить на основе любой поверхности с прямоугольной или треугольной областью путем изменения области определения ее параметров. Именно эта поверхность лежит в основе моделирования тел. Ее параметрам разрешается принимать значения только внутри области, ограниченной заданными на ней контурами. Границы поверхности представляют собой отображение в трехмерное пространство границ двумерной связной области.

Исходную поверхность с прямоугольной или треугольной областью определения параметров будем называть *базовой* поверхностью. Пусть $\mathbf{b}(u, v)$ есть радиус-вектор ее точек. Радиус-вектор ограниченной контурами поверхности описывается тем же выражением, что и базовая поверхность, но имеет другую область определения параметров

$$\mathbf{r}(u, v) = \mathbf{b}(u, v), \quad u, v \in \Omega, \quad (3.15.1)$$

где Ω — область определения параметров, которая представляет собой ограниченную замкнутыми кривыми часть плоскости, из каждой точки которой можно

попасть в любую другую ее часть, не пересекая границ. Область определения параметров определяется двумерными контурами (замкнутыми составными кривыми). Двухмерный контур $c(t) = [u(t) \ v(t)]^T$, $t_{\min} \leq t \leq t_{\max}$, состоит из нескольких двумерных кривых, стыкующихся между собой. Радиус-вектор контура определяется функцией (2.12.7). Таких контуров может быть несколько. Совокупность контуров ограничивает двумерную область параметров и придает границам поверхности заданную форму.

Область определения параметров ограниченной контурами поверхности представляет собой связную часть двумерного пространства и описывается набором непересекающихся двумерных контуров, один из которых является внешним, а остальные лежат внутри него и являются внутренними. Внутренние контуры не могут быть вложены друг в друга. Для того чтобы легче было определить, принадлежит ли некоторая двумерная точка $p = [u \ v]^T$ области определения параметров поверхности, контуры ориентируют определенным образом. Пусть внешний контур ориентирован так, что обход по нему в положительном направлении осуществлялся против часовой стрелки, если смотреть на поверхность навстречу ее нормали. Внутренние контуры должны быть ориентированы в противоположном направлении.

Принадлежность точки области определения параметров. Пусть требуется ответить на вопрос, принадлежит ли двумерная точка $p = [u \ v]^T$ пространства параметров области определения поверхности — классифицировать точку относительно области. Для областей прямоугольной формы ответить на него просто — для этого нужно проверить, удовлетворяют ли параметры u и v неравенствам $u_{\min} \leq u \leq u_{\max}$, $v_{\min} \leq v \leq v_{\max}$. Для ограниченной контурами поверхности, это сделать несколько сложнее.

Пусть требуется классифицировать точку $p = [u \ v]^T$ в параметрической области поверхности, ограниченной тремя контурами: одним внешним и двумя внутренними, как показано на рис. 3.15.1.

В общем случае для того чтобы установить принадлежность точки p некоторой двумерной области, можно найти все точки пересечения любой двумерной прямой линии (проще горизонтальной или вертикальной), проходящей

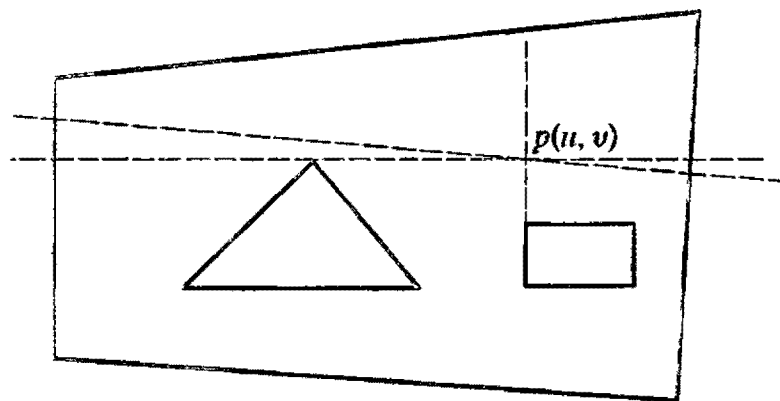


Рис. 3.15.1. Определение положения точки относительно области

через эту точку, с границами области. Таких точек пересечения должно быть четное число — каждой точке входа прямой в область должна соответствовать точка выхода из области. Но в некоторых случаях, как показано на рис. 3.15.1, их может быть нечетное число из-за касания или прохождения прямой через угол граничного контура. Если точек пересечения получено нечетное число, то

можно пересечь границу области другой двухмерной прямой, проходящей через проверяемую точку и составляющую небольшой угол с уже построенной прямой. Далее легко определить принадлежность рассматриваемой точки заданной области. Если точка p находится между точкой входа прямой в область и соответствующей ей точкой выхода из области, то рассматриваемая точка принадлежит данной области. В противном случае или в случае отсутствия точек пересечения прямой с границей области рассматриваемая точка лежит вне данной области.

Если известно, что внешний граничный контур области параметров поверхности ориентирован против часовой стрелки, а внутренние контуры ориентированы по часовой стрелке, если смотреть на поверхность навстречу ее нормали, то классифицировать точку $p = [u \ v]^T$ в параметрической области поверхности можно следующим образом. Найдем ближайшую к точке p точку p_0 на границе области. Построим двухмерный вектор w из найденной точки p_0 в точку p .

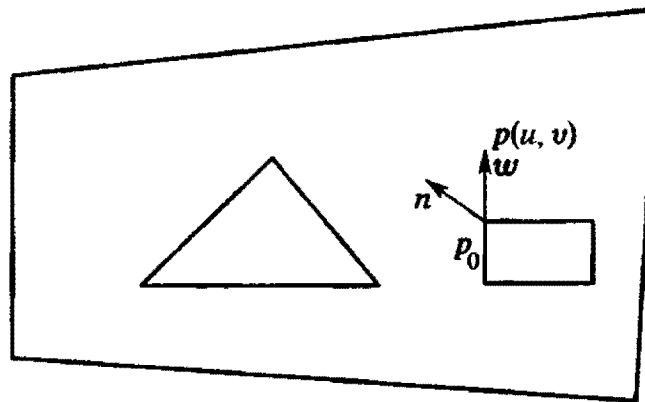


Рис. 3.15.2. Определение положения точки относительно области определения параметров поверхности.

Вычислим нормаль n к граничному контуру в точке p_0 . Нормаль n в правой системе координат всегда направлена влево, если смотреть вдоль направления контура. Если точка p_0 является точкой излома граничного контура, то в качестве нормали возьмем среднюю нормаль к границе в этой точке. Если $n \cdot w \geq 0$, то точка p принадлежит области определения поверхности. В противном случае — нет. Это отражено на рис. 3.15.2.

Примеры. На рис. 3.15.3 приведена NURBS поверхность, имеющая прямоугольную область определения параметров, и линии на ней. На рис. 3.15.4

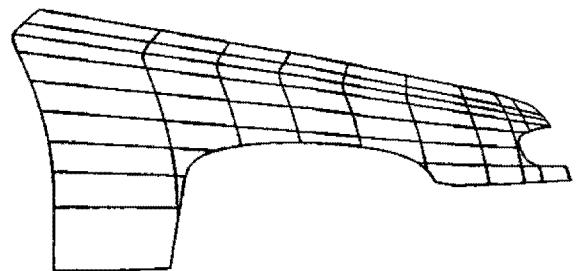
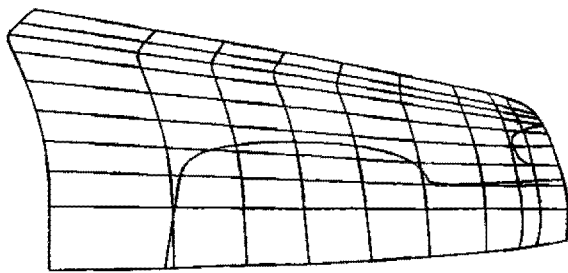


Рис. 3.15.3. NURBS поверхность и линии на ней

Рис. 3.15.4. Ограниченная контурами поверхность

приведена NURBS поверхность, ограниченная кривыми на ней. На рис. 3.15.5 приведена ограниченная контурами поверхность вращения.

При создании тонкостенного тела требуется строить ограниченные контурами поверхности, базирующиеся на продолженной поверхности, которая является эквидистантной поверхностью. Для таких поверхностей в качестве базовой

удобно ввести новый тип поверхности, который включает в себя одновременно свойства эквидистантной и продолженной поверхностей.

В общем случае область определения параметров Ω ограниченной контурами поверхности $\mathbf{r}(u, v)$ может выходить за область определения параметров базовой

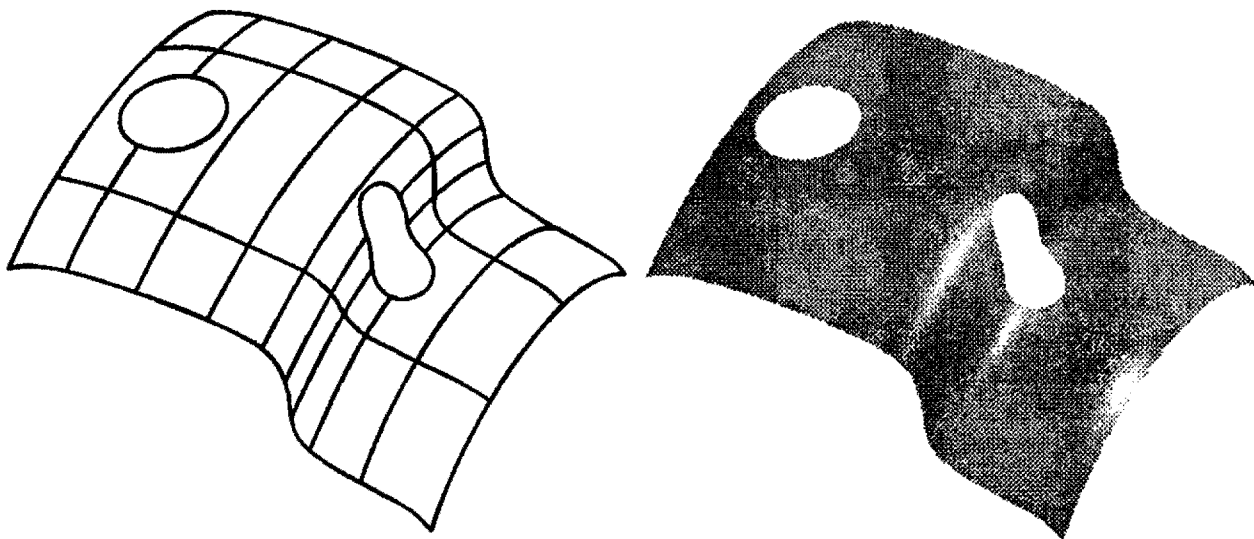


Рис. 3.15.5. Ограниченная контурами поверхность

вой поверхности $\mathbf{b}(u, v)$ при условии, что определено поведение поверхности $\mathbf{b}(u, v)$ за границами ее области определения. Это можно сделать, например, с помощью формул (3.14.8)–(3.14.10).

Общее правило. Ограниченные контурами поверхности не должны допускать многократного наследования своего же типа. В качестве базовой поверхности для поверхности, ограниченной контурами, не должна использоваться другая ограниченная контурами поверхность, а должна быть взята базовая поверхность последней. Ограниченные контурами поверхности должны создаваться в последнюю очередь. Например, если требуется построить эквидистантную поверхность на базе ограниченной контурами поверхности (3.15.1), то сначала должна быть построена эквидистантная поверхность на основе базовой поверхности $\mathbf{b}(u, v)$, а потом на базе эквидистантной должна быть создана ограниченная контурами поверхность.

3.16. Способы построения поверхностей

На рис. 3.16.1 приведены способы построения поверхностей. Здесь отражены лишь наиболее распространенные способы. В авиастроении широко используются различные поверхности сопряжения. Кроме того, можно построить поверхности с характерными особенностями — фичерные поверхности (*feature surface*). Способы их построения могут быть различными и зависят от специфики области применения.

Рисунок 3.16.1 отражает процесс построения поверхностей при чтении его слева направо. Изначально поверхности могут быть построены или по местной системе координат и скалярным параметрам, или по характерным точкам, или на основе линий. Далее эти поверхности могут быть продлены или усечены, на их основе могут быть получены эквидистантные поверхности. Все упомянутые поверхности имеют прямоугольную или треугольную область определения параметров. На базе поверхностей с прямоугольной или треугольной областью определения параметров можно построить ограниченные контурами поверхности произвольной формы.

Все поверхности являются равноправными по отношению к выполняемым ими функциям и выполняемым над ними операциям. Кроме общих функций произвольного геометрического объекта поверхности должны выполнять функции вычисления радиус-вектора поверхности, частных производных радиус-вектора по параметрам, нормали поверхности и ее частных производных,

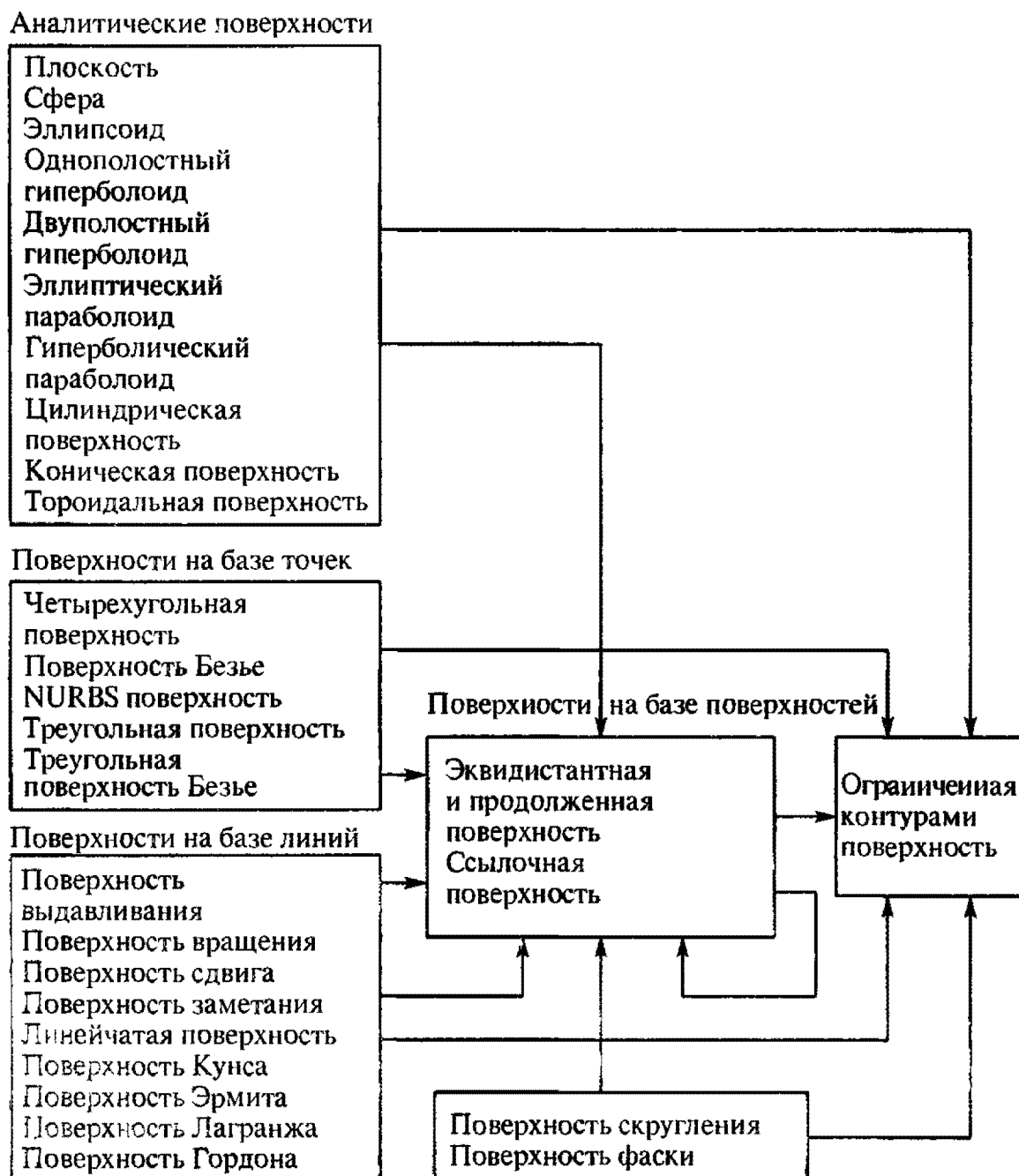


Рис. 3.16.1. Способы построения поверхностей

коэффициентов первой и второй квадратичных форм поверхности, функции предоставления информации об области определения параметров поверхности (замкнутости, граничных контурах), функции определения принадлежности заданных параметров области определения поверхности, функции изменения положения и ориентации в пространстве, функции модификации формы поверхности.

Математическая модель поверхности должна быть дополнена функциями, обеспечивающими взаимодействие поверхности с другими геометрическими объектами и выполнение над ними различных операций. В следующей главе мы рассмотрим операции над кривыми и поверхностями.

Глава 4

ОПЕРАЦИИ НАД КРИВЫМИ И ПОВЕРХНОСТЯМИ

4.1. Выполнение операций

В предыдущих главах были рассмотрены методы построения различных кривых линий и поверхностей. Этих методов не достаточно, чтобы моделировать сложные геометрические объекты, например, поверхности, ограниченные контурами. В большинстве случаев сложные геометрические объекты появляются как результат некоторой операции над более простыми объектами. В данной главе мы рассмотрим методы выполнения операций. *Операцией* будем называть совокупность действий над одним или несколькими исходными объектами, которая приводит к рождению нового геометрического объекта. Действия, которые изменяют объект, не изменяя его природы, будем называть модификацией или редактированием.

Редактирование объекта сводится к изменению значений его данных при неизменной их структуре. Все геометрические объекты в итоге описываются скалярными величинами и векторами. Редактирование объектов приводит к изменению этих скаляров и компонент векторов. Редактированием можно масштабировать, зеркально отразить, переместить, повернуть в пространстве геометрический объект или изменить его форму. Преобразования трансформации по матрице, перемещение, поворот в пространстве геометрического объекта сводятся к соответствующим преобразованиям радиус-векторов, лежащих в структуре данных объекта.

К операциям мы будем относить построение проекций точек на кривые и поверхности по нормали к ним, построение проекций точек на поверхности по заданному направлению, построение точек пересечения кривых, построение точек пересечения кривых и поверхностей, построение линий пересечения поверхностей, построение поверхностей скругления и поверхностей фасок, а также решение других задач. Операция пересечения двумерных кривых и операция пересечения поверхностей являются основополагающими, так как они присутствуют в большинстве других операций, выполняемых над геометрическими объектами.

Используемые нами математические модели линий и поверхностей построены так, что они выдают геометрическую информацию, если мы обратимся к ним с некоторыми параметрами. Если геометрический объект рассматривать как «черный ящик», то параметры подаются на его вход, а на выходе мы получим геометрическую информацию, например, радиус-вектор точки объекта, соответствующий введенным параметрам. В обратном направлении подобный «черный ящик» сигнал не пропускает. То есть, если мы возьмем некоторую точку в пространстве и захотим узнать, далеко ли она находится от данного геометрического объекта или какие параметры объекта ей соответствуют, то построенная нами математическая модель окажется в затруднительном положении. Подобные задачи возникают часто и математическая модель должна с

ними справляться. Математика располагает методами решения таких задач, но эти методы являются итерационными и для того, чтобы получить искомое решение, необходимо указать некоторое *нулевое* (начальное) *приближение* искомого решения.

В качестве примера возьмем задачу определения точки пересечения поверхности $s(u, v)$ и кривой линии $c(t)$. В точке их пересечения радиус-векторы кривой и поверхности имеют одинаковые значения: $s(u, v) = c(t)$. Это векторное равенство представляет собой систему трех скалярных уравнений, искомыми величинами которых являются параметры u , v и t . Решив эту систему уравнений, получим параметры u , v и t , по которым при необходимости можно вычислить точку на поверхности или на кривой. Система уравнений в общем случае является нелинейной и решается итерационными методами. Точек пересечения может быть несколько и от того, с какого приближения параметров начнется итерационный процесс решения, зависит, к каким параметрам сойдется решение. Из данного примера видно, что для решения задачи нужно уметь делать две вещи: находить нулевое приближение для каждого решения и решать системы нелинейных уравнений. Что касается решения системы уравнений, то существует много методов, которые приводят к искомым решениям, если корректно задано его нулевое приближение. Мы рассмотрим эти методы чуть позже.

Для нахождения нулевого приближения каждого решения используем метод проб и ошибок. Нулевое приближение может быть найдено довольно грубо, важно только, чтобы оно лежало в области сходимости решения системы уравнений к одному из точных решений. Важно также, чтобы каждому решению было найдено свое нулевое приближение. Для поиска нулевых приближений будем проверять некоторое событие в выбранных точках геометрических объектов. В упомянутом примере пересечения кривой и поверхности возьмем некоторые значения $u^{(0)}$, $v^{(0)}$ и $t^{(0)}$ параметров поверхности и кривой и заменим поверхность частью касательной к ней плоскости в выбранной точке, а кривую заменим частью касательной прямой линии в выбранной точке. Найдем точку пересечения касательной прямой и касательной плоскости и оценим, далеко ли от точек касания находится точка пересечения. Если точка пересечения находится недалеко от точек, в которых построены касательная прямая и касательная плоскость, то параметры $u^{(0)}$, $v^{(0)}$ и $t^{(0)}$ можно принять в качестве начального приближения для дальнейшего определения параметров точки пересечения. Если же точка пересечения находится далеко от точек, в которых построены касательная прямая и касательная плоскость, то возьмем другие точки на кривой и поверхности и выполним в них те же действия.

В данной схеме проверок остается неясным, какие точки можно считать близкими и какие точки следует выбрать для проверки следующими. Для кривых и поверхностей эти величины удобнее измерять в параметрических единицах. Расстояние близости и длина перемещения для следующей проверки в соседнюю точку являются связанными величинами. Действительно, если близким считать некоторое параметрическое расстояние Δ , то следующий поиск следует выполнять на параметрическом расстоянии Δ от данной точки. Длину параметрического шага будем связывать с искривленностью геометрического объекта. Приращения параметров Δ будем называть *параметрическими шагами*. Шаги перемещения по параметрической области кривой через кривизну связаны с допустимым углом отклонения касательной к ней. Шаги перемеще-

ния по параметрической области поверхности через соответствующую кривизну связаны с допустимым углом отклонения нормали поверхности. Максимально допустимую величину угла отклонения касательной или нормали обозначим $\Delta\alpha$.

Шагать по параметрической области нужно достаточно часто, чтобы не пропустить решение, и в то же время достаточно редко, чтобы потратить меньше времени. Это регулируется выбором угла $\Delta\alpha$, при этом шаг по параметру не зависит от абсолютных размеров геометрического объекта. Подобная схема выбора нулевых приближений применима при выполнении всех операций над кривыми и поверхностями.

Из предложенной схемы видно, что для выполнения операций нужно уметь корректно перемещаться по параметрическим областям кривых и поверхностей в поиске нулевых приближений решения, уметь находить нулевые приближения и уметь находить точное решение, отталкиваясь от некоторого приближения. Далее мы рассмотрим вычисление шага по параметрам для кривых и поверхностей, а также методы решения систем уравнений при известных нулевых приближениях решений и уравнения, описывающие ту или иную операцию.

4.2. Движение по параметрической области

Для определения нулевого приближения решения системы уравнений будем двигаться с некоторым шагом по параметрической области и в дискретных точках проверять выполнение некоторого условия, характерного для данной операции. Эти дискретные точки должны быть выбраны таким образом, чтобы было найдено нулевое приближение для каждого решения. Характерным условием для операции, как правило, является смена знака невязки уравнений рассматриваемой операции или уменьшение невязки уравнений ниже известного уровня.

Движение по кривой. Движение по параметрической области некоторой кривой $c(t)$ будем начинать с минимального значения ее параметра t_{\min} . Двигаться будем в направлении возрастания параметра до достижения его максимального значения t_{\max} с некоторым вычисляемым шагом Δt по параметру. Шаг Δt мы будем вычислять, исходя из условия, что при переходе от точки к точке касательная кривой отклонялась бы на угол, не превышающий некоторую величину $\Delta\alpha$. Приращение параметра или параметрический шаг Δt меняется от точки к точке. В каждой точке будет проверяться выполнение некоторого условия, по которому должно быть принято решение о дальнейших действиях.

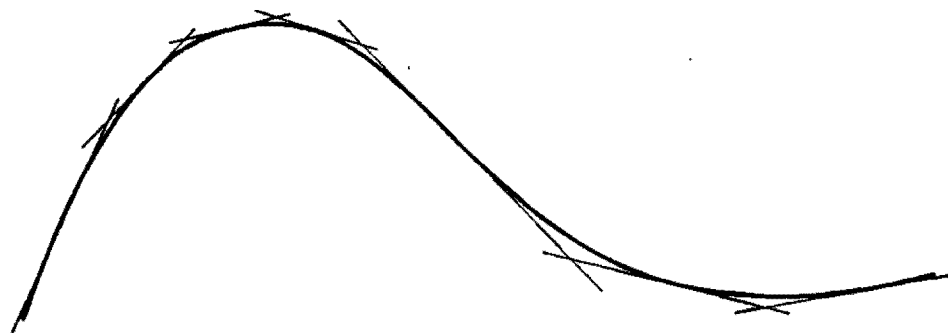


Рис. 4.2.1. Заменяющие отрезки касательных кривой при поиске начальных приближений параметра

В некоторых операциях условие будет проверяться на отрезке касательной прямой, построенной в данной точке. Фактически будет выполнена замена кривой отрезками прямых линий, касательных к кривой. На рис. 4.2.1 показана кривая

и заменяющие ее отрезки на начальной стадии операции. Направление отрезков совпадает с касательными кривой в точках касания. Касательные строятся в точках вычисления очередного шага. Дины отрезков и точки, в которых вычисляются касательные, зависят от кривизны кривой и определяются следующим образом.

Пусть в процессе движения мы находимся в некоторой точке кривой, описываемой параметром $t = t_0$. Следующее значение параметра $t = t_0 + \Delta t$ должно быть таким, чтобы касательный вектор в новой точке отклонился бы от касательного вектора кривой в предыдущей точке на угол, не превышающий $\Delta\alpha$. Значение предельного угла отклонения будем считать заданным. Для вычисления нулевых приближений параметров при выполнении операций можно принять $\pi/10 \leq \Delta\alpha \leq \pi/20$. Таким образом, нам нужно предвидеть искривление линии в некоторой окрестности точки $t = t_0$ и в соответствии с этим выбрать следующую точку. Если предположить, что кривизна кривой сохраняется постоянной в окрестности точки $t = t_0$, то кривая отклонится на угол $\Delta\alpha$, если точка сместится по дуге на расстояние

$$\Delta s = \frac{\Delta\alpha}{k}, \quad (4.2.1)$$

где k — кривизна кривой в данной точке. Если еще предположить, что и длина производной радиус-вектора кривой сохранится постоянной в окрестности рассматриваемой точки, то смещению Δs по дуге будет соответствовать приращение параметра

$$\Delta t = \frac{\Delta s}{|c'|} = \frac{\Delta\alpha}{k|c'|}, \quad (4.2.2)$$

где c' — производная радиус-вектора кривой по параметру. Подставив в последнее равенство выражение (1.5.28) для кривизны кривой, получим шаг по параметру

$$\Delta t = \frac{\Delta\alpha |c'|^2}{|c' \times c''|}. \quad (4.2.3)$$

Описанным способом пройдем по всей линии и проверим в каждой точке выполнение некоторого условия, по которому можно судить о близости того или иного события. В каждой точке остановки мы построим отрезок прямой линии длиной $|c'| \Delta t$, касающийся своей серединой кривой в рассматриваемой точке.

Все рассуждения базируются на том, что рассматриваемую окрестность кривой мы аппроксимируем дугой окружности конечного радиуса. При стремлении кривизны в рассматриваемой точке к нулю длина параметрического шага стремится к бесконечности. Кривизна кривой в рассматриваемой точке может быть очень малой или вовсе равной нулю, хотя в достаточно близкой к рассматриваемой точке области кривизна может иметь гораздо большее значение. В таком случае мы шагнем слишком далеко. Для учета подобных ситуаций можно ввести ограничение максимального параметрического шага или сделать небольшой шаг по параметру и повторить вычисления в новой точке. Если в рассматриваемой точке длина первой производной радиус-вектора стремится к нулю, то длина параметрического шага тоже стремится к нулю. Длина первой производной радиус-вектора кривой в рассматриваемой точке может быть очень малой или вовсе равной нулю, хотя в достаточно близкой к рассматриваемой

мой точке области длина производной может иметь гораздо большее значение. В таком случае приращение параметра будет неоправданно малым или вовсе равным нулю. Для учета подобных ситуаций введем ограничение минимального параметрического шага. Для разных кривых это ограничение может быть различным.

Параметрические шаги допускается определять приближенно, так как мы ничего не вычисляем окончательно, а лишь находим нулевые приближения для дальнейшего уточнения. Для некоторых кривых нет необходимости прибегать к вычислениям для определения шага Δt . Например, для окружности или ее дуги $\Delta t = \Delta\alpha$, если параметром является центральный угол, для отрезка прямой можно положить $\Delta t = t_{\max} - t_{\min}$.

Движение по поверхности. Движение по параметрической области некоторой поверхности $s(u, v)$ будем выполнять аналогично движению по кривой. Положим один из параметров постоянным, и будем двигаться вдоль другого параметра описанным выше способом. При движении будем заодно вычислять и шаг по параметру, который в данный момент считаем фиксированным с таким расчетом, чтобы в конце вычислить среднее значение шага и по неподвижному параметру. Для вычисления шага по параметрам поверхности в правую часть (4.2.2) подставим кривизны (1.8.5) вдоль координатных линий поверхности. При $v = \text{const}$ и $u = \text{const}$ кривизны линий соответственно равны $\mu_u = b_{11}/g_{11}$ и $\mu_v = b_{22}/g_{22}$, где g_{11} и g_{22} — коэффициенты первой основной квадратичной формы поверхности (1.7.8), b_{11} и b_{22} — коэффициенты второй основной квадратичной формы поверхности (1.7.21). В результате шаг по параметрам поверхности вычислим по формулам

$$\Delta u = \frac{\Delta\alpha\sqrt{g_{11}}}{b_{11}} = \frac{\Delta\alpha|s_1|}{\mathbf{m} \cdot \mathbf{s}_{11}}, \quad (4.2.4)$$

$$\Delta v = \frac{\Delta\alpha\sqrt{g_{22}}}{b_{22}} = \frac{\Delta\alpha|s_2|}{\mathbf{m} \cdot \mathbf{s}_{22}}, \quad (4.2.5)$$

где \mathbf{m} — нормаль поверхности, \mathbf{s}_1 — производная радиус-вектора поверхности по параметру u (по первому параметру), \mathbf{s}_2 — производная радиус-вектора поверхности по параметру v (по второму параметру).

Начинать движение по параметрической области поверхности будем с минимальных значений параметров u_{\min} и v_{\min} . Двигаться будем в направлении возрастания одного из параметров до его максимального значения, одновременно вычисляя среднее значение шага вдоль другого параметра. Затем сделаем усредненный шаг по другому параметру поверхности и повторим проход по первому параметру от его минимального значения до максимального. Так рядами пройдем всю параметрическую область поверхности. Если поверхность построена на базе кривых линий, то для вычисления шага по параметру поверхности можно привлекать соответствующий шаг по параметру кривой. Для аналитических поверхностей нет необходимости прибегать к вычислениям по формулам (4.2.4) и (4.2.5), так как шаг можно определить по аналитическим формулам поверхности.

Так же как для кривых, все рассуждения для поверхностей базируются на том, что рассматриваемая окрестность точки поверхности сохраняет постоянной нормальную кривизну вдоль параметрических направлений. Если же одна из

нормальных кривизн в рассматриваемой точке стремится к нулю, то длина соответствующего параметрического шага стремится к бесконечности. Нормальная кривизна поверхности в рассматриваемой точке может быть очень малой или вовсе равной нулю, хотя в достаточно близкой к рассматриваемой точке области кривизна может быть большой. Для того чтобы не перешагнуть через резкое изменение формы поверхности, можно ввести ограничение максимального шага по каждому из параметров или сделать небольшой шаг по соответствующему параметру и повторить вычисления в новой точке. Если же в рассматриваемой точке длина частной производной радиус-вектора стремится к нулю, то длина соответствующего параметрического шага тоже стремится к нулю. Чтобы параметрический шаг не мог стать как угодно малым, можно ввести ограничение минимального параметрического шага по параметрам поверхности. Для разных поверхностей это ограничение может быть различным.

Используя описанную схему движения по параметрической области, можно пройти по отдельным точкам поверхности и проверить в каждой точке выполнение некоторого условия, по которому можно судить о наличии решения вблизи точки. Можно получить разбиение параметрической области поверхности на прямоугольники со сторонами Δu и Δv внутри каждого такого прямоугольника поверхность аппроксимировать четырехугольной частью плоскости (4.1.2), которая в средней точке касается рассматриваемой поверхности. Другими словами, при необходимости можно покрыть поверхность плоскими четырехугольниками (рис. 4.2.2), со сторонами Δu и Δv (при этом они могут частично перекрывать друг друга и частично выходить за область определения параметров) и проверить на каждом четырехугольнике выполнение некоторого условия.

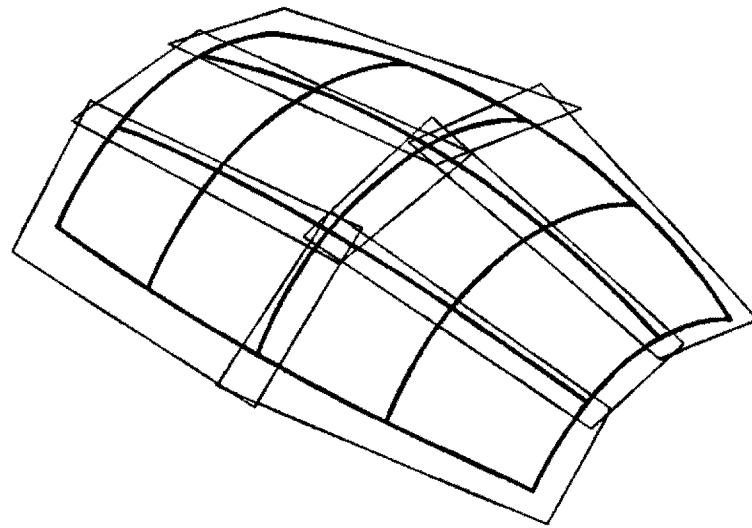


Рис. 4.2.2. Заменяющие поверхность касательные четырехугольники

Результаты этой проверки будут меняться от четырехугольника к четырехугольнику и примерно будут равны результатам аналогичной проверки этого же условия для самой поверхности. По результатам проверки некоторого условия можно определить нулевое приближение каждого решения.

Описанный способ движения по кривой или поверхности не зависит от размеров последних. Он преследует цель отследить близость некоторого события для геометрического объекта и используется в операциях над геометрическими объектами. Для других целей (например, визуализации объекта) используются другие способы вычисления параметрического шага и перемещения по параметрической области.

являющейся линейной системой относительно погрешностей Δx_i , $i = 1, 2, \dots, n$. Решив систему линейных алгебраических уравнений (4.3.3), вычислим следующее приближение для искомого решения:

$$x_i^{(r+1)} = x_i^{(r)} + \Delta x_i, \quad i = 1, 2, \dots, n. \quad (4.3.4)$$

Каждое приближение будет отличаться от точного решения, так как мы линеаризовали систему уравнений. Погрешности Δx_i на каждом приближении будут становиться все меньше и меньше. Мы не будем останавливаться на условиях и скорости сходимости метода Ньютона–Рафсона. Скажем только, что если начальное приближение выбрано достаточно хорошо и матрица системы линейных уравнений (4.3.3) на каждой итерации хорошо обусловлена и имеет обратную матрицу, то метод Ньютона–Рафсона сходится к единственному в данной окрестности решению

$$\lim_{r \rightarrow \infty} x_i^{(r)} = x_i^*, \quad i = 1, 2, \dots, n.$$

Метод Ньютона–Рафсона имеет квадратичную сходимость

$$|x_i^{(r+1)} - x_i^*| \leq M |x_i^{(r)} - x_i^*|^2, \quad i = 1, 2, \dots, n, \quad (4.3.5)$$

где M определяется матрицей системы (4.3.3) и удовлетворяет неравенству

$$M |x_i^{(r)} - x_i^*| < 1.$$

Это сравнительно быстрая сходимость — после очередной итерации погрешность каждой неизвестной уменьшается примерно на один или два порядка. Итерационный процесс решения системы нелинейных уравнений (4.3.1) заканчивается, когда на очередной итерации погрешности всех неизвестных становятся меньше заданной величины ε :

$$|\Delta x_i| < \varepsilon \quad \text{для всех } i = 1, 2, \dots, n.$$

Запишем систему линейных уравнений (4.3.3) в матричном виде

$$\mathbf{F}'(\mathbf{x}^{(r)}) \cdot (\mathbf{x}^{(r+1)} - \mathbf{x}^{(r)}) = -\mathbf{f}(\mathbf{x}^{(r)}), \quad (4.3.6)$$

где

$$\mathbf{F}'(\mathbf{x}^{(r)}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}, \quad \mathbf{x}^{(r)} = \begin{bmatrix} x_1^{(r)} \\ x_2^{(r)} \\ \dots \\ x_n^{(r)} \end{bmatrix}, \quad \mathbf{f}(\mathbf{x}^{(r)}) = \begin{bmatrix} f_1(\mathbf{x}^{(r)}) \\ f_2(\mathbf{x}^{(r)}) \\ \dots \\ f_n(\mathbf{x}^{(r)}) \end{bmatrix}.$$

$\mathbf{F}'(\mathbf{x}^{(r)})$ — матрица Якоби, которая одновременно является матрицей системы линейных уравнений (4.3.3), $\mathbf{f}(\mathbf{x}^{(r)}) = [f_1(\mathbf{x}^{(r)}) \ f_2(\mathbf{x}^{(r)}) \ \dots \ f_n(\mathbf{x}^{(r)})]^T$ — матрица-столбец значений функций (4.3.1), $\mathbf{x}^{(r)} = [x_1^{(r)} \ x_2^{(r)} \ \dots \ x_n^{(r)}]^T$ — матрица-столбец неизвестных на r -м приближении. Запись системы (4.3.3) в матричной форме (4.3.6) похожа на запись

$$\frac{\partial \mathbf{f}(\mathbf{x}^{(r)})}{\partial \mathbf{x}} (\mathbf{x}^{(r+1)} - \mathbf{x}^{(r)}) = -\mathbf{f}(\mathbf{x}^{(r)}).$$

метода Ньютона для одного нелинейного уравнения. Рекуррентное соотношение для одного нелинейного алгебраического уравнения $f(x) = 0$ имеет вид

$$x^{(r+1)} = x^{(r)} - \frac{f(x^{(r)})}{\partial f(x^{(r)})/\partial x}.$$

Рекуррентное соотношение (4.3.4) для системы уравнений (4.3.3) в матричной форме имеет вид

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} - (\mathbf{F}'(\mathbf{x}^{(r)}))^{-1} \mathbf{f}(\mathbf{x}^{(r)}).$$

Модифицированный метод Ньютона. В некоторых случаях для решения системы (4.3.1) применяется *модифицированный метод Ньютона*, который имеет вид

$$\mathbf{F}'(\mathbf{x}^{(0)})(\mathbf{x}^{(r+1)} - \mathbf{x}^{(r)}) = -\mathbf{f}(\mathbf{x}^{(r)}). \quad (4.3.7)$$

Модифицированный метод Ньютона отличается тем, что при решении системы линейных уравнений матрица \mathbf{F}' не вычисляется на каждой итерации заново, а используется матрица, вычисленная на первой итерации $\mathbf{F}'(\mathbf{x}^{(0)})$. Модифицированный метод Ньютона обладает линейной сходимостью, поэтому для получения той же точности он требует больше итераций, чем метод Ньютона. Но общие вычислительные затраты модифицированного метода Ньютона за счет расчета по готовой матрице могут оказаться меньшими, чем у метода Ньютона. Модифицированный метод Ньютона является представителем стационарных методов решения нелинейных систем уравнений в силу того, что матрица $\mathbf{F}'(\mathbf{x}^{(0)})$ не зависит от номера итерации.

Если система (4.3.1) имеет большую размерность, то хорошие результаты дает комбинация метода Ньютона и модифицированного метода Ньютона, когда после очередной итерации метода Ньютона выполняется несколько итераций модифицированного метода Ньютона.

Метод Ньютона (4.3.6) допускает еще одно видоизменение, которое называется *методом Ньютона с параметром* и имеет вид

$$\mathbf{F}'(\mathbf{x}^{(r)}) \cdot (\mathbf{x}^{(r+1)} - \mathbf{x}^{(r)}) = -\tau^{(r)} \mathbf{f}(\mathbf{x}^{(r)}), \quad (4.3.8)$$

где $\tau^{(r)}$ — изменяющийся на каждой итерации параметр. Параметр как бы корректирует правую часть системы (4.3.3), «улучшая» решение на итерации. Для использования метода Ньютона с параметром нужно иметь способ вычислять или оценивать параметр $\tau^{(r)}$. Оценку параметра можно выполнять по приращениям неизвестных на итерационном шаге. Параметр обычно имеет значения $1 \leq \tau^{(r)} \leq 2$. Метод Ньютона с параметром обладает линейной сходимостью.

Метод простой итерации. Рассмотренные методы Ньютона являются представителями неявных методов решения нелинейных систем уравнений в силу того, что матрица $\mathbf{F}'(\mathbf{x}^{(r)})$ не является единичной. К явным методам относится *метод простой итерации*, который получим из (4.3.8) заменой $\mathbf{F}'(\mathbf{x}^{(r)})$ на единичную матрицу

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} - \tau^{(r)} \mathbf{f}(\mathbf{x}^{(r)}). \quad (4.3.9)$$

Если в (4.3.9) $\tau^{(r)} = \tau = \text{const}$ не зависит от итерации, то получим *метод релаксации*. Метод простой итерации сходится, если $\|\mathbf{E} - \tau^{(r)} \mathbf{F}'(\mathbf{x}^{(r)})\| < 1$. Метод простой итерации не требует решения системы линейных алгебраических уравнений, но сходится медленнее других рассмотренных методов (в определенных случаях он не сходится).

Метод Якоби. Рассмотренные до сих пор методы являлись линейными относительно новой итерации $x_i^{(r)}$, $i = 1, 2, \dots, n$. Существуют и нелинейные методы, когда для вычисления $x_i^{(r)}$ приходится решать нелинейные уравнения. Одним из нелинейных

методов является *метод Якоби*, в котором решаются n нелинейных уравнений

$$\begin{aligned} f_1(x_1^{(r+1)}, x_2^{(r)}, \dots, x_n^{(r)}) &= 0, \\ f_2(x_1^{(r)}, x_2^{(r+1)}, \dots, x_n^{(r)}) &= 0, \\ \dots & \\ f_n(x_1^{(r)}, x_2^{(r)}, \dots, x_n^{(r+1)}) &= 0. \end{aligned} \quad (4.3.10)$$

В каждом уравнении неизвестной считается только одна величина. Любое уравнение $f_i(x_1^{(r)}, x_2^{(r)}, \dots, x_{i-1}^{(r)}, x_i^{(r+1)}, x_{i+1}^{(r)}, \dots, x_n^{(r)}) = 0$ может быть решено отдельно от других одним из описанных методов применительно к одному уравнению.

Метод Зейделя. Другим нелинейным методом является *метод Зейделя*, в котором решаются n нелинейных уравнений

$$\begin{aligned} f_1(x_1^{(r+1)}, x_2^{(r)}, \dots, x_n^{(r)}) &= 0, \\ f_2(x_1^{(r+1)}, x_2^{(r+1)}, \dots, x_n^{(r)}) &= 0, \\ \dots & \\ f_n(x_1^{(r+1)}, x_2^{(r+1)}, \dots, x_n^{(r+1)}) &= 0, \end{aligned} \quad (4.3.11)$$

в i -м уравнении неизвестной считается величина $x_i^{(r+1)}$. Уравнения (4.3.11) решаются последовательно, так что в i -м уравнении все $x_j^{(r+1)}$, $j = 1, 2, \dots, i - 1$ известны из решения предыдущих уравнений. Нелинейные методы сходятся гораздо медленнее метода Ньютона. В нелинейных методах не требуется строить матрицу системы уравнений. Решение каждого уравнения (4.3.10) или (4.3.11) выполняется независимо от других уравнений системы. Решение каждого уравнения в свою очередь выполняется итерационно. Эти итерации называются внутренними, а итерации для всей системы нелинейных уравнений называются внешними.

Большое распространение получили гибридные методы, когда внешние итерации осуществляются одним методом, а внутренние итерации — другим. Например, внешние итерации можно выполнять методом Ньютона, а внутренние (итерационное решение системы линейных уравнений) — методом Зейделя. При этом внутренние итерации необязательно выполнять до тех пор, пока решение будет удовлетворять заданной точности, а можно ограничиться некоторым заданным числом итераций.

Метод градиента. В *методе скорейшего спуска* или *методе градиента* решение системы (4.3.1) сводится к задаче отыскания минимумов функции $U(x_1, x_2, \dots, x_n)$, которую можно построить различными способами, например,

$$U(x_1, x_2, \dots, x_n) = \sum_{i=1}^n (f_i(x_1, x_2, \dots, x_n))^2 = \mathbf{f}^T \cdot \mathbf{f} \quad (4.3.12)$$

или

$$U(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} f_i(x_1, x_2, \dots, x_n) f_j(x_1, x_2, \dots, x_n) = \mathbf{f}^T \cdot \mathbf{A} \cdot \mathbf{f}, \quad (4.3.13)$$

где a_{ij} — элементы некоторой положительно определенной матрицы \mathbf{A} , $\mathbf{f}^T = [f_1 \ f_2 \ \dots \ f_n]$ и $\mathbf{f} = [f_1 \ f_2 \ \dots \ f_n]^T$ — матрица-строка и матрица-столбец, элементами которых являются правые части уравнений системы (4.3.1). Функцию (4.3.12) будем считать частным случаем функции (4.3.13). Если $x_1^*, x_2^*, \dots, x_n^*$ есть некоторое решение системы (4.3.1), то $U(x_1^*, x_2^*, \dots, x_n^*) = 0$. В других точках $U(x_1, x_2, \dots, x_n) > 0$ в силу положительной определенности матрицы \mathbf{A} . Таким образом, решение системы (4.3.1) сводится к поиску нулевых минимумов функции $U(x_1, x_2, \dots, x_n)$. Будем рассматривать функцию $U(x_1, x_2, \dots, x_n)$ как некоторую скалярную функцию n -мерного

векторного пространства и будем обозначать ее еще и как $U(\mathbf{x})$, где искомые величины считаются компонентами n -мерного вектора $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^\top$. Производные $\partial U / \partial x_j$, $j = 1, 2, \dots, n$ в этом пространстве являются компонентами вектора

$$\nabla U = \text{grad } U = \left[\frac{\partial U}{\partial x_1} \quad \frac{\partial U}{\partial x_2} \quad \dots \quad \frac{\partial U}{\partial x_n} \right]^\top, \quad (4.3.14)$$

для функции (4.3.12):

$$\frac{\partial U}{\partial x_j} = \sum_{i=1}^n 2f_i(x_1, x_2, \dots, x_n) \frac{\partial f_i(x_1, x_2, \dots, x_n)}{\partial x_j}.$$

В матричной записи на r -й итерации градиент (4.3.14) для функции (4.3.12) имеет вид

$$\nabla U = 2\mathbf{F}'^\top(\mathbf{x}^{(r)}) \cdot \mathbf{f}(\mathbf{x}^{(r)}), \quad (4.3.15)$$

где $\mathbf{F}'^\top(\mathbf{x}^{(r)})$ — транспонированная матрица Якоби $\mathbf{F}'(\mathbf{x}^{(r)})$, которая определена в (4.3.6). Вектор ∇U как градиент скалярной функции, вычисленный при некоторых значениях $x_1^{(r)}, x_2^{(r)}, \dots, x_n^{(r)}$, ортогонален поверхности уровня, для которой $U(x_1, x_2, \dots, x_n) = U(x_1^{(r)}, x_2^{(r)}, \dots, x_n^{(r)}) = \text{const}$, и направлен в сторону возрастания функции $U(x_1, x_2, \dots, x_n)$ в данной точке n -мерного пространства. Таким образом, итерационная зависимость для определения экстремума функции $U(x_1, x_2, \dots, x_n)$ имеет вид

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} - \lambda^{(r)} \nabla U(\mathbf{x}^{(r)}) = \mathbf{x}^{(r)} - \lambda^{(r)} 2\mathbf{F}'^\top(\mathbf{x}^{(r)}) \cdot \mathbf{f}(\mathbf{x}^{(r)}). \quad (4.3.16)$$

Остается определить множитель $\lambda^{(r)}$. Для этого подставим новое приближение (4.3.16) в исследуемую функцию $U(x_1, x_2, \dots, x_n)$ и рассмотрим ее как функцию параметра λ

$$U(\lambda) = U(\mathbf{x}^{(r)} - \lambda \nabla U(\mathbf{x}^{(r)})) = \sum_{i=1}^n (f_i(\mathbf{x}^{(r)} - \lambda \nabla U(\mathbf{x}^{(r)})))^2.$$

Параметр λ должен иметь такой знак, чтобы от итерации к итерации функция $U(\lambda)$ уменьшалась, а величина параметра λ должна позволять при данном векторе $\nabla U(\mathbf{x}^{(r)})$ как можно ближе продвинуться к минимуму функции $U(\lambda)$. Параметр λ приближенно можно найти следующим образом. Считая параметр λ малой величиной, разложим функцию $U(\lambda)$ в ряд Тейлора по степеням λ в окрестности нуля с точностью до слагаемых второй степени. В результате получим

$$U(\lambda) \approx \sum_{i=1}^n \left(f_i(\mathbf{x}^{(r)}) - \lambda (\nabla f_i(\mathbf{x}^{(r)}))^\top \cdot \nabla U(\mathbf{x}^{(r)}) \right)^2, \quad (4.3.17)$$

где

$$(\nabla f_i(\mathbf{x}^{(r)}))^\top = \left[\frac{\partial f_i}{\partial x_1} \quad \frac{\partial f_i}{\partial x_2} \quad \dots \quad \frac{\partial f_i}{\partial x_n} \right] \Big|_{\mathbf{x}^{(r)}}.$$

Функция $U(\lambda)$ дает изменение значения функции $U(\mathbf{x}^{(r)})$ вдоль нормали $\nabla U(\mathbf{x}^{(r)})$ к поверхности уровня в точке $\mathbf{x}^{(r)}$, поэтому значение параметра получим из условия минимума функции $U(\lambda)$ как корень уравнения $U'(\lambda) = 0$. Возьмем производную $U(\lambda)$

по λ , приравняем ее нулю и получим уравнение для определения λ

$$-2 \sum_{i=1}^n \left((\nabla f_i(\mathbf{x}^{(r)}))^{\top} \cdot \nabla U(\mathbf{x}^{(r)}) \right) \left(f_i(\mathbf{x}^{(r)}) - \lambda (\nabla f_i(\mathbf{x}^{(r)}))^{\top} \cdot \nabla U(\mathbf{x}^{(r)}) \right) = 0.$$

Отсюда получим параметр

$$\lambda^{(r)} = \frac{\sum_{i=1}^n \left((\nabla f_i(\mathbf{x}^{(r)}))^{\top} \cdot \nabla U(\mathbf{x}^{(r)}) \right) f_i(\mathbf{x}^{(r)})}{\sum_{i=1}^n \left((\nabla f_i(\mathbf{x}^{(r)}))^{\top} \cdot \nabla U(\mathbf{x}^{(r)}) \right)^2} = \frac{1}{2} \frac{(\mathbf{F}' \cdot \mathbf{F}'^{\top} \cdot \mathbf{f}) \cdot \mathbf{f}}{(\mathbf{F}' \cdot \mathbf{F}'^{\top} \cdot \mathbf{f}) \cdot (\mathbf{F}' \cdot \mathbf{F}'^{\top} \cdot \mathbf{f})^{\top}},$$

где мы использовали равенство (4.3.15) и для краткости обозначили $\mathbf{F}' = \mathbf{F}'(\mathbf{x}^{(r)})$, $\mathbf{f} = \mathbf{f}(\mathbf{x}^{(r)})$. Таким образом, для принятых упрощений равенство (4.3.16) примет вид

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} - \frac{(\mathbf{F}' \cdot \mathbf{F}'^{\top} \cdot \mathbf{f}) \cdot \mathbf{f}}{(\mathbf{F}' \cdot \mathbf{F}'^{\top} \cdot \mathbf{f}) \cdot (\mathbf{F}' \cdot \mathbf{F}'^{\top} \cdot \mathbf{f})^{\top}} \mathbf{F}'^{\top}(\mathbf{x}^{(r)}) \cdot \mathbf{f}(\mathbf{x}^{(r)}). \quad (4.3.18)$$

В процессе решения следует следить за поведением параметра $\lambda^{(r)}$. Он не должен возрастать от итерации к итерации и должен изменяться плавно. Существует много разновидностей метода градиента на основе управления этим параметром. Для дважды дифференцируемых функций $f_i(x_1, x_2, \dots, x_n)$, $i = 1, 2, \dots, n$, можно получить более точные формулы для параметра $\lambda^{(r)}$.

Если начальное приближение выбрано удачно и в окрестности искомого нулевого минимума функции $U(x_1, x_2, \dots, x_n)$ нет других минимумов, то метод градиента довольно быстро даст искомое решение с заданной точностью. Если в окрестности искомого решения функция $U(x_1, x_2, \dots, x_n)$ имеет другие минимумы, то итерационный процесс (4.3.16) сойдется, но не обязательно даст решение исходной системы уравнений (4.3.1).

Сжимающий оператор. Все рассмотренные методы решения системы нелинейных уравнений можно записать в виде

$$\mathbf{M}(x^{(r)}) (\mathbf{x}^{(r+1)} - \mathbf{x}^{(r)}) = -\tau^{(r)} \mathbf{b}(x^{(r)}), \quad (4.3.19)$$

где $\mathbf{M}(x^{(r)})$ — квадратная матрица размерности n , $\mathbf{b}(x^{(r)})$ — матрица-столбец, $\tau^{(r)}$ — скалярный параметр, $\mathbf{x}^{(r)}$ — n -мерный вектор некоторого линейного нормированного пространства, компонентами которого являются искомые величины на r -й итерации решения. Для всех методов в искомой точке $\mathbf{x}^* = [x_1^* \ x_2^* \ \dots \ x_n^*]^{\top}$ выполняется равенство $\mathbf{b}(x^*) = 0$. Если условно решить уравнение (4.3.19) относительно $\mathbf{x}^{(r+1)}$, то получим

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} - \tau^{(r)} \mathbf{M}^{-1}(x^{(r)}) \cdot \mathbf{b}(x^{(r)}) = \mathbf{S}(\mathbf{x}^{(r)}), \quad (4.3.20)$$

где $\mathbf{S}(\mathbf{x}^{(r)})$ будем называть оператором, отображающим n -мерное пространство само на себя. Точка \mathbf{x}^* является неподвижной точкой оператора, так как для нее выполняется равенство $\mathbf{x}^* = \mathbf{S}(\mathbf{x}^*)$. Оператор называется сжимающим в некоторой области n -мерного линейного нормированного пространства, если существует число $0 < q < 1$ такое, что для любых $\mathbf{x}^{(r+1)}$ и $\mathbf{x}^{(r)}$, принадлежащих этой области, выполняется неравенство

$$\|\mathbf{S}(\mathbf{x}^{(r+1)}) - \mathbf{S}(\mathbf{x}^{(r)})\| \leq q \|\mathbf{x}^{(r+1)} - \mathbf{x}^{(r)}\|.$$

Величина q называется коэффициентом сжатия. Существует доказательство того, что если оператор $S(x)$ является сжимающим на всем конечномерном линейном нормированном пространстве, то он имеет единственную неподвижную точку x^* и итерационный процесс (4.3.20) сходится к этой неподвижной точке при любом начальном $x^{(0)}$. Для погрешности решения на r -й итерации справедлива оценка

$$\|x^{(r)} - x^*\| \leq q^r \|x^{(0)} - x^*\|.$$

Операторы, соответствующие рассмотренным итерационным процессам, являются сжимающими только в некоторых окрестностях решений, но если в процессе вычислений мы не выходим из окрестности, где оператор является сжимающим, то итерационный процесс сойдется к единственному в этой окрестности решению.

4.4.* Решение системы линейных уравнений

В процессе решения системы нелинейных уравнений (4.3.1) на каждой итерации требуется решать систему линейных алгебраических уравнений

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n &= b_2, \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n &= b_3, \\ \dots &\dots \dots \dots \dots \dots \dots \dots \dots \dots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \dots + a_{nn}x_n &= b_n \end{aligned} \tag{4.4.1}$$

для неизвестных величин $x_i, i = 1, 2, \dots, n$. Если вернуться к системе уравнений (4.3.3), то из нее система (4.4.1) получена путем обозначения Δx_i через $x_i; \frac{\partial f_i(x^{(r)})}{\partial x_j}$ —

через $a_{ij}; -f_i(x^{(r)})$ — через b_i . Для удобства записи системы уравнений коэффициенты a_{ij} представим в виде квадратной матрицы A , неизвестные x_i и правые части b_i представим в виде матриц-столбцов $x = [x_1 \ x_2 \ \dots \ x_n]^T$ и $b = [b_1 \ b_2 \ \dots \ b_n]^T$. В матричной записи система линейных уравнений (4.4.1) имеет вид

$$A \cdot x = b. \tag{4.4.2}$$

Система (4.4.1) имеет единственное решение, если определитель матрицы A отличен от нуля. Существует много методов решения системы уравнений (4.4.1). Некоторые методы рассчитаны на матрицы A специального вида (трехдиагональные, симметричные, разреженные). При выполнении операций над геометрическими объектами матрица A имеет общий вид. Все методы решения систем линейных алгебраических уравнений условно можно разделить на две группы: прямые методы и итерационные методы. В прямых методах решение находится за конечное число арифметических действий. В итерационных методах решение находится как предел последовательных приближений искомых функций $x_i^{(r)}, i = 1, 2, \dots, n$. Как правило, прямые методы быстрее и с большей надежностью дают решение системы, но итерационные методы требуют во много раз меньше памяти во время вычислений. Эти обстоятельства играют главную роль при выборе метода решения.

Метод исключения Гаусса. Одним из наиболее часто применяемых прямых методов решения системы линейных уравнений является *метод исключения Гаусса*. Заметим, что решение системы уравнений (4.4.1) не зависит от порядка, в котором записаны составляющие систему уравнения, поэтому строки системы можно безболезненно переставлять местами, если это будет необходимо. Среди коэффициентов $a_{11}, a_{21}, \dots, a_{n1}$ найдем максимальный по абсолютной величине коэффициент. Пусть это будет

в которой только первое уравнение содержит неизвестную x_1 . Если бы из остальных уравнений удалось найти x_2, x_3, \dots, x_n , то неизвестная x_1 определилась бы равенством $x_1 = y_1 - u_{12}x_2 - u_{13}x_3 - \dots - u_{1n}x_n$. Матрицу системы уравнений (4.4.5) обозначим через $\mathbf{A}^{(1)}$. Рассмотрим систему уравнений (4.4.5) без первого уравнения. Она аналогична системе (4.4.1), только содержит на одно уравнение меньше и меньше на одну неизвестную величину

$$\begin{aligned} a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \dots + a_{2n}^{(1)}x_n &= b_2^{(1)}, \\ a_{32}^{(1)}x_2 + a_{33}^{(1)}x_3 + \dots + a_{3n}^{(1)}x_n &= b_3^{(1)}, \\ \dots & \\ a_{n2}^{(1)}x_2 + a_{n3}^{(1)}x_3 + \dots + a_{nn}^{(1)}x_n &= b_n^{(1)}. \end{aligned} \quad (4.4.6)$$

С системой (4.4.6) поступим аналогично тому, как мы от системы (4.4.1) перешли к системе (4.4.5). В результате придем к системе уравнений

$$\begin{aligned} x_1 + u_{12}x_2 + u_{13}x_3 + u_{14}x_4 + \dots + u_{1n}x_n &= y_1, \\ x_2 + u_{23}x_3 + u_{24}x_4 + \dots + u_{2n}x_n &= y_2, \\ a_{33}^{(2)}x_3 + a_{34}^{(2)}x_4 + \dots + a_{3n}^{(2)}x_n &= b_3^{(2)}, \\ a_{43}^{(2)}x_3 + a_{44}^{(2)}x_4 + \dots + a_{4n}^{(2)}x_n &= b_4^{(2)}, \\ \dots & \\ a_{n3}^{(2)}x_3 + a_{n4}^{(2)}x_4 + \dots + a_{nn}^{(2)}x_n &= b_n^{(2)}, \end{aligned} \quad (4.4.7)$$

где

$$\begin{aligned} u_{2j} &= \frac{a_{2j}^{(1)}}{a_{22}^{(1)}}, & y_2 &= \frac{b_2^{(1)}}{a_{22}^{(1)}}, \\ a_{kj}^{(2)} &= a_{kj}^{(1)} - a_{k2}^{(1)} \frac{a_{2j}^{(1)}}{a_{22}^{(1)}}, & j &= 3, 4, \dots, n, \\ b_k^{(2)} &= b_k^{(1)} - a_{k2}^{(1)} \frac{b_2^{(1)}}{a_{22}^{(1)}}, & k &= 3, 4, \dots, n. \end{aligned}$$

Из уравнений (4.4.7) можно выделить систему, содержащую еще на одну неизвестную величину меньше. Матрицу системы уравнений (4.4.7) обозначим через $\mathbf{A}^{(2)}$. Продолжая этот процесс далее, придем к системе линейных уравнений треугольного вида

$$\begin{aligned} x_1 + u_{12}x_2 + u_{13}x_3 + \dots + u_{1n-1}x_{n-1} + u_{1n}x_n &= y_1, \\ x_2 + u_{23}x_3 + \dots + u_{2n-1}x_{n-1} + u_{2n}x_n &= y_2, \\ x_3 + \dots + u_{3n-1}x_{n-1} + u_{3n}x_n &= y_3, \\ \dots & \\ x_{n-1} + u_{n-1n}x_n &= y_{n-1}, \\ x_n &= y_n, \end{aligned} \quad (4.4.8)$$

которая эквивалентна исходной системе (4.4.1). Таким образом, мы перешли от системы уравнений (4.4.2) к системе

$$\mathbf{U} \cdot \mathbf{x} = \mathbf{y}. \quad (4.4.9)$$

Матрица \mathbf{U} системы (4.4.8) содержит нули под главной диагональю и поэтому называется верхней треугольной матрицей. На главной диагонали этой матрицы стоят еди-

ницы. Коэффициенты матрицы вычисляются строка за строкой

$$U = \begin{bmatrix} 1 & u_{12} & \dots & u_{1n-1} & u_{1n} \\ 0 & 1 & \dots & u_{2n-1} & u_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & u_{n-1n} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{a_{12}}{a_{11}} & \dots & \frac{a_{1n-1}}{a_{11}} & \frac{a_{1n}}{a_{11}} \\ 0 & 1 & \dots & \frac{a_{2n-1}^{(1)}}{a_{22}^{(1)}} & \frac{a_{2n}^{(1)}}{a_{22}^{(1)}} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \frac{a_{n-1n}^{(n-2)}}{a_{n-1n-1}^{(n-2)}} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

Алгоритм вычисления коэффициентов матрицы U по коэффициентам матрицы A выражается формулами

$$a_{kj}^{(0)} = a_{kj}, \quad k = 1, 2, \dots, n,$$

$$u_{kj} = \frac{a_{kj}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad j = k, k+1, \dots, n, \quad (4.4.10)$$

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - a_{ik}^{(k-1)} \frac{a_{kj}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad i = k+1, k+2, \dots, n,$$

к которым нужно добавить поиск максимального коэффициента $a_{ik}^{(k)}$, $i = k+1, k+2, \dots, n$ и перестановку уравнений. Будем считать, что перестановка уравнений на каждом шаге выполняется описанным выше способом. Алгоритм вычисления правой части y по коэффициентам матрицы A и b выражается формулами

$$b_k^{(0)} = b_k, \quad k = 1, 2, \dots, n, \quad y_k = \frac{b_k^{(k-1)}}{a_{kk}^{(k-1)}}, \quad (4.4.11)$$

$$b_i^{(k)} = b_i^{(k-1)} - a_{ik}^{(k-1)} \frac{b_k^{(k-1)}}{a_{kk}^{(k-1)}}, \quad i = k+1, k+2, \dots, n.$$

Процесс перехода от системы (4.4.1) к системе (4.4.8) называется прямым ходом метода Гаусса. Обратным ходом метода Гаусса называется процесс определения неизвестных $x_n, x_{n-1}, \dots, x_2, x_1$ из системы (4.4.8). Это легко сделать, определяя $x_n, x_{n-1}, \dots, x_2, x_1$ последовательно, начиная с последнего неизвестного и используя соотношение

$$x_i = y_i - \sum_{j=i+1}^n u_{ij} x_j, \quad i = n-1, n-2, \dots, 2, 1. \quad (4.4.12)$$

Описанный метод решения системы линейных уравнений называется методом Гаусса с выбором главного элемента по столбцу, так как в процессе прямого хода мы искали уравнение с максимальным коэффициентом при «первом» неизвестном соответствующей системы уравнений.

В методе Гаусса можно искать главный элемент по строке. В этом случае меняются местами не уравнения, а неизвестные во всех уравнениях. Например, пусть требуется решить систему уравнений (4.4.1). Среди коэффициентов $a_{11}, a_{12}, \dots, a_{1n}$, найдем максимальный по абсолютной величине коэффициент. Пусть это будет коэффициент a_{1k} .

формул (4.4.10); из этих же формул можно в явном виде получить матрицу L . Переход от системы уравнений (4.4.1) к эквивалентной ее системе (4.4.5) в матричном виде можно представить как умножение уравнения (4.4.2) слева на матрицу S_1 :

$$S_1 \cdot A \cdot x = S_1 \cdot b, \quad (4.4.20)$$

где матрица S_1 имеет коэффициенты

$$S_1 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{11} & & & & \\ -\frac{a_{21}}{a_{11}} & 1 & 0 & \dots & 0 \\ \frac{a_{31}}{a_{11}} & & & & \\ -\frac{a_{n1}}{a_{11}} & 0 & 0 & \dots & 1 \end{bmatrix}$$

Переход от системы уравнений (4.4.5) к эквивалентной ее системе (4.4.7) в матричном виде можно представить как умножение уравнения (4.4.20) слева на матрицу S_2 :

$$S_2 \cdot S_1 \cdot A \cdot x = S_2 \cdot S_1 \cdot b, \quad (4.4.21)$$

где матрица S_2 имеет коэффициенты

$$S_2 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{a_{22}^{(1)}} & 0 & \dots & 0 \\ 0 & -\frac{a_{32}^{(1)}}{a_{22}^{(1)}} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & -\frac{a_{n2}^{(1)}}{a_{22}^{(1)}} & 0 & \dots & 1 \end{bmatrix}.$$

На k -м шаге исключения метода Гаусса система уравнений, полученная на предыдущем шаге, умножается на матрицу

$$S_k = \begin{bmatrix} 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & \frac{1}{a_{kk}^{(k-1)}} & 0 & \dots & 0 \\ 0 & \dots & 0 & -\frac{a_{k+1k}^{(k-1)}}{a_{kk}^{(k-1)}} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & -\frac{a_{nk}^{(k-1)}}{a_{kk}^{(k-1)}} & 0 & \dots & 1 \end{bmatrix}. \quad (4.4.22)$$

В результате прямого хода метода исключения Гаусса приходим к системе уравнений (4.4.8), которую можно представить в виде

$$S \cdot A \cdot x = S \cdot b, \quad (4.4.23)$$

где

$$\mathbf{S} = \mathbf{S}_n \cdot \mathbf{S}_{n-1} \cdot \dots \cdot \mathbf{S}_2 \cdot \mathbf{S}_1. \quad (4.4.24)$$

Таким образом, мы получили, что $\mathbf{U} = \mathbf{S}_n \cdot \mathbf{S}_{n-1} \cdot \dots \cdot \mathbf{S}_2 \cdot \mathbf{S}_1 \cdot \mathbf{A}$, откуда следует, что

$$\mathbf{A} = (\mathbf{S}_n \cdot \mathbf{S}_{n-1} \cdot \dots \cdot \mathbf{S}_2 \cdot \mathbf{S}_1)^{-1} \cdot \mathbf{U}. \quad (4.4.25)$$

Матрица типа (4.4.22) называется элементарной нижней треугольной матрицей. Обратная к (4.4.22) матрица также является элементарной нижней треугольной матрицей и имеет вид

$$\mathbf{S}_k^{-1} = \begin{bmatrix} 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & a_{kk}^{(k-1)} & 0 & \dots & 0 \\ 0 & \dots & 0 & a_{k+1k}^{(k-1)} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & a_{nk}^{(k-1)} & 0 & \dots & 1 \end{bmatrix}. \quad (4.4.26)$$

Произведение элементарных нижних треугольных матриц (4.4.26) и есть нижняя треугольная матрица, которая присутствует в разложении (4.4.14):

$$\begin{aligned} (\mathbf{S}_n \cdot \mathbf{S}_{n-1} \cdot \dots \cdot \mathbf{S}_2 \cdot \mathbf{S}_1)^{-1} &= \mathbf{S}_1^{-1} \cdot \mathbf{S}_2^{-1} \cdot \dots \cdot \mathbf{S}_{n-1}^{-1} \cdot \mathbf{S}_n^{-1} = \\ &= \mathbf{L} = \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22}^{(1)} & 0 & \dots & 0 \\ a_{31} & a_{32}^{(1)} & a_{33}^{(2)} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2}^{(1)} & a_{n3}^{(2)} & \dots & a_{nn}^{(n-1)} \end{bmatrix}. \end{aligned} \quad (4.4.27)$$

Докажем, что любую квадратную матрицу \mathbf{A} , все угловые миноры которой отличны от нуля, можно разложить в произведение нижней треугольной матрицы \mathbf{L} и верхней треугольной матрицы \mathbf{U} с единицами на главной диагонали. Будем называть i -м угловым минором M_i матрицы \mathbf{A} определитель, составленный из первых i элементов первых i строк этой матрицы:

$$M_i = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1i} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2i} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3i} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & a_{i3} & \dots & a_{ii} \end{vmatrix}. \quad (4.4.28)$$

Доказательство проведем методом индукции. Для матрицы второго порядка это разложение имеет вид

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \cdot \begin{bmatrix} 1 & u_{12} \\ 0 & 1 \end{bmatrix},$$

где $l_{11} = a_{11}$, $l_{21} = a_{21}$, $l_{22} = \frac{a_{11}a_{22} - a_{21}a_{12}}{a_{11}}$, $u_{12} = \frac{a_{12}}{a_{11}}$. По условию $a_{11} \neq 0$, $a_{11}a_{22} - a_{21}a_{12} \neq 0$, следовательно, $l_{11} \neq 0$, $l_{22} \neq 0$. Пусть существует разложение

(4.4.14) матриц $(k-1)$ -го порядка; докажем, что аналогичное разложение существует для матриц k -го порядка. Для этого представим матрицу k -го порядка в виде

$$A_k = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k-1} & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k-1} & a_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ a_{k-11} & a_{k-12} & \dots & a_{k-1k-1} & a_{k-1k} \\ a_{k1} & a_{k2} & \dots & a_{kk-1} & a_{kk} \end{bmatrix} = \begin{bmatrix} A_{k-1} & \mathbf{b}_k \\ \mathbf{a}_k & a_{kk} \end{bmatrix},$$

где через A_{k-1} обозначена матрица $(k-1)$ -го порядка, $\mathbf{a}_k = [a_{k1} \ a_{k2} \ \dots \ a_{kk-1}]$, $\mathbf{b}_k = [a_{1k} \ a_{2k} \ \dots \ a_{k-1k}]^T$. Согласно предположению индукции $A_{k-1} = L_{k-1} \cdot U_{k-1}$, где L_{k-1} — нижняя треугольная матрица $(k-1)$ -го порядка, U_{k-1} — верхняя треугольная матрица $(k-1)$ -го порядка с единичной диагональю. Будем искать разложение матрицы A_k в виде

$$\begin{bmatrix} A_{k-1} & \mathbf{b}_k \\ \mathbf{a}_k & a_{kk} \end{bmatrix} = \begin{bmatrix} L_{k-1} & \mathbf{0} \\ \mathbf{l}_k & l_{kk} \end{bmatrix} \cdot \begin{bmatrix} U_{k-1} & \mathbf{u}_k \\ \mathbf{0} & 1 \end{bmatrix}, \quad (4.4.29)$$

где $\mathbf{l}_k = [l_{k1} \ l_{k2} \ \dots \ l_{kk-1}]$, $\mathbf{u}_k = [u_{1k} \ u_{2k} \ \dots \ u_{k-1k}]^T$ — неизвестные пока матрицы. Выполнив умножение в (4.4.29), придем к уравнениям

$$\begin{aligned} L_{k-1} \cdot U_{k-1} &= A_{k-1}, & \mathbf{l}_k \cdot U_{k-1} &= \mathbf{a}_k, \\ L_{k-1} \cdot \mathbf{u}_k &= \mathbf{b}_k, & \mathbf{l}_k \cdot \mathbf{u}_k + l_{kk} &= a_{kk}, \end{aligned}$$

первое из которых выполняется из предположения индукции, а из остальных трех определяются \mathbf{l}_k , \mathbf{u}_k и l_{kk} . Так как матрицы U_{k-1} и L_{k-1} треугольные, то вычисления выполняются последовательной подстановкой

$$l_{k1} = a_{k1}, \quad l_{ki} = a_{ki} - \sum_{j=1}^{i-1} l_{kj} u_{ji}, \quad i = 2, 3, \dots, k-1, \quad (4.4.30)$$

$$u_{1k} = \frac{a_{1k}}{l_{11}}, \quad u_{ik} = \frac{1}{l_{ii}} \left(a_{ik} - \sum_{j=1}^{i-1} l_{ij} u_{jk} \right), \quad i = 2, 3, \dots, k-1, \quad (4.4.31)$$

$$l_{kk} = a_{kk} - \sum_{j=1}^{k-1} l_{kj} u_{jk}. \quad (4.4.32)$$

Равенство (4.4.32) представляет собой продолжение равенств (4.4.30) для $i = k$. Так как значение k не было определено, то формулы (4.4.30)–(4.4.32) являются формулами для вычисления элементов нижней треугольной матрицы L и верхней треугольной матрицы U разложения квадратной матрицы. Метод решения системы линейных алгебраических уравнений путем разложения матрицы системы на нижнюю треугольную и верхнюю треугольную матрицы часто называют *методом Холецкого*. Коэффициент l_{kk} не может быть равен нулю. Это следует из того, что $\det A_k \neq 0$ и

$$|A_k| = |L_{k-1}| l_{kk} |U_{k-1}| = l_{kk} |L_{k-1}| \neq 0.$$

Определитель треугольной матрицы (верхней или нижней) равен произведению ее диагональных элементов, поэтому $\det U_{k-1} = 1$, так как на главной диагонали треугольной матрицы U_{k-1} стоят единицы. Таким образом, $l_{kk} \neq 0$.

Формулы (4.4.29)–(4.4.32) доказывают, что любую квадратную матрицу A , все угловые миноры которой отличны от нуля, можно разложить в произведение нижней треугольной матрицы L и верхней треугольной матрицы U с единицами на главной диагонали. Такое разложение единственно. Единственность разложения можно доказать методом от противного.

Метод Якоби. Одним из итерационных методов решения системы линейных алгебраических уравнений (4.4.1) является *метод Якоби*. Выразим из каждого уравнения системы одну из неизвестных величин x_i , $i = 1, 2, \dots, n$. Например, из i -го уравнения выразим i -ю неизвестную и перепишем систему в виде

$$x_i = \frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} x_j, \quad i = 1, 2, \dots, n. \quad (4.4.33)$$

В (4.4.33) предполагается, что все a_{ii} не равны нулю. Этого всегда можно достичь с помощью перестановок уравнений местами, что приводит к перестановке строк матрицы. Через $x_1^{(r)}$, $x_2^{(r)}$, \dots , $x_n^{(r)}$ обозначим значения неизвестных на r -й итерации решения системы уравнений (4.4.1). Метод Якоби определяет неизвестные на $(r+1)$ -м приближении через неизвестные на r -м приближении с помощью равенства

$$x_i^{(r+1)} = \frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(r)} - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} x_j^{(r)}, \quad i = 1, 2, \dots, n. \quad (4.4.34)$$

Начальные приближения $x_1^{(0)}$, $x_2^{(0)}$, \dots , $x_n^{(0)}$ могут быть выбраны любыми, при полной неопределенности их можно положить равными нулю. Итерации выполняются до тех пор, пока на очередной итерации погрешности всех неизвестных не станут меньше заданной величины $\varepsilon > 0$

$$|x_i^{(r+1)} - x_i^{(r)}| < \varepsilon \quad \text{для всех } i = 1, 2, \dots, n. \quad (4.4.35)$$

Для n -мерного вектора, компонентами которого являются неизвестные на r -м приближении, введем обозначение $\mathbf{x}^{(r)} = [x_1^{(r)} \ x_2^{(r)} \ \dots \ x_n^{(r)}]^\top$. Через D обозначим диагональную матрицу, ненулевые элементы которой равны диагональным элементам матрицы A . Через C обозначим нижнюю треугольную матрицу, ненулевые элементы которой равны элементам матрицы A , стоящим под главной диагональю матрицы A . Через R обозначим верхнюю треугольную матрицу, ненулевые элементы которой равны элементам матрицы A , стоящим над главной диагональю матрицы A . Эти матрицы имеют вид

$$C = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ a_{21} & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n-11} & a_{n-12} & \dots & 0 & 0 \\ a_{n1} & a_{n2} & \dots & a_{nn-1} & 0 \end{bmatrix}, \quad D = \begin{bmatrix} a_{11} & 0 & \dots & 0 & 0 \\ 0 & a_{22} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{n-1n-1} & 0 \\ 0 & 0 & \dots & 0 & a_{nn} \end{bmatrix},$$

$$R = \begin{bmatrix} 0 & a_{12} & \dots & a_{1n-1} & a_{1n} \\ 0 & 0 & \dots & a_{2n-1} & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & a_{n-1n} \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Через \mathbf{b} обозначим вектор правой части: $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_n]^\top$. В принятых обозначениях уравнение (4.4.2) будет записано следующим образом:

$$(C + R + D) \cdot \mathbf{x} = \mathbf{b}.$$

Это неравенство также приводит к требованию преобладания диагональных элементов в матрице A , т. е. методы Якоби и Зейделя для сходимости предъявляют одинаковые требования к коэффициентам системы уравнений

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}|, \quad \text{для всех } i = 1, 2, \dots, n.$$

Обычно метод Зейделя сходится лучше, чем метод Якоби, хотя это происходит не всегда. Процесс Зейделя может сходиться, когда метод Якоби расходится, хотя известны и обратные случаи.

Метод релаксации. Иногда перестановка местами уравнений или перенумерация неизвестных улучшают сходимость метода Зейделя. Можно использовать еще один способ для улучшения сходимости. Этот способ приводит к *методу релаксации*. На r -м приближении каждое уравнение системы (4.4.1) удовлетворяется неточно и дает некоторую невязку

$$r_i = \frac{1}{a_{ii}} (a_{i1}x_1^{(r)} + a_{i2}x_2^{(r)} + a_{i3}x_3^{(r)} + \dots + a_{in}x_n^{(r)} - b_i), \quad i = 1, 2, \dots, n. \quad (4.4.40)$$

Следующее приближение первой неизвестной $x_1^{(r+1)}$ найдем из равенства нулю максимальной невязки, т. е. из уравнения

$$a_{k1}x_1^{(r+1)} + a_{k2}x_2^{(r)} + a_{k3}x_3^{(r)} + \dots + a_{kn}x_n^{(r)} - b_k = 0,$$

где k — номер уравнения с максимальной по модулю невязкой. Затем вычислим невязки при полученной неизвестной $x_1^{(r+1)}$

$$r_i = \frac{1}{a_{ii}} (a_{i1}x_1^{(r+1)} + a_{i2}x_2^{(r)} + a_{i3}x_3^{(r)} + \dots + a_{in}x_n^{(r)} - b_i), \quad i \neq k,$$

и найдем следующее приближение второй неизвестной $x_2^{(r+1)}$ из равенства нулю максимальной невязки

$$a_{m1}x_1^{(r+1)} + a_{m2}x_2^{(r+1)} + a_{m3}x_3^{(r)} + \dots + a_{mn}x_n^{(r)} - b_m = 0,$$

где m — номер уравнения с максимальной по модулю невязкой. Продолжим такие действия и дальше пока не используем все n уравнений, из которых уточним все n неизвестных величин. После этого аналогичным образом найдем следующее приближение решения системы. Итерационный процесс заканчивается, когда на очередной итерации погрешности всех неизвестных не станут меньше заданной величины $\epsilon > 0$. Начальные приближения $x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}$ могут быть выбраны произвольным образом. Критерии сходимости метода релаксации аналогичны критериям сходимости методов Якоби и Зейделя. Метод релаксации в отличие от методов Якоби и Зейделя является нестационарным методом.

Метод невязок. Рассмотрим еще один итерационный метод решения систем линейных алгебраических уравнений. Он относится к итерационным нелинейным методам. В этом методе новое приближение искомых функций $x_1^{(r+1)}, x_2^{(r+1)}, \dots, x_n^{(r+1)}$ связывается с предыдущим приближением искомых функций $x_1^{(r)}, x_2^{(r)}, \dots, x_n^{(r)}$ рекуррентным соотношением

$$x_i^{(r+1)} = x_i^{(r)} - \lambda^{(r)} p_i^{(r)}, \quad i = 1, 2, \dots, n, \quad (4.4.41)$$

где $p_i^{(r)}$ — некоторые коррекции неизвестных, $\lambda^{(r)}$ — скалярный множитель. В матричной записи с использованием обозначений $\mathbf{x}^{(r)} = [x_1^{(r)} \ x_2^{(r)} \ \dots \ x_n^{(r)}]^\top$ и $\mathbf{p}^{(r)} = [p_1^{(r)} \ p_2^{(r)} \ \dots \ p_n^{(r)}]^\top$ для неизвестных и коррекцией на r -й итерации соотношение

(4.4.41) имеет вид

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} - \lambda^{(r)} \mathbf{p}^{(r)}. \quad (4.4.42)$$

Рассмотрим матрицу-столбец невязок уравнений системы (4.4.1) $\mathbf{q}^{(r)} = \mathbf{A} \cdot \mathbf{x}^{(r)} - \mathbf{b}$, с компонентами $q_i^{(r)} = a_{i1}x_1^{(r)} + a_{i2}x_2^{(r)} + a_{i3}x_3^{(r)} + \dots + a_{in}x_n^{(r)} - b_i$. Невязки уравнений системы на соседних итерациях связаны соотношением

$$\mathbf{q}^{(r+1)} = \mathbf{A} \cdot \mathbf{x}^{(r+1)} - \mathbf{b} = \mathbf{A} \cdot \mathbf{x}^{(r)} - \lambda^{(r)} \mathbf{A} \cdot \mathbf{p}^{(r)} - \mathbf{b} = \mathbf{q}^{(r)} - \lambda^{(r)} \mathbf{A} \cdot \mathbf{p}^{(r)}. \quad (4.4.43)$$

Будем рассматривать неизвестные $\mathbf{x}^{(r)}$, коррекции $\mathbf{p}^{(r)}$ и невязки уравнений $\mathbf{q}^{(r)}$ как векторы n -мерного пространства. Скалярный множитель $\lambda^{(r)}$ найдем из условия минимума длины вектора невязки (4.4.43). Для этого умножим n -мерный вектор $\mathbf{q}^{(r+1)}$ скалярно на себя, продифференцируем это произведение по $\lambda^{(r)}$ и приравняем производную нулю, в результате получим

$$\mathbf{q}^{(r)\top} \cdot \mathbf{A} \cdot \mathbf{p}^{(r)} - \lambda^{(r)} \mathbf{p}^{(r)\top} \cdot \mathbf{A}^\top \cdot \mathbf{A} \cdot \mathbf{p}^{(r)} = 0.$$

Отсюда

$$\lambda^{(r)} = \frac{\mathbf{q}^{(r)\top} \cdot \mathbf{A} \cdot \mathbf{p}^{(r)}}{\mathbf{p}^{(r)\top} \cdot \mathbf{A}^\top \cdot \mathbf{A} \cdot \mathbf{p}^{(r)}}.$$

Выбор вектора коррекции $\mathbf{p}^{(r)}$ определяет ту или иную расчетную схему. Если положить $\mathbf{p}^{(r)} = \mathbf{q}^{(r)}$, то тогда

$$\lambda^{(r)} = \frac{\mathbf{q}^{(r)\top} \cdot \mathbf{A} \cdot \mathbf{q}^{(r)}}{\mathbf{q}^{(r)\top} \cdot \mathbf{A}^\top \cdot \mathbf{A} \cdot \mathbf{q}^{(r)}}$$

и мы получим *метод минимальных невязок*. Метод минимальных невязок сходится к точному решению примерно со скоростью метода Якоби. Если начальное приближение находится далеко от искомого решения, то метод минимальных невязок на начальной стадии может сходиться очень медленно.

4.5. Проекция точки на линию

Проекция точки на прямую линию. Проекция точки на прямую линию находится достаточно просто и при выполнении некоторых операций нулевое приближение вычисляется как проекция точки на касательную прямую. Рассмотрим этот частный случай общей задачи.

Пусть дана прямая

$$\mathbf{l}(t) = \mathbf{c}_0 + t\mathbf{w} \quad (4.5.1)$$

и точка \mathbf{p} . Будем считать, что вектор прямой \mathbf{w} имеет произвольную длину. Прямая линия проходит через точку \mathbf{c}_0 , в которой параметр t равен нулю, и имеет направление вектора \mathbf{w} . Требуется найти проекцию точки \mathbf{p} на прямую линию $\mathbf{l}(t)$. Эта задача имеет единственное решение. Построим вектор из точки прямой \mathbf{c}_0 в точку \mathbf{p} и вычислим скалярное произведение этого вектора и вектора прямой \mathbf{w} . На рис. 4.5.1 показаны направляющий вектор прямой \mathbf{w} , ее начальная точка \mathbf{c}_0 и проекция \mathbf{p}_l заданной точки. Если разделим это скалярное произведение на длину вектора \mathbf{w} , то получим длину проекции вектора $\mathbf{p} - \mathbf{c}_0$ на прямую линию. Если же разделим это скалярное произведение

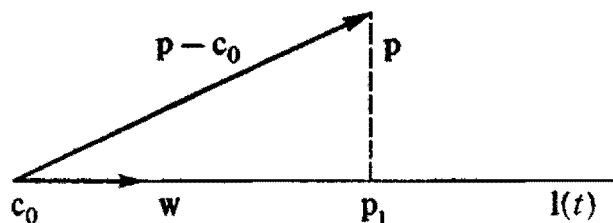


Рис. 4.5.1. Проекция точки \mathbf{p} на прямую линию $\mathbf{l}(t)$

скалярное произведение на длину вектора \mathbf{w} , то получим длину проекции вектора $\mathbf{p} - \mathbf{c}_0$ на прямую линию. Если же разделим это скалярное произведение

на квадрат длины вектора \mathbf{w} , то получим длину проекции вектора $\mathbf{p} - \mathbf{c}_0$ на прямую в единицах длины вектора \mathbf{w} , т. е. получим параметр t для проекции точки \mathbf{p} на прямую линию.

Таким образом, параметр t_p проекции точки \mathbf{p} на прямую линию и радиус-вектор проекции \mathbf{p}_l вычисляются по формулам

$$t_p = \frac{(\mathbf{p} - \mathbf{c}_0) \cdot \mathbf{w}}{|\mathbf{w}|^2}, \quad (4.5.2)$$

$$\mathbf{p}_l = \mathbf{l}(t_p) = \mathbf{c}_0 + t_p \mathbf{w}. \quad (4.5.3)$$

Если длина вектора \mathbf{w} равна единице, то в (4.5.2) не требуется выполнять деление на $|\mathbf{w}|^2$. Расстояние от точки до ее проекции на кривую в общем случае вычисляется как длина вектора $\mathbf{p}_l - \mathbf{p}$. Расстояние от точки до ее проекции на прямую линию можно определить, не вычисляя проекцию точки, а воспользовавшись формулой

$$l_p = \frac{|(\mathbf{p} - \mathbf{c}_0) \times \mathbf{w}|}{|\mathbf{w}|}. \quad (4.5.4)$$

Частные случаи. Проекция точки на аналитические кривые также может быть найдена без привлечения численных методов. Например, чтобы найти проекцию точки на коническое сечение, нужно перевести проектируемую точку в местную систему координат конического сечения, спроецировать эту точку на плоскость конического сечения и найти параметр двухмерной проекции заданной точки.

Общий случай. Пусть требуется найти все проекции точки \mathbf{p} на кривую линию $\mathbf{c}(t)$. Каждая искомая точка кривой $\mathbf{c}(t)$ удовлетворяет уравнению

$$(\mathbf{p} - \mathbf{c}(t)) \cdot \frac{d\mathbf{c}}{dt} = 0. \quad (4.5.5)$$

Это уравнение содержит одну неизвестную величину — параметр t . Как было уже сказано, решение этой задачи разобьем на два этапа. На первом этапе определим нулевые приближения параметров проекций точки на кривую, а на втором этапе найдем точные значения параметров кривой, определяющие проекции заданной точки \mathbf{p} на кривую линию $\mathbf{c}(t)$.

Здесь и в дальнейшем будем считать заданным значение предельного угла отклонения $\Delta\alpha$ кривой. С шагом Δt , определяемым по формуле (4.2.3), пройдем вдоль кривой описанным выше способом и в каждой точке t_i вычислим скалярные произведения векторов

$$a_i = (\mathbf{p} - \mathbf{c}(t_i - \Delta t)) \cdot \frac{d\mathbf{c}}{dt}, \quad b_i = (\mathbf{p} - \mathbf{c}(t_i + \Delta t)) \cdot \frac{d\mathbf{c}}{dt}. \quad (4.5.6)$$

Если искомое решение лежит на параметрическом расстоянии Δt от параметра t_i , то a_i и b_i будут иметь разные знаки. Смена знака скалярных произведений (4.5.6) и является тем событием, которое сигнализирует о том, что рядом находится искомое решение. Примем за нулевое приближение решения параметр $t^{(0)} = t_i$ и одним из методов решения нелинейных уравнений найдем решение с заданной точностью. Например, с помощью метода Ньютона $(r+1)$ -е

приближение параметра проекции точки на кривую вычисляется по формуле

$$t^{(r+1)} = t^{(r)} + \frac{(c - p) \cdot (dc/dt)}{(p - c) \cdot (d^2c/dt^2) - (dc/dt) \cdot (dc/dt)} \Big|_{t^{(r)}}. \quad (4.5.7)$$

Процесс уточнения параметра закончим, когда на очередной итерации $|t^{(r+1)} - t^{(r)}|$ станет меньше заданной величины ε . Таким же образом найдем все остальные корни уравнения (4.5.5).

4.6. Проекция точки на поверхность

Проекция точки на плоскость. Проекция точки на плоскость является частным случаем общей задачи нахождения проекции точки на поверхность. В силу простоты вычисления проекции точки на касательную к поверхности плоскость используется в качестве нулевого приближения при решении общей задачи.

Рассмотрим задачу проецирования точки p на плоскость, заданную радиус-вектором

$$s(u, v) = s_0 + us_1 + vs_2. \quad (4.6.1)$$

Будем считать, что векторы s_1 и s_2 не коллинеарные. Допустим, что в общем случае векторы s_1 и s_2 не ортогональны и имеют не единичную длину. Плоскость проходит через точку s_0 , в которой параметры равны нулю, а векторы s_1 и s_2 определяют параметрические направления. Заданная точка p имеет единственную проекцию на плоскость (4.6.1).

Построим единичную нормаль к плоскости

$$m = \frac{s_1 \times s_2}{|s_1||s_2|}. \quad (4.6.2)$$

Вычислим радиус-вектор p_s проекции точки на плоскость как разность радиус-вектора проецируемой точки и составляющей вектора $p - s_0$, параллельной нормали к плоскости,

$$p_s = p - m((p - s_0) \cdot m). \quad (4.6.3)$$

Рис. 4.6.1. Проекция p_s точки p на плоскость $s(u, v)$

Далее найдем параметры u_p и v_p , соответствующие проекции точки на плоскость. Для вычисления параметров проекции точки найдем длины проекций вектора $p - s_0$ на векторы s_1 и s_2 , которые обозначим через p_1 и p_2 :

$$p_1 = \frac{(p - s_0) \cdot s_1}{|s_1|}, \quad (4.6.4)$$

$$p_2 = \frac{(p - s_0) \cdot s_2}{|s_2|}. \quad (4.6.5)$$

На рис. 4.6.1 показаны векторы плоскости s_1 и s_2 , ее начальная точка s_0 и проекция p_s заданной точки.

Параметры u_p и v_p и длины проекций p_1 и p_2 связаны уравнениями

$$p_1 = u_p |s_1| + v_p |s_2| \cos \varphi, \quad p_2 = v_p |s_2| + u_p |s_1| \cos \varphi,$$

где $\cos \varphi = (\mathbf{s}_1 \cdot \mathbf{s}_2) / (|\mathbf{s}_1| |\mathbf{s}_2|)$ — косинус угла между векторами \mathbf{s}_1 и \mathbf{s}_2 определяется по формуле (1.7.13). Из системы этих уравнений найдем параметры проекции точки на плоскость

$$u_p = \frac{p_1 - p_2 \cos \varphi}{(1 - \cos^2 \varphi) |\mathbf{s}_1|} = \frac{g_{22} \mathbf{s}_1 \cdot (\mathbf{p}_s - \mathbf{s}_0) - g_{21} \mathbf{s}_2 \cdot (\mathbf{p}_s - \mathbf{s}_0)}{g_{11} g_{22} - g_{21}^2} = \\ = g^{11} \mathbf{s}_1 \cdot (\mathbf{p}_s - \mathbf{s}_0) + g^{12} \mathbf{s}_2 \cdot (\mathbf{p}_s - \mathbf{s}_0), \quad (4.6.6)$$

$$v_p = \frac{p_2 - p_1 \cos \varphi}{(1 - \cos^2 \varphi) |\mathbf{s}_2|} = \frac{g_{11} \mathbf{s}_2 \cdot (\mathbf{p}_s - \mathbf{s}_0) - g_{12} \mathbf{s}_1 \cdot (\mathbf{p}_s - \mathbf{s}_0)}{g_{11} g_{22} - g_{21}^2} = \\ = g^{21} \mathbf{s}_1 \cdot (\mathbf{p}_s - \mathbf{s}_0) + g^{22} \mathbf{s}_2 \cdot (\mathbf{p}_s - \mathbf{s}_0), \quad (4.6.7)$$

где $g_{11} = \mathbf{s}_1 \cdot \mathbf{s}_1$, $g_{22} = \mathbf{s}_2 \cdot \mathbf{s}_2$, $g_{12} = \mathbf{s}_1 \cdot \mathbf{s}_2$ — коэффициенты первой основной квадратичной формы плоскости (1.7.8), они же ковариантные компоненты метрического тензора поверхности, g^{11} , g^{22} , g^{12} , g^{21} — контравариантные компоненты метрического тензора поверхности. Если векторы \mathbf{s}_1 и \mathbf{s}_2 ортогональные, то $g_{12} = g^{21} = 0$ и формулы (4.6.6) и (4.6.7) примут вид

$$u_p = \frac{(\mathbf{p} - \mathbf{s}_0) \cdot \mathbf{s}_1}{\mathbf{s}_1 \cdot \mathbf{s}_1}, \quad v_p = \frac{(\mathbf{p} - \mathbf{s}_0) \cdot \mathbf{s}_2}{\mathbf{s}_2 \cdot \mathbf{s}_2}.$$

Расстояние от точки до ее проекции на плоскость в общем случае вычисляется как длина вектора $\mathbf{p}_s - \mathbf{p}$. Расстояние от точки до ее проекции на плоскость можно определить, не вычисляя проекцию точки, а вычислив проекцию вектора $\mathbf{p} - \mathbf{s}_0$ на нормаль к плоскости

$$l_s = (\mathbf{p} - \mathbf{s}_0) \cdot \mathbf{m}. \quad (4.6.8)$$

Частные случаи. Проекция точки на некоторые аналитические поверхности могут быть найдены без привлечения численных методов. Например, чтобы найти проекции точки на поверхность кругового цилиндра, конуса, сферы или тора, нужно перевести проецируемую точку в местную систему координат поверхности, где легко найти параметры проекций. Аналогично могут быть найдены проекции на поверхности выдавливания и вращения. В некоторых частных случаях положения проецируемой точки ее проекции могут быть легко найдены и на другие поверхности.

Общий случай. Рассмотрим задачу проецирования точки на поверхность в общем случае. Пусть требуется найти все проекции точки \mathbf{p} на поверхность $\mathbf{s}(u, v)$. Каждая искомая точка поверхности $\mathbf{s}(u, v)$ удовлетворяет системе двух уравнений

$$(\mathbf{p} - \mathbf{s}(u, v)) \cdot \frac{\partial \mathbf{s}}{\partial u} = 0, \quad (\mathbf{p} - \mathbf{s}(u, v)) \cdot \frac{\partial \mathbf{s}}{\partial v} = 0. \quad (4.6.9)$$

Система уравнений (4.6.9) содержит две неизвестные величины — параметры u и v . Эта задача решается так же, как и задача нахождения проекций заданной точки на кривую. На первом этапе определим нулевые приближения параметров

поверхности для проекций точки, а на втором этапе найдем точные значения параметров, определяющие проекции заданной точки \mathbf{p} на поверхность $\mathbf{s}(u, v)$.

Пройдем по поверхности с шагами Δu и Δv , вычисляемыми по формулам (4.2.4) и (4.2.5), описанным выше способом движения по параметрической области. Обозначим параметры точек, через которые мы пройдем, через u_i и v_j . В каждой точке будем вычислять скалярные произведения векторов

$$a_{ij} = (\mathbf{p} - \mathbf{s}(u_i - \Delta u, v_j)) \cdot \frac{\partial \mathbf{s}}{\partial u}, \quad b_{ij} = (\mathbf{p} - \mathbf{s}(u_i + \Delta u, v_j)) \cdot \frac{\partial \mathbf{s}}{\partial u}, \quad (4.6.10)$$

$$c_{ij} = (\mathbf{p} - \mathbf{s}(u_i, v_j - \Delta v)) \cdot \frac{\partial \mathbf{s}}{\partial v}, \quad d_{ij} = (\mathbf{p} - \mathbf{s}(u_i, v_j + \Delta v)) \cdot \frac{\partial \mathbf{s}}{\partial v}. \quad (4.6.11)$$

Если искомое решение лежит вблизи точки с параметрами u_i, v_j , то a_{ij} и b_{ij} будут иметь разные знаки, а также c_{ij} и d_{ij} будут иметь разные знаки. Смена знаков скалярных произведений ($a_{ij}b_{ij} \leq 0$ и $c_{ij}d_{ij} \leq 0$) говорит о том, что рядом находится искомое решение. За нулевое приближение параметров примем значения $u^{(0)} = u_i, v^{(0)} = v_j$. Начиная с нулевого приближения параметров, одним из методов решения нелинейных уравнений найдем решение задачи с заданной точностью. Например, в методе Ньютона на r -й итерации приращения параметров $\Delta u^{(r)}$ и $\Delta v^{(r)}$ проекции найдутся из системы линейных уравнений

$$((\mathbf{p} - \mathbf{s}) \cdot \mathbf{s}_{11} - \mathbf{s}_1 \cdot \mathbf{s}_1) \Delta u^{(r)} + ((\mathbf{p} - \mathbf{s}) \cdot \mathbf{s}_{12} - \mathbf{s}_2 \cdot \mathbf{s}_1) \Delta v^{(r)} = (\mathbf{s} - \mathbf{p}) \cdot \mathbf{s}_1,$$

$$((\mathbf{p} - \mathbf{s}) \cdot \mathbf{s}_{21} - \mathbf{s}_1 \cdot \mathbf{s}_2) \Delta u^{(r)} + ((\mathbf{p} - \mathbf{s}) \cdot \mathbf{s}_{22} - \mathbf{s}_2 \cdot \mathbf{s}_2) \Delta v^{(r)} = (\mathbf{s} - \mathbf{p}) \cdot \mathbf{s}_2,$$

где $\mathbf{s}_1 = \frac{\partial \mathbf{s}}{\partial u}$, $\mathbf{s}_2 = \frac{\partial \mathbf{s}}{\partial v}$, $\mathbf{s}_{11} = \frac{\partial^2 \mathbf{s}}{\partial u^2}$, $\mathbf{s}_{12} = \mathbf{s}_{21} = \frac{\partial^2 \mathbf{s}}{\partial u \partial v}$, $\mathbf{s}_{22} = \frac{\partial^2 \mathbf{s}}{\partial v^2}$ — частные производные радиус-вектора по параметрам. Следующее приближение параметров проекции точки равны $u^{(r+1)} = u^{(r)} + \Delta u^{(r)}$, $v^{(r+1)} = v^{(r)} + \Delta v^{(r)}$. Процесс уточнения параметров закончим, когда на очередной итерации выполняются неравенства $|u^{(r+1)} - u^{(r)}| < \varepsilon$ и $|v^{(r+1)} - v^{(r)}| < \varepsilon$, где ε — заданная погрешность. Таким же образом найдем все остальные корни системы уравнений (4.6.9).

Если требуется найти только ближайшую проекцию заданной точки на поверхность, то можно пройти по тем же точкам геометрического объекта и выбрать из них ближайшую к заданной точке. Параметры ближайшей точки и следует выбрать в качестве нулевого приближения решения задачи.

Проекция точки на поверхность в заданном направлении. В определенных случаях возникает задача определения проекции точки \mathbf{p} на поверхность не по нормали к ней, а вдоль заданного направления. Пусть направление проецирования задано вектором единичной длины \mathbf{q} . Построим прямую линию

$$\mathbf{l}(t) = \mathbf{p} + t\mathbf{q}, \quad (4.6.12)$$

проходящую через заданную точку и имеющую направление заданного вектора. Проекция точки на поверхность в заданном направлении определим как точки пересечения поверхности с прямой (4.6.12), проходящей через заданную точку в заданном направлении.

4.7. Определение точек пересечения линии и поверхности

Простейшим представителем кривых является прямая линия, а простейшим представителем поверхности является плоскость. Точка их пересечения может быть найдена аналитически. Для определения нулевого приближения каждой точки пересечения кривой и поверхности будет решаться задача пересечения касательной линии кривой и касательной плоскости поверхности, поэтому сначала рассмотрим задачу определения точки пересечения прямой и плоскости.

Пересечение прямой и плоскости. Пусть радиус-векторы прямой линии и плоскости описываются зависимостями

$$\mathbf{c}(z) = \mathbf{c}_0 + z\mathbf{c}', \quad (4.7.1)$$

$$\mathbf{s}(x, y) = \mathbf{s}_0 + x\mathbf{s}_1 + y\mathbf{s}_2. \quad (4.7.2)$$

Найдем их точку пересечения. Такая точка существует и является единственной, если прямая не параллельна плоскости, т. е. если $\mathbf{c}' \cdot (\mathbf{s}_1 \times \mathbf{s}_2) \neq 0$. Построим единичную нормаль \mathbf{m} к плоскости по формуле (4.6.2) и вычислим значение параметра прямой, соответствующее точке пересечения плоскости и прямой

$$z_0 = \frac{\mathbf{m} \cdot (\mathbf{s}_0 - \mathbf{c}_0)}{\mathbf{m} \cdot \mathbf{c}'}. \quad (4.7.3)$$

На рис. 4.7.1 показаны прямая и плоскость при взгляде параллельно плоскости. В числителе правой части (4.7.3) стоит расстояние от начальной точки \mathbf{c}_0

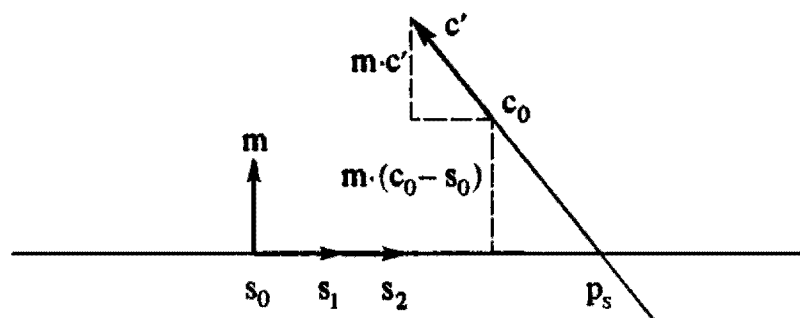


Рис. 4.7.1. Пересечение прямой и плоскости

прямой до плоскости, а в знаменателе стоит нормальная к плоскости составляющая производной прямой: отношение этих величин даст параметр пересечения прямой с плоскостью. Подставим (4.7.3) в (4.7.1) и получим радиус-вектор \mathbf{p}_s точки пересечения плоскости и прямой

$$\mathbf{p}_s = \mathbf{c}_0 + \mathbf{c}' \frac{\mathbf{m} \cdot (\mathbf{s}_0 - \mathbf{c}_0)}{\mathbf{m} \cdot \mathbf{c}'}. \quad (4.7.4)$$

Далее, по формулам (4.6.6) и (4.6.7) вычислим параметры плоскости x_0 и y_0 , соответствующие точке пересечения

$$x_0 = g^{11}\mathbf{s}_1 \cdot (\mathbf{p}_s - \mathbf{s}_0) + g^{12}\mathbf{s}_2 \cdot (\mathbf{p}_s - \mathbf{s}_0), \quad (4.7.5)$$

$$y_0 = g^{21}\mathbf{s}_1 \cdot (\mathbf{p}_s - \mathbf{s}_0) + g^{22}\mathbf{s}_2 \cdot (\mathbf{p}_s - \mathbf{s}_0). \quad (4.7.6)$$

Общий случай. Пусть требуется найти все точки пересечения поверхности, радиус-вектор которой описывается векторной функцией $\mathbf{s}(u, v)$, и кривой, радиус-вектор которой описывается векторной функцией $\mathbf{c}(t)$. Для каждой точки

пересечения должно удовлетворяться векторное уравнение

$$\mathbf{s}(u, v) - \mathbf{c}(t) = 0, \quad (4.7.7)$$

которое, будучи записанным по координатам, представляет собой систему трех скалярных уравнений

$$s_1(u, v) - c_1(t) = 0, \quad s_2(u, v) - c_2(t) = 0, \quad s_3(u, v) - c_3(t) = 0,$$

где $\mathbf{s}(u, v) = [s_1(u, v) \ s_2(u, v) \ s_3(u, v)]^T$, $\mathbf{c}(t) = [c_1(t) \ c_2(t) \ c_3(t)]^T$. В качестве неизвестных выступают три параметра — u , v и t . Решив эту систему уравнений, получим параметры u , v и t , по которым при необходимости можно вычислить точки на поверхности и на кривой. Система уравнений в общем случае является нелинейной и решается итерационными методами. Точек пересечения может быть несколько и от того, с какого приближения параметров начнется итерационный процесс решения, зависит то, к какой тройке параметров сойдется решение.

Для нахождения нулевого приближения каждого решения будем двигаться описанным выше способом по поверхности и кривой, т. е. будем переходить от точки к точке с вычисляемыми по формулам (4.2.4) и (4.2.5) шагами Δu и Δv поверхности и вычисляемым по формуле (4.2.3) шагом Δt кривой. При каждом сочетании точек выполним следующую проверку. Пусть в некоторой точке поверхности значения параметров равны u_0 и v_0 , а в некоторой точке кривой значение параметра равно t_0 . Построим касательную плоскость (4.7.2) в данной точке поверхности и касательную прямую (4.7.1) в данной точке кривой. Начальными точками плоскости и прямой являются точки $\mathbf{s}_0 = \mathbf{s}(u_0, v_0)$ и $\mathbf{c}_0 = \mathbf{c}(t_0)$. Векторами плоскости \mathbf{s}_1 и \mathbf{s}_2 являются частные производные радиус-вектора поверхности по ее параметрам $\mathbf{s}_1 = \partial \mathbf{s} / \partial u$, $\mathbf{s}_2 = \partial \mathbf{s} / \partial v$ в данной точке. Вектором \mathbf{c}' прямой является производная радиус-вектора кривой по ее параметру $\mathbf{c}' = d\mathbf{c} / dt$ в данной точке. По формулам (4.7.3)–(4.7.6) найдем параметры точки пересечения касательной плоскости и касательной прямой линии.

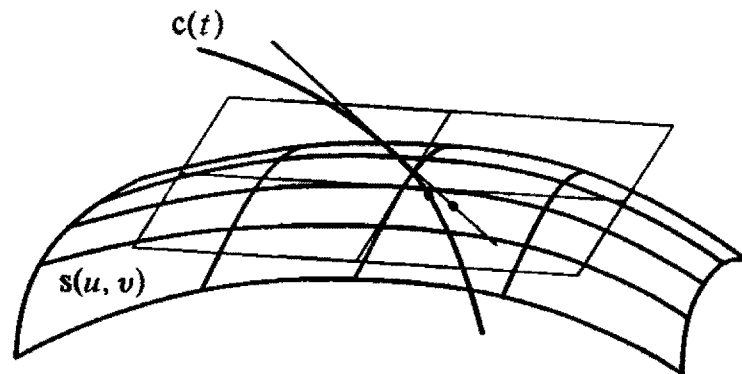


Рис. 4.7.2. Поиск пересечений кривой и поверхности начинается с пересечения касательной линии кривой и касательной плоскости поверхности

Точка пересечения может быть найдена, если выбранные точки поверхности и кривой являются обыкновенными и если касательная прямая не параллельна касательной плоскости. Вблизи точки пересечения параметры x_0 и y_0 приблизительно соответствуют приращениям параметров поверхности, а параметр z_0 приблизительно соответствует приращению параметра кривой. На рис. 4.7.2 показаны поверхность, касательная к ней плоскость, кривая и касательная к ней линия.

Далее предстоит решить, можно ли принять значения параметров $u_0 + x_0$, $v_0 + y_0$, $t_0 + z_0$ в качестве нулевого приближения при решении системы уравнений или нельзя. Для этого нужно знать, каких размеров часть поверхности можно приближенно аппроксимировать частью касательной к ней плоскостью и какого размера участок кривой можно аппроксимировать частью касательной к кривой линией так, чтобы при этом не перешагнуть через соседнюю точку пересечения. Эти размеры определяются шагами параметров Δu , Δv и Δt . Если будут выполнены условия

$$-\Delta u \leq x_0 \leq \Delta u, \quad -\Delta v \leq y_0 \leq \Delta v, \quad -\Delta t \leq z_0 \leq \Delta t,$$

то параметры

$$u^{(0)} = u_0 + x_0, \quad v^{(0)} = v_0 + y_0, \quad t^{(0)} = t_0 + z_0 \quad (4.7.8)$$

могут быть приняты в качестве начального приближения при решении системы уравнений (4.7.7) определения точки пересечения поверхности $\mathbf{s}(u, v)$ и кривой линии $\mathbf{c}(t)$.

Далее, начиная с нулевого приближения параметров, одним из методов решения нелинейных уравнений найдем решение с заданной точностью. Например, в методе Ньютона на r -й итерации приращения параметров $\Delta u^{(r)}$, $\Delta v^{(r)}$ и $\Delta t^{(r)}$ точки пересечения найдутся из системы линейных уравнений

$$\begin{aligned} \frac{\partial s_1}{\partial u} \Delta u^{(r)} + \frac{\partial s_1}{\partial v} \Delta v^{(r)} - \frac{\partial c_1}{\partial t} \Delta t^{(r)} &= c_1 - s_1, \\ \frac{\partial s_2}{\partial u} \Delta u^{(r)} + \frac{\partial s_2}{\partial v} \Delta v^{(r)} - \frac{\partial c_2}{\partial t} \Delta t^{(r)} &= c_2 - s_2, \\ \frac{\partial s_3}{\partial u} \Delta u^{(r)} + \frac{\partial s_3}{\partial v} \Delta v^{(r)} - \frac{\partial c_3}{\partial t} \Delta t^{(r)} &= c_3 - s_3, \end{aligned}$$

где s_1, s_2, s_3 — компоненты радиус-вектора поверхности \mathbf{s} , c_1, c_2, c_3 — компоненты радиус-вектора кривой \mathbf{c} . Следующее приближение параметров точки пересечения равны $u^{(r+1)} = u^{(r)} + \Delta u^{(r)}$, $v^{(r+1)} = v^{(r)} + \Delta v^{(r)}$, $t^{(r+1)} = t^{(r)} + \Delta t^{(r)}$. Процесс уточнения параметров закончим, когда на очередной итерации выполняются неравенства $|u^{(r+1)} - u^{(r)}| < \varepsilon$, $|v^{(r+1)} - v^{(r)}| < \varepsilon$ и $|t^{(r+1)} - t^{(r)}| < \varepsilon$, где ε — заданная погрешность.

После этого продолжим поиск других точек пересечения. Сделаем шаг Δu или Δv по одному из параметров поверхности или шаг Δt по параметру кривой и повторим аппроксимацию поверхности касательной плоскостью, аппроксимацию кривой касательной прямой, поиск пересечения плоскости и прямой, нахождение приближенных параметров пересечения и принятие решения, но уже для других точек кривой и поверхности. Так мы должны дискретно пройти всю поверхность и всю кривую в поиске нулевых приближений параметров точек пересечения. Возможны случаи, когда, начав итерационный процесс с различными нулевыми приближениями, мы придем к одному и тому же решению. Поэтому все полученные решения необходимо проверить на совпадение. Решения u^*, v^*, t^* и u^{**}, v^{**}, t^{**} будем считать одинаковыми, если выполняются одновременно три неравенства: $|u^* - u^{**}| < \xi$, $|v^* - v^{**}| < \xi$, $|t^* - t^{**}| < \xi$, где $\xi > \varepsilon$. О величине ξ мы расскажем в конце главы.

Если в некоторой точке кривая касается поверхности, то определитель матрицы системы уравнений (4.7.7) на некоторой итерации станет недопустимо малым или равным нулю и мы не сможем из этой системы найти точку касания. В этом случае вместо векторного уравнения (4.7.7) можно решать систему трех скалярных уравнений

$$\begin{aligned} (\mathbf{s}(u, v) - \mathbf{c}(t)) \cdot \frac{\partial \mathbf{s}}{\partial u} &= 0, \\ (\mathbf{s}(u, v) - \mathbf{c}(t)) \cdot \frac{\partial \mathbf{s}}{\partial v} &= 0, \\ (\mathbf{s}(u, v) - \mathbf{c}(t)) \cdot \frac{\partial \mathbf{c}}{\partial t} &= 0, \end{aligned} \quad (4.7.9)$$

определитель которой не равен нулю в точке касания. Но системе (4.7.9) удовлетворяют не только точки касания, а еще и другие точки, в которых касательная прямая параллельна касательной плоскости и отрезок, соединяющий точки касания, ортогонален касательной прямой и касательной плоскости. Поэтому после решения системы уравнений (4.7.9) следует проверить, являются ли полученные точки являются точками касания.

Итерационный процесс решения системы (4.7.9) закончим, когда на очередной итерации приращение каждого из параметров станет меньше некоторой заданной величины ε . Предположим, что максимальное отклонение полученных в результате решения параметров от их точных решений u^* , v^* , t^* составляет $\pm\varepsilon$, тогда максимальное отклонение радиус-вектора полученной точки пересечения от его истинного значения будет равно произведению максимальной по модулю производной радиус-вектора объекта на погрешность параметра

$$\Delta\rho = \varepsilon \max(|s_1|, |s_2|, |c'|). \quad (4.7.10)$$

Величина $\Delta\rho$ есть радиус сферы погрешности: если ее центром сделать полученную в результате решения точку кривой или поверхности, то любая другая точка геометрического объекта внутри сферы будет считаться неразличимой с центральной точкой сферы. Значение $\Delta\rho$ примем за критерий для ответа на вопрос, является ли найденное решение системы уравнений (4.7.9) точкой касания или нет. Если расстояние между найденными точками на кривой и поверхности меньше $\Delta\rho$, то решение дало точку касания, в противном случае решение не является ни точкой касания, ни точкой пересечения и должно быть опущено. Для различных геометрических объектов и даже различных точек одного и того же объекта радиусы сфер погрешности будут разными в разных точках, тем не менее, они удобны для оценки точности решения. Нам не требуется знать истинные размеры объектов и масштабы расстояний между ними, требуется знать только погрешность определения параметров при выполнении операции, а радиус сферы погрешности в любой точке объекта мы всегда найдем. Данная оценка близости точек не зависит от размеров геометрических объектов, так как размеры области определения параметров не зависят от реальных размеров геометрического объекта. В дальнейшем будет важно знать, с какой точностью решена та или иная задача.

Задача определения пересечения поверхности и прямой линии, а также приводимая к ней задача проекции точки на поверхность в заданном направлении, решаются описанным методом при неподвижной пробной точке на прямой линии.

4.8. Определение точек пересечения линий

Сначала рассмотрим задачу определения точек пересечения двумерных кривых, а далее перейдем к задаче пересечения пространственных кривых. При решении задачи пересечения двумерных кривых для определения нулевого приближения будет решаться задача определения точек пересечения прямых линий.

Пересечение двумерных прямых линий. Точка пересечения двумерных прямых линий может быть найдена аналитически. Пусть даны двумерные прямые

$$l_1(w_1) = p_1 + w_1 c_1', \quad (4.8.1)$$

$$l_2(w_2) = p_2 + w_2 c_2'. \quad (4.8.2)$$

Прямые имеют точку пересечения, если их направляющие векторы не коллинеарны. Через n_1 и n_2 обозначим единичные нормали к прямым $l_1(w_1)$ и $l_2(w_2)$. Если касательный вектор $c' = [x \ y]^T$ кривой имеет координаты x и y , то ортогональный ему вектор n имеет координаты $-ky$ и kx , где $k = \frac{1}{\sqrt{x^2 + y^2}}$. На

рис. 4.8.1 показаны прямые (4.8.1) и (4.8.2), нормали к ним и точка пересечения p . Из рисунка видно, что, например, параметр второй прямой для точки пересечения может быть найден как отношение проекции вектора $p_1 - p_2$ на нормаль первой прямой к проекции вектора c_2' на ту же нормаль.

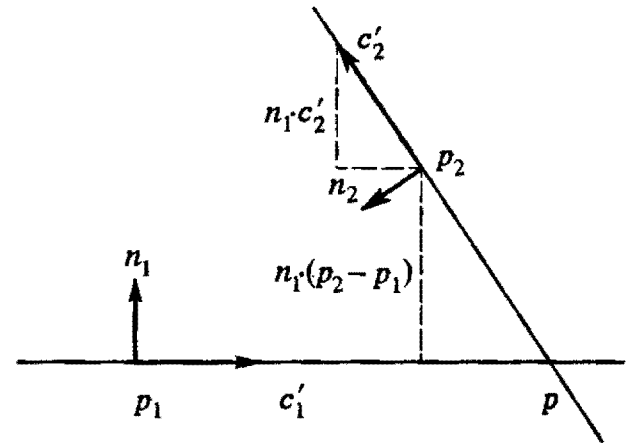


Рис. 4.8.1. Пересечение прямых

Аналогично найдем параметры точки пересечения для первой прямой:

$$w_1 = \frac{n_2 \cdot (p_2 - p_1)}{n_2 \cdot c_1'}, \quad (4.8.3)$$

$$w_2 = \frac{n_1 \cdot (p_1 - p_2)}{n_1 \cdot c_2'}. \quad (4.8.4)$$

Радиус-вектор точки пересечения найдем, подставив в (4.8.1) или (4.8.2) соответствующий параметр,

$$p = p_1 + \frac{n_2 \cdot (p_2 - p_1)}{n_2 \cdot c_1'} c_1' = p_2 + \frac{n_1 \cdot (p_1 - p_2)}{n_1 \cdot c_2'} c_2'. \quad (4.8.5)$$

Пересечение двумерных отрезков прямых. Аналитически можно найти точку пересечения отрезков прямых. Пусть даны два отрезка прямых

$$l_1(t_1) = (1 - t_1)p_1 + t_1q_1, \quad l_2(t_2) = (1 - t_2)p_2 + t_2q_2,$$

построенных по точкам $p_1 = [x_1 \ y_1]^T$, $p_2 = [x_2 \ y_2]^T$, $q_1 = [a_1 \ b_1]^T$ и $q_2 = [a_2 \ b_2]^T$. Параметры их точки пересечения найдем из векторного уравнения

$$(1 - t_1)p_1 + t_1q_1 = (1 - t_2)p_2 + t_2q_2,$$

которое представляет собой систему двух скалярных уравнений

$$\begin{aligned}(x_1 - a_1)t_1 - (x_2 - a_2)t_2 &= x_1 - x_2, \\ (y_1 - b_1)t_1 - (y_2 - b_2)t_2 &= y_1 - y_2.\end{aligned}$$

Если $(x_2 - a_2)(y_1 - b_1) - (x_1 - a_1)(y_2 - b_2) \neq 0$, то по формуле Крамера параметры точки пересечения равны

$$\begin{aligned}t_1 &= \frac{(x_2 - a_2)(y_1 - y_2) - (x_1 - x_2)(y_2 - b_2)}{(x_2 - a_2)(y_1 - b_1) - (x_1 - a_1)(y_2 - b_2)}, \\ t_2 &= \frac{(x_1 - x_2)(y_1 - b_1) - (x_1 - a_1)(y_1 - y_2)}{(x_2 - a_2)(y_1 - b_1) - (x_1 - a_1)(y_2 - b_2)}.\end{aligned}$$

Отрезки пересекаются, если $0 \leq t_1 \leq 1$ и $0 \leq t_2 \leq 1$, в противном случае пересекаются продолжения отрезков.

Частные случаи. Аналитически могут быть найдены точки пересечения прямой с кривыми второго порядка, точки пересечения кривых второго порядка между собой и точки пересечения других кривых, для которых известно аналитическое описание. В указанных случаях задача решается в координатном представлении и сводится к вычислению корней многочленов. В общем случае можно найти корни многочлена, степень которого не превосходит 4. Например, задача нахождения точек пересечения двух эллипсов приводит к решению уравнения четвертой степени. Если аналитическая формула кривой неизвестна или степень разрешающего уравнения превосходит 4, то задачу пересечения кривых следует решать численно.

Общий случай пересечения двумерных кривых. Пусть требуется найти все точки пересечения двух двумерных кривых, радиус-векторы которых описываются функциями $c_1(t_1)$ и $c_2(t_2)$. Для каждой точки пересечения должно удовлетворяться векторное уравнение

$$c_1(t_1) - c_2(t_2) = 0, \quad (4.8.6)$$

которое, будучи записанным по координатно, представляет собой систему двух скалярных уравнений

$$x_1(t_1) - x_2(t_2) = 0, \quad y_1(t_1) - y_2(t_2) = 0. \quad (4.8.7)$$

Неизвестными являются параметры кривых t_1 и t_2 . Система уравнений в общем случае является нелинейной и решается итерационными методами. Точек пересечения может быть несколько и от того, с какого приближения параметров начнется итерационный процесс решения, зависит, к какой точке сойдется решение. Итерационный процесс определения параметров точки пересечения закончим, когда на очередной итерации не только параметры кривых удовлетворят заданным условиям, но и координаты точек удовлетворят заданным условиям точности. Процесс уточнения решения закончим, когда выполнятся неравенства $|t_1^{(r+1)} - t_1^{(r)}| < \varepsilon$, $|t_2^{(r+1)} - t_2^{(r)}| < \varepsilon$, $|x_1^{(r+1)} - x_1^{(r)}| < \varepsilon$, $|y_1^{(r+1)} - y_1^{(r)}| < \varepsilon$, $|x_2^{(r+1)} - x_2^{(r)}| < \varepsilon$, $|y_2^{(r+1)} - y_2^{(r)}| < \varepsilon$. Это связано с тем, что координаты двумерных кривых сами служат параметрами поверхностей. Для нахождения нулевого приближения каждой точки пересечения будем

двигаться с вычисляемыми по формуле (4.2.3) шагами Δt_1 и Δt_2 по кривым, в каждом сочетании точек строить касательные линии (4.8.1) и (4.8.2) к кривым и определять параметры w_1 и w_2 точки их пересечения. Для этого в (4.8.1) и (4.8.2) подставим значения: $p_1 = c_1(t_1)$, $p_2 = c_2(t_2)$ — точки на кривых, $c_1' = dc_1/dt_1$, $c_2' = dc_2/dt_2$ — производные кривых в этих точках. Параметры касательных прямых в точке их пересечения определим по формулам (4.8.3) и (4.8.4). Вблизи точки пересечения параметры w_1 и w_2 приблизительно соответствуют приращениям параметров кривых, поэтому величины

$$t_1^{(0)} = t_1 + w_1, \quad t_2^{(0)} = t_2 + w_2 \quad (4.8.8)$$

могут быть приняты в качестве нулевого приближения при решении системы уравнений (4.8.6), если удовлетворяются условия

$$-\Delta t_1 \leq w_1 \leq \Delta t_1, \quad -\Delta t_2 \leq w_2 \leq \Delta t_2.$$

Шаги по параметрам Δt_1 и Δt_2 вычислены исходя из условия, что касательная к каждой кривой при данном шаге отклонится на угол, не превышающий заданную величину $\Delta\alpha$. Шаги определяют, какого размера участок кривой можно аппроксимировать частью касательной к кривой прямой линией так, чтобы при этом не перешагнуть через соседнюю точку пересечения.

После этого продолжим поиск других точек пересечения. Сделаем шаг Δt_1 или Δt_2 по параметру одной из кривых и повторим аппроксимацию кривой касательной прямой, поиск пересечения прямых, нахождение приближенных параметров пересечения и принятие решения о поиске точного решения. Таким образом, мы должны дискретно пройти обе кривые в поиске нулевых приближений параметров точек пересечения.

Полученные решения необходимо проверить на совпадение. Решения t_1^* , t_2^* и t_1^{**} , t_2^{**} будем считать одинаковыми, если выполняются одновременно неравенства: $|t_1^* - t_1^{**}| < \xi$, $|t_2^* - t_2^{**}| < \xi$, где $\xi < \varepsilon$. Величину ξ мы рассмотрим в конце главы.

Если заданные кривые имеют точку касания, то уравнения системы (4.8.7) являются линейно зависимыми и на некоторой итерации определитель системы станет недопустимо малым или равным нулю, и мы не сможем найти точку касания. В этом случае вместо векторного уравнения (4.8.6) можно решать систему скалярных уравнений

$$(c_1(t_1) - c_2(t_2)) \cdot \frac{dc_1}{dt_1} = 0, \quad (c_2(t_2) - c_1(t_1)) \cdot \frac{dc_2}{dt_2} = 0, \quad (4.8.9)$$

опредетель которой не равен нулю в точке касания. Но системе (4.8.9) удовлетворяют не только точки касания, а еще и точки, в которых касательные прямые параллельны друг другу и ортогональны вектору, соединяющему точки, в которых построены касательные прямые. Поэтому после решения системы уравнений (4.8.9) следует проверить тот факт, что полученные точки являются точками касания.

Проверку выполним аналогично тому, как это было сделано при определении точек касания кривой и поверхности. Итерационный процесс решения системы нелинейных уравнений (4.8.9) будет закончен, когда на очередной итерации приращение каждого из параметров станет меньше некоторой заданной величины ε . Предположим, что максимальное отклонение полученных в результате

решения параметров от их точных решений составляет $\pm \varepsilon$, тогда максимальное отклонение радиус-вектора полученной точки пересечения от его истинного значения будет равно произведению максимальной по модулю производной радиус-вектора объекта на погрешность параметра

$$\Delta \rho = \varepsilon \max (|c_1'|, |c_2'|). \quad (4.8.10)$$

Если расстояние между найденными точками кривых и поверхности меньше $\Delta \rho$, то решение дало точку касания, в противном случае решение не является ни точкой касания, ни точкой пересечения и должно быть опущено.

Общий случай пересечения пространственных кривых. С помощью системы уравнений (4.8.9) можно найти точки пересечения или касания пространственных кривых $c_1(t_1)$ и $c_2(t_2)$. Системе уравнений

$$(c_1(t_1) - c_2(t_2)) \cdot \frac{dc_1}{dt_1} = 0, \quad (c_2(t_2) - c_1(t_1)) \cdot \frac{dc_2}{dt_2} = 0 \quad (4.8.11)$$

удовлетворяют точки пересечения пространственных кривых, точки касания пространственных кривых, а также точки, в которых касательные к кривым прямые ортогональны вектору, соединяющему точки построения касательных прямых. Система уравнений (4.8.11) получена из системы уравнений (4.8.9) путем замены в ней двумерных кривых $c_1(t_1)$ и $c_2(t_2)$ на трехмерные кривые $c_1(t_1)$ и $c_2(t_2)$. После решения системы уравнений (4.8.11) следует проверить полученные точки на близость по формуле (4.8.10). Проверкой коллинеарности производных радиус-векторов кривых можно выделить среди полученных точек точки касания. Метод определения близости найденных точек с помощью радиуса погрешности (4.8.10) инвариантен по отношению к геометрическим размерам объектов и к единицам измерения координатной системы.

4.9. Построение линий пересечения поверхностей

Пересечение плоскостей. Линия пересечения двух плоскостей

$$r(u, v) = r_0 + ur_1 + vr_2, \quad s(a, b) = s_0 + as_1 + bs_2,$$

с единичными нормальными векторами r_3 и s_3 находится достаточно просто. Ее направление определяется вектором $z = r_3 \times s_3$. Если длина вектора z отлична от нуля, то плоскости пересекаются, в противном случае они параллельны или совпадают. Если плоскости пересекаются, то теперь нужно найти точку на линии пересечения. Построим вектор $w = r_0 - s_0$ и вычислим его составляющую, параллельную плоскости $s(a, b)$: $w' = w - s_3(w \cdot s_3)$. Далее по формуле (4.7.4) определим точку p пересечения линии, проходящей через точку s_0 в направлении вектора w' , с плоскостью $r(u, v)$

$$p = s_0 + w' \frac{r_3 \cdot (r_0 - s_0)}{r_3 \cdot w'}.$$

Плоскости пересекаются по линии

$$l(t) = p + tz. \quad (4.9.1)$$

Параметры плоскостей, соответствующие точки \mathbf{p} , определим по формулам

$$u_p = \frac{\mathbf{r}_1 \cdot (\mathbf{p} - \mathbf{r}_0)}{\mathbf{r}_1 \cdot \mathbf{r}_1}, \quad v_p = \frac{\mathbf{r}_2 \cdot (\mathbf{p} - \mathbf{r}_0)}{\mathbf{r}_2 \cdot \mathbf{r}_2},$$

$$a_p = \frac{\mathbf{s}_1 \cdot (\mathbf{p} - \mathbf{s}_0)}{\mathbf{s}_1 \cdot \mathbf{s}_1}, \quad b_p = \frac{\mathbf{s}_2 \cdot (\mathbf{p} - \mathbf{s}_0)}{\mathbf{s}_2 \cdot \mathbf{s}_2},$$

при условии ортогональности пар векторов $\mathbf{r}_1, \mathbf{r}_2$ и $\mathbf{s}_1, \mathbf{s}_2$, или по формулам (4.6.6) и (4.6.7) в противном случае. Аналогичным образом можно найти параметры для еще одной точки линии пересечения $\mathbf{q} = \mathbf{l}(1) = \mathbf{p} + \mathbf{z}$. Пусть параметры u_q и v_q соответствуют точке \mathbf{q} на плоскости $\mathbf{r}(u, v)$, а параметры a_q и b_q соответствуют этой же точке на плоскости $\mathbf{s}(a, b)$. Построим на плоскостях двумерные линии

$$l_{uv}(t) = [u(t) \ v(t)]^\top = [u_p(1-t) + u_q t \ v_p(1-t) + v_q t]^\top, \quad l_{uv}(t) \in \mathbf{r}(u, v),$$

$$l_{ab}(t) = [a(t) \ b(t)]^\top = [a_p(1-t) + a_q t \ b_p(1-t) + b_q t]^\top, \quad l_{ab}(t) \in \mathbf{s}(a, b).$$

(4.9.2)

Эти линии при каждом параметре t совпадают в пространстве друг с другом и с линией $\mathbf{l}(t)$. Линию пересечения плоскостей будем строить в виде совокупности двух двумерных линий на двух пересекаемых плоскостях. Аналогично линию пересечения поверхностей будем строить в виде совокупности двух двумерных линий на них.

Частные случаи. Во многих частных случаях задачу пересечения поверхностей можно решить достаточно просто. Пусть поверхность (3.5.1) получена выдавливанием кривой на плоскости. Если плоская кривая задана координатными функциями $\mathbf{c}(t) = [x(t) \ y(t)]^\top$, $t_0 \leq t \leq t_1$, и плоскостью $\mathbf{s}(x, y) = \mathbf{p} + x\mathbf{i}_x + y\mathbf{i}_y$, то поверхность выдавливания будет описываться радиус-вектором

$$\mathbf{r}(u, v) = \mathbf{p} + x(u)\mathbf{i}_x + y(u)\mathbf{i}_y + v\mathbf{d}, \quad t_0 \leq u \leq t_1.$$

При пересечении этой поверхности плоскостью $\mathbf{w}(a, b) = \mathbf{q} + a\mathbf{j}_x + b\mathbf{j}_y$, параллельной плоскости $\mathbf{s}(x, y)$, линию пересечения построим из трансформированной копии кривой $\mathbf{c}(t)$ на плоскости $\mathbf{w}(a, b)$ и отрезка прямой линии $l_{uv}(t) = [t \ v]^\top$ на поверхности $\mathbf{r}(u, v)$. Трансформированную копию кривой $\mathbf{c}(t)$ на плоскости $\mathbf{w}(a, b)$ обозначим через $\mathbf{c}_{ab}(t)$. Кривая $\mathbf{c}_{ab}(t)$ должна быть повернута и перемещена на плоскости $\mathbf{w}(a, b)$ в соответствии с взаимной ориентацией базисных векторов и начальных точек плоскостей $\mathbf{w}(a, b)$ и $\mathbf{s}(x, y)$. Отрезок прямой $l_{uv}(t)$ представляет собой координатную линию $v = \text{const}$ на поверхности $\mathbf{r}(u, v)$. Необходимо, чтобы отрезок прямой $l_{uv}(t)$ и кривая $\mathbf{c}_{ab}(t)$, расположенные на разных поверхностях, имели одинаковую параметрическую длину $t_0 \leq t \leq t_1$.

Аналогично можно получить линию пересечения поверхности, полученной вращением кривой на плоскости вокруг оси, принадлежащей этой плоскости, и плоскости, проходящей через ось вращения. Кривую пересечения поверхности вращения плоскостью, ортогональной оси вращения, построим из окружности на плоскости и линии $u = \text{const}$ на поверхности. При построении линии пересе-

чения такими способами следует соблюдать одно правило: линии на пересекаемых поверхностях должны иметь одинаковые области определения параметра и их соответствующие точки на поверхностях должны совпадать в пространстве.

Общий случай пересечения поверхностей. Обратимся к общему случаю пересечения поверхностей. Пусть требуется найти линию пересечения поверхностей, описываемых радиус-векторами $\mathbf{r}(u, v)$ и $\mathbf{s}(a, b)$. Каждая точка линии пересечения поверхностей должна удовлетворять векторному уравнению

$$\mathbf{r}(u, v) - \mathbf{s}(a, b) = 0. \quad (4.9.3)$$

Это векторное уравнение содержит три скалярных уравнения для компонент радиус-векторов поверхностей:

$$\begin{aligned} r_1(u, v) - s_1(a, b) &= 0, \\ r_2(u, v) - s_2(a, b) &= 0, \\ r_3(u, v) - s_3(a, b) &= 0; \end{aligned} \quad (4.9.4)$$

и четыре искомого параметра: u, v, a, b . Данная задача отличается от рассмотренных тем, что из трех уравнений нужно найти четыре неизвестные величины. Если определить один из четырех параметров, то остальные три можно найти из системы уравнений (4.9.4), т. е. один из параметров может в то же

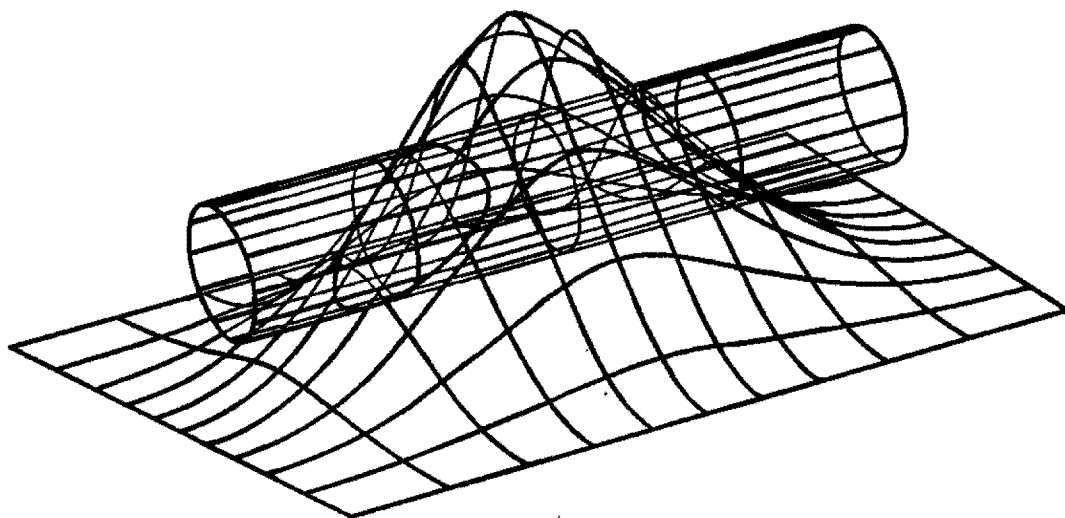


Рис. 4.9.1. Линии пересечения цилиндрической поверхности и NURBS поверхности

время служить параметром решения или являться однозначной функцией от него. Таким образом, три параметра являются функциями четвертого параметра. Чтобы все четыре параметра были равноправными, представим их все в виде функции некоторого общего параметра t :

$$u = u(t), \quad v = v(t), \quad a = a(t), \quad b = b(t), \quad (4.9.5)$$

и придем к тому, что результатом решения системы уравнений (4.9.4) являются две двумерные линии

$$\begin{aligned} l_{uv}(t) &= [u(t) \ v(t)]^T, & l_{uv}(t) &\in \mathbf{r}(u, v), \\ l_{ab}(t) &= [a(t) \ b(t)]^T, & l_{ab}(t) &\in \mathbf{s}(a, b). \end{aligned} \quad (4.9.6)$$

Эти линии связаны друг с другом общим параметром и определяют одну и ту же линию в пространстве — линию пересечения поверхностей. На рис. 4.9.1 приведен пример пересечения поверхностей.

Алгоритм построения линий пересечения. Две поверхности могут пересекаться по нескольким линиям. Мы ставим своей целью построить все линии пересечения, принадлежащие области определения обеих поверхностей. Алгоритм построения линий пересечения поверхностей разобьем на два этапа.

На *первом этапе*, шагая по поверхностям и исследуя их близость, найдем отдельные точки кривых пересечения. Под точками в данном случае будем иметь в виду пары параметров u, v и a, b на параметрических плоскостях поверхностей. Точки из полученного на первом этапе набора никак не связаны друг с другом и могут принадлежать различным кривым пересечения. Важно только, чтобы от каждой кривой пересечения 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 затем в другую, найдем точку за точкой искомую совокупность точек линии пересечения. Направление движения дает векторное произведение нормалей к поверхностям. Шаг движения вычислим в соответствии с кривизнами поверхностей в текущей точке. Движение по кривой пересечения 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 пересечения 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 и $a_i, b_i, i = 0, 1, 2, \dots, n$, одной кривой пересечения. При этом в наборе точек, полученном на первом этапе, не будет содержаться ни одной точки данной кривой пересечения. Если в наборе останутся еще точки, то данные поверхности имеют, по крайней мере, еще одну линию пересечения. Совокупность ее точек найдем, взяв любую точку из набора и повторив все сказанное о втором этапе построения. Построение линий закончим, когда в наборе не останется ни одной точки.

Рассмотрим подробнее отдельные элементы алгоритма построения линий пересечения двух поверхностей.

Для определения начальных точек пересечения на первом этапе решим задачу пересечения линий $r(u, \text{const})$ и $r(\text{const}, v)$ поверхности $r(u, v)$ с поверхностью $s(a, b)$. Не изменяющемуся параметру будем сообщать последовательно значения $\text{const} = v_{\min} + \Delta v$ (при этом он проходит с некоторым шагом значения от v_{\min} до v_{\max}), а затем — $\text{const} = u_{\min} + \Delta u$ (при этом он проходит с некоторым шагом значения от u_{\min} до u_{\max}). При одном неподвижном параметре из системы уравнений (4.9.4) можно найти остальные три параметра. Далее поменяем поверхности ролями и найдем точки пересечения координатных линий $s(a, \text{const})$ и $s(\text{const}, b)$ поверхности $s(a, b)$ с поверхностью $r(u, v)$. Таким образом, мы получим набор некоторых точек, принадлежащих в общем случае различным линиям пересечения заданных поверхностей. В этом наборе нам понадобится по одной точке для построения каждой линии пересечения. Эти точки не должны быть точками касания поверхностей. Если известна некоторая точка линии пересечения, то по ней можно построить и другие точки этой линии.

Возьмем из полученного набора произвольную точку пересечения поверхностей. Пусть ее параметры на первой поверхности равны u_k, v_k , а на второй поверхности равны a_k, b_k . Построим нормали к поверхностям $m_r(u_k, v_k)$ и

$\mathbf{m}_s(a_k, b_k)$. Линия пересечения в каждой точке ортогональна нормальям обеих поверхностей, поэтому направление линии пересечения коллинеарно векторному произведению нормалей к поверхностям

$$\mathbf{t} = \mathbf{m}_r(u_k, v_k) \times \mathbf{m}_s(a_k, b_k). \quad (4.9.9)$$

Выберем в качестве направления движения к следующей точке линии пересечения положительное направление вектора \mathbf{t} . Далее, по формулам (4.2.4) и (4.2.5) определим возможный шаг смещения по каждому из четырех параметров Δu , Δv , Δa , Δb . Если одному из параметров дать приращение и сделать его неподвижным, то из системы уравнений (4.9.4) можно найти остальные параметры. Важно только, чтобы изменение каждого из параметров по модулю не превышало найденные возможные шаги. В качестве фиксированного параметра, которому будет дано приращение, выберем тот, направление движения вдоль которого ближе к направлению кривой пересечения, т. е. составляет меньший угол с вектором \mathbf{t} . Для этого вычислим частные производные радиус-векторов поверхностей по параметрам

$$\mathbf{r}_1 = \frac{\partial \mathbf{r}}{\partial u}, \quad \mathbf{r}_2 = \frac{\partial \mathbf{r}}{\partial v}, \quad \mathbf{s}_1 = \frac{\partial \mathbf{s}}{\partial a}, \quad \mathbf{s}_2 = \frac{\partial \mathbf{s}}{\partial b}, \quad (4.9.10)$$

которые и дают направления движения вдоль параметров.

На рис. 4.9.3 показаны отправная точка $\mathbf{p}_k = \mathbf{r}(u_k, v_k) = \mathbf{s}(a_k, b_k)$, векторы производных \mathbf{r}_1 , \mathbf{r}_2 , \mathbf{s}_1 , \mathbf{s}_2 и сфера смещения, внутри которой приращения всех параметров не превышают, соответственно, Δu , Δv , Δa , Δb . Следующая

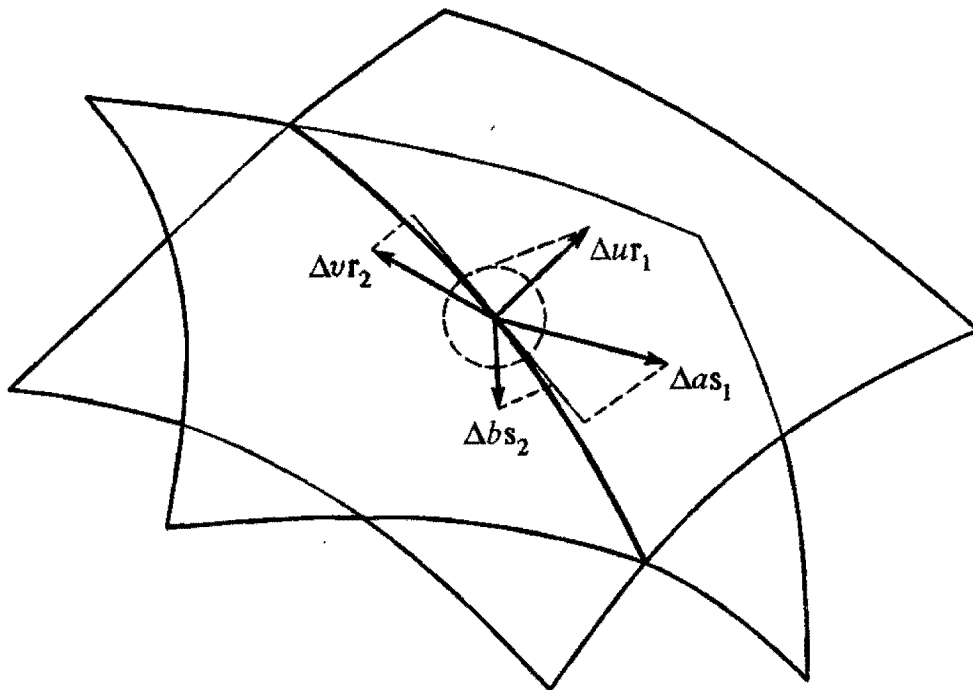


Рис. 4.9.3. Определение неподвижного параметра и начальных приближений для следующей точки линии пересечения поверхностей

точка линии пересечения не должна выйти за сферу смещения. Радиус сферы смещения r_0 возьмем равным наименьшей по абсолютной величине проекции векторов $\Delta u \mathbf{r}_1$, $\Delta v \mathbf{r}_2$, $\Delta a \mathbf{s}_1$, $\Delta b \mathbf{s}_2$ на вектор касательной \mathbf{t} :

$$r_0 = \min (|\Delta u \mathbf{r}_1 \cdot \mathbf{t}|, |\Delta v \mathbf{r}_2 \cdot \mathbf{t}|, |\Delta a \mathbf{s}_1 \cdot \mathbf{t}|, |\Delta b \mathbf{s}_2 \cdot \mathbf{t}|) \frac{1}{|\mathbf{t}|}. \quad (4.9.11)$$

Для следующей точки линии пересечения в качестве нулевых приближений примем значения параметров:

$$u_{k+1}^{(0)} = u_k + r_0 \frac{\mathbf{t} \cdot \mathbf{r}_1}{|\mathbf{t}|(\mathbf{r}_1 \cdot \mathbf{r}_1)}, \quad v_{k+1}^{(0)} = v_k + r_0 \frac{\mathbf{t} \cdot \mathbf{r}_2}{|\mathbf{t}|(\mathbf{r}_2 \cdot \mathbf{r}_2)}, \quad (4.9.12)$$

$$a_{k+1}^{(0)} = a_k + r_0 \frac{\mathbf{t} \cdot \mathbf{s}_1}{|\mathbf{t}|(\mathbf{s}_1 \cdot \mathbf{s}_1)}, \quad b_{k+1}^{(0)} = b_k + r_0 \frac{\mathbf{t} \cdot \mathbf{s}_2}{|\mathbf{t}|(\mathbf{s}_2 \cdot \mathbf{s}_2)}. \quad (4.9.13)$$

Тот параметр, производная по которому составляет меньший угол (по абсолютной величине) с направлением линии пересечения, сделаем неподвижным. Далее, решив систему уравнений (4.9.4), в которой неподвижный параметр будет считаться известным и неизменным, получим остальные три параметра. Таким образом, найдем параметры u_{k+1} , v_{k+1} на первой поверхности и параметры a_{k+1} , b_{k+1} на второй поверхности для следующей точки линии пересечения. Продолжим этот процесс и дальше пока не дойдем до края одной из поверхностей или не окажемся вблизи точки отправления.

Если в процессе движения мы оказались вблизи точки отправления, это означает, что кривая пересечения является замкнутой. Близость точек определяется с помощью тех же шагов по параметрам, которые используются в процессе движения. Кривая пересечения должна замкнуться, если на очередном шаге параметры отличаются не более, чем на величины Δu , Δv , Δa , Δb соответственно.

Если в процессе движения мы оказались вблизи границы одной из поверхностей и движемся в ее направлении (имеется в виду прямоугольная граница), то в качестве неподвижного параметра следует выбрать тот, который находится рядом со своей границей. Важно, чтобы были найдены параметры, соответствующие крайним точкам линии пересечения. Для этого функции поверхностей, определяющие векторы точек и их производных, должны корректировать параметры, если они вышли за область определения.

После определения крайней точки линии пересечения поверхностей продолжим движение и поиск параметров линии пересечения в другом направлении от точки старта (в противоположном вектору \mathbf{t} направлении). Такая линия не может быть замкнутой, поэтому движение будем продолжать до тех пор, пока не будут найдены параметры другого края линии пересечения поверхностей.

Если одна из поверхностей или обе имеют не прямоугольную область определения параметров, а ограниченную контурами область на параметрической плоскости, то при поиске параметров линии пересечения в качестве границы можно использовать габаритный прямоугольник внешнего контура области. После построения линий пересечения их следует проверить на пересечение с граничными контурами и усечь этими контурами, сохранив только те части линий пересечения, которые лежат внутри областей определения параметров обеих поверхностей.

Если на некотором участке пересекаемые поверхности имеют близкие (почти коллинеарные) нормали, то кривая пересечения в этом месте может сильно менять свое направление. Шаг по фиксированному параметру в этой области должен быть уменьшен.

Если линии пересечения поверхностей проходят через точку касания, то в этой точке поверхности имеют коллинеарные нормали. Матрица линеаризованной системы уравнений (4.9.4) в этой точке будет иметь нулевой определитель

и из этой системы уравнений нельзя найти параметры точки касания поверхностей. В этом случае вместо системы уравнений (4.9.4) можно решать систему четырех скалярных уравнений

$$\begin{aligned} (\mathbf{r}(u, v) - \mathbf{s}(a, b)) \cdot \frac{\partial \mathbf{r}}{\partial u} &= 0, \\ (\mathbf{r}(u, v) - \mathbf{s}(a, b)) \cdot \frac{\partial \mathbf{r}}{\partial v} &= 0, \\ (\mathbf{s}(a, b) - \mathbf{r}(u, v)) \cdot \frac{\partial \mathbf{s}}{\partial a} &= 0, \\ (\mathbf{s}(a, b) - \mathbf{r}(u, v)) \cdot \frac{\partial \mathbf{s}}{\partial b} &= 0, \end{aligned} \quad (4.9.14)$$

определитель которой не равен нулю в точке касания. Но системе (4.9.14) удовлетворяют не только точки касания, а еще и другие точки, в которых касательные к поверхностям плоскости параллельны друг другу и ортогональны отрезку, соединяющему точки касания. Поэтому после решения системы уравнений (4.9.14) следует проверить то, что полученные точки являются точками касания. Для этого будем использовать тот же критерий, что применяется в задаче пересечения кривой и поверхности.

Итерационный процесс решения системы (4.9.14) закончим, когда на очередной итерации приращение каждого из параметров станет меньше некоторой заданной величины ε . Отклонению $\pm\varepsilon$ полученных в результате решения параметров от их точных значений соответствует максимальное отклонение радиус-вектора полученной точки от его истинного значения. Это отклонение равно произведению погрешности параметров и длины максимальной по модулю производной радиус-векторов:

$$\Delta\rho = \varepsilon \max(|\mathbf{r}_1|, |\mathbf{r}_2|, |\mathbf{s}_1|, |\mathbf{s}_2|). \quad (4.9.15)$$

Если расстояние между найденными точками на поверхностях меньше $\Delta\rho$, то будем считать эти точки неразличимыми и что решение дало точку касания. В противном случае решение не является ни точкой касания, ни точкой пересечения и должно быть опущено.

Итак, в общем случае описанным выше способом каждую линию пересечения двух поверхностей будем представлять упорядоченной совокупностью параметров u_i, v_i и $a_i, b_i, i = 0, 1, 2, \dots, n$, или упорядоченной совокупностью координат двумерных точек на параметрических плоскостях. Оформим эти совокупности в виде двух ломаных (2.4.1) линий $l_{uv}(t) = [u(t) \ v(t)]^T$ и $l_{ab}(t) = [a(t) \ b(t)]^T$ на соответствующих поверхностях. Эти ломаные имеют общий параметр, при целочисленных значениях параметра они совпадают с линией пересечения, при других значениях параметра они проходят вблизи линии пересечения, их концы совпадают с концами линии пересечения поверхностей. Для любого значения параметра ломаных могут быть получены параметры точки пересечения поверхностей с любой заданной точностью из решения системы (4.9.8). Кривая пересечения поверхностей в данном представлении ничем не отличается от любой другой кривой или сплайна. При обращении к кривой пересечения с некоторым параметром она может выдать любую геометрическую информацию о себе.

Пусть мы имеем для линии пересечения поверхностей $\mathbf{r}(u, v)$ и $\mathbf{s}(a, b)$ точные параметрические зависимости (4.9.5). Тогда радиус-вектор $\mathbf{l}(t)$ линии пересечения поверхностей вычислим как среднее арифметическое радиус-векторов точек на пересекаемых поверхностях

$$\mathbf{l}(t) = \frac{\mathbf{r}(u(t), v(t)) + \mathbf{s}(a(t), b(t))}{2}. \quad (4.9.16)$$

Производные радиус-вектора линии пересечения поверхностей вычислим как среднее арифметическое производных кривых на поверхностях по формулам

$$\frac{d\mathbf{l}}{dt} = \frac{1}{2} \left(\frac{d\mathbf{r}}{dt} + \frac{d\mathbf{s}}{dt} \right), \quad (4.9.17)$$

$$\frac{d^2\mathbf{l}}{dt^2} = \frac{1}{2} \left(\frac{d^2\mathbf{r}}{dt^2} + \frac{d^2\mathbf{s}}{dt^2} \right), \quad (4.9.18)$$

$$\frac{d^3\mathbf{l}}{dt^3} = \frac{1}{2} \left(\frac{d^3\mathbf{r}}{dt^3} + \frac{d^3\mathbf{s}}{dt^3} \right). \quad (4.9.19)$$

Производные радиус-векторов кривых на поверхностях вычисляются по формулам (1.7.20)

$$\frac{d\mathbf{r}}{dt} = \mathbf{r}_1 \frac{du}{dt} + \mathbf{r}_2 \frac{dv}{dt}, \quad (4.9.20)$$

$$\frac{d^2\mathbf{r}}{dt^2} = \mathbf{r}_{11} \left(\frac{du}{dt} \right)^2 + 2\mathbf{r}_{12} \frac{du}{dt} \frac{dv}{dt} + \mathbf{r}_{22} \left(\frac{dv}{dt} \right)^2 + \mathbf{r}_1 \frac{d^2u}{dt^2} + \mathbf{r}_2 \frac{d^2v}{dt^2}, \quad (4.9.21)$$

$$\begin{aligned} \frac{d^3\mathbf{r}}{dt^3} = & \mathbf{r}_{111} \left(\frac{du}{dt} \right)^3 + 3\mathbf{r}_{112} \left(\frac{du}{dt} \right)^2 \frac{dv}{dt} + 3\mathbf{r}_{122} \frac{du}{dt} \left(\frac{dv}{dt} \right)^2 + \mathbf{r}_{222} \left(\frac{dv}{dt} \right)^3 + \\ & + 3\mathbf{r}_{11} \frac{d^2u}{dt^2} \frac{du}{dt} + 3\mathbf{r}_{12} \left(\frac{d^2u}{dt^2} \frac{dv}{dt} + \frac{d^2v}{dt^2} \frac{du}{dt} \right) + 3\mathbf{r}_{22} \frac{d^2v}{dt^2} \frac{dv}{dt} + \mathbf{r}_1 \frac{d^3u}{dt^3} + \mathbf{r}_2 \frac{d^3v}{dt^3}, \end{aligned} \quad (4.9.22)$$

$$\frac{d\mathbf{s}}{dt} = \mathbf{s}_1 \frac{da}{dt} + \mathbf{s}_2 \frac{db}{dt}, \quad (4.9.23)$$

$$\frac{d^2\mathbf{s}}{dt^2} = \mathbf{s}_{11} \left(\frac{da}{dt} \right)^2 + 2\mathbf{s}_{12} \frac{da}{dt} \frac{db}{dt} + \mathbf{s}_{22} \left(\frac{db}{dt} \right)^2 + \mathbf{s}_1 \frac{d^2a}{dt^2} + \mathbf{s}_2 \frac{d^2b}{dt^2}, \quad (4.9.24)$$

$$\begin{aligned} \frac{d^3\mathbf{s}}{dt^3} = & \mathbf{s}_{111} \left(\frac{da}{dt} \right)^3 + 3\mathbf{s}_{112} \left(\frac{da}{dt} \right)^2 \frac{db}{dt} + 3\mathbf{s}_{122} \frac{da}{dt} \left(\frac{db}{dt} \right)^2 + \mathbf{s}_{222} \left(\frac{db}{dt} \right)^3 + \\ & + 3\mathbf{s}_{11} \frac{d^2a}{dt^2} \frac{da}{dt} + 3\mathbf{s}_{12} \left(\frac{d^2a}{dt^2} \frac{db}{dt} + \frac{d^2b}{dt^2} \frac{da}{dt} \right) + 3\mathbf{s}_{22} \frac{d^2b}{dt^2} \frac{db}{dt} + \mathbf{s}_1 \frac{d^3a}{dt^3} + \mathbf{s}_2 \frac{d^3b}{dt^3}. \end{aligned} \quad (4.9.25)$$

Если параметрические зависимости (4.9.5) не являются точными и дают точки линий пересечения при целочисленных значениях параметра t (при других значениях параметра проходят вблизи линии пересечения), то формулы (4.9.16)–(4.9.19) будут верны только приближенно. Точные значения радиус-вектора линии пересечения поверхностей и ее производных можно получить из системы уравнений (4.9.8).

4.10. Поверхность скругления

Многие моделируемые детали имеют скругления, поэтому при построении тел требуется выполнять операцию скругления ребер тела. Рассмотрим построение поверхности, которая в дальнейшем будет использоваться для скругления ребер тел. Пока будем строить поверхности скругления, не связывая их с телами.

Пусть имеются две пересекающиеся поверхности, описываемые радиус-векторами $r(u, v)$ и $s(a, b)$. Вблизи линии пересечения пространство делится поверхностями на четыре сектора.

Сектор 1: перпендикуляры, восстановленные от поверхностей к точкам первого сектора, имеют направление, совпадающее с нормальными обеими поверхностями.

Сектор 2: перпендикуляр, восстановленный от первой поверхности к точкам второго сектора, совпадает по направлению с нормалью первой поверхности, а перпендикуляр, восстановленный от второй поверхности к точкам второго сектора, противоположен по направлению нормали второй поверхности.

Сектор 3: перпендикуляры, восстановленные от поверхностей к точкам третьего сектора, противоположны по направлению нормальным обеим поверхностям.

Сектор 4: перпендикуляр, восстановленный от первой поверхности к точкам четвертого сектора, противоположен по направлению нормали первой поверхности, а перпендикуляр, восстановленный от второй поверхности к точкам четвертого сектора, совпадает по направлению с нормалью ко второй поверхности.

Построим поверхность скругления, представляющую собой след от качения сферы радиуса ρ , касающейся одновременно двух поверхностей. Сфера будет

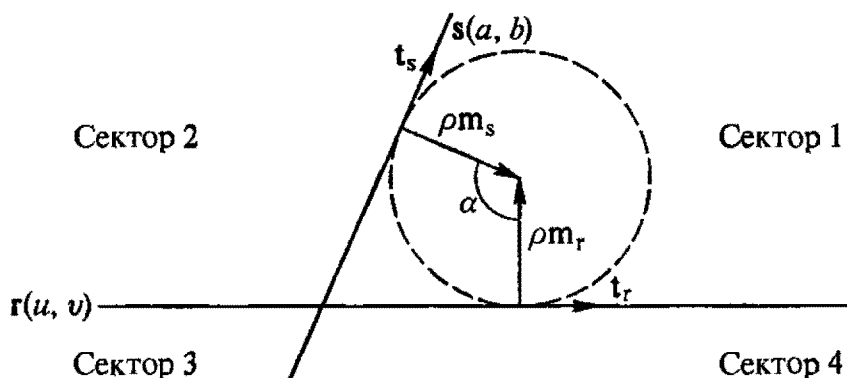


Рис. 4.10.1. Скругление плоских граней

двигаться около линии пересечения поверхностей в одном из четырех упомянутых секторов. На рис. 4.10.1 показано сечение поверхностей и сферы.

Частные случаи. Если скругляемыми поверхностями являются плоскости, то угол α между поверхностями остается постоянным при движении вдоль линии их пересечения. Пусть радиус скругления остается постоянным и равным ρ . В этом случае линии перехода с поверхности скругления на сопрягаемые плоскости можно получить как эквидистантные линии к линии пересечения. Имея

согласованные по параметру линии перехода $l_r(t)$, $l_s(t)$ и линию пересечения плоскостей $l(t)$, поверхность скругления можно представить в виде (3.10.3)

$$q(t, z) = \frac{(1-z)^2 l_r(t) + 2(1-z)z \cos(\alpha/2) l(t) + z^2 l_s(t)}{(1-z)^2 + 2(1-z)z \cos(\alpha/2) + z^2}, \quad 0 \leq z \leq 1.$$

Линии перехода $l_r(t)$, $l_s(t)$ построим в виде линий на поверхностях. Каждая из них представляет собой двумерную линию и поверхность (в данном случае — плоскость). Двухмерные линии могут быть получены как эквидистантные линии к линии пересечения плоскостей $l(t)$, отстоящие от нее на расстоянии $d = \pm \rho \operatorname{tg}(\alpha/2)$. Знак d зависит от ориентации линии пересечения и от сектора, в котором строится поверхность скругления. Область определения параметра t поверхности скругления зависит от дальнейшего ее использования. Полученная поверхность скругления по форме совпадает с частью цилиндрической поверхности. Как правило, в рассмотренном случае строится именно часть цилиндрической поверхности. Аналогичным образом в качестве поверхности скругления между цилиндрической поверхностью и ортогональной ее оси плоскостью может быть использована часть поверхности тора.

Общий случай. Рассмотрим построение поверхности скругления в общем случае. Построим точки касания катящейся сферы радиуса ρ с поверхностями. Продолжение нормалей к поверхностям в точках касания пересекутся в центре катящейся сферы. Обозначим нормали (1.7.18) поверхностей $r(u, v)$ и $s(a, b)$ через $\mathbf{m}_r(u, v)$ и $\mathbf{m}_s(a, b)$, а проекции на эти нормали векторов из точек касания до центра сферы — через ρ_r и ρ_s соответственно. Величины ρ_r и ρ_s по модулю равны радиусу сферы ρ , но имеют знак, характеризующий упомянутый сектор. Параметры точек касания сферы связаны уравнением

$$\mathbf{r}(u, v) + \rho_r \mathbf{m}_r(u, v) = \mathbf{s}(a, b) + \rho_s \mathbf{m}_s(a, b). \quad (4.10.1)$$

Это векторное уравнение содержит три скалярных уравнения для компонент нормалей поверхностей и четыре искомого параметра u, v, a, b . Построение поверхности скругления по уравнению (4.10.1) сходно с задачей построения линии пересечения поверхностей. В обоих случаях результатом решения являются две двумерные линии на соответствующих поверхностях.

Переменный радиус скругления. Пусть требуется построить поверхность скругления переменного радиуса. Для этого нам потребуется кривая пересечения поверхностей. Величины радиуса скругления $\rho_r = \rho_r(s)$ и $\rho_s = \rho_s(s)$ будем считать функциями длины дуги s линии пересечения $\mathbf{c}(s)$ скругляемых поверхностей. В данном случае катящаяся сфера будет иметь переменный радиус. Кроме того, положение центра катящейся сферы связано с точкой на линии пересечения. Расположим центр катящейся сферы в нормальной плоскости кривой пересечения. Нормальная плоскость ортогональна касательному вектору кривой. Вместо векторного уравнения (4.10.1) параметры точек касания сферы свяжем уравнениями

$$\mathbf{r}(u, v) + \rho_r(s) \mathbf{m}_r(u, v) = \mathbf{s}(a, b) + \rho_s(s) \mathbf{m}_s(a, b), \quad (4.10.2)$$

$$\left(\mathbf{c}(s) - \frac{1}{2} (\mathbf{r}(u, v) + \rho_r(s) \mathbf{m}_r(u, v) + \mathbf{s}(a, b) + \rho_s(s) \mathbf{m}_s(a, b)) \right) \cdot \frac{d\mathbf{c}}{ds} = 0. \quad (4.10.3)$$

Эти уравнения содержат четыре скалярных уравнения относительно четырех искомых параметров u, v, a, b . Параметр s линии пересечения является известной величиной. По текущему параметру s мы вычислим радиусы $\rho_r(s)$ и $\rho_s(s)$, точку $\mathbf{c}(s)$ и касательный вектор кривой в ней $\mathbf{c}'(s)$. Решив систему уравнений (4.10.2) и (4.10.3), получим параметры u, v, a, b , касания катящейся сферы и поверхностей.

Система уравнений (4.10.2), (4.10.3) может быть использована вместо системы уравнений (4.10.1) для построения поверхности скругления постоянного радиуса. В этом случае необязательно в качестве параметра кривой пересечения использовать длину ее дуги.

Результатом решения системы уравнений (4.10.1) или системы уравнений (4.10.2) и (4.10.3) являются две двумерные линии на поверхностях

$$\begin{aligned} l_{uv}(t) &= [u(t) \ v(t)]^T, & l_{uv}(t) &\in \mathbf{r}(u, v), \\ l_{ab}(t) &= [a(t) \ b(t)]^T, & l_{ab}(t) &\in \mathbf{s}(a, b), \end{aligned} \quad (4.10.4)$$

на соответствующих поверхностях. В общем случае линии $l_{uv}(t)$ и $l_{ab}(t)$ могут быть получены как сплайны, проходящие через заданные точки. Пространственные линии, построенные по этим линиям на поверхностях, обозначим соответственно через

$$l_r(t) = \mathbf{r}(u(t), v(t)), \quad (4.10.5)$$

$$l_s(t) = \mathbf{s}(a(t), b(t)). \quad (4.10.6)$$

Они определяют края поверхности скругления, полученной качением сферы одновременно по двум поверхностям.

По двум кривым на поверхностях (4.10.5) и (4.10.6), являющимися следами касания катящейся сферы, построим поверхность скругления. Первый параметр поверхности скругления совместим с параметром t граничных кривых (4.10.5) и (4.10.6). При движении вдоль второго параметра поверхности скругления при фиксированном первом параметре должна быть описана дуга окружности. Построим эту дугу окружности в виде рациональной кривой Безье (2.6.16). Для этого при каждом значении параметра кривых на поверхности нужно знать радиус-вектор средней точки и ее вес. Вес средней точки рациональной кривой Безье (2.6.16) равен косинусу половины угла между векторами $\rho_r \mathbf{m}_r$ и $\rho_s \mathbf{m}_s$. Косинус угла между этими векторами равен их скалярному произведению, деленному на произведение длин векторов: $\cos \alpha = \frac{\rho_r \mathbf{m}_r \cdot \mathbf{m}_s \rho_s}{\rho^2}$. Используя формулу $2 \cos^2(\alpha/2) = 1 + \cos \alpha$, вычислим вес средней точки

$$w(t) = \cos \frac{\alpha}{2} = \frac{1}{\sqrt{2}} \sqrt{1 + \frac{(\rho_r \mathbf{m}_r \cdot \mathbf{m}_s \rho_s)}{\rho^2}}. \quad (4.10.7)$$

На рис. 2.6.5 видно, что радиус-вектор средней точки, который обозначим через $\mathbf{l}(t)$, можно вычислить по формуле

$$\mathbf{l}(t) = \frac{1}{2} (\mathbf{l}_r(t) + \rho_r \mathbf{m}_r(t) + \mathbf{l}_s(t) + \rho_s \mathbf{m}_s(t)) - \frac{1}{2} \frac{\rho_r \mathbf{m}_r(t) + \rho_s \mathbf{m}_s(t)}{w^2(t)}. \quad (4.10.8)$$

Радиус-вектор средней точки (4.10.8) и ее вес (4.10.7) являются функциями параметра t кривых на поверхности (4.10.5) и (4.10.6), так как от параметра t зависят \mathbf{l}_r , \mathbf{l}_s и нормали к поверхностям

$$\mathbf{m}_r(t) = \mathbf{m}_r(u(t), v(t)), \quad (4.10.9)$$

$$\mathbf{m}_s(t) = \mathbf{m}_s(a(t), b(t)). \quad (4.10.10)$$

Таким образом, радиус-вектор поверхности скругления при постоянном или переменном радиусе описывается выражением

$$\mathbf{q}(t, z) = \frac{(1-z)^2 \mathbf{l}_r(t) + 2w(t)(1-z)z \mathbf{l}(t) + z^2 \mathbf{l}_s(t)}{(1-z)^2 + 2w(t)(1-z)z + z^2}, \quad 0 \leq z \leq 1, \quad (4.10.11)$$

где вес $w(t)$ и радиус-вектор $\mathbf{l}(t)$ определяются равенствами (4.10.7) и (4.10.8), а через z обозначен второй параметр поверхности. Рассмотренная поверхность скругления не имеет четких границ в направлении первого параметра. Эти границы будут определены при дальнейшем использовании поверхности для скругления ребер тел. На рис. 4.10.2 приведен пример поверхности скругления. В зависимости от замкнутости скругляемых поверхностей и линий (4.10.5) и (4.10.6) поверхность скругления может быть замкнутой или незамкнутой.

При решении системы уравнений (4.10.1) и (4.10.2) требуется вычислять производные нормалей поверхностей по параметрам. Эти производные дают

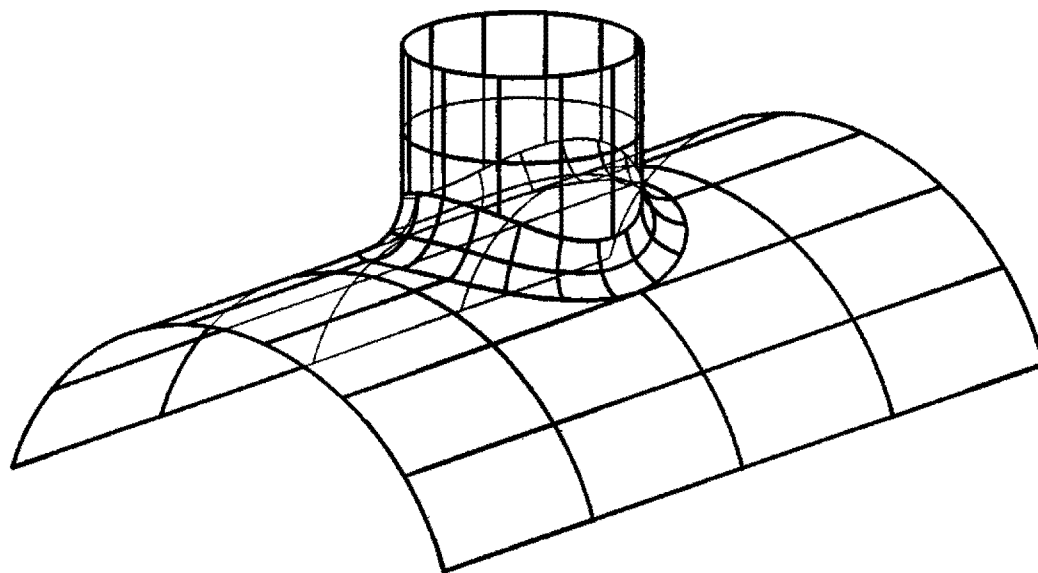


Рис. 4.10.2. Поверхность скругления

формулы Вейнгартена (1.7.26). Радиус-вектор точки поверхности за ее пределами может быть вычислен по одной из формул (3.14.8)–(3.14.10) в зависимости от замкнутости поверхности. Эти же формулы позволяют определить нормали поверхности и их производные за пределами поверхности.

4.11. Поверхность фаски

Наряду со скруглениями ребер тела требуется выполнять снятие фасок вдоль его ребер. Построение поверхности фаски имеет много общего с построением поверхности скругления. Поверхность фаски несколько проще поверхности скругления. Остановимся на том, что их отличает.

Пусть имеются две пересекающиеся поверхности, описываемые радиус-векторами $\mathbf{r}(u, v)$ и $\mathbf{s}(a, b)$. Радиус-вектор поверхности фаски представим в форме линейчатой поверхности

$$\mathbf{q}(t, z) = (1 - z)\mathbf{l}_r(t) + z\mathbf{l}_s(t), \quad 0 \leq z \leq 1, \quad (4.11.1)$$

построенной по двум линиям на заданных поверхностях

$$\mathbf{l}_r(t) = \mathbf{r}(u(t), v(t)), \quad (4.11.2)$$

$$\mathbf{l}_s(t) = \mathbf{s}(a(t), b(t)), \quad (4.11.3)$$

проходящим вдоль их линии пересечения. Эти двумерные линии определяют края поверхности фаски (4.11.1). Линии (4.11.2) и (4.11.3) должны иметь общий параметр t . Первый параметр поверхности фаски совместим с параметром кривых, на которых она построена. Границы поверхности фаски по первому параметру будут определены при дальнейшем использовании поверхности.

Если известна линия пересечения поверхностей (например, при снятии фаски на ребре тела) и угол α между поверхностями остается постоянным при движении вдоль линии пересечения, то линии $\mathbf{l}_r(t)$, $\mathbf{l}_s(t)$ можно построить как эквидистантные линии к линии пересечения. Рассмотрим построение поверхности фаски в общем виде.

Общий случай. Вблизи линии пересечения пространство будем делить поверхностями на четыре сектора таким же образом, как и при построении поверхности скругления. Это нужно для того, чтобы определить, с какой стороны от линии пересечения проходят линии (4.11.2) и (4.11.3). Через $\mathbf{m}_r(u, v)$ и $\mathbf{m}_s(a, b)$ обозначим нормали (1.7.18) поверхностей $\mathbf{r}(u, v)$ и $\mathbf{s}(x, y)$, а через

$$\mathbf{t}_r(u, v) = \pm \frac{\mathbf{m}_r \times (\mathbf{m}_r \times \mathbf{m}_s)}{|\mathbf{m}_r \times \mathbf{m}_s|}, \quad (4.11.4)$$

$$\mathbf{t}_s(a, b) = \pm \frac{\mathbf{m}_s \times (\mathbf{m}_s \times \mathbf{m}_r)}{|\mathbf{m}_s \times \mathbf{m}_r|} \quad (4.11.5)$$

обозначим единичные векторы трансверселей, которые лежат в касательных к поверхностям плоскостях и ортогональны линиям $\mathbf{l}_r(t)$ и $\mathbf{l}_s(t)$ соответственно. На рис. 4.11.1 приведены нормали и трансверсели (4.11.4) и (4.11.5).

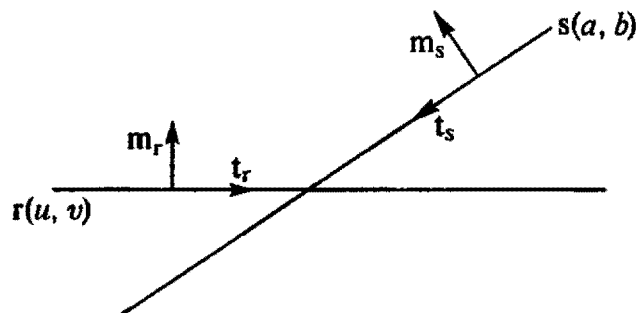


Рис. 4.11.1

Задача построения поверхности фаски сходна с задачей построения поверхности скругления. Параметры $u(t)$, $v(t)$, $a(t)$, $b(t)$, определяющие линии (4.11.2) и (4.11.3) и являющиеся координатами точек на поверхностях, связаны вектор-

ным уравнением

$$\mathbf{r}(u, v) + d_r \mathbf{t}_r(u, v) = \mathbf{s}(a, b) + d_s \mathbf{t}_s(a, b). \quad (4.11.6)$$

Величины d_r и d_s определяют расстояние от линий (4.11.2) и (4.11.3), соответственно, до точки пересечения трансверсалей. Знаки величин d_r и d_s определяют, с какой стороны от линии пересечения проходят линии $l_r(t)$ и $l_s(t)$, т.е. определяют сектор, в котором лежит поверхность фаски. Если в области пересечения поверхности плоские и ортогональны друг другу, то абсолютные величины d_r и d_s равны катетам фаски.

Фаска с переменными катетами. Поверхность фаски можно связать с линией пересечения поверхностей $\mathbf{c}(s)$, а величины d_r и d_s можно поставить в зависимость от длины дуги s этой линии. В этом случае параметры $u(t)$, $v(t)$, $a(t)$, $b(t)$, определяющие линии (4.11.2) и (4.11.3), связаны уравнениями

$$\mathbf{r}(u, v) + d_r(s) \mathbf{t}_r(u, v) = \mathbf{s}(a, b) + d_s(s) \mathbf{t}_s(a, b), \quad (4.11.7)$$

$$\left(\mathbf{c}(s) - \frac{1}{2} (\mathbf{r}(u, v) + d_r \mathbf{t}_r(u, v) + \mathbf{s}(a, b) + d_s \mathbf{t}_s(a, b)) \right) \cdot \frac{d\mathbf{c}}{ds} = 0. \quad (4.11.8)$$

Данная система содержит четыре уравнения относительно четырех параметров: u , v , a , b .

При решении системы уравнений (4.11.7) и (4.11.8) требуется вычислять производные нормалей поверхностей по параметрам. Эти производные определяют формулы Вейнгартена (1.7.26). Радиус-вектор точки поверхности за ее

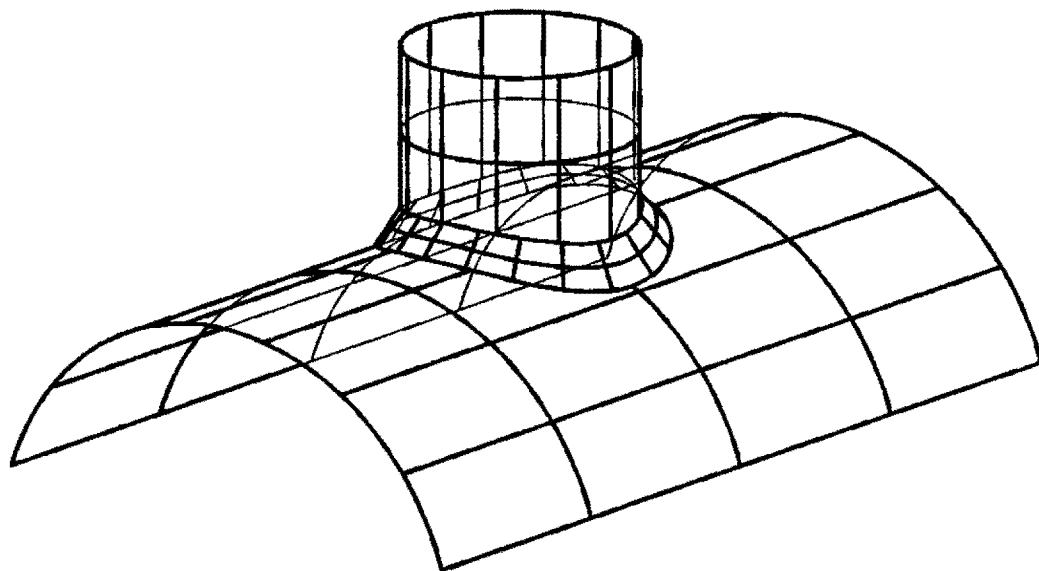


Рис. 4.11.2. Поверхность фаски

пределами может быть вычислен по одной из формул (3.14.8)–(3.14.10) в зависимости от замкнутости поверхности. Эти же формулы позволяют определить нормали и векторы (4.11.4) и (4.11.5) поверхности и их производные за пределами поверхности.

На рис. 4.11.2 приведен пример поверхности фаски.

В зависимости от замкнутости поверхностей и линий (4.11.2) и (4.11.3) поверхность фаски может быть замкнутой или незамкнутой.

4.12. Определение точек пересечения трех поверхностей

В некоторых случаях требуется найти точку (точки) пересечения трех поверхностей. Рассмотрим сначала частный случай — пересечение трех плоскостей.

Пересечение трех плоскостей. Пусть даны три плоскости

$$\begin{aligned} \mathbf{r}(u, v) &= \mathbf{r}_0 + u\mathbf{r}_1 + v\mathbf{r}_2, \\ \mathbf{s}(a, b) &= \mathbf{s}_0 + a\mathbf{s}_1 + b\mathbf{s}_2, \\ \mathbf{q}(x, y) &= \mathbf{q}_0 + x\mathbf{q}_1 + y\mathbf{q}_2 \end{aligned} \quad (4.12.1)$$

с единичными нормальными векторами $\mathbf{m}_r, \mathbf{m}_s, \mathbf{m}_q$ соответственно. Плоскости имеют единственную точку пересечения, если смешанное произведение нормалей отлично от нуля: $(\mathbf{m}_r \times \mathbf{m}_s) \cdot \mathbf{m}_q \neq 0$. Пусть искомая точка имеет радиус-вектор \mathbf{p} . Тогда вектор $\mathbf{p} - \mathbf{r}_0$ коллинеарен плоскости $\mathbf{r}(u, v)$, вектор $\mathbf{p} - \mathbf{s}_0$ коллинеарен плоскости $\mathbf{s}(a, b)$, вектор $\mathbf{p} - \mathbf{q}_0$ коллинеарен плоскости $\mathbf{q}(x, y)$. Следовательно, должны выполняться равенства

$$\begin{aligned} (\mathbf{p} - \mathbf{r}_0) \cdot \mathbf{m}_r &= 0, \\ (\mathbf{p} - \mathbf{s}_0) \cdot \mathbf{m}_s &= 0, \\ (\mathbf{p} - \mathbf{q}_0) \cdot \mathbf{m}_q &= 0. \end{aligned} \quad (4.12.2)$$

Обозначим через p_1, p_2, p_3 координаты точки \mathbf{p} , через m_1, m_2, m_3 — компоненты вектора нормали \mathbf{m}_r , через n_1, n_2, n_3 — компоненты вектора нормали \mathbf{m}_s , через l_1, l_2, l_3 — компоненты вектора нормали \mathbf{m}_q . Тогда система уравнений (4.12.2) примет вид

$$\begin{aligned} m_1 p_1 + m_2 p_2 + m_3 p_3 &= \mathbf{r}_0 \cdot \mathbf{m}_r, \\ n_1 p_1 + n_2 p_2 + n_3 p_3 &= \mathbf{s}_0 \cdot \mathbf{m}_s, \\ l_1 p_1 + l_2 p_2 + l_3 p_3 &= \mathbf{q}_0 \cdot \mathbf{m}_q. \end{aligned} \quad (4.12.3)$$

Решив систему линейных алгебраических уравнений (4.12.3) относительно p_1, p_2, p_3 , найдем общую точку трех плоскостей. Параметры этой точки для каждой из плоскостей найдем по формулам

$$\begin{aligned} u_p &= \frac{\mathbf{r}_1 \cdot (\mathbf{p} - \mathbf{r}_0)}{\mathbf{r}_1 \cdot \mathbf{r}_1}, & v_p &= \frac{\mathbf{r}_2 \cdot (\mathbf{p} - \mathbf{r}_0)}{\mathbf{r}_2 \cdot \mathbf{r}_2}, \\ a_p &= \frac{\mathbf{s}_1 \cdot (\mathbf{p} - \mathbf{s}_0)}{\mathbf{s}_1 \cdot \mathbf{s}_1}, & b_p &= \frac{\mathbf{s}_2 \cdot (\mathbf{p} - \mathbf{s}_0)}{\mathbf{s}_2 \cdot \mathbf{s}_2}, \\ x_p &= \frac{\mathbf{q}_1 \cdot (\mathbf{p} - \mathbf{q}_0)}{\mathbf{q}_1 \cdot \mathbf{q}_1}, & y_p &= \frac{\mathbf{q}_2 \cdot (\mathbf{p} - \mathbf{q}_0)}{\mathbf{q}_2 \cdot \mathbf{q}_2}, \end{aligned} \quad (4.12.4)$$

если ортогональны пары векторов $\mathbf{r}_1, \mathbf{r}_2$; $\mathbf{s}_1, \mathbf{s}_2$; $\mathbf{q}_1, \mathbf{q}_2$, или по формулам (4.6.6) и (4.6.7) в противном случае.

Общий случай. В общем случае пересечения трех поверхностей $\mathbf{r}(u, v)$, $\mathbf{s}(a, b)$ и $\mathbf{q}(x, y)$ каждая точка пересечения должна удовлетворять векторным уравнениям

$$\begin{aligned} \mathbf{r}(u, v) - \mathbf{q}(x, y) &= 0, \\ \mathbf{s}(a, b) - \mathbf{q}(x, y) &= 0. \end{aligned} \quad (4.12.5)$$

Эти векторные уравнения содержат шесть скалярных нелинейных уравнений относительно шести параметров u, v, a, b, x, y поверхностей. Если в точках пересечения смешанное произведение нормалей поверхностей отлично от нуля, то система может быть решена одним из описанных численных методов. Как правило, мы будем обращаться к решению этой задачи с известным начальным приближением. Например, нам будут известны линии пересечения поверхностей $\mathbf{q}(x, y)$ и $\mathbf{r}(u, v)$, $\mathbf{q}(x, y)$ и $\mathbf{s}(a, b)$, и будет требоваться найти пересечение этих линий. Начальное приближение может быть найдено из решения задачи пересечения соответствующих двумерных линий на поверхности $\mathbf{q}(x, y)$.

Пусть нам известно начальное приближение параметров трех поверхностей в районе их общей точки пересечения. Тогда в методе Ньютона на r -й итерации приращения параметров $\Delta u^{(r)}, \Delta v^{(r)}, \Delta a^{(r)}, \Delta b^{(r)}, \Delta x^{(r)}, \Delta y^{(r)}$ точки пересечения будем искать из системы линейных уравнений

$$\begin{aligned} \frac{\partial r_1}{\partial u} \Delta u^{(r)} + \frac{\partial r_1}{\partial v} \Delta v^{(r)} - \frac{\partial q_1}{\partial x} \Delta x^{(r)} - \frac{\partial q_1}{\partial y} \Delta y^{(r)} &= q_1 - r_1, \\ \frac{\partial r_2}{\partial u} \Delta u^{(r)} + \frac{\partial r_2}{\partial v} \Delta v^{(r)} - \frac{\partial q_2}{\partial x} \Delta x^{(r)} - \frac{\partial q_2}{\partial y} \Delta y^{(r)} &= q_2 - r_2, \\ \frac{\partial r_3}{\partial u} \Delta u^{(r)} + \frac{\partial r_3}{\partial v} \Delta v^{(r)} - \frac{\partial q_3}{\partial x} \Delta x^{(r)} - \frac{\partial q_3}{\partial y} \Delta y^{(r)} &= q_3 - r_3, \\ \frac{\partial s_1}{\partial a} \Delta a^{(r)} + \frac{\partial s_1}{\partial b} \Delta b^{(r)} - \frac{\partial q_1}{\partial x} \Delta x^{(r)} - \frac{\partial q_1}{\partial y} \Delta y^{(r)} &= q_1 - s_1, \\ \frac{\partial s_2}{\partial a} \Delta a^{(r)} + \frac{\partial s_2}{\partial b} \Delta b^{(r)} - \frac{\partial q_2}{\partial x} \Delta x^{(r)} - \frac{\partial q_2}{\partial y} \Delta y^{(r)} &= q_2 - s_2, \\ \frac{\partial s_3}{\partial a} \Delta a^{(r)} + \frac{\partial s_3}{\partial b} \Delta b^{(r)} - \frac{\partial q_3}{\partial x} \Delta x^{(r)} - \frac{\partial q_3}{\partial y} \Delta y^{(r)} &= q_3 - s_3, \end{aligned} \tag{4.12.6}$$

где r_1, r_2, r_3 — компоненты радиус-вектора поверхности $\mathbf{r}(u^{(r)}, v^{(r)})$, s_1, s_2, s_3 — компоненты радиус-вектора поверхности $\mathbf{s}(a^{(r)}, b^{(r)})$, q_1, q_2, q_3 — компоненты радиус-вектора поверхности $\mathbf{q}(x^{(r)}, y^{(r)})$. Следующее приближение параметров точки пересечения равны $u^{(r+1)} = u^{(r)} + \Delta u^{(r)}$, $v^{(r+1)} = v^{(r)} + \Delta v^{(r)}$, $a^{(r+1)} = a^{(r)} + \Delta a^{(r)}$, $b^{(r+1)} = b^{(r)} + \Delta b^{(r)}$, $x^{(r+1)} = x^{(r)} + \Delta x^{(r)}$, $y^{(r+1)} = y^{(r)} + \Delta y^{(r)}$. Процесс уточнения параметров закончим, когда на очередной итерации выполняются неравенства $|u^{(r+1)} - u^{(r)}| < \varepsilon$, $|v^{(r+1)} - v^{(r)}| < \varepsilon$, $|a^{(r+1)} - a^{(r)}| < \varepsilon$, $|b^{(r+1)} - b^{(r)}| < \varepsilon$, $|x^{(r+1)} - x^{(r)}| < \varepsilon$, $|y^{(r+1)} - y^{(r)}| < \varepsilon$.

4.13. Точность выполнения операций

Для описания кривых линий и поверхностей мы используем параметрическое представление. Геометрические объекты построены так, что они выдают необходимую геометрическую информацию для заданных параметров. Поэтому поиск некоторого решения заключается в поиске соответствующих ему параметров. Если известны параметры, то можно считать, что известно все остальное. Параметры решения не изменяются при одновременном перемещении, повороте или трансформировании взаимодействующих геометрических объектов.

В итерационных методах выполнения операций параметры геометрических объектов мы находим с некоторой точностью ε . Мы заканчиваем процесс уточнения параметров, когда на очередной итерации изменение каждого параметра по абсолютной величине становится меньше ε : ($|x^{(k+1)} - x^{(k)}| < \varepsilon$). На самом деле, точность решения ниже величины ε . Вблизи точек касания решение сходится очень медленно, а в самих точках касания решение может расходиться. Значение некоторого параметра $x^{(k+1)}$ на последней итерации отличается от искомого точного значения x^* на величину $\xi \approx |x^{(k+1)} - x^*|$, которая в несколько десятков или сотен раз может быть больше величины ε . В то же время, величина ε не должна превосходить 10^{-n} , где n — число точных чисел в мантиссе представления параметров. Общее количество $m > n$ значащих цифр в мантиссе представления действительных чисел в памяти компьютера является отправной точкой для определения максимальной точности, которую можно достичь численными методами.

Пусть мы определились с величиной ξ , которую назовем *заданной точностью*. Две точки геометрического объекта мы будем считать совпадающими, если соответствующие им параметры отличаются меньше, чем на величину ξ . Кроме величины ξ нам потребуется *точностью выполнения операций* — величина ε , которая связана с величиной ξ и должна быть меньше нее ($\varepsilon < \xi$).

Там, где это возможно, операции следует выполнять без привлечения численных методов. В этих случаях точность выполнения операций будет зависеть от точности диагностики частных случаев. Например, при пересечении цилиндрической поверхности с ортогональной ее оси плоскостью точность построения линии пересечения будет определяться точностью, при которой вектор нормали плоскости и ось цилиндра считаются параллельными. Два вектора можно считать параллельными, если длина их векторного произведения меньше некоторой величины. Если по результатам проверки векторов на параллельность мы будем выполнять операцию, то проверка должна выполняться с участием точности выполнения операций ε . Если проверка выполняется для других целей, то она должна выполняться с участием точности ξ .

Возникает вопрос, каким образом с помощью величин ξ и ε , связанных с параметрами, оценить близость точек или векторов в пространстве? Для ответа на этот вопрос нам придется привлечь некоторый геометрический размер модели. В качестве такого геометрического размера можно взять диагональ d габаритного куба модели. Предположим, что параметрическая длина некоторой кривой или поверхности модели не меньше единицы ($t_{\max} - t_{\min} \geq 1$) и размеры соизмеримы с диагональю d . Тогда две точки геометрического объекта, находящиеся на параметрическом расстоянии ξ или ε , в пространстве будут находиться на расстоянии, не превышающем ξd или εd соответственно. Если нам известно значение параметра точки \mathbf{r} геометрического объекта, в которой параметр определен с точностью ε , то в пространстве эта точка будет определена с точностью $\varepsilon r'$. Точно так же отклонение вектора длиной d на угол ξ или ε даст перемещение в пространстве, примерно равное ξd или εd соответственно. Через габаритный размер d геометрические погрешности связаны с параметрическими и угловыми погрешностями ξ и ε .

При выполнении операций важно знать, с какой точностью они были выполнены. Мы будем пользоваться двумя величинами: ξ и ε . С точностью ε мы будем выполнять операции, а с точностью ξ мы будем проверять величины на равенство.

Глава 5

ТОПОЛОГИЯ ОБОЛОЧЕК

5.1. 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 потребуются 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 несклеиваемого 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. Одним

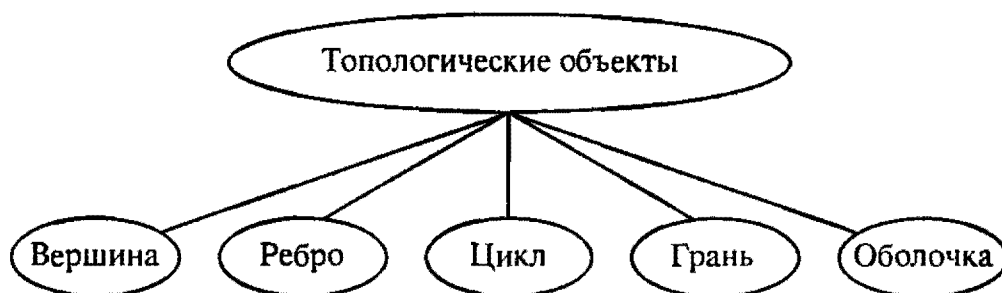


Рис. 5.1.1. Топологические объекты

из топологических объектов является оболочка. При построении оболочки будем использовать такие топологические объекты, как грани, ребра, вершины и циклы (рис. 5.1.1). 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. Таким образом, внешний цикл грани ориентирован против

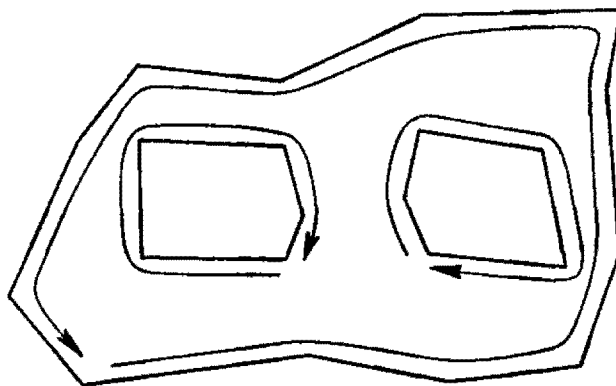


Рис. 5.1.2. Ориентация внешнего и внутренних циклов грани

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.

5.2.* 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

$$F - E + V + (F - L) = H, \quad (5.2.1)$$

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.

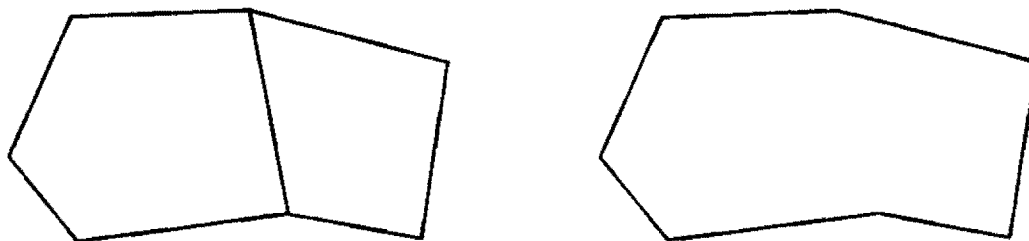


Рис. 5.2.1. Эйлерова характеристика не изменяется при ликвидации ребра оболочки

При желании оболочкой можно считать только показанный на рисунках фрагмент. На рис. 5.2.1 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

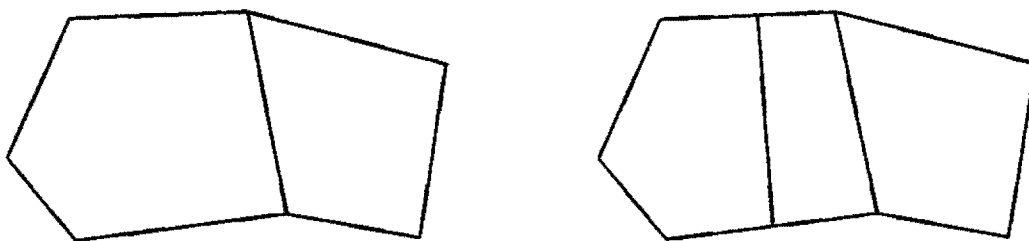


Рис. 5.2.2. Эйлерова характеристика не изменяется при добавлении ребра в оболочку

ребер увеличится на единицу. 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. 5.2.2.

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. Число вырезов

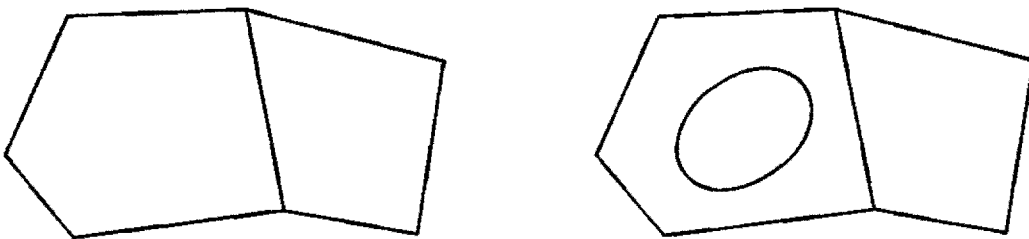


Рис. 5.2.3. Добавление грани и двух циклов в оболочку

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. 5.2.3 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). Эйлерова характеристика оболочки при этом также не изменится. Эйлерова характеристика оболочки не изменится, если мы преобразуем грань с двумя циклами в грань с одним циклом, что

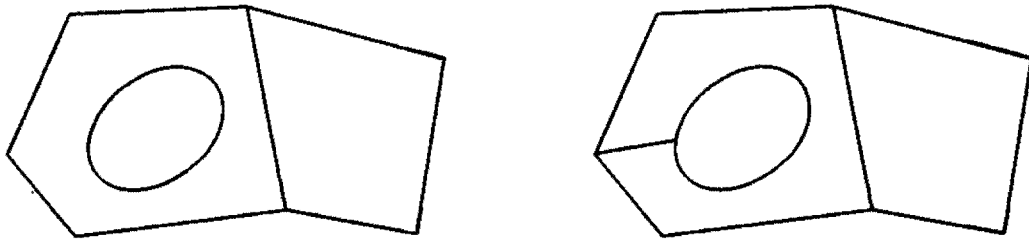


Рис. 5.2.4. Добавление ребра и ликвидация внутреннего цикла грани

приведено на рис. 5.2.4. Покажем это. Пусть новое ребро начинается и оканчивается в уже существующих вершинах, тогда число ребер увеличится на единицу, число циклов уменьшится на единицу, а эйлерова характеристика не изменится.

Все перечисленные модификации оболочки не изменяют ее эйлеровой характеристики. Это иллюстрирует то, что эйлерова характеристика не зависит от способа разбиения оболочки на грани, а зависит только от природы оболочки.

Рассмотрим еще один пример, показывающий, что эйлерова характеристика зависит от топологии оболочки, и не зависит от способа раскроя ее на грани. Возьмем

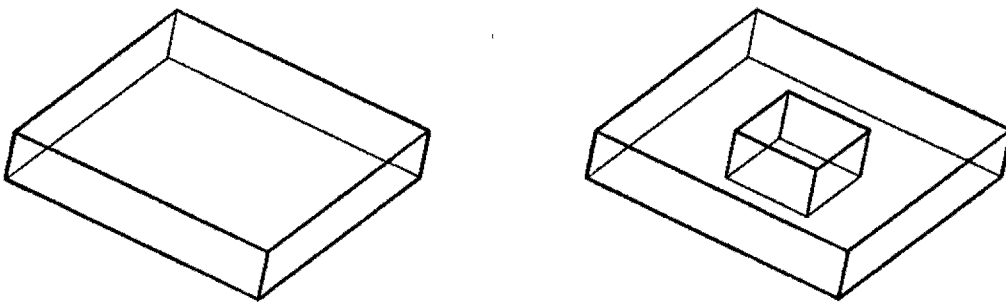


Рис. 5.2.5. Эйлерова характеристика оболочек различна

оболочку в форме четырехугольной призмы и превратим ее в оболочку в форме четырехугольной призмы с четырехугольным отверстием, что показано на рис. 5.2.5.

Исходная оболочки имела следующие числа граней, циклов, ребер и вершин: $F = 6$, $L = 6$, $E = 12$, $V = 8$, и ее эйлерова характеристика равна $H = 2$. Результирующая оболочки имеет: граней $F = 10$, циклов $L = 12$, ребер $E = 24$, вершин $V = 16$, а ее эйлерова характеристика равна $H = 0$. В двух гранях увеличилось число ци-

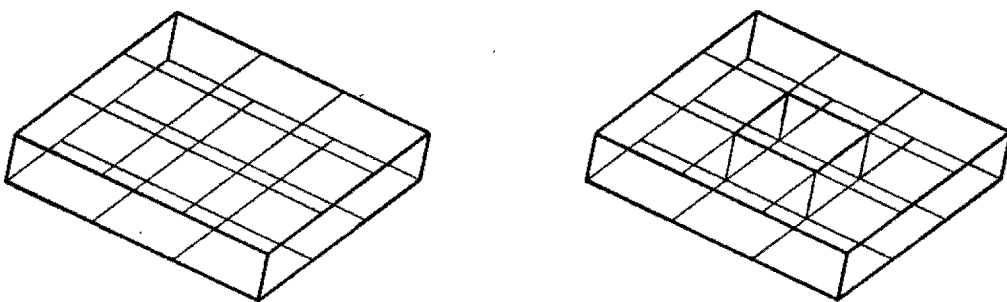


Рис. 5.2.6

клов (было по одному циклу, а стало по два). Эйлерова 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. 5.2.6.

Исходная оболочка имела следующие числа граней, циклов, ребер и вершин: $F = 22$, $L = 22$, $E = 52$, $V = 32$, и ее эйлерова характеристика равна $H = 2$. Результирующая оболочка имеет: граней $F = 24$, циклов $L = 24$, 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

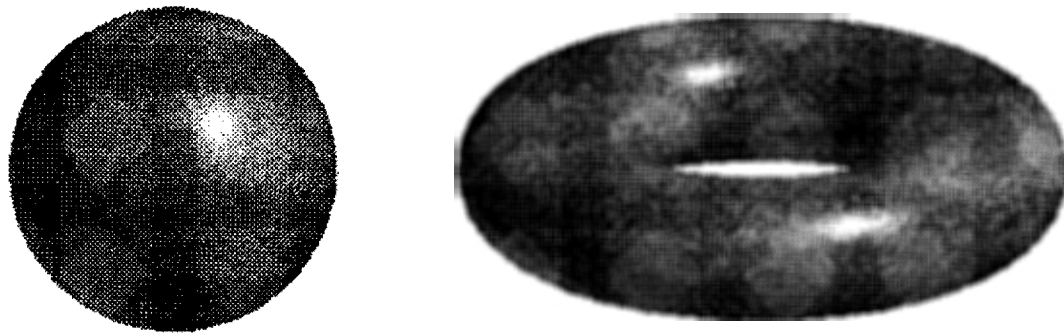


Рис. 5.2.7. Сфера и тор имеют различную топологию

углы, из первой оболочки получим сферу, а из второй — тор (рис. 5.2.7). 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*.

5.3.* 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 несклеиваемого 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. 5.3.1 приведены примеры плоских оболочек с различным числом циклов. Если

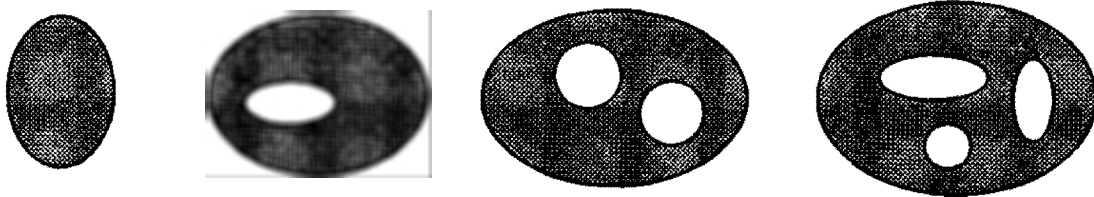


Рис. 5.3.1. Плоские оболочки различной связности

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 раскрыть 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 раскрытия ее в односвязную грань, потребуется, как минимум, две замкнутые кривые на ней. Способ раскрытия тороидальной оболочки показан на рис. 5.3.2.

Используя определение, получим, что связность сферы равна единице, а связность тора равна трем.

На замкнутой оболочке связности h можно построить $h - 1$ замкнутых кривых, которые не нарушают ее целостности (которые не разрезают ее на отдельные части), но нельзя построить h таких кривых. Для того, чтобы замкнутую оболочку связности



Рис. 5.3.2. Раскрой тороидальной оболочки

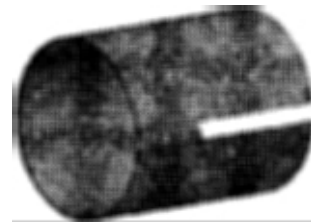


Рис. 5.3.3. Раскрой цилиндрической оболочки

h раскрыть 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. 5.3.4 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. 5.3.4 незамкнутых оболочек соответственно

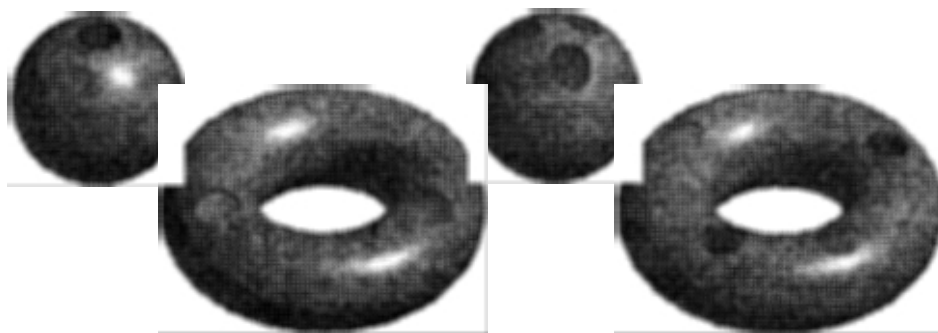


Рис. 5.3.4. Незамкнутые оболочки

равна 1, 4, 3, 6. 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*.

5.4.* 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. 5.4.1. 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 неориентируема. 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

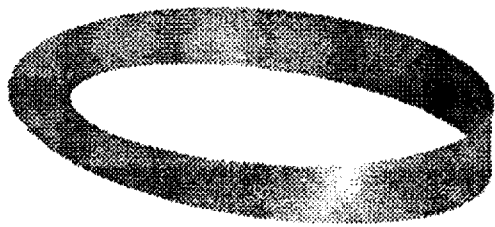


Рис. 5.4.1. Лист Мёбиуса — односторонняя незамкнутая оболочка

Рис. 5.4.2. Ориентируемая самопересекающаяся незамкнутая оболочка

направлении. 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. 5.4.2, самопересекающаяся, но не односторонняя.

Бутылка Клейна. Примером замкнутой односторонней оболочки является бутылка Клейна, которая показана на рис. 5.4.3. Бутылка Клейна имеет одну замкнутую линию

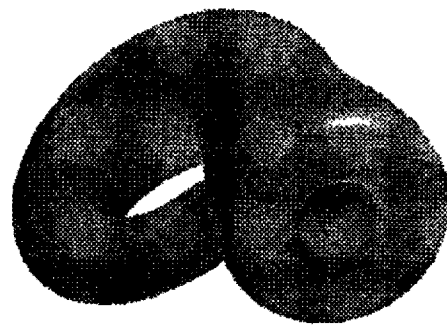
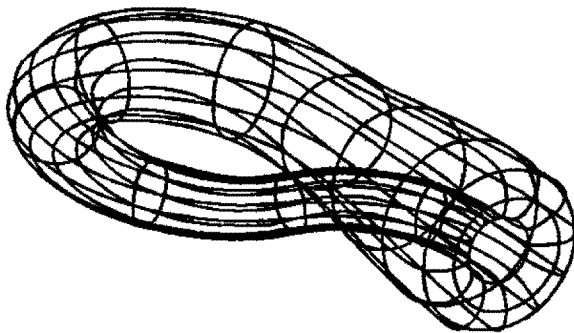


Рис. 5.4.3. Бутылка Клейна — односторонняя замкнутая оболочка

самопересечения. Она не может служить сосудом. Связность бутылки Клейна равна трем. Система линии, 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. 5.4.4. 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. 5.4.5.

Если мы начнем движение по поверхности гептаэдра из точки A , то, двигаясь по указанной на рис. 5.4.5 траектории, попадем в точку B , расположенную на другой стороне поверхности.

Таким образом, кроме связности оболочки характеризуются еще и ориентируемостью. Тор и бутылка Клейна обладают одинаковой связностью, обе замкнутые оболочки,

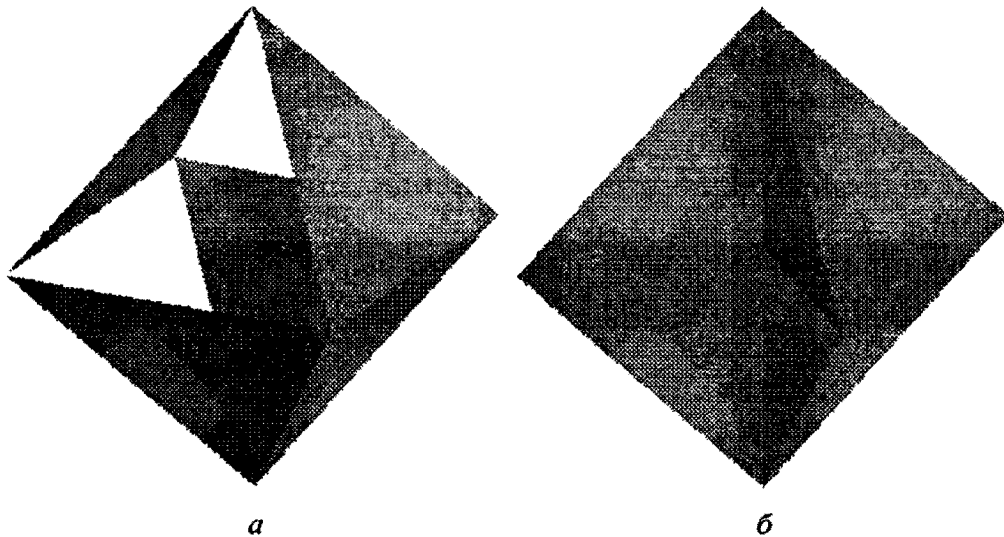


Рис. 5.4.4. Четыре грани октаэдра (а), грани на диагоналях октаэдра (б)

но тор является ориентируемой оболочкой, а бутылка Клейна — неориентируемой. Известно еще несколько неориентируемых оболочек. Замкнутые неориентируемые оболочки пересекают сами себя, и поэтому не могут быть использованы для моделирования

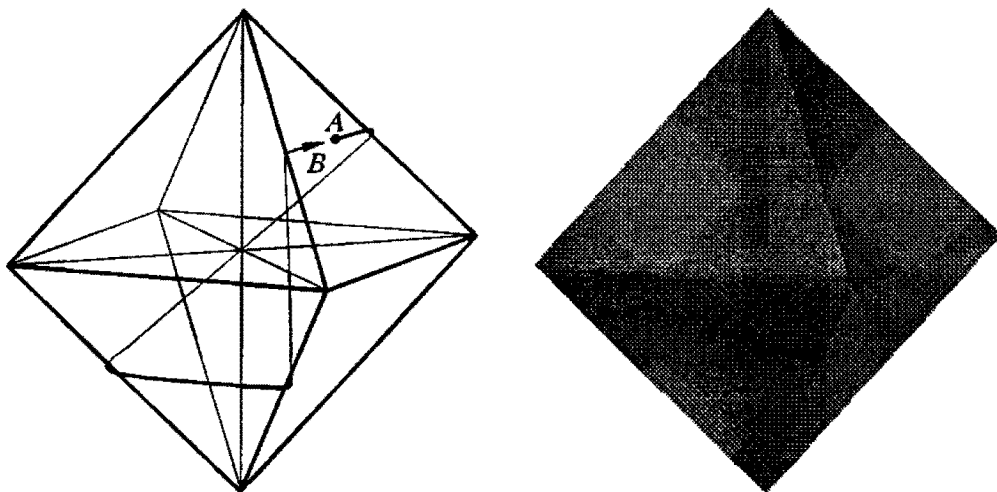


Рис. 5.4.5. Гептаэдр (семигранник) — самопересекающаяся односторонняя оболочка

деталей. Оболочки, с которыми мы сталкиваемся в реальности, являются ориентируемыми. Тем не менее, неориентируемые оболочки с математической точки зрения равноправны с ориентируемыми оболочками.

5.5. Оболочки для моделирования тел

Рассмотрим взаимно однозначное и непрерывное отображение одной оболочки на другую. При этом отображении соседние точки остаются соседними. Такое отображение может искажать 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

$$H = 3 - h. \quad (5.5.1)$$

Используя это соотношение, получим формулу, связывающую число граней F , число циклов L , число ребер E и число вершин V оболочки с ее связностью h :

$$F - E + V + (F - L) = 3 - h. \quad (5.5.2)$$

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 гирю, 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 $h = 2G + 1$. Оболочка реальной детали топологически эквивалентна сфере с некоторым числом ручек. На рис. 5.5.1 приведена сфера с четырьмя ручками.

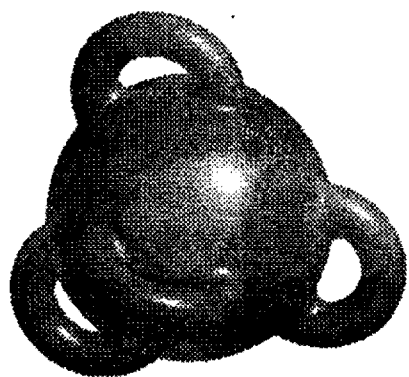


Рис. 5.5.1. Сфера с четырьмя ручками

Сфера с G ручками имеет связность $h = 2G + 1$. Оболочка реальной детали топологически эквивалентна сфере с некоторым числом ручек. На рис. 5.5.1 приведена 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 $h = 2G + 1$, и получим формулу, связывающую число граней F , ребер E , вершин V и циклов L с характерной величиной G ,

$$F - E + V + (F - L) - 2(1 - G) = 0. \quad (5.5.3)$$

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:

$$G = 1 - F + \frac{E + L - V}{2}. \quad (5.5.4)$$

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 (5.5.3), а для объекта с S замкнутыми оболочками формула Эйлера–Пуанкаре (5.5.3) примет вид

$$F - E + V + (F - L) - 2(S - G) = 0, \quad (5.5.5)$$

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

$$F - E + V + (F - L) + h = 2. \quad (5.5.6)$$

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. 5.5.2, некорректна, так как в ребре AB стыкуются четыре грани.

Такая оболочка топологически должна быть представлена одним из способов, приведенных на рис. 5.5.3 и 5.5.4. 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 пере-

сечь все стыкующиеся в вершине ребра и посетить все примыкающие к вершине грани. Такие вершины будем называть простыми. Оболочка, приведенная на

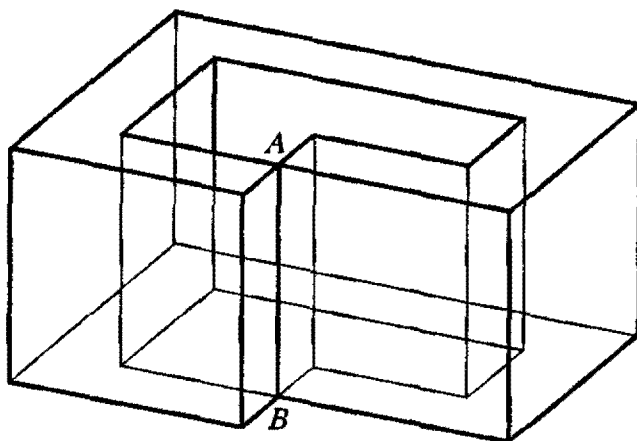


Рис. 5.5.2. В каждом ребре должны пересекаться только две грани

рис. 5.5.5, 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. 5.5.6.

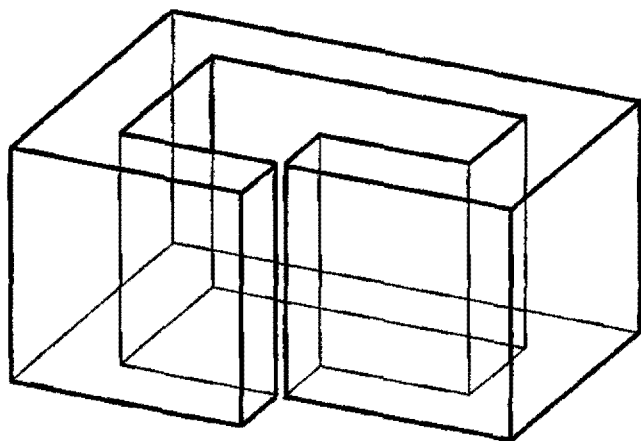


Рис. 5.5.3. Корректная оболочка

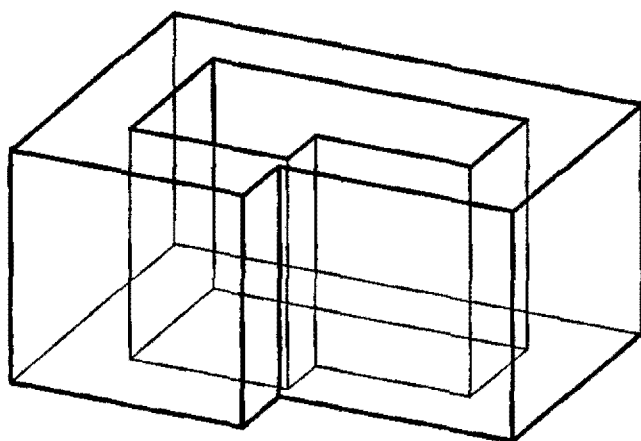


Рис. 5.5.4. Корректная оболочка

Таким образом, для моделирования мы будем использовать оболочки, вершины которых являются простыми, грани пересекаются по ребрам, причем

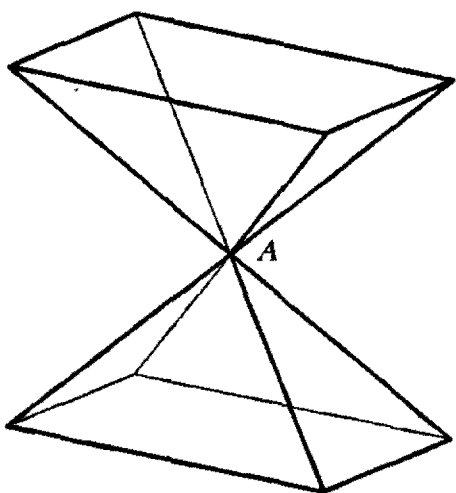


Рис. 5.5.5. Некорректная оболочка

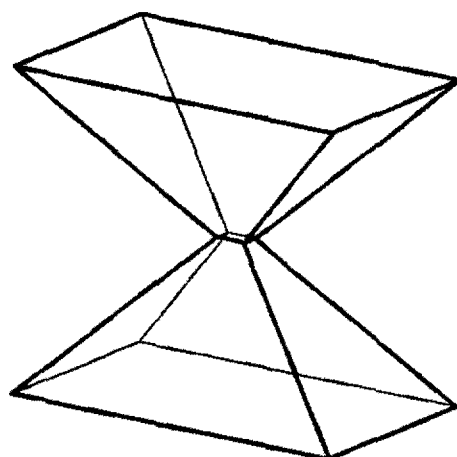


Рис. 5.5.6. Корректная оболочка

в каждом ребре стыкуются только две грани. Такие оболочки будем называть *однородными*. У однородных оболочек все топологические элементы построены по единым правилам.

5.6. 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.

Глава 6

МОДЕЛИРОВАНИЕ ТЕЛ

6.1. 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 делают 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. Но для редактирования тела необходима

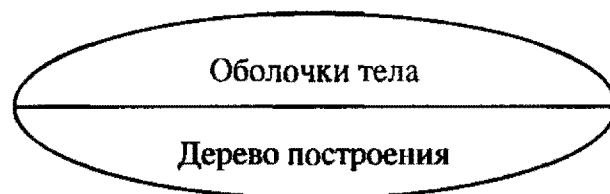


Рис. 6.1.1. Математическая модель тела

информация о последовательности и способах его построения, поэтому в модель тела включают еще и *дерево построения* (или протокол построения) тела (рис. 6.1.1).

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 коррект-

ного 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 $r(u, v)$, двумерную линию обозначим векторной функцией $l_{uv}(t) = [u(t) \ v(t)]^T$, а на грани, базирующейся на поверхности $s(a, b)$, двумерную линию обозначим векторной функцией $l_{ab}(t) = [a(t) \ b(t)]^T$. Таким образом, линию пересечения граней, на которой базируется ребро, будем записывать в виде

$$\begin{aligned} l_{uv}(t) &= \begin{bmatrix} u(t) \\ v(t) \end{bmatrix} \equiv [u(t) \ v(t)]^T, & l_{uv}(t) \in r(u, v), \\ l_{ab}(t) &= \begin{bmatrix} a(t) \\ b(t) \end{bmatrix} \equiv [a(t) \ b(t)]^T, & l_{ab}(t) \in s(a, b), \end{aligned} \quad (6.1.1)$$

$$t_{\min} \leq t \leq t_{\max}.$$

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

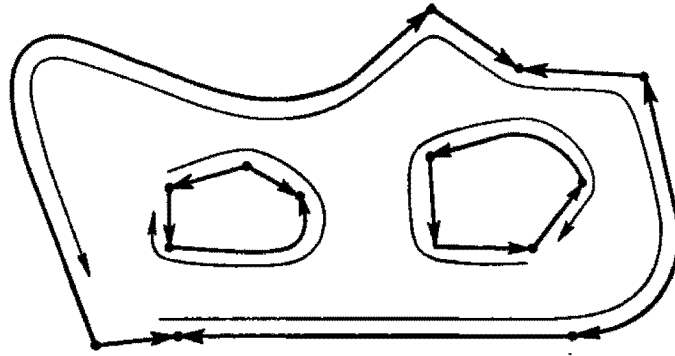


Рис. 6.1.2. Грань и ее структура данных: поверхность, ее ориентация и список циклов

часовой стрелке, если смотреть навстречу нормали грани. По каждому циклу можно построить контур на поверхности, который описывает границу грани. На рис. 6.1.2 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. Ребро должно

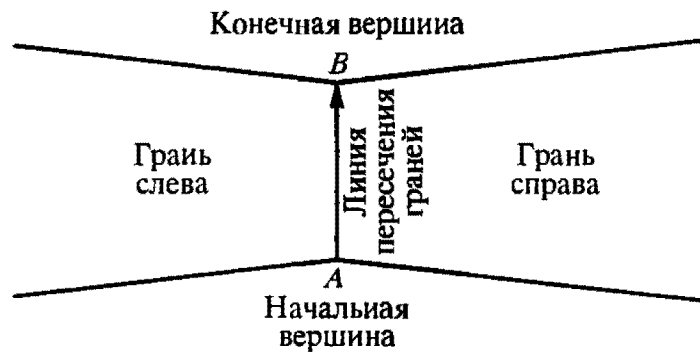


Рис. 6.1.3. Ребро AB и его структура данных

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. 6.1.3 показано ребро, начинающееся в вершине A и оканчивающееся в вершине B .

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.

6.2. 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 браться 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 $\mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z$.

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 \mathbf{i}_x тело имеет длину, равную x , в направлении орта \mathbf{i}_y — длину, равную y , а в направлении орта \mathbf{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. Контур состоит из отрезков прямых (2.2.3). 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 двумерной линии на ней. Прямоугольная призма с ориентацией циклов граней показана на рис. 6.2.1.

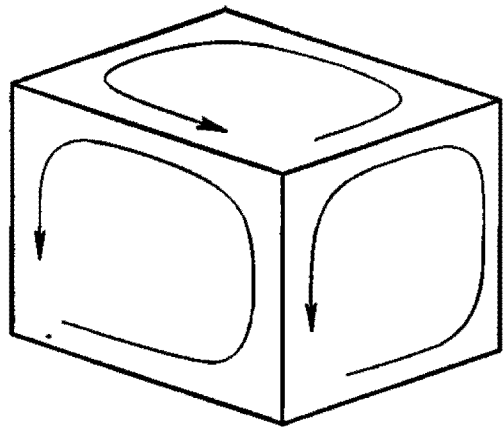


Рис. 6.2.1. Ориентация циклов граней призматического тела

Грани тела будут описываться поверхностями:

$$\begin{aligned}
 \mathbf{r}_1(u_1, v_1) &= \mathbf{p} + u_1 \mathbf{i}_x + v_1 \mathbf{i}_y, & 0 \leq u_1 \leq x, & & 0 \leq v_1 \leq y, \\
 \mathbf{r}_2(u_2, v_2) &= \mathbf{p} + u_2 \mathbf{i}_y + v_2 \mathbf{i}_z, & 0 \leq u_2 \leq y, & & 0 \leq v_2 \leq z, \\
 \mathbf{r}_3(u_3, v_3) &= \mathbf{p} + u_3 \mathbf{i}_x + v_3 \mathbf{i}_z, & 0 \leq u_3 \leq x, & & 0 \leq v_3 \leq z, \\
 \mathbf{r}_4(u_4, v_4) &= \mathbf{p} + z \mathbf{i}_z + u_4 \mathbf{i}_x + v_4 \mathbf{i}_y, & 0 \leq u_4 \leq x, & & 0 \leq v_4 \leq y, \\
 \mathbf{r}_5(u_5, v_5) &= \mathbf{p} + x \mathbf{i}_x + u_5 \mathbf{i}_y + v_5 \mathbf{i}_z, & 0 \leq u_5 \leq y, & & 0 \leq v_5 \leq z, \\
 \mathbf{r}_6(u_6, v_6) &= \mathbf{p} + y \mathbf{i}_y + u_6 \mathbf{i}_x + v_6 \mathbf{i}_z, & 0 \leq u_6 \leq x, & & 0 \leq v_6 \leq z.
 \end{aligned} \tag{6.2.1}$$

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. Например, ребро между первой и второй гранями описывается линией пересечения этих граней, состоящей из двух

двухмерных отрезков:

$$\begin{aligned} l_{uv1}(t) &= [u_1(t) \ v_1(t)]^T = [0 \ yt]^T, & l_{uv1}(t) &\in r_1(u_1, v_1), \\ l_{uv2}(t) &= [u_2(t) \ v_2(t)]^T = [yt \ 0]^T, & l_{uv2}(t) &\in r_2(u_2, v_2), \\ & & 0 &\leq t \leq 1, \end{aligned} \quad (6.2.2)$$

и двух поверхностей. Выражение $l_{uv1}(t) \in r_1(u_1, v_1)$ означает, что отрезок $l_{uv1}(t)$ лежит на поверхности $r_1(u_1, v_1)$, а выражение $l_{uv2}(t) \in r_2(u_2, v_2)$ означает, что отрезок $l_{uv2}(t)$ лежит на поверхности $r_2(u_2, v_2)$. Пусть направление ребра совпадает с направлением кривой (6.2.2), 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. 6.2.2). Начало местной системы координат \mathbf{p} поместим в центр одного из торцев 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 $v = v_{\min}$ или $v = v_{\max}$ на цилиндрической поверхности. Напомним, что обе линии на поверхности, составляющие линию пересечения, должны иметь одинаковую параметрическую длину.

Боковая грань тела базируется на цилиндрической поверхности

$$\begin{aligned} r_1(u_1, v_1) &= \mathbf{p} + r \cos u_1 \mathbf{i}_x + r \sin u_1 \mathbf{i}_y + hv_1 \mathbf{i}_z, \\ 0 &\leq u_1 \leq 2\pi, \quad 0 \leq v_1 \leq 1. \end{aligned} \quad (6.2.3)$$

Нормали поверхности (6.2.3) 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

$$r_2(u_2, v_2) = \mathbf{p} + u_2 \mathbf{i}_x + v_2 \mathbf{i}_y, \quad (6.2.4)$$

$$r_3(u_3, v_3) = \mathbf{p} + u_3 \mathbf{i}_x + v_3 \mathbf{i}_y + h \mathbf{i}_z, \quad (6.2.5)$$

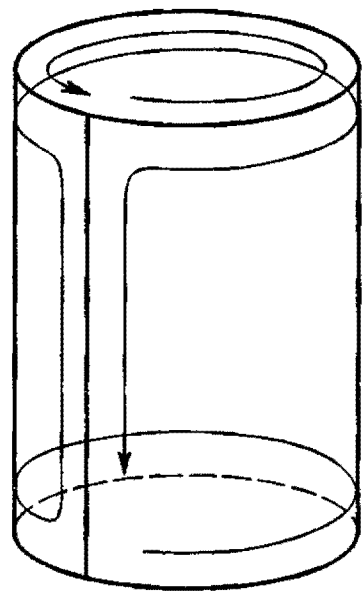


Рис. 6.2.2. Ориентация циклов граней цилиндрического тела

где параметры u_2, v_2 лежат внутри окружности $l_{uv2}(t) = [r \cos t \ r \sin t]^\top$, $0 \leq t \leq 2\pi$, на плоскости $\mathbf{r}_2(u_2, v_2)$, а параметры u_3, v_3 лежат внутри окружности $l_{uv3}(t) = [r \cos t \ r \sin t]^\top$, $0 \leq t \leq 2\pi$, на плоскости $\mathbf{r}_3(u_3, v_3)$. Нормали поверхности (6.2.4) и ее грани противоположны по направлению, а нормали поверхности (6.2.5) 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 двухмерными кривыми

$$\begin{aligned} l_{uv01}(t) &= [u_1(t) \ v_1(t)]^\top = [0 \ ht]^\top, & l_{uv01}(t) &\in \mathbf{r}_1(u_1, v_1), \\ l_{uv02}(t) &= [u_1(t) \ v_1(t)]^\top = [2\pi \ ht]^\top, & l_{uv02}(t) &\in \mathbf{r}_1(u_1, v_1), \\ & & 0 &\leq t \leq 1, \end{aligned} \quad (6.2.6)$$

являющимися линиями $u = u_{\min}$ и $u = u_{\max}$ на цилиндрической поверхности.

Ребро между боковой гранью и основанием (6.2.4) описывается линией пересечения этих граней, состоящей из двух двухмерных кривых (отрезка и окружности)

$$\begin{aligned} l_{uv1}(t) &= [u_1(t) \ v_1(t)]^\top = [t \ 0]^\top, & l_{uv1}(t) &\in \mathbf{r}_1(u, v_1), \\ l_{uv2}(t) &= [u_2(t) \ v_2(t)]^\top = [r \cos t \ r \sin t]^\top, & l_{uv2}(t) &\in \mathbf{r}_2(u_2, v_2), \\ & & 0 &\leq t \leq 2\pi. \end{aligned} \quad (6.2.7)$$

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 (6.2.7), что отметим соответствующим знаком 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 (6.2.5) описывается линией пересечения этих граней, состоящей из двух двухмерных кривых (отрезка и окружности)

$$\begin{aligned} l_{uv1}(t) &= [u_1(t) \ v_1(t)]^\top = [t \ h]^\top, & l_{uv1}(t) &\in \mathbf{r}_1(u_1, v_1), \\ l_{uv3}(t) &= [u_3(t) \ v_3(t)]^\top = [r \cos t \ r \sin t]^\top, & l_{uv3}(t) &\in \mathbf{r}_3(u_3, v_3), \\ & & 0 &\leq t \leq 2\pi. \end{aligned} \quad (6.2.8)$$

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 (6.2.8), что отметим соответствующим знаком 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 (6.2.7) — флаг положительный,
ребро на базе кривой (6.2.6) — флаг положительный,
ребро на базе кривой (6.2.8) — флаг отрицательный,
ребро на базе кривой (6.2.6) — флаг отрицательный.

Цилиндрическое тело и ориентация циклов его граней показаны на рис. 6.2.2.

Для построения граней тела достаточно знать местную систему координат p, i_x, i_y, i_z , радиус r 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 (3.2.10)

$$\begin{aligned} \mathbf{r}_1(u_1, v_1) &= \mathbf{p} + (r + hv_1 \operatorname{tg} \gamma)(\cos u_1 \mathbf{i}_x + \sin u_1 \mathbf{i}_y) + hv_1 \mathbf{i}_z, \\ 0 &\leq u_1 \leq 2\pi, \quad 0 \leq v_1 \leq 1. \end{aligned} \quad (6.2.9)$$

Грани основания базируются на ограниченных окружностями поверхностях

$$\mathbf{r}_2(u_2, v_2) = \mathbf{p} + u_2 \mathbf{i}_x + v_2 \mathbf{i}_y, \quad (6.2.10)$$

$$\mathbf{r}_3(u_3, v_3) = \mathbf{p} + u_3 \mathbf{i}_x + v_3 \mathbf{i}_y + h \mathbf{i}_z, \quad (6.2.11)$$

где параметры u_2, v_2 лежат внутри окружности

$$l_{uv2}(t) = [r \cos t \quad r \sin t]^T, \quad 0 \leq t \leq 2\pi,$$

а параметры u_3, v_3 лежат внутри окружности

$$l_{uv3}(t) = [(r + h \operatorname{tg} \gamma) \cos t \quad (r + h \operatorname{tg} \gamma) \sin t]^T, \quad 0 \leq t \leq 2\pi.$$

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 p, i_x, i_y, i_z , радиус r одного из оснований конуса, высоту h и угол конусности γ .

Сферическое тело. Сферическая поверхность (3.2.3) может служить оболочкой сферического тела, но эта оболочка не является замкнутой, так как имеет два отверстия нулевого радиуса в полюсах. Начало местной системы координат 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 двумерные кривые на сферической поверхности в полюсах имеют ненулевую длину. Цикл грани сферического тела составляют три ребра, одно из которых является швом

$$\begin{aligned} l_{uv1}(t) &= [u(t) \ v(t)]^T = [0 \ t]^T, & l_{uv1}(t) &\in \mathbf{r}(u, v), \\ l_{uv2}(t) &= [u(t) \ v(t)]^T = [2\pi \ t]^T, & l_{uv2}(t) &\in \mathbf{r}(u, v), \\ & & -\frac{\pi}{2} &\leq t \leq \frac{\pi}{2}, \end{aligned}$$

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. 6.2.3.

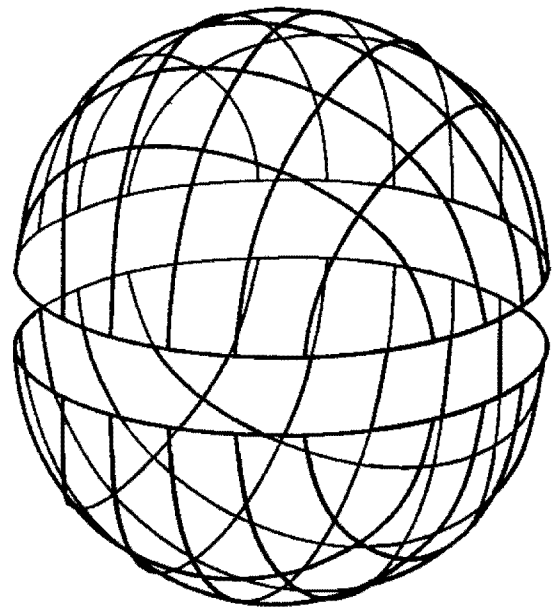
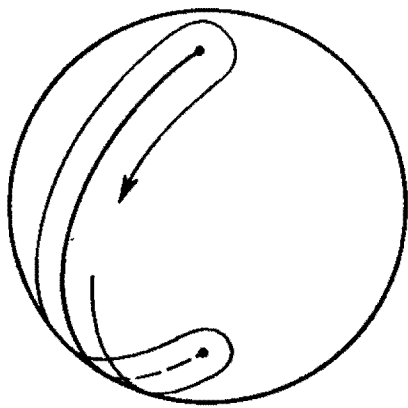


Рис. 6.2.3. Ориентация цикла грани сферического тела

Рис. 6.2.4. Построение сферического тела по двум полусферам

Сферическое тело можно построить из двух полусфер, описываемых векторными функциями

$$\mathbf{r}_1(u, v) = \mathbf{p} + x\mathbf{i}_x + y\mathbf{i}_y + \sqrt{r^2 - x^2 - y^2}\mathbf{i}_z, \quad (6.2.12)$$

$$\mathbf{r}_2(u, v) = \mathbf{p} + x\mathbf{i}_x + y\mathbf{i}_y - \sqrt{r^2 - x^2 - y^2}\mathbf{i}_z, \quad (6.2.13)$$

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 (6.2.12) описывает верхнюю полусферу, а поверхность (6.2.13) описывает нижнюю полусферу. Эти поверхности аналогичны поверхности (3.6.5). 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:

$$l_{uv11}(t) = [u(t) \ v(t)]^\top = \left[\frac{\pi t}{2} \ 0 \right]^\top, \quad l_{uv11}(t) \in \mathbf{r}_1(u, v),$$

$$l_{uv21}(t) = [u(t) \ v(t)]^\top = \left[\frac{\pi t}{2} \ 0 \right]^\top, \quad l_{uv21}(t) \in \mathbf{r}_2(u, v),$$

$$0 \leq t \leq 1;$$

$$l_{uv12}(t) = [u(t) \ v(t)]^\top = \left[\frac{\pi}{2} \ \frac{\pi t}{2} \right]^\top, \quad l_{uv12}(t) \in \mathbf{r}_1(u, v),$$

$$l_{uv22}(t) = [u(t) \ v(t)]^\top = \left[\frac{\pi}{2} \ \frac{\pi t}{2} \right]^\top, \quad l_{uv22}(t) \in \mathbf{r}_2(u, v),$$

$$0 \leq t \leq 1;$$

$$l_{uv13}(t) = [u(t) \ v(t)]^\top = \left[\frac{\pi(1-t)}{2} \ \frac{\pi}{2} \right]^\top, \quad l_{uv13}(t) \in \mathbf{r}_1(u, v),$$

$$l_{uv23}(t) = [u(t) \ v(t)]^\top = \left[\frac{\pi(1-t)}{2} \ \frac{\pi}{2} \right]^\top, \quad l_{uv23}(t) \in \mathbf{r}_2(u, v),$$

$$0 \leq t \leq 1;$$

$$l_{uv14}(t) = [u(t) \ v(t)]^\top = \left[0 \ \frac{\pi(1-t)}{2} \right]^\top, \quad l_{uv14}(t) \in \mathbf{r}_1(u, v),$$

$$l_{uv24}(t) = [u(t) \ v(t)]^\top = \left[0 \ \frac{\pi(1-t)}{2} \right]^\top, \quad l_{uv24}(t) \in \mathbf{r}_2(u, v),$$

$$0 \leq t \leq 1.$$

Поверхности (6.2.12) и (6.2.13) для построения сферического тела приведены на рис. 6.2.4.

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

$$\mathbf{r}(u, v) = \mathbf{p} + (R + r \cos v) \cos u \mathbf{i}_x + (R + r \cos v) \sin u \mathbf{i}_y + r \sin v \mathbf{i}_z, \quad (6.2.14)$$

$$0 \leq u \leq 2\pi, \quad -\pi \leq v \leq \pi,$$

два ребра

$$l_{uvR1}(t) = [u(t) \ v(t)]^\top = [t \ -\pi]^\top, \quad (6.2.15)$$

$$l_{uvR2}(t) = [u(t) \ v(t)]^\top = [t \ \pi]^\top,$$

$$0 \leq t \leq 2\pi,$$

$$l_{uvr1}(t) = [u(t) \ v(t)]^\top = [0 \ t]^\top, \quad (6.2.16)$$

$$l_{uvr2}(t) = [u(t) \ v(t)]^\top = [2\pi \ t]^\top,$$

$$-\pi \leq t \leq \pi,$$

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 (6.2.15) — флаг положительный,*
- ребро на базе кривой (6.2.16) — флаг положительный,*
- ребро на базе кривой (6.2.15) — флаг отрицательный,*
- ребро на базе кривой (6.2.16) — 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. 6.2.5.

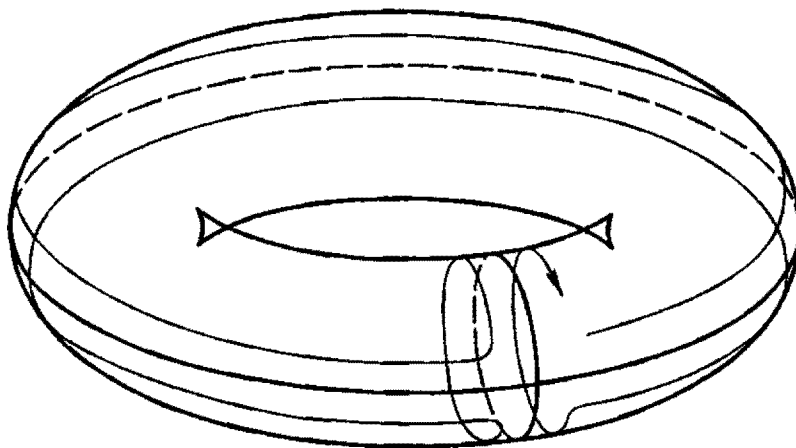


Рис. 6.2.5. Ориентация цикла грани тороидального тела

Для построения граней тела достаточно знать местную систему координат \mathbf{p} , \mathbf{i}_x , \mathbf{i}_y , \mathbf{i}_z , больший 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.

6.3. 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

$$r(x, y) = p + x i_x + y i_y, \quad x, y \in \Omega, \quad (6.3.1)$$

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 $c_i(u) = [x_i(u) \ y_i(u)]^T$, $u_{i \min} \leq u \leq u_{i \max}$, $i = 1, 2, \dots, n$. Радиус-вектор контура описывается формулой (2.12.7). Каждая кривая контура в пространстве описывается радиус-вектором

$$c_i(u) = p + x_i(u) i_x + y_i(u) i_y, \quad u_{i \min} \leq u \leq u_{i \max}. \quad (6.3.2)$$

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 (3.15.1). 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 $g(v) = p + vhd$, $0 \leq v \leq 1$, то мы получим тело выдавливания. Одна из его торцевых граней будет описываться ограниченной плоскостью (6.3.1). Другая его торцевая грань будет описываться аналогичной плоскостью, только с началом в точке $p' = p + hd$. Боковая грань тела выдавливания, соответствующая кривой контура $c_i(u)$, описывается поверхностями

$$\begin{aligned} r_i(u, v) &= p + x_i(u)i_x + y_i(u)i_y + vhd = c_i(u) + vhd, \\ u_{i \min} &\leq u \leq u_{i \max}, \quad 0 \leq v \leq 1. \end{aligned} \quad (6.3.3)$$

Тело выдавливания, построенное по замкнутому контуру, приведено на рис. 6.3.1. Для контура можно построить эквидистантный контур. По заданному контуру и эквидистантному к нему контуру может быть построено тело

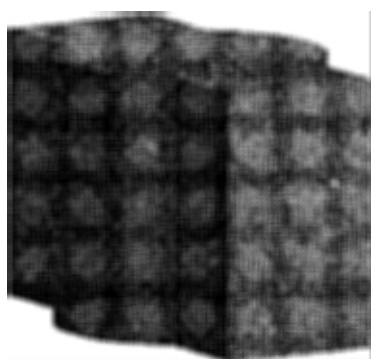


Рис. 6.3.1. Тело выдавливания

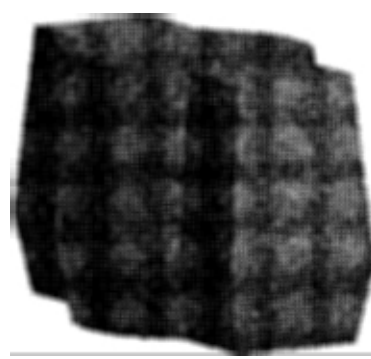


Рис. 6.3.2. Тело выдавливания с уклоном

выдавливания с уклоном. Угол уклона определяется эквидистантой и длиной выдавливания. Тело выдавливания с уклоном приведено на рис. 6.3.2. Выдавливание выполнено в обе стороны от плоскости (6.3.1). Боковые грани тела с уклоном построены на линейчатых поверхностях.



Рис. 6.3.3. Тонкостенное тело выдавливания

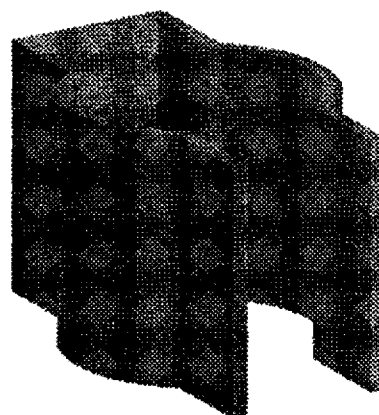


Рис. 6.3.4. Тонкостенное тело выдавливания

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. 6.3.3 и 6.3.4.

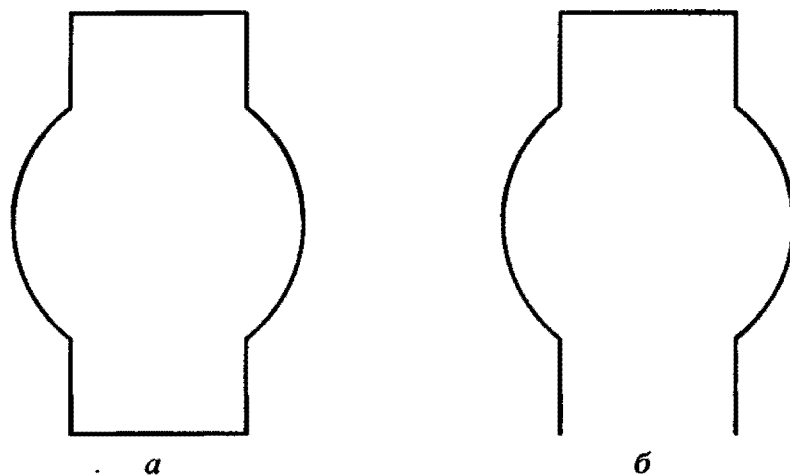


Рис. 6.3.5. Замкнутая (а) и незамкнутая (б) составная кривая

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 p_0 , а ее направление характеризуется единичным вектором \mathbf{i} . Тогда боковая грань тела вращения, соответствующая кривой $c_i(u)$, будет описываться поверхностью

$$\begin{aligned} \mathbf{r}_i(u, v) &= \mathbf{p}_0 + \mathbf{r}_{3i}(u) + \cos v \mathbf{r}_{1i}(u) + \sin v \mathbf{r}_{2i}(u), \\ u_{i \min} &\leq u \leq u_{i \max}, \quad 0 \leq v \leq \alpha, \end{aligned} \quad (6.3.4)$$

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 $c_i(u)$ описывается радиус-вектором (6.3.2).

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. 6.3.6. Тело вращения с топологией тора, построенное по замкнутому контуру, приведено на рис. 6.3.7.

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. 6.3.8 и 6.3.9.

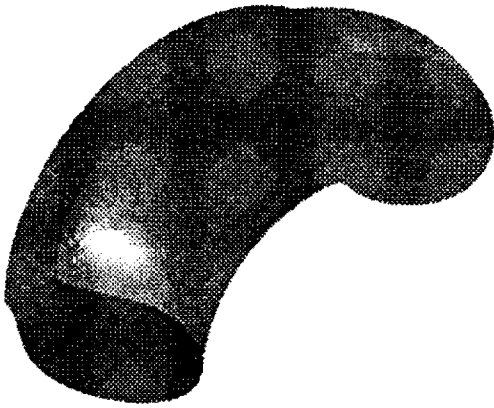


Рис. 6.3.6. Тело вращения

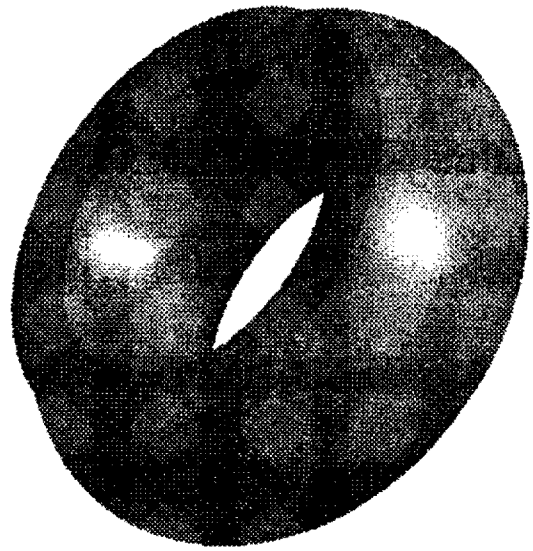


Рис. 6.3.7. Тело вращения с топологией тора

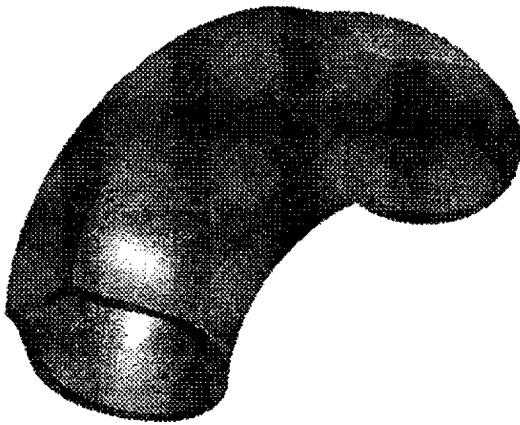


Рис. 6.3.8. Тонкостенное тело вращения

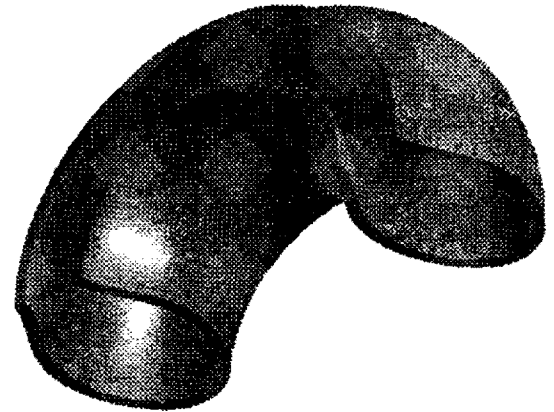


Рис. 6.3.9. Тонкостенное тело вращения

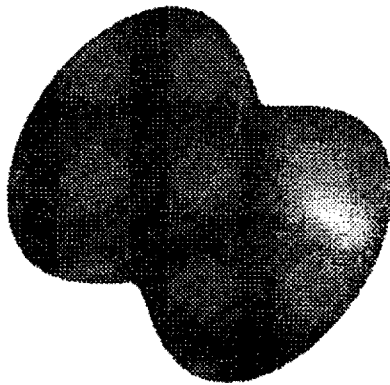


Рис. 6.3.10. Сфероид

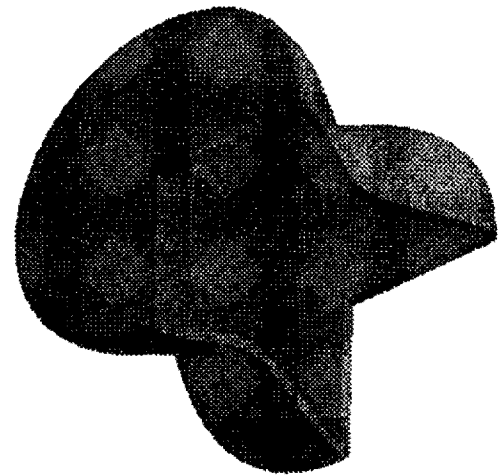


Рис. 6.3.11. Тонкостенный сфероид

По незамкнутой составной плоской кривой могут быть построены тела вращения, приведенные на рис. 3.6.10 и 3.6.11. 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. 3.6.10 и 3.6.11, 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 (6.3.2) и описываются поверхностью (3.4.3)

$$\begin{aligned} \mathbf{r}_i(u, v) &= \mathbf{g}(v) + (\mathbf{c}_i(u) - \mathbf{g}(v_{\min}) - \mathbf{h}), \\ u_{i \min} &\leq u \leq u_{i \max}, \quad v_{\min} \leq v \leq v_{\max}, \end{aligned} \quad (6.3.5)$$

где \mathbf{h} — вектор привязки образующей к направляющей. Вектор \mathbf{h} смещает кривую $\mathbf{c}_i(u)$ 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 (6.3.2) и описываются поверхностью (3.4.7)

$$\begin{aligned} \mathbf{r}_i(u, v) &= \mathbf{g}(v) + \mathbf{M}(v) \cdot (\mathbf{c}_i(u) - \mathbf{g}(v_{\min}) - \mathbf{h}), \\ u_{i \min} &\leq u \leq u_{i \max}, \quad v_{\min} \leq v \leq v_{\max}, \end{aligned} \quad (6.3.6)$$

где $\mathbf{M}(v)$ — матричная функция, определяемая формулой (3.4.8).

Если направляющая кривая замкнута, то получим кинематическое тело с топологией тора. Если направляющая кривая не замкнута, то получим кинематическое тело с топологией призмы. Боковые грани кинематического тела

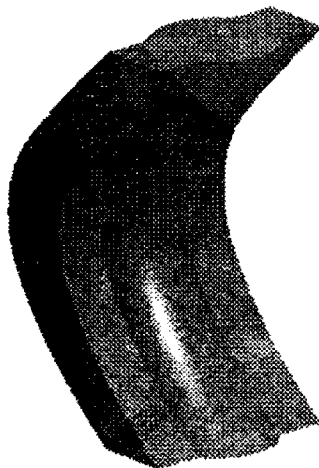


Рис. 6.3.12. Тело сдвига

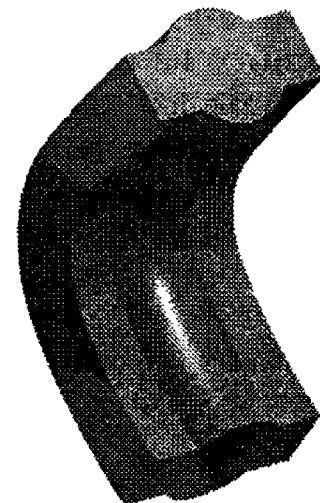


Рис. 6.3.13. Тело заметания

(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. 6.3.12. Тело заметания приведено на рис. 6.3.13.

На рис. 6.3.14 приведено тонкостенное кинематическое тело, построенное по незамкнутой образующей и замкнутой направляющей.

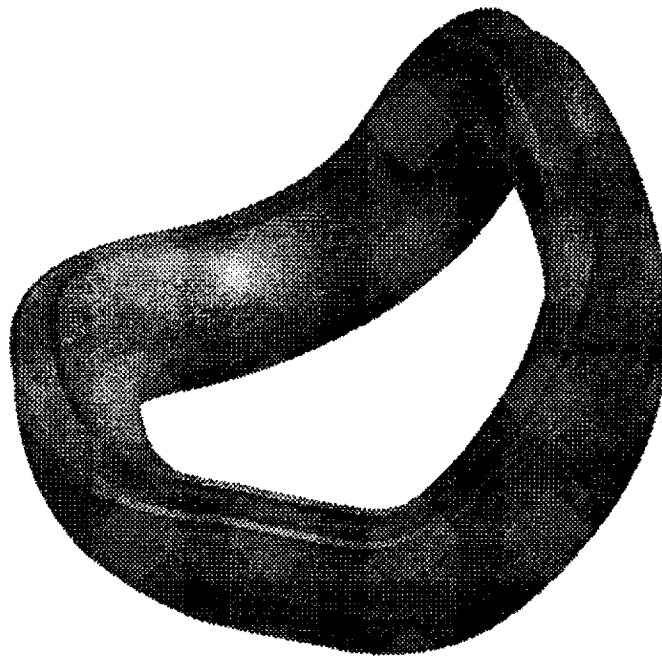


Рис. 6.3.14. Тонкостенное тело заметания

Для построения граней рассмотренных тел нужно знать двухмерный контур $c(t)$, местную его плоскость p , i_x , i_y , направляющую кривую $g(v)$ и другие необходимые параметры (толщину стенки для тонкостенного тела, способ движения контура — плоскопараллельный или ортогональный).

6.4. Построение тела по плоским сечениям

Многие детали можно построить по их плоским сечениям в определенных местах. Пусть имеется несколько плоских контуров $c_i(t)$, $i = 0, 1, 2, \dots, n$, 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 (3.7.1) или поверхности (3.7.2). Их количество равно количеству кривых в контуре. В каждой боковой

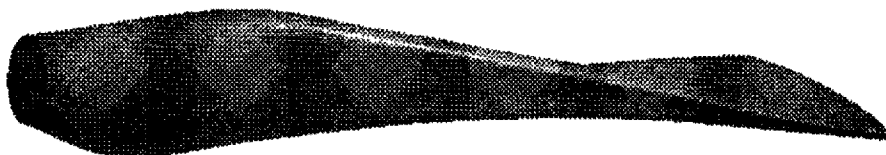


Рис. 6.4.1. Тело, построенное по плоским сечениям

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. 6.4.1.

В общем случае все контуры, за исключением торцевых контуров, могут быть не плоскими. Тело, построенное по плоским сечениям, может быть замкнутым

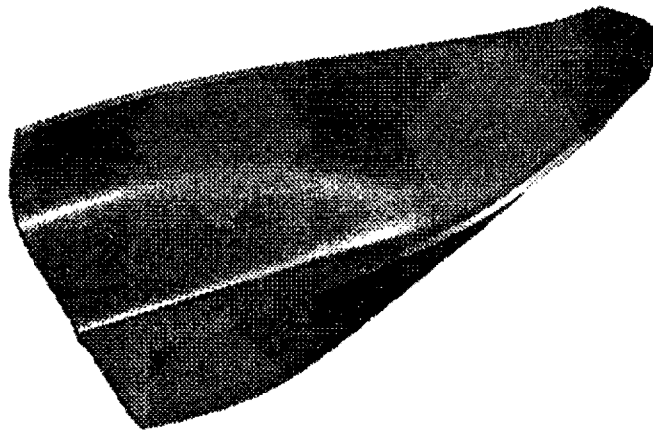


Рис. 6.4.2. Тело, построенное по плоским сечениям

и иметь топологию тора. Тело, построенное по сечениям, описываемым незамкнутыми линиями, приведено на рис. 6.4.2. По незамкнутым линиям в любом случае строятся замкнутые контуры, в которых участвуют эквидистантные кривые.

6.5. 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

$$\mathbf{r}(u, v) = \mathbf{b}(u, v) + hm(u, v), \quad (6.5.1)$$

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 (3.5.1), одной базовой линией которых является граничная линия на поверхности $\mathbf{b}(u, v)$, а второй — соответствующая ей линия на

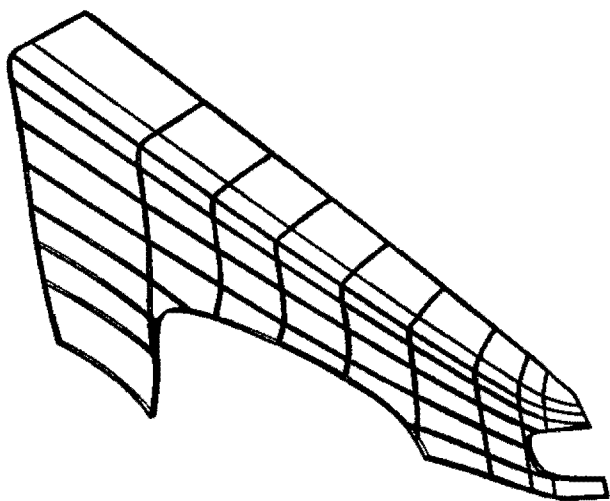


Рис. 6.5.1. Тело в форме листа

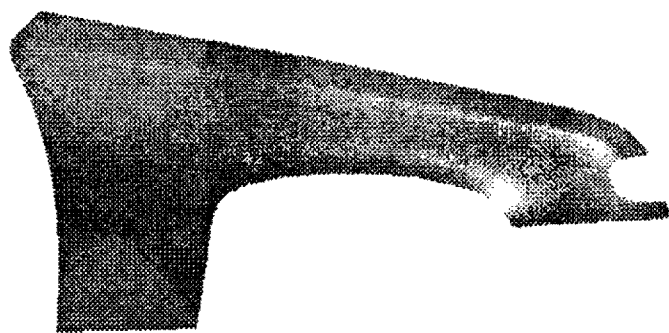


Рис. 6.5.2. Тело в форме листа

поверхности $\mathbf{r}(u, v)$. 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.

6.6. 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 и 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. Результатом

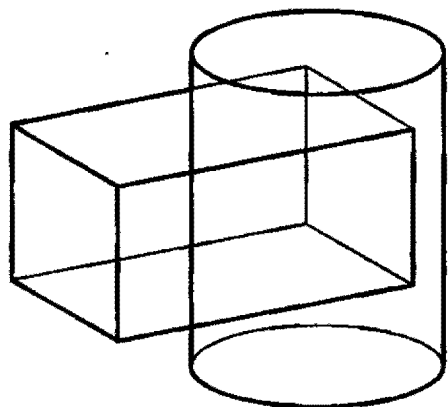


Рис. 6.6.1. Два исходных тела

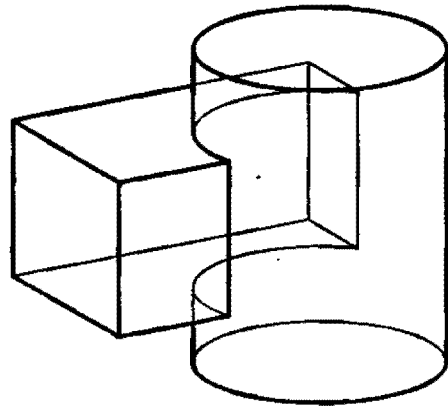


Рис. 6.6.2. Объединение тел

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. 6.6.1 приведены исходные для булевой операции тела. На рис. 6.6.2 приведен результат операции объединения тел, на рис. 6.6.3 приведен результат операции пересечения, на рис. 6.6.4 и 6.6.5 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^- . При выворачивании тела наизнанку внутренние стороны

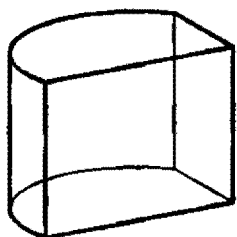


Рис. 6.6.3. Пересечение тел

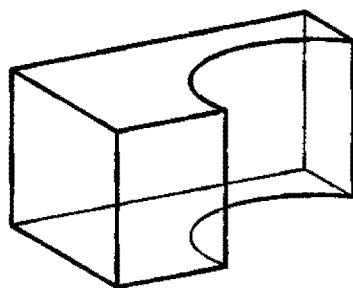


Рис. 6.6.4. Разность тел

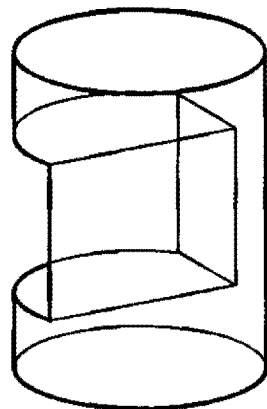


Рис. 6.6.5. Разность тел

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 $S = S_1 - S_2 = S_1 \cap S_2^-$. 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

$$\mathbf{r}_i(u_i, v_i), \quad i = 1, 2, \dots, m, \quad (6.6.1)$$

а грани второго тела описываются поверхностями

$$\mathbf{s}_j(a_j, b_j), \quad j = 1, 2, \dots, n. \quad (6.6.2)$$

На базе линий пересечения граней первого и второго тел

$$\begin{aligned} \mathbf{l}_{uv}(t) &= [u_i(t) \ v_i(t)]^\top, & \mathbf{l}_{uv}(t) &\in \mathbf{r}_i(u_i, v_i), \\ \mathbf{l}_{ab}(t) &= [a_j(t) \ b_j(t)]^\top, & \mathbf{l}_{ab}(t) &\in \mathbf{s}_j(a_j, b_j), \\ t_{\min} &\leq t \leq t_{\max}, \end{aligned} \quad (6.6.3)$$

построим ребра пересечения. При этом ребрам пересечения дадим направление векторного произведения нормали грани первого тела с нормалью грани второго тела: $\mathbf{t}_{\text{edge}} = \mathbf{m}_r \times \mathbf{m}_s$. Направление ребра определяется знаком 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. 6.6.6 показаны направления ребер пересечения грани первого тела с двумя гранями второго тела.

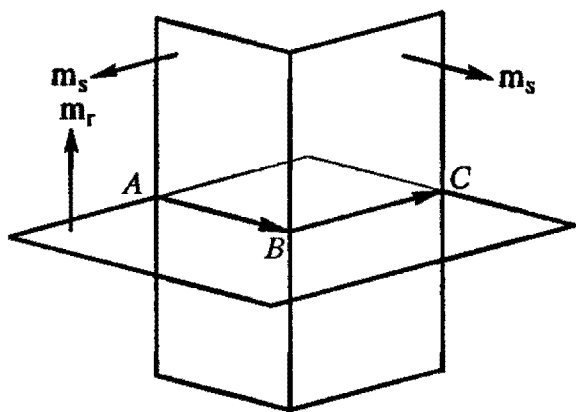


Рис. 6.6.6. Направление ребер пересечения граней тел

только своими концами. В точках A , B и C (рис. 6.6.6) 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 (6.1.1). Кривую пересечения поверхностей составляют две поверхности и две соответствующие им двумерных 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 (4.8.6) 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 (двухмерные 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 $l_{uv}(q) = [u_i(q) \ v_i(q)]^T$ и $l_{uv}(q) = [u_n(q) \ v_n(q)]^T$ нужно вставить дополнительные двухмерные 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 (4.12.5) 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. 6.6.6 граней разрезать на части, нужно перестроить ее циклы и в соответствии с циклами изменить контуры, описывающие область определения параметров поверхности грани. На рис. 6.6.7 показаны две пересекшиеся грани (тонкими линиями со стрелками показано направление циклов граней исходных тел) и ребро пересечения. На рис. 6.6.8 показаны те части граней,

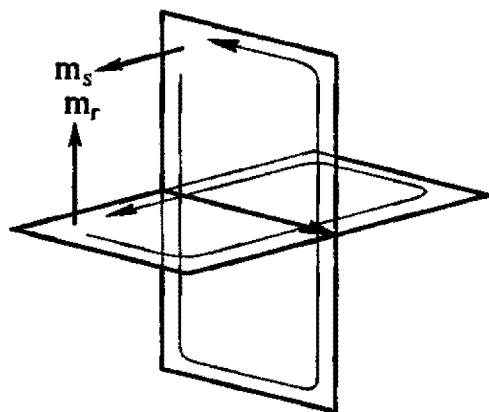


Рис. 6.6.7. Исходные грани тела

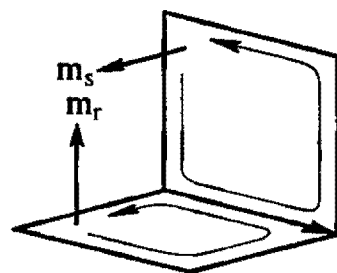


Рис. 6.6.8. Обрезанные грани

которые войдут в объединение тел. 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 ($t_{\text{edge}} = m_r \times m_s$) 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.

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. 6.6.9 показано, что с новым ребром BA в точке A стыкуются несколько ребер.

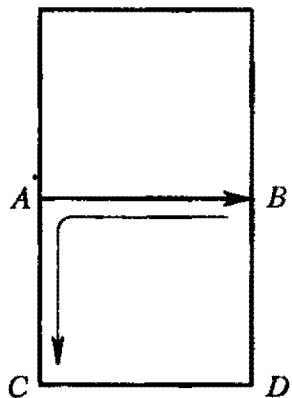


Рис. 6.6.9

В данном случае для продолжения цикла, начатого с ребра пересечения BA , следует выбрать ребро AC .

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. 6.6.10а–6.6.13а приведены варианты исходных граней

первого тела. На рис. 6.6.10б–6.6.13б приведены грани с добавлением ребер пересечения (ребра пересечения выделены). На рис. 6.6.10в–6.6.13в приведены грани, которые получились в результате операции.

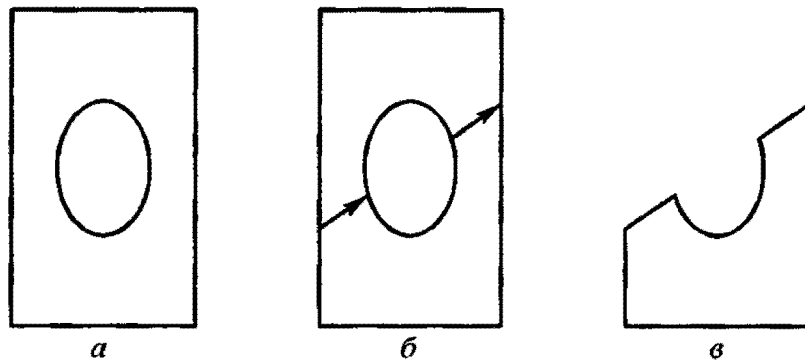


Рис. 6.6.10. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

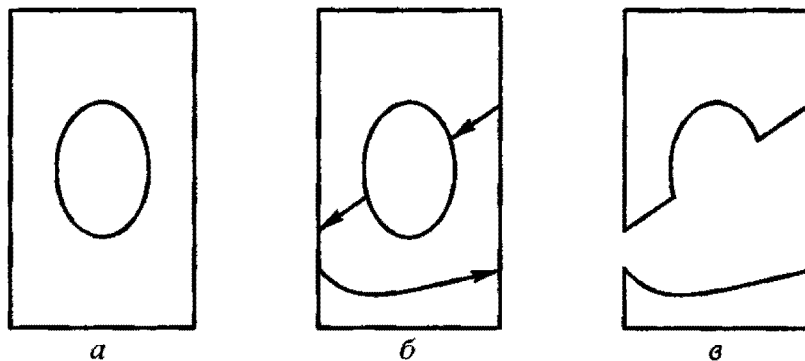


Рис. 6.6.11. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

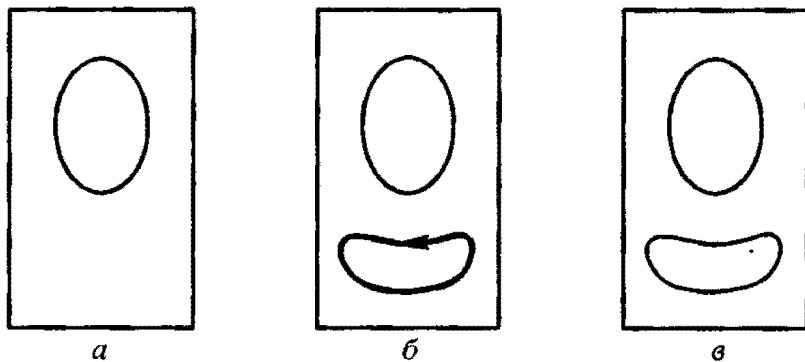


Рис. 6.6.12. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

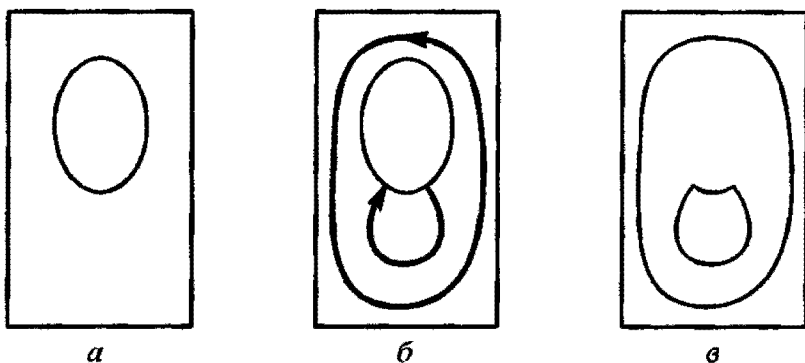


Рис. 6.6.13. Исходная грань (а), грань с добавлением ребер пересечения (б), результат операции (в)

На рис. 6.6.10 грань с двумя циклами была разрезана и получилась одна грань с одним циклом. На рис. 6.6.11 грань с двумя циклами дала две грани. В примере, приведенном на рис. 6.6.12, потребовалось использовать старый внешний и внутренний циклы. На рис. 6.6.13 из одной грани получено две грани, причем для одной из них потребовалось использовать старый внешний цикл.

Описанное перестроение циклов выполняется для каждой пересекающейся грани первого тела.

С пересекающимися 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 пересекающиеся грани объединяемых тел. Все их включим в оболочку результирующего тела. Для получения результирующего тела остается к этим граням в оболочку тела добавить не пересекающиеся в операции грани, которые топологически связаны с пересекающимися 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. 6.6.6. 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. 6.6.14)

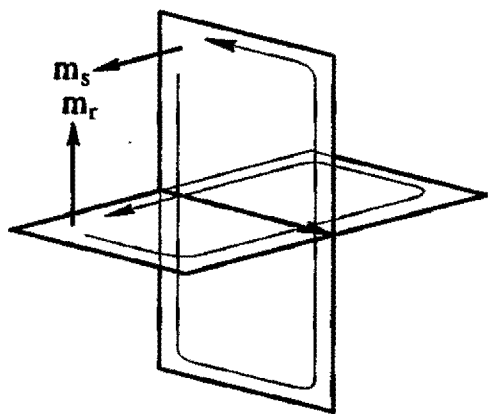


Рис. 6.6.14. Исходные грани

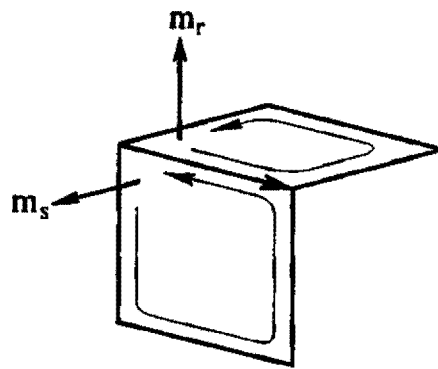


Рис. 6.6.15. Перестроенные грани пересечения тел

(в объединение тел вошла часть грани первого тела, лежащая вне второго тела, и часть грани второго тела, лежащая вне первого тела). 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. 6.6.16 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: A и B , 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 A и B 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.

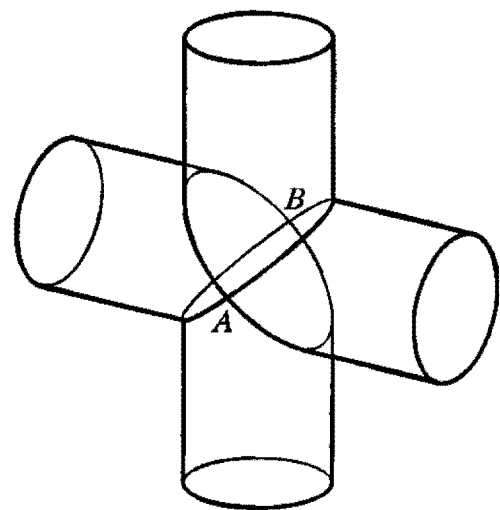


Рис. 6.6.16. В точках A и B нормали граней совпадают

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.

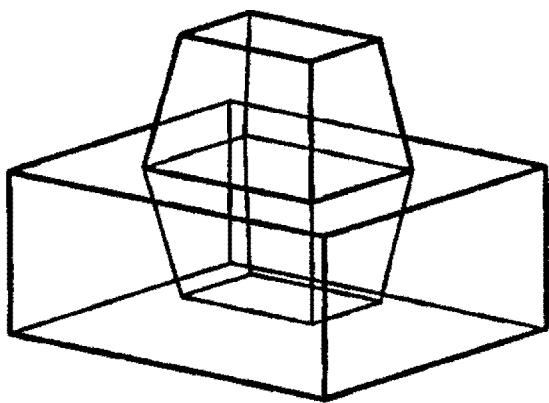


Рис. 6.6.17. Совпадение ребер пересечения с ребрами меньшего тела

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. 6.6.17 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. 6.6.18 это векторы a_1 и b_1 для грани с нормалью m_1 и векторы a_2 и b_2 для грани с нормалью m_2 . Векторы a_1 и a_2 поворачивают влево от ребра пересечения в соответствующих плоскостях, а векторы b_1 и b_2 поворачивают вправо от ребра пересечения в соответствующих плоскостях. Если грань не имеет продолже-

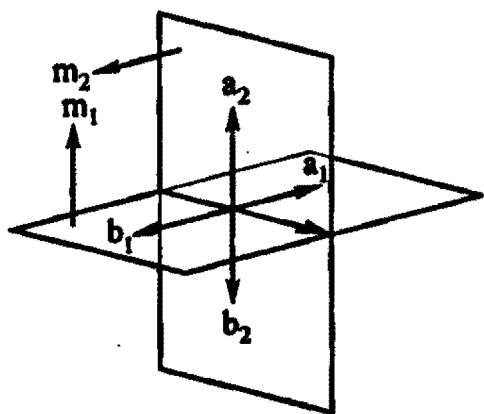


Рис. 6.6.18. Пересечение граней

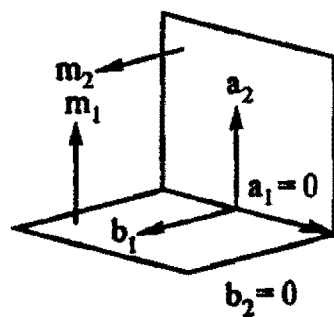


Рис. 6.6.19. Пересечение граней по ребру

ния за ребро пересечения (ребро пересечения частично или полностью совпадает с ребром тела), то соответствующий вектор положим равным нулю (рис. 6.6.19). Используя векторы a_1 , a_2 , b_1 , b_2 и нормали m_1 , 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 выпол-

нены условия

$$m_1 \cdot a_2 > 0 \quad \text{и} \quad m_2 \cdot b_1 > 0. \quad (6.6.7)$$

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 внутрь второго тела, а грань второго тела мы сможем перестроить, если она имеет продолжение справа от ребра пересечения внутрь первого тела. Таким образом, в булевой операции пересечения тел для ребра пересечения должны быть выполнены условия

$$m_1 \cdot b_2 < 0 \quad \text{и} \quad m_2 \cdot a_1 < 0. \quad (6.6.8)$$

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 (6.6.8) 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 p пространству внутри тела — классифицировать точку относительно тела. Нам известно, что нормаль каждой грани направлена вне объема тела. Для точки p найдем ближайшую точку p_0 на ближайшей грани. Построим вектор n из найденной точки p_0 в точку p . Вычислим нормаль m тела в точке p_0 . Если точка p_0 лежит на границе граней (является точкой ребра или вершины), то в качестве нормали m тела возьмем среднюю нормаль граничащих граней. Если $m \cdot n \geq 0$, то точка 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. 6.6.20 приведены два тела, некоторые грани которых частично перекрывают

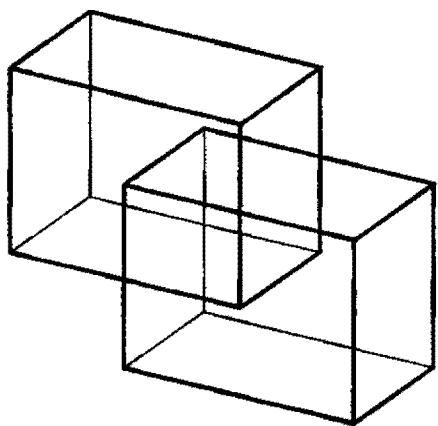


Рис. 6.6.20. Совпадающие грани тел

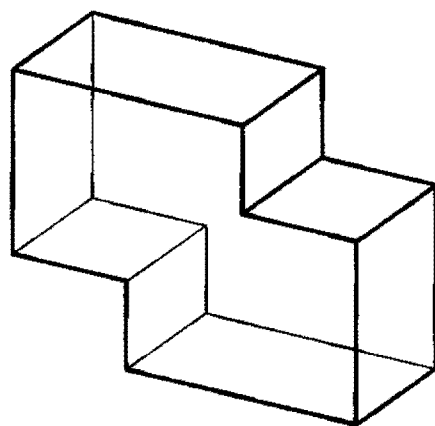


Рис. 6.6.21. Объединение тел

друг друга. 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. 6.6.21.

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 m_1 и 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. 6.6.22. В узлах дерева находятся тела. Дерево имеет несколько ярусов. Операции между телами обозначены соответствующими знаками. Операции выполняются между телами одного яруса.

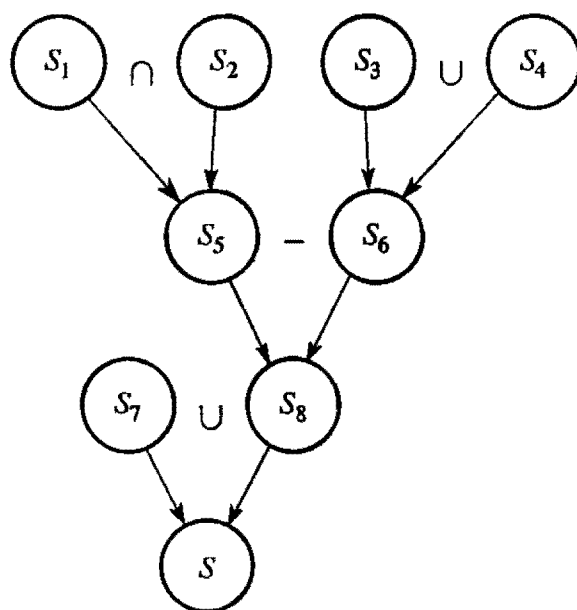


Рис. 6.6.22. Дерево построения тела

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.

6.7. 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. 6.7.1 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.

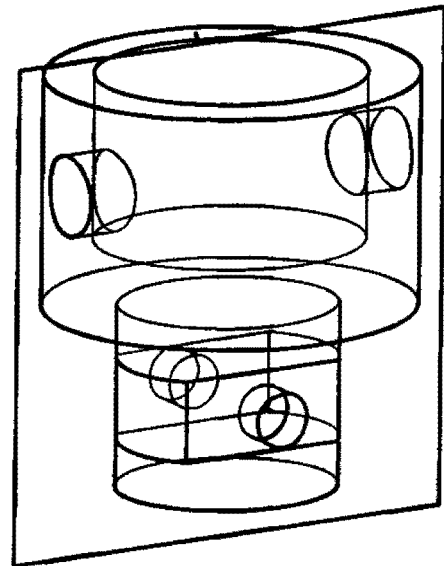
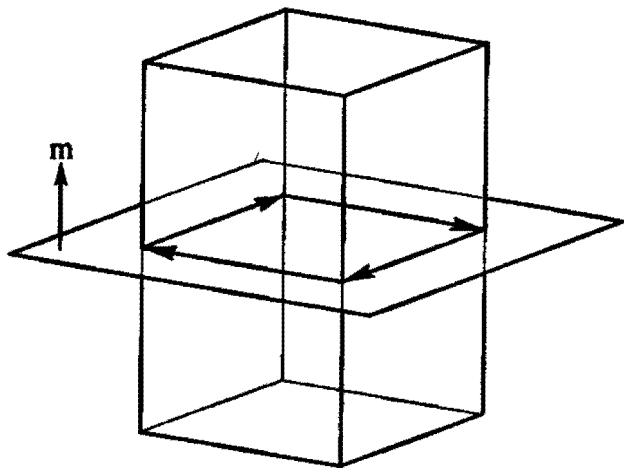


Рис. 6.7.1. Направление ребер пересечения

Рис. 6.7.2. Исходное тело и режущая поверхность

В частном случае тело можно разрезать плоскостью. На рис. 6.7.2 приведены исходное тело и режущая плоскость. На рис. 6.7.3 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. Для получения сложного разреза возьмем

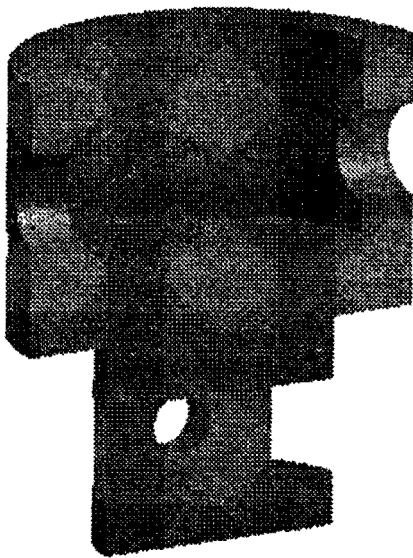


Рис. 6.7.2. Разрезанное тело

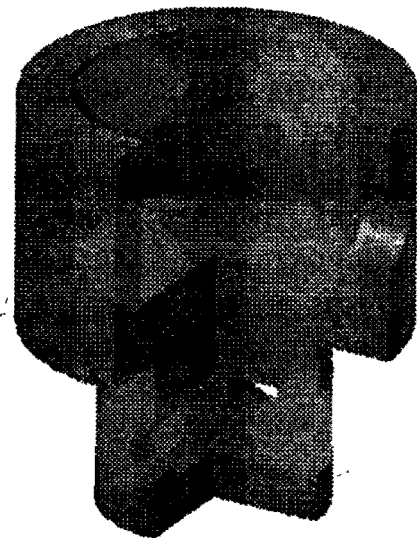


Рис. 6.7.4. Сложный разрез тела

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. 6.7.4.

6.8. 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 и \mathbf{i}_2 — ортогональные векторы единичной длины. Тогда матрица преобразования симметрии (1.3.8) тела относительно этой плоскости определяется формулой

$$\mathbf{A} = 2\mathbf{i}_1\mathbf{i}_1 + 2\mathbf{i}_2\mathbf{i}_2 - \mathbf{E}, \quad (6.8.1)$$

где $\mathbf{i}_1\mathbf{i}_1$ и $\mathbf{i}_2\mathbf{i}_2$ — диадные произведения векторов. Свободный вектор \mathbf{v}_0 после преобразования симметрии относительно этой плоскости будет описываться радиус-вектором

$$\mathbf{v} = \mathbf{A} \cdot \mathbf{v}_0. \quad (6.8.2)$$

Произвольная точка \mathbf{r}_0 после преобразования симметрии относительно этой плоскости будет описываться радиус-вектором

$$\mathbf{r} = \mathbf{p}_0 + \mathbf{A} \cdot (\mathbf{r}_0 - \mathbf{p}_0). \quad (6.8.3)$$

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 (6.8.1), то получим зеркальное отражение тела, вывернутое наизнанку. Таким образом, построение симметричного тела сводится к преобразованию его копии по матрице (6.8.1) 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

$$\begin{aligned} l_{uv}(t) &= [u_i(t) \ v_i(t)]^\top, & l_{uv}(t) &\in \mathbf{r}_i(u_i, v_i), \\ l_{xy}(t) &= [x(t) \ y(t)]^\top, & l_{xy}(t) &\in \mathbf{p}(x, y), \end{aligned} \quad (6.8.4)$$

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: $\mathbf{t}_{\text{edge}} = \mathbf{m}_r \times \mathbf{m}_p$.

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 (6.8.1) 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 (6.8.4) в ребра положим кривые

$$\begin{aligned} l_{uv}(t) &= [u_i(t) \ v_i(t)]^\top, & l_{uv}(t) &\in \mathbf{r}_i(u_i, v_i), \\ l_{uv}'(t) &= [u_i(t) \ v_i(t)]^\top, & l_{uv}'(t) &\in \mathbf{r}_i'(u_i, v_i). \end{aligned} \quad (6.8.5)$$

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.

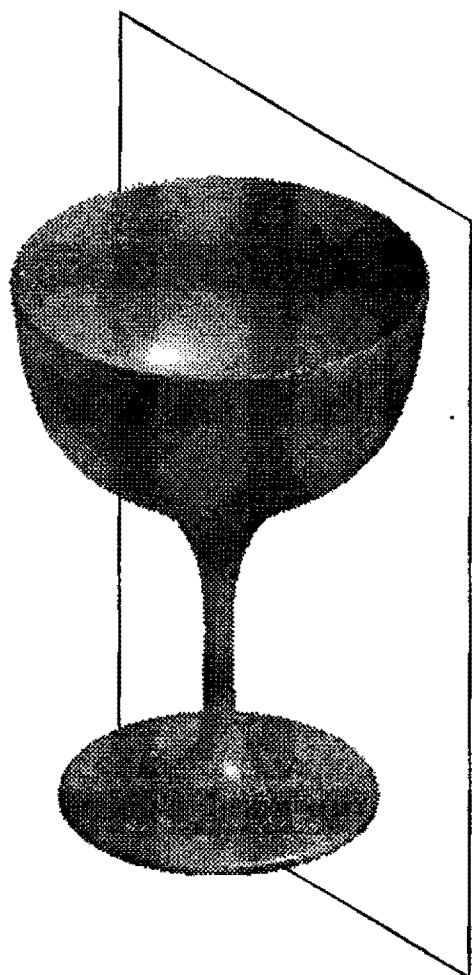


Рис. 6.8.1. Тело и плоскость симметрии

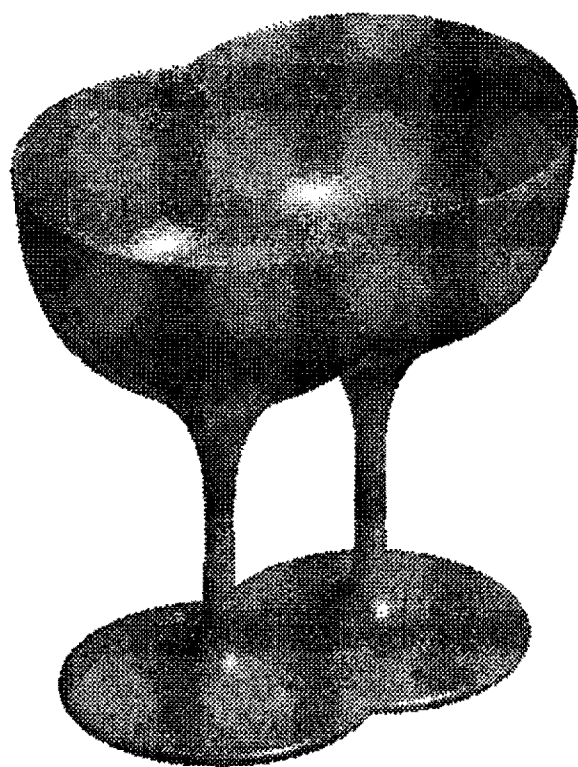


Рис. 6.8.2. Симметричное тело

На рис. 6.8.1 приведено исходное тело и плоскость симметрии. На рис. 6.8.2 приведен результат построения симметричного тела.

6.9. 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. 6.9.1 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 формулой (3.14.1). На продолжении эквидистантной поверхности за пределы области определения параметров ее радиус-вектор

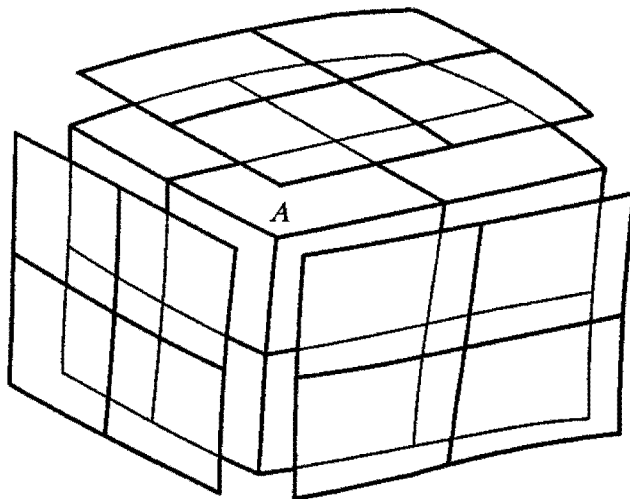


Рис. 6.9.1. Эквидистантные поверхности

будем вычислять по одной из формул (3.14.8)–(3.14.10) в зависимости от замкнутости базовой 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. Вершине базового тела будут соответствовать

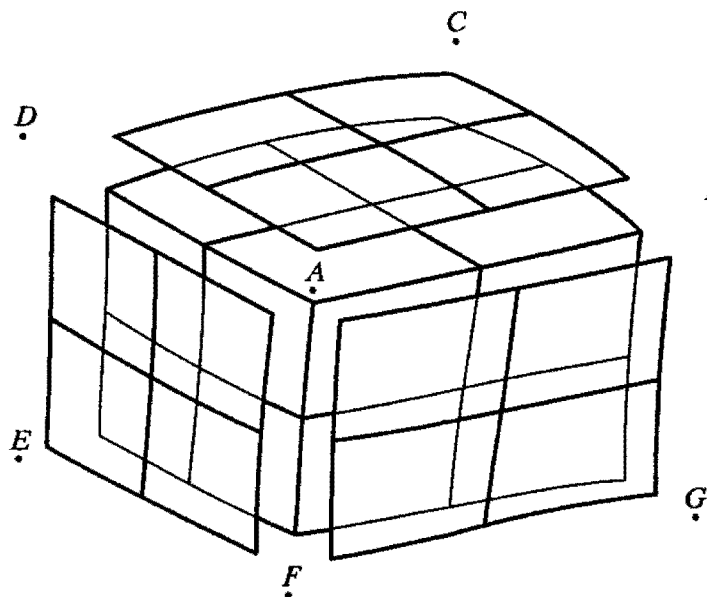


Рис. 6.9.2. Вершины эквидистантного тела

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. 6.9.2 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

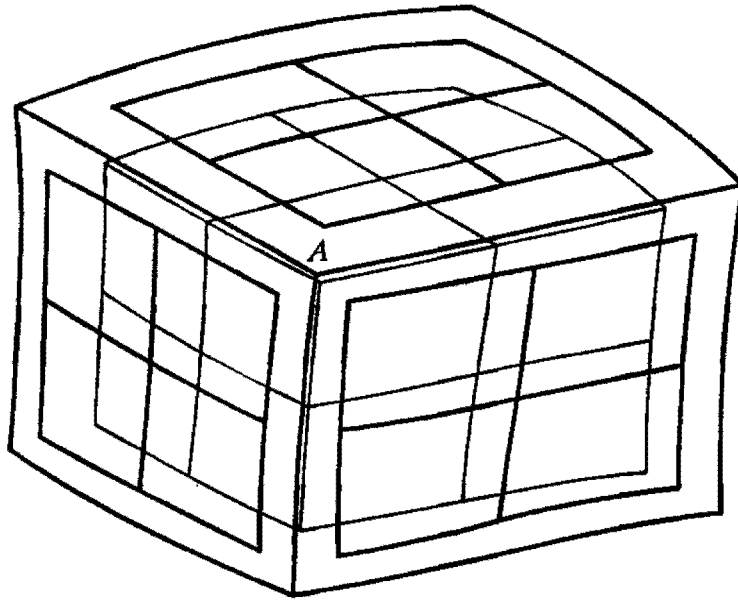


Рис. 6.9.3. Линии пересечения эквидистантных поверхностей

пересечения эквидистантных поверхностей. Начальные и конечные точки ребер нам известны из второго этапа. На рис. 6.9.3 приведены линии пересечения продолженных эквидистантных поверхностей, на которых будут базироваться ребра.

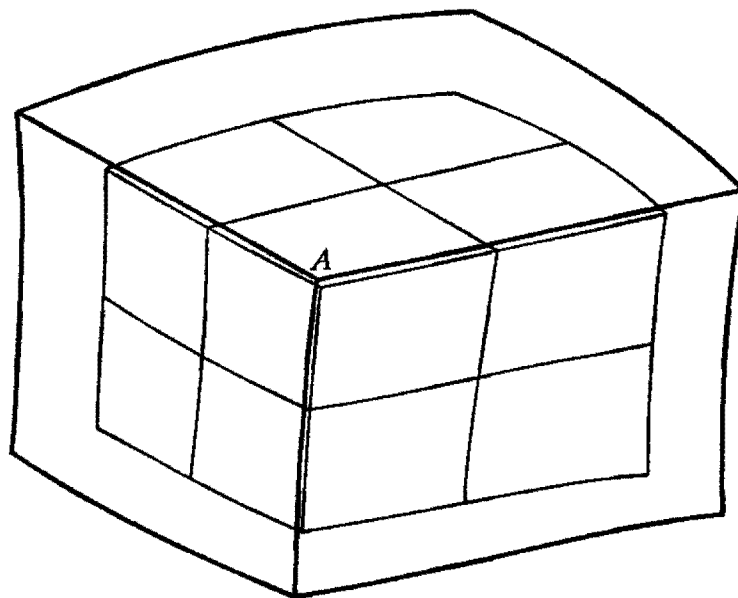


Рис. 6.9.4. Эквидистантные грани

На последнем, *четвертом этапе* построим циклы эквидистантных граней (рис. 6.9.4). Этот процесс аналогичен процессу перестроения циклов граней в булевых операциях.

Таким образом, мы получим оболочку эквидистантного тела.

Пример построения эквидистантного тела приведен на рис. 6.9.5 (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. Если

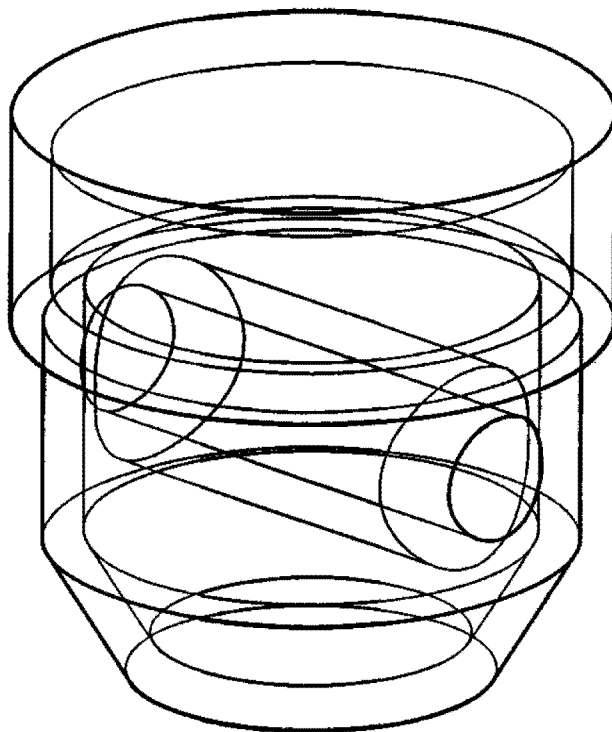


Рис. 6.9.5. Эквидистантное тело

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. 6.9.6 приведена пирамида и эквидистантное пирамиде тело с отрицательным параметром эквидистанты, а на рис. 6.9.7 приведена аналогичная пирамида и эквидистантное пирамиде тело с положительным

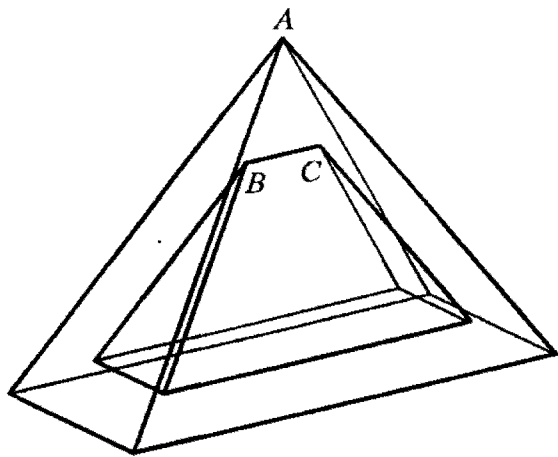


Рис. 6.9.6. Эквидистантная пирамида ($h < 0$)

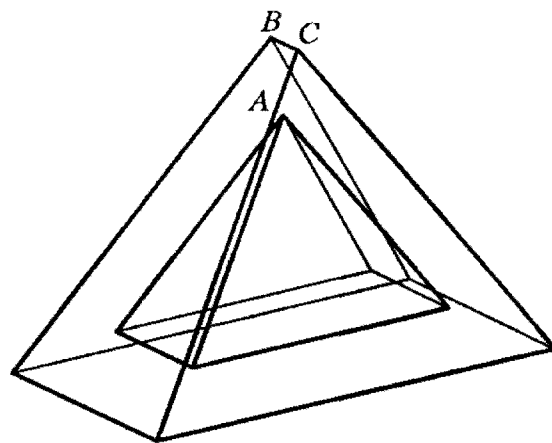


Рис. 6.9.7. Эквидистантная пирамида ($h > 0$)

параметром эквидистанты. В вершине A пирамиды стыкуется четыре ребра. В эквидистантном теле данной вершине соответствует две вершины B и C и одно дополнительное ребро BC , что мы и наблюдаем на рис. 6.9.6 и 6.9.7.

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.

6.10. 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. 6.9.5.

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

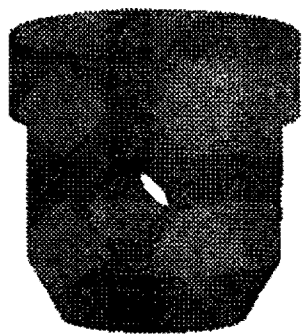


Рис. 6.10.1. Исходное тело



Рис. 6.10.2. Тонкостенное тело

оболочка. На рис. 6.10.2 приведен пример открытого тонкостенного тела ($h > 0$), построенного путем вскрытия одной грани тела, показанного на рис. 6.10.1.

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 формулой (3.14.1). На продолжении эквидистантной поверхности за пределы области определения параметров ее радиус-вектор будем вычислять по одной из формул (3.14.8)–(3.14.10) 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. Если $h > 0$, то изменим на противоположные направления циклов и нормалей сохраняемых граней базового

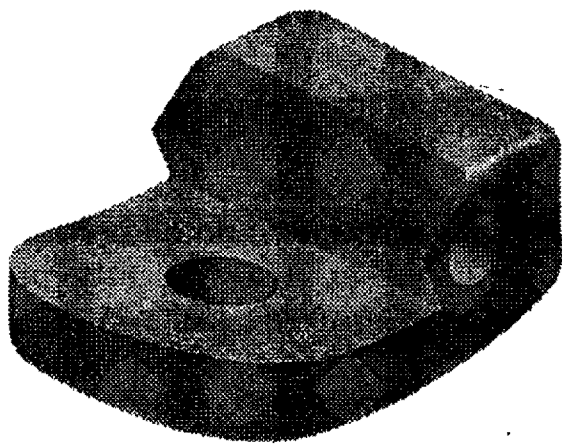


Рис. 6.10.3. Базовое тело

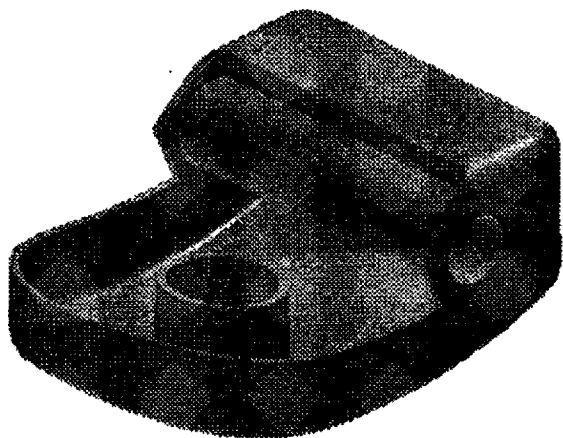


Рис. 6.10.4. Тонкостенное тело, построенное внутрь от базового тела



Рис. 6.10.5. Тонкостенное тело, построенное наружу от базового тела

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. 6.10.4 приведено открытое тонкостенное тело, построенное внутрь от базового тела путем вскрытия трех граней. Базовое тело приведено на рис. 6.10.3. На рис. 6.10.5 приведено открытое тонкостенное тело, построенное на основе этого же базового тела наружу от него 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. 6.10.6 приведено открытое тонкостенное тело, построенное на базе тела, показанного

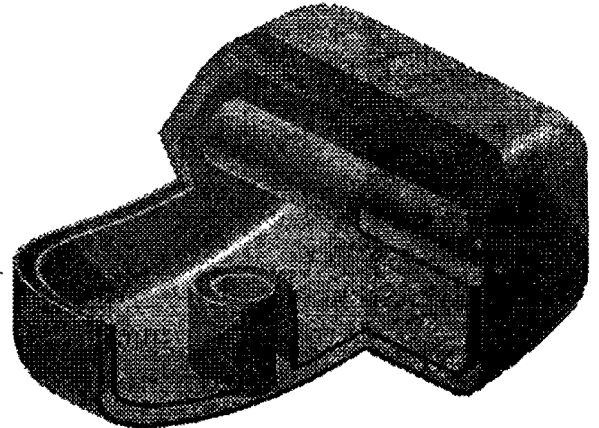
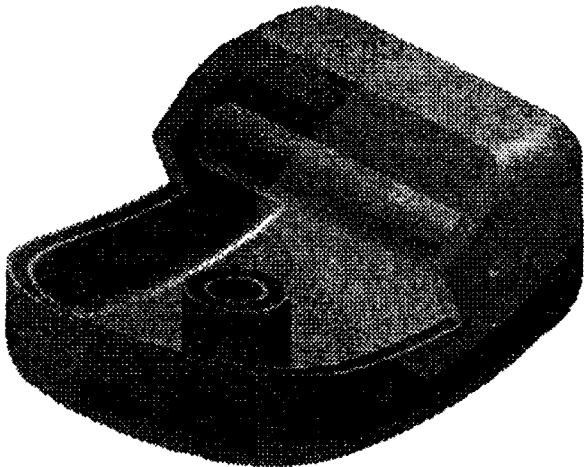


Рис. 6.10.6. Тонкостенное тело на базе тонкостенного тела

Рис. 6.10.7. Разрез тонкостенного тела

на рис. 6.10.5, путем вскрытия двух граней. На рис. 6.10.7 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.

6.11. Скругление ребер тела

Операция скругления ребер тела позволяет построить плавный переход от одной грани к другой. На рис. 6.11.1–6.11.4 приведены примеры скругления ребер тел. На рис. 6.11.1, 6.11.2 скруглены выпуклые ребра.

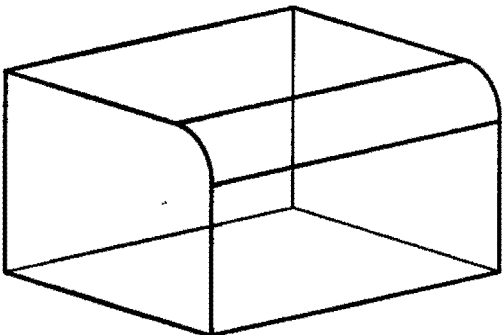


Рис. 6.11.1. Призма со скругленным ребром

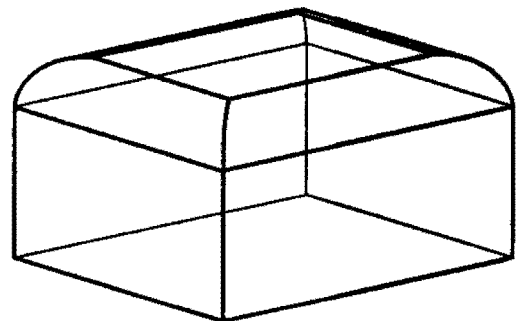


Рис. 6.11.2. Призма со скругленными ребрами

На рис. 6.11.3, 6.11.4 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 $r(u, v)$, а грань, лежащая слева от

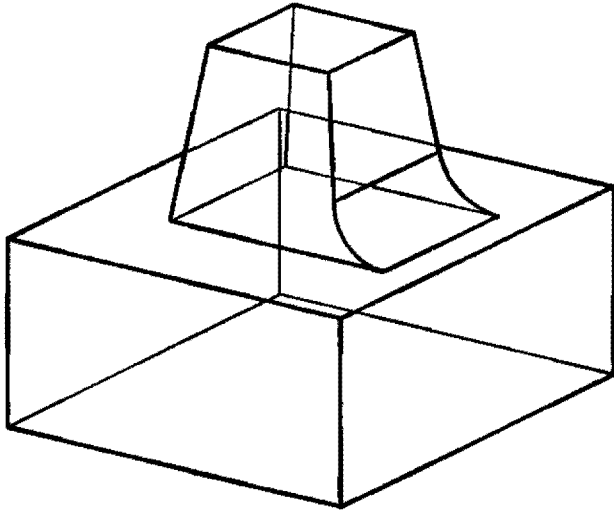


Рис. 6.11.3. Скругление ребра

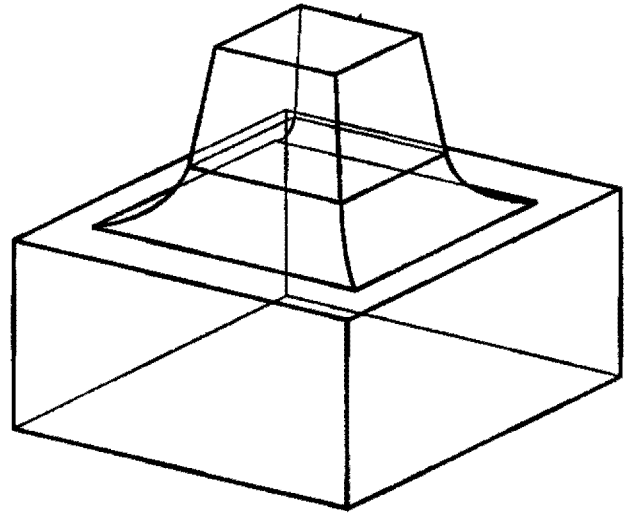


Рис. 6.11.4. Вариант скругления ребер

скругляемого ребра, базируется на поверхности $s(a, b)$. Эти грани будем называть *сопрягаемыми*. Пусть скругляемое ребро базируется на линии пересечения $c(s)$ поверхностей $r(u, v)$ и $s(a, b)$. По кривой $c(s)$ построим поверхность скругления $q(t, z)$, которая определяется формулой (4.10.11). Краями поверхности скругления являются линии (4.10.5) и (4.10.6), совпадающие с двумерными линиями (4.10.4)

$$\begin{aligned} l_{uv}(t) &= [u(t) \ v(t)]^\top, & l_{uv}(t) &\in r(u, v), \\ l_{ab}(t) &= [a(t) \ b(t)]^\top, & l_{ab}(t) &\in s(a, b) \end{aligned} \quad (6.11.1)$$

на сопрягаемых поверхностях. Построим два ребра вдоль краев поверхности скругления на базе линий пересечения

$$\begin{aligned} l_{uv}(t) &= [u(t) \ v(t)]^\top, & l_{uv}(t) &\in r(u, v), \\ l_{tz}(t) &= [t \ 0]^\top, & l_{tz}(t) &\in q(t, z), \end{aligned} \quad (6.11.2)$$

$$\begin{aligned} l_{ab}(t) &= [a(t) \ b(t)]^\top, & l_{ab}(t) &\in s(a, b), \\ l_{tz}(t) &= [t \ 1]^\top, & l_{tz}(t) &\in q(t, z). \end{aligned} \quad (6.11.3)$$

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

$$\begin{aligned} l_{tz1}(w) &= [t_{\min} \ w]^\top, & l_{tz1}(w) &\in q(t, z), \\ l_{tz2}(w) &= [t_{\max} \ w]^\top, & l_{tz2}(w) &\in q(t, z), \\ & & 0 \leq w \leq 1, \end{aligned} \quad (6.11.4)$$

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 $q(t, z)$ 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 $r(u, v)$ и $s(a, b)$ и циклы граней на них. Для этого найдем пересечение линий (6.11.2) и (6.11.3) с ребрами сопрягаемых граней или их продолжениями и изменим эти ребра. В одних случаях упомянутые ребра нужно обрезать (рис. 6.11.1), в других случаях их нужно продлить (рис. 6.11.3). Вместо скругляемого ребра в цикл грани $r(u, v)$ поставим ребро на базе линии пересечения (6.11.2), а в цикл грани $s(a, b)$ поставим ребро на базе линии пересечения (6.11.3) (рис. 6.11.5).

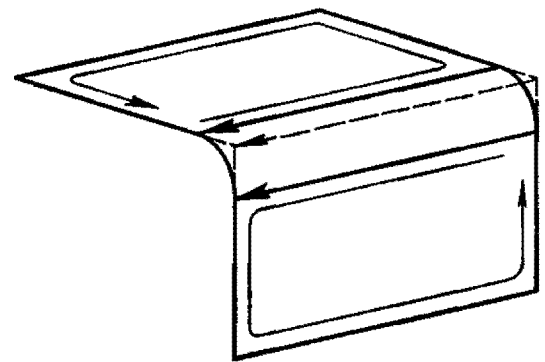


Рис. 6.11.5. Перестроение циклов граней при скруглении ребра

Далее перестроим циклы остальных граней, пересекающихся с поверхностью скругления $q(t, z)$. 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

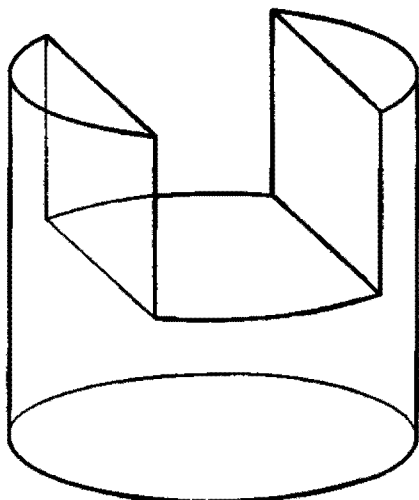


Рис. 6.11.6. Исходное тело

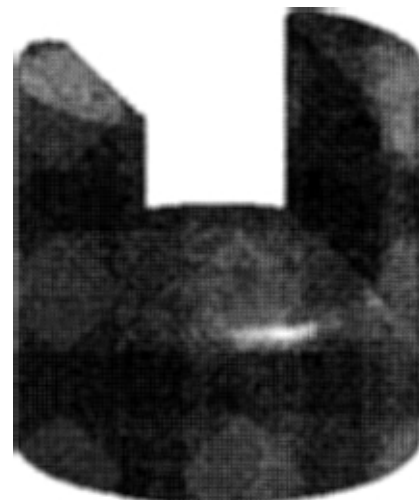


Рис. 6.11.7. Тело со скруглением нескольких ребер

несопряженными. Если скругляются сразу несколько несопряженных ребер тела, то выполним их скругление последовательно одно за другим. Скругление ребер тел, приведенных на рис. 6.11.2, 6.11.4, 6.11.7, 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. Случай скругления

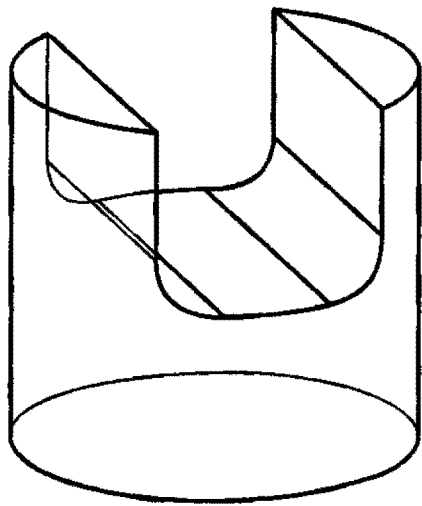


Рис. 6.11.8. Исходное тело

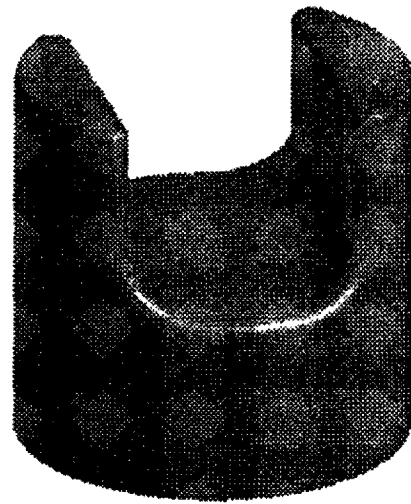


Рис. 6.11.9. Тело со скруглением цепочек ребер

нескольких сопряженных ребер приведен на рис. 6.11.9. Исходное тело, показанное на рис. 6.11.8, получено скруглением двух ребер тела, приведенного на рис. 6.11.6.

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. 6.11.11 приведено тело со скруглением сопряженных ребер, у которых на концах продольные ребра грани скругления сходятся в одну точку

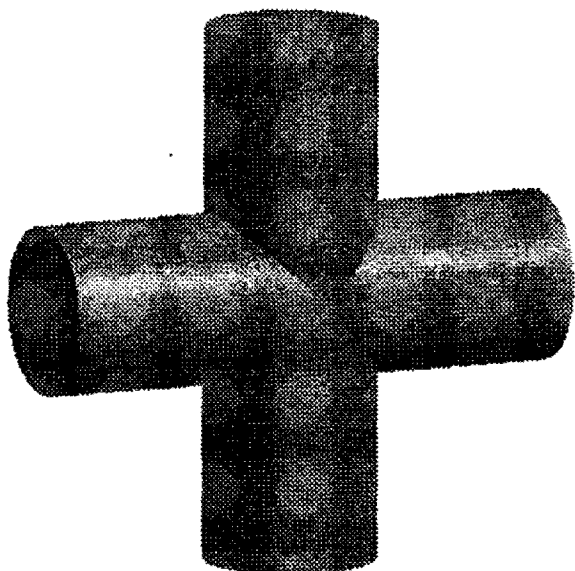


Рис. 6.11.10. Исходное тело

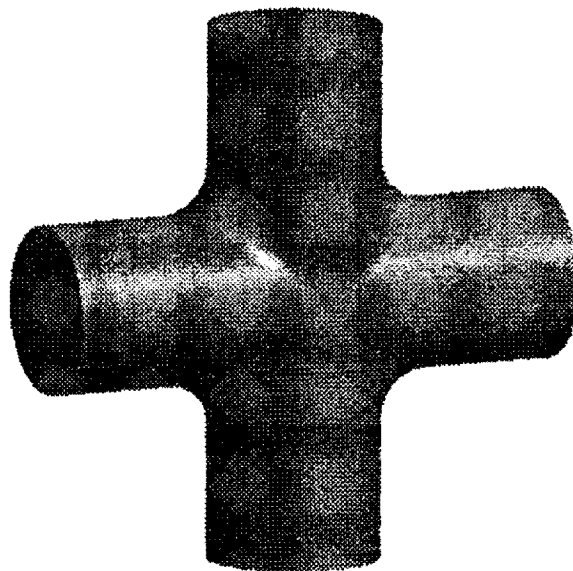


Рис. 6.11.11. Тело со скруглением цепочек ребер

(рис. 6.11.10). 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 гранью граней.

Скругление вершин. Если скруглить три ребра, стыкующиеся в одной вершине, то в вершине получим картину, приведенную на рис. 6.11.12. 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.

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 $q(t, z)$. Пусть продольная линия $l_r(t)$ этой поверхности скругления пересекается с продольной линией другой поверхности скругления в точке с параметром t_r , а продольная линия $l_s(t)$ рассматриваемой поверхности скругления пересекается с продольной линией третьей поверхности скругления в точке с параметром t_s . Построим на рассматриваемой поверхности скругления двумерный отрезок прямой из точки $[t_t \ 0]^T$ в точку $[t_s \ 1]^T$:

$$l_{tz}(z) = [(1-z)t_r \ zt_s]^T, \quad l_{tz}(z) \in q(t, z), \quad 0 \leq z \leq 1. \quad (6.11.5)$$

По отрезку и поверхности построим пространственную линию $l_q(z)$. Линии (6.11.5) построим на каждой из трех поверхностей скругления, имеющих общую точку пересечения. Отрежем и опустим ту часть каждой грани скругления, которая лежит за построенной линией (6.11.5) (со стороны общей вершины). По трем линиям (6.11.5) построим поверхность (3.11.11), а на ее базе построим грань скругления вершины.

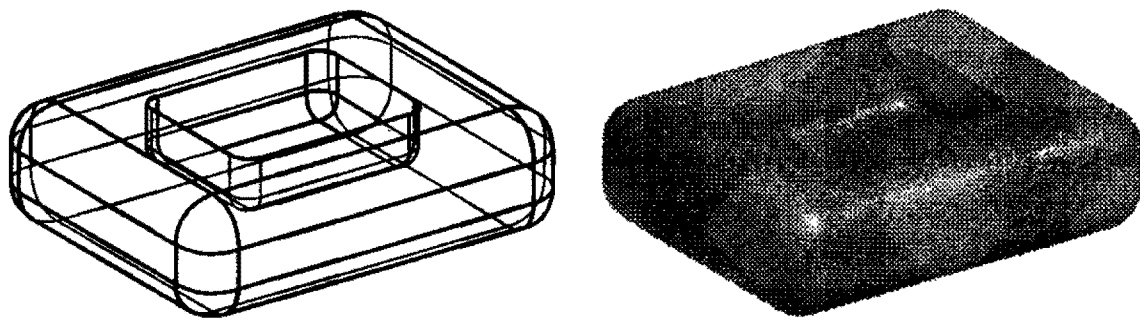


Рис. 6.11.13. Скругление ребер и вершин тела

На рис. 6.11.13 приведен пример скругления ребер и вершин призмы описанным способом.

Аналогично можно скруглить вершину, в которой стыкуются четыре скругляемых ребра. Для этого следует использовать поверхность (3.6.1), построенную по четырем кривым.

6.12. 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 (4.11.1), построенной по двум линиям (4.11.2) и (4.11.3) на пересекающихся по ребру поверхностях.

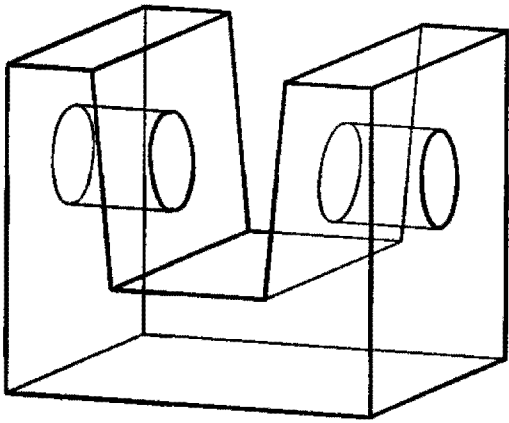


Рис. 6.12.1. Исходное тело

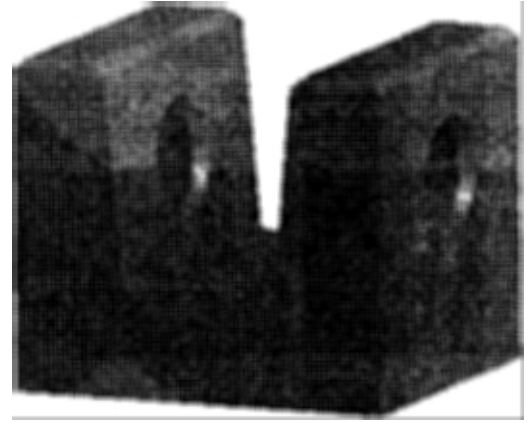


Рис. 6.12.2. Тело с фасками ребер

Результат построения фасок ребер тела, показанного на рис. 6.12.1, приведен на рис. 6.12.2.

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. Для этого найдем три

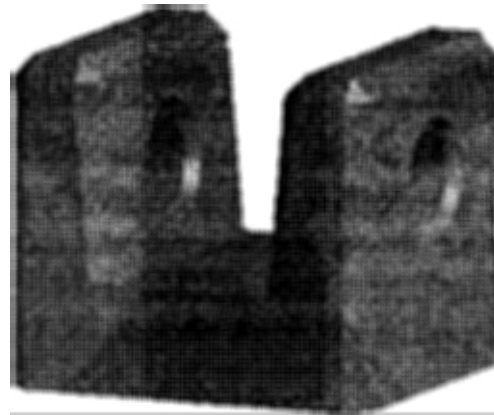


Рис. 6.12.3. Тело с фасками ребер и вершин

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. 6.12.3.

6.13. 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. 6.13.1).

Результат такой операции приведен на рис. 6.13.2.

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.

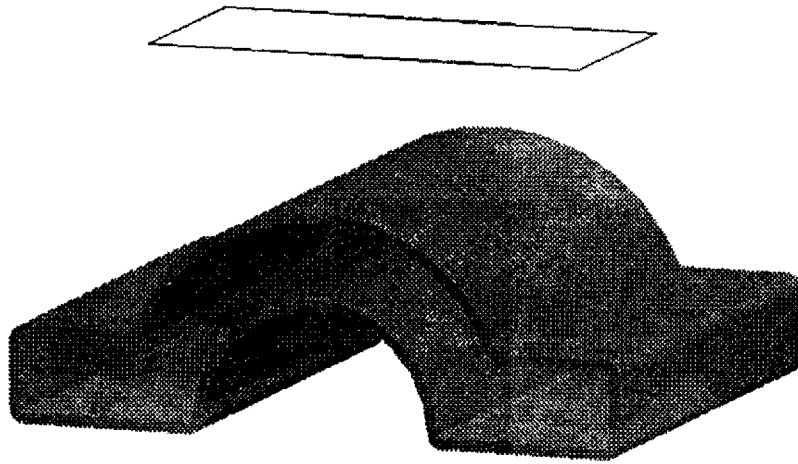


Рис. 6.13.1. Исходное тело и плоский контур

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

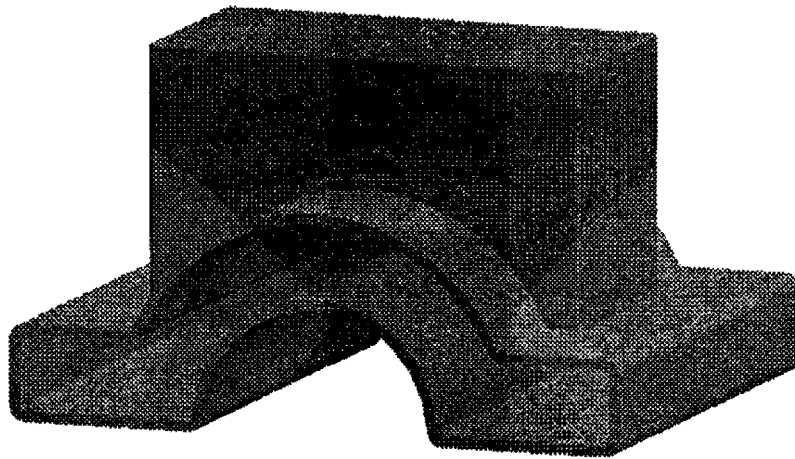


Рис. 6.13.2. Выдавливание контура до ближайшей поверхности тела

них ближайшую к заданному контуру. 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. 6.13.3).

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.

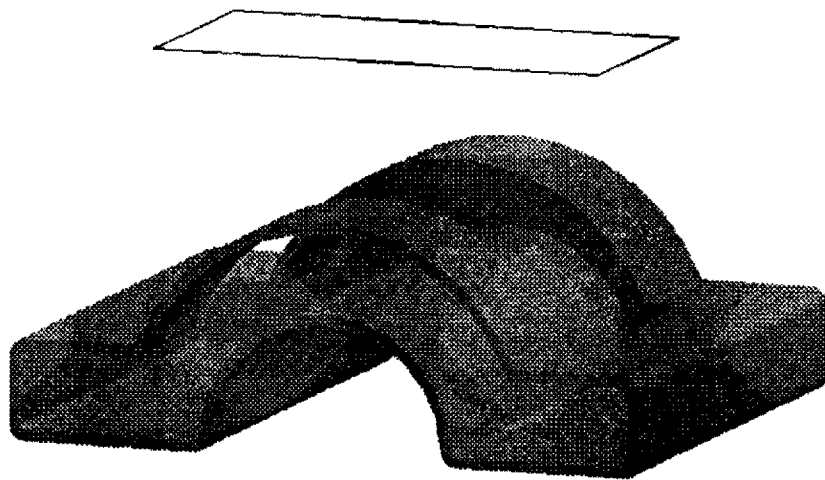


Рис. 6.13.3. Вырез контуром ближайших поверхностей тела

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. 6.13.4.

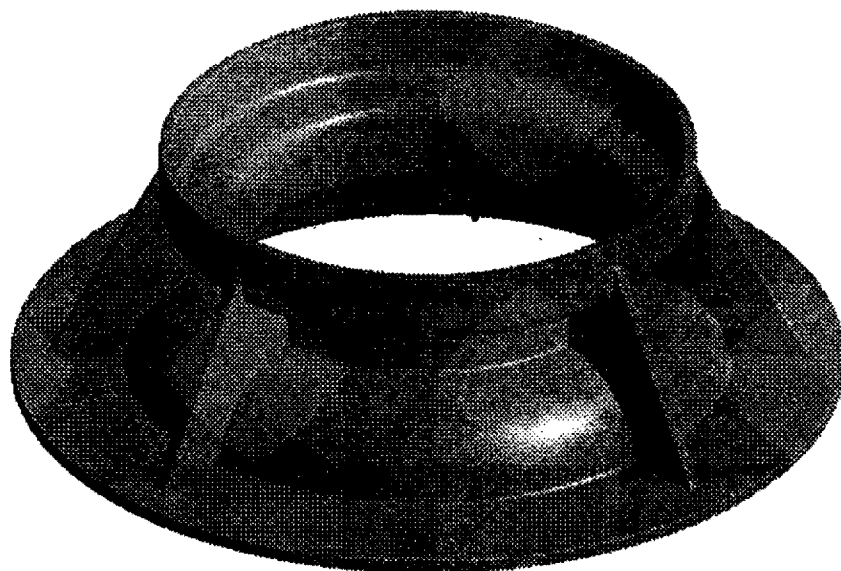


Рис. 6.13.4. Тело с ребрами жесткости

Когда оболочка одного из тел-операндов должна быть построена в процессе выполнения операции, после выполнения операций следует проверить оболочку результирующего тела на замкнутость.

6.14. 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. 6.14.1. 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. 6.14.1, из исходного тела можно получить тело с более сложной оболочкой. Операции над телом могут выполняться многократно и в произвольной последовательности, что отражено на рис. 6.14.1.

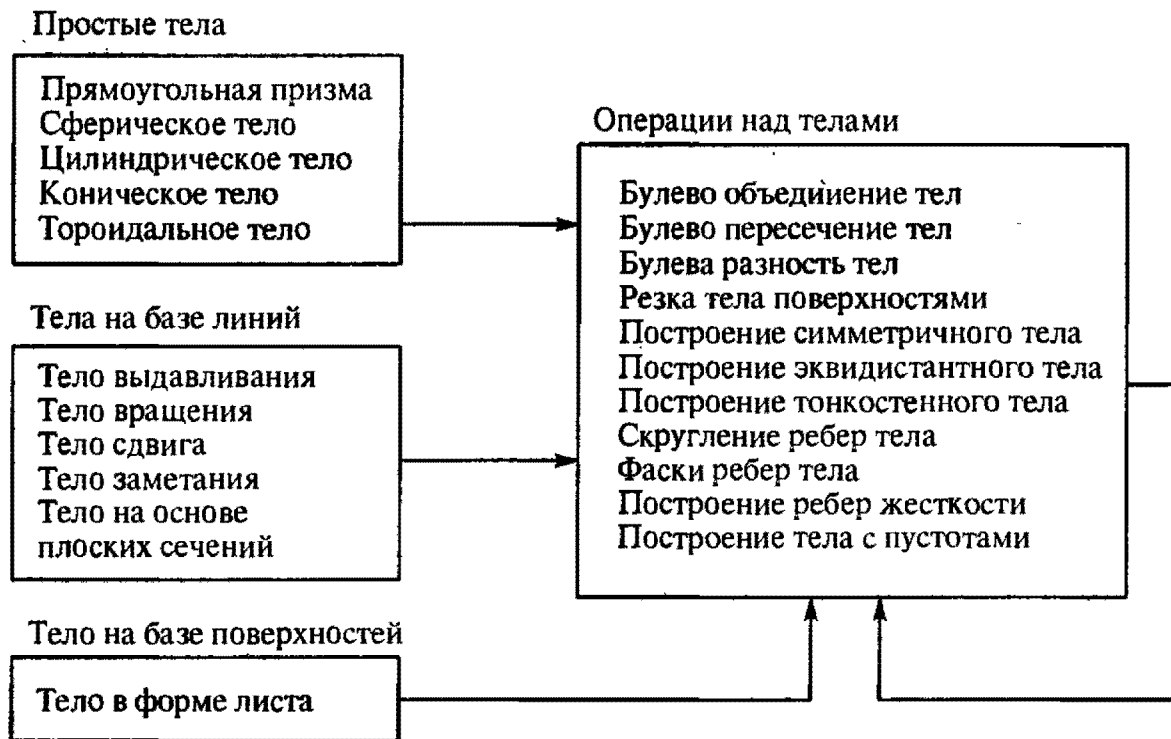


Рис. 6.14.1. Способы построения тел

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.

Глава 7

ВАРИАЦИОННЫЕ СВЯЗИ ГЕОМЕТРИЧЕСКИХ ОБЪЕКТОВ

7.1. Наложение вариационных связей

В данной главе мы рассмотрим один из эффективных способов управления геометрическими объектами. 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 геометрическими объектами.

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

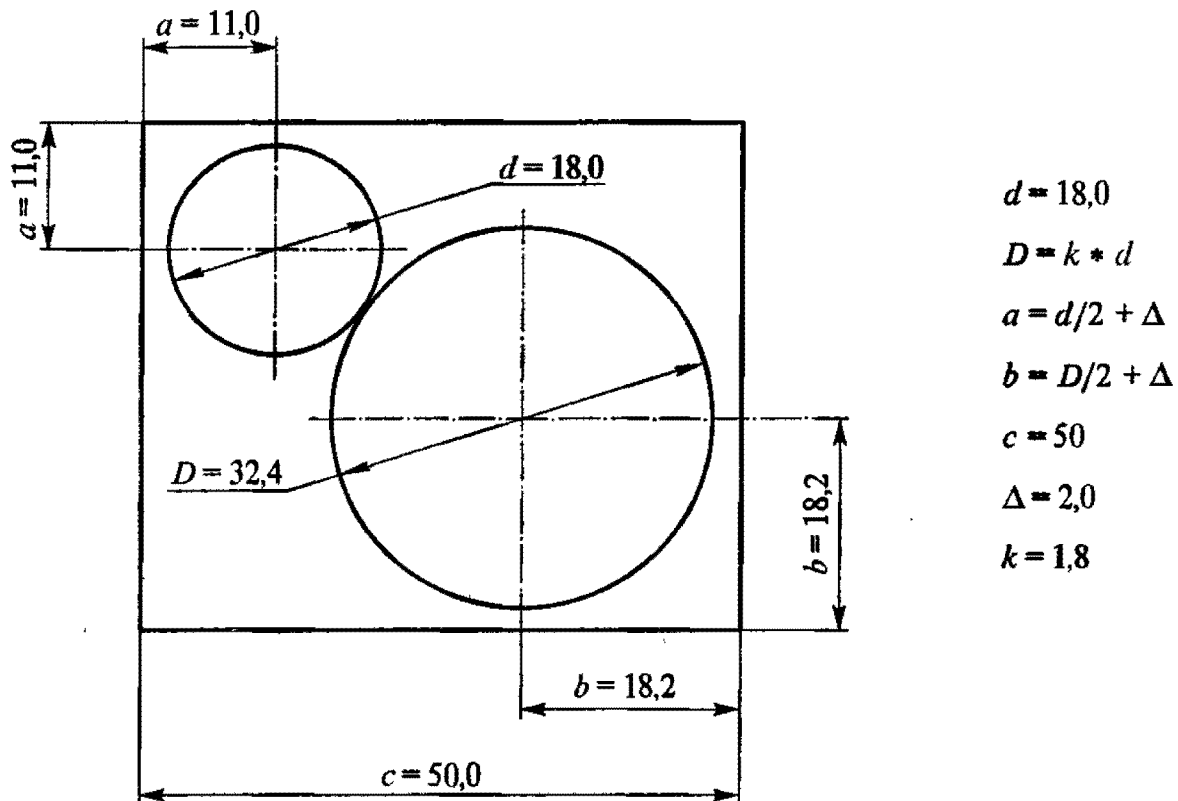


Рис. 7.1.1. Вариационные связи геометрических объектов

на параметры геометрических объектов является очень полезным. На рис. 7.1.1 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 пред-

ставляя 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. 7.1.1. Эскиз автоматически перестраивается при изменении одной или нескольких зависимостей. На рис. 7.1.2 приведены те же геометрические объекты при другом отношении диаметров окружностей. Изменение коэффициента отношения диаметров k привело

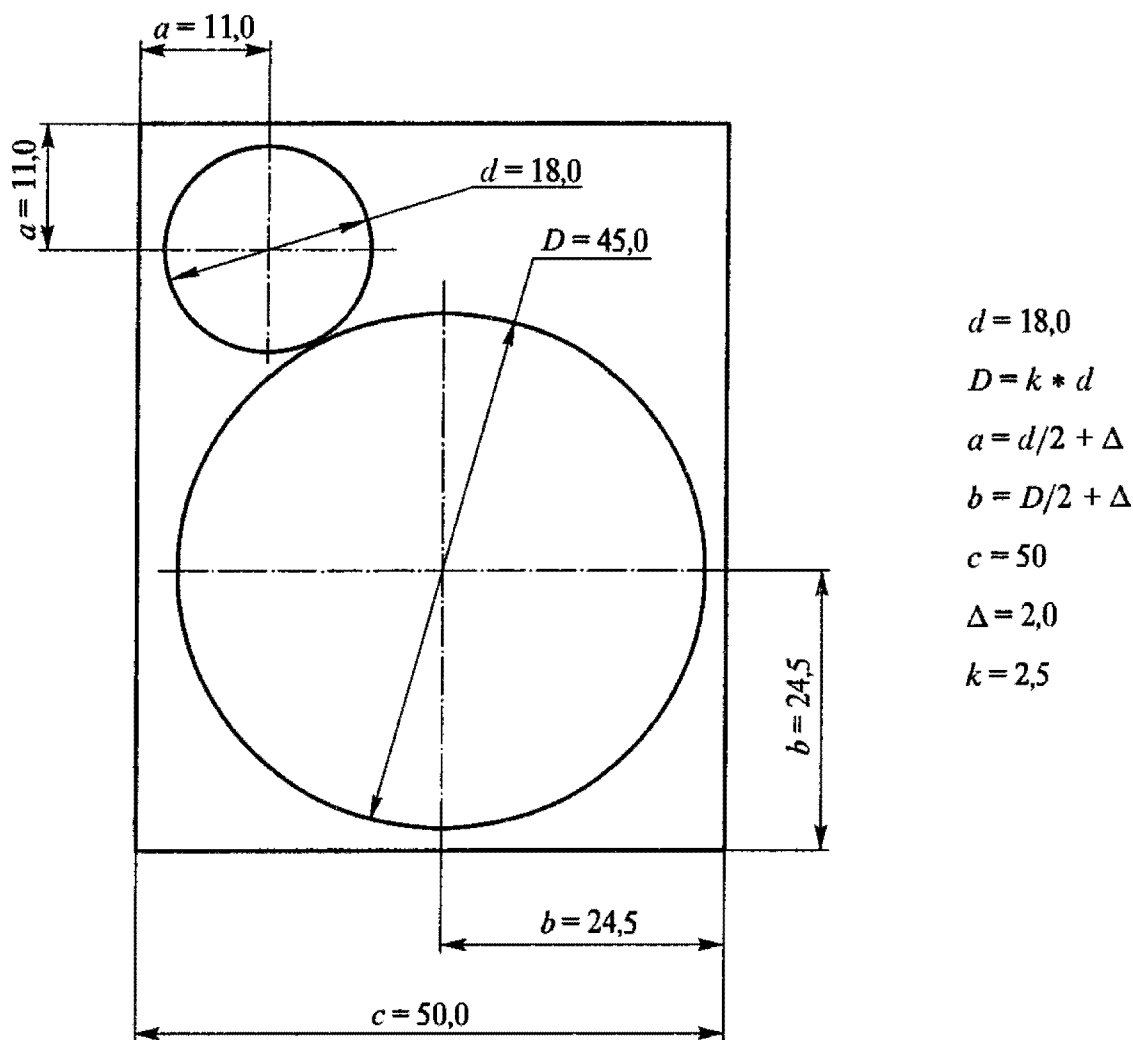


Рис. 7.1.2. Результат изменения одной из связей геометрических объектов

к изменению размера вертикальной стороны прямоугольника, изменению диаметра и положения центра большей окружности. Все заданные уравнения при новом коэффициенте k 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

$$\begin{aligned}
 f_1(x_1, x_2, \dots, x_n) &= 0, \\
 f_2(x_1, x_2, \dots, x_n) &= 0, \\
 &\dots\dots\dots \\
 f_k(x_1, x_2, \dots, x_n) &= 0,
 \end{aligned}
 \tag{7.1.1}$$

относительно параметров x_1, x_2, \dots, x_n объектов. Будем называть связывающие параметры уравнения *вариационными связями*. Кроме координат точек, компонент векторов, скалярных величин из структуры данных геометрических объектов параметрами могут служить дополнительные скалярные величины и координаты точек. Они могут существовать отдельно или могут быть введены в структуру данных геометрических объектов для удобства управления объектами посредством вариационных связей. Для наложения вариационных связей нужно, чтобы геометрические объекты предоставляли свои параметры «во внешнее пользование» и перестраивались по измененным параметрам. Вариационными связями в общем случае могут служить любые алгебраические уравнения. Естественно, что каждой геометрической задаче соответствует определенное уравнение (или несколько уравнений) для определенных параметров.

Управление геометрическими объектами с помощью вариационных связей осуществляется следующим образом. Пусть задан набор вариационных связей. Каждая связь накладывает одно или несколько уравнений на определенные параметры геометрических объектов. Связями формируется система уравнений относительно участвующих в связях параметров. Пусть в исходном состоянии параметры удовлетворяют этой системе уравнений. Предположим, что мы изменили некоторую константу в уравнении связи (например, изменили значение размера) или изменили некоторые параметры геометрического объекта (например, переместили объект) и зафиксировали их. Чтобы после этого система уравнений связей удовлетворялась, необходимо изменить остальные параметры. Для этого мы должны решить систему уравнений связей и определить новые значения параметров. О результатах решения следует сообщить геометрическим объектам. Они должны перестроиться в соответствии с этим решением (в соответствии с новыми значениями своих данных). Так как уравнения в общем случае являются нелинейными, то их 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 двумерных объектов.

7.2. Фиксирующие связи

Простейшей связью является фиксирующая связь. Она описывается одним уравнением, содержащим один параметр. Если параметр обозначить через q , то уравнение имеет вид

$$q = q^{(0)} = \text{const}, \quad (7.2.1)$$

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 (7.2.1).

Примером фиксирующей связи может служить вариационная зависимость, называемая *закреплением точки*. Она содержит три уравнения типа (7.2.1) и фиксирует три координаты точки $\mathbf{p} = [x \ y \ z]^T$. Закрепление точки описывается одним векторным уравнением

$$\mathbf{p} = \mathbf{p}^{(0)}, \quad (7.2.2)$$

состоящим из трех скалярных уравнений

$$x = x^{(0)}, \quad y = y^{(0)}, \quad z = z^{(0)}, \quad (7.2.3)$$

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 (7.2.1) 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.

7.3. 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 $p_1 = [x_1 \ y_1 \ z_1]^T$ и $p_2 = [x_2 \ y_2 \ z_2]^T$. Наложим на точки вариационную связь, называемую *линейным размером*. Она описывается уравнением

$$|p_2 - p_1| - d = 0, \quad (7.3.1)$$

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. 7.3.1. 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. В общем

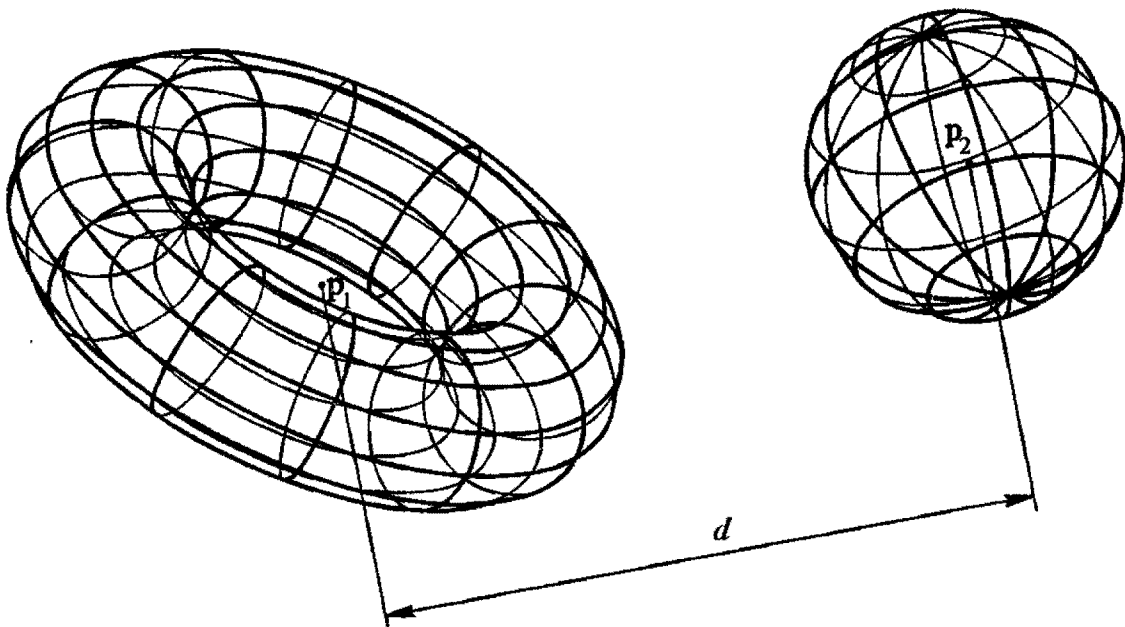


Рис. 7.3.1. Линейный размер между точками центров тора и сферы

случае это могут быть любые точки из структуры данных геометрических объектов. Мы имеем одно уравнение (7.3.1), связывающее в общем случае шесть параметров: $x_1, y_1, z_1, x_2, y_2, z_2$. В уравнении (7.3.1) 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

$$(x_2 - x_1)(y_2^{(0)} - y_1^{(0)}) = (y_2 - y_1)(x_2^{(0)} - x_1^{(0)}), \quad (7.3.2)$$

$$(y_2 - y_1)(z_2^{(0)} - z_1^{(0)}) = (z_2 - z_1)(y_2^{(0)} - y_1^{(0)}), \quad (7.3.3)$$

$$x_2 + x_1 = x_2^{(0)} + x_1^{(0)}, \quad (7.3.4)$$

$$y_2 + y_1 = y_2^{(0)} + y_1^{(0)}, \quad (7.3.5)$$

$$z_2 + z_1 = z_2^{(0)} + z_1^{(0)}, \quad (7.3.6)$$

где $\mathbf{p}_1^{(0)} = [x_1^{(0)} \ y_1^{(0)} \ z_1^{(0)}]^T$ и $\mathbf{p}_2^{(0)} = [x_2^{(0)} \ y_2^{(0)} \ z_2^{(0)}]^T$ — исходное положение точек, которое в общем случае не удовлетворяет уравнению (7.3.1). Добавив (7.3.2)–(7.3.6) к (7.3.1), 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, описанным в гл. 4. Организуем итерационный процесс

$$\begin{aligned} x_i^{(r+1)} &= x_i^{(r)} + \Delta x_i, \\ y_i^{(r+1)} &= y_i^{(r)} + \Delta y_i, \\ z_i^{(r+1)} &= z_i^{(r)} + \Delta z_i, \\ i &= 1, 2, \end{aligned} \quad (7.3.7)$$

где r — номер итерации. Приращения координат Δx_i , Δy_i , Δz_i , $i = 1, 2$, на каждой итерации определим из системы линейных алгебраических уравнений

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & 0 & a_{24} & a_{25} & 0 \\ 0 & a_{32} & a_{33} & 0 & a_{35} & a_{36} \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \Delta x_1 \\ \Delta y_1 \\ \Delta z_1 \\ \Delta x_2 \\ \Delta y_2 \\ \Delta z_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix}, \quad (7.3.8)$$

где

$$a_{11} = -a_{14} = \frac{x_1 - x_2}{d(r)}, \quad a_{12} = -a_{15} = \frac{y_1 - y_2}{d(r)}, \quad a_{13} = -a_{16} = \frac{z_1 - z_2}{d(r)},$$

$$a_{21} = -a_{24} = y_1^{(0)} - y_2^{(0)}, \quad a_{22} = -a_{25} = x_2^{(0)} - x_1^{(0)},$$

$$a_{32} = -a_{35} = z_1^{(0)} - z_2^{(0)}, \quad a_{33} = -a_{36} = y_2^{(0)} - y_1^{(0)},$$

$$b_1 = d - d^{(r)},$$

$$b_2 = (y_2^{(r)} - y_1^{(r)})(x_2^{(0)} - x_1^{(0)}) - (x_2^{(r)} - x_1^{(r)})(y_2^{(0)} - y_1^{(0)}),$$

$$b_3 = (z_2^{(r)} - z_1^{(r)})(y_2^{(0)} - y_1^{(0)}) - (y_2^{(r)} - y_1^{(r)})(z_2^{(0)} - z_1^{(0)}),$$

$$b_4 = x_2^{(0)} + x_1^{(0)} - x_2^{(r)} - x_1^{(r)},$$

$$b_5 = y_2^{(0)} + y_1^{(0)} - y_2^{(r)} - y_1^{(r)},$$

$$b_6 = z_2^{(0)} + z_1^{(0)} - z_2^{(r)} - z_1^{(r)},$$

$$d^{(r)} = \sqrt{(x_2^{(r)} - x_1^{(r)})^2 + (y_2^{(r)} - y_1^{(r)})^2 + (z_2^{(r)} - z_1^{(r)})^2}.$$

Система уравнений (7.3.8) в матричной записи имеет вид

$$\mathbf{A} \cdot \Delta \mathbf{x} = \mathbf{b}. \quad (7.3.9)$$

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 (7.3.8) 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 (7.3.7) 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 (7.3.2), (7.3.3) 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 равноправие 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

$$x_2 - x_1 = d_x, \quad (7.3.10.1)$$

$$y_2 - y_1 = d_y, \quad (7.3.10.2)$$

$$z_2 - z_1 = d_z, \quad (7.3.10.3)$$

где d_x, d_y, d_z — требуемые размеры. Уравнения (7.3.10), 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 (7.3.10) 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

$$|(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{m}| = d, \quad (7.3.11)$$

где $\mathbf{m} = [m_x \ m_y \ m_z]^\top$ — задающий направление вектор единичной длины, d — требуемый размер. Уравнение (7.3.11) 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 (7.3.11) имеет вид

$$|(x_2 - x_1)m_x + (y_2 - y_1)m_y + (z_2 - z_1)m_z| = d. \quad (7.3.12)$$

Уравнение (7.3.12) связывает шесть параметров. Если $\mathbf{m} = [m_x \ m_y \ m_z]^\top$, $\mathbf{l} = [l_x \ l_y \ l_z]^\top$, $\mathbf{n} = [n_x \ n_y \ n_z]^\top$ — три линейно независимых вектора единичной длины, то можно задать относительное положение двух точек \mathbf{p}_2 и \mathbf{p}_1 сразу тремя размерами вдоль трех некопланарных направлений \mathbf{m} , \mathbf{l} , \mathbf{n} :

$$\begin{aligned} |(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{m}| &= d_m, \\ |(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{l}| &= d_l, \\ |(\mathbf{p}_2 - \mathbf{p}_1) \cdot \mathbf{n}| &= d_n. \end{aligned} \quad (7.3.13)$$

В координатной записи уравнения (7.3.13) 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

$$\mathbf{p}_2 = \mathbf{p}_1. \quad (7.3.14)$$

Данная вариационная связь содержит три скалярных уравнения

$$x_2 - x_1 = 0, \quad (7.3.15)$$

$$y_2 - y_1 = 0, \quad (7.3.16)$$

$$z_2 - z_1 = 0. \quad (7.3.17)$$

Симметрия точек относительно плоскости. Рассмотрим вариационную связь симметрии точек \mathbf{p}_2 и \mathbf{p}_1 относительно плоскости

$$\mathbf{r}(x, y) = \mathbf{p} + xi_1 + yi_2. \quad (7.3.18)$$

Плоскость будем считать неподвижной, поэтому ее параметры варьироваться не будут.

Симметрия точек относительно плоскости описывается тремя уравнениями

$$(\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{i}_1 = (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{i}_1, \quad (7.3.19)$$

$$(\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{i}_2 = (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{i}_2, \quad (7.3.20)$$

$$(\mathbf{p}_1 - \mathbf{p}) \cdot \mathbf{i}_3 + (\mathbf{p}_2 - \mathbf{p}) \cdot \mathbf{i}_3 = 0, \quad (7.3.21)$$

где $\mathbf{i}_3 = \mathbf{i}_1 \times \mathbf{i}_2$. Три уравнения (7.3.19)–(7.3.21) связывают шесть параметров. Симметрия точек показана на рис. 7.3.2.

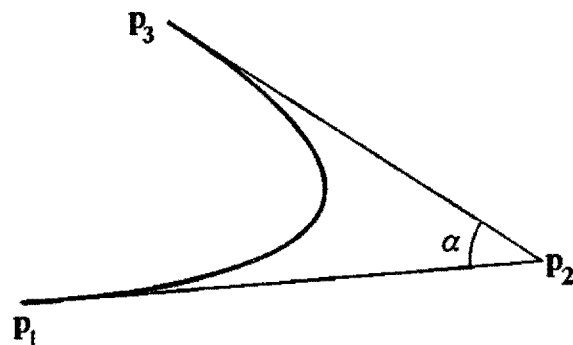
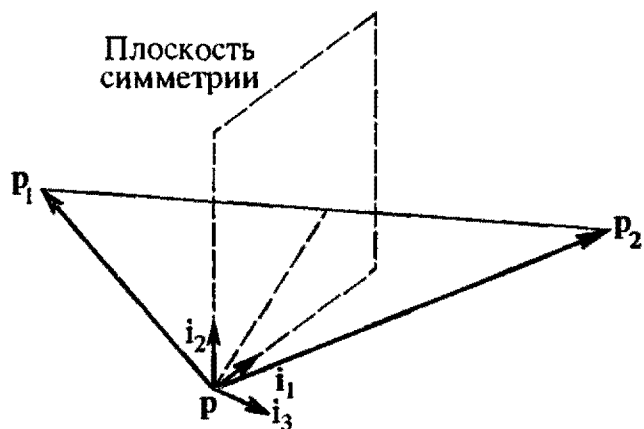


Рис. 7.3.2. Симметрия точек относительно плоскости

Рис. 7.3.3. Угловой размер между тремя точками NURBS кривой

Угловой размер. Для трех точек можно установить угловую зависимость. Пусть имеется три точки $\mathbf{p}_1 = [x_1 \ y_1 \ z_1]^T$, $\mathbf{p}_2 = [x_2 \ y_2 \ z_2]^T$ и $\mathbf{p}_3 = [x_3 \ y_3 \ z_3]^T$. 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. 7.3.3 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

$$\operatorname{arctg} \frac{s}{c} = \alpha, \quad (7.3.22)$$

где

$$s = |(\mathbf{p}_1 - \mathbf{p}_2) \times (\mathbf{p}_3 - \mathbf{p}_2)|, \quad (7.3.23)$$

$$\begin{aligned} c &= (\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_2) = \\ &= (x_1 - x_2)(x_3 - x_2) + (y_1 - y_2)(y_3 - y_2) + (z_1 - z_2)(z_3 - z_2), \end{aligned} \quad (7.3.24)$$

α — заданный угол. Величина s равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_2$ на $\sin \alpha$. Величина c равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_2$ на $\cos \alpha$. Квадрат s определяется формулой

$$s^2 = ((y_1 - y_2)(z_3 - z_2) - (z_1 - z_2)(y_3 - y_2))^2 + \\ + ((z_1 - z_2)(x_3 - x_2) - (x_1 - x_2)(z_3 - z_2))^2 + \\ + ((x_1 - x_2)(y_3 - y_2) - (y_1 - y_2)(x_3 - x_2))^2. \quad (7.3.25)$$

Уравнение (7.3.22) связывает девять параметров.

Другой вариационной связью трех точек \mathbf{p}_1 , \mathbf{p}_2 и \mathbf{p}_3 является *угловой размер в плоскости*. Эта связь устанавливает угловой размер между проекциями векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_2$ на плоскость, ортогональную единичному вектору \mathbf{m} , и описывается уравнением вида (7.3.22), где

$$s = \pm |\mathbf{q}_1 \times \mathbf{q}_3|, \quad (7.3.26)$$

$$c = \mathbf{q}_1 \cdot \mathbf{q}_3, \quad (7.3.27)$$

$$\mathbf{q}_1 = \mathbf{p}_1 - \mathbf{p}_2 - \mathbf{m}((\mathbf{p}_1 - \mathbf{p}_2) \cdot \mathbf{m}),$$

$$\mathbf{q}_2 = \mathbf{p}_3 - \mathbf{p}_2 - \mathbf{m}((\mathbf{p}_3 - \mathbf{p}_2) \cdot \mathbf{m}).$$

Векторы \mathbf{q}_1 и \mathbf{q}_3 являются составляющими векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_2$, ортогональными единичному вектору \mathbf{m} . Знак плюс в (7.3.26) выбирается в случае совпадения направления векторов $\mathbf{q}_1 \times \mathbf{q}_3$ и \mathbf{m} , знак минус — в противном случае.

Угол между векторами. Пусть имеется четыре точки $\mathbf{p}_1 = [x_1 \ y_1 \ z_1]^\top$, $\mathbf{p}_2 = [x_2 \ y_2 \ z_2]^\top$, $\mathbf{p}_3 = [x_3 \ y_3 \ z_3]^\top$ и $\mathbf{p}_4 = [x_4 \ y_4 \ z_4]^\top$. 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 (7.3.22). Уравнение, описывающее угол между четырьмя точками, имеет тот же вид, что и уравнение (7.3.22):

$$\operatorname{arctg} \frac{s}{c} = \alpha, \quad (7.3.28)$$

где

$$s = |(\mathbf{p}_1 - \mathbf{p}_2) \times (\mathbf{p}_3 - \mathbf{p}_4)|, \quad (7.3.29)$$

$$c = (\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_4) = \\ = (x_1 - x_2)(x_3 - x_4) + (y_1 - y_2)(y_3 - y_4) + (z_1 - z_2)(z_3 - z_4), \quad (7.3.30)$$

α — заданный угол. Величина s равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_4$ на $\sin \alpha$. Величина c равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2$ и

$\mathbf{p}_3 - \mathbf{p}_4$ на $\cos \alpha$. Квадрат s определяется формулой

$$s^2 = ((y_1 - y_2)(z_3 - z_4) - (z_1 - z_2)(y_3 - y_4))^2 + \\ + ((z_1 - z_2)(x_3 - x_4) - (x_1 - x_2)(z_3 - z_4))^2 + \\ + ((x_1 - x_2)(y_3 - y_4) - (y_1 - y_2)(x_3 - x_4))^2.$$

Уравнение (7.3.28) связывает двенадцать параметров.

Потребуем, чтобы при изменении угла между векторами точки перемещались ортогонально вектору \mathbf{n} . Для нахождения точек скрещения построим две линии

$$\mathbf{r}_1(t_1) = (1 - t_1)\mathbf{p}_1 + t_1\mathbf{p}_2, \quad (7.3.31)$$

$$\mathbf{r}_3(t_3) = (1 - t_3)\mathbf{p}_3 + t_3\mathbf{p}_4. \quad (7.3.32)$$

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

$$(\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{r}_1 - \mathbf{r}_3) = 0, \quad (7.3.33)$$

$$(\mathbf{p}_3 - \mathbf{p}_4) \cdot (\mathbf{r}_1 - \mathbf{r}_3) = 0. \quad (7.3.34)$$

Вектор, построенный по точкам скрещения, ортогонален обеим прямым. Параметры t_1 и t_3 прямых, соответствующие точкам скрещения, определим из системы уравнений

$$(\mathbf{p}_1 - \mathbf{p}_2) \cdot ((1 - t_1)\mathbf{p}_1 + t_1\mathbf{p}_2 - (1 - t_3)\mathbf{p}_3 - t_3\mathbf{p}_4) = 0, \quad (7.3.35)$$

$$(\mathbf{p}_3 - \mathbf{p}_4) \cdot ((1 - t_1)\mathbf{p}_1 + t_1\mathbf{p}_2 - (1 - t_3)\mathbf{p}_3 - t_3\mathbf{p}_4) = 0 \quad (7.3.36)$$

или

$$(\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_2 - \mathbf{p}_1)t_1 + (\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_4)t_3 = (\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_1),$$

$$(\mathbf{p}_3 - \mathbf{p}_4) \cdot (\mathbf{p}_2 - \mathbf{p}_1)t_1 + (\mathbf{p}_3 - \mathbf{p}_4) \cdot (\mathbf{p}_3 - \mathbf{p}_4)t_3 = (\mathbf{p}_3 - \mathbf{p}_4) \cdot (\mathbf{p}_3 - \mathbf{p}_1).$$

По найденным параметрам найдем точки скрещения.

Ортогональность векторов. Частным случаем вариационной связи (7.3.28) является ортогональность векторов, построенных по точкам $\mathbf{p}_1, \mathbf{p}_2$ и $\mathbf{p}_3, \mathbf{p}_4$. Данная вариационная связь определяется уравнением

$$(\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_3 - \mathbf{p}_4) = 0. \quad (7.3.37)$$

Если на точках построены отрезки прямых (7.3.31) и (7.3.32), то уравнение (7.3.37) делает эти отрезки ортогональными.

Параллельность векторов. Вариационная связь, делающая векторы $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_4$ параллельными, определяется векторным уравнением

$$(\mathbf{p}_1 - \mathbf{p}_2) \times (\mathbf{p}_3 - \mathbf{p}_4) = \mathbf{0}. \quad (7.3.38)$$

Связь (7.3.38) будем называть параллельностью векторов. Уравнение (7.3.38) содержит три скалярных уравнения

$$(y_1 - y_2)(z_3 - z_4) = 0, \quad (z_1 - z_2)(x_3 - x_4) = 0, \quad (x_1 - x_2)(y_3 - y_4) = 0.$$

7.4. 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 (2.2.11) 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 и w_0 есть постоянные величины, если не будет объявлено, что они также являются варьируемыми параметрами уравнения. Линейный размер описывается уравнением (7.3.1), в котором теперь $\mathbf{p}_1 = \mathbf{a}(t_0)$, $\mathbf{p}_2 = \mathbf{b}(w_0)$. Координаты связываемых точек $\mathbf{p}_1 = [x_1 \ y_1 \ z_1]^\top$ и $\mathbf{p}_2 = [x_2 \ y_2 \ z_2]^\top$ являются функциями координат точек кривых $\mathbf{a}_1 = [x_{a1} \ y_{a1} \ z_{a1}]^\top$, $\mathbf{a}_2 = [x_{a2} \ y_{a2} \ z_{a2}]^\top$, $\mathbf{b}_1 = [x_{b1} \ y_{b1} \ z_{b1}]^\top$, $\mathbf{b}_2 = [x_{b2} \ y_{b2} \ z_{b2}]^\top$. 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 $\mathbf{p}_1 = \mathbf{a}(t_0)$ и $\mathbf{p}_2 = \mathbf{b}(w_0)$.

Аналогично можно установить размер по одной из координат (7.3.10) или размер вдоль заданного направления (7.3.12) между определенной точкой кривой $\mathbf{a}(t)$ и определенной точкой кривой $\mathbf{b}(w)$. Можно связать определенные точки кривых вариационной зависимостью (7.3.14). Можно установить угловой размер (7.3.28) 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

$$\mathbf{p} = \mathbf{s}_p, \quad (7.4.1)$$

где \mathbf{s}_p — проекция точки на геометрический объект. Векторное уравнение (7.4.1) 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.

7.5. Алгебраические связи

В уравнения связей (7.3.1), (7.3.12), (7.3.22), (7.3.28) входят величины d и α , которые до сих пор читались константами. На рис. 7.1.1 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 $\mathbf{p}_1 = [x_1 \ y_1 \ z_1]^\top$ и $\mathbf{p}_2 = [x_2 \ y_2 \ z_2]^\top$, а другой связывает точки $\mathbf{p}_3 = [x_3 \ y_3 \ z_3]^\top$ и $\mathbf{p}_4 = [x_4 \ y_4 \ z_4]^\top$. Пусть линейные размеры описываются уравнениями

$$|\mathbf{p}_2 - \mathbf{p}_1| = a, \quad (7.5.1)$$

$$|\mathbf{p}_3 - \mathbf{p}_4| = b. \quad (7.5.2)$$

Пусть значения размеров a и b связаны некоторыми алгебраическими уравнениями

$$g(a, b) = 0, \quad (7.5.3)$$

$$h(a, b) = 0. \quad (7.5.4)$$

Уравнения (7.5.1) и (7.5.2) отличаются от уравнения (7.3.1) тем, что в последнем размер d не варьируется, а в уравнениях (7.5.1) и (7.5.2) параметры a и b являются варьируемыми. На $(r+1)$ -й итерации решения системы уравнений изменяются как координаты точек, так и значения размеров:

$$x_i^{(r+1)} = x_i^{(r)} + \Delta x_i, \quad y_i^{(r+1)} = y_i^{(r)} + \Delta y_i, \quad z_i^{(r+1)} = z_i^{(r)} + \Delta z_i,$$

$$a^{(r+1)} = a^{(r)} + \Delta a, \quad b^{(r+1)} = b^{(r)} + \Delta b,$$

$$i = 1, 2, 3, 4.$$

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 (7.5.3) алгебраическая связь может определяться неравенством

$$a > b. \quad (7.5.5)$$

На каждой итерации решения системы уравнений неравенство заменяется соответствующим уравнением. Неравенство (7.5.5) на каждой итерации должно быть заменено уравнением

$$a - b = \varepsilon, \quad (7.5.6)$$

где

$$\varepsilon = a^{(r)} - b^{(r)}, \quad \text{если } a^{(r)} > b^{(r)},$$

$$\varepsilon = 0, \quad \text{если } a^{(r)} \leq b^{(r)}.$$

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*.

7.6. 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 (7.3.1). Если ввести дополнительные уравнения (7.3.2)–(7.3.6), 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 (7.3.2)–(7.3.6) 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

$$\psi = \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (z_1 - z_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 + (z_2 - z_2^{(0)})^2 \right). \quad (7.6.1)$$

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 (7.3.1). Используем метод неопределенных множителей Лагранжа для отыскания минимума функции (7.6.1) при условии (7.3.1). Необходимым условием минимума функции (7.6.1) при условии (7.3.1) является равенство нулю частных производных по параметрам функции

$$\begin{aligned} F = \psi + \lambda f &= \frac{1}{2} \left(|\mathbf{p}_1 - \mathbf{p}_1^{(0)}|^2 + |\mathbf{p}_2 - \mathbf{p}_2^{(0)}|^2 \right) + \lambda (|\mathbf{p}_2 - \mathbf{p}_1| - d) = \\ &= \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (z_1 - z_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 + \right. \\ &\quad \left. + (z_2 - z_2^{(0)})^2 \right) + \lambda \left(\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} - d \right), \quad (7.6.2) \end{aligned}$$

где $f = |\mathbf{p}_2 - \mathbf{p}_1| - d$, λ — подлежащий определению множитель. Искомые координаты точек и множитель λ найдем из системы уравнений

$$\begin{aligned} \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} - d &= 0, \\ x_1 - x_1^{(0)} + \lambda \frac{x_1 - x_2}{|\mathbf{p}_1 - \mathbf{p}_2|} &= 0, \\ y_1 - y_1^{(0)} + \lambda \frac{y_1 - y_2}{|\mathbf{p}_1 - \mathbf{p}_2|} &= 0, \\ z_1 - z_1^{(0)} + \lambda \frac{z_1 - z_2}{|\mathbf{p}_1 - \mathbf{p}_2|} &= 0, \\ x_2 - x_2^{(0)} + \lambda \frac{x_2 - x_1}{|\mathbf{p}_1 - \mathbf{p}_2|} &= 0, \\ y_2 - y_2^{(0)} + \lambda \frac{y_2 - y_1}{|\mathbf{p}_1 - \mathbf{p}_2|} &= 0, \\ z_2 - z_2^{(0)} + \lambda \frac{z_2 - z_1}{|\mathbf{p}_1 - \mathbf{p}_2|} &= 0. \end{aligned} \quad (7.6.3)$$

Данная система уравнений имеет решение

$$\begin{aligned}
 x_1 &= x_1^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(x_1^{(0)} - x_2^{(0)}), \\
 y_1 &= y_1^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(y_1^{(0)} - y_2^{(0)}), \\
 z_1 &= z_1^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(z_1^{(0)} - z_2^{(0)}), \\
 x_2 &= x_2^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(x_2^{(0)} - x_1^{(0)}), \\
 y_2 &= y_2^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(y_2^{(0)} - y_1^{(0)}), \\
 z_2 &= z_2^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}}(z_2^{(0)} - z_1^{(0)}), \\
 \lambda &= -\frac{d - d^{(0)}}{2},
 \end{aligned} \tag{7.6.4}$$

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 (7.6.2). Точки переместятся вдоль прямой, проходящей через точки $\mathbf{p}_1^{(0)}$ и $\mathbf{p}_2^{(0)}$, 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 (7.2.2) и ее координаты не варьируются. Пусть точка \mathbf{p}_1 закреплена. Тогда функция (7.6.2) будет иметь вид

$$\begin{aligned}
 F &= \frac{1}{2}|\mathbf{p}_2 - \mathbf{p}_2^{(0)}|^2 + \lambda(|\mathbf{p}_2 - \mathbf{p}_1| - d) = \\
 &= \frac{1}{2} \left((x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 + (z_2 - z_2^{(0)})^2 \right) + \\
 &+ \lambda \left(\sqrt{(x_2 - x_1^{(0)})^2 + (y_2 - y_1^{(0)})^2 + (z_2 - z_1^{(0)})^2} - d \right). \tag{7.6.5}
 \end{aligned}$$

Координаты закрепленной точки в функцию (7.6.5) не входят. Функция (7.6.5) достигает минимума при

$$\begin{aligned}
 x_2 &= x_2^{(0)} + \frac{d - d^{(0)}}{d^{(0)}}(x_2^{(0)} - x_1^{(0)}), \\
 y_2 &= y_2^{(0)} + \frac{d - d^{(0)}}{d^{(0)}}(y_2^{(0)} - y_1^{(0)}), \\
 z_2 &= z_2^{(0)} + \frac{d - d^{(0)}}{d^{(0)}}(z_2^{(0)} - z_1^{(0)}), \\
 \lambda &= -(d - d^{(0)}).
 \end{aligned} \tag{7.6.6}$$

ские 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

$$\psi = \psi(x_1, x_2, \dots, x_n). \quad (7.7.2)$$

По определению действительная функция (7.7.2) 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 δ , 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

$$\Delta\psi = \psi(a_1 + \Delta x_1, a_2 + \Delta x_2, \dots, a_n + \Delta x_n) - \psi(a_1, a_2, \dots, a_n) \quad (7.7.3)$$

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 (7.7.3) и получим

$$\Delta\psi = \sum_{i=1}^n \frac{\partial\psi}{\partial x_i} \Delta x_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2\psi}{\partial x_i \partial x_j} \Delta x_i \Delta x_j + o(\Delta x^3), \quad (7.7.4)$$

где $o(\Delta x^3)$ — слагаемые третьего порядка относительно Δx_i . Слагаемые первого и второго порядка относительно Δx_i в (7.7.4) 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

$$d\psi = \frac{\partial\psi}{\partial x_1} dx_1 + \frac{\partial\psi}{\partial x_2} dx_2 + \dots + \frac{\partial\psi}{\partial x_n} dx_n = 0. \quad (7.7.5)$$

Для этого необходимо выполнение равенств

$$\frac{\partial\psi}{\partial x_1} = 0, \quad \frac{\partial\psi}{\partial x_2} = 0, \quad \dots, \quad \frac{\partial\psi}{\partial x_n} = 0, \quad (7.7.6)$$

при $x_1 = a_1, x_2 = a_2, \dots, x_n = a_n$. Выполнение равенств (7.7.6) является необходимым условием экстремума. Достаточным условием экстремума функции $\psi(x_1, x_2, \dots, x_n)$ является знакоопределенность второго дифференциала

$$d^2\psi = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2\psi}{\partial x_i \partial x_j} dx_i dx_j, \quad (7.7.7)$$

при $x_1 = a_1, x_2 = a_2, \dots, x_n = a_n$. Если второй дифференциал (7.7.7) представляет собой положительно определенную квадратичную форму, то функция в рассматриваемой точке имеет локальный минимум. Если второй дифференциал (7.7.7) есть отрицательно

(7.7.18). Система уравнений (7.7.18) при использовании критерия (7.7.20) примет вид

$$\begin{aligned} x_1 - x_1^{(0)} + \lambda_1 \frac{\partial f_1}{\partial x_1} + \lambda_2 \frac{\partial f_2}{\partial x_1} + \dots + \lambda_m \frac{\partial f_m}{\partial x_1} &= 0, \\ x_2 - x_2^{(0)} + \lambda_1 \frac{\partial f_1}{\partial x_2} + \lambda_2 \frac{\partial f_2}{\partial x_2} + \dots + \lambda_m \frac{\partial f_m}{\partial x_2} &= 0, \\ \dots & \\ x_n - x_n^{(0)} + \lambda_1 \frac{\partial f_1}{\partial x_n} + \lambda_2 \frac{\partial f_2}{\partial x_n} + \dots + \lambda_m \frac{\partial f_m}{\partial x_n} &= 0, \end{aligned} \quad (7.7.21)$$

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0, \\ f_2(x_1, x_2, \dots, x_n) &= 0, \\ \dots & \\ f_m(x_1, x_2, \dots, x_n) &= 0. \end{aligned}$$

Критерий (7.7.20) будем использовать при наложении вариационных связей на координаты точек и на любые другие параметры геометрических объектов.

7.8.* 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 придем к системе уравнений (7.7.21), 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 (1.10.5) метрического тензора. Пусть каждая координата является функцией некоторого параметра t . Тогда функции $x(t)$, $y(t)$, $z(t)$ 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 перемещения точек (7.6.1) заменим общим требованием минимума функционала

$$\Phi = \frac{1}{2} \int_{t_{\min}}^{t_{\max}} \left(\left(\frac{dx_1}{dt} \right)^2 + \left(\frac{dy_1}{dt} \right)^2 + \left(\frac{dz_1}{dt} \right)^2 + \left(\frac{dx_2}{dt} \right)^2 + \left(\frac{dy_2}{dt} \right)^2 + \left(\frac{dz_2}{dt} \right)^2 \right) dt, \quad (7.8.1)$$

где $x_1, y_1, z_1, x_2, y_2, z_2$ — координаты точек \mathbf{p}_1 и \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 (7.8.1) данной задачи от задачи в постановке (7.6.1) заключается в следующем. Функция (7.6.1) предполагает, что точки перемещаются по прямой линии, а в функционале (7.8.1) такое предположение отсутствует, и перемещение точек может быть произвольным. Функция (7.6.1) не зависит от пути изменения параметров, тогда как функционал (7.8.1) 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 (7.6.1) и условный экстремум функционала (7.8.1) 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

$$\Phi = \int_0^1 h \left(t, x_1, x_2, \dots, x_n, \frac{dx_1}{dt}, \frac{dx_2}{dt}, \dots, \frac{dx_n}{dt} \right) dt. \quad (7.8.2)$$

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 $x_1(t), x_2(t), \dots, x_n(t)$, и по каждой первой производной $x_1'(t), x_2'(t), \dots, x_n'(t)$ каждой функции по параметру t . В функционале (7.8.2) 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

$$x_i(t, \alpha) = x_i(t) + \alpha \delta x_i(t), \quad 0 \leq \alpha \leq 1, \quad i = 1, 2, \dots, n. \quad (7.8.3)$$

Величины $\delta x_i(t)$ называются вариациями соответствующих функций. Эти функции можно дифференцировать по параметру t :

$$\begin{aligned} \frac{d(\delta x_i(t))}{dt} &= \delta x_i' \equiv \delta \left(\frac{dx_i}{dt} \right), & \frac{d^2(\delta x_i(t))}{dt^2} &= \delta x_i'' \equiv \delta \left(\frac{d^2 x_i}{dt^2} \right), & \dots \\ \dots, & & \frac{d^k(\delta x_i(t))}{dt^k} &= \delta x_i^{(k)} \equiv \delta \left(\frac{d^k x_i}{dt^k} \right), & i = 1, 2, \dots, n. \end{aligned}$$

Подставим (7.8.3) в (7.8.2) 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 $\Phi'(0) = 0$. 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

$$\begin{aligned} \left. \frac{d\Phi(\alpha)}{d\alpha} \right|_{\alpha=0} &= \\ &= \int_0^1 (h_1 \delta x_1 + h_2 \delta x_2 + \dots + h_n \delta x_n + h_{1(1)} \delta x_1' + h_{2(1)} \delta x_2' + \dots + h_{n(1)} \delta x_n') dt, \quad (7.8.4) \end{aligned}$$

где

$$\begin{aligned}
 h_1 &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial x_1}, \\
 h_2 &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial x_2}, \\
 &\dots\dots\dots \\
 h_n &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial x_n}, \\
 h_{1(1)} &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial (dx_1/dt)}, \\
 h_{2(1)} &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial (dx_2/dt)}, \\
 &\dots\dots\dots \\
 h_{n(1)} &= \frac{\partial h(t, x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n')}{\partial (dx_n/dt)}.
 \end{aligned}$$

Таким образом, необходимым условием экстремума функционала является равенство нулю его вариации, что выражается формулой

$$\int_0^1 (h_1 \delta x_1 + h_2 \delta x_2 + \dots + h_n \delta x_n + h_{1(1)} \delta x_1' + h_{2(1)} \delta x_2' + \dots + h_{n(1)} \delta x_n') dt = 0. \quad (7.8.5)$$

Проинтегрируем по частям вторую половину слагаемых в (7.8.5)

$$\begin{aligned}
 \int_0^1 h_{i(1)} \delta x_i' dt &= \int_0^1 h_{i(1)} \frac{d(\delta x_i)}{dt} dt = \int_0^1 \left(\frac{d(h_{i(1)} \delta x_i)}{dt} - \frac{dh_{i(1)}}{dt} \delta x_i \right) dt = \\
 &= h_{i(1)} \delta x_i \Big|_0^1 - \int_0^1 \frac{dh_{i(1)}}{dt} \delta x_i dt, \quad i = 1, 2, \dots, n.
 \end{aligned}$$

В результате получим, что необходимое условие экстремума функционала (7.8.2) выражается равенством

$$\sum_{i=1}^n h_{i(1)} \delta x_i \Big|_0^1 + \int_0^1 \sum_{i=1}^n \left(\left(h_i - \frac{dh_{i(1)}}{dt} \right) \delta x_i \right) dt = 0. \quad (7.8.6)$$

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

$$\sum_{i=1}^n h_{i(1)} \delta x_i \Big|_0^1 = 0, \quad i = 1, 2, \dots, n. \quad (7.8.7)$$

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 (7.8.6) необходимо выполнение равенств

$$\begin{aligned} h_1 - \frac{dh_{1(1)}}{dt} &= 0, \\ h_2 - \frac{dh_{2(1)}}{dt} &= 0, \\ \dots\dots\dots \\ h_n - \frac{dh_{n(1)}}{dt} &= 0, \end{aligned} \quad (7.8.8)$$

так как вариации δx_i произвольны и в общем случае не равны нулю. Дифференциальные уравнения (7.8.8) называются *уравнениями Эйлера*. После дифференцирования по параметру t как сложной функции каждое из равенств (7.8.8) 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

$$\begin{aligned} \frac{\partial h}{\partial x_i} - \frac{\partial^2 h}{\partial x_i' \partial t} - \sum_{j=1}^n \frac{\partial^2 h}{\partial x_j \partial x_i'} \frac{dx_j}{dt} - \sum_{j=1}^n \frac{\partial^2 h}{\partial x_j' \partial x_i'} \frac{d^2 x_j}{dt^2} &= 0, \\ i &= 1, 2, \dots, n. \end{aligned} \quad (7.8.9)$$

Проинтегрировав систему (7.8.8), получим общее решение для экстремалей, каждая из которых содержит по две константы. Константы могут быть найдены из граничных условий $\delta x_i|_0^1 = 0$. Заметим, что система (7.8.8) не всегда имеет решение, а если решение существует, то оно может быть не единственным или не удовлетворять достаточным условиям экстремума. Для функционала (7.8.1) уравнения (7.8.9) примут более простой вид:

$$\frac{d^2 x_i}{dt^2} = 0, \quad \frac{d^2 y_i}{dt^2} = 0, \quad \frac{d^2 z_i}{dt^2} = 0, \quad i = 1, 2. \quad (7.8.10)$$

Решением системы (7.8.10) являются линейные функции $x_i(t) = C_{0xi} + C_{1xit}$, $y_i(t) = C_{0yi} + C_{1yit}$, $z_i(t) = C_{0zi} + C_{1zit}$. Константы C_{0xi} , C_{1xi} , C_{0yi} , C_{1yi} , C_{0zi} и C_{1zi} должны быть найдены из краевых условий (7.8.7).

В функционале (7.8.2) 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

$$\Phi(h) = \int_0^1 h(t, x_1, x_2, \dots, x_n, x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}, x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}, \dots, x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}) dt, \quad (7.8.11)$$

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 (7.8.3) функций и их производные в (7.8.11), найдем производную $\Phi'(\alpha)$ при $\alpha = 0$ и приравняем ее нулю. В результате получим, что необходимое условие экстремума функционала (7.8.11) выражается равенством

$$\int_0^1 \sum_{i=1}^n (h_i \delta x_i + h_{i(1)} \delta x_i^{(1)} + h_{i(2)} \delta x_i^{(2)} + \dots + h_{i(k)} \delta x_i^{(k)}) dt = 0. \quad (7.8.12)$$

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 $\delta x_i^{(k)}$

$$\int_0^1 h_{i(k)} \delta x_i^{(k)} dt = h_{i(k)} \delta x_i^{(k-1)} \Big|_0^1 - \frac{dh_{i(k)}}{dt} \delta x_i^{(k-2)} \Big|_0^1 + \frac{d^2 h_{i(k)}}{dt^2} \delta x_i^{(k-3)} \Big|_0^1 - \dots$$

$$\dots + (-1)^{k-1} \frac{d^{k-1} h_{i(k)}}{dt^{k-1}} \delta x_i \Big|_0^1 + (-1)^k \int_0^1 \frac{d^k h_{i(k)}}{dt^k} \delta x_i dt, \quad i = 1, 2, \dots, n. \quad (7.8.13)$$

В результате получим, что необходимое условие экстремума функционала (7.8.11) выражается равенством

$$\sum_{i=1}^n \left\{ h_{i(k)} \delta x_i^{(k-1)} \Big|_0^1 + \left(h_{i(k-1)} - \frac{dh_{i(k)}}{dt} \right) \delta x_i^{(k-2)} \Big|_0^1 + \right.$$

$$\left. + \left(h_{i(k-2)} - \frac{dh_{i(k-1)}}{dt} + \frac{d^2 h_{i(k)}}{dt^2} \right) \delta x_i^{(k-3)} \Big|_0^1 + \dots \right.$$

$$\left. \dots + \left(h_{i(1)} - \frac{dh_{i(2)}}{dt} + \dots + (-1)^{k-1} \frac{d^{k-1} h_{i(k)}}{dt^{k-1}} \right) \delta x_i \Big|_0^1 \right\} +$$

$$+ \sum_{i=1}^n \int_0^1 \left(h_i - \frac{dh_{i(1)}}{dt} + \frac{d^2 h_{i(2)}}{dt^2} + \dots + (-1)^k \frac{d^k h_{i(k)}}{dt^k} \right) \delta x_i dt = 0, \quad (7.8.14)$$

$$i = 1, 2, \dots, n.$$

В результате мы получили, что необходимое условие экстремума функционала (7.8.2) с условиями (7.8.17) на границе, требует выполнения уравнений Эйлера (7.8.8), равенств нулю вариаций начального положения (7.8.18), выполнения условий трансверсальности (7.8.22) и уравнений связей (7.8.17) 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

$$\begin{aligned} &\left(\frac{\partial h}{\partial(dx_1/dt)} + \lambda_1 \frac{\partial f_1}{\partial x_1} + \lambda_2 \frac{\partial f_2}{\partial x_1} + \dots + \lambda_m \frac{\partial f_m}{\partial x_1} \right) \Big|_{t=1} = 0, \\ &\left(\frac{\partial h}{\partial(dx_2/dt)} + \lambda_1 \frac{\partial f_1}{\partial x_2} + \lambda_2 \frac{\partial f_2}{\partial x_2} + \dots + \lambda_m \frac{\partial f_m}{\partial x_2} \right) \Big|_{t=1} = 0, \\ &\dots\dots\dots \\ &\left(\frac{\partial h}{\partial(dx_n/dt)} + \lambda_1 \frac{\partial f_1}{\partial x_n} + \lambda_2 \frac{\partial f_2}{\partial x_n} + \dots + \lambda_m \frac{\partial f_m}{\partial x_n} \right) \Big|_{t=1} = 0, \tag{7.8.23} \\ &f_1(x_1, x_2, \dots, x_n) \Big|_{t=1} = 0, \\ &f_2(x_1, x_2, \dots, x_n) \Big|_{t=1} = 0, \\ &\dots\dots\dots \\ &f_m(x_1, x_2, \dots, x_n) \Big|_{t=1} = 0. \end{aligned}$$

Эта система и ее решение симметричны относительно искомых 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 (7.8.17).

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

$$\Phi = \frac{1}{2} \int_0^t \sum_{i=1}^n \left(\frac{dx_i}{dt} \right)^2 dt. \tag{7.8.24}$$

Подинтегральная функция в данном случае равна половине суммы квадратов производных

$$h = \frac{1}{2} \left(\left(\frac{dx_1}{dt} \right)^2 + \left(\frac{dx_2}{dt} \right)^2 + \dots + \left(\frac{dx_n}{dt} \right)^2 \right). \tag{7.8.25}$$

В соответствии с (7.8.9) и (7.8.18) общее решение системы уравнений Эйлера для функции (7.8.25) имеет линейный вид

$$x_i(t) = tx_i + (1 - t)x_i^{(0)}, \quad i = 1, 2, \dots, n, \tag{7.8.26}$$

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 (7.8.17). Конечные значения функций должны быть найдены из системы уравнений (7.8.23), которая после подстановки (7.8.26) в функцию

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 (7.9.1) принимает минимальное значение. В декартовой прямоугольной системе координат $u^1 = x$, $u^2 = y$, $u^3 = z$, а функционал (7.9.1) 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 (7.9.1),

$$\begin{aligned} \Phi = \frac{1}{2} \int_0^1 g_{ij} \frac{du^i}{dt} \frac{du^j}{dt} dt = \frac{1}{2} \int_0^1 & \left(g_{11} \left(\frac{du^1}{dt} \right)^2 + g_{22} \left(\frac{du^2}{dt} \right)^2 + g_{33} \left(\frac{du^3}{dt} \right)^2 + \right. \\ & \left. + 2g_{12} \frac{du^1}{dt} \frac{du^2}{dt} + 2g_{23} \frac{du^2}{dt} \frac{du^3}{dt} + 2g_{13} \frac{du^1}{dt} \frac{du^3}{dt} \right) dt, \quad (7.9.2) \end{aligned}$$

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 (7.9.1) выражается формулой

$$\delta \Phi = \frac{1}{2} \int_0^1 \left(\frac{\partial g_{ij}}{\partial u^k} \frac{du^i}{dt} \frac{du^j}{dt} \delta u^k + 2g_{ij} \frac{du^i}{dt} \delta \left(\frac{du^j}{dt} \right) \right) dt, \quad (7.9.3)$$

где использовалась симметрия метрического тензора $g_{ij} = g_{ji}$. Проинтегрируем по частям вторую половину слагаемых в (7.9.3) и получим

$$\begin{aligned} \int_0^1 g_{ij} \frac{du^i}{dt} \delta \left(\frac{du^j}{dt} \right) dt &= g_{ij} \frac{du^i}{dt} \delta u^j \Big|_0^1 - \int_0^1 \frac{d(g_{ij} (du^i/dt))}{dt} \delta u^j dt = \\ &= g_{ij} \frac{du^i}{dt} \delta u^j \Big|_0^1 - \int_0^1 \left(\frac{\partial g_{ij}}{\partial u^m} \frac{du^i}{dt} \frac{du^m}{dt} + g_{ij} \frac{d^2 u^i}{dt^2} \right) \delta u^j dt. \quad (7.9.4) \end{aligned}$$

В силу равенства нулю вариаций $\delta u^j|_0^1 = 0$ в конечных точках вариация функционала (7.9.1) выразится формулой

$$\delta\Phi = \frac{1}{2} \int_0^1 \left(\frac{\partial g_{ij}}{\partial u^k} \frac{du^i}{dt} \frac{du^j}{dt} - 2 \frac{\partial g_{ik}}{\partial u^m} \frac{du^i}{dt} \frac{du^m}{dt} - 2g_{ik} \frac{d^2 u^i}{dt^2} \right) \delta u^k dt. \quad (7.9.5)$$

Необходимое условие экстремума функционала выражается равенством нулю его вариации. Отсюда в силу произвольности выбора δu^k следует, что функционал (7.9.1) достигает экстремума на кривых $u^i(t)$, $i = 1, 2, 3$, удовлетворяющих дифференциальным уравнениям

$$\frac{\partial g_{ij}}{\partial u^k} \frac{du^i}{dt} \frac{du^j}{dt} - 2 \frac{\partial g_{ik}}{\partial u^m} \frac{du^i}{dt} \frac{du^m}{dt} - 2g_{ik} \frac{d^2 u^i}{dt^2} = 0. \quad (7.9.6)$$

Изменим обозначения индексов (индексы, по которым выполняется суммирование, мы имеем право обозначать любыми буквами) и преобразуем первое и второе слагаемые в левой части равенства (7.9.6) 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 (1.10.16).

С учетом полученного вариация функционала (7.9.1) будет равна нулю при выполнении равенства

$$g_{mk} \frac{d^2 u^m}{dt^2} + \Gamma_{ij,k} \frac{du^i}{dt} \frac{du^j}{dt} = 0. \quad (7.9.7)$$

Умножив равенства (7.9.7) на g^{rk} , выполнив суммирование и используя свойство $g^{rk} g_{mk} = \delta_m^r$, где δ_k^r — символы Кронекера (1.10.7), получим

$$\frac{d^2 u^r}{dt^2} + \Gamma_{ij}^r \frac{du^i}{dt} \frac{du^j}{dt} = 0. \quad (7.9.8)$$

где Γ_{ij}^r — символы Кристоффеля 2-го рода (1.10.19). Таким образом, геодезические линии должны удовлетворять уравнениям (7.9.8). Эти уравнения представляют собой уравнения Эйлера (7.8.8) для частного случая функционала (7.9.1), но в общем случае криволинейных координат. Геодезические линии являются кратчайшими линиями, соединяющими две заданные точки. В евклидовом пространстве кратчайшей линией является прямая линия. Если в евклидовом пространстве мы построим криволинейную систему координат, то прямые линии в этой системе координат будут описываться уравнениями (7.9.8).

Для фиксации скалярных параметров, например, радиуса окружности или угла дуги окружности, используются *фиксирующие связи* типа (7.2.1). Для закрепления двумерной точки применяется векторное уравнение (7.2.2), состоящее из двух скалярных уравнений (7.2.3).

Размер вдоль координаты. *Размер вдоль одной координаты* между двумямерными точками $p_1 = [x_1 \ y_1]^T$ и $p_2 = [x_2 \ y_2]^T$ описывается одним из уравнений

$$x_2 - x_1 - d = 0, \quad \text{или} \quad y_2 - y_1 - d = 0, \quad (7.10.1)$$

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

$$\begin{aligned} \psi &= \frac{1}{2} \left(|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 \right) = \\ &= \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 \right) \end{aligned} \quad (7.10.2)$$

была минимальной. Для определенности будем считать, что размер поставлен вдоль координат x . Необходимым условием минимума функции (7.10.2) при условии (7.10.1) является равенство нулю частных производных по параметрам $x_1, y_1, x_2, y_2, \lambda$ функции

$$\begin{aligned} F \doteq \psi + \lambda f &= \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 \right) + \\ &+ \lambda(x_2 - x_1 - d). \end{aligned} \quad (7.10.3)$$

Искомые координаты точек и параметр λ найдем из системы уравнений

$$\begin{aligned} x_2 - x_1 &= d, \\ x_1 - x_1^{(0)} - \lambda &= 0, \quad y_1 - y_1^{(0)} = 0, \\ x_2 - x_2^{(0)} + \lambda &= 0, \quad y_2 - y_2^{(0)} = 0. \end{aligned} \quad (7.10.4)$$

Данная система уравнений имеет решение

$$\begin{aligned} x_1 &= \frac{1}{2}(x_2^{(0)} + x_1^{(0)} - d), \quad y_1 = y_1^{(0)}, \\ x_2 &= \frac{1}{2}(x_2^{(0)} + x_1^{(0)} + d), \quad y_2 = y_2^{(0)}, \\ \lambda &= \frac{1}{2}(x_2^{(0)} - x_1^{(0)} - d). \end{aligned} \quad (7.10.5)$$

Можно доказать, что найденное решение является точкой минимума функции (7.10.2). 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 (7.2.2) и ее координаты не варьируются. Пусть точка p_2 закреплена. Тогда система будет иметь решение

$$\begin{aligned} x_1 &= x_2^{(0)} - d, & y_1 &= y_1^{(0)}, & x_2 &= x_2^{(0)}, & y_2 &= y_2^{(0)}, \\ \lambda &= x_2^{(0)} - x_1^{(0)} - d. \end{aligned} \quad (7.10.6)$$

Размер вдоль направления. *Размер вдоль заданного направления* между двумерными точками $p_1 = [x_1 \ y_1]^\top$ и $p_2 = [x_2 \ y_2]^\top$ описывается уравнением (7.3.12), которое в координатной записи имеет вид

$$(x_2 - x_1)m_x + (y_2 - y_1)m_y - d = 0, \quad (7.10.7)$$

где $m = [m_x \ m_y]^\top$ — задающий направление двумерный вектор единичной длины. Найдем положение точек из условия минимума их суммарного перемещения (7.10.2). В соответствии с методом Лагранжа составим функцию

$$\begin{aligned} F &= \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 \right) + \\ &\quad + \lambda \left((x_2 - x_1)m_x + (y_2 - y_1)m_y - d \right). \end{aligned} \quad (7.10.8)$$

Искомые координаты точек и параметр λ найдем из равенство нулю частных производных этой функции по параметрам $x_1, y_1, x_2, y_2, \lambda$:

$$\begin{aligned} (x_2 - x_1)m_x + (y_2 - y_1)m_y - d &= 0, \\ x_1 - x_1^{(0)} - \lambda m_x &= 0, & y_1 - y_1^{(0)} - \lambda m_y &= 0, \\ x_2 - x_2^{(0)} + \lambda m_x &= 0, & y_2 - y_2^{(0)} + \lambda m_y &= 0. \end{aligned} \quad (7.10.9)$$

Данная система уравнений имеет решение

$$\begin{aligned} x_1 &= x_1^{(0)} + \frac{m_x}{2} \left((x_2^{(0)} - x_1^{(0)})m_x + (y_2^{(0)} - y_1^{(0)})m_y - d \right), \\ y_1 &= y_1^{(0)} + \frac{m_y}{2} \left((x_2^{(0)} - x_1^{(0)})m_x + (y_2^{(0)} - y_1^{(0)})m_y - d \right), \\ x_2 &= x_2^{(0)} - \frac{m_x}{2} \left((x_2^{(0)} - x_1^{(0)})m_x + (y_2^{(0)} - y_1^{(0)})m_y - d \right), \\ y_2 &= y_2^{(0)} - \frac{m_y}{2} \left((x_2^{(0)} - x_1^{(0)})m_x + (y_2^{(0)} - y_1^{(0)})m_y - d \right), \\ \lambda &= \frac{1}{2} \left((x_2^{(0)} - x_1^{(0)})m_x + (y_2^{(0)} - y_1^{(0)})m_y - d \right). \end{aligned} \quad (7.10.10)$$

Перемещение точек производится вдоль вектора 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 $p_1 = [x_1 \ y_1]^\top$ и $p_2 = [x_2 \ y_2]^\top$ описывается уравнением (7.3.1), которое в координатной записи имеет вид

$$f = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d = 0. \quad (7.10.11)$$

Для симметричного поведения точек при установки размера потребуем, чтобы суммарное перемещение связанных точек было минимальным. Необходимым условием минимума функции (7.10.2) при условии (7.10.11) является равенство нулю частных производных по параметрам функции

$$\begin{aligned} F &= \frac{1}{2} \left(|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 \right) + \lambda (|p_2 - p_1| - d) = \\ &= \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 \right) + \\ &\quad + \lambda \left(\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d \right). \end{aligned} \quad (7.10.12)$$

Система уравнений, из которой найдется искомое положение точек, аналогична системе (7.6.3). Решение такой системы уравнений имеет вид

$$\begin{aligned} x_1 &= x_1^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}} (x_1^{(0)} - x_2^{(0)}), \\ y_1 &= y_1^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}} (y_1^{(0)} - y_2^{(0)}), \\ x_2 &= x_2^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}} (x_2^{(0)} - x_1^{(0)}), \\ y_2 &= y_2^{(0)} + \frac{d - d^{(0)}}{2d^{(0)}} (y_2^{(0)} - y_1^{(0)}), \\ \lambda &= -\frac{1}{2} (d - d^{(0)}). \end{aligned} \quad (7.10.13)$$

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

$$F = \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 \right) + \lambda (x_2 - x_1) \quad (7.10.14)$$

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

$$\begin{aligned} x_1 &= \frac{1}{2} (x_1^{(0)} + x_2^{(0)}), & x_2 &= \frac{1}{2} (x_2^{(0)} + x_1^{(0)}), \\ \lambda &= \frac{1}{2} (x_2^{(0)} - x_1^{(0)}). \end{aligned} \quad (7.10.15)$$

Симметрия двух точек. Рассмотрим симметрию точек $p_1 = [x_1 \ y_1]^T$ и $p_2 = [x_2 \ y_2]^T$ относительно прямой линии

$$l(t) = p + ti, \quad (7.10.16)$$

что показано на рис. 7.10.1. Прямую будем считать неподвижной, поэтому ее параметры варьироваться не будут.

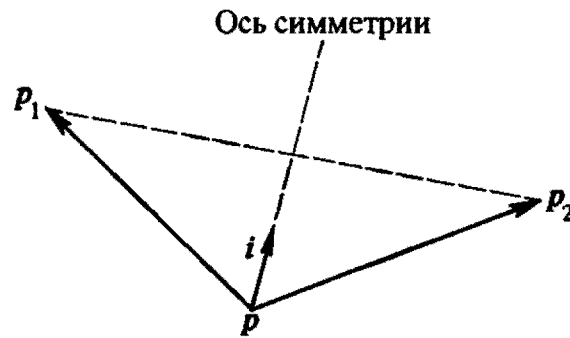


Рис. 7.10.1. Симметрия двумерных точек относительно прямой

Симметрия точек относительно прямой линии описывается двумя уравнениями

$$\begin{aligned} (p_1 - p) \cdot i &= (p_2 - p) \cdot i, \\ (p_1 - p) \cdot j + (p_2 - p) \cdot j &= 0, \end{aligned} \quad (7.10.17)$$

где вектор $j = [-b \ a]^T$ ортогонален вектору $i = [a \ b]^T$. Пусть точка p имеет координаты x и y . Составим функцию

$$\begin{aligned} F &= \frac{1}{2} \left((x_1 - x_1^{(0)})^2 + (y_1 - y_1^{(0)})^2 + (x_2 - x_2^{(0)})^2 + (y_2 - y_2^{(0)})^2 \right) + \\ &+ \lambda_1 \left((x_1 - x_2)a + (y_1 - y_2)b \right) + \lambda_2 \left((y_1 + y_2 - 2y)a - (x_1 + x_2 - 2x)b \right). \end{aligned} \quad (7.10.18)$$

Систему уравнений для определения положения точек получим из равенства нулю частных производных функции (7.10.18)

$$\begin{aligned} x_1 - x_1^{(0)} + \lambda_1 a - \lambda_2 b &= 0, \\ y_1 - y_1^{(0)} + \lambda_1 b + \lambda_2 a &= 0, \\ x_2 - x_2^{(0)} - \lambda_1 a - \lambda_2 b &= 0, \\ y_2 - y_2^{(0)} - \lambda_1 b + \lambda_2 a &= 0, \\ (x_1 - x_2)a + (y_1 - y_2)b &= 0, \\ (y_1 + y_2 - 2y)a - (x_1 + x_2 - 2x)b &= 0. \end{aligned}$$

Решение этой системы определит новое положение точек.

Если мы хотим, чтобы точка p_1 осталась неподвижной, то введем уравнение

$$p_1 = p_1^{(0)} \quad (7.10.19)$$

или в (7.10.18) ее координаты будем считать константами. Если мы хотим, чтобы точка \mathbf{p}_2 осталась неподвижной, а точка \mathbf{p}_1 стала ей симметрична, то введем уравнение

$$\mathbf{p}_2 = \mathbf{p}_2^{(0)} \quad (7.10.20)$$

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 $\mathbf{p}_1 = [x_1 \ y_1]^\top$, $\mathbf{p}_2 = [x_2 \ y_2]^\top$ и $\mathbf{p}_3 = [x_3 \ y_3]^\top$ описывается уравнением (7.3.28), где

$$s = (x_1 - x_2)(y_3 - y_2) - (y_1 - y_2)(x_3 - x_2), \quad (7.10.21)$$

$$c = (x_1 - x_2)(x_3 - x_2) + (y_1 - y_2)(y_3 - y_2). \quad (7.10.22)$$

Величина s равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_2$ на $\sin \alpha$. Величина c равна произведению длин векторов $\mathbf{p}_1 - \mathbf{p}_2$ и $\mathbf{p}_3 - \mathbf{p}_2$ на $\cos \alpha$. Угловой размер на плоскости связывает шесть параметров. Составим функцию

$$F = \frac{1}{2} \left(|\mathbf{p}_1 - \mathbf{p}_1^{(0)}|^2 + |\mathbf{p}_2 - \mathbf{p}_2^{(0)}|^2 + |\mathbf{p}_3 - \mathbf{p}_3^{(0)}|^2 \right) + \lambda \left(\operatorname{arctg} \frac{s}{c} - \alpha \right), \quad (7.10.23)$$

где $\mathbf{p}_1^{(0)} = [x_1^{(0)} \ y_1^{(0)}]^\top$, $\mathbf{p}_2^{(0)} = [x_2^{(0)} \ y_2^{(0)}]^\top$ и $\mathbf{p}_3^{(0)} = [x_3^{(0)} \ y_3^{(0)}]^\top$ — исходные положения точек. Систему уравнений для определения параметров получим из равенства нулю частных производных функции (7.10.23)

$$\begin{aligned} x_1 - x_1^{(0)} + \lambda \frac{c(y_3 - y_2) - s(x_3 - x_2)}{s^2 + c^2} &= 0, \\ y_1 - y_1^{(0)} + \lambda \frac{-c(x_3 - x_2) - s(y_3 - y_2)}{s^2 + c^2} &= 0, \\ x_2 - x_2^{(0)} + \lambda \frac{c(y_1 - y_3) - s(2x_2 - x_1 - x_3)}{s^2 + c^2} &= 0, \\ y_2 - y_2^{(0)} + \lambda \frac{c(x_3 - x_1) - s(2y_2 - y_1 - y_3)}{s^2 + c^2} &= 0, \\ x_3 - x_3^{(0)} + \lambda \frac{-c(y_1 - y_2) - s(x_1 - x_2)}{s^2 + c^2} &= 0, \\ y_3 - y_3^{(0)} + \lambda \frac{c(x_1 - x_2) - s(y_1 - y_2)}{s^2 + c^2} &= 0, \\ \operatorname{arctg} \left(\frac{(x_1 - x_2)(y_3 - y_2) - (x_3 - x_2)(y_1 - y_2)}{(x_1 - x_2)(x_3 - x_2) + (y_1 - y_2)(y_3 - y_2)} \right) &= \alpha. \end{aligned} \quad (7.10.24)$$

Решение этой системы определит положение точек при заданном угле α .

7.11. 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

$$r_1(t_1) = (1 - t_1)p_1 + t_1p_2, \quad (7.11.1)$$

$$r_3(t_3) = (1 - t_3)p_3 + t_3p_4, \quad (7.11.2)$$

построенных по точкам $p_1 = [x_1 \ y_1]^T$, $p_2 = [x_2 \ y_2]^T$, $p_3 = [x_3 \ y_3]^T$ и $p_4 = [x_4 \ y_4]^T$. На данные четыре точки можно наложить вариационную связь, определяющую угол между отрезками $r_1(t_1)$ и $r_3(t_3)$. Эта связь аналогична угловому размеру и описывается тем же уравнением

$$\operatorname{arctg} \frac{s}{c} = \alpha, \quad (7.11.3)$$

где α — заданный угол. Синус и косинус угла пропорциональны величинам:

$$s = (x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3), \quad (7.11.4)$$

$$c = (x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3). \quad (7.11.5)$$

Уравнение (7.11.3) связывает восемь параметров.

Ортогональность отрезков. Как частный случай связи (7.11.3) может рассматриваться вариационная связь устанавливающая *ортогональность отрезков*. Данная вариационная связь описывается уравнением

$$(p_2 - p_1) \cdot (p_4 - p_3) = (x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3) = 0. \quad (7.11.6)$$

Систему уравнений для определения положения точек получим из равенства нулю частных производных функции

$$F = \frac{1}{2} \left(|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 + |p_3 - p_3^{(0)}|^2 + |p_4 - p_4^{(0)}|^2 \right) + \lambda \left((x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3) \right), \quad (7.11.7)$$

где $p_1^{(0)} = [x_1^{(0)} \ y_1^{(0)}]^T$, $p_2^{(0)} = [x_2^{(0)} \ y_2^{(0)}]^T$, $p_3^{(0)} = [x_3^{(0)} \ y_3^{(0)}]^T$ и $p_4^{(0)} = [x_4^{(0)} \ y_4^{(0)}]^T$ — исходные положения точек, в общем случае не удовлетворяющие уравнению (7.11.6).

Параллельность отрезков. Другим частным случаем связи (7.11.3) является вариационная связь, устанавливающая *параллельность отрезков*. Она определяется уравнением

$$(x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3) = 0. \quad (7.11.8)$$

Пусть в исходном состоянии координаты точек $p_1^{(0)} = [x_1^{(0)} \ y_1^{(0)}]^\top$, $p_2^{(0)} = [x_2^{(0)} \ y_2^{(0)}]^\top$, $p_3^{(0)} = [x_3^{(0)} \ y_3^{(0)}]^\top$ и $p_4^{(0)} = [x_4^{(0)} \ y_4^{(0)}]^\top$ не удовлетворяют уравнению (7.11.7). Применим метод минимизации суммарного изменения параметров. Систему уравнений для определения положения точек получим из равенства нулю частных производных функции

$$F = \frac{1}{2} \left(|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 + |p_3 - p_3^{(0)}|^2 + |p_4 - p_4^{(0)}|^2 \right) + \lambda \left((x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3) \right). \quad (7.11.9)$$

Для простоты предположим, что точки p_2 , p_3 и p_4 зафиксированы и, следовательно, не изменяют своего положения. Параллельность отрезков (7.11.1) и (7.11.2) будет достигнута перемещением точки p_1 . Система уравнений будет иметь вид

$$\begin{aligned} x_1 - x_1^{(0)} - \lambda(y_4 - y_3) &= 0, \\ y_1 - y_1^{(0)} + \lambda(x_4 - x_3) &= 0, \\ (x_2 - x_1)(y_4 - y_3) - (y_2 - y_1)(x_4 - x_3) &= 0. \end{aligned} \quad (7.11.10)$$

Эта система имеет решение

$$\begin{aligned} x_1 &= x_1^{(0)} + (y_4^{(0)} - y_3^{(0)}) \frac{h^{(0)}}{d^{(0)}}, \\ y_1 &= y_1^{(0)} - (x_4^{(0)} - x_3^{(0)}) \frac{h^{(0)}}{d^{(0)}}, \\ \lambda &= \frac{h^{(0)}}{d^{(0)}}, \end{aligned} \quad (7.11.11)$$

где $d^{(0)} = |p_4^{(0)} - p_3^{(0)}|^2$, $h^{(0)} = (x_2^{(0)} - x_1^{(0)})(y_4^{(0)} - y_3^{(0)}) - (y_2^{(0)} - y_1^{(0)}) \times (x_4^{(0)} - x_3^{(0)})$. Перемещение точки p_1 происходит по нормали к отрезку, проходящему через точки $p_4^{(0)} - p_3^{(0)}$ (рис. 7.11.1).

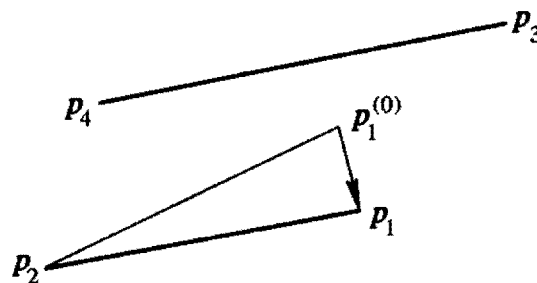


Рис. 7.11.1. Параллельность отрезков прямой линии

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 и 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

$$\begin{aligned}x_3^{(0)} &= x_4^{(0)} + a \cos \varphi, \\y_3^{(0)} &= y_4^{(0)} + a \sin \varphi.\end{aligned}\tag{7.11.12}$$

В соответствии с решением (7.11.11) 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 $r = b \cos \varphi$. 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 . При изменении угла φ от 0 до π точка p_3 сделает

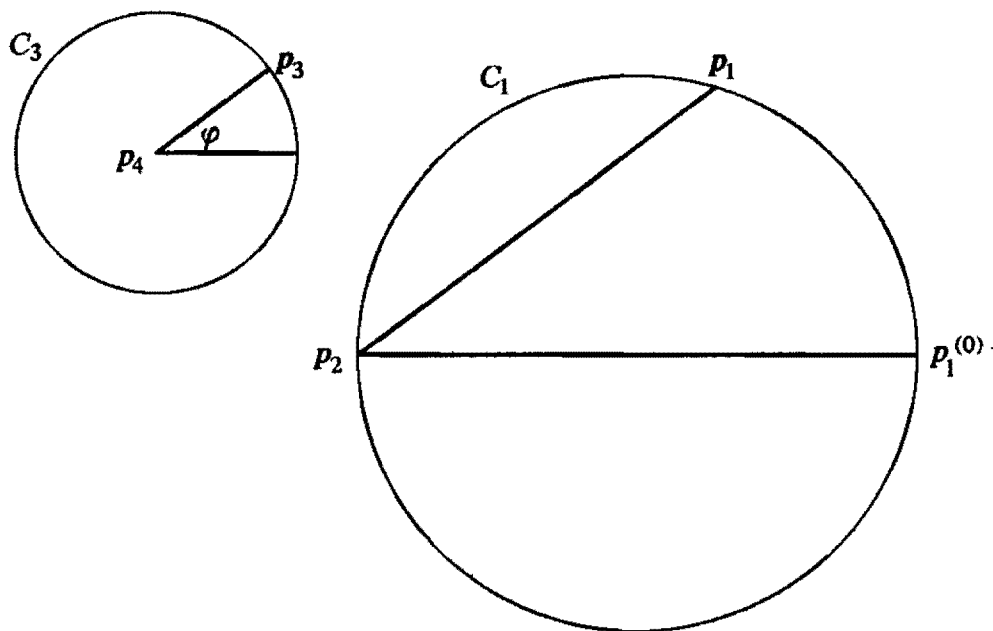


Рис. 7.11.2. Траектории движения отрезка p_1p_2 , параллельно зависящего от отрезка p_3p_4 , при закрепленных точках p_2 и p_4

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 и p_3 показаны на рис. 7.11.2. 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

$$r_p(t) = \frac{\sum_{i=1}^n N_{i,m}(t)w_i p_i}{\sum_{i=1}^n N_{i,m}(t)w_i}, \quad t_{\min} \leq t \leq t_{\max}, \quad (7.11.13)$$

и построена на точках p_i , $i = 1, 2, \dots, n$ с весами w_i , а другая описывается функцией

$$r_q(u) = \frac{\sum_{j=1}^k N_{j,l}(u)z_j q_j}{\sum_{j=1}^k N_{j,l}(u)z_j}, \quad u_{\min} \leq u \leq u_{\max}, \quad (7.11.14)$$

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,

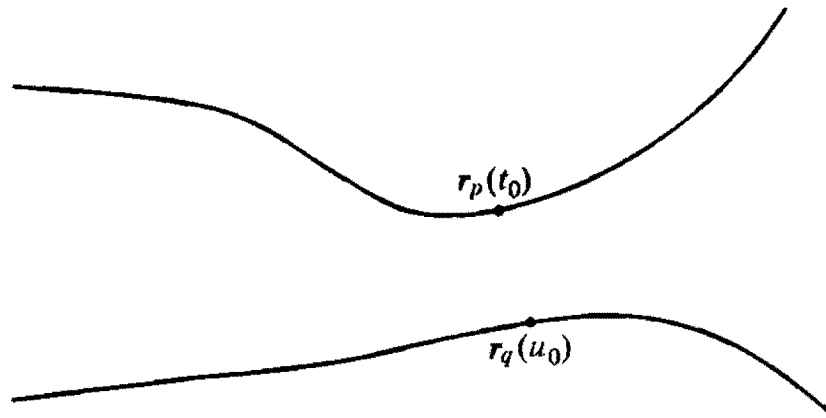


Рис. 7.11.3. NURBS кривые будут касаться ближайшими точками

другая — на второй, касательные в которых параллельны. Эти точки должны удовлетворять уравнениям (4.8.9), в данном случае имеющим вид

$$\begin{aligned} (r_p(t) - r_q(u)) \cdot \frac{dr_p}{dt} &= 0, \\ (r_q(u) - r_p(t)) \cdot \frac{dr_q}{du} &= 0. \end{aligned} \quad (7.11.15)$$

Пусть этим точкам соответствуют параметры t_0 и u_0 (рис. 7.11.3).

Точки на кривых выражаются в виде сумм

$$r_p(t_0) = P_1 p_1 + P_2 p_2 + \dots + P_n p_n = \sum_{i=1}^n P_i p_i, \quad (7.11.16)$$

$$r_q(u_0) = Q_1 q_1 + Q_2 q_2 + \dots + Q_k q_k = \sum_{j=1}^k Q_j q_j, \quad (7.11.17)$$

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 $p_i = [x_i \ y_i]^\top$, $i = 1, 2, \dots, n$, $q_j = [a_j \ b_j]^\top$, $j = 1, 2, \dots, k$. Если бы точки $r_p(t_0)$ и $r_q(u_0)$ совпали, то являлись бы решением задачи касания кривых (7.11.13) и (7.11.14). Поэтому в качестве уравнения связи возьмем уравнение

$$\begin{aligned} |r_p(t_0) - r_q(u_0)| &= |(P_1 p_1 + P_2 p_2 + \dots + P_n p_n) - (Q_1 q_1 + Q_2 q_2 + \dots + Q_k q_k)| = \\ &= \sqrt{\left(\sum_{i=1}^n P_i x_i - \sum_{j=1}^k Q_j a_j\right)^2 + \left(\sum_{i=1}^n P_i y_i - \sum_{j=1}^k Q_j b_j\right)^2} = 0. \end{aligned} \quad (7.11.18)$$

Составим функцию суммарного перемещения точек касающихся кривых совместно с уравнением связи (7.11.18). Эта функция имеет вид

$$\begin{aligned} F = \frac{1}{2} &\left(|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 + \dots + |p_n - p_n^{(0)}|^2 + \right. \\ &\left. + |q_1 - q_1^{(0)}|^2 + |q_2 - q_2^{(0)}|^2 + \dots + |q_k - q_k^{(0)}|^2 \right) + \\ &+ \lambda |r_p(t_0) - r_q(u_0)|. \end{aligned} \quad (7.11.19)$$

Систему уравнений для определения положения точек получим из равенства нулю частных производных функции (7.11.19)

$$\begin{aligned} x_i - x_i^{(0)} + \lambda P_i \frac{x_p(t_0) - a_q(u_0)}{\sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2}} &= 0, \quad i = 1, 2, \dots, n, \\ y_i - y_i^{(0)} + \lambda P_i \frac{y_p(t_0) - b_q(u_0)}{\sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2}} &= 0, \quad i = 1, 2, \dots, n, \\ a_j - a_j^{(0)} - \lambda Q_j \frac{x_p(t_0) - a_q(u_0)}{\sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2}} &= 0, \quad j = 1, 2, \dots, k, \\ b_j - b_j^{(0)} - \lambda Q_j \frac{y_p(t_0) - b_q(u_0)}{\sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2}} &= 0, \quad j = 1, 2, \dots, k, \\ \sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2} &= 0, \end{aligned} \quad (7.11.20)$$

где

$$x_p(t_0) = P_1x_1 + P_2x_2 + \dots + P_nx_n = \sum_{i=1}^n P_i x_i,$$

$$y_p(t_0) = P_1y_1 + P_2y_2 + \dots + P_ny_n = \sum_{i=1}^n P_i y_i,$$

$$a_q(u_0) = Q_1a_1 + Q_2a_2 + \dots + Q_ka_k = \sum_{j=1}^k Q_j a_j,$$

$$b_q(u_0) = Q_1b_1 + Q_2b_2 + \dots + Q_kb_k = \sum_{j=1}^k Q_j b_j.$$

Система уравнений (7.11.20) на $(r+1)$ -й итерации метода Ньютона дает приращения искомых функций, равные:

$$\Delta x_i^{(r+1)} = -\frac{P_i}{s} \left(\sum_{m=1}^n P_m x_m^{(r)} - \sum_{m=1}^k Q_m a_m^{(r)} \right), \quad i = 1, 2, \dots, n,$$

$$\Delta y_i^{(r+1)} = -\frac{P_i}{s} \left(\sum_{m=1}^n P_m y_m^{(r)} - \sum_{m=1}^k Q_m b_m^{(r)} \right), \quad i = 1, 2, \dots, n,$$

$$\Delta a_j^{(r+1)} = \frac{Q_j}{s} \left(\sum_{m=1}^n P_m x_m^{(r)} - \sum_{m=1}^k Q_m a_m^{(r)} \right), \quad j = 1, 2, \dots, k,$$

$$\Delta b_j^{(r+1)} = \frac{Q_j}{s} \left(\sum_{m=1}^n P_m y_m^{(r)} - \sum_{m=1}^k Q_m b_m^{(r)} \right), \quad j = 1, 2, \dots, k.$$

(7.11.21)

Значение коэффициента λ на текущей итерации равно

$$\lambda = \frac{1}{s} \sqrt{\left(\sum_{i=1}^n P_i x_i - \sum_{j=1}^k Q_j a_j \right)^2 + \left(\sum_{i=1}^n P_i y_i - \sum_{j=1}^k Q_j b_j \right)^2} =$$

$$= \frac{1}{s} \sqrt{(x_p(t_0) - a_q(u_0))^2 + (y_p(t_0) - b_q(u_0))^2}, \quad (7.11.22)$$

где $s = \sum_{i=1}^n P_i^2 + \sum_{j=1}^k Q_j^2$. 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 (7.11.18), что обращает в нуль числители и знаменатели последних слагаемых уравнений системы (7.11.20). Если на очередной r -й итерации выполнилось ра-

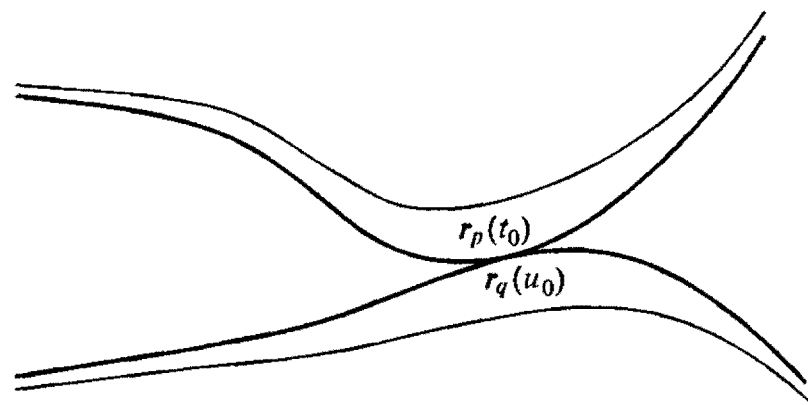


Рис. 7.11.4. Модификация NURBS кривых при наложении связи касания

венство (7.11.18), 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. 7.11.4.

Касание отрезка и окружности. Решим задачу касания в частных случаях. Касание отрезка $r_1(t)$ и окружности $r_0(u)$,

$$\begin{aligned} r_1(t) &= (1-t)p_1 + tp_2, \\ r_0(u) &= p_0 + i_x r \cos u + i_y r \sin u, \end{aligned} \quad (7.11.23)$$

описывается уравнением

$$|(1-t_0)p_1 + t_0p_2 - p_0| = r, \quad (7.11.24)$$

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, $t_0 = \frac{(p_0 - p_1) \cdot (p_2 - p_1)}{(p_2 - p_1) \cdot (p_2 - p_1)}$ — значение параметра отрезка, соответствующее про-

екции 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

$$\begin{aligned} F &= \frac{1}{2} \left(|p_1 - p_1^{(0)}|^2 + |p_2 - p_2^{(0)}|^2 + |p_0 - p_0^{(0)}|^2 + (r - r^{(0)})^2 \right) + \\ &\quad + \lambda (|(1-t_0)p_1 + t_0p_2 - p_0| - r), \end{aligned} \quad (7.11.25)$$

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

$$x_0 - x_0^{(0)} - \lambda \frac{x_0}{\sqrt{((1-t_0)x_1 + t_0x_2 - x_0)^2 + ((1-t_0)y_1 + t_0y_2 - y_0)^2}} = 0,$$

$$y_0 - y_0^{(0)} - \lambda \frac{y_0}{\sqrt{((1-t_0)x_1 + t_0x_2 - x_0)^2 + ((1-t_0)y_1 + t_0y_2 - y_0)^2}} = 0,$$

$$x_1 - x_1^{(0)} + \lambda \frac{(1-t_0)x_1}{\sqrt{((1-t_0)x_1 + t_0x_2 - x_0)^2 + ((1-t_0)y_1 + t_0y_2 - y_0)^2}} = 0,$$

$$y_1 - y_1^{(0)} + \lambda \frac{(1-t_0)y_1}{\sqrt{((1-t_0)x_1 + t_0x_2 - x_0)^2 + ((1-t_0)y_1 + t_0y_2 - y_0)^2}} = 0,$$

$$x_2 - x_2^{(0)} + \lambda \frac{t_0x_2}{\sqrt{((1-t_0)x_1 + t_0x_2 - x_0)^2 + ((1-t_0)y_1 + t_0y_2 - y_0)^2}} = 0,$$

$$y_2 - y_2^{(0)} + \lambda \frac{t_0y_2}{\sqrt{((1-t_0)x_1 + t_0x_2 - x_0)^2 + ((1-t_0)y_1 + t_0y_2 - y_0)^2}} = 0,$$

$$r - r^{(0)} - \lambda = 0,$$

$$\sqrt{((1-t_0)x_1 + t_0x_2 - x_0)^2 + ((1-t_0)y_1 + t_0y_2 - y_0)^2} - r = 0.$$

(7.11.26)

Система уравнений (7.11.26) решается итерационно. Перед началом новой итерации необходимо заново вычислить параметр t_0 . Поведение отрезка и окружности приведено на рис. 7.11.5.

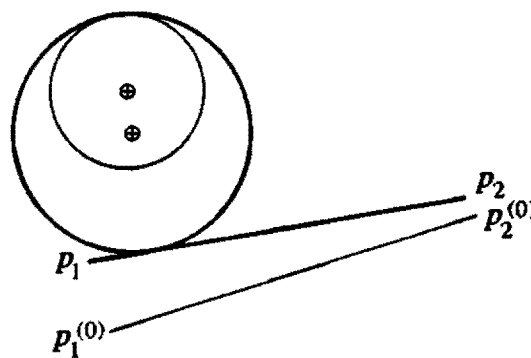


Рис. 7.11.5. Поведение окружности и отрезка при касании

Касание окружностей. Касание двух окружностей

$$r_1(t_1) = q_1 + i_x r_1 \cos t_1 + i_y r_1 \sin t_1, \quad (7.11.27)$$

$$r_2(t_2) = q_2 + i_x r_2 \cos t_2 + i_y r_2 \sin t_2, \quad (7.11.28)$$

описывается уравнением

$$|q_1 - q_2| = |r_1 \pm r_2|. \quad (7.11.29)$$

В данном случае радиусы окружностей также являются варьируемыми параметрами. Уравнение (7.11.29) связывает шесть параметров. Систему уравнений для определения параметров связи получим, приравняв нулю частные производ-

ные по параметрам функции

$$F = \frac{1}{2} \left(|q_1 - q_1^{(0)}|^2 + |q_2 - q_2^{(0)}|^2 + (r_1 - r_1^{(0)})^2 + (r_2 - r_2^{(0)})^2 \right) + \lambda (|q_1 - q_2| - |r_1 \pm r_2|), \quad (7.11.30)$$

где $q_1^{(0)}$, $q_2^{(0)}$ — исходные положения центров, $r_1^{(0)}$ и $r_2^{(0)}$ — 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

$$\begin{aligned} x_1 - x_1^{(0)} + \lambda \frac{x_1 - x_2}{d} &= 0, & y_1 - y_1^{(0)} + \lambda \frac{y_1 - y_2}{d} &= 0, \\ x_2 - x_2^{(0)} - \lambda \frac{x_1 - x_2}{d} &= 0, & y_2 - y_2^{(0)} - \lambda \frac{y_1 - y_2}{d} &= 0, \end{aligned} \quad (7.11.31)$$

$$r_1 - r_1^{(0)} - \lambda = 0, \quad r_2 - r_2^{(0)} - \lambda = 0, \quad |q_1 - q_2| - (r_1 + r_2) = 0,$$

где $d = |q_1 - q_2|$. Эта система имеет решение

$$\begin{aligned} x_1 &= x_1^{(0)} - \frac{1}{4} a^{(0)} \frac{x_1^{(0)} - x_2^{(0)}}{d^{(0)}}, & y_1 &= y_1^{(0)} - \frac{1}{4} a^{(0)} \frac{y_1^{(0)} - y_2^{(0)}}{d^{(0)}}, \\ x_2 &= x_2^{(0)} + \frac{1}{4} a^{(0)} \frac{x_1^{(0)} - x_2^{(0)}}{d^{(0)}}, & y_2 &= y_2^{(0)} + \frac{1}{4} a^{(0)} \frac{y_1^{(0)} - y_2^{(0)}}{d^{(0)}}, \\ r_1 &= r_1^{(0)} + \frac{1}{4} a^{(0)}, & r_2 &= r_2^{(0)} + \frac{1}{4} a^{(0)}, & \lambda &= \frac{1}{4} a^{(0)}, \end{aligned} \quad (7.11.32)$$

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

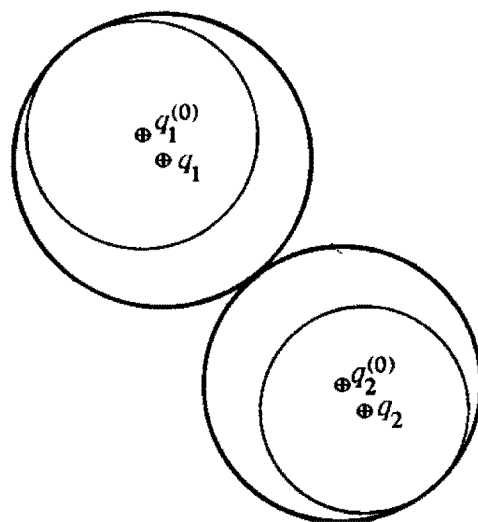


Рис. 7.11.6. Касание окружностей

центров окружностей. Поведение окружностей приведено на рис. 7.11.6. При расположении окружностей одна внутри другой в уравнениях (7.11.31) изменяется знак перед одним из радиусов.

7.12. 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. 7.12.1. Каждая связь имеет одно или несколько уравнений и перечень параметров, участвующих в каждом уравнении.

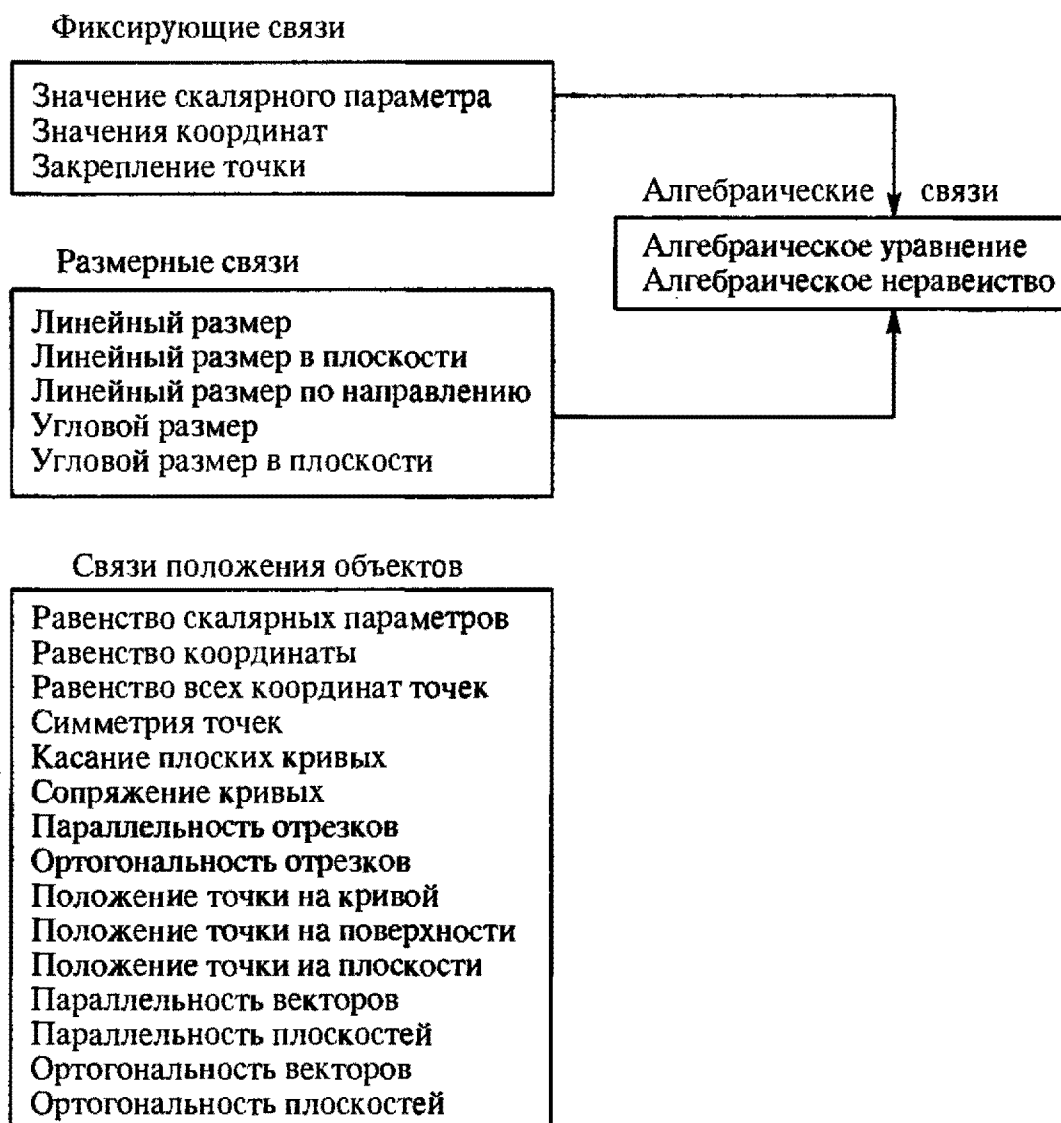


Рис. 7.12.1. Типы вариационных связей

Нормальным состоянием для вариационных связей является состояние, когда их уравнения выполняются. Будем называть это состояние *равновесным*. В процессе моделирования придется модифицировать геометрические объекты или их взаимное положение. Таким образом, вариационные связи выводятся

описанных в гл. 4. Если на какие-либо из параметров наложены фиксирующие связи (7.2.1), то их можно исключить из функции (7.12.2) и при решении общей системы уравнений считать константами.

В большинстве случаев управление данными происходит следующим образом. На некотором этапе система уравнений связей (7.12.1) выполняется. Далее одному или нескольким параметрам присваивается новое значение, и эти параметры фиксируются. Например, мы переместили некоторую точку геометрического объекта и хотим узнать, как это повлияет на другие варьируемые параметры. Управление данными геометрических объектов может происходить путем изменения констант уравнений, что также выведет систему уравнений связей из состояния равновесия. Например, мы изменили значение некоторого размера. В результате для исходных значений параметров некоторые уравнения связей перестанут выполняться. Необходимо найти новые значения параметров, удовлетворяющие системе уравнений связей, и перестроить геометрические объекты. Система вариационных связей выходит из равновесия также при добавлении к ней новых связей. Например, мы хотим, чтобы некоторые две плоскости геометрической модели были параллельными. При этом рождается новая вариационная связь, которая добавляет в систему уравнений свои уравнения. Нужно решить систему уравнений связей и перестроить геометрические объекты.

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 потребуется 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 (7.12.4). 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 (7.12.1) и (7.12.4) 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.

Глава 8

ВЫЧИСЛЕНИЕ ГЕОМЕТРИЧЕСКИХ ХАРАКТЕРИСТИК

8.1. Возможности геометрической модели

Результатом геометрического моделирования некоторого объекта является математическая модель его геометрии. Математическая модель позволяет *графически отобразить* моделируемый объект, получить его *геометрические характеристики*, выполнить *исследование* многих *физических свойств* объекта путем постановки численных экспериментов, *подготовить производство* и, наконец, *изготовить* объект.

Для того чтобы увидеть, как выглядит объект, нужно смоделировать поток падающих и возвращающихся от его поверхностей лучей света. При этом граням модели можно придать необходимый цвет, прозрачность, фактуру и другие физические свойства. Модель можно осветить с разных сторон светом различного цвета и интенсивности.

Геометрическая модель позволяет определить массово-центровочные и инерционные характеристики проектируемого объекта, выполнить измерения длин и углов его элементов. Она дает возможность произвести расчет размерных цепей и определить собираемость проектируемого объекта. Если объект представляет собой механизм, то на модели можно проверить его работоспособность и выполнить расчет кинематических характеристик.

Используя геометрическую модель, можно поставить численный эксперимент по определению напряженно-деформированного состояния, частот и форм собственных колебаний, устойчивости элементов конструкции, тепловых, оптических и других свойств объекта. Для этого нужно дополнить геометрическую модель физическими свойствами, смоделировать внешние условия ее работы и, используя физические законы, выполнить соответствующий расчет.

По геометрической модели можно вычислить траекторию режущего инструмента для механической обработки объекта. При выбранной технологии изготовления объекта геометрическая модель позволяет спроектировать оснастку и выполнить подготовку производства, а также проверить саму возможность изготовления объекта данным способом и качество этого изготовления. Кроме того, возможна графическая имитация процесса изготовления. Но для того, чтобы изготовить объект, кроме геометрической информации нужна информация о технологическом процессе, производственном оборудовании и многом другом, связанном с производством.

Многие из перечисленных проблем образуют самостоятельные разделы прикладной науки и по своей сложности не уступают, а в большинстве случаев и превосходят проблему создания геометрической модели. Геометрическая модель является отправной точкой для дальнейших действий. При построении геометрической модели мы не использовали физические законы, радиус-вектор каждой точки границы раздела внешней и внутренней частей моделируемого объекта является известным, поэтому при построении геометрической модели

нам приходится составлять и решать алгебраические уравнения. Задачи, в которых используются физические законы, приводят к дифференциальным и интегральным уравнениям, решение которых сложнее решения алгебраических уравнений.

В данной главе остановимся на выполнении расчетов, не связанных с физическими процессами. Мы рассмотрим вычисление чисто геометрических характеристик тел и их плоских сечений: площади поверхности, объема, центра масс, моментов инерции и ориентации главных осей инерции. Эти расчеты не требуют привлечения дополнительной информации. Кроме этого, мы рассмотрим проблемы численного интегрирования, которые приходится решать при определении геометрических характеристик.

Определение площади, центра масс и моментов инерции плоского сечения тела приводит к вычислению интегралов по площади сечения. Для плоских сечений мы располагаем информацией об их границах. Интегралы по площади плоского сечения мы сведем к криволинейным интегралам, которые в свою очередь сводятся к определенным интегралам. Определение площади поверхности, объема, центра масс, моментов инерции тела приводит к вычислению поверхностных и объемных интегралов. Мы будем опираться на представление тела с помощью границ (*B-rep*), т. е. на описание тела совокупностью ограничивающих его поверхностей и топологическую информацию о взаимном соседстве этих поверхностей. Мы сведем интегралы по объему тела к поверхностным интегралам по поверхностям граней тела, которые в свою очередь сводятся к двойным интегралам. В общем случае область интегрирования представляет собой связную двумерную область. Вычисление двойных интегралов численными методами можно выполнить для областей простых типов — четырехугольной или треугольной формы. В связи с этим в конце главы рассмотрены методы вычисления определенных интегралов и двойных интегралов по четырехугольным и треугольным областям. Методы разбивки областей определения параметров поверхностей на совокупности треугольных подобластей рассмотрены в следующей главе.

В начале главы рассмотрим сведение интегралов по площади к криволинейным интегралам и сведение объемных интегралов к поверхностным интегралам. На этом будут базироваться вычисления геометрических характеристик моделей.

8.2.* Криволинейные интегралы

Определение геометрических характеристик кривых линий и плоских сечений приводит к вычислению криволинейных интегралов, так как всю геометрическую информацию о кривой несет функция ее радиус-вектора от некоторого внутреннего параметра.

Криволинейными интегралами первого рода называются интегралы вида

$$\int_L f(\mathbf{r}) ds, \quad (8.2.1)$$

где $f(\mathbf{r})$ — некоторая заданная на кривой L функция ее точки $\mathbf{r}(t)$. Интегрирование в (8.2.1) выполняется по длине кривой s . Криволинейный интеграл (8.2.1) сводится к обыкновенному определенному интегралу

$$\int_L f(\mathbf{r}) ds = \int_{t_{\min}}^{t_{\max}} f(\mathbf{r}) \sqrt{\frac{d\mathbf{r}}{dt} \cdot \frac{d\mathbf{r}}{dt}} dt. \quad (8.2.2)$$

Область, на которой задана функция точки, является областью изменения параметра кривой.

Криволинейными интегралами второго рода называются интегралы вида

$$\int_L \mathbf{t} \cdot \mathbf{F}(\mathbf{r}) ds, \quad (8.2.3)$$

где $\mathbf{t} = d\mathbf{r}/ds = \mathbf{r}'/|\mathbf{r}'|$ — касательный вектор кривой, $\mathbf{F}(\mathbf{r})$ — векторная функция точки кривой $\mathbf{r}(t)$. Криволинейный интеграл (8.2.3) сводится к обыкновенному определенному интегралу

$$\int_L \mathbf{t} \cdot \mathbf{F}(\mathbf{r}) ds = \int_{t_{\min}}^{t_{\max}} \frac{d\mathbf{r}}{dt} \cdot \mathbf{F}(\mathbf{r}(t)) dt. \quad (8.2.4)$$

При изменении направления кривой на противоположное направление криволинейный интеграл второго рода меняет знак на противоположный.

Так как координаты радиус-вектора $\mathbf{r}(t) = [x(t) \ y(t) \ z(t)]^T$ кривой являются функциями параметра кривой, то мы можем преобразовать величину $\mathbf{t} ds$ к следующему виду:

$$\mathbf{t} ds = \frac{d\mathbf{r}}{dt} dt = \mathbf{i} \frac{dx}{dt} dt + \mathbf{j} \frac{dy}{dt} dt + \mathbf{k} \frac{dz}{dt} dt = \mathbf{i} dx + \mathbf{j} dy + \mathbf{k} dz. \quad (8.2.5)$$

Пусть в пространстве задана векторная функция

$$\mathbf{F}(\mathbf{r}) = \mathbf{i}F_x(x, y, z) + \mathbf{j}F_y(x, y, z) + \mathbf{k}F_z(x, y, z).$$

Тогда, используя преобразование (8.2.5), получим

$$\int_L \mathbf{t} \cdot \mathbf{F}(\mathbf{r}) ds = \int_L F_x dx + F_y dy + F_z dz, \quad (8.2.6)$$

где координаты x, y, z принадлежат кривой.

Формула Стокса. Пусть кривая $\mathbf{r}(t)$ замкнутая и построена на некоторой гладкой поверхности S , и пусть в пространстве задано векторное поле $\mathbf{F}(\mathbf{r})$. Пусть при движении вдоль кривой в положительном направлении ее обход происходит против часовой стрелки, если смотреть навстречу нормали поверхности \mathbf{m} . Рассмотрим поверхностный интеграл

$$\begin{aligned} & \iint_S \mathbf{m} \cdot (\nabla \times \mathbf{F}(\mathbf{r})) dS = \\ & = \iint_S \left(\left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) \mathbf{m} \cdot \mathbf{i} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) \mathbf{m} \cdot \mathbf{j} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) \mathbf{m} \cdot \mathbf{k} \right) dS, \end{aligned} \quad (8.2.7)$$

где ∇ — набла-оператор или оператор Гамильтона (1.11.13). Предположим, что нам известно явное уравнение поверхности S и кривой $\mathbf{r}(t)$ на ней. Возьмем, например, уравнение поверхности в виде $z = Z(x, y)$, а уравнение кривой на ней в виде двух уравнений, одним из которых является уравнение поверхности $z = Z(x, y)$, а вторым — уравнение цилиндрической поверхности $C(x, y) = 0$ с образующей параллельной орту \mathbf{k} . Пусть линию пересечения цилиндрической поверхности $C(x, y) = 0$ с

координатной плоскостью Oxy можно описать двумя двухмерными кривыми $y = Y_1(x)$ и $y = Y_2(x)$, изменяющимися от x_1 до x_2 , что показано на рис. 8.2.1. Кривые $y = Y_1(x)$ и $y = Y_2(x)$ в совокупности описывают проекцию заданной пространственной кривой

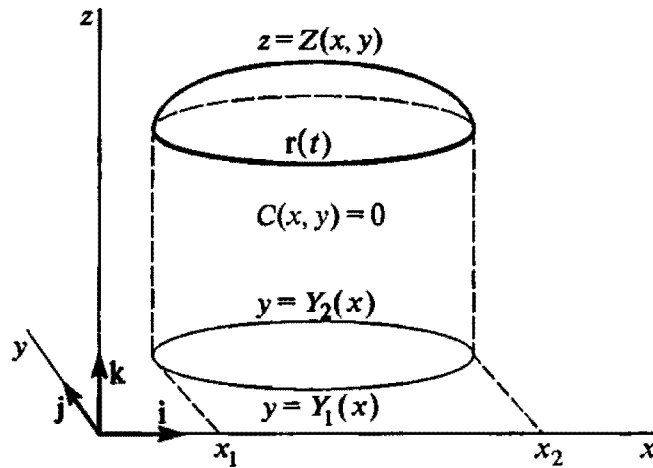


Рис. 8.2.1. Проекция замкнутой кривой $r(t)$ на поверхности $z = Z(x, y)$ на плоскость xy

на плоскость Oxy , только направление кривой $y = Y_1(x)$ совпадает с направлением пространственной кривой $r(t)$, а направление кривой $y = Y_2(x)$ противоположно направлению пространственной кривой.

Преобразуем часть интеграла (8.2.7), используя (8.2.5), следующим образом

$$\begin{aligned} \iint_S \left(\frac{\partial F_x}{\partial z} (\mathbf{m} \cdot \mathbf{j}) - \frac{\partial F_x}{\partial y} (\mathbf{m} \cdot \mathbf{k}) \right) dS &= \iint_S \frac{\partial F_x}{\partial z} dz dx - \iint_S \frac{\partial F_x}{\partial y} dx dy = \\ &= - \iint_S \left(\frac{\partial F_x}{\partial y} - \frac{\partial F_x}{\partial z} \frac{\mathbf{m} \cdot \mathbf{j}}{\mathbf{m} \cdot \mathbf{k}} \right) dx dy. \end{aligned} \quad (8.2.8)$$

Найдем значение $(\mathbf{m} \cdot \mathbf{j})/(\mathbf{m} \cdot \mathbf{k})$ для поверхности, которую можно описать уравнением $z = Z(x, y)$. Если за внутренние параметры поверхности принять координаты x и y , то для радиус-вектора и его частных производных по параметрам x и y получим значения

$$\mathbf{r} = [x \ y \ Z(x, y)]^T, \quad \mathbf{r}_1 = [1 \ 0 \ \partial Z/\partial x]^T, \quad \mathbf{r}_2 = [0 \ 1 \ \partial Z/\partial y]^T.$$

Используя эти соотношения, получим

$$\frac{\mathbf{m} \cdot \mathbf{j}}{\mathbf{m} \cdot \mathbf{k}} = \frac{(\mathbf{r}_1 \times \mathbf{r}_2) \cdot \mathbf{j}}{(\mathbf{r}_1 \times \mathbf{r}_2) \cdot \mathbf{k}} = -\frac{\partial Z}{\partial y}.$$

Подставим последнее равенство в (8.2.8) и продолжим далее его преобразование, приняв координаты x и y в качестве независимых параметров интегрирования:

$$- \iint_S \left(\frac{\partial F_x}{\partial y} - \frac{\partial F_x}{\partial z} \frac{\mathbf{m} \cdot \mathbf{j}}{\mathbf{m} \cdot \mathbf{k}} \right) dx dy = - \iint_{\Omega} \left(\frac{\partial F_x}{\partial y} + \frac{\partial F_x}{\partial z} \frac{\partial Z}{\partial y} \right) dx dy.$$

От поверхностного интеграла мы перешли к двойному интегралу по проекции Ω поверхности S на плоскость Oxy . Сведем этот интеграл к повторному интегралу по области на плоскости Oxy , ограниченной кривыми $y = Y_1(x)$ и $y = Y_2(x)$, изменяющимися от x_1

до x_2 (рис. 8.2.1). Величину $\frac{\partial F_x}{\partial y} + \frac{\partial F_x}{\partial z} \frac{\partial Z}{\partial y}$, стоящую под интегралом, можно рассматривать как полную производную $\frac{dF_x}{dy} = \frac{\partial F_x}{\partial y} + \frac{\partial F_x}{\partial z} \frac{\partial Z}{\partial y}$ функции $F_x = F_x(\text{const}, y, z)$ (параметр x при интегрировании по параметру y считается неизменным). Тогда

$$\begin{aligned} - \iint_{\Omega} \left(\frac{\partial F_x}{\partial y} + \frac{\partial F_x}{\partial z} \frac{\partial Z}{\partial y} \right) dx dy &= - \int_{x_1}^{x_2} \int_{Y_1(x)}^{Y_2(x)} \left(\frac{\partial F_x}{\partial y} + \frac{\partial F_x}{\partial z} \frac{\partial Z}{\partial y} \right) dy dx = \\ &= - \int_{x_1}^{x_2} \left(\int_{Y_1(x)}^{Y_2(x)} \frac{dF_x}{dy} dy \right) dx = - \int_{x_1}^{x_2} F_x(x, Y_2(x), z) dx + \int_{x_1}^{x_2} F_x(x, Y_1(x), z) dx. \quad (8.2.9) \end{aligned}$$

От повторного интеграла мы перешли к двум обычным определенным интегралам. Функция F_x принимает значения на пространственной кривой $\mathbf{r}(t)$. В первом интеграле правой части (8.2.9) переставим местами пределы интегрирования x_1 и x_2 , в результате чего интеграл изменит знак на противоположный. После такой замены интегрирование в правой части (8.2.9) будет выполняться по замкнутому контуру, являющемуся проекцией заданной пространственной кривой на плоскость Oxy , и мы можем перейти от двух обыкновенных интегралов к криволинейному интегралу. Таким образом, мы выполнили преобразование

$$\begin{aligned} \iint_S \left(\frac{\partial F_x}{\partial z} (\mathbf{m} \cdot \mathbf{j}) - \frac{\partial F_x}{\partial y} (\mathbf{m} \cdot \mathbf{k}) \right) dS &= \\ &= \int_{x_2}^{x_1} F_x(x, Y_2(x), z) dx + \int_{x_1}^{x_2} F_x(x, Y_1(x), z) dx = \oint_O F_x(x, y, z) dx = \oint_O \mathbf{F} \cdot \mathbf{i} dx. \end{aligned}$$

Обозначение области определения криволинейного интеграла означает, что кривая, по которой выполняется интегрирование, является замкнутой. С помощью аналогичных рассуждений и преобразований получим еще два равенства

$$\begin{aligned} \iint_S \left(\frac{\partial F_y}{\partial x} (\mathbf{m} \cdot \mathbf{k}) - \frac{\partial F_y}{\partial z} (\mathbf{m} \cdot \mathbf{i}) \right) dS &= \oint_O F_y(x, y, z) dy = \oint_O \mathbf{F} \cdot \mathbf{j} dy, \\ \iint_S \left(\frac{\partial F_z}{\partial y} (\mathbf{m} \cdot \mathbf{i}) - \frac{\partial F_z}{\partial x} (\mathbf{m} \cdot \mathbf{j}) \right) dS &= \oint_O F_z(x, y, z) dz = \oint_O \mathbf{F} \cdot \mathbf{k} dz. \end{aligned}$$

Сложив три последние равенства, получим *формулу Стокса*

$$\begin{aligned} \iint_S \left(\left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) \mathbf{m} \cdot \mathbf{i} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) \mathbf{m} \cdot \mathbf{j} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) \mathbf{m} \cdot \mathbf{k} \right) dS &= \\ &= \oint_O F_x(x, y, z) dx + F_y(x, y, z) dy + F_z(x, y, z) dz = \oint_O \mathbf{t} \cdot \mathbf{F}(\mathbf{r}) ds. \quad (8.2.10) \end{aligned}$$

Можно показать, что формула Стокса справедлива и в случае, когда поверхность является многосвязной, т. е. ограниченной одной внешней замкнутой кривой на ней и несколькими внутренними замкнутыми кривыми, лежащими внутри внешней кривой. В последнем случае внутренние кривые должны иметь ориентацию, противоположную ориентации внешней кривой. То есть внешняя кривая должна быть связана с нормалью поверхности правилом правого винта, а внутренние кривые должны быть связаны с нормалью поверхности правилом левого винта. При соблюдении этих правил поверхность всегда будет находиться *слева*, если двигаться вдоль ее границы с положительной стороны по отношению к нормали. При многоконтурной границе в правой части равенства (8.2.10) должна стоять сумма криволинейных интегралов по всем замкнутым кривым, ограничивающим поверхность.

Если кривая ограничивает на поверхности S связную область, является замкнутой и регулярной, а функция $\mathbf{F}(\mathbf{r})$ является однозначной, непрерывной и имеет непрерывные частные производные на заданной области поверхности, то поверхностный интеграл (8.2.7) связан с криволинейным интегралом (8.2.6) равенством

$$\oint \mathbf{t} \cdot \mathbf{F}(\mathbf{r}) ds = \pm \iint_S \mathbf{m} \cdot (\nabla \times \mathbf{F}(\mathbf{r})) dS. \quad (8.2.11)$$

Знак плюс соответствует случаю, когда ориентация замкнутой кривой и нормали к поверхности согласованы правилом правого винта, в противном случае в (8.2.11) должен стоять знак минус. Эта формула выражает *теорему Стокса*, утверждающую, что криволинейный интеграл от векторной функции $\mathbf{F}(\mathbf{r})$ по замкнутому контуру равен потоку ротора этой функции через поверхность, натянутую на контур. Для замкнутых кривых, построенных на поверхности S , справедливы следующие равенства:

$$\oint \mathbf{t} \times \mathbf{F}(\mathbf{r}) ds = \pm \iint_S (\mathbf{m} \times \nabla) \times \mathbf{F}(\mathbf{r}) dS, \quad (8.2.12)$$

$$\oint \varphi(\mathbf{r}) \mathbf{t} ds = \pm \iint_S \mathbf{m} \times \nabla \varphi(\mathbf{r}) dS, \quad (8.2.13)$$

где $\varphi(\mathbf{r})$ — скалярная функция точки пространства. Знак в (8.2.12) и (8.2.13), как и в (8.2.11), зависит от ориентации положительного обхода по кривой относительно нормали поверхности. Знак плюс берется тогда, когда обход по замкнутой кривой в положительном направлении при взгляде навстречу \mathbf{m} осуществляется против часовой стрелки.

Формула Грина. Для двумерных кривых $\mathbf{r}(t) = [x(t) \ y(t)]^T$ криволинейные интегралы первого и второго рода определяются аналогично криволинейным интегралам для пространственных кривых

$$\int_L f(\mathbf{r}) ds = \int_{t_{\min}}^{t_{\max}} f(t) \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt, \quad (8.2.14)$$

$$\int_L \mathbf{t} \cdot \mathbf{F}(\mathbf{r}) ds = \int_{t_{\min}}^{t_{\max}} \frac{d\mathbf{r}}{dt} \cdot \mathbf{F}(\mathbf{r}(t)) dt = \int_{t_{\min}}^{t_{\max}} F_x dx + F_y dy, \quad (8.2.15)$$

где $\mathbf{t} = \frac{d\mathbf{r}}{ds} = \frac{\mathbf{r}'}{|\mathbf{r}'|}$ — касательный вектор кривой, $f(\mathbf{r}) = [f_x(x, y) \ f_y(x, y)]^T$ —

двухмерная скалярная функция, $\mathbf{F}(\mathbf{r}) = [F_x(x, y) \ F_y(x, y)]^T$ — двухмерная векторная функция точек кривой. Для двумерных замкнутых кривых существует формула,

являющаяся аналогом формулы Стокса. Эта формула называется формулой Грина и связывает интеграл по некоторой двумерной (плоской) области с криволинейным интегралом второго рода (8.2.15) по границе этой области.

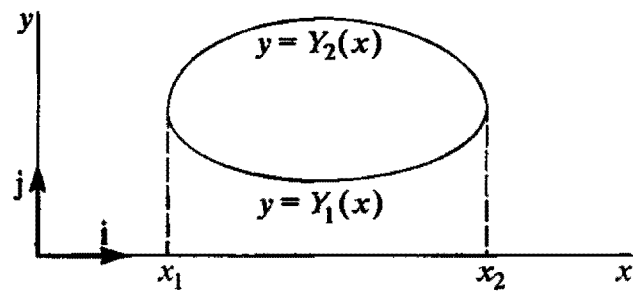


Рис. 8.2.2. Разбиение границы плоской области

Пусть дана некоторая плоская область Ω , ограниченная двумерной кривой L , и двумерная векторная функция $\mathbf{F}(\mathbf{r})$ на ней. Пусть $\mathbf{r}(t)$ есть радиус-вектор граничной кривой, которая является замкнутой и направлена так, что обход области Ω осуществляется против часовой стрелки. Пусть границу области Ω можно описать двумя кривыми $y = Y_1(x)$ и $y = Y_2(x)$, изменяющимися от x_1 до x_2 , что показано на рис. 8.2.2. Кривые $y = Y_1(x)$ и $y = Y_2(x)$ в совокупности описывают границу L заданной

плоской области, только направление кривой $y = Y_1(x)$ совпадает с направлением кривой $\mathbf{r}(t)$, а направление кривой $y = Y_2(x)$ противоположно направлению кривой.

Рассмотрим интеграл по области

$$\iint_{\Omega} \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) dS.$$

Представим часть этого интеграла в виде двукратного интеграла

$$\begin{aligned} - \iint_{\Omega} \frac{\partial F_x}{\partial y} dS &= - \int_{x_1}^{x_2} \left(\int_{Y_1(x)}^{Y_2(x)} \frac{\partial F_x}{\partial y} dy \right) dx = \\ &= - \int_{x_1}^{x_2} F_x(x, Y_2(x)) dx + \int_{x_1}^{x_2} F_x(x, Y_1(x)) dx = \\ &= \int_{x_2}^{x_1} F_x(x, Y_2(x)) dx + \int_{x_1}^{x_2} F_x(x, Y_1(x)) dx = \oint_L F_x(x, y) dx. \end{aligned}$$

Мы перешли от интеграла по области к криволинейному интегралу по границе этой области. Аналогично получим равенство

$$\iint_{\Omega} \frac{\partial F_y}{\partial x} dS = \oint_L F_y(x, y) dy.$$

Сложив два последних равенства, получим формулу Грина

$$\iint_{\Omega} \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) dS = \oint_L F_x(x, y) dx + F_y(x, y) dy \equiv \oint_L \mathbf{t} \cdot \mathbf{F}(\mathbf{r}) ds. \quad (8.2.16)$$

Можно показать, что формула Грина справедлива и в случае, когда область Ω является многосвязной. При этом в левой части равенства (8.2.16) должна стоять сумма криволинейных интегралов по всем замкнутым кривым, а обход по ограничивающим область кривым должен осуществляться таким образом, чтобы область всегда находилась слева. Формулу Грина можно получить как частный случай формулы Стокса, положив в (8.2.10) $\mathbf{F} = [F_x(x, y) \ F_y(x, y) \ 0]^T$, а за поверхность принять плоскость Oxy .

Рассмотрим преобразование векторной функции $\mathbf{F}(\mathbf{r})$ с помощью матрицы \mathbf{N} , поворачивающей вектор \mathbf{F} на прямой угол против часовой стрелки и равной

$$\mathbf{N} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \quad (8.2.17)$$

После такого преобразования получим функцию

$$\mathbf{N} \cdot \mathbf{F}(\mathbf{r}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} -F_y \\ F_x \end{bmatrix}. \quad (8.2.18)$$

После двойного преобразования получим ту же функцию, но с противоположным знаком

$$\mathbf{N} \cdot \mathbf{N} \cdot \mathbf{F}(\mathbf{r}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -F_y \\ F_x \end{bmatrix} = - \begin{bmatrix} F_x \\ F_y \end{bmatrix} = -\mathbf{F}(\mathbf{r}). \quad (8.2.19)$$

С помощью данного преобразования формула Грина примет вид

$$\begin{aligned} \oint_L \mathbf{t} \cdot \mathbf{F}(\mathbf{r}) ds &= - \oint_L \mathbf{t} \cdot (\mathbf{N} \cdot (\mathbf{N} \cdot \mathbf{F}(\mathbf{r}))) ds = \\ &= - \iint_{\Omega} \nabla \cdot (\mathbf{N} \cdot \mathbf{F}(\mathbf{r})) dS = \iint_{\Omega} \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) dS, \end{aligned} \quad (8.2.20)$$

где ∇ — двухмерный набла-оператор. Приведенные формулы позволяют свести поверхностные интегралы к криволинейным интегралам. Заменим в равенстве

$$- \oint_L \mathbf{t} \cdot (\mathbf{N} \cdot (\mathbf{N} \cdot \mathbf{F}(\mathbf{r}))) ds = - \iint_{\Omega} \nabla \cdot (\mathbf{N} \cdot \mathbf{F}(\mathbf{r})) dS$$

$\mathbf{N} \cdot \mathbf{F}(\mathbf{r})$ на $\mathbf{F}(\mathbf{r})$ и получим формулу Грина в следующем виде:

$$\iint_{\Omega} \nabla \cdot \mathbf{F}(\mathbf{r}) dS = \oint_L \mathbf{t} \cdot (\mathbf{N} \cdot \mathbf{F}(\mathbf{r})) ds = \oint_L \frac{d\mathbf{r}}{dt} \cdot (\mathbf{N} \cdot \mathbf{F}(\mathbf{r})) dt. \quad (8.2.21)$$

Эта формула будет использоваться при вычислении геометрических характеристик плоских сечений.

8.3. Геометрические характеристики плоских сечений

При решении некоторых задач, например, расчете напряжений в сечении балки, требуется оперировать геометрическими характеристиками сечений тел. Сечения тел описываются двухмерными контурами (2.12.7). Плоские контуры представляют собой двухмерные составные замкнутые кривые. Пусть дан некоторый контур в декартовой прямоугольной системе координат на плоскости, который описывается двухмерным радиус-вектором $\mathbf{r}(t) = [x(t) \ y(t)]^T$. За положительное направление контура примем такое направление его обхода, при котором сечение остается слева.

Периметр сечения. Периметр плоского сечения равен длине ограничивающего сечение контура и определяется интегралом

$$l = \int_L ds = \int_{t_{\min}}^{t_{\max}} \sqrt{\frac{dr}{dt} \cdot \frac{dr}{dt}} dt = \int_{t_{\min}}^{t_{\max}} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt.$$

В общем случае периметр можно вычислить как сумму длин сегментов контура.

Площадь и центр масс сечения. Площадь и статические моменты сечения относительно осей координат определяется формулами

$$F = \iint_F dF, \quad (8.3.1)$$

$$S_x = \iint_F y dF, \quad S_y = \iint_F x dF, \quad (8.3.2)$$

где x и y — текущие координаты бесконечно малой площади dF , а интегрирование ведется по площади сечения. При параллельном переносе осей координат на

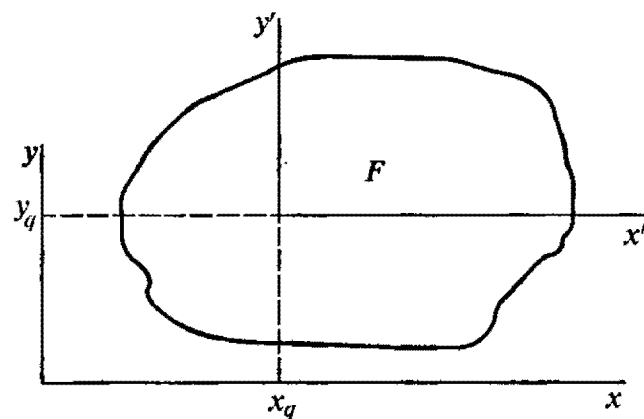


Рис. 8.3.1. Параллельный перенос осей координат

вектор $\mathbf{q} = [x_q \ y_q]^T$ (рис. 8.3.1), статические моменты в новой системе координат связаны со статическими моментами инерции в исходной системе следующими равенствами:

$$S_{x'} = S_x - y_q F, \quad S_{y'} = S_y - x_q F. \quad (8.3.3)$$

Точка, при переносе в которую начала координат статические моменты сечения становятся равными нулю, является центром масс сечения. Координаты центра масс сечения определяются формулами

$$x_c = \frac{S_y}{F}, \quad y_c = \frac{S_x}{F}. \quad (8.3.4)$$

Ось координат, относительно которой статический момент сечения равен нулю, называется *центральной*.

Моменты инерции сечения. Осевые и центробежный моменты инерции сечения определяются интегралами

$$J_x = \iint_F y^2 dF, \quad J_y = \iint_F x^2 dF, \quad J_{xy} = \iint_F xy dF. \quad (8.3.5)$$

Осевые моменты инерции всегда положительны, а центробежный момент инерции может быть как положительным, так и отрицательным, в зависимости от расположения осей относительно сечения. Моменты инерции сечения в системе координат Oxy и системе $O'x'y'$, сдвинутой относительно первой на вектор $\mathbf{q} = [x_q \ y_q]^\top$, связаны равенствами

$$J_{x'} = \iint_F (y - y_q)^2 dF = J_x - 2y_q S_x + y_q^2 F, \quad (8.3.6)$$

$$J_{y'} = \iint_F (x - x_q)^2 dF = J_y - 2x_q S_y + x_q^2 F, \quad (8.3.7)$$

$$J_{x'y'} = \iint_F (x - x_q)(y - y_q) dF = J_{xy} - x_q S_x - y_q S_y + x_q y_q F. \quad (8.3.8)$$

При переходе от системы координат Oxy к системе координат $O'x'y'$, повернутой относительно первой на угол φ (рис. 8.3.2), радиус-вектор $\mathbf{r} = [x \ y]^\top$ преобразуется по формулам

$$x' = x \cos \varphi + y \sin \varphi, \quad y' = -x \sin \varphi + y \cos \varphi. \quad (8.3.9)$$

Моменты инерции сечения в системе координат Oxy и системе $O'x'y'$, повернутой относительно первой на угол φ , связаны равенствами

$$J_{x'} = J_x \cos^2 \varphi - J_{xy} \sin 2\varphi + J_y \sin^2 \varphi, \quad (8.3.10)$$

$$J_{y'} = J_x \sin^2 \varphi + J_{xy} \sin 2\varphi + J_y \cos^2 \varphi, \quad (8.3.11)$$

$$J_{x'y'} = J_{xy} \cos 2\varphi + \frac{J_x - J_y}{2} \sin 2\varphi. \quad (8.3.12)$$

Заметим, что величина $J_\rho = J_x + J_y = J_{x'} + J_{y'}$ одинакова в обеих системах координат. Эта величина называется полярным моментом инерции сечения.

С изменением угла поворота φ осевые моменты инерции меняются, а их сумма остается неизменной. Следовательно, существует такой угол φ , при котором один из моментов инерции сечения достигает своего максимального значения, в то время как другой момент инерции принимает минимальное значение. Дифференцируя выражение для $J_{x'}$ по φ и приравнявая производную нулю, находим

$$\operatorname{tg} 2\varphi = \frac{2J_{xy}}{J_y - J_x}. \quad (8.3.13)$$

В соответствии с формулой (8.3.12) центробежный момент инерции при данном угле равен нулю. Система координат, в которой центробежный момент инерции равен нулю, называется *главной системой координат*. Если к тому же

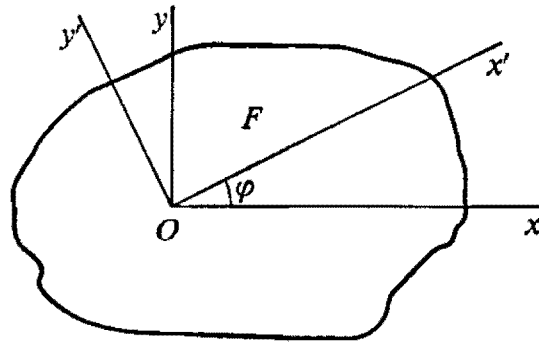


Рис. 8.3.2. Поворот осей координат

эта система является центральной, то она называется *главной центральной системой координат*. Если сечение имеет ось симметрии, то эта ось всегда будет главной. Осевые моменты инерции относительно главной системы координат называются *главными моментами инерции*. Для их определения перепишем (8.3.10) и (8.3.11) в виде

$$J_{x'} = \frac{J_x + J_y}{2} - \frac{J_y - J_x}{2} \cos 2\varphi - J_{xy} \sin 2\varphi, \quad (8.3.14)$$

$$J_{y'} = \frac{J_x + J_y}{2} + \frac{J_y - J_x}{2} \cos 2\varphi + J_{xy} \sin 2\varphi. \quad (8.3.15)$$

Учитывая, что

$$\cos 2\varphi = \frac{1}{\sqrt{1 + \operatorname{tg}^2 2\varphi}}, \quad \sin 2\varphi = \frac{\operatorname{tg} 2\varphi}{\sqrt{1 + \operatorname{tg}^2 2\varphi}}, \quad (8.3.16)$$

исключим при помощи (8.3.13) угол φ и получим

$$J_{\max} = \frac{J_x + J_y}{2} + \sqrt{\left(\frac{J_y - J_x}{2}\right)^2 + J_{xy}^2}, \quad (8.3.17)$$

$$J_{\min} = \frac{J_x + J_y}{2} - \sqrt{\left(\frac{J_y - J_x}{2}\right)^2 + J_{xy}^2}. \quad (8.3.18)$$

Вычисление моментов инерции сечения. Так же как и все кривые, контур описан в параметрической форме и не имеет своих явных уравнений, связывающих координаты его радиус-вектора. Без явных уравнений мы не можем воспользоваться непосредственно формулами (8.3.1), (8.3.2), (8.3.5) для определения геометрических характеристик сечения. Вся геометрическую информацию о контуре несет функция его радиус-вектора от некоторого внутреннего параметра. Для вычисления геометрических характеристик воспользуемся формулой Грина (8.2.20), позволяющей свести поверхностный интеграл к криволинейному интегралу.

Положим в (8.2.20) последовательно: $\mathbf{F}(\mathbf{r}) = \mathbf{r}$, $\mathbf{F}(\mathbf{r}) = x\mathbf{r}$, $\mathbf{F}(\mathbf{r}) = y\mathbf{r}$, $\mathbf{F}(\mathbf{r}) = x^2\mathbf{r}$, $\mathbf{F}(\mathbf{r}) = y^2\mathbf{r}$, $\mathbf{F}(\mathbf{r}) = xy\mathbf{r}$, где $\mathbf{r} = [x \ y]^\top$ — радиус-вектор точки сечения. Тогда $\nabla \cdot (\mathbf{r}) = 2$, $\nabla \cdot (x\mathbf{r}) = 3x$, $\nabla \cdot (y\mathbf{r}) = 3y$, $\nabla \cdot (x^2\mathbf{r}) = 4x^2$, $\nabla \cdot (y^2\mathbf{r}) = 4y^2$, $\nabla \cdot (xy\mathbf{r}) = 4xy$. Для каждого случая вычислим правую часть равенства (8.2.20):

$$\begin{aligned} \iint_F \nabla \cdot \mathbf{r} dF &= \iint_F 2 dF = 2F, \\ \iint_F \nabla \cdot (x\mathbf{r}) dF &= \iint_F 3x dF = 3S_y, \\ \iint_F \nabla \cdot (y\mathbf{r}) dF &= \iint_F 3y dF = 3S_x, \\ \iint_F \nabla \cdot (x^2\mathbf{r}) dF &= \iint_F 4x^2 dF = 4J_y, \\ \iint_F \nabla \cdot (y^2\mathbf{r}) dF &= \iint_F 4y^2 dF = 4J_x, \\ \iint_F \nabla \cdot (xy\mathbf{r}) dF &= \iint_F 4xy dF = 4J_{xy}. \end{aligned}$$

Подставим вычисленные значения в формулу Грина (8.2.20) и получим формулы для определения площади, статических моментов и моментов инерции плоского сечения через криволинейные интегралы по ограничивающим его контурам:

$$F = \frac{1}{2} \oint \frac{d\mathbf{r}}{dt} \cdot (\mathbf{N} \cdot \mathbf{r}) dt = \frac{1}{2} \oint \left(-y \frac{dx}{dt} + x \frac{dy}{dt} \right) dt, \quad (8.3.19)$$

$$S_x = \frac{1}{3} \oint \frac{d\mathbf{r}}{dt} \cdot (\mathbf{N} \cdot (y\mathbf{r})) dt = \frac{1}{3} \oint \left(-y^2 \frac{dx}{dt} + xy \frac{dy}{dt} \right) dt, \quad (8.3.20)$$

$$S_y = \frac{1}{3} \oint \frac{d\mathbf{r}}{dt} \cdot (\mathbf{N} \cdot (x\mathbf{r})) dt = \frac{1}{3} \oint \left(-xy \frac{dx}{dt} + x^2 \frac{dy}{dt} \right) dt, \quad (8.3.21)$$

$$J_x = \frac{1}{4} \oint \frac{d\mathbf{r}}{dt} \cdot (\mathbf{N} \cdot (y^2\mathbf{r})) dt = \frac{1}{4} \oint \left(-y^3 \frac{dx}{dt} + xy^2 \frac{dy}{dt} \right) dt, \quad (8.3.22)$$

$$J_y = \frac{1}{4} \oint \frac{d\mathbf{r}}{dt} \cdot (\mathbf{N} \cdot (x^2\mathbf{r})) dt = \frac{1}{4} \oint \left(-yx^2 \frac{dx}{dt} + x^3 \frac{dy}{dt} \right) dt, \quad (8.3.23)$$

$$J_{xy} = \frac{1}{4} \oint \frac{d\mathbf{r}}{dt} \cdot (\mathbf{N} \cdot (xy\mathbf{r})) dt = \frac{1}{4} \oint \left(-xy^2 \frac{dx}{dt} + x^2y \frac{dy}{dt} \right) dt. \quad (8.3.24)$$

Если плоское сечение ограничено одним внешним контуром, то криволинейный интеграл должен быть взят при обходе контура против часовой стрелки. Если сечение кроме внешнего контура ограничено еще и внутренними контурами, то к интегралу по внешнему контуру должны быть добавлены интегралы по внутренним контурам, вычисленные при их обходе по часовой стрелке. Таким образом, в формулах (8.3.19)–(8.3.24) стоят суммы интегралов по всем контурам, ограничивающим плоское сечение и имеющим соответствующую ориентацию. Будем считать, что интегрирование выполняется по всем границам плоского сечения, и опустим знак суммы. Полученные определенные интегралы могут быть вычислены с помощью квадратурных формул, которые мы рассмотрим ниже.

8.4. Длина и центр масс кривой линии

Длина кривой. Длина кривой определяется интегралом

$$L = \int_L ds = \int_{t_{\min}}^{t_{\max}} \sqrt{\mathbf{r}' \cdot \mathbf{r}'} dt = \int_{t_{\min}}^{t_{\max}} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} dt. \quad (8.4.1)$$

Для двумерной кривой координату z радиус-вектора кривой следует опустить (положить равной нулю). Если для кривой определить функцию распределенной по ней плотности $\rho(t)$, то интеграл

$$M = \int_{t_{\min}}^{t_{\max}} \rho(t) \sqrt{\mathbf{r}' \cdot \mathbf{r}'} dt = \int_{t_{\min}}^{t_{\max}} \rho(t) \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} dt \quad (8.4.2)$$

определит массу кривой.

Центр масс кривой. Пусть на кривой определена функция распределенной по ней плотности $\rho = \rho(t)$. Тогда интегралы

$$S_x = \int_L \rho x ds = \int_{t_{\min}}^{t_{\max}} \rho x \sqrt{\mathbf{r}' \cdot \mathbf{r}'} dt = \int_{t_{\min}}^{t_{\max}} \rho x \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} dt, \quad (8.4.3)$$

$$S_y = \int_L \rho y ds = \int_{t_{\min}}^{t_{\max}} \rho y \sqrt{\mathbf{r}' \cdot \mathbf{r}'} dt = \int_{t_{\min}}^{t_{\max}} \rho y \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} dt, \quad (8.4.4)$$

$$S_z = \int_L \rho z ds = \int_{t_{\min}}^{t_{\max}} \rho z \sqrt{\mathbf{r}' \cdot \mathbf{r}'} dt = \int_{t_{\min}}^{t_{\max}} \rho z \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} dt \quad (8.4.5)$$

определяют статические моменты кривой относительно осей координат. Статические моменты являются компонентами некоторого вектора, так как при преобразовании координат они изменяются так же, как компоненты вектора. При переносе начала координат в точку $\mathbf{q}_c = [x_c \ y_c \ z_c]^T$ статические моменты инерции в новой системе координат будут равны

$$S_{x'} = S_x - x_c M, \quad S_{y'} = S_y - y_c M, \quad S_{z'} = S_z - z_c M.$$

Точка, при переносе в которую начала координат статические моменты становятся равными нулю, является центром масс кривой. Таким образом, координаты центра масс кривой определяются формулами

$$x_c = \frac{S_x}{M}, \quad y_c = \frac{S_y}{M}, \quad z_c = \frac{S_z}{M}. \quad (8.4.6)$$

Для кривой, так же как и для тела, можно определить моменты инерции. Формулы для их вычисления аналогичны формулам для вычисления моментов инерции тел, с той лишь разницей, что интегрирование ведется не по объему, а по кривой линии. Вычисление моментов инерции мы рассмотрим ниже.

8.5.* Поверхностные интегралы

Определение геометрических характеристик тел приводит к вычислению поверхностных и объемных интегралов. Так как всю геометрическую информацию о теле несут его грани, которые описываются поверхностями в параметрическом виде, то объемные интегралы мы сведем к поверхностным интегралам. Пусть дано некоторое тело в декартовой прямоугольной системе координат $Oxyz$. Орты этой системы координат обозначим через \mathbf{i} , \mathbf{j} , \mathbf{k} . Пусть в этой системе координат задана некоторая скалярная $f(\mathbf{r})$ или векторная $\mathbf{F}(\mathbf{r})$ функция точки $\mathbf{r} = [x \ y \ z]^T$. Для точек, принадлежащих граням тела, эти функции являются функциями параметров поверхностей граней u и v , так как $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$.

Поверхностными интегралами первого рода называются интегралы вида

$$\iint_S f(\mathbf{r}) dS, \quad (8.5.1)$$

где $f(\mathbf{r}) \equiv f(x, y, z)$ — заданная скалярная функция точки поверхности тела \mathbf{r} , а интегрирование выполняется по всей поверхности S тела. Воспользуемся равенством (1.7.15) для бесконечно малого участка площади поверхности и перейдем от поверхностного интеграла (8.5.1) к двойному интегралу

$$\iint_S f(\mathbf{r}) dS = \iint_{\Omega} f(u, v) \sqrt{g_{11}g_{22} - g_{12}^2} du dv, \quad (8.5.2)$$

где g_{11} , g_{22} , g_{12} — коэффициенты первой квадратичной формы поверхности (1.7.8), Ω — область, на которой задана функция точки, является областью изменения параметров u и v поверхности, $f(u, v) = f(\mathbf{r}(u, v))$ — функция параметров поверхности.

Поверхностными интегралами второго рода называются интегралы вида

$$\iint_S \mathbf{m} \cdot \mathbf{F}(\mathbf{r}) dS, \quad \iint_S \mathbf{m} \times \mathbf{F}(\mathbf{r}) dS, \quad \iint_S \varphi(\mathbf{r}) \mathbf{m} dS, \quad (8.5.3)$$

где \mathbf{m} — нормаль поверхности тела, $\mathbf{F}(\mathbf{r}) \equiv \mathbf{F}(x, y, z)$ — заданная векторная функция точки поверхности тела \mathbf{r} , $\varphi(\mathbf{r})$ — заданная скалярная функция точки поверхности тела \mathbf{r} . Интегрирование выполняется по всей поверхности S тела. Подставим в интегралы (8.5.3) выражение для нормали поверхности (1.7.18) и выражение (1.7.15) для бесконечно малого участка площади и перейдем к двойным интегралам по области параметров граней тела

$$\begin{aligned} & \iint_{\Omega} (\mathbf{r}_1 \times \mathbf{r}_2) \cdot \mathbf{F}(u, v) du dv, \\ & \iint_{\Omega} (\mathbf{r}_1 \times \mathbf{r}_2) \times \mathbf{F}(u, v) du dv, \\ & \iint_{\Omega} \varphi(u, v)(\mathbf{r}_1 \times \mathbf{r}_2) du dv, \end{aligned} \quad (8.5.4)$$

где $\mathbf{r}_1 = \frac{\partial \mathbf{r}}{\partial u}$, $\mathbf{r}_2 = \frac{\partial \mathbf{r}}{\partial v}$ — частные производные радиус-вектора поверхности по параметрам поверхности, $\mathbf{F}(u, v) = \mathbf{F}(\mathbf{r}(u, v))$ — векторная функция параметров поверхности.

Запишем интегралы (8.5.4) в несколько ином виде. Рассмотрим величину $\mathbf{i} dy dz + \mathbf{j} dz dx + \mathbf{k} dx dy$ для точек поверхности. Координаты радиус-вектора $\mathbf{r} = [x \ y \ z]^T$ поверхности являются функциями параметров поверхности $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$. Используя равенства

$$dy dz = \begin{vmatrix} \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{vmatrix} du dv, \quad dz dx = \begin{vmatrix} \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \\ \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \end{vmatrix} du dv, \quad dx dy = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} du dv,$$

перейдем в величине $\mathbf{i} dy dz + \mathbf{j} dz dx + \mathbf{k} dx dy$ к параметрам поверхности

$$\begin{aligned} & \mathbf{i} dy dz + \mathbf{j} dz dx + \mathbf{k} dx dy = \\ & = \mathbf{i} \left(\frac{\partial y}{\partial u} \frac{\partial z}{\partial v} - \frac{\partial y}{\partial v} \frac{\partial z}{\partial u} \right) du dv + \mathbf{j} \left(\frac{\partial z}{\partial u} \frac{\partial x}{\partial v} - \frac{\partial z}{\partial v} \frac{\partial x}{\partial u} \right) du dv + \mathbf{k} \left(\frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} \right) du dv = \\ & = \left(\frac{\partial x}{\partial u} \mathbf{i} + \frac{\partial y}{\partial u} \mathbf{j} + \frac{\partial z}{\partial u} \mathbf{k} \right) du \times \left(\frac{\partial x}{\partial v} \mathbf{i} + \frac{\partial y}{\partial v} \mathbf{j} + \frac{\partial z}{\partial v} \mathbf{k} \right) dv = \\ & = (\mathbf{r}_1 \times \mathbf{r}_2) du dv = \mathbf{m} \sqrt{g_{11}g_{22} - g_{12}g_{21}} du dv = \mathbf{m} dS. \end{aligned} \quad (8.5.5)$$

Мы получили величину, стоящую в поверхностных интегралах второго рода (8.5.4) или (8.5.3). При переходе от переменных интегрирования x, y, z к переменным интегрирования u, v предполагалось, что отображения $(y, z) \rightarrow (u, v)$, $(z, x) \rightarrow (u, v)$, $(x, y) \rightarrow (u, v)$ взаимно однозначны, непрерывны и имеют отличный от нуля якобиан. Пусть в пространстве задана векторная функция

$$\mathbf{F}(\mathbf{r}) = \mathbf{i}F_x(x, y, z) + \mathbf{j}F_y(x, y, z) + \mathbf{k}F_z(x, y, z). \quad (8.5.6)$$

Тогда, используя преобразование (8.5.5), получим

$$\iint_S \mathbf{m} \cdot \mathbf{F}(\mathbf{r}) dS = \iint_S F_x dy dz + F_y dz dx + F_z dx dy, \quad (8.5.7)$$

$$\begin{aligned} \iint_S \mathbf{m} \times \mathbf{F}(\mathbf{r}) dS &= \\ &= \iint_S \mathbf{i}(F_z dx dz - F_y dx dy) + \mathbf{j}(F_x dy dx - F_z dy dz) + \mathbf{k}(F_y dz dy - F_x dz dx), \end{aligned} \quad (8.5.8)$$

$$\iint_S \varphi(\mathbf{r}) \mathbf{m} dS = \iint_S \varphi \mathbf{i} dy dz + \varphi \mathbf{j} dz dx + \varphi \mathbf{k} dx dy, \quad (8.5.9)$$

где координаты x, y, z принадлежат поверхности. Найдем связь поверхностных интегралов второго рода с объемными интегралами.

Формула Остроградского–Гаусса. Оболочка тела ограничивает односвязную область пространства. Рассмотрим объемный интеграл от дивергенции векторной функции (8.5.6)

$$\iiint_V \nabla \cdot \mathbf{F}(\mathbf{r}) dV \equiv \iiint_V \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \right) dV, \quad (8.5.10)$$

где ∇ — набла-оператор или оператор Гамильтона (1.11.13), который в принятых обозначениях данной системы координат имеет вид

$$\nabla = \frac{\partial}{\partial x} \mathbf{i} + \frac{\partial}{\partial y} \mathbf{j} + \frac{\partial}{\partial z} \mathbf{k}. \quad (8.5.11)$$

Предположим, что нам известно явное уравнение поверхности оболочки в любом из видов: $x = X_i(y, z)$, $y = Y_i(z, x)$, $z = Z_i(x, y)$. Индекс i явных уравнений говорит о том, что замкнутая оболочка в общем случае описывается несколькими поверхностями.

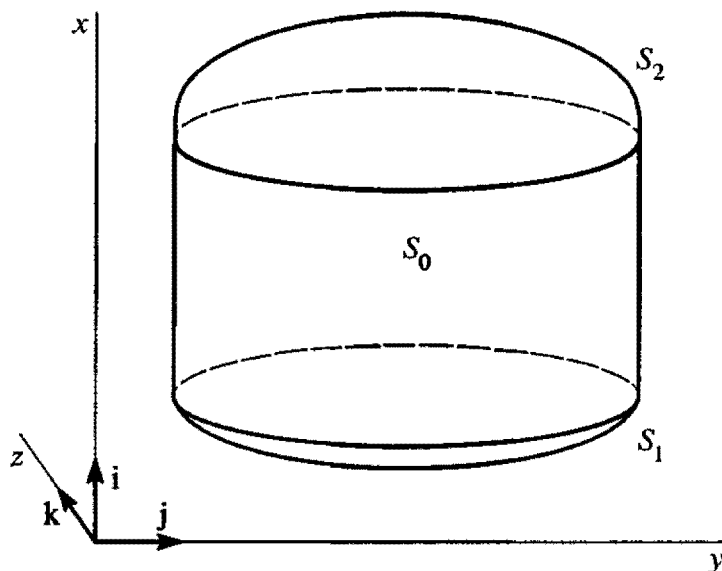


Рис. 8.5.1. Разбиение замкнутой оболочки на три характерные части

Например, каждая прямая, проходящая через тело, пересекает его четное число раз, сколько раз входя в него, столько раз и выходя из него. Кроме того, могут существовать участки поверхности тела, касающиеся указанных прямых. Пусть \mathbf{m} есть нормаль

тела. Предположим, что в одном из вариантов поверхность тела можно описать тремя уравнениями $x = X_0(y, z)$, $x = X_1(y, z)$, $x = X_2(y, z)$. Пусть уравнение $x = X_1(y, z)$ описывает часть S_1 поверхности тела, нормаль которой образует тупой угол с ортом \mathbf{i} ($\mathbf{m} \cdot \mathbf{i} < 0$), уравнение $x = X_2(y, z)$ описывает часть S_2 поверхности тела, нормаль которой образует острый угол с ортом \mathbf{i} ($\mathbf{m} \cdot \mathbf{i} > 0$), уравнение $x = X_0(y, z)$ описывает цилиндрическую часть S_0 поверхности тела, нормаль которой ортогональна орту \mathbf{i} ($\mathbf{m} \cdot \mathbf{i} = 0$). Поверхности $x = X_0(y, z)$, $x = X_1(y, z)$, $x = X_2(y, z)$ приведены на рис. 8.5.1.

Тогда для части объемного интеграла (8.5.11) можно выполнить следующие преобразования

$$\begin{aligned} \iiint_V \frac{\partial F_x}{\partial x} dV &= \iint_S \left(\int_{X_1(y, z)}^{X_2(y, z)} \frac{\partial F_x}{\partial x} dx \right) dy dz = \\ &= \iint_{S_2} F_x(X_2(y, z), y, z) dy dz - \iint_{S_1} F_x(X_1(y, z), y, z) dy dz. \end{aligned}$$

Нормаль поверхности $x = X_2(y, z)$ совпадает с нормалью оболочки тела, а нормаль поверхности $x = X_1(y, z)$ направлена против нормали оболочки тела, поэтому при переходе от интегрирования по перечисленным поверхностям к интегрированию по поверхности тела, следует изменить знак перед вторым интегралом в правой части последнего равенства. Добавим к интегралам по поверхностям S_1 и S_2 интеграл по части поверхности S_0 $x = X_0(y, z)$, который равен нулю. Таким образом, имеет место равенство

$$\iiint_V \frac{\partial F_x}{\partial x} dV = \iint_S F_x(x, y, z) \mathbf{i} \cdot \mathbf{m} dS.$$

Интегрирование в правой части последнего равенства выполняется по всей поверхности оболочки. Нетрудно доказать, что данное равенство справедливо при любой форме тела и любом представлении его поверхности. Аналогично получим еще два равенства

$$\iiint_V \frac{\partial F_y}{\partial y} dV = \iint_S F_y(x, y, z) \mathbf{j} \cdot \mathbf{m} dS,$$

$$\iiint_V \frac{\partial F_z}{\partial z} dV = \iint_S F_z(x, y, z) \mathbf{k} \cdot \mathbf{m} dS.$$

Сложив три последние равенства, получим формулу Остроградского–Гаусса

$$\iiint_V \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \right) dV = \iint_S (F_x \mathbf{i} + F_y \mathbf{j} + F_z \mathbf{k}) \cdot \mathbf{m} dS. \quad (8.5.12)$$

Поверхностные интегралы (8.5.8) и (8.5.9) также можно связать с интегралами по объему. Формулы, связывающие поверхностные интегралы (8.5.8) и (8.5.9) с интегралами по объему, приведем без вывода.

Если объем тела V является ограниченным и пространственно односвязным, поверхность тела S — замкнутой и регулярной, а функции $\mathbf{F}(\mathbf{r})$ и $\varphi(\mathbf{r})$ — однозначными, непрерывными и имеющими непрерывные частные производные внутри оболочки тела и на ее поверхности, то поверхностные интегралы (8.5.4) связаны с интегралами по объему тела соответственно теоремой Остроградского–Гаусса (теоремой

о дивергенции), теоремой о роторе и теоремой о градиенте:

$$\iint_S \mathbf{m} \cdot \mathbf{F}(\mathbf{r}) dS = \iiint_V \nabla \cdot \mathbf{F}(\mathbf{r}) dV, \quad (8.5.13)$$

$$\iint_S \mathbf{m} \times \mathbf{F}(\mathbf{r}) dS = \iiint_V \nabla \times \mathbf{F}(\mathbf{r}) dV, \quad (8.5.14)$$

$$\iint_S \varphi(\mathbf{r}) \mathbf{m} dS = \iiint_V \nabla \varphi(\mathbf{r}) dV, \quad (8.5.15)$$

Если тело описывается несколькими гранями, то в формулах подразумевается суммирование по граням тела. Заметим, что \mathbf{m} есть нормаль грани тела, а не поверхности этой грани. Нормаль грани всегда направлена наружу тела, а нормаль поверхности, на которой базируется грань, может совпадать с нормалью грани или может иметь противоположное направление. При выводе формулы Остроградского–Гаусса мы опирались на то, что система координат является декартовой. Равенства (8.5.14) и (8.5.15) справедливы в евклидовых пространствах.

Теорема Остроградского–Гаусса гласит, что интеграл от дивергенции векторного поля по некоторому объему равен потоку поля через поверхность, ограничивающую этот объем. Векторная функция $\mathbf{F}(\mathbf{r})$ в (8.5.13) сама может быть получена с участием оператора, например, $\mathbf{F}(\mathbf{r}) = \nabla \varphi(\mathbf{r})$, $\mathbf{F}(\mathbf{r}) = \psi \nabla \varphi(\mathbf{r})$ или $\mathbf{F}(\mathbf{r}) = \psi \nabla \varphi(\mathbf{r}) - \varphi \nabla \psi(\mathbf{r})$. Для приведенных функций теорема Остроградского–Гаусса имеет вид

$$\iiint_V \nabla^2 \varphi(\mathbf{r}) dV = \iint_S \mathbf{m} \cdot \nabla \varphi(\mathbf{r}) dS, \quad (8.5.16)$$

$$\iiint_V (\nabla \psi(\mathbf{r}) \cdot \nabla \varphi(\mathbf{r}) + \psi(\mathbf{r}) \nabla^2 \varphi(\mathbf{r})) dV = \iint_S \psi(\mathbf{r}) \mathbf{m} \cdot \nabla \varphi(\mathbf{r}) dS, \quad (8.5.17)$$

$$\iiint_V (\nabla \psi(\mathbf{r}) \nabla^2 \varphi(\mathbf{r}) - \varphi(\mathbf{r}) \nabla^2 \psi(\mathbf{r})) dV = \iint_S \mathbf{m} \cdot (\psi(\mathbf{r}) \nabla \varphi(\mathbf{r}) - \varphi(\mathbf{r}) \nabla \psi(\mathbf{r})) dS. \quad (8.5.18)$$

Теорема Остроградского–Гаусса позволит нам при вычислении геометрических характеристик перейти от интегралов по объему тела к интегралам по его поверхности.

8.6. Площадь поверхности, объем и центр масс тела

Площадь поверхности тела. Площадь поверхности тела складывается из площадей поверхности его граней. Площадь S каждой поверхности определяется интегралом

$$S = \iint_S dS = \iint_{\Omega} \sqrt{g_{11}g_{22} - g_{12}^2} du dv. \quad (8.6.1)$$

Площадь поверхности тела равна сумме площадей граней

$$S = \sum_i \iint_{S_i} dS = \sum_i \iint_{\Omega_i} \sqrt{g_{11}g_{22} - g_{12}^2} du_i dv_i, \quad (8.6.2)$$

где u_i, v_i — параметры поверхности i -й грани тела, Ω_i — область определения параметров поверхности i -й грани тела. В дальнейших формулах геометриче-

ских характеристик тела мы опустим знак суммирования и индексы граней, считая, что суммирование, естественно, выполняется.

Объем тела. Положим в (8.5.13) $\mathbf{F}(\mathbf{r}) = \mathbf{r}$. Тогда $\nabla \cdot (\mathbf{r}) = 3$ и из теоремы Остроградского–Гаусса получим

$$\iint_S \mathbf{m} \cdot \mathbf{r} dS = 3 \iiint_V dV = 3V.$$

Таким образом, объем тела определится формулой

$$V = \frac{1}{3} \iint_S \mathbf{m} \cdot \mathbf{r} dS = \frac{1}{3} \iint_{\Omega} \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11}g_{22} - g_{12}^2} du dv. \quad (8.6.3)$$

Пусть плотность тела ρ постоянна по всему его объему. Тогда масса тела определится формулой

$$M = \rho V = \frac{\rho}{3} \iint_S \mathbf{m} \cdot \mathbf{r} dS = \frac{\rho}{3} \iint_{\Omega} \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11}g_{22} - g_{12}^2} du dv. \quad (8.6.4)$$

Если плотность тела не является постоянной, то нам не удастся перейти от интеграла по объему к интегралу по поверхности тела при вычислении его массы. В формулах (8.6.3) и (8.6.4) подразумевается суммирование по граням тела. Заметим, что \mathbf{m} есть нормаль грани тела, а не поверхности этой грани. Нормаль грани всегда направлена наружу тела, а нормаль поверхности, на которой базируется грань, может совпадать с нормалью грани или может иметь противоположное нормали грани направление. В общем случае тело может иметь пустоты внутри. Тогда оно описывается несколькими замкнутыми оболочками, одна из которых является внешней, а остальные внутренними и полностью лежат внутри внешней оболочки. Нормаль к любой оболочке направлена вне объема тела. Интегрирование в (8.6.3) и (8.6.4) выполняется по всем оболочкам, как внешней оболочке, так и внутренним оболочкам.

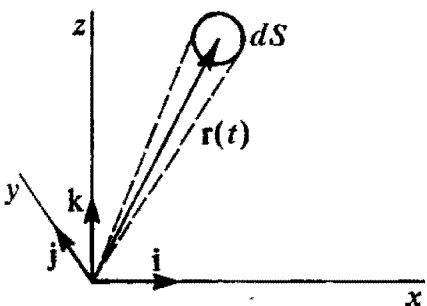


Рис. 8.6.1. Пирамида с бесконечно малым основанием dS

Формулу (8.6.3) можно получить и другим способом. Рассмотрим бесконечно малый участок dS поверхности некоторой грани тела. Построим элементарную пирамиду на базе бесконечно малого участка dS с вершиной в начале координат и прямолинейными образующими. Такая элементарная пирамида приведена на рис. 8.6.1. Основанием пирамиды является рассматриваемый бесконечно малый участок dS поверхности.

Пусть поверхность рассматриваемой грани описывается векторной функцией $\mathbf{r} = \mathbf{r}(u, v)$. Рассмотрим выражение

$$\frac{1}{3} \mathbf{m} \cdot \mathbf{r} dS, \quad (8.6.5)$$

где \mathbf{m} — единичный вектор нормали рассматриваемой поверхности.

Абсолютная величина выражения $\mathbf{m} \cdot \mathbf{r} dS$ равна объему элементарной пирамиды, построенной на участке поверхности dS , так как модуль $\mathbf{m} \cdot \mathbf{r}$ представляет собой высоту, опущенную из вершины пирамиды на ее основание. Знак

$\mathbf{m} \cdot \mathbf{r} dS$ зависит от ориентации нормали поверхности и ее радиус-вектора. Если вычислить интеграл от выражения (8.6.5) по всем граням тела, то в силу того, что оболочка тела всегда является замкнутой и ориентируемой, получим объем этого тела.

Статические моменты тела. Центр масс тела определяется с помощью статических моментов. Статические моменты тела относительно координатных плоскостей соответственно равны объемным интегралам

$$M_x = \iiint_V \rho x dV, \quad (8.6.6)$$

$$M_y = \iiint_V \rho y dV, \quad (8.6.7)$$

$$M_z = \iiint_V \rho z dV, \quad (8.6.8)$$

где ρ — плотность тела. Интегрирование выполняется по объему тела. Пусть плотность тела постоянна по всему его объему ($\rho = \text{const}$). Это позволит нам при вычислении геометрических характеристик тела перейти от интегрирования по объему к интегрированию по поверхности тела. Для вычисления статических моментов тела положим в (8.5.13) последовательно $\mathbf{F}(\mathbf{r}) = x\mathbf{r}$, $\mathbf{F}(\mathbf{r}) = y\mathbf{r}$, $\mathbf{F}(\mathbf{r}) = z\mathbf{r}$. Тогда $\nabla \cdot (x\mathbf{r}) = 4x$, $\nabla \cdot (y\mathbf{r}) = 4y$, $\nabla \cdot (z\mathbf{r}) = 4z$ и из теоремы о дивергенции получим

$$\iint_S \mathbf{m} \cdot (x\mathbf{r}) dS = \iiint_V \nabla \cdot (x\mathbf{r}) dV = 4 \iiint_V x dV,$$

$$\iint_S \mathbf{m} \cdot (y\mathbf{r}) dS = \iiint_V \nabla \cdot (y\mathbf{r}) dV = 4 \iiint_V y dV,$$

$$\iint_S \mathbf{m} \cdot (z\mathbf{r}) dS = \iiint_V \nabla \cdot (z\mathbf{r}) dV = 4 \iiint_V z dV.$$

Таким образом, статические моменты тела можно вычислить с помощью поверхностных интегралов по формулам

$$M_x = \frac{\rho}{4} \iint_S \mathbf{m} \cdot (x\mathbf{r}) dS = \frac{\rho}{4} \iint_{\Omega} \mathbf{m} \cdot (x\mathbf{r}) \sqrt{g_{11}g_{22} - g_{12}^2} du dv, \quad (8.6.9)$$

$$M_y = \frac{\rho}{4} \iint_S \mathbf{m} \cdot (y\mathbf{r}) dS = \frac{\rho}{4} \iint_{\Omega} \mathbf{m} \cdot (y\mathbf{r}) \sqrt{g_{11}g_{22} - g_{12}^2} du dv, \quad (8.6.10)$$

$$M_z = \frac{\rho}{4} \iint_S \mathbf{m} \cdot (z\mathbf{r}) dS = \frac{\rho}{4} \iint_{\Omega} \mathbf{m} \cdot (z\mathbf{r}) \sqrt{g_{11}g_{22} - g_{12}^2} du dv. \quad (8.6.11)$$

В формулах (8.6.9)–(8.6.11) подразумевается суммирование по граням тела. Так же как и в (8.6.6) \mathbf{m} — нормаль граней тела. Полученные двойные интегралы могут быть вычислены с помощью кубатурных формул.

Статические моменты представляют собой компоненты вектора, так как при преобразованиях координат ведут себя аналогично компонентам вектора. Вектор $\mathbf{M} = [M_x \ M_y \ M_z]^T$ можно получить, положив в (8.5.13) $\mathbf{F}(\mathbf{r}) = (1/4)\mathbf{r}\mathbf{r}$, где $\mathbf{r}\mathbf{r}$ — диадное произведение векторов. Тогда

$$\begin{aligned} \nabla \cdot \left(\frac{1}{4} \mathbf{r}\mathbf{r} \right) &= \frac{1}{4} \begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \end{bmatrix} \cdot \begin{bmatrix} xx & xy & xz \\ yx & yy & yz \\ zx & zy & zz \end{bmatrix} = \\ &= \frac{1}{4} \begin{bmatrix} \frac{\partial(xx)}{\partial x} + \frac{\partial(xy)}{\partial y} + \frac{\partial(xz)}{\partial z} \\ \frac{\partial(yx)}{\partial x} + \frac{\partial(yy)}{\partial y} + \frac{\partial(yz)}{\partial z} \\ \frac{\partial(zx)}{\partial x} + \frac{\partial(zy)}{\partial y} + \frac{\partial(zz)}{\partial z} \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{r} \end{aligned}$$

и, используя теорему о дивергенции, получим

$$\begin{aligned} \mathbf{M} &= \rho \iiint_V \mathbf{r} dV = \frac{\rho}{4} \iiint_V \nabla \cdot (\mathbf{r}\mathbf{r}) dV = \frac{\rho}{4} \iint_S \mathbf{m} \cdot (\mathbf{r}\mathbf{r}) dS = \\ &= \frac{\rho}{4} \iint_S \mathbf{r}(\mathbf{m} \cdot \mathbf{r}) dS = \frac{\rho}{4} \iint_{\Omega} \mathbf{r}(\mathbf{m} \cdot \mathbf{r}) \sqrt{g_{11}g_{22} - g_{12}^2} du dv. \quad (8.6.12) \end{aligned}$$

Формулу (8.6.12) можно получить, рассмотрев элементарную пирамиду, приведенную на рис. 8.6.1. Центр масс элементарной пирамиды, показанной на рис. 8.6.1, определяется радиус-вектором $(3/4)\mathbf{r}$. Произведение массы на радиус-вектор центра масс элементарной пирамиды определяет статический момент элементарной пирамиды. Предел суммы статических моментов всех элементарных пирамид тела равен статическому моменту этого тела. При этом объем элементарной пирамиды должен быть взят со знаком, соответствующим знаку скалярного произведения $\mathbf{m} \cdot \mathbf{r}$. В результате мы приходим к формуле (8.6.12).

Центр масс тела. Рассмотрим, как изменятся статические моменты при параллельном переносе начала декартовой прямоугольной системы координат. Пусть начало декартовой прямоугольной системы координат $Oxyz$ перемещено на вектор \mathbf{q} , имеющий в данной системе компоненты $\mathbf{q} = [x_q \ y_q \ z_q]^T$. В новой системе координат $O'x'y'z'$ радиус-вектор точки тела равен $\mathbf{r}' = \mathbf{r} - \mathbf{q}$, а вектор статического момента будет равен

$$\mathbf{M}' = \iiint_V \rho \mathbf{r}' dV = \iiint_V \rho(\mathbf{r} - \mathbf{q}) dV = \mathbf{M} - \mathbf{q}\rho V. \quad (8.6.13)$$

Если мы перенесем начало координат $Oxyz$ на вектор

$$\mathbf{r}_c = \frac{\mathbf{M}}{\rho V} = \frac{\mathbf{M}}{M}, \quad (8.6.14)$$

то статический момент в новой системе координат будет равен нулю. Точка, при параллельном переносе в которую начала координат статические моменты тела становятся равными нулю, называется центром масс тела. Таким образом, формула (8.6.14) определяет радиус-вектор центра масс тела. Координаты центра масс вычисляются формулами

$$x_c = \frac{M_x}{\rho V}, \quad y_c = \frac{M_y}{\rho V}, \quad z_c = \frac{M_z}{\rho V}. \quad (8.6.15)$$

Система координат, в которой статические моменты тела равны нулю, называется *центральной системой координат*.

При преобразовании координат $\mathbf{r}' = \mathbf{A} \cdot \mathbf{r}$, где

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

вектор статического момента и его компоненты преобразуются по формулам

$$\mathbf{M}' = \iiint_V \rho \mathbf{r}' dV = \iiint_V \rho \mathbf{A} \cdot \mathbf{r} dV = \mathbf{A} \cdot \mathbf{M} \quad (8.6.16)$$

или

$$\begin{aligned} M_x' &= a_{11}M_x + a_{12}M_y + a_{13}M_z, \\ M_y' &= a_{21}M_x + a_{22}M_y + a_{23}M_z, \\ M_z' &= a_{31}M_x + a_{32}M_y + a_{33}M_z. \end{aligned}$$

Из формул преобразования следует, что в центральной системе координат статические моменты тела всегда равны нулю независимо от ориентации координатных осей.

8.7. Моменты инерции тела

Моменты инерции тела определяются объемными интегралами

$$\begin{aligned} J_{xx} &= \iiint_V \rho(y^2 + z^2) dV, \\ J_{yy} &= \iiint_V \rho(z^2 + x^2) dV, \\ J_{zz} &= \iiint_V \rho(x^2 + y^2) dV, \end{aligned} \quad (8.7.1)$$

а также

$$\begin{aligned} J_{xy} = J_{yx} &= - \iiint_V \rho xy \, dV, \\ J_{yz} = J_{zy} &= - \iiint_V \rho yz \, dV, \\ J_{zx} = J_{xz} &= - \iiint_V \rho xz \, dV, \end{aligned} \quad (8.7.2)$$

где x, y, z — компоненты радиус-вектора \mathbf{r} . Величины J_{xx}, J_{yy}, J_{zz} называются моментами инерции тела относительно соответствующих координатных осей, а величины $-J_{xy}, -J_{yz}, -J_{zx}$ называются центробежными моментами инерции тела. Интегрирование в (8.7.1)–(8.7.2) выполняется по объему тела. Пусть плотность тела постоянна по всему его объему, это позволит вычислить объемные интегралы через поверхностные интегралы.

Тензор инерции. Моменты инерции вводятся при рассмотрении движения (вращения) твердого тела, имеющего неподвижную точку. Поместим начало координат в неподвижную точку, тогда скорость некоторой точки тела определяется соотношением $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, где $\boldsymbol{\omega}$ — вектор мгновенной угловой скорости тела. Вектор момента количества движения тела равен

$$\mathbf{L} = \iiint_V \mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r}) \rho \, dV = \iiint_V (\mathbf{r}^2 \boldsymbol{\omega} - (\mathbf{r} \cdot \boldsymbol{\omega}) \mathbf{r}) \rho \, dV. \quad (8.7.3)$$

Проекции вектора момента количества движения тела на оси декартовой системы координат равны

$$\begin{aligned} L_x &= \omega_x \iiint_V \rho (y^2 + z^2) \, dV - \omega_y \iiint_V \rho xy \, dV - \omega_z \iiint_V \rho xz \, dV = \\ &= \omega_x J_{xx} + \omega_y J_{xy} + \omega_z J_{xz}, \end{aligned}$$

$$\begin{aligned} L_y &= -\omega_x \iiint_V \rho yx \, dV + \omega_y \iiint_V \rho (z^2 + x^2) \, dV - \omega_z \iiint_V \rho yz \, dV = \\ &= \omega_x J_{yx} + \omega_y J_{yy} + \omega_z J_{yz}, \end{aligned}$$

$$\begin{aligned} L_z &= -\omega_x \iiint_V \rho zx \, dV - \omega_y \iiint_V \rho zy \, dV + \omega_z \iiint_V \rho (x^2 + y^2) \, dV = \\ &= \omega_x J_{zx} + \omega_y J_{zy} + \omega_z J_{zz}, \end{aligned}$$

где $\omega_x, \omega_y, \omega_z$ — компоненты вектора угловой скорости. В результате вектор момента количества движения тела можно определить формулой

$$\mathbf{L} = \mathbf{J} \cdot \boldsymbol{\omega}, \quad (8.7.4)$$

где величина \mathbf{J} определяется матрицей, элементами которой являются осевые моменты инерции (8.7.1) и центробежные моменты инерции (8.7.2), взятые с противоположным знаком:

$$\mathbf{J} = \begin{bmatrix} J_{xx} & J_{xy} & J_{xz} \\ J_{yx} & J_{yy} & J_{yz} \\ J_{zx} & J_{zy} & J_{zz} \end{bmatrix}. \quad (8.7.5)$$

Так как $\boldsymbol{\omega}$ и \mathbf{L} являются аксиальными векторами (аксиальный вектор или псевдовектор изменяет свое направление на противоположное при переходе от правой системы координат к левой), то величина \mathbf{J} должна представлять собой тензор второго ранга. Действительно, при преобразовании координат, описываемых некоторой матрицей \mathbf{A} , величины $\boldsymbol{\omega}$ и \mathbf{L} преобразуются по формулам $\boldsymbol{\omega}' = \mathbf{A} \cdot \boldsymbol{\omega}$ (или $\boldsymbol{\omega}' = -\mathbf{A} \cdot \boldsymbol{\omega}$) и $\mathbf{L}' = \mathbf{A} \cdot \mathbf{L}$ (или $\mathbf{L}' = -\mathbf{A} \cdot \mathbf{L}$), тогда $\mathbf{L}' = \pm \mathbf{A} \cdot \mathbf{L} = \pm \mathbf{A} \cdot \mathbf{J} \cdot \boldsymbol{\omega} = \pm \mathbf{A} \cdot \mathbf{J} \cdot \mathbf{A}^T \cdot \mathbf{A} \cdot \boldsymbol{\omega} = \mathbf{A} \cdot \mathbf{J} \cdot \mathbf{A}^T \cdot \boldsymbol{\omega}' = \mathbf{J}' \cdot \boldsymbol{\omega}'$ и величина \mathbf{J} должна преобразовываться по формуле

$$\mathbf{J}' = \mathbf{A} \cdot \mathbf{J} \cdot \mathbf{A}^T. \quad (8.7.6)$$

Таким образом, моменты инерции тела (с точностью до знака) являются компонентами тензора \mathbf{J} , который называется *тензором инерции*. Компоненты тензора инерции зависят от ориентации системы координат, но тензор инерции представляет собой физический объект, не зависящий от ориентации системы координат. Тензор инерции характеризует инертность тела при его вращении вокруг некоторой неподвижной точки и является аналогом массы, которая характеризует инертность тела при поступательном движении. Тензор инерции имеет различные значения в различных точках тела (является функцией неподвижной точки тела). Он образует тензорное поле.

Собственные значения матрицы инерции. В формуле (8.7.4) вектор количества движения \mathbf{L} в общем случае не совпадает по направлению с вектором угловой скорости $\boldsymbol{\omega}$. Это приводит к тому, что при вращении тела вокруг заданной оси, в точках крепления тела возникают зависящие от модуля угловой скорости реакции. Такое вращение является динамически неуравновешенным. Выясним, существуют ли для данного тензора инерции \mathbf{J} такие направления, при которых вектор количества движения \mathbf{L} совпадает по направлению с вектором угловой скорости $\boldsymbol{\omega}$. Пусть это направление определяется единичным вектором $\mathbf{e} = [x \ y \ z]^T$, тогда вектор $\mathbf{J} \cdot \mathbf{e}$, должен быть пропорционален вектору \mathbf{e} . Запишем это условие в виде равенства

$$\mathbf{J} \cdot \mathbf{e} - \lambda \mathbf{e} \equiv (\mathbf{J} - \lambda \mathbf{E}) \cdot \mathbf{e} = 0, \quad (8.7.7)$$

где λ — пока неизвестный скаляр, \mathbf{E} — тензорная единица ранга тензора \mathbf{J} . Проецируя это равенство на оси координат, получим три равенства

$$\begin{aligned} (J_{xx} - \lambda)x + J_{xy}y + J_{xz}z &= 0, \\ J_{xy}x + (J_{yy} - \lambda)y + J_{yz}z &= 0, \\ J_{xz}x + J_{yz}y + (J_{zz} - \lambda)z &= 0, \end{aligned} \quad (8.7.8)$$

которые будем рассматривать как однородную систему линейных алгебраических уравнений относительно x, y, z . Тривиальное решение нам не подходит, так как необходимо найти вектор, для которого

$$x^2 + y^2 + z^2 = 1. \quad (8.7.9)$$

Система (8.7.8) имеет нетривиальные решения, когда определитель системы равен нулю:

$$\begin{vmatrix} J_{xx} - \lambda & J_{xy} & J_{xz} \\ J_{yx} & J_{yy} - \lambda & J_{yz} \\ J_{zx} & J_{zy} & J_{zz} - \lambda \end{vmatrix} = 0. \quad (8.7.10)$$

Это равенство выполняется, когда скаляр λ удовлетворяет уравнению

$$\lambda^3 - I_1\lambda^2 + I_2\lambda - I_3 = 0, \quad (8.7.11)$$

где

$$I_1 = J_{xx} + J_{yy} + J_{zz},$$

$$I_2 = \begin{vmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{vmatrix} + \begin{vmatrix} J_{yy} & J_{yz} \\ J_{zy} & J_{zz} \end{vmatrix} + \begin{vmatrix} J_{zz} & J_{zx} \\ J_{xz} & J_{xx} \end{vmatrix},$$

$$I_3 = \begin{vmatrix} J_{xx} & J_{xy} & J_{xz} \\ J_{yx} & J_{yy} & J_{yz} \\ J_{zx} & J_{zy} & J_{zz} \end{vmatrix}.$$

Уравнение (8.7.11) называется *характеристическим уравнением тензора инерции*. Корни $\lambda_1, \lambda_2, \lambda_3$ характеристического уравнения (8.7.11) являются действительными, так как матрица тензора инерции является симметричной. Скаляры $\lambda_i, i = 1, 2, 3$ являются *собственными значениями матрицы моментов инерции*. Каждому λ_i соответствует собственный вектор \mathbf{e}_i .

Главные оси инерции. Предположим, что собственные значения матрицы инерции λ_1 и λ_2 различны. Покажем, что соответствующие им собственные векторы \mathbf{e}_1 и \mathbf{e}_2 ортогональны. Действительно, для данных собственных значений выполняются равенства

$$\mathbf{J} \cdot \mathbf{e}_1 - \lambda_1 \mathbf{e}_1 = 0, \quad \mathbf{J} \cdot \mathbf{e}_2 - \lambda_2 \mathbf{e}_2 = 0.$$

Умножим первое равенство скалярно на \mathbf{e}_2 , а второе — на \mathbf{e}_1 и вычтем второе из первого. Используя симметрию тензора инерции, получим

$$\mathbf{e}_2 \cdot \mathbf{J} \cdot \mathbf{e}_1 - \lambda_1 \mathbf{e}_2 \cdot \mathbf{e}_1 - \mathbf{e}_1 \cdot \mathbf{J} \cdot \mathbf{e}_2 + \lambda_2 \mathbf{e}_1 \cdot \mathbf{e}_2 = 0$$

или

$$(\lambda_2 - \lambda_1) \mathbf{e}_1 \cdot \mathbf{e}_2 = 0.$$

Из последнего равенства в силу предположения $\lambda_2 \neq \lambda_1$ следует, что $\mathbf{e}_1 \cdot \mathbf{e}_2 = 0$ и собственные векторы \mathbf{e}_1 и \mathbf{e}_2 ортогональны. Аналогично доказывается ортогональность других собственных векторов. Будем говорить, что собственный вектор определяет главное направление тензора инерции в пространстве. Если два

из трех собственных значения матрицы инерции равны, то мы сможем определить только одно из главных направлений, а другими главными направлениями могут быть любые направления, ортогональные найденному направлению. Если все три собственных значения матрицы инерции равны между собой, то главными направлениями могут быть любые направления в пространстве. В любом случае можно получить три взаимно ортогональных главных направления.

Три взаимно ортогональных главных направления, соответствующих собственным векторам $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ определяют *главные оси инерции* тела и являются базисными векторами главной системы координат. Если орты новой декартовой прямоугольной системы координат ориентировать вдоль главных направлений, то осевые моменты инерции тела будут равны собственным значениям матрицы инерции, а центробежные моменты инерции будут равны нулю, т.е. $J_{11} = \lambda_1$, $J_{22} = \lambda_2$, $J_{33} = \lambda_3$, $J_{12} = 0$, $J_{23} = 0$, $J_{13} = 0$. Докажем это. Для главных направлений выполняются равенства

$$\mathbf{J} \cdot \mathbf{e}_i = \lambda_i \mathbf{e}_i, \quad i = 1, 2, 3.$$

Тогда

$$\mathbf{e}_k \cdot \mathbf{J} \cdot \mathbf{e}_i = \lambda_i \delta_{ki}, \quad (8.7.12)$$

что доказывает высказанное утверждение. Матрица тензора инерции в главной системе координат имеет вид

$$\mathbf{J} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \equiv \begin{bmatrix} J_{11} & 0 & 0 \\ 0 & J_{22} & 0 \\ 0 & 0 & J_{33} \end{bmatrix}. \quad (8.7.13)$$

Система координат, в которой все центробежные моменты инерции тела равны нулю, называется *главной системой координат*. Моменты инерции J_{11}, J_{22}, J_{33} в главной системе координат называются *главными моментами инерции*. Главные направления тензора инерции зависят от неподвижной точки тела (от выбора начала координат). Система координат, в которой статические моменты и все центробежные моменты инерции тела равны нулю, называется *главной центральной системой координат*. Моменты инерции в главной центральной системе координат называются *главными центральными моментами инерции*. Направления главных осей инерции тела лежат в плоскостях симметрии тела. Начало главной центральной системы координат находится в центре масс тела.

Найдем направления главных осей инерции. Пусть главному моменту инерции λ_i соответствует главное направление, определяемое ортом $\mathbf{e}_i = [x_i \ y_i \ z_i]^T$. Возможны три случая соотношений между корнями характеристического уравнения. Первый случай — все корни характеристического уравнения различны, второй случай — два из трех корней характеристического уравнения равны и отличны от третьего корня, третий случай — все три корня характеристического уравнения равны между собой

Пусть $\lambda_1 \neq \lambda_2 \neq \lambda_3$ (корни характеристического уравнения различные). Для каждого λ_i составим систему уравнений (8.7.8)

$$\begin{aligned} (J_{11} - \lambda_i)x_i + J_{12}y_i + J_{13}z_i &= 0, \\ J_{12}x_i + (J_{22} - \lambda_i)y_i + J_{23}z_i &= 0, \\ J_{13}x_i + J_{23}y_i + (J_{33} - \lambda_i)z_i &= 0. \end{aligned}$$

Среди данных трех уравнений одно является линейной комбинацией двух других, так как определитель системы (8.7.10) равен нулю. Нам нужно выяснить, какие два из трех уравнений являются линейно независимыми. Для этого вычислим определители

$$A_1 = \begin{vmatrix} J_{22} - \lambda_i & J_{23} \\ J_{23} & J_{33} - \lambda_i \end{vmatrix}, \quad A_2 = \begin{vmatrix} J_{33} - \lambda_i & J_{13} \\ J_{13} & J_{11} - \lambda_i \end{vmatrix},$$

$$A_3 = \begin{vmatrix} J_{11} - \lambda_i & J_{12} \\ J_{12} & J_{22} - \lambda_i \end{vmatrix}.$$

По крайней мере, один из них должен быть отличен от нуля. Пусть $A_2 \neq 0$. Тогда компоненты соответствующего главного вектора найдем из системы уравнений

$$\begin{aligned} (J_{11} - \lambda_i)x_i + J_{12}y_i + J_{13}z_i &= 0, \\ J_{13}x_i + J_{23}y_i + (J_{33} - \lambda_i)z_i &= 0, \\ x_i^2 + y_i^2 + z_i^2 &= 1. \end{aligned}$$

Для ее решения из первых двух уравнений выразим x_i и z_i через y_i и подставим их в третье уравнение. В результате получим y_i . Далее из первых двух уравнений найдем x_i и z_i . Так мы можем определить все три собственных вектора. Вектор $e_i = [x_i \ y_i \ z_i]^T$ соответствует λ_i .

Пусть два из трех корней характеристического уравнения равны между собой, например, $\lambda_1 = \lambda_2 \neq \lambda_3$. В этом случае базисный вектор $e_3 = [x_3 \ y_3 \ z_3]^T$ может быть найден описанным выше способом. Двумя другими базисными векторами могут служить любые ортогональные друг другу и вектору e_3 орта.

Если $\lambda_1 = \lambda_2 = \lambda_3$, то собственными векторами могут служить любые три взаимно ортогональные орта, а $\lambda_i = I_1/3$.

Эллипсоид инерции. Рассмотрим поведение компонент тензора инерции при произвольном повороте осей координатной системы. Пусть мы переходим от исходной системы координат к новой системы координат с тем же центром и с ортами $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$. Радиус-вектор произвольной точки в новой системе координат будет равен $\mathbf{r}' = \mathbf{A} \cdot \mathbf{r}$, где

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix}.$$

Компоненты матрицы преобразования \mathbf{A} являются координатами ортов $\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3$ новой системы координат: $\mathbf{i}_1 = [x_1 \ y_1 \ z_1]^T$, $\mathbf{i}_2 = [x_2 \ y_2 \ z_2]^T$, $\mathbf{i}_3 = [x_3 \ y_3 \ z_3]^T$. Соотношение (8.7.6) в развернутой записи имеет вид

$$\begin{bmatrix} J_{11} & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \\ J_{31} & J_{32} & J_{33} \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix} \cdot \begin{bmatrix} J_{xx} & J_{xy} & J_{xz} \\ J_{yx} & J_{yy} & J_{yz} \\ J_{zx} & J_{zy} & J_{zz} \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix}.$$

Компоненты тензора инерции при переходе к новой системе координат преобразуются по формулам

$$\begin{aligned}
 J_{11} &= x_1^2 J_{xx} + y_1^2 J_{yy} + z_1^2 J_{zz} + 2x_1 y_1 J_{xy} + 2y_1 z_1 J_{yz} + 2z_1 x_1 J_{zx}, \\
 J_{22} &= x_2^2 J_{xx} + y_2^2 J_{yy} + z_2^2 J_{zz} + 2x_2 y_2 J_{xy} + 2y_2 z_2 J_{yz} + 2z_2 x_2 J_{zx}, \\
 J_{33} &= x_3^2 J_{xx} + y_3^2 J_{yy} + z_3^2 J_{zz} + 2x_3 y_3 J_{xy} + 2y_3 z_3 J_{yz} + 2z_3 x_3 J_{zx}, \\
 J_{12} &= J_{21} = x_1 x_2 J_{xx} + y_1 y_2 J_{yy} + z_1 z_2 J_{zz} + \\
 &\quad + (x_1 y_2 + x_2 y_1) J_{xy} + (y_1 z_2 + y_2 z_1) J_{yz} + (z_1 x_2 + z_2 x_1) J_{zx}, \quad (8.7.14) \\
 J_{23} &= J_{32} = x_2 x_3 J_{xx} + y_2 y_3 J_{yy} + z_2 z_3 J_{zz} + \\
 &\quad + (x_2 y_3 + x_3 y_2) J_{xy} + (y_2 z_3 + y_3 z_2) J_{yz} + (z_2 x_3 + z_3 x_2) J_{zx}, \\
 J_{31} &= J_{13} = x_3 x_1 J_{xx} + y_3 y_1 J_{yy} + z_3 z_1 J_{zz} + \\
 &\quad + (x_3 y_1 + x_1 y_3) J_{xy} + (y_3 z_1 + y_1 z_3) J_{yz} + (z_3 x_1 + z_1 x_3) J_{zx}.
 \end{aligned}$$

Можно показать, что величины I_1, I_2, I_3 не изменяются при повороте системы координат. Они являются инвариантами относительно преобразования координат, сохраняющего начало. Неизменность главных направлений, главных моментов инерции J_{11}, J_{22}, J_{33} и величин I_1, I_2, I_3 при изменении направления координатных осей является проявлением независимости тензора инерции от ориентации системы координат. Любая из первых трех формул (8.7.14) может считаться формулой для определения осевого момента инерции тела относительно произвольной оси, заданной своими проекциями на оси исходной системы координат. Пусть некоторая ось проходит через начало координат и имеет компоненты $\mathbf{l} = [x_l \ y_l \ z_l]^T$. Тогда осевой момент инерции тела относительно этой оси равен

$$J_{ll} = x_l^2 J_{xx} + y_l^2 J_{yy} + z_l^2 J_{zz} + 2x_l y_l J_{xy} + 2y_l z_l J_{yz} + 2z_l x_l J_{zx}.$$

Разделим компоненты оси на осевой момент инерции J_{ll} , введем обозначения $x = x_l/\sqrt{J_{ll}}, y = y_l/\sqrt{J_{ll}}, z = z_l/\sqrt{J_{ll}}$ и получим равенство

$$x^2 J_{xx} + y^2 J_{yy} + z^2 J_{zz} + 2xy J_{xy} + 2yz J_{yz} + 2zx J_{zx} = 1. \quad (8.7.15)$$

Равенство (8.7.15) можно рассматривать как уравнение поверхности второго порядка. Длина вектора, проведенного из начала координат к любой точке поверхности (8.7.15), обратно пропорциональна квадратному корню из момента инерции тела относительно оси, совпадающей с данным вектором по направлению. Момент инерции тела относительно любой оси есть величина положительная (все собственные значения матрицы инерции положительные), поэтому поверхность (8.7.15) представляет собой эллипсоид. В главной системе координат уравнение эллипсоида инерции имеет канонический вид

$$x^2 J_{11} + y^2 J_{22} + z^2 J_{33} = 1. \quad (8.7.16)$$

Если два из трех главных момента инерции равны, то эллипсоид инерции является эллипсоидом вращения. Если все три главных момента инерции равны, то эллипсоид инерции превращается в сферу.

Тензорное поле. При параллельном переносе начала декартовой прямоугольной системы координат изменяются не только компоненты тензора инерции тела, но и сам тензор инерции. Пусть начало декартовой прямоугольной системы координат $Oxyz$ перемещено на вектор \mathbf{q} , имеющий в данной системе компоненты $\mathbf{q} = [x_q \ y_q \ z_q]^T$. В новой системе координат $O'x'y'z'$ радиус-вектор точки тела равен $\mathbf{r}' = \mathbf{r} - \mathbf{q}$, а компоненты тензора инерции будут равны

$$J_{xx'} = \iiint_V (y^2 + z^2 - 2yy_q - 2zz_q + y_q^2 + z_q^2) \rho dV = \\ = J_{xx} - 2y_q M_y - 2z_q M_z + (y_q^2 + z_q^2) M,$$

$$J_{yy'} = \iiint_V (z^2 + x^2 - 2zz_q - 2xx_q + z_q^2 + x_q^2) \rho dV = \\ = J_{yy} - 2z_q M_z - 2x_q M_x + (z_q^2 + x_q^2) M,$$

$$J_{zz'} = \iiint_V (x^2 + y^2 - 2xx_q - 2yy_q + x_q^2 + y_q^2) \rho dV = \\ = J_{zz} - 2x_q M_x - 2y_q M_y + (x_q^2 + y_q^2) M,$$

$$J_{yz'} = - \iiint_V (yz - yz_q - zy_q + y_q z_q) \rho dV = J_{yz} + z_q M_y + y_q M_z - y_q z_q M,$$

$$J_{zx'} = - \iiint_V (zx - zx_q - xz_q + z_q x_q) \rho dV = J_{zx} + x_q M_z + z_q M_x - z_q x_q M,$$

$$J_{xy'} = - \iiint_V (xy - xy_q - yx_q + x_q y_q) \rho dV = J_{xy} + y_q M_x + x_q M_y - x_q y_q M.$$

Пусть система координат $Oxyz$ является главной центральной системой координат тела. Тогда в ней $M_x = M_y = M_z = 0$, $J_{xy} = J_{yz} = J_{zx} = 0$. При перемещении вдоль главной центральной оси две из трех компонент вектора сдвига равны нулю и, следовательно, главные направления сохраняются. В новой системе координат в общем случае центробежные моменты инерции будут отличны от нуля и, следовательно, она не будет являться ни центральной, ни главной. Это говорит о том, что тензор инерции является функцией положения неподвижной точки тела, образуя в теле *тензорное поле*.

Вычисление моментов инерции тела. Не имея явных уравнений граней тела, мы не можем непосредственно вычислить объемные интегралы (8.7.1) и (8.7.2). В нашем распоряжении есть параметрическое представление граней тела. Грани тела образуют одну или несколько замкнутых оболочек, поэтому объемный интеграл для каждой замкнутой оболочки мы сведем к поверхностному интегралу. Для вычисления компонент тензора инерции положим в (8.5.13) соответственно $\mathbf{F}(\mathbf{r}) = x^2 \mathbf{r}$, $\mathbf{F}(\mathbf{r}) = y^2 \mathbf{r}$, $\mathbf{F}(\mathbf{r}) = z^2 \mathbf{r}$, $\mathbf{F}(\mathbf{r}) = yz \mathbf{r}$, $\mathbf{F}(\mathbf{r}) = zx \mathbf{r}$, $\mathbf{F}(\mathbf{r}) = xy \mathbf{r}$. Тогда $\nabla \cdot (x^2 \mathbf{r}) = 5x^2$, $\nabla \cdot (y^2 \mathbf{r}) = 5y^2$, $\nabla \cdot (z^2 \mathbf{r}) = 5z^2$, $\nabla \cdot (yz \mathbf{r}) = 5yz$, $\nabla \cdot (zx \mathbf{r}) = 5zx$, $\nabla \cdot (xy \mathbf{r}) = 5xy$ и из теоремы Остроградского–Гаусса (8.5.13)

ПОЛУЧИМ

$$\iint_S \mathbf{m} \cdot (x^2 \mathbf{r}) dS = 5 \iiint_V x^2 dV,$$

$$\iint_S \mathbf{m} \cdot (y^2 \mathbf{r}) dS = 5 \iiint_V y^2 dV,$$

$$\iint_S \mathbf{m} \cdot (z^2 \mathbf{r}) dS = 5 \iiint_V z^2 dV,$$

$$\iint_S \mathbf{m} \cdot (xy \mathbf{r}) dS = 5 \iiint_V xy dV,$$

$$\iint_S \mathbf{m} \cdot (yz \mathbf{r}) dS = 5 \iiint_V yz dV,$$

$$\iint_S \mathbf{m} \cdot (zx \mathbf{r}) dS = 5 \iiint_V zx dV.$$

Таким образом, компоненты тензора инерции тела определяются формулами

$$J_{xx} = \frac{\rho}{5} \iint_S (y^2 + z^2) \mathbf{m} \cdot \mathbf{r} dS = \frac{\rho}{5} \iint_{\Omega} (y^2 + z^2) \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11}g_{22} - g_{11}^2} du dv, \quad (8.7.17)$$

$$J_{yy} = \frac{\rho}{5} \iint_S (z^2 + x^2) \mathbf{m} \cdot \mathbf{r} dS = \frac{\rho}{5} \iint_{\Omega} (z^2 + x^2) \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11}g_{22} - g_{11}^2} du dv, \quad (8.7.18)$$

$$J_{zz} = \frac{\rho}{5} \iint_S (x^2 + y^2) \mathbf{m} \cdot \mathbf{r} dS = \frac{\rho}{5} \iint_{\Omega} (x^2 + y^2) \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11}g_{22} - g_{11}^2} du dv, \quad (8.7.19)$$

$$J_{xy} = -\frac{\rho}{5} \iint_S xy \mathbf{m} \cdot \mathbf{r} dS = -\frac{\rho}{5} \iint_{\Omega} xy \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11}g_{22} - g_{11}^2} du dv, \quad (8.7.20)$$

$$J_{yz} = -\frac{\rho}{5} \iint_S yz \mathbf{m} \cdot \mathbf{r} dS = -\frac{\rho}{5} \iint_{\Omega} yz \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11}g_{22} - g_{11}^2} du dv, \quad (8.7.21)$$

$$J_{xz} = -\frac{\rho}{5} \iint_S xz \mathbf{m} \cdot \mathbf{r} dS = -\frac{\rho}{5} \iint_{\Omega} xz \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11}g_{22} - g_{11}^2} du dv. \quad (8.7.22)$$

В формулах (8.7.17)–(8.7.22) подразумевается суммирование по граням тела. Так же как и в (8.6.12), \mathbf{m} есть нормаль граней тела. Формулы (8.7.17)–(8.7.22) позволяют вычислить моменты инерции в той системе координат, в которой задано тело.

Часто тело описывается в некоторой глобальной системе координат, начало которой не совпадает с центром масс тела, а моменты инерции требуется определить в центральной системе координат. В этом случае можно поступить следующим образом. Пусть центр масс имеет координаты x_c, y_c, z_c . Вычислим интегралы

$$N_{xx} = \rho \iiint_V (x - x_c)^2 dV = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} x^2 dS - 2x_c M_x + x_c^2 \rho V,$$

$$N_{yy} = \rho \iiint_V (y - y_c)^2 dV = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} y^2 dS - 2y_c M_y + y_c^2 \rho V,$$

$$N_{zz} = \rho \iiint_V (z - z_c)^2 dV = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} z^2 dS - 2z_c M_z + z_c^2 \rho V,$$

$$N_{xy} = \rho \iiint_V (x - x_c)(y - y_c) dV = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} xy dS - x_c M_y - y_c M_x + x_c y_c \rho V,$$

$$N_{yz} = \rho \iiint_V (y - y_c)(z - z_c) dV = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} yz dS - y_c M_z - z_c M_y + y_c z_c \rho V,$$

$$N_{zx} = \rho \iiint_V (z - z_c)(x - x_c) dV = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} zx dS - z_c M_x - x_c M_z + z_c x_c \rho V.$$

С учетом формул (8.6.15) получим

$$N_{xx} = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} x^2 dS - x_c^2 \rho V = \frac{\rho}{5} \iint_{\Omega} x^2 \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11} g_{22} - g_{11}^2} du dv - x_c^2 \rho V,$$

$$N_{yy} = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} y^2 dS - y_c^2 \rho V = \frac{\rho}{5} \iint_{\Omega} y^2 \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11} g_{22} - g_{11}^2} du dv - y_c^2 \rho V,$$

$$N_{zz} = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} z^2 dS - z_c^2 \rho V = \frac{\rho}{5} \iint_{\Omega} z^2 \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11} g_{22} - g_{11}^2} du dv - z_c^2 \rho V,$$

$$N_{xy} = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} xy dS - x_c y_c \rho V = \frac{\rho}{5} \iint_{\Omega} xy \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11} g_{22} - g_{11}^2} du dv - x_c y_c \rho V,$$

$$N_{yz} = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} yz dS - y_c z_c \rho V = \frac{\rho}{5} \iint_{\Omega} yz \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11} g_{22} - g_{11}^2} du dv - y_c z_c \rho V,$$

$$N_{zx} = \frac{\rho}{5} \iint_S \mathbf{m} \cdot \mathbf{r} zx dS - z_c x_c \rho V = \frac{\rho}{5} \iint_{\Omega} zx \mathbf{m} \cdot \mathbf{r} \sqrt{g_{11} g_{22} - g_{11}^2} du dv - z_c x_c \rho V.$$

Компоненты тензора инерции тела относительно центральной системы координат определяются формулами

$$J_{xx} = N_{yy} + N_{zz}, \quad (8.7.23)$$

$$J_{yy} = N_{zz} + N_{xx}, \quad (8.7.24)$$

$$J_{zz} = N_{xx} + N_{yy}, \quad (8.7.25)$$

$$J_{xy} = -N_{xy}, \quad (8.7.26)$$

$$J_{yz} = -N_{yz}, \quad (8.7.27)$$

$$J_{zx} = -N_{zx}. \quad (8.7.28)$$

По этим компонентам могут быть вычислены главные моменты инерции тела и направления осей главной центральной системы координат.

Полученные двойные интегралы могут быть вычислены с помощью кубатурных формул, которые мы рассмотрим ниже.

8.8.* Решение кубического уравнения

Для определения канонических коэффициентов поверхности второго порядка и собственных значений матрицы инерции нужно решить кубическое уравнение (3.3.8) и (8.7.11):

$$\lambda^3 - I_1\lambda^2 + I_2\lambda - I_3 = 0.$$

Это уравнение может быть решено аналитически. Заменим в уравнении (8.7.11) неизвестную λ новой неизвестной x , связанной с λ равенством

$$\lambda = x + \frac{I_1}{3}. \quad (8.8.1)$$

После подстановки (8.8.1) в (8.7.11) получим неполное кубическое уравнение относительно x , не содержащее квадрата новой неизвестной:

$$x^3 + px + q = 0, \quad (8.8.2)$$

где введены обозначения

$$p = -\frac{1}{3}I_1^2 + I_2, \quad (8.8.3)$$

$$q = -\frac{2}{27}I_1^3 + \frac{1}{3}I_1I_2 - I_3. \quad (8.8.4)$$

В соответствии с основной теоремой алгебры уравнение (8.8.2) имеет три корня, в общем случае комплексных. Пусть x_i есть один из этих корней. Рассмотрим многочлен

$$f(y) = y^2 - x_i y - \frac{1}{3}p$$

некоторой вспомогательной неизвестной y . Этот многочлен обладает двумя корнями y_1 и y_2 , удовлетворяющими формулам Виета

$$y_1 + y_2 = x_i, \quad (8.8.5)$$

$$y_1 y_2 = -\frac{1}{3}p. \quad (8.8.6)$$

Подставим (8.8.5) в (8.8.2) и получим $(y_1 + y_2)^3 + p(y_1 + y_2) + q = 0$ или

$$y_1^3 + y_2^3 + (3y_1y_2 + p)(y_1 + y_2) + q = 0.$$

С учетом (8.8.6) последнее равенство примет вид

$$y_1^3 + y_2^3 = -q. \quad (8.8.7)$$

С другой стороны, из (8.8.6) следует, что

$$y_1^3 y_2^3 = -\frac{1}{27}p^3. \quad (8.8.8)$$

Равенства (8.8.7) и (8.8.8) свидетельствуют, что числа y_1^3 и y_2^3 являются корнями квадратного уравнения

$$z^2 + qz - \frac{1}{27}p^3 = 0. \quad (8.8.9)$$

Решив это уравнение, получим его корни

$$z_1 = y_1^3 = -\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}, \quad z_2 = y_2^3 = -\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} \quad (8.8.10)$$

В соответствии с (8.8.5) сумма кубических корней z_1 и z_2 является решением неполного кубического уравнения (8.8.2):

$$x_i = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} \quad (8.8.11)$$

Формула (8.8.11) называется *формулой Кардано*.

Так как кубический радикал на множестве комплексных чисел имеет три значения, то z_1 даст три значения для y_1 , а z_2 даст три значения для y_2 . В формуле (8.8.11) должны быть суммированы кубические радикалы (значения y_1 и y_2), удовлетворяющие условию (8.8.6). Пусть y_1 — любое из трех значений кубического радикала z_1 и y_2 — удовлетворяющее равенству (8.8.6) значение кубического радикала z_2 . Тогда корнями неполного кубического уравнения (8.8.2) будут

$$x_1 = y_1 + y_2, \quad x_2 = \varepsilon y_1 + \varepsilon^2 y_2, \quad x_3 = \varepsilon^2 y_1 + \varepsilon y_2, \quad (8.8.12)$$

где $\varepsilon = \cos(2\pi/3) + i \sin(2\pi/3) = -1/2 + i\sqrt{3}/2$ — кубический корень из единицы, а $\varepsilon^2 = \cos(4\pi/3) + i \sin(4\pi/3) = -1/2 - i\sqrt{3}/2$ — квадрат этого кубического корня из единицы.

Формула Кардано позволяет найти три корня, в общем случае комплексных, кубического уравнения. Коэффициенты кубического уравнения в общем случае могут быть комплексными. В случае вычисления главных моментов инерции коэффициенты (8.8.3) и (8.8.4) неполного кубического уравнения, так же как и инварианты I_1, I_2, I_3 , являются действительными числами. В формуле Кардано величина

$$D = -4p^3 - 27q^2 = -108 \left(\frac{q^2}{4} + \frac{p^3}{27} \right)$$

называется *дискриминантом* неполного кубического уравнения (8.8.2). Если коэффициенты p и q являются действительными числами, то дискриминант также является действительным числом и может принимать значения: $D = 0$, $D < 0$, $D > 0$.

Пусть $D = 0$. Тогда $y_1 = y_2 = \sqrt[3]{-q/2}$, а $\varepsilon + \varepsilon^2 = -1$. Подставив данные значения в (8.8.12), получим решение

$$x_1 = 2\sqrt[3]{-\frac{q}{2}}, \quad x_2 = -\sqrt[3]{-\frac{q}{2}}, \quad x_3 = -\sqrt[3]{-\frac{q}{2}}. \quad (8.8.13)$$

Таким образом, если $D = 0$, то кубическое уравнение с действительными коэффициентами имеет три действительных корня, два из которых равны между собой, а третий равен сумме первых двух, взятой с противоположным знаком.

Пусть $D < 0$. В этом случае под знаком каждого из квадратных радикалов стоит положительное действительное число, и, следовательно, под знаком каждого из кубических радикалов стоят действительные числа. Кубический корень из действительного числа имеет одно действительное и два комплексных сопряженных значения. Пусть действительными значениями этих кубических радикалов являются y_1 и y_2 . Подставив данные значения в (8.8.12), получим

$$\begin{aligned} x_1 &= y_1 + y_2, \\ x_2 &= -\frac{y_1 + y_2}{2} + i\sqrt{3}\frac{y_1 - y_2}{2}, \\ x_3 &= -\frac{y_1 + y_2}{2} - i\sqrt{3}\frac{y_1 - y_2}{2}. \end{aligned} \quad (8.8.14)$$

Таким образом, если $D < 0$, то кубическое уравнение с действительными коэффициентами имеет один действительный корень и два сопряженных комплексных корня.

Пусть $D > 0$. В этом случае под знаком каждого из квадратных радикалов стоит отрицательное действительное число, и, следовательно, под знаками кубических радикалов стоят сопряженные комплексные числа. Кубические корни из комплексных чисел в общем случае также являются комплексными числами. Но среди корней кубического уравнения с действительными коэффициентами обязательно один из корней должен быть действительным числом. Пусть действительным корнем кубического уравнения является число $x_1 = y_1 + y_2$. В соответствии с (8.8.5) и (8.8.6) сумма и произведение комплексных чисел y_1 и y_2 равны действительным числам, следовательно, y_1 и y_2 являются сопряженными комплексными числами как корни квадратного уравнения с действительными коэффициентами. Введем обозначение $y_1 = a + ib$, $y_2 = a - ib$ и подставим данные значения в (8.8.12). В результате получим

$$x_1 = 2a, \quad x_2 = -a - b\sqrt{3}, \quad x_3 = -a + b\sqrt{3}. \quad (8.8.15)$$

Таким образом, если $D > 0$, то кубическое уравнение с действительными коэффициентами имеет три действительных корня.

Можно показать, что все корни λ_1 , λ_2 и λ_3 характеристического кубического уравнения (8.7.11) являются действительными числами, поэтому рассмотрим подробнее этот случай. Итак, величина $\frac{q^2}{4} + \frac{p^3}{27}$ является отрицательной. Тогда

$$y_1^3 = -\frac{q}{2} + i\sqrt{-\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3} = -r^3(\cos \gamma + i \sin \gamma),$$

$$y_2^3 = -\frac{q}{2} - i\sqrt{-\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3} = -r^3(\cos \gamma - i \sin \gamma),$$

где

$$r = \frac{q}{|q|} \sqrt{\frac{|p|}{3}}, \quad \gamma = \arccos \frac{q}{2r^3}. \quad (8.8.16)$$

Знак r должен совпадать со знаком q . После подстановки данных формул в (8.8.11) получим $a = -r \cos(\gamma/3)$, $b = -r \sin(\gamma/3)$ и

$$y_1 = -r \left(\cos \frac{\gamma}{3} + i \sin \frac{\gamma}{3} \right), \quad y_2 = -r \left(\cos \frac{\gamma}{3} - i \sin \frac{\gamma}{3} \right).$$

Тогда три действительных корня неполного кубического уравнения выразятся через его коэффициенты формулами

$$x_1 = -2r \cos \frac{\gamma}{3}, \quad (8.8.17)$$

$$x_2 = r \left(\cos \frac{\gamma}{3} + \sqrt{3} \sin \frac{\gamma}{3} \right) = 2r \cos \left(\frac{\pi}{3} - \frac{\gamma}{3} \right), \quad (8.8.18)$$

$$x_3 = r \left(\cos \frac{\gamma}{3} - \sqrt{3} \sin \frac{\gamma}{3} \right) = 2r \cos \left(\frac{\pi}{3} + \frac{\gamma}{3} \right). \quad (8.8.19)$$

Разыскание корней кубического уравнения с действительными коэффициентами по формуле Кардано требует извлечение кубических корней из комплексных чисел, что можно делать путем перехода к тригонометрической записи комплексных чисел. Запись корней (8.8.11) с помощью радикалов следует рассматривать именно с таких позиций.

Корни характеристического уравнения (3.3.8) и (8.7.11) определяются формулами

$$\lambda_1 = \frac{I_1}{3} - 2r \cos \frac{\gamma}{3}, \quad (8.8.20)$$

$$\lambda_2 = \frac{I_1}{3} + 2r \cos \left(\frac{\pi}{3} - \frac{\gamma}{3} \right), \quad (8.8.21)$$

$$\lambda_3 = \frac{I_1}{3} + 2r \cos \left(\frac{\pi}{3} + \frac{\gamma}{3} \right), \quad (8.8.22)$$

где

$$p = -\frac{1}{3}I_1^2 + I_2, \quad q = -\frac{2}{27}I_1^3 + \frac{1}{3}I_1I_2 - I_3, \quad r = \frac{q}{|q|} \sqrt{\frac{|p|}{3}}, \quad \gamma = \arccos \frac{q}{2r^3}.$$

Возможны частные случаи.

Случай 1: $p = 0$, тогда $\lambda_1 = \lambda_2 = \lambda_3 = \frac{I_1}{3}$.

Случай 2: $D = 0$, тогда $\lambda_1 = \frac{I_1}{3} - 2\sqrt[3]{\frac{q}{2}}$, $\lambda_2 = \lambda_3 = \frac{I_1}{3} + \sqrt[3]{\frac{q}{2}}$.

8.9.* Квадратурные формулы

Выше мы показали, как с помощью криволинейных интегралов вычисляются геометрические характеристики линий и плоских сечений. Рассмотрим некоторые методы приближенного вычисления определенных интегралов.

Квадратурная формула Гаусса. Пусть требуется приближенно вычислить определенный интеграл от некоторой функции $f(t)$. Если эта функция однозначна и непрерывна на интервале интегрирования, тогда по теореме Вейерштрасса для любого положительного достаточно малого ε существует такой полином $p(t)$, что $|f(t) - p(t)| < \varepsilon$ всюду на

заданном интервале. Если бы нам была известна формула точного вычисления определенного интеграла от полинома $p(t)$, то мы могли бы с известной точностью принять интеграл от полинома за приближенное значение интеграла от заданной функции $f(t)$. Поставим задачу найти значения аргументов t_1, t_2, \dots, t_n и весовые множители w_1, w_2, \dots, w_n , чтобы формула

$$\int_{-1}^1 f(t) dt \approx \int_{-1}^1 p(t) dt = \sum_{i=1}^n w_i p(t_i) \quad (8.9.1)$$

была бы точной для всех полиномов $p(t)$ степени $2n - 1$. Полином степени $2n - 1$ содержит $2n$ коэффициентов, по которым можно найти n аргументов t_1, t_2, \dots, t_n и n весовых множителей w_1, w_2, \dots, w_n . К пределам интегрирования от -1 до 1 всегда можно свести любые пределы путем линейной замены аргумента.

Пусть

$$p(t) = C_0 + C_1 t + C_2 t^2 + \dots + C_{2n-1} t^{2n-1}. \quad (8.9.2)$$

В соответствии с формулой (8.9.1) должны выполняться равенства

$$\int_{-1}^1 t^m dt = \sum_{i=1}^n w_i t_i^m, \quad m = 0, 1, \dots, 2n - 1. \quad (8.9.3)$$

В то же время

$$\int_{-1}^1 t^m dt = \frac{2}{m+1}, \quad \text{если } m \text{ четное, и } 0, \text{ если } m \text{ нечетное.} \quad (8.9.4)$$

Подставим (8.9.2) в (8.9.1) и получим

$$\begin{aligned} C_0 \int_{-1}^1 dt + C_1 \int_{-1}^1 t dt + C_2 \int_{-1}^1 t^2 dt + \dots + C_{2n-1} \int_{-1}^1 t^{2n-1} dt = \\ = C_0 \sum_{i=1}^n w_i + C_1 \sum_{i=1}^n w_i t_i + C_2 \sum_{i=1}^n w_i t_i^2 + \dots + C_{2n-1} \sum_{i=1}^n w_i t_i^{2n-1}. \end{aligned} \quad (8.9.5)$$

Приравняем выражения при $C_i, i = 0, 1, 2, \dots, 2n - 1$ в правой и левой частях равенства (8.9.5) и, учитывая равенства (8.9.4), получим систему $2n$ уравнений

$$\begin{aligned} w_1 + w_2 + \dots + w_n &= 2, \\ w_1 t_1 + w_2 t_2 + \dots + w_n t_n &= 0, \\ w_1 t_1^2 + w_2 t_2^2 + \dots + w_n t_n^2 &= \frac{2}{3}, \\ \dots \dots \dots \end{aligned} \quad (8.9.6)$$

$$w_1 t_1^{2n-2} + w_2 t_2^{2n-2} + \dots + w_n t_n^{2n-2} = \frac{2}{2n-1},$$

$$w_1 t_1^{2n-1} + w_2 t_2^{2n-1} + \dots + w_n t_n^{2n-1} = 0.$$

Из этой системы уравнений могут быть найдены значения аргументов t_1, t_2, \dots, t_n и весовые множители w_1, w_2, \dots, w_n . Система является нелинейной и может быть решена численным методом.

Если в качестве полинома вместо (8.9.2) взять полиномы

$$p(t) = t^m P_n(t), \quad m = 0, 1, 2, \dots, n-1, \quad (8.9.7)$$

где $P_n(t)$ — полиномы Лежандра, то окажется, что в качестве значений t_1, t_2, \dots, t_n достаточно взять решения уравнения $P_n(t) = 0$. Полиномы Лежандра определяются формулой Родрига

$$P_n(t) = \frac{1}{2^n n!} \frac{d^n (t^2 - 1)^n}{dt^n} \quad (8.9.8)$$

или формулой

$$P_n(t) = \frac{1}{2^n} \sum_{m=0}^{[n/2]} \frac{(-1)^m (2n-2m)!}{m!(n-m)!(n-2m)!} t^{n-2m},$$

где $[n/2] = n/2$ для четных n и $[n/2] = (n-1)/2$ для нечетных n . Первые восемь полиномов Лежандра имеют вид

$$\begin{aligned} P_0(t) &= 1, & P_1(t) &= t, & P_2(t) &= \frac{3t^2 - 1}{2}, & P_3(t) &= \frac{5t^3 - 3t}{2}, \\ P_4(t) &= \frac{35t^4 - 30t^2 + 3}{8}, & P_5(t) &= \frac{63t^5 - 70t^3 + 15t}{8}, \\ P_6(t) &= \frac{231t^6 - 315t^4 + 105t^2 - 5}{16}, \\ P_7(t) &= \frac{429t^7 - 693t^5 + 315t^3 - 35t}{16}. \end{aligned}$$

Полиномы Лежандра удовлетворяют уравнению Лежандра

$$(1-t^2) \frac{d^2 y}{dt^2} - 2t \frac{dy}{dt} + n(n+1) = 0,$$

которое появляется при решении уравнения Лапласа в сферических координатах методом разделения переменных.

Полиномы Лежандра обладают следующими свойствами.

$$\begin{aligned} P_n(-t) &= (-1)^n P_n(t), & P_n(1) &= 1, & P_n(-1) &= (-1)^n, \\ (n+1)P_{n+1}(t) &= (2n+1)tP_n(t) - nP_{n-1}(t), \\ P_{n+1}'(t) - P_{n-1}'(t) &= (2n+1)P_n(t), \end{aligned}$$

$$\int_{-1}^1 P_n(t)p(t) dt = 0 \quad (8.9.9)$$

для любых полиномов $p(t)$ степени $m = 0, 1, \dots, n-1$. Полином Лежандра $P_n(t)$ имеет n различных действительных корней t_1, t_2, \dots, t_n , которые расположены на интервале от -1 до 1 .

Полиномы (8.9.7) имеют степень, не превышающую $2n-1$, поэтому для них должно выполняться равенство (8.9.1), т. е.

$$\int_{-1}^1 t^m P_n(t) dt = \sum_{i=1}^n w_i t_i^m P_n(t_i), \quad m = 0, 1, \dots, n-1. \quad (8.9.10)$$

В силу свойства (8.9.9) полиномов Лежандра равенство (8.9.10) примет вид

$$\sum_{i=1}^n w_i t_i^m P_n(t_i) = 0, \quad m = 0, 1, \dots, n-1. \quad (8.9.11)$$

Эти равенства будут выполнены при любых w_i , $i = 1, 2, \dots, n$, если положить

$$P_n(t_i) = 0, \quad i = 1, 2, \dots, n. \quad (8.9.12)$$

То есть в качестве аргументов t_1, t_2, \dots, t_n могут быть взяты нули соответствующего полинома Лежандра. Эти нули действительны, различны и расположены в интервале $-1 \leq t_i \leq 1$. По известным аргументам t_1, t_2, \dots, t_n из первых n уравнений (8.9.6) могут быть найдены весовые множители w_1, w_2, \dots, w_n . Эта система уравнений линейная относительно весовых множителей и ее определитель является определителем Вандермонда

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_n \\ \dots & \dots & \dots & \dots \\ t_1^n & t_2^n & \dots & t_n^n \end{vmatrix},$$

который в силу различности корней полинома Лежандра не равен нулю. Формула (8.9.1), где t_1, t_2, \dots, t_n — нули полиномов Лежандра, называется *квадратурной формулой Гаусса*.

Рассмотрим квадратурные формулы Гаусса для интеграла

$$I = \int_a^b f(x) dx. \quad (8.9.13)$$

Путем замены переменной $x = (a+b)/2 + t(b-a)/2$ придем к интегралу по переменной t с пределами интегрирования от -1 до 1 , для которого применим формулу Гаусса

$$I = \int_a^b f(x) dx = \frac{1}{2}(b-a) \int_{-1}^1 f(x(t)) dt \approx \frac{1}{2}(b-a) \sum_{i=1}^n w_i f(x_i), \quad (8.9.14)$$

где $x_i = (a+b)/2 + t_i(b-a)/2$ и t_i , $i = 1, 2, \dots, n$ — корни полинома Лежандра $P_n(t)$. Путем разложения в ряд интегрируемой функции можно показать, что погрешность

квадратурной формулы Гаусса равна

$$R_n = \frac{(b-a)^{2n+1} (n!)^4 d^{2n} f(\xi) / dx^{2n}}{((2n)!)^3 (2n+1)}, \quad \text{где } a < \xi < b. \quad (8.9.15)$$

В табл. 8.9.1 приведены значения аргументов t_1, t_2, \dots, t_n и весовые множители w_1, w_2, \dots, w_n для квадратурных формул Гаусса на интервале $-1 \leq t \leq 1$.

Таблица 8.9.1

n	t_i	w_i
1	0	2
2	$\pm 0,57735026918962576450$	1
3	0	8/9
	$\pm 0,77459666924148337703$	5/9
4	$\pm 0,33998104358485626480$	0,65214515486254614262
	$\pm 0,86113631159405257522$	0,34785484513745385737
5	0	0,56888888888888888889
	$\pm 0,53846931010568309103$	0,47862867049936646804
	$\pm 0,90617984593866399279$	0,23692688505618908751
6	$\pm 0,23861918608319690863$	0,46791393457269104738
	$\pm 0,66120938646626451366$	0,36076157304813860756
	$\pm 0,93246951420315202781$	0,17132449237917034504
7	0	0,41795918367346938775
	$\pm 0,40584515137739716690$	0,38183005050511894495
	$\pm 0,74153118559939443986$	0,27970539148927666790
	$\pm 0,94910791234275852452$	0,12948496616886969327
8	$\pm 0,18343464249564980493$	0,36268378337836198296
	$\pm 0,52553240991632898581$	0,31370664587788728733
	$\pm 0,79666647741362673959$	0,22238103445337447054
	$\pm 0,96028985649753623168$	0,10122853629037625915

Квадратурные формулы Ньютона–Котеса. Если, на участке интегрирования функция может быть аппроксимирована полиномом, то за приближенное значение интеграла может быть принят интеграл от этого полинома. В некоторых случаях точки аппроксимации отстоят друг от друга на равном расстоянии по параметру, а в качестве интерполяционного полинома используется полином Лагранжа (2.4.15).

Пусть требуется вычислить интеграл

$$I = \int_a^b f(t) dt. \quad (8.9.16)$$

Разобьем отрезок интегрирования равноотстоящими точками

$$t_0 = a, \quad t_i = t_0 + ih, \quad i = 1, 2, \dots, n-1, \quad t_n = b$$

на n равных участках, где $h = (b - a)/n$. Заменяем функцию $f(t)$ интерполяционным полиномом Лагранжа

$$\Lambda_n(t) = \sum_{i=0}^n L_i(t) f(t_i), \quad (8.9.17)$$

где

$$L_i(t) = \frac{(t - t_0)(t - t_1) \dots (t - t_{i-1})(t - t_{i+1}) \dots (t - t_n)}{(t_i - t_0)(t_i - t_1) \dots (t_i - t_{i-1})(t_i - t_{i+1}) \dots (t_i - t_n)}.$$

При введении обозначения $q = n \frac{t - t_0}{b - a} = \frac{t - t_0}{h}$ полином Лагранжа примет вид

$$\Lambda_n(t) = \sum_{i=0}^n \frac{(-1)^{n-i} q(q-1) \dots (q-n)}{i!(n-1)!(q-i)} f(t_i). \quad (8.9.18)$$

За приближенное значение интеграла (8.9.16) примем интеграл от интерполяционного полинома Лагранжа

$$I = \int_a^b f(t) dt \approx \int_{t_0}^{t_n} \Lambda_n(t) dt = \sum_{i=0}^n a_i f(t_i), \quad (8.9.19)$$

где введено обозначение

$$a_i = \int_{t_0}^{t_n} \frac{(-1)^{n-i} q(q-1) \dots (q-n)}{i!(n-1)!(q-i)} dt = \frac{(-1)^{n-i} (b-a)}{i!(n-1)!n} \int_0^n \frac{q(q-1) \dots (q-n)}{q-i} dq. \quad (8.9.20)$$

Обычно полагают $a_i = (b-a)H_i$, где постоянные величины

$$H_i = \frac{(-1)^{n-i}}{i!(n-1)!n} \int_0^n \frac{q(q-1) \dots (q-n)}{q-i} dq, \quad (8.9.21)$$

называются *коэффициентами Котеса*.

Полученные таким образом формулы вычисления определенных интегралов называются *квадратурными формулами Ньютона-Котеса* и имеют вид

$$\int_{t_0}^{t_n} f(t) dt \approx (b-a) \sum_{i=0}^n H_i f(t_i), \quad (8.9.22)$$

где $t_i = a + i \frac{b-a}{n}$.

Частные случаи формул Ньютона-Котеса. Положим в формуле (8.9.22) $n = 1$ и получим *формулу трапеций* для приближенного вычисления определенного интеграла

$$\int_a^b f(t) dt \approx \frac{b-a}{2} (f(a) + f(b)). \quad (8.9.23)$$

Можно показать, что погрешность формулы трапеций равна

$$R_1 = \frac{h^3}{12} \frac{d^2 f(\xi)}{dx^2}, \quad \text{где } a < \xi < b. \quad (8.9.24)$$

Если на участке интегрирования $d^2 f/dx^2 > 0$, формула трапеций дает значение интеграла и избытком, а если на участке интегрирования $d^2 f/dx^2 < 0$, формула трапеций дает значение интеграла с недостатком. Для более точного вычисления область интегрирования разбивают на несколько участков и на каждом из них применяется формула трапеций. Это равносильно аппроксимации кривой, описываемой данной функцией, ломаной линией, и вычислению интеграла от этой ломаной.

Квадратурная формула, в которой интегрируемая функция заменяется параболой, называется *формулой Симпсона*. Формула Симпсона может быть получена как частный случай формулы (8.9.22) при $n = 2$. Из формулы (8.9.22) при $n = 2$ получим $H_0 = 1/6$, $H_1 = 2/3$, $H_2 = 1/6$ и

$$\int_a^b f(t) dt \approx \frac{b-a}{6} (f(t_0) + 4f(t_1) + f(t_2)), \quad (8.9.25)$$

где $t_i = t_0 + i \frac{b-a}{2}$, $i = 0, 1, 2$. Можно показать, что погрешность формулы Симпсона равна

$$R_2 = -\frac{h^5}{90} \frac{d^4 f(\xi)}{dx^4}, \quad \text{где } a < \xi < b. \quad (8.9.26)$$

Формула Симпсона является точной для полиномов не только второй, но и третьей степени.

Из формулы (8.9.22) при $n = 3$ получим *квадратурную формулу Ньютона*

$$\int_a^b f(t) dt \approx \frac{b-a}{8} (f(t_0) + 3f(t_1) + 3f(t_2) + f(t_3)), \quad (8.9.27)$$

где $t_i = t_0 + i \frac{b-a}{3}$, $i = 0, 1, 2, 3$. Можно показать, что погрешность формулы Ньютона равна

$$R_3 = -3 \frac{h^5}{80} \frac{d^4 f(\xi)}{dx^4}, \quad \text{где } a < \xi < b. \quad (8.9.28)$$

Формула Ньютона при одинаковом шаге $h = (b-a)/n$ менее точна, чем формула Симпсона. С точки зрения точности формулы Ньютона–Котеса с четным n являются более эффективными.

В общем случае область интегрирования может разбиваться на несколько участков и на каждом участке использоваться интегрирование одним из описанных методов. Полученные таким способом формулы называются обобщенными формулами Ньютона–Котеса и обобщенными формулами Симпсона.

Рассмотренные квадратурные формулы используются для вычисления геометрических характеристик кривых линий и плоских областей, ограниченных линиями на плоскости. Из всех приведенных формул при одном и том же числе точек n квадратурная формула Гаусса имеет наименьшую погрешность при тех же вычислительных затратах.

8.10.* Кубатурные формулы

При вычислении геометрических характеристик тел мы свели объемные интегралы к поверхностным интегралам. Рассмотрим методы приближенного вычисления двойных интегралов по параметрическим областям поверхностей. В некоторых случаях двойной интеграл можно вычислить сразу по всей области, но в большинстве случаев области определения поверхности мы будем разбивать на совокупность небольших подобластей четырехугольной или треугольной формы и вычислять поверхностный интеграл как сумму интегралов по областям простейших форм. Формулы приближенного вычисления двойных интегралов называются кубатурными формулами.

Кубатурная формула Гаусса. Пусть требуется приближенно вычислить двойной интеграл от некоторой непрерывной и однозначной функции $f(u, v)$ на области Ω

$$I = \iint_{\Omega} f(u, v) du dv. \quad (8.10.1)$$

Пусть область интегрирования можно ограничить двумя непрерывными и однозначными кривыми

$$v = \varphi(u), \quad v = \psi(u), \quad \varphi(u) \leq \psi(u), \quad a \leq u \leq b,$$

как показано на рис. 8.10.1.

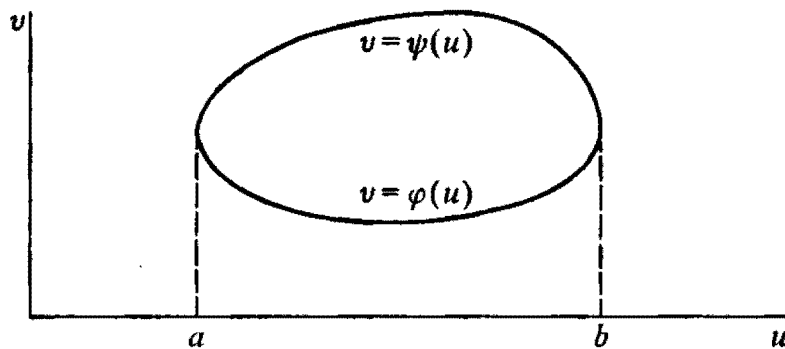


Рис. 8.10.1. Область интегрирования

Двойной интеграл (8.10.1) можно свести к повторному интегралу:

$$I = \iint_{\Omega} f(u, v) du dv = \int_a^b \left(\int_{\varphi(u)}^{\psi(u)} f(u, v) dv \right) du = \int_a^b F(u) du, \quad (8.10.2)$$

где

$$F(u) = \int_{\varphi(u)}^{\psi(u)} f(u, v) dv. \quad (8.10.3)$$

Применим к последнему интегралу квадратурную формулу и получим

$$I = \iint_{\Omega} f(u, v) du dv \approx \sum_{i=1}^m a_i F(u_i), \quad (8.10.4)$$

где a_i — весовые множители используемой квадратурной формулы. В свою очередь значения $F(u_i)$ так же могут быть найдены с помощью квадратурных формул

$$F(u_i) = \int_{\varphi(u_i)}^{\psi(u_i)} f(u_i, v) dv \approx \sum_{j=1}^n b_j f(u_i, v_j).$$

Таким образом, двойной интеграл (8.10.15) приближенно определится формулой

$$I = \iint_{\Omega} f(u, v) du dv \approx \sum_{i=1}^m a_i \sum_{j=1}^n b_j f(u_i, v_j) = \sum_{i=1}^m \sum_{j=1}^n w_{ij} f(u_i, v_j), \quad (8.10.5)$$

где $w_{ij} = a_i b_j$ — постоянные коэффициенты, вычисленные по весовым множителям. В случае использования квадратурных формул Гаусса (8.9.14) кубатурная формула (8.10.5) примет вид

$$I = \iint_{\Omega} f(u, v) du dv = \int_a^b \left(\int_{\varphi(u)}^{\psi(u)} f(u, v) dv \right) du \approx \approx \frac{b-a}{2} \sum_{i=1}^m w_i \frac{\psi(u_i) - \varphi(u_i)}{2} \sum_{j=1}^n z_j f(u_i, v_j), \quad (8.10.6)$$

где $u_i = \frac{a+b}{2} + x_i \frac{b-a}{2}$, $v_j = \frac{\varphi(u_i) + \psi(u_i)}{2} + y_j \frac{\psi(u_i) - \varphi(u_i)}{2}$ и x_i , $i = 1, 2, \dots, m$ — параметры корней полинома Лежандра $P_m(x)$, w_1, w_2, \dots, w_m — соответствующие им весовые множители, y_j , $j = 1, 2, \dots, n$ — параметры корней полинома Лежандра $P_n(y)$, z_1, z_2, \dots, z_n — соответствующие им весовые множители.

Кубатурная формула Гаусса на четырехугольной области. Пусть при интегрировании по поверхности общего вида мы разбили ее область определения параметров на выпуклые четырехугольники. Рассмотрим вычисление двойного интеграла на четырехугольной области с использованием квадратурных формул Гаусса. Пусть вершинам

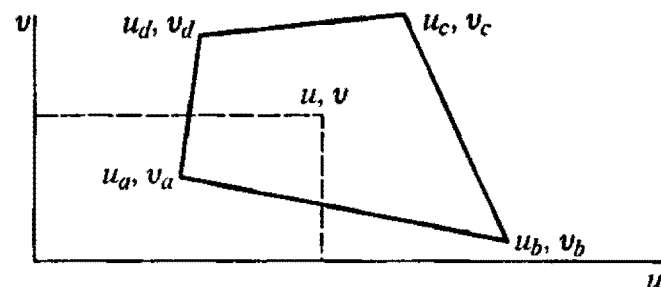


Рис. 8.10.2. Четырехугольная область интегрирования

четырехугольника соответствуют параметры $u_a, v_a, u_b, v_b, u_c, v_c, u_d, v_d$, которые объединим в точки $\mathbf{p}_a = [u_a \ v_a]^T$, $\mathbf{p}_b = [u_b \ v_b]^T$, $\mathbf{p}_c = [u_c \ v_c]^T$, $\mathbf{p}_d = [u_d \ v_d]^T$ на плоскости параметров (рис. 8.10.2).

Пусть требуется вычислить определенный интеграл

$$I = \iint_{\diamond} f(u, v) du dv \quad (8.10.7)$$

на данной четырехугольной области. Если область интегрирования представляет собой квадрат со стороной 2, то кубатурная формула Гаусса (8.10.6) примет вид

$$I = \int_{-1}^1 \int_{-1}^1 f(u, v) dv du \approx \sum_{i=1}^m \sum_{j=1}^n w_i z_j f(u_i, v_j), \quad (8.10.8)$$

где u_1, u_2, \dots, u_m — корни полинома Лежандра $P_m(u)$, w_1, w_2, \dots, w_m — соответствующие им весовые множители, v_1, v_2, \dots, v_n — корни полинома Лежандра $P_n(v)$, z_1, z_2, \dots, z_n — соответствующие им весовые множители. Заменяем переменные в двойном интеграле (8.10.7). Для произвольной выпуклой четырехугольной области введем новые параметры x и y , связанные с параметрами u и v соотношениями

$$\begin{aligned} u(x, y) &= u_a \frac{1-x-y+xy}{4} + u_b \frac{1+x-y-xy}{4} + u_c \frac{1+x+y+xy}{4} + u_d \frac{1-x+y-xy}{4}, \\ v(x, y) &= v_a \frac{1-x-y+xy}{4} + v_b \frac{1+x-y-xy}{4} + v_c \frac{1+x+y+xy}{4} + v_d \frac{1-x+y-xy}{4}. \end{aligned} \quad (8.10.9)$$

Точкам p_a, p_b, p_c, p_d соответствуют параметры $x_a = -1, y_a = -1; x_b = 1, y_b = -1; x_c = 1, y_c = 1; x_d = -1, y_d = 1$. Функции (8.10.9) являются непрерывными, имеют непрерывные частные производные первого порядка и отображают квадрат с вершинами в точках $x = \pm 1, y = \pm 1$ на заданную четырехугольную область интегрирования. Это отображение взаимно однозначно и имеет отличный от нуля якобиан

$$\begin{aligned} J(x, y) &= \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \\ &= \frac{1}{16} (-u_a + u_b + u_c - u_d + (u_a - u_b + u_c - u_d)y) (-v_a - v_b + v_c + v_d + (v_a - v_b + v_c - v_d)x) - \\ &- \frac{1}{16} (-u_a - u_b + u_c + u_d + (u_a - u_b + u_c - u_d)x) (-v_a + v_b + v_c - v_d + (v_a - v_b + v_c - v_d)y) = \\ &= \frac{1}{8} ((u_c - u_a)(v_d - v_b) + (u_b - u_a)(v_c - v_d)x + (u_c - u_b)(v_d - v_a)y - \\ &- (u_d - u_b)(v_c - v_a) - (u_c - u_d)(v_b - v_a)x - (u_d - u_a)(v_c - v_b)y). \end{aligned} \quad (8.10.10) \end{aligned}$$

После замены параметров в интеграле (8.10.7) получим

$$\begin{aligned} \iint_{\diamond} f(u, v) du dv &= \int_{-1}^1 \int_{-1}^1 f(u(x, y), v(x, y)) J(x, y) dx dy \approx \\ &\approx \sum_{i=1}^m \sum_{j=1}^n w_i z_j f(u(x_i, y_j), v(x_i, y_j)) J(x_i, y_j), \end{aligned} \quad (8.10.11)$$

где x_1, x_2, \dots, x_m — корни полинома Лежандра $P_m(x)$, w_1, w_2, \dots, w_m — соответствующие им весовые множители, y_1, y_2, \dots, y_n — корни полинома Лежандра $P_n(y)$, z_1, z_2, \dots, z_n — соответствующие им весовые множители.

Кубатурные формулы на треугольной области. Пусть мы разбили область определения параметров поверхности на треугольники. Покрытие некоторой области треугольниками называется *триангуляцией*. Рассмотрим вычисление интеграла (8.10.7) по треугольной области параметров поверхности u и v для одного такого треугольника.

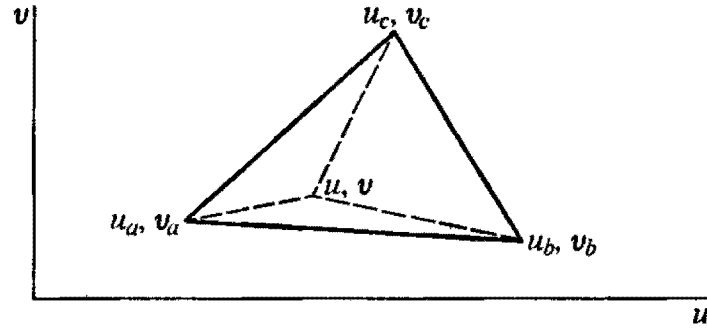


Рис. 8.10.3. Треугольная область интегрирования

Пусть его вершинам соответствуют параметры $u_a, v_a, u_b, v_b, u_c, v_c$, которые объединим в точки $p_a = [u_a \ v_a]^T$, $p_b = [u_b \ v_b]^T$, $p_c = [u_c \ v_c]^T$ на плоскости параметров (рис. 8.10.3).

Площадь S треугольника $p_a p_b p_c$ равна половине определителя (3.11.5)

$$S = \frac{1}{2} \begin{vmatrix} 1 & 1 & 1 \\ u_a & u_b & u_c \\ v_a & v_b & v_c \end{vmatrix} = \frac{1}{2} (u_a v_b + u_b v_c + u_c v_a - v_a u_b - v_b u_c - v_c u_a).$$

Для треугольных областей удобно перейти от координат u и v к трем барицентрическим координатам a, b, c , определяемым равенствами (3.11.4). Барицентрические координаты произвольной точки $p = [u \ v]^T$ равны площади треугольника, полученного из p_a, p_b, p_c заменой соответствующей вершины точкой p , деленной на площадь S . Барицентрические координаты в сумме равны единице. Обратный переход от барицентрических координат треугольника к параметрам u и v осуществляется по формулам

$$\begin{aligned} u &= a u_a + b u_b + c u_c, \\ v &= a v_a + b v_b + c v_c. \end{aligned} \quad (8.10.12)$$

Таким образом, интегрируемую функцию $f(u, v)$ можно представить в виде функции барицентрических координат $f(a, b, c) \equiv f(u(a, b, c), v(a, b, c))$.

Поставим задачу найти значения барицентрических координат a_i, b_i, c_i и весовые множители w_i , $i = 1, 3, 4, \dots, 3n, 3n + 1$, чтобы формула

$$I = \iint_{\Delta} f(u, v) du dv = S \sum_i w_i f(u(a_i, b_i, c_i), v(a_i, b_i, c_i)) \quad (8.10.13)$$

была бы точной для некоторых полиномов $p(u(a, b, c), v(a, b, c))$ на треугольной области. Из трех барицентрических координат только две являются независимыми. В качестве независимых параметров, не теряя общности, выберем первые две барицентрические координаты. Третья координата связана с двумя другими равенством (3.11.2). Используя равенства (8.10.12) в форме

$$\begin{aligned} u &= a(u_a - u_c) + b(u_b - u_c) + u_c, \\ v &= a(v_a - v_c) + b(v_b - v_c) + v_c \end{aligned} \quad (8.10.14)$$

перепишем интеграл (8.9.15) в виде

$$\begin{aligned} \iint_{\Delta} f(u, v) du dv &= \iint_{\Delta} f(u(a, b, c), v(a, b, c)) J(a, b, c) da db = \\ &= \int_0^1 \left(\int_0^{1-a} f(u(a, b), v(a, b)) J(a, b) db \right) da, \end{aligned} \quad (8.10.15)$$

где интегрирование проводится по треугольной области, а $J(a, b)$ — якобиан преобразования координат

$$\begin{aligned} J(a, b) &= \begin{vmatrix} \frac{\partial u}{\partial a} & \frac{\partial u}{\partial b} \\ \frac{\partial v}{\partial a} & \frac{\partial v}{\partial b} \end{vmatrix} = \begin{vmatrix} u_a - u_c & u_b - u_c \\ v_a - v_c & v_b - v_c \end{vmatrix} = (u_a - u_c)(v_b - v_c) - (v_a - v_c)(u_b - u_c) = \\ &= u_a v_b + u_b v_c + u_c v_a - v_a u_b - v_b u_c - v_c u_a = 2S. \end{aligned}$$

Потребуем, чтобы барицентрические координаты a_i, b_i, c_i в (8.10.13) обладали некоторой треугольной симметрией, например, чтобы две из трех координат были бы равны между собой. При такой симметрии число точек, по которым производится суммирование, должно быть кратно трем. Для определения параметров трех симметрично расположенных точек в сумме (8.10.13) нужно знать одну из барицентрических координат. Все три барицентрические координаты трех точек будут определены, если найдена одна из них. Пусть одна из барицентрических координат в правой части (8.10.13) равна z , тогда три симметрично расположенные точки будут иметь координаты

$$\begin{aligned} a_1 &= z, & b_1 &= \frac{1-z}{2}, & c_1 &= \frac{1-z}{2}, \\ a_2 &= \frac{1-z}{2}, & b_2 &= z, & c_2 &= \frac{1-z}{2}, \\ a_3 &= \frac{1-z}{2}, & b_3 &= \frac{1-z}{2}, & c_3 &= z. \end{aligned} \quad (8.10.16)$$

В силу симметрии все три точки должны иметь одинаковые весовые множители $w_1 = w_2 = w_3$. В случае, если $z = 1/3$, три точки сливаются в одну. Таким образом, на треугольнике в правой части (8.10.13) могут быть использованы следующие количества точек: $1, 4, \dots, 3(n-1) + 1$ или $3, 6, \dots, 3n$. Для определения барицентрических координат и весовых множителей требуется соответствующее число уравнений: $1, 3, \dots, 2n-1$ или $2, 4, \dots, 2n$. Потребуем, чтобы выполнялось равенство (8.10.13) для всех функций вида

$$f(u(a, b, c), v(a, b, c)) = \sum_{m=1}^n (A_m a^m + B_m b^m + C_m c^m), \quad (8.10.17)$$

где A_m, B_m, C_m — коэффициенты полиномиального представления заданной функции, и найдем независимые барицентрические координаты и весовые множители. При $n > 2$ функция вида (8.10.17) не представляет собой полную функцию степени n двух переменных u и v , но обладает треугольной симметричной на заданной области. Для функций $f(u, v) = f(u(a, b, c), v(a, b, c)) = f(a, b)$ вида

$$f(a, b) = a^m, \quad f(a, b) = b^m, \quad f(a, b) = c^m = (1-a-b)^m, \quad m = 1, 2, \dots, n,$$

в соответствии с формулой (8.10.13) и (8.10.15) должны выполняться равенства

$$2S \int_0^1 \left(\int_0^{1-a} f(a, b) db \right) da = S \sum_{i=1}^k w_i f(a_i, b_i)$$

или

$$\int_0^1 \left(\int_0^{1-a} a^m db \right) da = \frac{1}{2} \sum_{i=1}^k w_i a_i^m, \quad m = 1, 2, \dots, n, \quad (8.10.18)$$

$$\int_0^1 \left(\int_0^{1-a} b^m db \right) da = \frac{1}{2} \sum_{i=1}^k w_i b_i^m, \quad m = 1, 2, \dots, n, \quad (8.10.19)$$

$$\int_0^1 \left(\int_0^{1-a} c^m db \right) da = \frac{1}{2} \sum_{i=1}^k w_i c_i^m, \quad m = 1, 2, \dots, n, \quad (8.10.20)$$

где $k = \frac{3}{2}(n-1) + 1$ при нечетных n и $k = \frac{3}{2}n$ при четных n .

В то же время путем интегрирования по треугольной области получим

$$\int_0^1 \left(\int_0^{1-a} a^m db \right) da = \frac{1}{(m+1)(m+2)}, \quad (8.10.21)$$

$$\int_0^1 \left(\int_0^{1-a} b^m db \right) da = \frac{1}{(m+1)(m+2)}, \quad (8.10.22)$$

$$\int_0^1 \left(\int_0^{1-a} c^m db \right) da = \frac{1}{(m+1)(m+2)}. \quad (8.10.23)$$

Подставим (8.10.18)–(8.10.20) в (8.10.13) и получим

$$\begin{aligned} & \iint_{\Delta} (A_1 a + B_1 b + C_1 c) da db + \iint_{\Delta} (A_2 a^2 + B_2 b^2 + C_2 c^2) da db + \dots \\ & \dots + \iint_{\Delta} (A_n a^n + B_n b^n + C_n c^n) da db = \frac{1}{2} \left(A_1 \sum_{i=1}^k w_i a_i + B_1 \sum_{i=1}^k w_i b_i + C_1 \sum_{i=1}^k w_i c_i \right) + \\ & + \frac{1}{2} \left(A_2 \sum_{i=1}^k w_i a_i^2 + B_2 \sum_{i=1}^k w_i b_i^2 + C_2 \sum_{i=1}^k w_i c_i^2 \right) + \dots \\ & \dots + \frac{1}{2} \left(A_n \sum_{i=1}^k w_i a_i^n + B_n \sum_{i=1}^k w_i b_i^n + C_n \sum_{i=1}^k w_i c_i^n \right). \quad (8.10.24) \end{aligned}$$

Для нахождения барицентрических и весовых множителей приравняем выражения при $A_i, B_i, C_i, i = 1, 2, \dots, n$ в правой и левой частях равенства (8.10.24). В силу треугольной симметрии это достаточно сделать для набора коэффициентов при одной из барицентрических координат. Например, из равенства выражений при $A_i, i = 1, 2, \dots, n$ в правой и левой частях равенства (8.10.24) с учетом равенств (8.10.21)–(8.10.23) получим систему n уравнений

$$\begin{aligned} w_1 a_1 + w_2 a_2 + w_3 a_3 + \dots + w_k a_k &= \frac{1}{3}, \\ w_1 a_1^2 + w_2 a_2^2 + w_3 a_3^2 + \dots + w_k a_k^2 &= \frac{1}{6}, \\ w_1 a_1^3 + w_2 a_2^3 + w_3 a_3^3 + \dots + w_k a_k^3 &= \frac{1}{10}, \\ &\dots \dots \dots \\ w_1 a_1^n + w_2 a_2^n + w_3 a_3^n + \dots + w_k a_k^n &= \frac{2}{(n+1)(n+2)}, \end{aligned}$$

где $k = \frac{3}{2}(n-1) + 1$ при n нечетном и $k = \frac{3}{2}n$ при n четном. При n нечетном барицентрические координаты одной из точек известны, так как она совпадает с центром треугольника. При $n = 1$ получим одно уравнение для веса центральной точки и будем иметь:

$$a_0 = \frac{1}{3}, \quad b_0 = \frac{1}{3}, \quad c_0 = \frac{1}{3}, \quad w_0 = 1.$$

При $n = 2$ будем иметь два уравнения

$$\begin{aligned} wz + w \frac{1-z}{2} + w \frac{1-z}{2} &= \frac{1}{3}, \\ wz^2 + w \left(\frac{1-z}{2} \right)^2 + w \left(\frac{1-z}{2} \right)^2 &= \frac{1}{6}, \end{aligned}$$

из которых с учетом (8.10.16) получим три симметрично расположенные точки на треугольной области:

$$\begin{aligned} a_1 = z = \frac{2}{3}, & \quad b_1 = \frac{1-z}{2} = \frac{1}{6}, & \quad c_1 = \frac{1-z}{2} = \frac{1}{6}, & \quad w_1 = w = \frac{1}{3}, \\ a_2 = \frac{1-z}{2} = \frac{1}{6}, & \quad b_2 = z = \frac{2}{3}, & \quad c_2 = \frac{1-z}{2} = \frac{1}{6}, & \quad w_2 = w = \frac{1}{3}, \\ a_3 = \frac{1-z}{2} = \frac{1}{6}, & \quad b_3 = \frac{1-z}{2} = \frac{1}{6}, & \quad c_3 = z = \frac{2}{3}, & \quad w_3 = w = \frac{1}{3}, \end{aligned}$$

или три равноценные им точки

$$\begin{aligned} a_1 = z = 0, & \quad b_1 = \frac{1-z}{2} = \frac{1}{2}, & \quad c_1 = \frac{1-z}{2} = \frac{1}{2}, & \quad w_1 = w = \frac{1}{3}, \\ a_2 = \frac{1-z}{2} = \frac{1}{2}, & \quad b_2 = z = 0, & \quad c_2 = \frac{1-z}{2} = \frac{1}{2}, & \quad w_2 = w = \frac{1}{3}, \\ a_3 = \frac{1-z}{2} = \frac{1}{2}, & \quad b_3 = \frac{1-z}{2} = \frac{1}{2}, & \quad c_3 = z = 0, & \quad w_3 = w = \frac{1}{3}. \end{aligned}$$

При $n = 3$ будем иметь три уравнения

$$\begin{aligned} wz + w \frac{1-z}{2} + w \frac{1-z}{2} + w_0 \frac{1}{3} &= \frac{1}{3}, \\ wz^2 + w \left(\frac{1-z}{2} \right)^2 + w \left(\frac{1-z}{2} \right)^2 + w_0 \frac{1}{9} &= \frac{1}{6}, \\ wz^3 + w \left(\frac{1-z}{2} \right)^3 + w \left(\frac{1-z}{2} \right)^3 + w_0 \frac{1}{27} &= \frac{1}{10}, \end{aligned}$$

из которых с учетом (8.10.16) получим вес w_0 для центральной точки и координаты z , $(1-z)/2$, $(1-z)/2$ и веса w еще для трех симметрично расположенных точек на треугольной области. При $n = 4$ будем иметь уравнения для шести точек и их весовых коэффициентов:

$$\begin{aligned} w_1 + w_2 &= \frac{1}{3}, \\ w_1 \left(z_1^2 + \frac{(1-z_1)^2}{2} \right) + w_2 \left(z_2^2 + \frac{(1-z_2)^2}{2} \right) &= \frac{1}{6}, \\ w_1 \left(z_1^3 + \frac{(1-z_1)^3}{4} \right) + w_2 \left(z_2^3 + \frac{(1-z_2)^3}{4} \right) &= \frac{1}{10}, \\ w_1 \left(z_1^4 + \frac{(1-z_1)^4}{8} \right) + w_2 \left(z_2^4 + \frac{(1-z_2)^4}{8} \right) &= \frac{1}{15}. \end{aligned}$$

При $n = 5$ и будем иметь уравнения для весового коэффициента центральной точки и шести точек и их весовых коэффициентов:

$$\begin{aligned} w_1 + w_2 + w_0 \frac{1}{3} &= \frac{1}{3}, \\ w_1 \left(z_1^2 + \frac{(1-z_1)^2}{2} \right) + w_2 \left(z_2^2 + \frac{(1-z_2)^2}{2} \right) + w_0 \frac{1}{9} &= \frac{1}{6}, \\ w_1 \left(z_1^3 + \frac{(1-z_1)^3}{4} \right) + w_2 \left(z_2^3 + \frac{(1-z_2)^3}{4} \right) + w_0 \frac{1}{27} &= \frac{1}{10}, \\ w_1 \left(z_1^4 + \frac{(1-z_1)^4}{8} \right) + w_2 \left(z_2^4 + \frac{(1-z_2)^4}{8} \right) + w_0 \frac{1}{81} &= \frac{1}{15}, \\ w_1 \left(z_1^5 + \frac{(1-z_1)^5}{16} \right) + w_2 \left(z_2^5 + \frac{(1-z_2)^5}{16} \right) + w_0 \frac{1}{243} &= \frac{1}{21}. \end{aligned}$$

В табл. 8.10.1 приведены значения барицентрических координат и весовые множители для кубатурных формул по треугольной области при $n = 1, 2, 3, 4, 5$. При $n > 5$ система уравнений для определения барицентрических координат не имеет решения. Для более точного вычисления двойных интегралов можно разбить треугольную область интегрирования на несколько треугольных подобластей.

Таблица 8.10.1

k	a_i	b_i	c_i	w_i
1	1/3	1/3	1/3	1
3	2/3	1/6	1/6	1/3
	1/6	2/3	1/6	1/3
	1/6	1/6	2/3	1/3
4	1/3	1/3	1/3	-0,5625
	a	b	b	0,5208333333333333
	b	a	b	0,5208333333333333
	b	b	a	0,5208333333333333
	где $a = 0,6$		$b = 0,2$	
6	a_1	b_1	b_1	0,1099517436553224
	b_1	a_1	b_1	0,1099517436553224
	b_1	b_1	a_1	0,1099517436553224
	a_2	b_2	b_2	0,2233815896780112
	b_2	a_2	b_2	0,2233815896780112
	b_2	b_2	a_2	0,2233815896780112
	где $a_1 = 0,8168475729804581$		$b_1 = 0,0915762135097711$	
	где $a_2 = 0,1081030181680702$		$b_2 = 0,4459484909159648$	
7	1/3	1/3	1/3	0,225
	a_1	b_1	b_1	0,1259391805448265
	b_1	a_1	b_1	0,1259391805448265
	b_1	b_1	a_1	0,1259391805448265
	a_2	b_2	b_2	0,1323941527885112
	b_2	a_2	b_2	0,1323941527885112
	b_2	b_2	a_2	0,1323941527885112
	где $a_1 = 0,7974269853530880$		$b_1 = 0,1012865073234560$	
	где $a_2 = 0,0597158717897742$		$b_2 = 0,4701420641051129$	

Погрешность кубатурных формул для треугольных областей имеет порядок $S \cdot \max(\partial^{n+1} f / \partial u^{n+1}, \partial^{n+1} f / \partial v^{n+1})$ и при тех же вычислительных затратах она больше погрешности соответствующих кубатурных формул Гаусса.

Рассмотренные кубатурные формулы применяются для вычисления геометрических характеристик тел.

8.11. Разбиение поверхностей при интегрировании

При численном интегрировании по поверхности требуется разбить область определения ее параметров на небольшие четырехугольные или треугольные подобласти. Размер каждой подобласти определим условием: угловое изменение нормали поверхности в подобласти не должно превышать некоторую заданную величину $\Delta\alpha$. В соответствии с формулами (4.2.4) и (4.2.5) изменение параметра

тров поверхности в каждой подобласти не должны превосходить величин

$$\Delta u = \frac{\Delta \alpha \sqrt{g_{11}}}{b_{11}}, \quad (8.11.1)$$

$$\Delta v = \frac{\Delta \alpha \sqrt{g_{22}}}{b_{22}}, \quad (8.11.2)$$

где g_{11} и g_{22} — коэффициенты первой основной квадратичной формы поверхности (1.7.8), b_{11} и b_{22} — коэффициенты второй основной квадратичной формы поверхности (1.7.21). Таким образом, мы заменим двухмерную область определения параметров совокупностью стыкующихся между собой четырехугольников и треугольников.

Рассмотрим разбиение поверхности, показанной на рис. 8.11.1.

Поверхность получена путем усечения контурами NURBS поверхности.

Разобьем прямоугольную область $u_{\min} \leq u \leq u_{\max}$, $v_{\min} \leq v \leq v_{\max}$, где u_{\min} , u_{\max} , v_{\min} , v_{\max} определяют габаритный прямоугольник внешнего граничного контура, на прямоугольные подобласти двухмерными линиями $u_i = \text{const}$

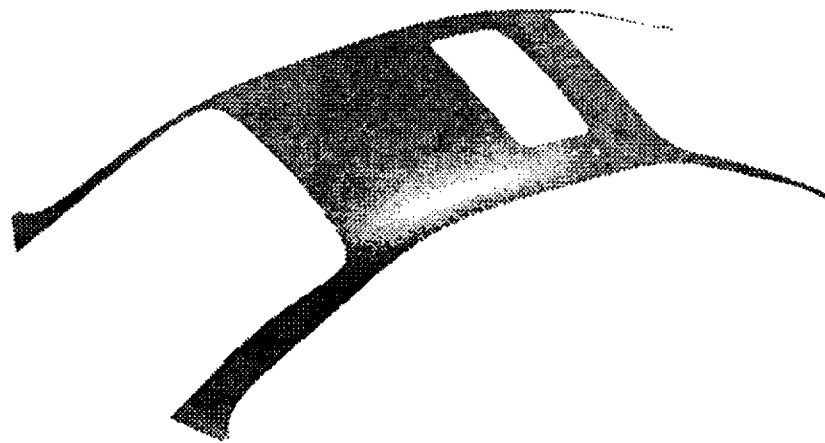


Рис. 8.11.1. Поверхность для численного интегрирования

и $v_j = \text{const}$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$. Параметрические расстояния $\Delta u_i = u_{i+1} - u_i$ между соседними линиями $u_i = \text{const}$ в соответствии с формулой (8.11.1) возьмем равными

$$\Delta u_i = \min \left(\frac{\Delta \alpha \sqrt{g_{11}(u_i, v)}}{b_{11}(u_i, v)} \right) \quad \text{для всех } v_{\min} \leq v \leq v_{\max}. \quad (8.11.3)$$

Параметрические расстояния $\Delta v_j = v_{j+1} - v_j$ между соседними линиями $v_j = \text{const}$ в соответствии с формулой (8.11.2) возьмем равными

$$\Delta v_j = \min \left(\frac{\Delta \alpha \sqrt{g_{22}(u, v_j)}}{b_{22}(u, v_j)} \right) \quad \text{для всех } u_{\min} \leq u \leq u_{\max}. \quad (8.11.4)$$

На рис. 8.11.2 приведено разбиение области $u_{\min} \leq u \leq u_{\max}$, $v_{\min} \leq v \leq v_{\max}$ базовой NURBS поверхности описанным способом.

Область определения параметров поверхности разбита на прямоугольники, но из-за искривленности поверхности каждая подобласть представляет собой криволинейный четырехугольник.

Далее определим те прямоугольники, которые целиком лежат внутри ограничивающих поверхность контуров (рис. 8.11.3).

Не вошедшую в прямоугольные подобласти часть поверхности $r(u, v)$ разобьем на треугольные подобласти (рис. 8.11.4).

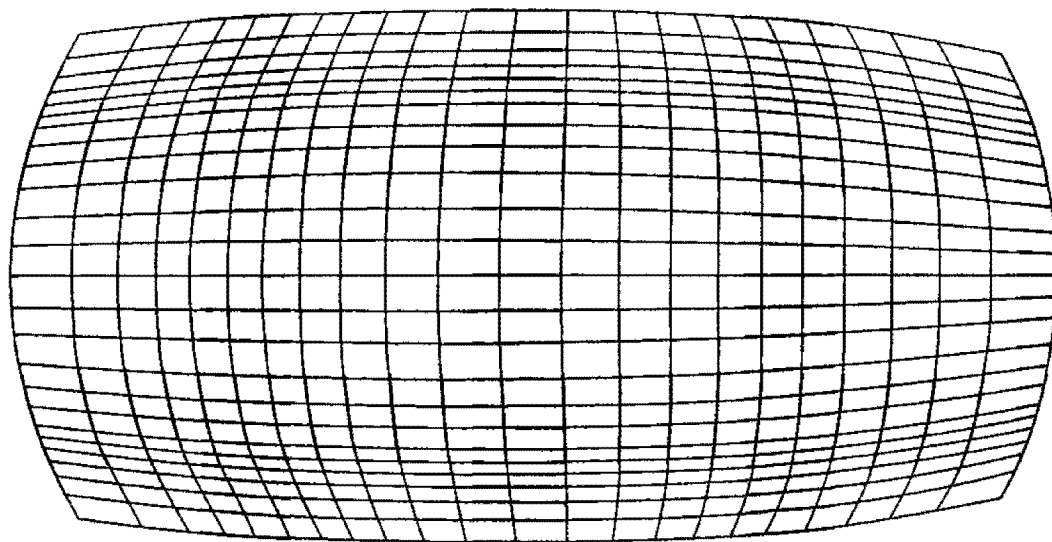


Рис. 8.11.2. Разбиение прямоугольной области определения параметров поверхности

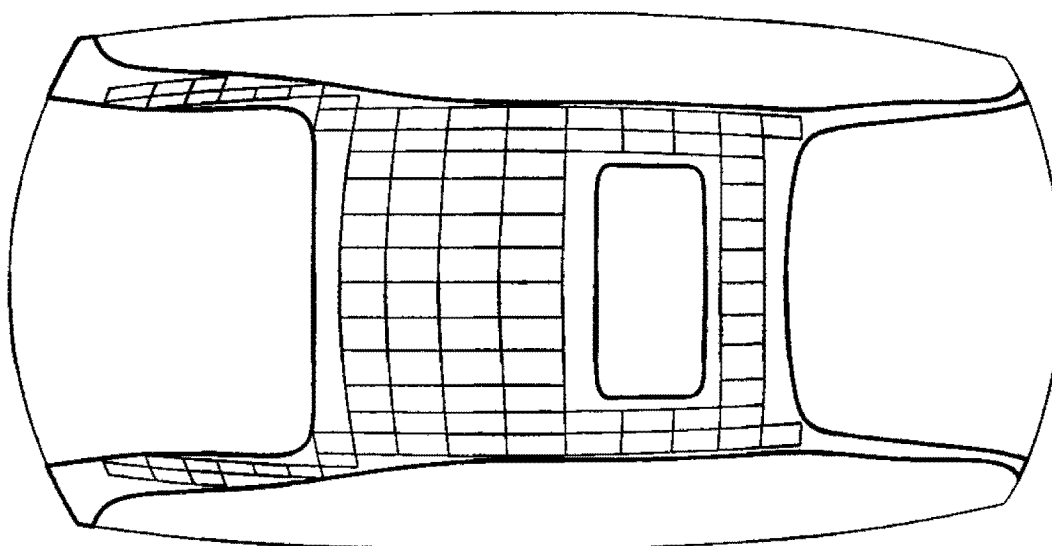


Рис. 8.11.3. Незаконченное разбиение поверхности

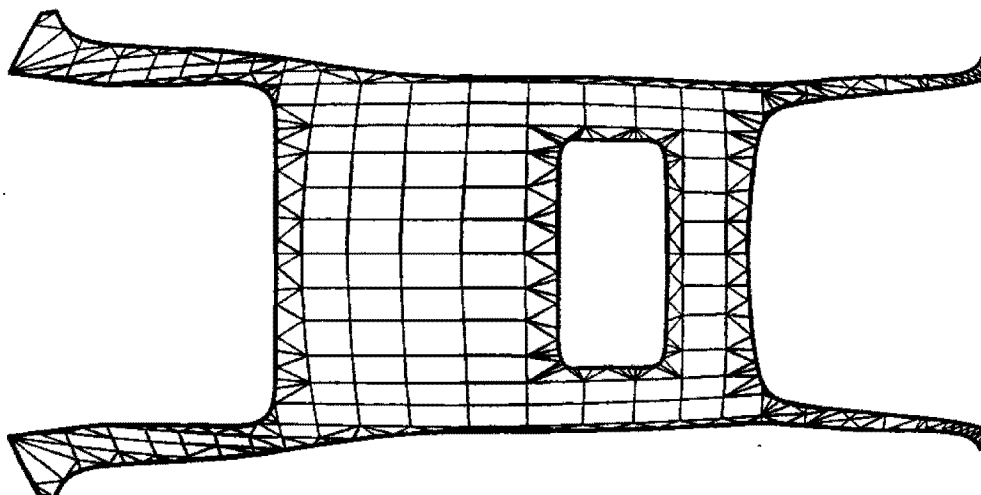


Рис. 8.11.4. Разбиение поверхности при численном интегрировании

В результате разбиения двумерные граничные контуры будут заменены ломаными линиями. Угол между касательными к участкам этих ломаных линий не должен превосходить заданную величину $\Delta\alpha$. Вершины ломаной лежат на граничных контурах. Параметрическое расстояние Δt между вершинами будем вычислять в соответствии с (4.2.3) по формуле

$$\Delta t = \frac{\Delta\alpha |dc/dt|^2}{|dc/dt \times d^2c/dt^2|}, \quad (8.11.5)$$

где $c(t) = r(u(t), v(t))$ — кривая граничного контура поверхности $r(u, v)$. Вершинами треугольных подобластей являются вершины ломаных и вершины четырехугольных подобластей. Построение треугольников по заданным вершинам можно выполнить с помощью алгоритма триангуляции Делоне.

Методы разбивки области определения параметров поверхности на треугольные подобласти (методы триангуляции поверхности) мы рассмотрим в следующей главе.

Глава 9

КОМПЬЮТЕРНАЯ ГРАФИКА

9.1. 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 видеofilm.

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. 9.1.1 и 9.1.2.

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. 9.1.3), вторые — *центральными* проекциями или перспективными изображениями (рис. 9.1.4). 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 реалистичные 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.

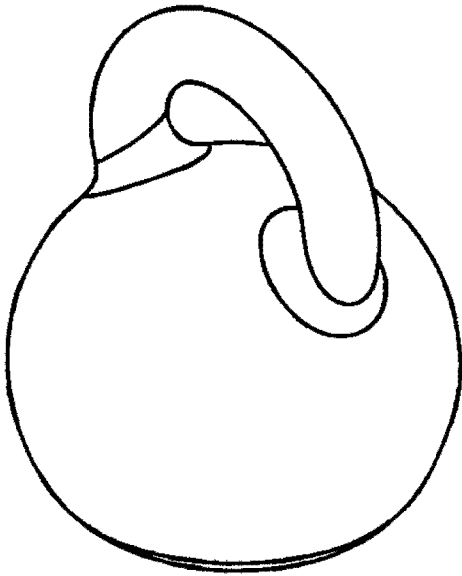


Рис. 9.1.1. Векторное изображение

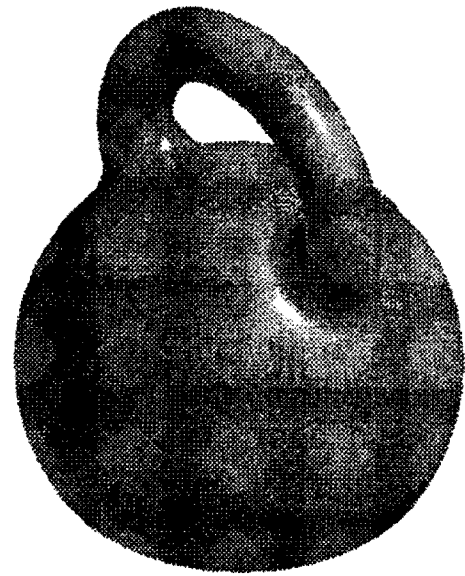


Рис. 9.1.2. Растровое изображение

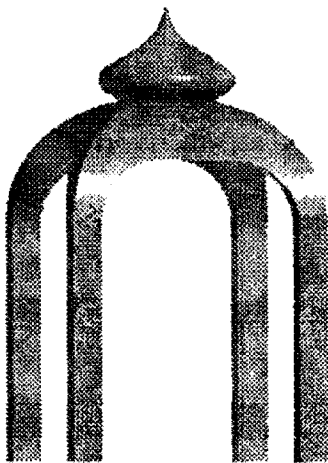


Рис. 9.1.3. Параллельная проекция

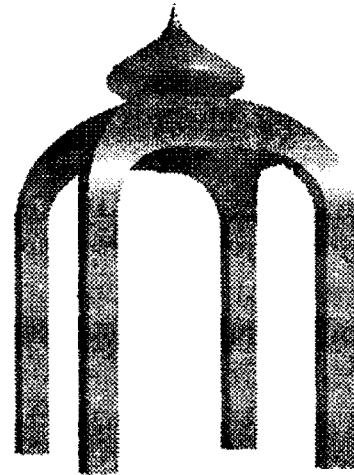


Рис. 9.1.4. Центральная проекция

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.

9.2. 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 i_x и 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 ($\mathbf{i}_z = -\mathbf{i}_x \times \mathbf{i}_y$). На рис. 9.2.1 приведен пример

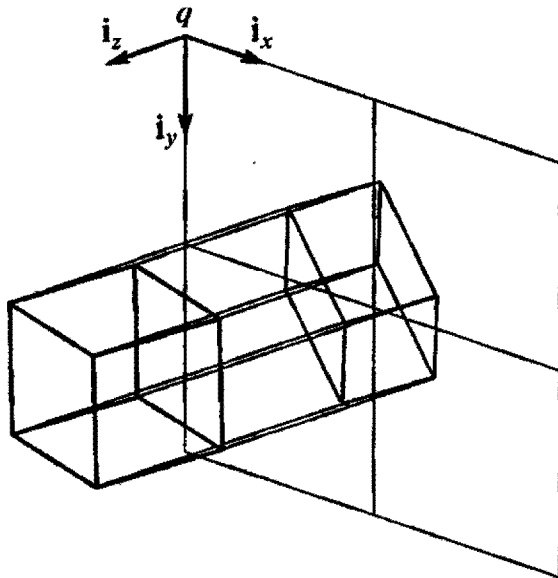


Рис. 9.2.1. Проекционная плоскость

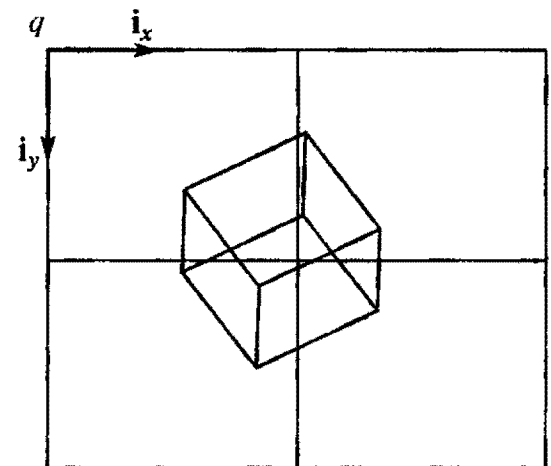


Рис. 9.2.2. Проекция объекта

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. 9.2.2 приведены параллельные проекции ребер прямого параллелепипеда на плоскости.

Проекция точки. Рассмотрим произвольную точку $\mathbf{r} = [r_1 \ r_2 \ r_3]^T$. Определить двухмерную точку $\mathbf{p} = [x \ y]^T$, являющуюся ее проекцией на проекционной плоскости

$$\mathbf{p}(x, y) = \mathbf{q} + x\mathbf{i}_x + y\mathbf{i}_y. \quad (9.2.1)$$

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 (1.2.8), которые в данном случае имеют вид

$$\begin{aligned} x &= x_1(r_1 - q_1) + x_2(r_2 - q_2) + x_3(r_3 - q_3) = x_1r_1 + x_2r_2 + x_3r_3 + o_x, \\ y &= y_1(r_1 - q_1) + y_2(r_2 - q_2) + y_3(r_3 - q_3) = y_1r_1 + y_2r_2 + y_3r_3 + o_y, \\ z &= z_1(r_1 - q_1) + z_2(r_2 - q_2) + z_3(r_3 - q_3) = z_1r_1 + z_2r_2 + z_3r_3 + o_z, \end{aligned} \quad (9.2.2)$$

где

$$o_x = -x_1q_1 - x_2q_2 - x_3q_3,$$

$$o_y = -y_1q_1 - y_2q_2 - y_3q_3,$$

$$o_z = -z_1q_1 - z_2q_2 - z_3q_3.$$

В однородных координатах преобразование (9.2.2) описывается равенством (1.4.8)

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & o_x \\ y_1 & y_2 & y_3 & o_y \\ z_1 & z_2 & z_3 & o_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} = \mathbf{A}_0 \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix}. \quad (9.2.3)$$

Коэффициенты расширенной матрицы преобразования выражаются через компоненты векторов, определяющих положение проекционной плоскости.

Проекция точки на плоскость (9.2.1) описывается первыми двумя координатами вектора (9.2.3). Таким образом, первые две координаты расширенного радиус-вектора, полученного по формуле (9.2.3), являются координатами искомой параллельной проекцией $r = [x \ y]^T$ рассматриваемой точки.

Масштаб проекции. С помощью преобразования (9.2.3) 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 $r = [x \ y]^T$ точки проекции необходимо умножить на матрицу

$$M = \begin{bmatrix} m_x & 0 \\ 0 & m_y \end{bmatrix}. \quad (9.2.4)$$

Если требуется получить изображение объекта в масштабе $m:1$, то радиус-вектор двумерной точки проекции следует преобразовать по формуле (1.3.17). Таким образом, координаты x' и y' проекции некоторой точки r геометрического объекта с учетом масштаба равны

$$\begin{aligned} \begin{bmatrix} x' \\ y' \end{bmatrix} &= \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + m \cdot \left(\begin{bmatrix} m_x & 0 \\ 0 & m_y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \right) = \\ &= \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + m \cdot \left(\begin{bmatrix} m_x & 0 \\ 0 & m_y \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 & x_3 & o_x \\ y_1 & y_2 & y_3 & o_y \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \right), \quad (9.2.5) \end{aligned}$$

где x_0 и y_0 — координаты точки, которая при масштабировании должна остаться неподвижной. Вместо матрицы \mathbf{A}_0 в (9.2.5) используется матрица, содержащая ее первые две строки.

Мы видим, что для получения координат x' и y' плоской проекции некоторой точки $r = [r_1 \ r_2 \ r_3]^T$ 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 .

9.3. 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 i_x и i_y . Свяжем с проекционной плоскостью *местную систему координат* $Qxyz$. Пусть начальная точка q местной системы координат в глобальной системе координат описывается радиус-вектором $q = [q_1 \ q_2 \ q_3]^T$, а орты $i_x, i_y, i_z = \pm i_x \times i_y$ местной системы координат в глобальной системе координат описываются векторами $i_x = [x_1 \ x_2 \ x_3]^T$, $i_y = [y_1 \ y_2 \ y_3]^T$, $i_z = [z_1 \ z_2 \ z_3]^T$.

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 . На рис. 9.3.1 приведены местная

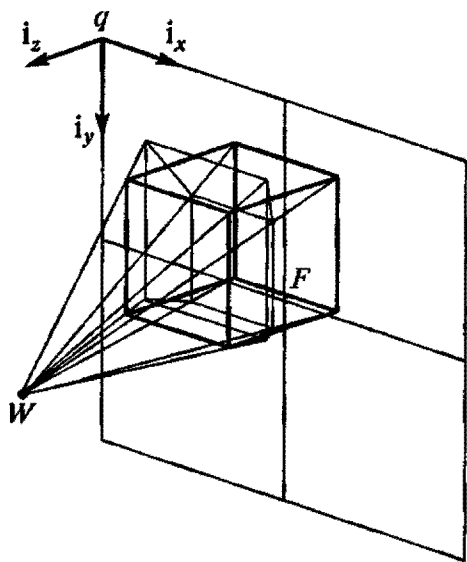


Рис. 9.3.1. Построение центральной проекции

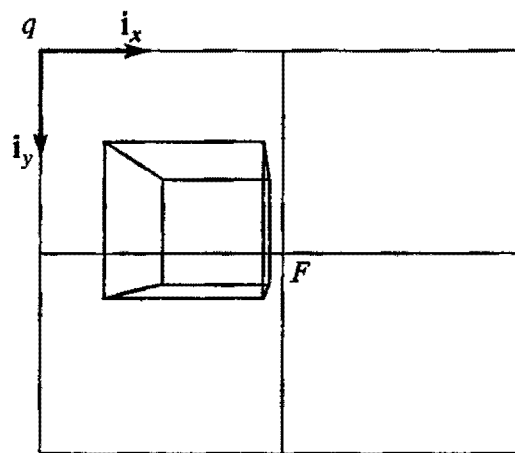


Рис. 9.3.2. Центральная проекция объекта

система координат, точка наблюдения и центральные проекции ребер прямого параллелепипеда. На рис. 9.3.2 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 (9.2.1). На рис. 9.3.3 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 .

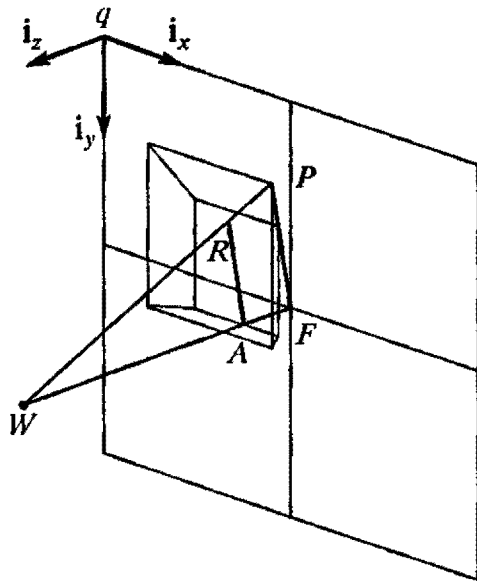


Рис. 9.3.3. Построение центральной проекции точки R

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 (9.2.3). 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. 9.3.3 два подобных треугольника: WFP и WAR . Из подобия треугольников следует, что $PW/FW = RW/AW$. Запишем последнее равенство в векторном виде:

$$\frac{\mathbf{p} - \mathbf{w}}{z_w} = \frac{\mathbf{r} - \mathbf{w}}{z_w - z}. \quad (9.3.1)$$

Используя это равенство, найдем радиус-вектор \mathbf{p} центральной проекции точки \mathbf{r}

$$\mathbf{p} = \mathbf{w} + (\mathbf{r} - \mathbf{w}) \frac{z_w}{z_w - z} = \mathbf{r} \frac{z_w}{z_w - z} - \mathbf{w} \frac{z}{z_w - z}. \quad (9.3.2)$$

Чтобы получить координаты точки \mathbf{p} в проекционной плоскости, нужно с помощью равенства (9.2.3) 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 (9.2.1). Запишем равенство (9.3.2) в местной системе координат:

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = \frac{z_w}{z_w - z} \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \frac{z}{z_w - z} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \begin{bmatrix} \frac{x z_w - z x_w}{z_w - z} \\ \frac{y z_w - z y_w}{z_w - z} \\ 0 \end{bmatrix}. \quad (9.3.3)$$

Таким образом, декартовы координаты двумерной точки $\mathbf{p} = [x_p \ y_p]^T$, являющейся центральной проекцией рассматриваемой точки \mathbf{r} в пространстве, на

плоскости (9.2.1) равны

$$x_p = \frac{xz_w - zx_w}{z_w - z}, \quad y_p = \frac{yz_w - zy_w}{z_w - z}. \quad (9.3.4)$$

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 (9.2.3) равенство (9.3.3) можно записать следующим образом:

$$\begin{aligned} \begin{bmatrix} x_p d \\ y_p d \\ z_p d \\ d \end{bmatrix} &= \begin{bmatrix} z_w & 0 & -x_w & 0 \\ 0 & z_w & -y_w & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & z_w \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \\ &= \begin{bmatrix} z_w & 0 & -x_w & 0 \\ 0 & z_w & -y_w & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & z_w \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 & x_3 & o_x \\ y_1 & y_2 & y_3 & o_y \\ z_1 & z_2 & z_3 & o_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 1 \end{bmatrix}. \quad (9.3.5) \end{aligned}$$

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 (9.3.5) 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

$$\mathbf{A}_w = \begin{bmatrix} x_1 & y_1 & z_1 & q_x \\ x_2 & y_2 & z_2 & q_y \\ x_3 & y_3 & z_3 & q_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} z_w & 0 & -x_w & 0 \\ 0 & z_w & -y_w & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & z_w \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 & x_3 & o_x \\ y_1 & y_2 & y_3 & o_y \\ z_1 & z_2 & z_3 & o_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (9.3.6)$$

После такого преобразования объект исказится так, что его параллельная проекция на плоскость (9.2.1) будет совпадать с его центральной проекцией на эту же плоскость. Определитель матрицы \mathbf{A}_w не равен нулю, и преобразование (9.3.6) 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 (9.2.4) и масштаб изображения.

На рис. 9.3.4 и 9.3.5 приведены перспективные изображения геометрического объекта. Для обоих рисунков использовалась одна и та же проекционная

плоскость, но на рис. 9.3.4 проекция точки наблюдения лежит слева и выше центра рисунка, а на рис. 9.3.5 проекция точки наблюдения лежит правее центра

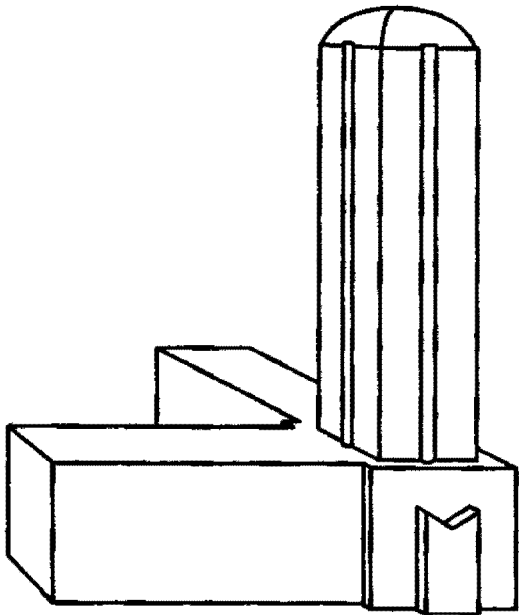


Рис. 9.3.4. Центральная проекция объекта

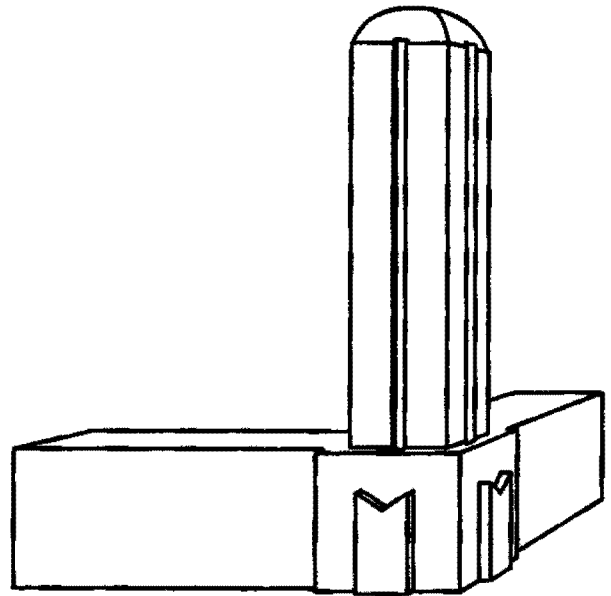


Рис. 9.3.5. Центральная проекция объекта

рисунка. Чем дальше от точки наблюдения находится объект, тем меньше его перспективное изображение.

9.4. 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 \mathbf{p}_k , 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. 9.4.1).

Радиус ρ соприкасающейся окружности равен радиусу кривизны кривой и определяется формулой

$$\rho = \frac{|\mathbf{r}'|^3}{|\mathbf{r}' \times \mathbf{r}''|}. \quad (9.4.1)$$

По теореме Пифагора квадрат половины длины хорды равен $\rho^2 - (\rho - \delta)^2 = \delta(2\rho - \delta)$. Длина всей хорды будет равна

$$h = 2\sqrt{\delta(2\rho - \delta)}. \quad (9.4.2)$$

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

$$\Delta t \approx \frac{h}{|\mathbf{r}'|} = 2 \frac{\sqrt{\delta(2\rho - \delta)}}{|\mathbf{r}'|}. \quad (9.4.3)$$

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 (9.4.3) 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.

Формула вычисления параметрического шага между точками полигона (9.4.3) 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 $m:1$ с точностью отображения δ , то мы должны построить полигон, точки которого находятся на параметрической расстоянии, не превышающем

$$\Delta t \approx 2 \frac{\sqrt{(\delta/m)(2\rho - \delta/m)}}{|\mathbf{r}'|}. \quad (9.4.4)$$

Шаг по параметру кривой мы вычисляли при выполнении операций. Шаг (4.2.3) определяется углом отклонения кривой и не зависит от ее размеров и масштаба отображения. Шаг (9.4.4) 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 и v_0 . Найдем параметры u_1 и 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

$$\rho = \frac{g_{11} du^2 + 2g_{12} du dv + g_{22} dv^2}{b_{11} du^2 + 2b_{12} du dv + b_{22} dv^2}. \quad (9.4.5)$$

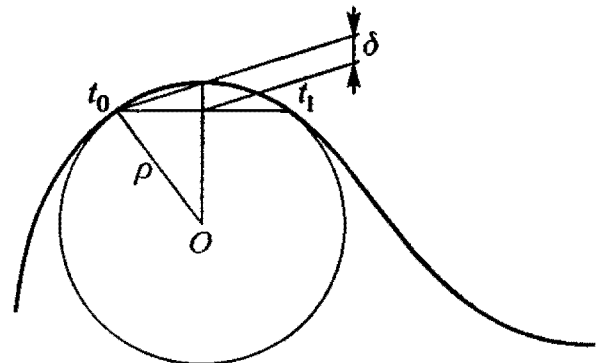


Рис. 9.4.1. Вычисление шага по параметру кривой по заданному прогибу δ

Подставим (9.4.5) в (9.4.4) и получим приращения для параметров соседней точки полигона

$$\Delta u \approx 2 \frac{\sqrt{(\delta/m)(2\rho - \delta/m)}}{|\mathbf{r}_1 + \mathbf{r}_2(dv/du)|}, \quad \Delta v \approx 2 \frac{\sqrt{(\delta/m)(2\rho - \delta/m)}}{|\mathbf{r}_1(du/dv) + \mathbf{r}_2|}. \quad (9.4.6)$$

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 — $\rho_v = g_{22}/b_{22}$. Соответственно ненулевые шаги вдоль координатных u -линий и v -линий поверхности равны

$$\Delta u \approx 2 \frac{\sqrt{(\delta/m)(2(g_{11}/b_{11}) - \delta/m)}}{|\partial \mathbf{r} / \partial u|}, \quad \Delta v = 0, \quad (9.4.7)$$

$$\Delta u = 0, \quad \Delta v \approx 2 \frac{\sqrt{(\delta/m)(2(g_{22}/b_{22}) - \delta/m)}}{|\partial \mathbf{r} / \partial v|}, \quad (9.4.8)$$

где $g_{11}, g_{12}, g_{21}, g_{22}, b_{11}, b_{12}, b_{21}, b_{22}$ — коэффициенты первой (1.7.8) и второй (1.7.21) 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. 9.4.2 приведен пример отображения сетки поверхности, построенной по координатным u -линиям и v -линиям.

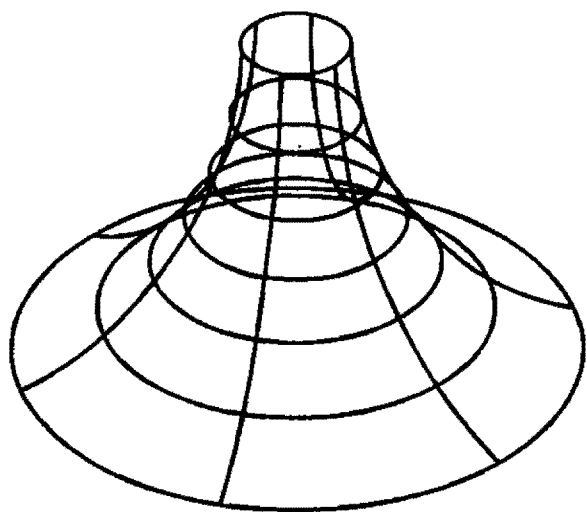


Рис. 9.4.2. Проекция сетки полигонов поверхности

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*.

9.5. 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

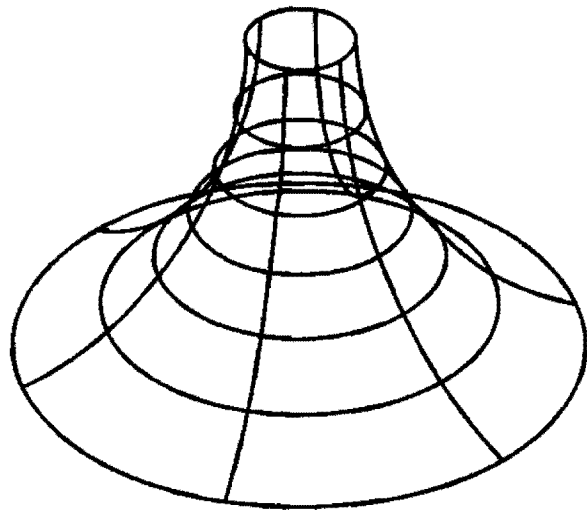
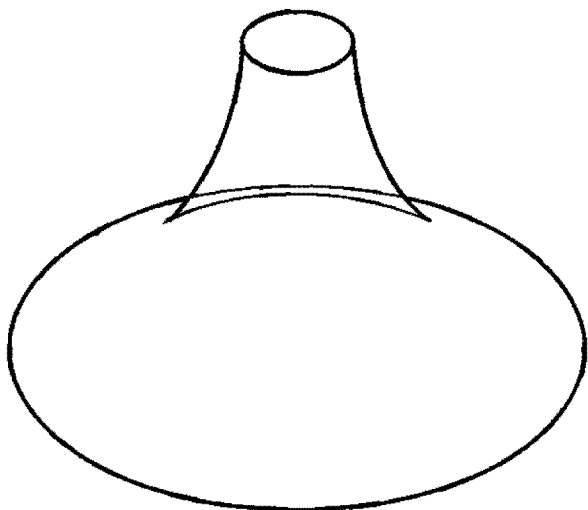


Рис. 9.5.1. Проекция линий очерка поверхности

Рис. 9.5.2. Проекция сетки полигонов и линий очерка

видимую часть поверхности или грани от невидимой ее части. На рис. 9.5.1 приведены линии очерка поверхности. На рис. 9.5.2 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 (9.2.1) должна удовлетворять уравнению

$$\mathbf{i}_z \cdot \mathbf{m} = 0, \quad (9.5.1)$$

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 (9.5.1) содержит два искомых параметра u, v . Если задать один из параметров, то другой можно найти из уравнения (9.5.1), т.е. один из параметров является функцией от другого. Для равноправия параметров их можно представить в виде функций некоторого общего параметра t :

$$u = u(t), \quad v = v(t). \quad (9.5.2)$$

Результатом решения уравнения (9.5.1) является двумерная линия

$$l_{uv}(t) = [u(t) \ v(t)]^T, \quad l_{uv}(t) \in \mathbf{r}(u, v) \quad (9.5.3)$$

на поверхности $\mathbf{r}(u, v)$. Эта линия и есть линия очерка поверхности.

Мы построим линию очерка по упорядоченной совокупности точек, удовлетворяющих уравнению (9.5.1). Точками мы называем пару параметров u_i и 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 (9.5.1). 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 $\mathbf{m} \cdot \mathbf{i}_z$ 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

$$u^{(k+1)} = u^{(k)}, \quad v^{(k+1)} = v^{(k)} - \left(\frac{\mathbf{m} \cdot \mathbf{i}_z}{\mathbf{i}_z \cdot (\partial \mathbf{m} / \partial v)} \right)^{(k)} \quad (9.5.4)$$

или итерационного процесса

$$u^{(k+1)} = u^{(k)} - \left(\frac{\mathbf{m} \cdot \mathbf{i}_z}{\mathbf{i}_z \cdot (\partial \mathbf{m} / \partial u)} \right)^{(k)}, \quad v^{(k+1)} = v^{(k)} \quad (9.5.5)$$

найдем параметры одной из точек линии очерка. Производные нормали \mathbf{m}_1 и \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

$$\mathbf{t} = \pm \frac{\mathbf{m}_1 \times \mathbf{m}}{|\mathbf{r}_1|} \pm \frac{\mathbf{m}_2 \times \mathbf{m}}{|\mathbf{r}_2|}, \quad (9.5.6)$$

где \mathbf{m}_1 и \mathbf{m}_2 — частные производные нормали $\mathbf{m}(u, v)$, \mathbf{r}_1 и \mathbf{r}_2 — частные производные радиус-вектора $\mathbf{r}(u, v)$ поверхности по параметрам u и 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 (9.4.7) или по формуле (9.4.8). Если

$$\frac{\mathbf{t} \cdot \mathbf{r}_1}{|\mathbf{r}_1|} > \frac{\mathbf{t} \cdot \mathbf{r}_2}{|\mathbf{r}_2|},$$

то по формуле (9.4.7) дадим приращение параметру u и по формуле (9.5.4) найдем соответствующий ему параметр v поверхности. В противном случае по формуле (9.4.8) дадим приращение параметру v и по формуле (9.5.5) 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 $[u_i \ v_i]^T$, $i = 0, 1, 2, \dots, n$ 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. Совокупность ее точек найдем, взяв

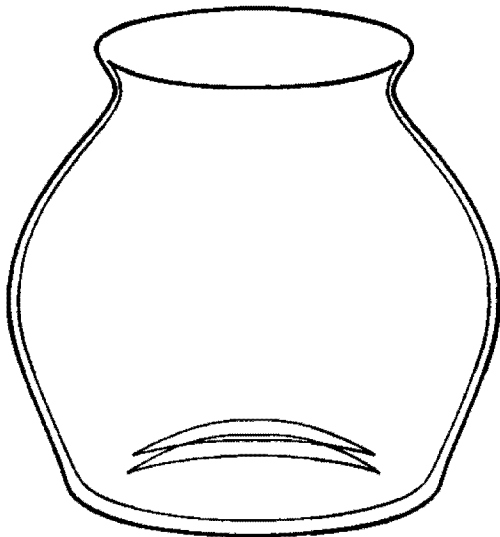


Рис. 9.5.3. Линии очерка тела



Рис. 9.5.4. Тело вращения

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

$$\mathbf{m} \cdot (\mathbf{r} - \mathbf{w}) = 0, \quad (9.5.7)$$

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 (9.5.7) 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 (9.5.3), и является пространственной кривой. Эта линия должна быть спроецирована на плоскость по правилам построения центральной проекции пространственной линии.

На рис. 9.5.5 приведена параллельная проекция линий очерка тора, а на рис. 9.5.6 для сравнения приведена центральная проекция линий очерка тора. Как можно видеть, эти проекции отличаются.

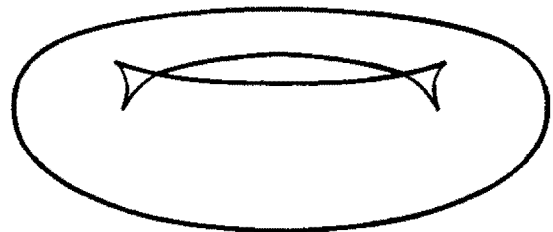
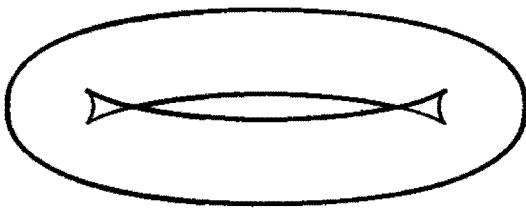


Рис. 9.5.5. Параллельная проекция линий очерка тора

Рис. 9.5.6. 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 (9.5.4) и (9.5.5) следует использовать формулы

$$u^{(k+1)} = u^{(k)}, \quad v^{(k+1)} = v^{(k)} - \left(\frac{\mathbf{m} \cdot (\mathbf{r} - \mathbf{w})}{(\partial \mathbf{m} / \partial v) \cdot (\mathbf{r} - \mathbf{w}) + \mathbf{m} \cdot (\partial \mathbf{r} / \partial v)} \right)^{(k)} \quad (9.5.8)$$

и формулы

$$u^{(k+1)} = u^{(k)} - \left(\frac{\mathbf{m} \cdot (\mathbf{r} - \mathbf{w})}{(\partial \mathbf{m} / \partial u) \cdot (\mathbf{r} - \mathbf{w}) + \mathbf{m} \cdot (\partial \mathbf{r} / \partial u)} \right)^{(k)}, \quad v^{(k+1)} = v^{(k)} \quad (9.5.9)$$

соответственно. В остальном алгоритм построения линий очерка для центральной проекции поверхности не отличается от алгоритма построения линий очерка для параллельной проекции.

9.6. 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. 9.6.1 приведена проекция поверхности, показанной на рис. 9.4.2, без невидимых линий.

Линия очерка будет видимой, если она не закрыта другой гранью или частью своей же грани.

Рассмотрим один из алгоритмов удаления невидимых линий. Сначала построим пространственные полигоны геометрического объекта, включая полигоны линий очерка. Совокупность пространственных линий $c_i(t_i)$, $i = 1, 2, \dots, n$, соответствующих всем полигонам объекта, будем называть *каркасной моделью*. Далее построим проекцию (параллельную или центральную) каркасной модели на заданную плоскость. Эта проекция будет состоять из двумерных линий $c_i(t_i)$, $i = 1, 2, \dots, n$. Каждой пространственной кривой $c_i(t_i)$ соответствует двумерная кривая $c_i(t_i)$. Каждой точке $\mathbf{r} = \mathbf{c}_i(t)$ пространственной кривой соответствует точка $r = c_i(t)$ двумерной 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 двумерных кривых $l_j(t_j)$, $j = 1, 2, \dots, m$ между собой и со всеми кривыми $c_i(t_i)$, $i = 1, 2, \dots, n$ 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 двумерные кривые. Все линии контура будут видимыми. Для определения видимости остальных линий возьмем произвольную точку r на двумерной кривой. Найдем соответствующую ей точку \mathbf{r} на соответствующей пространственной линии. Построим прямую $l(t) = \mathbf{r} + t\mathbf{i}$, 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 двумерными кривыми и удалим те части проекции, которые окажутся невидимыми.

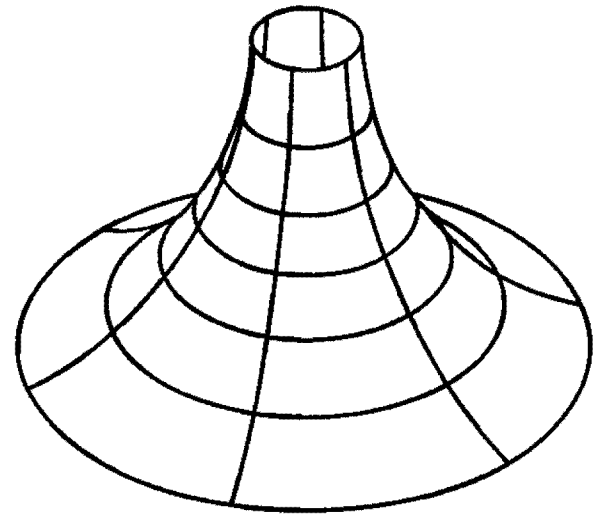


Рис. 9.6.1. Проекция поверхности с удалением невидимых линий

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

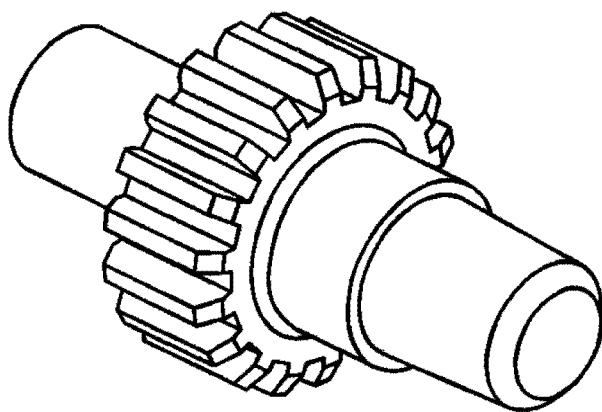


Рис. 9.6.2. Проекция детали с удалением невидимых линий

невидимыми. На рис. 9.6.2 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.

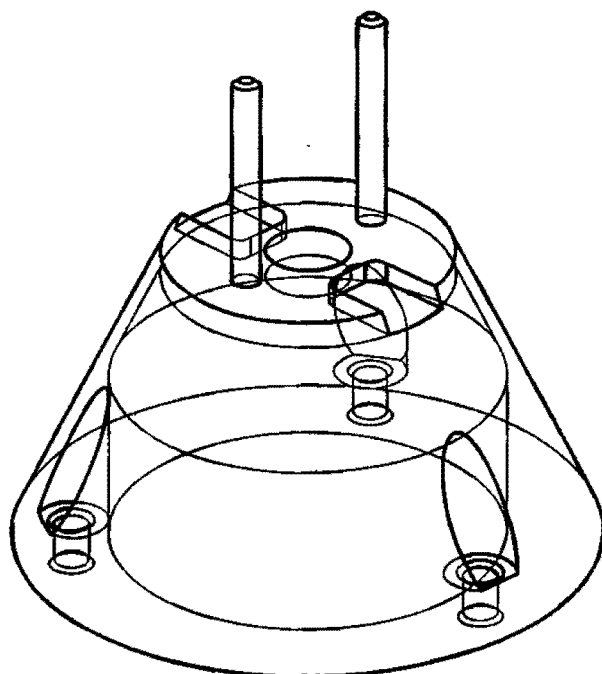


Рис. 9.6.3. Проекция детали с утонением невидимых линий

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 построением развертки 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. 9.6.4 приведено растровое изображение детали, построенное данным способом.

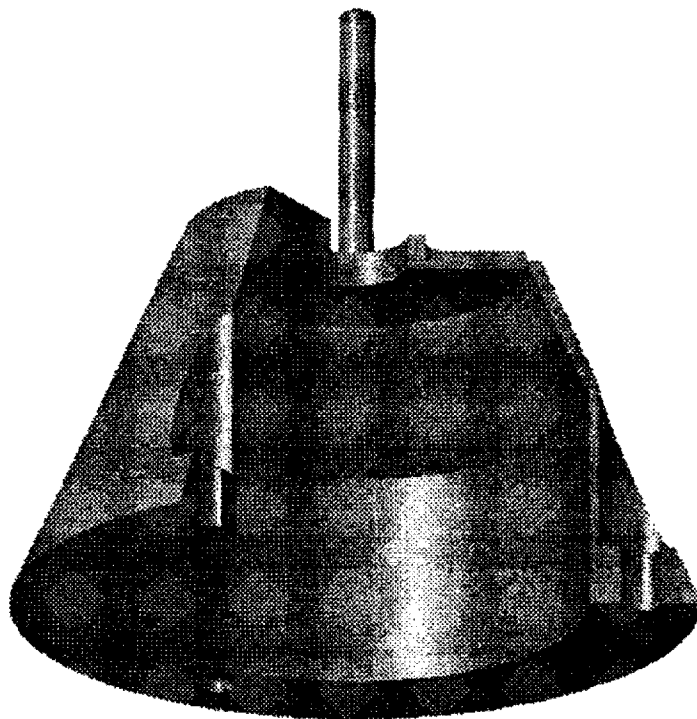


Рис. 9.6.4. Растровое изображение с определением яркости видимых точек

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*.

9.7. 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 $p_i(u_i, v_i)$ и массива нормалей $m_i(u_i, v_i)$ к граням тела в этих точках по массиву двумерных точек $p_i = [u_i \ v_i]^T$. Для быстрого отображения тел их грани аппроксимируют треугольными пластинами, построенными на точках 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 $p_i = [u_i \ v_i]^T$ 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 $p_i(u_i, v_i)$ на поверхности и нормаль $m_i(u_i, v_i)$ 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.

Пусть даны некоторая выпуклая двумерная область, ограниченная замкнутой ломаной линией, и набор точек внутри этой области (рис. 9.7.1).

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. *Триангуляция области будет триангуляцией Делоне,*

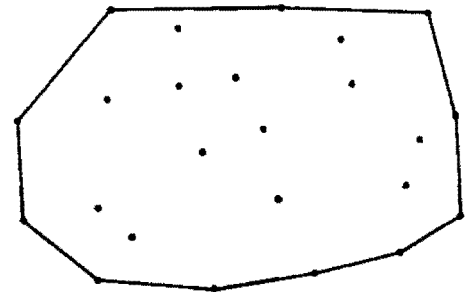


Рис. 9.7.1. Выпуклая область с заданными точками внутри

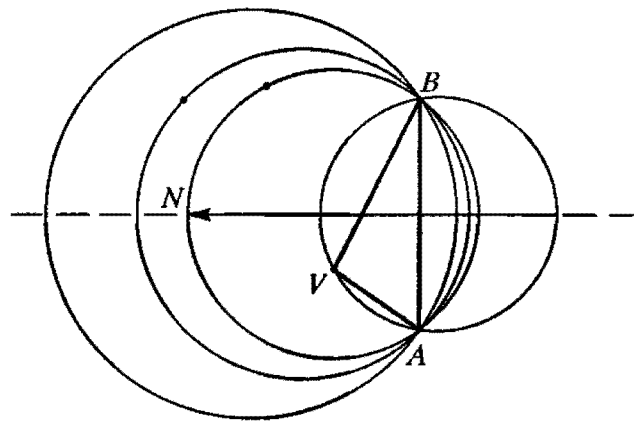


Рис. 9.7.2. Выбор третьей точки алгоритма Делоне

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 AB , показанное на рис. 9.7.2.

Через вершины A , B 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 AB , 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 A , B и V , лежит на пересечении перпендикуляров к серединам отрезков AB , BV и VA . Положение центра окружности будем характеризовать параметром t отрезка MN , перпендикулярного ребру AB , равного с ним по длине и проходящего

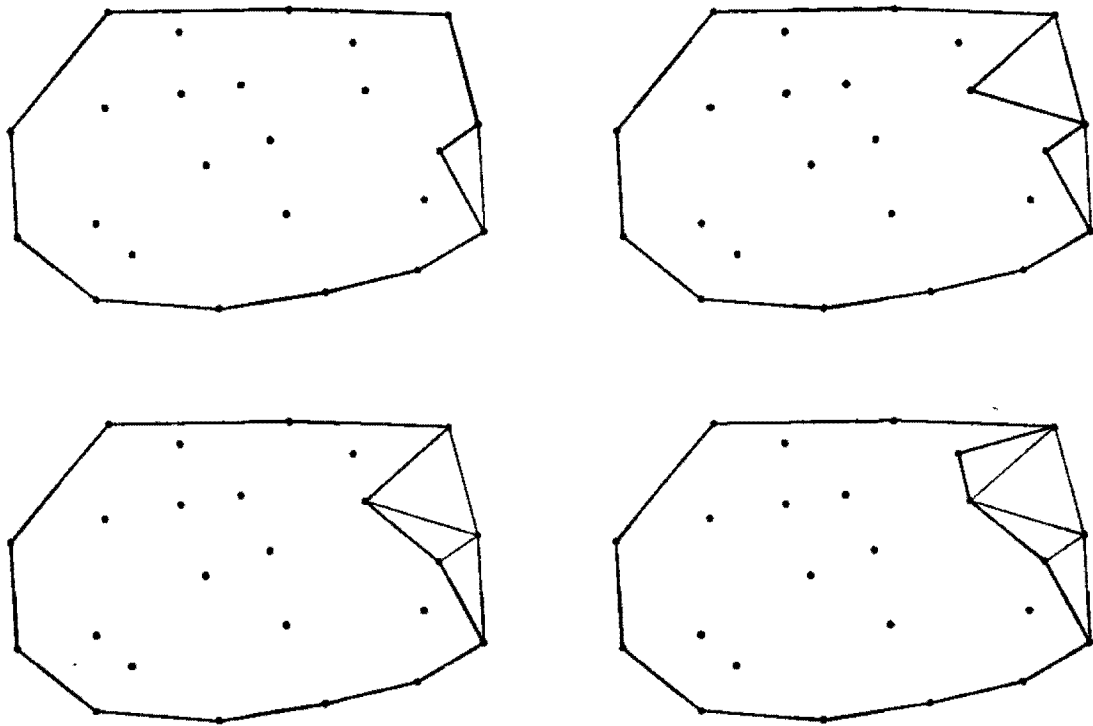


Рис. 9.7.3. Процесс триангуляции Делоне

через середину ребра AB . Для всех вершин, лежащих слева от отрезка AB , ближайшая вершина имеет наименьший параметр t . Соответствующая ближайшей вершине окружность не содержит других вершин слева от отрезка AB . Пусть вершины A , B и V описываются двумерными радиус-векторами $\mathbf{a} = [x_a \ y_a]^\top$, $\mathbf{b} = [x_b \ y_b]^\top$, $\mathbf{v} = [x_v \ y_v]^\top$ соответственно. Радиус-векторы середин отрезков AB и BV будут равны

$$\mathbf{m} = [x_m \ y_m]^\top = \begin{bmatrix} \frac{1}{2}(x_a + x_b) \\ \frac{1}{2}(y_a + y_b) \end{bmatrix}, \quad \mathbf{q} = [x_q \ y_q]^\top = \begin{bmatrix} \frac{1}{2}(x_v + x_b) \\ \frac{1}{2}(y_v + y_b) \end{bmatrix}.$$

Значение параметра t прямой $MN = (1-t)m + tn$, соответствующее положению на ней центра окружности, проходящей через точки A , B и V , равно

$$t = \frac{(x_v - x_b)(x_q - x_m) + (y_v - y_b)(y_q - y_m)}{(y_v - y_b)(x_b - x_a) - (x_v - x_b)(y_b - y_a)}. \quad (9.7.1)$$

Для ближайшей слева к отрезку AB 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 AB найдена ближайшая вершина V . Тогда построим треугольник ABV и переведем ребро AB 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 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. Процесс триангуляции показан на рис. 9.7.3.

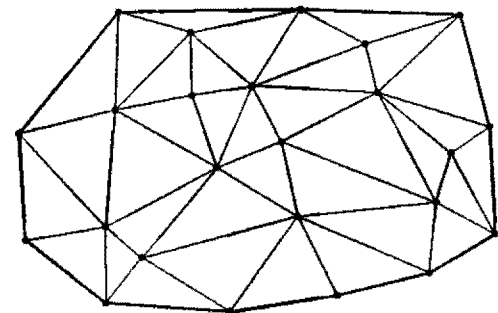


Рис. 9.7.4. Триангуляция Делоне

Триангуляцию закончим, когда все вершины и ребра станут неактивными. Результат триангуляции заданной области приведен на рис. 9.7.4.

Триангуляция методом коррекции. Рассмотрим триангуляцию некоторой поверхности $r(u, v)$ 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* двумерными линиями $u_i = \text{const}$ и $v_j = \text{const}$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$. 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 (9.4.7) возьмем равными

$$\Delta u_i = \min \left(2 \frac{\sqrt{(\delta/m)(2(g_{11}(u_i, v)/b_{11}(u_i, v)) - \delta/m)}}{\sqrt{g_{11}(u_i, v)}} \right) \quad \text{для всех } v_{\min} \leq v \leq v_{\max}. \quad (9.7.2)$$

Параметрические расстояния $\Delta v_j = v_{j+1} - v_j$ между соседними линиями $v_j = \text{const}$ в соответствии с формулой (9.4.8) возьмем равными

$$\Delta v_j = \min \left(2 \frac{\sqrt{(\delta/m)(2(g_{22}(u, v_j)/b_{22}(u, v_j)) - \delta/m)}}{\sqrt{g_{22}(u, v_j)}} \right) \quad \text{для всех } u_{\min} \leq u \leq u_{\max}. \quad (9.7.3)$$

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 $r(u, v)$ с произвольной границей. Метод триангуляции построим на коррекции граничными контурами описанной

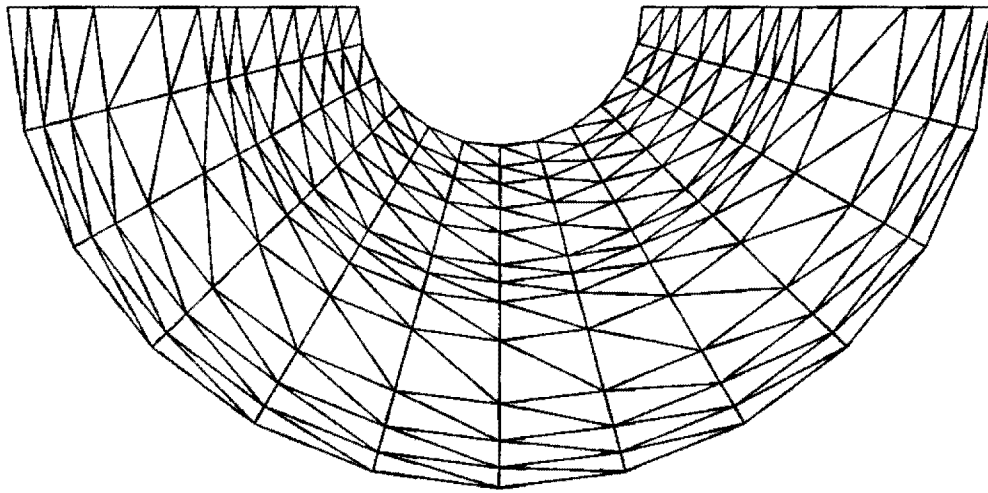


Рис. 9.7.5. Триангуляция поверхности с прямоугольной областью определения параметров

выше триангуляции поверхности с прямоугольной областью определения параметров. Пусть граница поверхности в области определения параметров описывается несколькими непересекающимися двумерными контурами (2.12.7). 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 двумерных 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 $r(t) = r(u(t), v(t))$ граничного контура поверхности вычислим по формуле (9.4.4).

Каждый полигон состоит из упорядоченного набора двумерных точек $p_i = [u_i \ v_i]^T$. Каждый участок полигона можно рассматривать как отрезок двумерной 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).

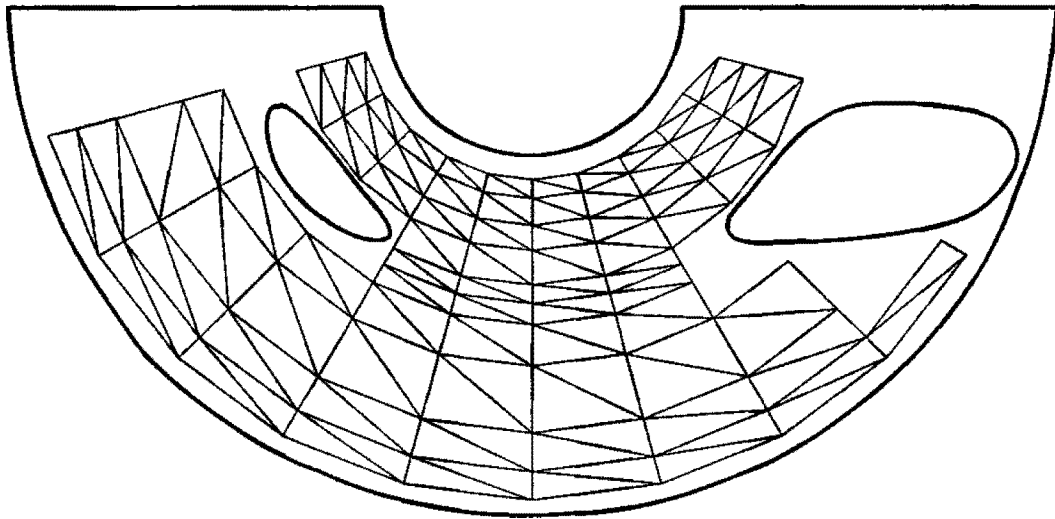


Рис. 9.7.6. Незаконченная триангуляция поверхности

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. 9.7.6.

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 AB найдена ближайшая вершина V и проверено, что отрезки BV и VA не пересекают другие граничные ребра. Тогда построим треугольник ABV и переведем ребро AB 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.

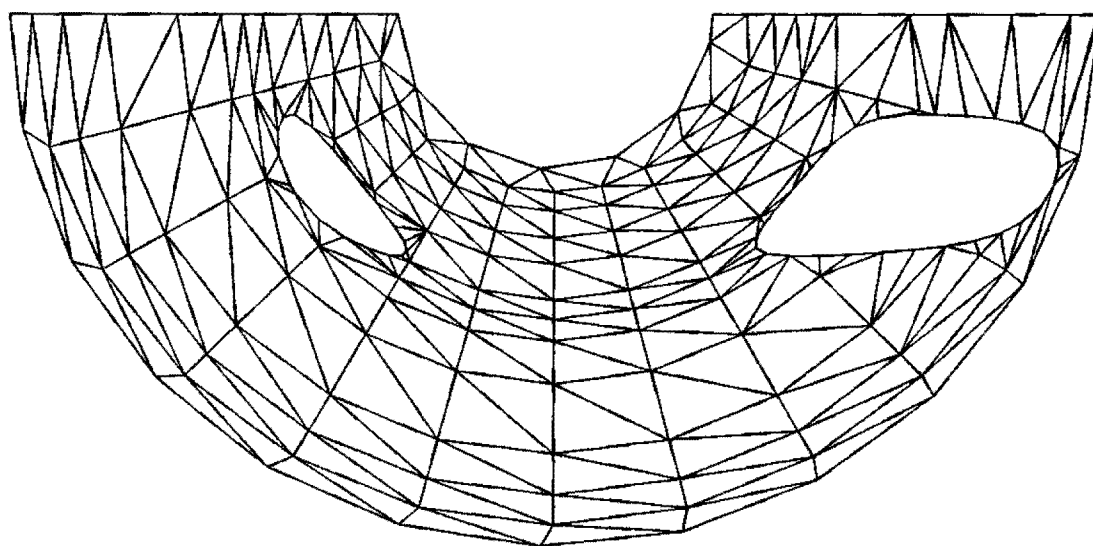


Рис. 9.7.7. Триангуляция методом коррекции

На рис. 9.7.7 приведена триангуляция поверхности методом коррекции треугольников в ячейках, пересеченных граничными контурами. На рис. 9.7.8 с помощью полученной триангуляции отображена сама поверхность.

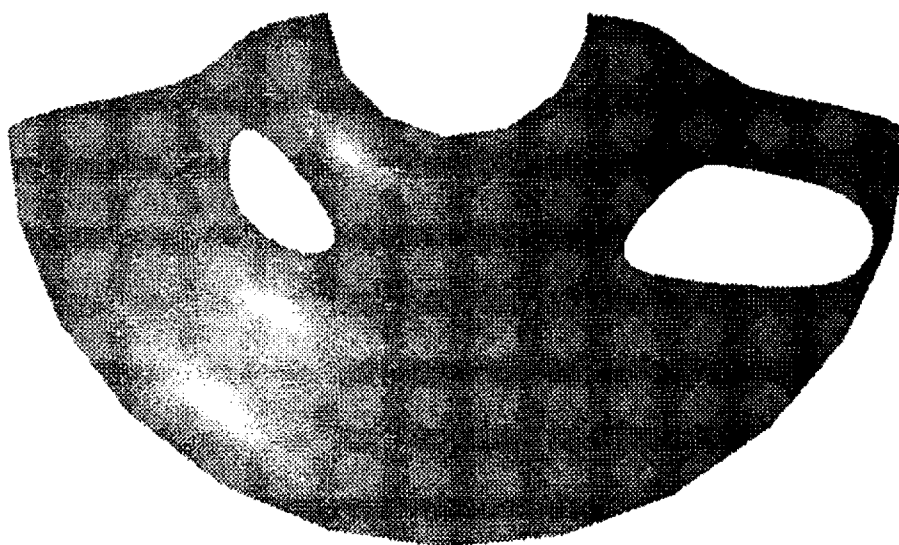


Рис. 9.7.8

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 и Δv . При известном направлении движения эти шаги определяются формулами (9.4.6). Приблизительно шаги по параметрам поверхности Δu и 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

$$\Delta u \approx 2 \frac{\sqrt{(\delta/m)(2(g_{11}/b_{11}) - \delta/m)}}{|\partial \mathbf{r} / \partial u|}, \quad \Delta v \approx 2 \frac{\sqrt{(\delta/m)(2(g_{22}/b_{22}) - \delta/m)}}{|\partial \mathbf{r} / \partial v|}, \quad (9.7.4)$$

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 и Δv . Этот эллипс имеет уравнение

$$\left(\frac{u - u_0}{\Delta u}\right)^2 + \left(\frac{v - v_0}{\Delta v}\right)^2 = 1. \quad (9.7.5)$$

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

$$u_1 = u_0 + \Delta u \cos \varphi, \quad v_1 = v_0 + \Delta v \sin \varphi. \quad (9.7.6)$$

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 и 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. 9.7.9). В противном случае построим два треугольника на данной, двух соседних и одной новой точках (рис. 9.7.11). Новая точка обозначена через P . Точку P будем искать на диагонали параллелограмма $BOCP$. Если вершина параллелограмма лежит внутри эллипса (рис. 9.7.10), 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 и v_p точки P определяются через координаты точек OBC

$$u_p = u_b + u_c - u_0, \quad v_p = v_b + v_c - v_0. \quad (9.7.7)$$

Угол отрезка OP с горизонталью определяется равенством

$$\operatorname{tg} \varphi = \frac{v_p - v_0}{u_p - u_0}. \quad (9.7.8)$$

Эти данные позволяют определить положение точки P относительно эллипса (9.7.5).

В случае, показанном на рис. 9.7.9, построим треугольник (запомним номера его вершин) и в рабочем полигоне удалим точку O . В случае, показанном на рис. 9.7.11, построим два треугольника и в рабочем полигоне точку O заменим точкой P и поместим последнюю в результирующий массив точек. На рис. 9.7.12 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.

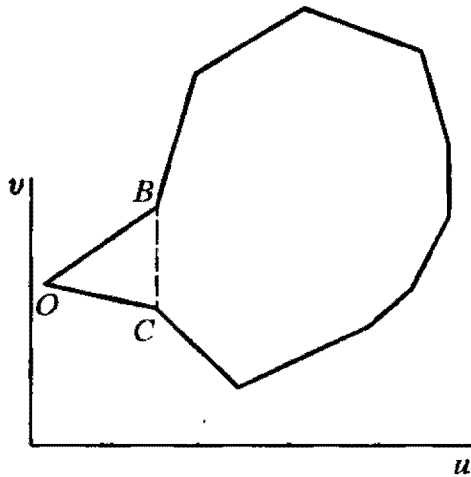


Рис. 9.7.9. Построение треугольника

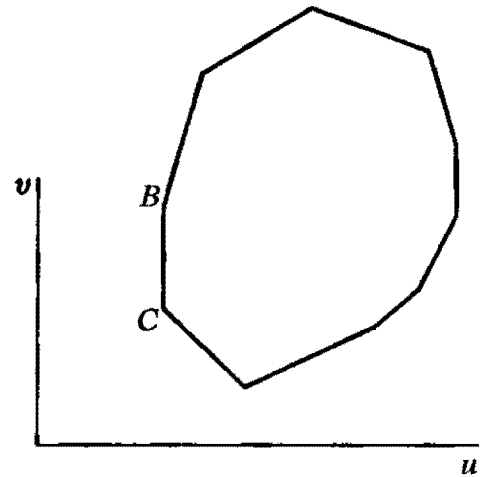


Рис. 9.7.10. Результирующий полигон

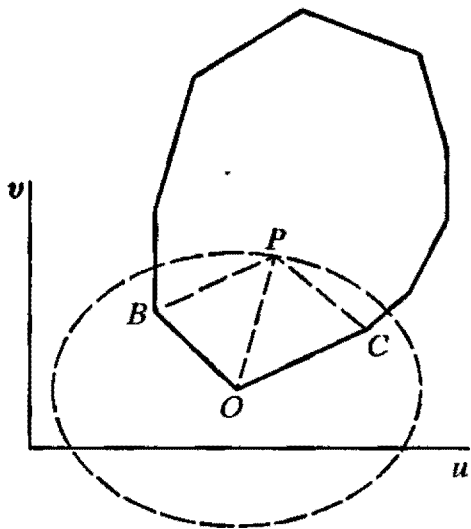


Рис. 9.7.11. Построение двух треугольников

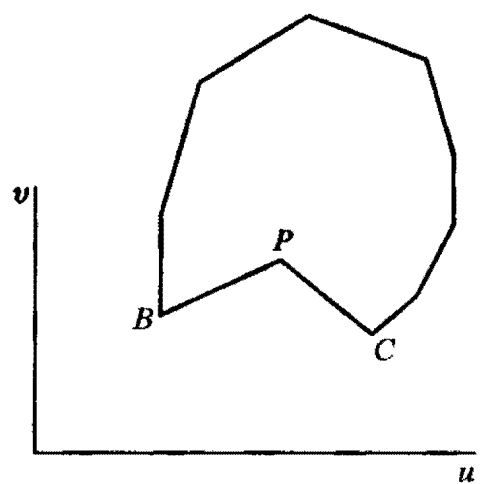


Рис. 9.7.12. Результирующий полигон

Такие ситуации показаны на рис. 9.7.13. Они могут возникнуть, когда стороны построенных треугольников пересекут несмежные с ними стороны рабочего полигона. Перед построением нового треугольника как в случае, показанном на рис. 9.7.9, так и в случае, показанном на рис. 9.7.11, должна быть выполнена проверка на отсутствие самопересечения результирующего полигона.

Более того, при определении положения точки P 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 O 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

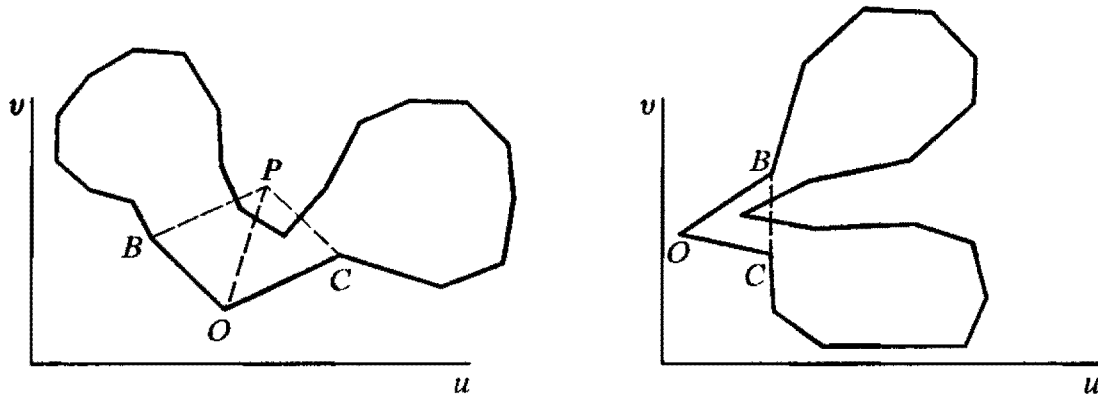


Рис. 9.7.13. В данном углу строить треугольники нельзя

треугольников и сузим полигон. 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. 9.7.11), 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. 9.7.14.

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 CEF 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

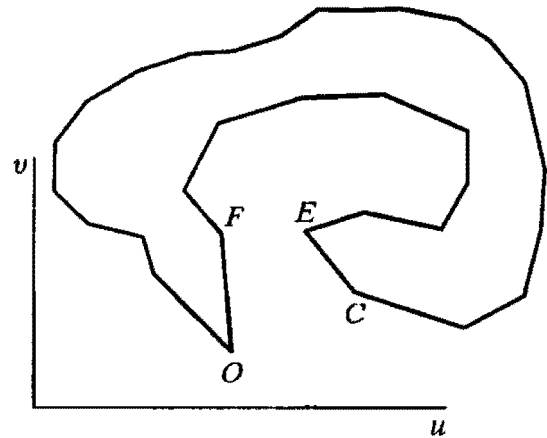
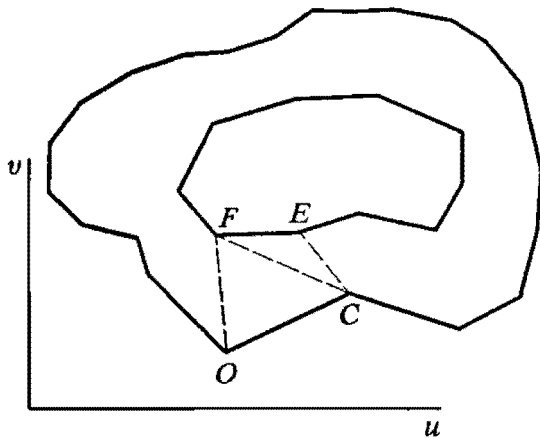


Рис. 9.7.14. Построение двух треугольников

Рис. 9.7.15. Слияние внешнего и внутреннего полигонов

отрезке EF и объединим их отрезками OF и EC . Результат слияния приведен на рис. 9.7.15. 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.

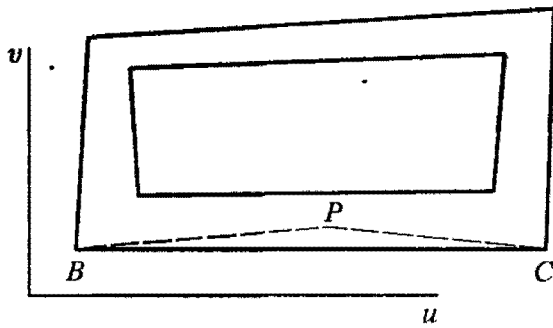


Рис. 9.7.16. В данном углу строить треугольники нельзя

Возможны ситуации, когда нельзя построить ни одного треугольника на заданных полигонах. На рис. 9.7.16 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 B и C полигона, для которых можно построить треугольник BSP . Точка P проецируется на середину стороны BC 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

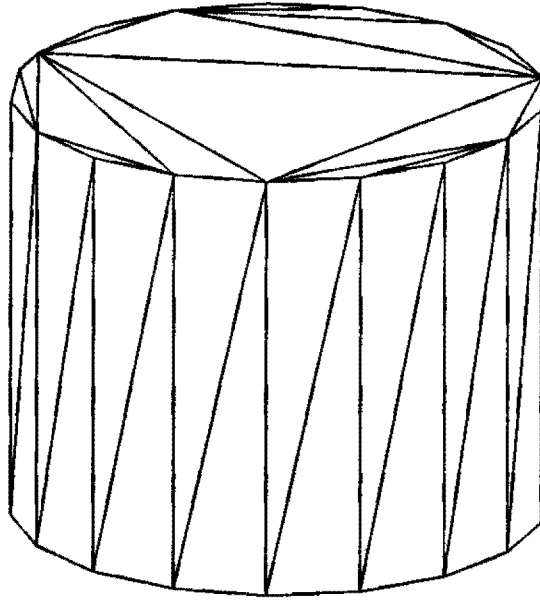


Рис. 9.7.17. Согласованность граничных полигонов граней тела

для смежных граней (рис. 9.7.17). 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. 6.5.1.

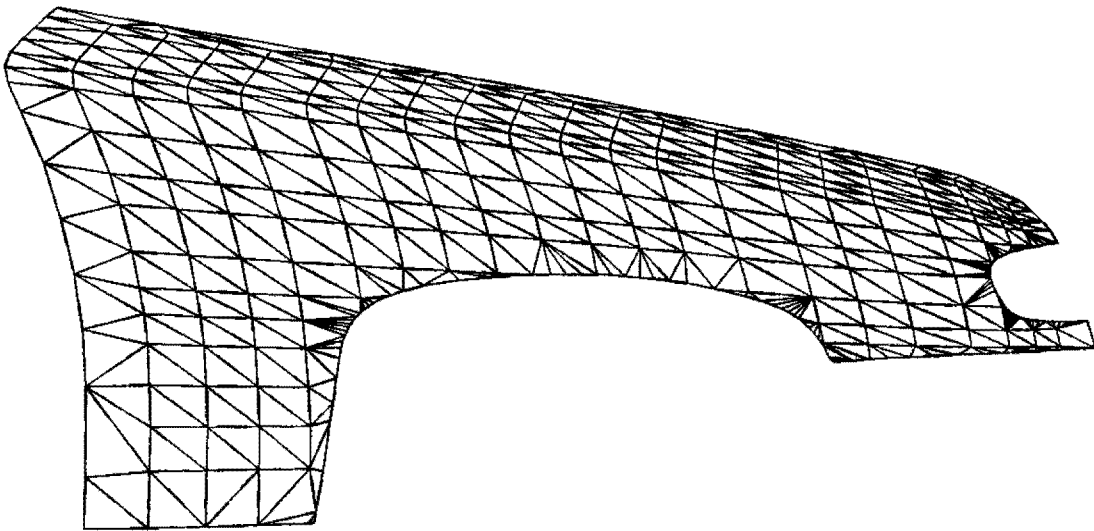


Рис. 9.7.18. Триангуляция грани тела методом коррекции

Разбиение области определения параметров поверхности на треугольники может быть использовано в интегралах (8.6.2), (8.6.3), (8.6.12), (8.7.17)–(8.7.22) при вычислении геометрических характеристик тел. При численном интегрировании параметрический шаг Δt для кривых следует вычислять по формуле

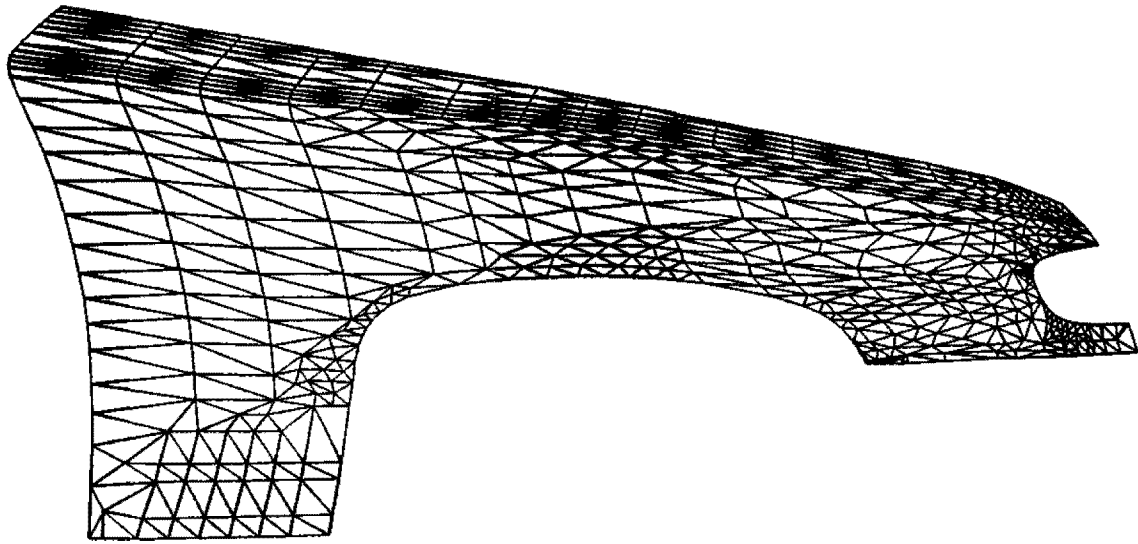


Рис. 9.7.19. Триангуляция грани тела методом поглощения

(8.11.5), а для поверхностей параметрические шаги Δu и Δv следует вычислять по формулам (8.11.1) и (8.11.2). Приближенно интеграл для треугольной области поверхности вычисляется по кубатурной формуле (8.10.13).

9.8. 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 светом поверхностью можно описать законом диффузного отражения, то поверхность называют *матовой*, если же большую часть отраженного светом поверхностью 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 I_a интенсивность рассеянного света. При наличии только рассеянного света в точку наблюдения от каждой точки объекта независимо от ориентации в ней поверхности придет свет интенсивности

$$I = k_a I_a, \quad (9.8.1)$$

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

$$I_d = k_d I_p \cos \theta, \quad (9.8.2)$$

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. 9.8.1). 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.

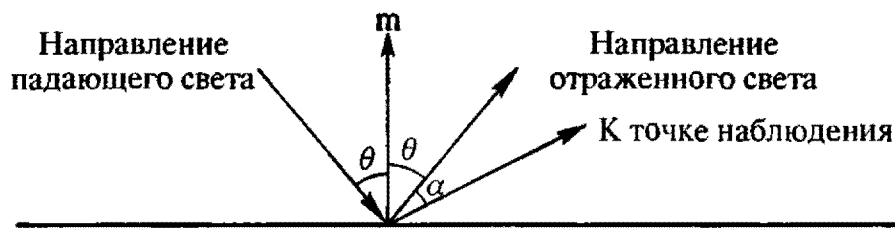


Рис. 9.8.1. Зеркальное отражение света

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

$$I_s = k_s I_p \cos^n \alpha, \quad (9.8.3)$$

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

$$I = I_t e^{-\mu l}, \quad (9.8.4)$$

где I_t — интенсивность света на входе в поглощающее вещество, l — 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_t 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

$$I_p = I_d + I_r + I_t = k_d I_p + k_s I_p + k_t I_p. \quad (9.8.5)$$

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. На-

правленно пропущенный свет претерпевает преломление на поверхности раздела двух сред. Преломленный луч лежит в одной плоскости с падающим лучом и нормалью в точке падения (рис. 9.8.2).

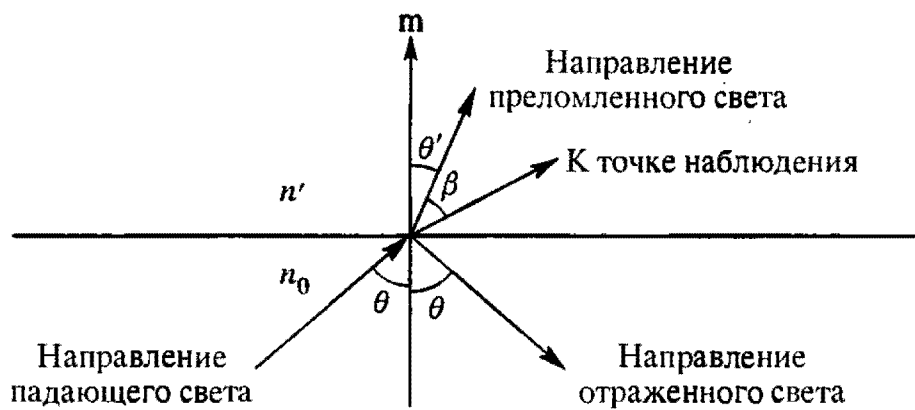


Рис. 9.8.2. Преломление света

Его направление составляет угол θ' с нормалью и описывается законом преломления

$$n_0 \sin \theta = n' \sin \theta', \quad (9.8.6)$$

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 (9.8.3) для направленно отраженного света

$$I_t = k_t I_p \cos^n \beta, \quad (9.8.7)$$

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

$$I = k_a I_a + I_p k_d \cos \theta + I_p k_s \cos^n \alpha. \quad (9.8.8)$$

Результаты закраски, полученные с помощью формулы (9.8.8), 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 от их расстояния 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

$$I = \frac{k_a I_a}{1 + k_r (r - r_0)} + I_p (k_d \cos \theta + k_s \cos^n \alpha), \quad (9.8.9)$$

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 (9.8.8). Интенсивность света от ближайшей точки также будет совпадать с (9.8.8). 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

$$I = k_a I_a + I_p \frac{k_d \cos \theta + k_s \cos^n \alpha}{1 + k_r (r - r_0)}, \quad (9.8.10)$$

Она также позволяет получать достаточно реалистичные изображения.

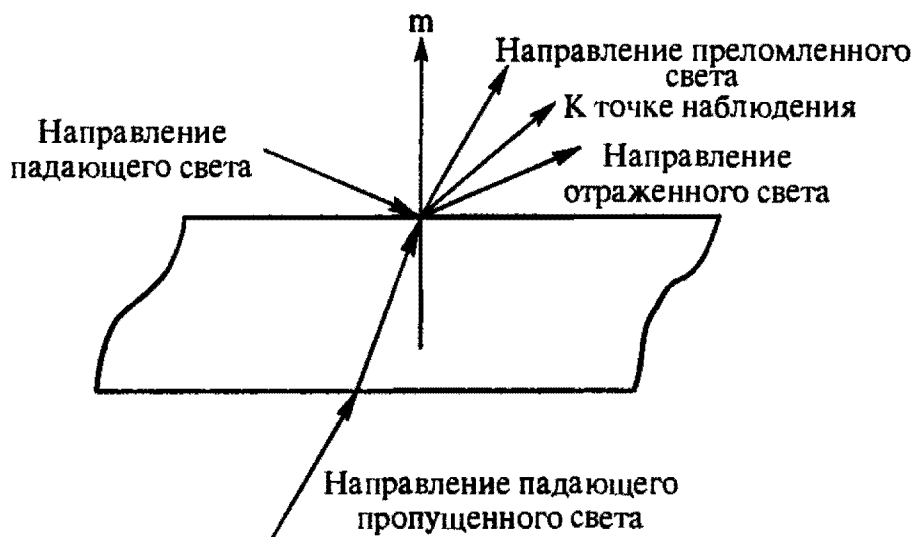


Рис. 9.8.3

Рассмотрим вычисление интенсивности света, пришедшего в точку наблюдения, от прозрачного объекта, показанного на рис. 9.8.3. Для построения реалистичных изображений интенсивность пришедшего от некоторой точки объекта света опишем формулой

$$\begin{aligned} I &= \frac{k_a I_a}{1 + k_r (r - r_0)} + I_p (k_d \cos \theta + k_s \cos^n \alpha) + I_p k_t \cos^n \beta = \\ &= \frac{k_a I_a}{1 + k_r (r - r_0)} + I_p (k_d \cos \theta + k_s \cos^n \alpha) + I_0 k_t^2 e^{-\mu l} \cos^n \beta. \end{aligned} \quad (9.8.11)$$

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 $I_p = k_t I_0 e^{-\mu l}$ и, 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 (9.8.9) примет вид

$$I = \frac{k_a I_a}{1 + k_r (r - r_0)} + \sum_i I_{pi} (k_d \cos \theta_i + k_s \cos^n \alpha_i). \quad (9.8.12)$$

Аналогичным образом изменится формула (9.8.11) вычисления интенсивности света, пришедшего от прозрачной поверхности. Она примет вид

$$I = \frac{k_a I_a}{1 + k_r (r - r_0)} + \sum_i I_{pi} (k_d \cos \theta_i + k_s \cos^n \alpha_i) + \sum_j I_{pj'} k_t \cos^n \beta_j, \quad (9.8.13)$$

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.

9.9. 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. Так,

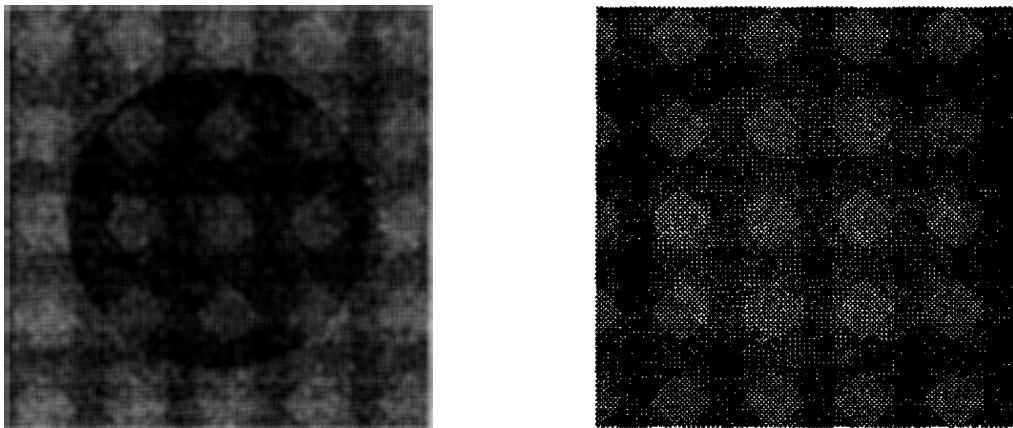


Рис. 9.9.1

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. 9.9.1.

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 меняющими 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 входящего 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

$$I = I_R + I_G + I_B. \quad (9.9.1)$$

Интенсивность каждого цвета определяется по одной из приведенных формул, в которой каждая цветовая компонента описывается своими коэффициентами, например,

$$\begin{aligned} I_R &= \frac{k_{aR} I_a}{1 + k_r(r - r_0)} + I_p(k_{dR} \cos \theta + k_{sR} \cos^n \alpha) + I_p' k_{tR} \cos^n \beta, \\ I_G &= \frac{k_{aG} I_a}{1 + k_r(r - r_0)} + I_p(k_{dG} \cos \theta + k_{sG} \cos^n \alpha) + I_p' k_{tG} \cos^n \beta, \\ I_B &= \frac{k_{aB} I_a}{1 + k_r(r - r_0)} + I_p(k_{dB} \cos \theta + k_{sB} \cos^n \alpha) + I_p' k_{tB} \cos^n \beta. \end{aligned} \quad (9.9.2)$$

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 (9.9.2) 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.

9.10. 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 треугольники переводятся в местную систему координат. Координаты вершин треугольников преобразуются по формуле (9.2.3). Компоненты векторов нормалей преобразуются по формуле

$$\begin{bmatrix} x_m \\ y_m \\ z_m \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & 0 \\ y_1 & y_2 & y_3 & 0 \\ z_1 & z_2 & z_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ 1 \end{bmatrix} = A_0 \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ 1 \end{bmatrix}, \quad (9.10.1)$$

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 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 (9.3.5), а координата z остается без изменений. Аналогичным образом уточним и нормали \mathbf{m} в вершинах треугольников: они должны быть повернуты на угол

$$\eta = \arccos \left(-\frac{(\mathbf{r} - \mathbf{w}) \cdot \mathbf{i}_z}{|\mathbf{r} - \mathbf{w}|} \right) \quad (9.10.2)$$

вокруг вектора

$$\mathbf{v} = -\frac{(\mathbf{r} - \mathbf{w}) \times \mathbf{i}_z}{|(\mathbf{r} - \mathbf{w}) \times \mathbf{i}_z|}, \quad (9.10.3)$$

где w — радиус-вектор точки наблюдения, r радиус-вектор точки поверхности, в которой вычислена нормаль, i_z — нормаль к проекционной плоскости. Поворот вектора m описывается формулой

$$m_p = (vv) \cdot m + \cos \eta (E - vv)m + \sin \eta v \times m. \quad (9.10.4)$$

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 двухмерной, что уменьшает объем вычислений. Для уменьшения количества просмотров проекции треугольников можно сгруппировать по отдельным областям.

На практике используются три основных способа закраски: *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 закраски 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 закраска. При однотонной закраске для каждого треугольника вычисляется уровень интенсивности отраженного света, который используется для закраски всего треугольника. При этом для простоты полагается, что θ и α являются постоянными в пределах своего треугольника. Фактически получается, что триангуляционная сетка является реальной моделируемой поверхностью, а не аппроксимацией последней. Однотонная закраска дает граненое изображение кривых поверхностей (рис. 9.10.1).

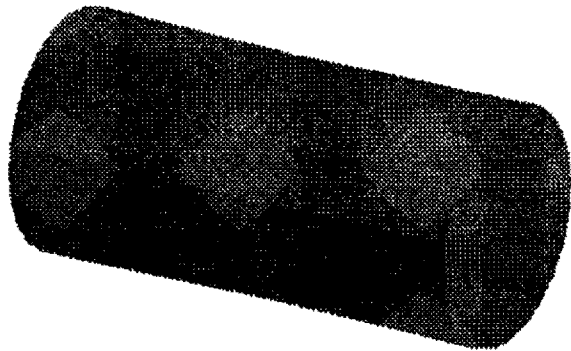


Рис. 9.10.1. Однотонная закраска

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

$$I(a, b, c) = aI_a + bI_b + cI_c, \quad (9.10.5)$$

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

$$\mathbf{m}(a, b, c) = a\mathbf{m}_a + b\mathbf{m}_b + c\mathbf{m}_c, \quad (9.10.6)$$

где a, b, c — барицентрические координаты произвольной точки $\mathbf{p} = [u \ v]^\top$ треугольника на параметрической плоскости, которые вычисляются по формулам (3.11.4). Несмотря на то, что векторы $\mathbf{m}_a, \mathbf{m}_b, \mathbf{m}_c$ нормированы, вектор $\mathbf{m}(a, b, c)$, в общем случае, не имеет единичной длины и его следует нормировать. На рис. 9.10.2 приведен треугольник и нормали в его вершинах.

Значение радиус-вектора точки треугольника определяется формулой

$$\mathbf{p}(a, b, c) = a\mathbf{p}_a + b\mathbf{p}_b + c\mathbf{p}_c. \quad (9.10.7)$$

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 (9.10.6) нормали плавно меняют свое направление при переходе от одного треугольника к другому, поэтому на тоновых изображениях поверхность выглядит

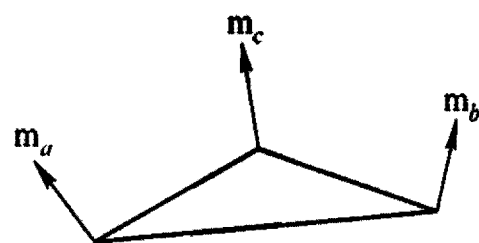


Рис. 9.10.2. Направление нормалей треугольника

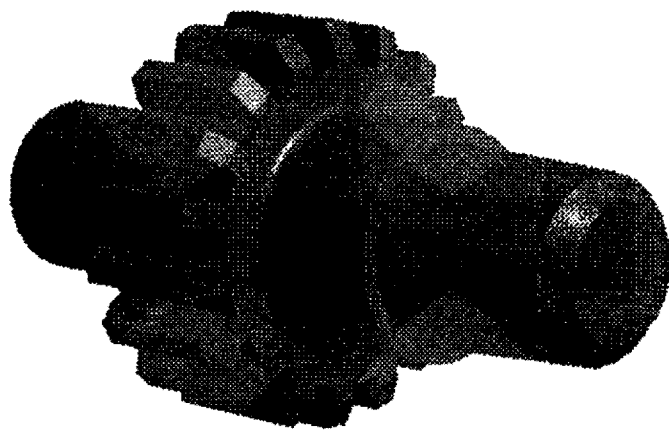


Рис. 9.10.3. Закраска методом Фонга

гладкой. Метод Фонга полностью устраняет полосы Маха, однако, значительно увеличивает объем вычислений. На рис. 9.10.3 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. 9.10.4 и 9.10.5 приведены тела, нормальям поверхностей которых сообщены случайные отклонения. На рис. 9.10.4 случайные отклонения нормали от исходного положения достигают 10° , а на рис. 9.10.5 эти отклонения достигают 60° .

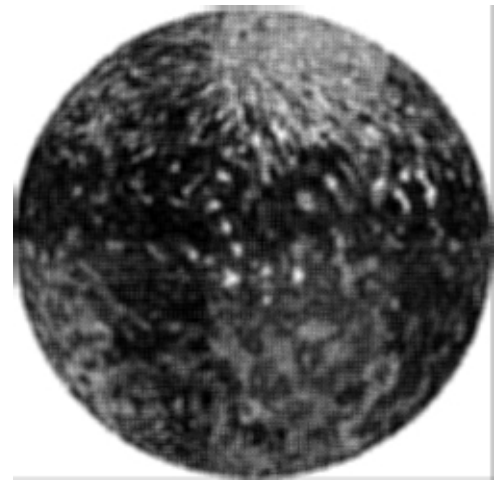


Рис. 9.10.4. Неровности на поверхности тора

Рис. 9.10.5. Неровности на поверхности сферы

На поверхность моделируемого объекта может быть нанесен рисунок, подобный рисунку на стенных 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

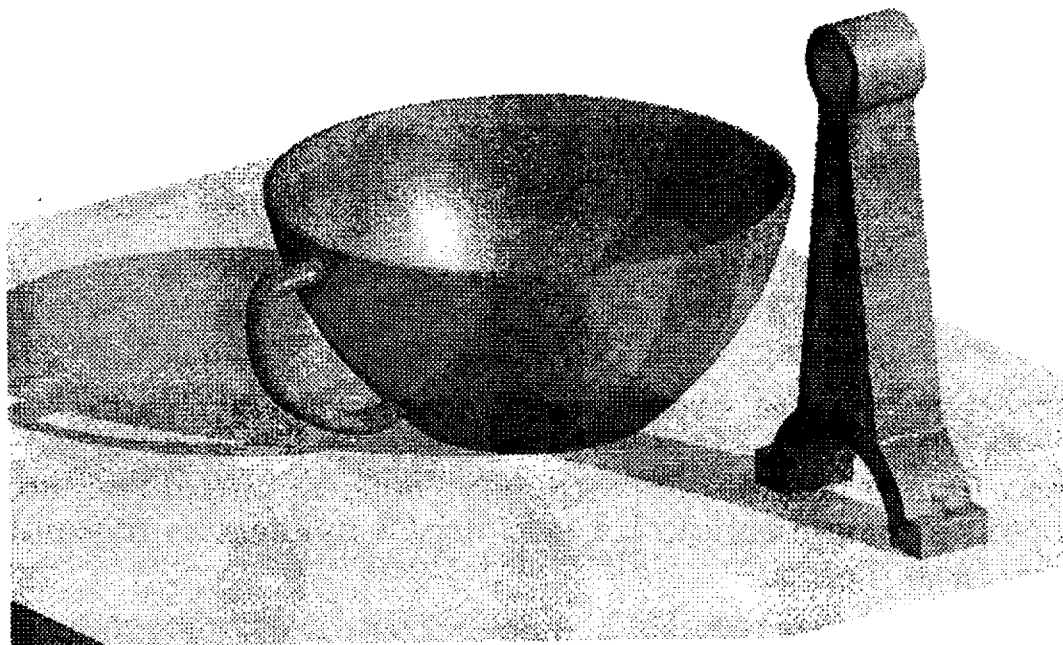


Рис. 9.10.6. Объекты и их тени

интенсивности источника I_p . 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. 9.10.6 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. 9.10.7 и 9.10.8 приведены изображения одного



Рис. 9.10.7. Непрозрачный объект

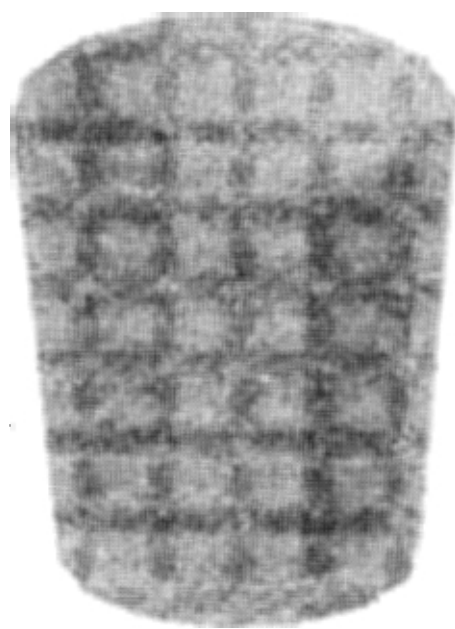


Рис. 9.10.8. Прозрачный объект

и того же объекта с различной прозрачностью материала. Материал объекта, приведенного на рис. 9.10.7, непрозрачный. Материал объекта, приведенного на рис. 9.10.8, 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 рутинной и сосредоточиться на творчестве.

СПИСОК ЛИТЕРАТУРЫ

1. 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).
2. Березин И.С., Жидков Н.П. Методы вычислений. Т. 1, 2.—М.: Наука, 1962, 1966.
3. Будаков Б.М., Фомин С.В. Кратные интегралы и ряды.—М.: Наука, 1967.
4. Гардан И., Люка М. Машинная графика и автоматизация конструирования.—М.: Мир, 1987 (Techniques Graphiques Interactives et C.A.O./par Michel Lucas et Yvon Gardan.—France: Hermes Publishing, 1983).
5. Гильберт Д., Кон-Фоссен С. Наглядная геометрия.—М.: Наука, 1981 (Hilbert D., Cohn-Vossen S. Anschauliche Geometrie.—Berlin: 1932).
6. Де Бор К. Практическое руководство по сплайнам.—М.: Радио и связь, 1985 (De Boor C. A Practical Guide to Splines.—Berlin: Springer, 1979).
7. Демидович Б.П., Марон И.А. Основы вычислительной математики.—М.: Наука, 1970.
8. Дубровин Б.А., Новиков С.П., Фоменко А.Т. Современная геометрия.—М.: Наука, 1986.
9. Ильин В.А., Позняк Э.Г. Аналитическая геометрия.—М.: Наука, 1981.
10. Ильин В.А., Позняк Э.Г. Линейная алгебра.—М.: Наука, 1984.
11. Калиткин Н.Н. Численные методы.—М.: Наука, 1978.
12. Курош А.Г. Курс высшей алгебры.—М.: Наука, 1971.
13. Лаврентьев М.А., Шабат Б.В. Методы теории функций комплексного переменного.—М.: Наука, 1987.
14. Ласло М. Вычислительная геометрия и компьютерная графика на C++.—М.: Бином, 1997 (Laszlo M.J. Computational Geometry and Computer Graphics in C++.—Prentice Hall, 1996).
15. Лойцянский Л.Г., Лурье А.И. Курс теоретической механики. Т. 1, 2.—М.: Наука, 1982, 1983.
16. Препарата Ф., Шеймос М. Вычислительная геометрия.—М.: Мир, 1989 (Preparata F.P., Shamos M. Computational Geometry: An Introduction.—New York, Berlin, Tokyo: Springer-Verlag, 1985).
17. Рашевский П.К. Курс дифференциальной геометрии.—М-Л.: Физматгиз, 1950.
18. Рашевский П.К. Риманова геометрия и тензорный анализ.—М.: Наука, 1967.
19. Роджерс Д., Адамс Дж. Математические основы машинной графики.—М.: Машиностроение, 1980 (Rogers D.F., Adams J.A. Mathematical Elements for Computer Graphics.—McGraw-Hill, 1976).
20. Розенфельд А. Распознавание и обработка изображений с помощью вычислительных машин.—М.: Мир, 1972.
21. Самарский А.А., Гулин А.В. Численные методы.—М.: Наука, 1989.
22. Сокольников И.С. Тензорный анализ.—М.: Наука, 1971.
23. Стечкин С.Б., Субботин Ю.Н. Сплайны в вычислительной математике.—М.: Наука, 1976.
24. Тихомиров Ю. Программирование трехмерной графики.—СПб.: ВHV-Санкт-Петербург, 1998.
25. Фаворин М.В. Моменты инерции тел.—М.: Машиностроение, 1977.

26. Феодосьев В. И. Сопротивление материалов.—М.: Наука, 1972.
27. Фокс А., Пратт М. Вычислительная геометрия.—М.: Мир, 1982 (Faux I. E., Pratt M. J. Computational Geometry for Design and Manufacture.—Ellis Horwood, Chichester, 1979).
28. Фоли Дж., Вен Дем А. Основы интерактивной машинной графики. Т. 1, 2.—М.: Мир, 1985 (James D. Foley, Andries Van Dam. Fundamentals Interactive Computer Graphics.—Massachusetts, Menlo Park, California, London, Amsterdam, Don Mills, Ontario, Sydney: Addison-Wesley Publishing Company Reading, 1982).
29. Фоменко А. Т. Наглядная геометрия и топология. Математические образы в реальном мире.—М.: Изд-во Моск. ун-та, Изд-во ЧеРо, 1998.
30. Эльсгольц Л. Э. Дифференциальные уравнения и вариационное исчисление.—М.: Наука, 1969.
31. Bezier P. Numerical control: Mathematics and application.—N.Y.: Wiley, 1972.
32. Barnhill E. R., Bohm W. Surfaces in CAGD.—Amsterdam: North-Holland, 1983.
33. Bohm W. A. Survey of Curve and Surface Methods in CAGD.—Computer Aided Geometric Design.—Amsterdam: North-Holland, July 1984. V. 1, № 1.
34. Christoph M. Hoffman. Geometric and Solid Modeling.—San Ateo (California): Morgan Kaufmann Publishers Inc., 1989.
35. Coons S. A. Modification of the Shape of Piecewise Curves // Computer Aided Design. 1977. 9, 3. P. 178–180.
36. Coons S. A. Surfaces for Computer Aided Design of Space Form // Report MAC-TR-41, Project MAC, M.I.T., 1967.
37. Cox M. G. An Algorithm for Spline Interpolation // J. Inst. Math-Applics. 1975. 15.
38. Cox M. G. The Numerical Evaluation of a Spline from its B-spline Representation.—Middlesex, 1976.
39. Dahmen W., Micchelli C. A., Seidel H-P. Blossoming Begests B-splines Built Better by B-patches // Mathematics of Computation. July, 1992. 59(199). P. 97–115.
40. De Boor C. On Calculating with B-splines // J. Approx. Theory. 1972. 6.
41. De Boor C. Package for Calculating with B-splines // SIAM J. Number. Anal. 1977. 14, 3.
42. De Casteljaeu F. Outillage methodes calcul.—Paris: Andre Citroen Automobiles SA, 1959.
43. De Casteljaeu F. Courbes et surfaces a poles.—Paris: Andre Citroen Automobiles SA, 1963.
44. Gordon W. J., Riesenfeld R. F. B-spline Curves and Surfaces // Computer Aided Geometric Design.—Academic Press, 1974.
45. Ferguson J. C. Multivariate Curve Interpolation // Journal ACM. 1964. II, 2. P. 221–228.
46. Forrest A. R. Mathematical Principles for Curve and Surface Representation. Curved Surfaces in Engineering.—IPC Science and Technology Press Ltd.
47. Liming R. A. Practical Analitical Geometry with Applications to Aircraft.—New York: Macmillan, 1944.
48. Richard H. Bartels, John C. Beatty, Brian A. Barsky. An Introduction to Splines for use in Computer Graphics and Geometric Modeling.—San Mateo, California: Morgan Kaufmann Publishers Inc., 1987.
49. Schoenberg I. J. Cardinal Spline Interpolation.—Philadelphia: SIAM, 1973.