

КОМПЬЮТЕРНАЯ ГЕОМЕТРИЯ И АЛГОРИТМЫ МАШИНОЙ ГРАФИКИ

- *Геометрические основы компьютерной графики*
- *Аффинные и проективные преобразования*
- *Математические модели поверхностей и объектов*
- *Геометрические задачи визуализации*



Е. А. Никулин

КОМПЬЮТЕРНАЯ ГЕОМЕТРИЯ И АЛГОРИТМЫ МАШИННОЙ ГРАФИКИ

Санкт-Петербург

«БХВ-Петербург»

2003

УДК 681.3.06
ББК 32.973
Н65

Никулин Е. А.

Н65 Компьютерная геометрия и алгоритмы машинной графики. — СПб.: БХВ-Петербург, 2003. — 560 с.: ил.

ISBN 5-94157-264-6

В книге приведено максимально полное изложение геометрических и алгоритмических основ современной компьютерной графики: математические модели графических элементов на плоскости и в пространстве, фундаментальные законы геометрической оптики и основанные на них алгоритмы построения оптических эффектов, методы геометрических преобразований, анализа и синтеза моделей линий, поверхностей и объектов, геометрические задачи визуализации — комплекс алгоритмов 2d- и 3d-отсечения и удаления. Материал сопровождается большим числом иллюстраций, блок-схем алгоритмов и примеров их реализации.

Для студентов вузов и специалистов, изучающих компьютерную графику и занимающихся разработкой новых алгоритмов и прикладных графических программ

УДК 681.3.06
ББК 32.973

Группа подготовки издания:

Главный редактор	<i>Екатерина Кондукова</i>
Зав. редакцией	<i>Григорий Добин</i>
Редактор	<i>Нина Седых</i>
Компьютерная верстка	<i>Натальи Смирновой</i>
Корректор	<i>Наталья Першакова</i>
Дизайн обложки	<i>Игоря Цырульникова</i>
Зав. производством	<i>Николай Тверских</i>

Рецензент:

кафедра вычислительной техники Нижегородского государственного технического университета (зав. кафедрой член-корреспондент РАН, профессор В. В. Кондратьев)

Лицензия ИД № 02429 от 24.07.00. Подписано в печать 20.03.03.

Формат 70×100¹/₁₆. Печать офсетная. Усл. печ. л. 45,15.

Тираж 3000 экз. Заказ №

"БХВ-Петербург", 198005, Санкт-Петербург, Измайловский пр., 29.

Гигиеническое заключение на продукцию, товар № 77.99.02.953.Д.001537.03.02 от 13.03.2002 г. выдано Департаментом ГСЭН Минздрава России.

Отпечатано с готовых диапозитивов
в Академической типографии "Наука" РАН
199034, Санкт-Петербург, 9 линия, 12.

ISBN 5-94157-264-6

© Никулин Е. А., 2003

© Оформление, издательство "БХВ-Петербург", 2003

Содержание

Введение.....	1
Основные обозначения и соотношения.....	5
Глава 1. Геометрические основы компьютерной графики.....	9
1.1. Графические элементы на плоскости	12
1.1.1. Модели прямой линии на плоскости.....	13
1.1.1.1. неявное уравнение прямой.....	13
1.1.1.2. Нормальное уравнение прямой	15
1.1.1.3. Параметрическая функция прямой.....	16
1.1.1.4. Уравнения прямой, проходящей через две точки	17
1.1.1.5. Уравнения прямой в отрезках	18
1.1.2. Взаимное положение графических элементов на плоскости	18
1.1.2.1. Коллинеарность точек	18
1.1.2.2. Взаимное расположение прямых	19
1.1.2.3. Взаимное расположение точки и прямой.....	19
1.1.2.4. Построение прямой, наименее удаленной от совокупности точек.....	21
1.1.2.5. Пересечение двух прямых	24
1.1.2.6. Уравнения пучка прямых и биссектриса угла	25
1.1.2.7. Тесты свойств графических элементов на плоскости	27
1.1.2.8. Тесты ориентации точки относительно полигона	32
1.1.2.9. Алгоритмы пересечения на плоскости	42
1.1.2.10. Площадь и геометрический центр полигона	48
1.1.2.11. Алгоритмы генерирования случайных полигонов	51
1.1.3. Квадратичные и параметрические кривые	54
1.2. Графические элементы в пространстве.....	60
1.2.1. Модели плоскости в пространстве.....	62
1.2.1.1. неявное уравнение плоскости.....	62
1.2.1.2. Нормальное уравнение плоскости.....	63
1.2.1.3. Параметрическая функция плоскости.....	64
1.2.1.4. Уравнения плоскости, проходящей через три точки	66
1.2.1.5. Уравнения плоскости в отрезках	66
1.2.1.6. Модели линии в пространстве	67
1.2.2. Взаимное положение графических элементов в пространстве	69
1.2.2.1. Коллинеарность точек	69
1.2.2.2. Компланарность точек.....	69

1.2.2.3. Точка и прямая	70
1.2.2.4. Точка и плоскость	70
1.2.2.5. Построение плоскости, наименее удаленной от совокупности точек.....	71
1.2.2.6. Взаимное расположение двух прямых.....	73
1.2.2.7. Взаимное расположение прямой и плоскости	74
1.2.2.8. Две плоскости	75
1.2.2.9. Пучок плоскостей и биссекторная плоскость.....	76
1.2.2.10. Пересечение плоскостей.....	77
1.2.2.11. Модель полиэдра	77
1.2.2.12. Тесты свойств графических элементов в пространстве	80
1.2.2.13. Тесты ориентации точки относительно полиэдра	83
1.2.2.14. Алгоритмы пересечения в пространстве	85
1.2.3. Квадратичные и параметрические поверхности	89
1.3. Основные задачи геометрической оптики.....	99
1.3.1. Пересечение луча с поверхностью	100
1.3.2. Отражение луча от поверхности.....	106
1.3.3. Преломление луча на поверхности.....	107
1.3.4. Прямая и обратная трассировка лучей	110
1.3.5. Лучевые методы построения оптических эффектов	112
1.3.5.1. Тень	116
1.3.5.2. Отражение	121
1.3.5.3. Преломление	128

Глава 2. Геометрические преобразования 139

2.1. Аффинные преобразования.....	140
2.1.1. Основные понятия и соотношения	140
2.1.2. Элементарные аффинные преобразования.....	144
2.1.2.1. Перенос.....	144
2.1.2.2. Масштабирование	144
2.1.2.3. Сдвиг	145
2.1.2.4. Вращение	148
2.1.2.5. Табличный расчет тригонометрических функций.....	149
2.1.3. Сложные аффинные преобразования.....	154
2.1.3.1. Методы расчета матрицы сложного преобразования.....	155
2.1.3.2. Кинематический метод построения объектов	170
2.1.3.3. Кинематическая задача перемещения в пространстве	182
2.2. Проективные преобразования	194
2.2.1. Ортографические проекции	196
2.2.2. Аксинометрические проекции	197
2.2.3. Косоугольные проекции	207
2.2.4. Центральные (перспективные) проекции	211
2.2.5. Проективные алгоритмы сложных преобразований.....	221
2.2.5.1. Проецирование пространственных линий на плоскость	223
2.2.5.2. Стереографические проекции	228

2.2.5.3. Картографические проекции.....	231
2.2.5.4. Построение сцены с подвижным наблюдателем	242
2.2.5.5. Проективные алгоритмы построения оптических эффектов.....	247

Глава 3. Математические модели поверхностей и объектов 261

3.1. Методы моделирования поверхностей	261
3.1.1. Методы изображения поверхностей	262
3.1.1.1. Выбор проекции изображения	263
3.1.1.2. Каркасные поверхности.....	264
3.1.1.3. Точечные поверхности.....	268
3.1.1.4. Модели освещенности и закрашивание поверхностей.....	271
3.1.2. Кинематические поверхности	278
3.1.2.1. Поверхности вращения, переноса и комбинированные.....	282
3.1.2.2. Линейчатые поверхности и их развертки	289
3.1.2.3. Нелинейчатые поверхности.....	307
3.1.3. Кусочно-определенные поверхности	324
3.1.4. Сплайны	329
3.1.4.1. Сплайновые кривые	330
3.1.4.2. Сплайновые поверхности	339
3.1.5. Фрактальные множества	347
3.1.5.1. Фрактал Мандельброта и алгоритмические фракталы.....	348
3.1.5.2. Геометрические фракталы.....	353
3.1.5.3. Свойства фракталов.....	370
3.1.6. Графические поверхности	376
3.2. Модели объектов в пространстве.....	379
3.2.1. Каркасные модели. Платоновы тела.....	381
3.2.2. Граничные модели	393
3.2.3. Сплошные модели	395

Глава 4. Геометрические задачи визуализации 405

4.1. Логические операции со списками	405
4.1.1. Объединение списков отрезков.....	408
4.1.2. Пересечение списков отрезков.....	411
4.1.3. Исключение списков отрезков	413
4.2. Методы отсечения	416
4.2.1. Регулярное плоское отсечение отрезка	420
4.2.2. Произвольное плоское отсечение отрезка.....	423
4.2.3. Произвольное плоское отсечение полигона.....	429
4.2.4. Объемное отсечение отрезка.....	432
4.2.5. Объемное отсечение полигона и полиэдра.....	434
4.2.6. Логическое конструирование 3d-объектов.....	442
4.2.7. Дополнительные задачи отсечения на плоскости.....	448
4.2.7.1. Отсечение выпуклого полигона полуплоскостью.....	448
4.2.7.2. Расчет ядра произвольного полигона	452
4.2.7.3. Пересечение выпуклых полигонов	453

4.2.7.4. Отсечение проекции выпуклого полигона.....	454
4.2.7.5. Выпуклая полигональная оболочка массива точек.....	461
4.2.7.6. Полигонализация массива точек.....	464
4.2.7.7. Разрезание невыпуклого полигона.....	468
4.2.7.8. Триангуляция полигона.....	472
4.2.8. Дополнительные задачи отсечения в пространстве.....	484
4.2.8.1. Отсечение выпуклого полиэдра полупространством.....	484
4.2.8.2. Сечение выпуклого полиэдра плоскостью.....	493
4.2.8.3. Расчет ядра полиэдра.....	495
4.2.8.4. Пересечение выпуклых полиэдров.....	496
4.2.8.5. Выпуклая полиэдральная оболочка массива точек.....	498
4.3. Методы удаления.....	504
4.3.1. Предварительная обработка моделей объектов.....	509
4.3.1.1. Выбор мировой системы координат.....	509
4.3.1.2. Построение охватывающих оболочек.....	513
4.3.1.3. Разбиение граней.....	523
4.3.1.4. Отбраковка нелицевых граней и нормировка векторов.....	530
4.3.2. Удаление невидимых линий.....	533
4.3.3. Удаление невидимых граней.....	539
Заключение.....	545
Список литературы.....	549

Введение

В основании бесконечно развивающегося здания компьютерной графики краеугольными камнями лежат фундаментальные дисциплины — аналитическая геометрия и оптика, скрепленные раствором — искусством программирования. Возникнув из потребностей рынка, развития информатики и вычислительной техники, компьютерная графика изучает методы построения изображений различных геометрических объектов и сцен.

Главными этапами построения изображения являются:

- *моделирование* как искусство применения методов математического описания объектов и сцен, природа которых может быть самой различной: обычные геометрические фигуры и тела в двух- и трехмерном пространстве; естественные явления природы — горы, деревья, облака и другие объекты; огромные массивы чисел, полученных в различных экспериментах, и многое другое;
- *визуализация* (отображение) как искусство построения реалистичных изображений объемного мира на плоском экране дисплея ЭВМ, заключается в преобразовании моделей объектов и сцен в статическое изображение или фильм — последовательность статических кадров.

В отличие от плоских изображений реальных трехмерных объектов, которые на фотографиях автоматически выглядят натурально благодаря действию в природе оптических законов, синтезированное компьютерное изображение будет похоже на реальное лишь при мастерском владении как методами моделирования геометрических форм, так и средствами их достоверной визуализации. Таким образом, компьютерную графику можно рассматривать как искусство создания реалистичной иллюзии действительного мира. Эта сторона ее применения нашла наиболее яркое воплощение в системах виртуальной реальности — комплексах аппаратных и программных средств имитации окружающей среды с помощью визуальных, акустических, тактильных и других эффектов.

С другой стороны, методами компьютерного синтеза стало возможным создание и нереальных, фантастических изображений и эффектов, без чего не обходится ни одна популярная компьютерная игра или кассовая голливудская кинолента. Подлинное второе дыхание компьютерная графика придала рекламному бизнесу благодаря умению создавать анимированные, т. е. оживленные, персонажи рекламируемых товаров.

Из более серьезных областей применения компьютерной графики можно назвать:

- системы автоматизированного проектирования (САПР), в которых осуществляется интерактивное взаимодействие конструктора и синтезированного с помощью компьютера изображения создаваемого изделия или сооружения;

- автоматизированные системы научных исследований (АСНИ), в которых осуществляется визуализация результатов экспериментов в виде трехмерных статических или динамических изображений, интерпретирующих огромные массивы первичных данных;
- распознавание и обработка изображений в системах искусственного зрения, авиационной и космической картографии, медицинской томографии и т. д. В качестве исходного материала могут быть отсканированные изображения, космические снимки, данные локации и т. п. Компьютерное преобразование этих данных в зримые геометрические образы помогает достаточно просто понимать суть происходящих процессов людям, далеким от вычислительной техники.

Таким образом, интерес к машинной графике проявляют представители самых различных специальностей: программисты компьютерных игр, конструкторы, технологи, физики, биологи, медики, дизайнеры, кинорежиссеры, художники-мультипликаторы, клипмейкеры и др. Неудивительно поэтому, что постоянно ощущается острый дефицит литературы, раскрывающей возможности применения научных методов в современной компьютерной графике. К такого рода изданиям и относится настоящая книга. Наиболее близким к ней по духу является фундаментальный тандем книг [25, 26] по основам машинной графики. К удачным изданиям последнего времени хотелось бы отнести также [5, 18, 23] и особенно [36].

Глубоко уважаемый автор последней книги, как и большинство других авторов, отстаивает "подход программиста" при овладении знаниями в области компьютерной графики. Цель его книги — научить приемам решения простейших задач трехмерной графики, не останавливаясь на детальном исследовании излагаемых алгоритмов и оптимальных методов их реализации.

Конечно, можно успешно водить автомобиль, не зная в деталях, что происходит под капотом. Но компьютерная графика — не массовое увлечение, а удел специалистов своего дела, знающих предмет изнутри. Основное содержание настоящей книги составляет детальное изложение **методов и алгоритмов** реалистичного изображения как статических, так и динамических объектов и сцен. Автор глубоко убежден в том, что гораздо интереснее и важнее понимать и уметь самому реализовать скрытые закономерности изображения сложных объектов, чем с восхищением наблюдать, как эти изображения создаются чужими, пусть и более совершенными графическими программами.

Автор сознательно не приводит текстов программ на каком-либо алгоритмическом языке, т. к. мода на языки программирования преходяща (сохранившиеся кое-где книги с программами на раритетных ныне языках Basic, Fortran и Algol [20, 24, 31] не вызывают ничего, кроме умиления), стиль работы каждого программиста уникален, а алгоритмы, изложенные на понятном для всех языке формул и блок-схем, универсальны и существенно более долговечны. Книга призвана вооружить читателя удочкой — методами, алгоритмами и особенностями их реализации, а не готовой рыбкой — текстами программ; побудить его работать своей головой, а не использовать чужие решения.

По аналогичной причине книга не является инструкцией пользователя какого-нибудь популярного графического пакета. Более того, она не содержит ни одного названия подобных программ, простое упоминание которых лет через пять вызовет лишь снисходительную улыбку читателя по причине их безнадёжной древности. Единственное исключение сделано уникальной программе MathCAD [22], которую автор использовал для создания геометрически достоверных рисунков и решения многочисленных примеров, иллюстрирующих работу алгоритмов. И снова читатель будет лишен удовольствия сравнить свое искусство программирования в среде MathCAD с авторским, поскольку в книге сознательно не приведено ни одного листинга конкретной программы.

Этому есть еще одно объяснение. Курс "Компьютерная графика", читаемый автором на протяжении ряда лет в Нижегородском государственном техническом университете, предполагает выполнение лабораторных работ, где излагаемые в книге алгоритмы программируются в среде MathCAD. Для того чтобы доставить студентам удовольствие от самого процесса получения самостоятельного решения, тексты программ не включены в содержание книги.

Автор выражает благодарность студентам факультета информационных систем и технологий Нижегородского государственного технического университета за роль подопытных кроликов, которую они доверчиво играли в процессе изучения курса, являясь одновременно экспертами методов и алгоритмов, излагаемых в книге. Нет границ признательности автора к членам своей семьи, в течение ряда лет терпеливо наблюдавшим спину своего мужа и отца, работающего за компьютером над созданием и бесчисленными исправлениями этой книги. Спасибо всем неравнодушным за моральную поддержку!

Основные обозначения и соотношения

☞ — символ, обозначающий начало алгоритма или примера.

☐ — символ, обозначающий конец алгоритма или примера.

\mathbb{R}^n — евклидово пространство размерности n .

x^0 , y^0 и z^0 — ортогональные векторы единичной длины (*орты*), направленные по осям декартовой системы координат xuz .

f, **h** и **p** — фронтальная (xu), горизонтальная (xz) и профильная (yz) *координатные плоскости* декартовой системы координат.

$V = [V_x \ V_y] = V_x x^0 + V_y y^0 \in \mathbb{R}^2$ — вектор в плоскости xu с координатами V_x , V_y

и модулем (длиной) $|V| = \sqrt{V_x^2 + V_y^2}$.

$V = [V_x \ V_y \ V_z] = V_x x^0 + V_y y^0 + V_z z^0 \in \mathbb{R}^3$ — вектор в пространстве xuz с коор-

динатами V_x , V_y , V_z и модулем $|V| = \sqrt{V_x^2 + V_y^2 + V_z^2}$.

$O_n \in \mathbb{R}^n$ — нулевой вектор или блок вектора размерности n .

$O_{n \times m} \in \mathbb{R}^{n \times m}$ — нулевая матрица или блок матрицы размерности $n \times m$.

$\tilde{V} = [V \ 1] \in \mathbb{R}^{n+1}$ — расширенный вектор V , дополненный координатой, равной единице.

$\check{V} = [V \ 0] \in \mathbb{R}^{n+1}$ — расширенный вектор V , дополненный координатой, равной нулю.

$\bar{V} = \frac{V}{|V|}$ — нормированный вектор V единичной длины.

$V \parallel W$ и $V \# W$ — соответственно, параллельные и непараллельные векторы V и W .

$V \perp W$ и $V \perp\!\!\!\perp W$ — ортогональные и неортогональные векторы V и W .

$\angle VW$ или $\angle(V, W)$ — угол между векторами V и W .

$V \circ W = V_x \cdot W_x + V_y \cdot W_y + V_z \cdot W_z$ — скалярное произведение векторов V и W . Другое выражение для скалярного произведения $V \circ W = |V| \cdot |W| \cdot \cos(\angle VW)$ позволяет найти косинус угла между векторами V и W :

$$\cos(\angle VW) = \frac{V \circ W}{|V| \cdot |W|}.$$

$V \circ W = 0$ — тест ортогональности векторов V и W .

$\bar{V} \circ \bar{W} = \pm 1$ — тесты параллельности (при $+1$) и антипараллельности (при -1) векторов V и W .

$\text{pr}_W V = |V| \cdot \cos(\angle VW) \bar{W} = \frac{V \circ W}{|W|^2} W = (V \circ \bar{W}) \bar{W}$ — проекция вектора V на вектор W .

$\text{pr}_\Pi V = V - \text{pr}_N V = V - (V \circ \bar{N}) \bar{N}$ — проекция вектора $V \in \mathbb{R}^3$ на плоскость Π , имеющую вектор нормали N .

$N = V \times W = \begin{vmatrix} x^0 & y^0 & z^0 \\ V_x & V_y & V_z \\ W_x & W_y & W_z \end{vmatrix} = N_x x^0 + N_y y^0 + N_z z^0$ — векторное произведение векто-

ров $V \in \mathbb{R}^3$ и $W \in \mathbb{R}^3$, имеющее координаты

$$N_x = \begin{vmatrix} V_y & V_z \\ W_y & W_z \end{vmatrix}, \quad N_y = \begin{vmatrix} V_z & V_x \\ W_z & W_x \end{vmatrix}, \quad N_z = \begin{vmatrix} V_x & V_y \\ W_x & W_y \end{vmatrix}$$

и модуль $|N| = |V| \cdot |W| \cdot \sin(\angle VW)$. Отсюда получаем значение синуса угла между векторами в пространстве:

$$\sin(\angle VW) = \frac{|V \times W|}{|V| \cdot |W|}.$$

$|V \times W| = 0$ — тест параллельности (антипараллельности) векторов $V \in \mathbb{R}^3$ и $W \in \mathbb{R}^3$.

$\{p_1 p_2 \dots p_n\}$ — ломаная линия (полилиния), заданная списком точек в порядке их соединения $p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$.

$\{p_1 p_2 \dots p_n p_1\}$ — многоугольник (полигон), заданный списком точек-вершин в порядке их обхода по замкнутому контуру $p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n \rightarrow p_1$.

$\nabla f(V) = \frac{df}{dV}$ — вектор градиента скалярной функции f векторного аргумента V .

$\text{line}(a, b)$ — функция рисования отрезка прямой линии ab на плоскости xy .

$\text{diag}[M_{11} \dots M_{mm}]$ — квадратная матрица $M \in \mathbb{R}^{n \times n}$ с диагональными элементами M_{ii} и недиагональными элементами $M_{ij} = 0 \forall i \neq j$.

$E_n = \text{diag}[1 \dots 1]$ — единичная $n \times n$ -матрица.

$\tilde{M} = \begin{bmatrix} M & \vdots & O_{n \times 1} \\ \hline O_{1 \times n} & \vdots & 1 \end{bmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}$ — расширенная матрица M , дополненная нижней диагональной единицей.

$\text{tr}(M) = \sum_{i=1}^n M_{ii}$ — след квадратной матрицы $M \in \mathbb{R}^{n \times n}$.

$\text{rang}(M)$ — ранг матрицы M , равный числу ее линейно независимых строк или столбцов.

$\{cond\} = \begin{cases} 1, & \text{если выполняется условие } cond; \\ 0, & \text{если не выполняется условие } cond. \end{cases}$

$\text{if}(cond : a, b) = \begin{cases} a & \text{при } cond \neq 0 \\ b & \text{при } cond = 0 \end{cases}$ — условный оператор.

$\text{sgn}(x) = \text{if}(x > 0 : 1, \text{if}(x < 0 : -1, 0)) = \begin{cases} 1 & \text{при } x > 0 \\ 0 & \text{при } x = 0 \\ -1 & \text{при } x < 0 \end{cases}$ — знак действительного числа x .

$\text{mod}(a, b)$ — остаток от деления действительного числа a на действительное число $b \neq 0$. Знак остатка равен знаку числа a . Дробная часть числа x равна $\text{mod}(x, 1)$.

$\lfloor x \rfloor$ — округление числа x до ближайшего целого числа $m \leq x$.

$\lceil x \rceil$ — округление числа x до ближайшего целого числа $M \geq x$.

$\lfloor x + 0.5 \rfloor$ или $\lceil x - 0.5 \rceil$ — округление числа x до ближайшего целого числа.

$\text{rnd}(x)$ — случайное число, равномерно распределенное в интервале $(0, x)$. Случайное число в интервале (a, b) равно $a + \text{rnd}(b - a)$. Случайное целое число в интервале $[n, m]$ генерируется как $\lfloor n + \text{rnd}(m + 1 - n) \rfloor$.

$L = \{l_1, \dots, l_n\}$ — список n скалярных переменных.

$Lp = \{p_1, \dots, p_n\}$ — список n точек, элемент которого есть вектор координат точки.

$Ls = \{a_1 b_1, \dots, a_n b_n\}$ — список n отрезков, элемент которого есть пара векторов концевых точек отрезка.

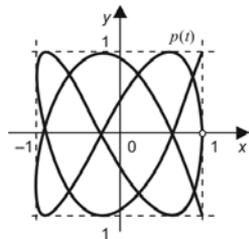
$Lm = \{L_1, \dots, L_m\}$ — группа m списков, элемент которого есть список типа L или Lp .

$\Gamma m = \{\Gamma_1, \dots, \Gamma_m\}$ — массив m групп списков, элемент которого есть массив списков типа Lm .

$\text{size}(S)$ — функция расчета числа элементов структурированного объекта S .

$\text{sort}_c(L)$ — функция сортировки элементов списка (вектора) L по критерию c .

Глава 1



Геометрические основы компьютерной графики

Реалистичность восприятия человеком сложного компьютерного изображения определяется умением разработчика математической модели изображаемого объекта или процесса достоверно повторить на экране его развитие в пространстве и во времени. Модель включает в себя систему уравнений и алгоритмов их реализации. Математической основой построения модели являются уравнения, описывающие форму и движение объектов. Все многообразие геометрических объектов является комбинацией различных *примитивов* — простейших фигур, которые в свою очередь состоят из *графических элементов* — точек, линий и поверхностей.

Положение точки в R^n задается *радиус-вектором* $p = [p_1 \ p_2 \ \dots \ p_n]$, имеющим n координат $p_1 \div p_n$ и разложение

$$p = [p_1 \ p_2 \ \dots \ p_n] = \sum_{i=1}^n p_i e_i \quad (1.1)$$

по n линейно-независимым *базисным векторам* $e_1 \div e_n$. Совокупность базисных векторов и единиц измерения расстояний вдоль этих векторов составляет *систему координат (СК)*. В пространствах размерностей $n \leq 3$ вместо цифр используются синонимы в виде буквенных обозначений координат. Например, первой координате можно сопоставить букву x , второй — букву y , а третьей — букву z .

Если при этом векторы $e_1 \equiv x^0$, $e_2 \equiv y^0$ и $e_3 \equiv z^0$ нормированы и взаимно ортогональны, то xyz является декартовой СК — наиболее распространенной системой координат.

Для описания формы графических объектов, задания расположения объектов в пространстве и их проекций на экране дисплея используются различные СК, наиболее удобные в каждом конкретном случае. Перечислим системы, применяемые в настоящей книге, и кратко охарактеризуем их назначение:

- неподвижная *мировая система координат (МСК)* xyz содержит точку отсчета (*начало координат*) и линейно независимый *базис*, благодаря которым становится возможным *цифровое* описание геометрических свойств любого графического объекта в *абсолютных* единицах;

- *экранный система координат (ЭСК)* $x_э, y_э, z_э$, в которой задается положение проекций геометрических объектов на экране дисплея. Проекция точки в ЭСК имеет координату $z_э = 0$. Тем не менее, не следует отбрасывать эту координату, поскольку МСК и ЭСК часто выбираются совпадающими, а, кроме того, вектор проекции $[x_э, y_э, 0]$ может участвовать в преобразованиях, где нужны не две, а три координаты;
- *система координат сцены (СКС)* $x_с, y_с, z_с$, в которой описывается положение всех объектов сцены — некоторой части мирового пространства с собственными началом отсчета и базисом, которые используются для описания положения объектов независимо от МСК;
- *объектная система координат (ОСК)* $x_о, y_о, z_о$, связанная с конкретным объектом и совершающая с ним все движения в СКС или МСК.

По умолчанию будем полагать, что все координаты задаются в МСК.

В R^2 наиболее распространены ортогональная *декартова* (x, y) и *полярная* (r, φ) системы (рис. 1.1, а), связанные друг с другом следующими соотношениями:

$$\begin{cases} x = r \cos(\varphi), & y = r \sin(\varphi), \\ r = \sqrt{x^2 + y^2}, & \operatorname{tg}(\varphi) = \frac{y}{x}. \end{cases} \quad (1.2)$$

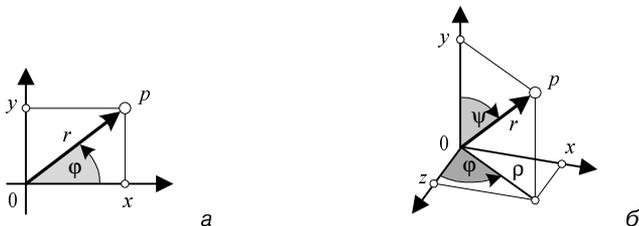


Рис. 1.1

В R^3 наиболее распространены ортогональная *декартова* (x, y, z) , *цилиндрическая* (ρ, y, φ) и *сферическая* (r, φ, ψ) системы (рис. 1.1, б), связанные друг с другом следующими соотношениями координат:

- декартова СК с цилиндрической СК:

$$\begin{cases} x = \rho \sin(\varphi), & z = r \cos(\varphi), \\ \rho = \sqrt{x^2 + z^2}, & \operatorname{tg}(\varphi) = \frac{x}{z}; \end{cases} \quad (1.3, а)$$

- декартова СК со сферической СК:

$$\begin{cases} x = r \sin(\varphi) \sin(\psi), & y = r \cos(\psi), & z = r \cos(\varphi) \sin(\psi), \\ r = \sqrt{x^2 + y^2 + z^2}, & \operatorname{tg}(\varphi) = \frac{x}{z}, & \operatorname{tg}(\psi) = \frac{\sqrt{x^2 + z^2}}{y}; \end{cases} \quad (1.3, б)$$

□ цилиндрическая СК со сферической СК:

$$\begin{cases} \rho = r \sin(\psi), & y = r \cos(\psi), \\ r = \sqrt{\rho^2 + y^2}, & \operatorname{tg}(\psi) = \frac{\rho}{y}. \end{cases} \quad (1.3, \text{в})$$

Для единообразия терминов и алгоритмов мы будем работать в правой декартовой СК, оси которой ориентированы в пространстве так, что вращения ортов $x^0 \rightarrow y^0$, $y^0 \rightarrow z^0$ и $z^0 \rightarrow x^0$ происходят в положительном направлении, т. е. против часовой стрелки с точки зрения наблюдателя, находящегося на конце третьего свободного орта.

Взаимосвязь координат графического элемента может быть описана *неявным уравнением $f(p) = 0$ и параметрической функцией $p(t)$* . Частным случаем неявного описания является явная форма, в которой одна из координат выражена через другие. Это возможно лишь для некоторых простейших линий и поверхностей. Не все графические элементы могут быть описаны неявными функциями. Наиболее универсальным является параметрическое описание каждой из n координат графического элемента соответствующей функцией времени, угла поворота или других параметров.

Целью настоящего краткого экскурса в вычислительную геометрию, не претендующего на исчерпывающую полноту и строгость, является систематизация формул, необходимых для работы с графическими объектами, приведение их к *алгоритмическому* виду. Это отличает компьютерную графику от обычной, в которой свойства и взаимосвязь графических объектов визуально воспринимаются глазами человека и интеллектуально интерпретируются его мозгом. С методами решения задач классической геометрии можно познакомиться, например, в [6, 7].

Компьютерная или вычислительная геометрия оперирует не с рисунками, а с числами (координатами точек, векторов, матрицами преобразований, списками данных) и отношениями между ними, выраженными также в цифровой форме. Исходная информация, обработанная специальными алгоритмами, преобразуется в другие цифровые данные, интерпретируемые как искомый результат — модель объекта, его экранное изображение, отношения между объектами сцены и т. п. Подробное изложение основ вычислительной геометрии приведено в [24, 29].

Рассмотрим основные соотношения между графическими элементами на плоскости R^2 и в пространстве R^3 . Все основные формулы будут записаны в координатной и векторной формах. Векторная форма компактнее координатной и предпочтительнее для понимания и программирования алгоритмов, хотя и требует несколько больших машинных ресурсов — объема памяти и времени счета. Увеличение ресурсов объясняется наличием в векторах и матрицах фиксированных элементов (единиц и нулей), с которыми иногда выполняются бесполезные арифметические операции.

Выше приведены основные обозначения и соотношения векторной алгебры. Самые трудоемкие матричные операции — обращение квадратной $n \times n$ -матрицы

M и расчет ее определителя — рекомендуется выполнять с максимально возможной точностью простым в программировании рекуррентным алгоритмом **Левверье—Фаддеева** [10]:

$$\begin{cases} L_1 = E_n, & \alpha_i = \frac{\text{tr}(ML_i)}{i}, & L_{i+1} = ML_i - \alpha_i E_n \quad \forall i = \overline{1, n}, \\ M^{-1} = \frac{L_n}{\alpha_n}, & |M| = (-1)^{n-1} \alpha_n. \end{cases} \quad (1.4)$$

1.1. Графические элементы на плоскости

Точка на плоскости имеет две степени свободы. *Расстояние между двумя точками* $p_1 = [x_1 \ y_1] \equiv [r_1 \ \varphi_1]$ и $p_2 = [x_2 \ y_2] \equiv [r_2 \ \varphi_2]$ равно:

$$d = \begin{cases} |p_1 - p_2| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} & \text{— в декартовой СК,} \\ \sqrt{r_1^2 + r_2^2 - 2r_1r_2 \cos(\varphi_1 - \varphi_2)} & \text{— в полярной СК.} \end{cases} \quad (1.5)$$

Линия на плоскости имеет одну степень свободы. *Уравнения линии* в неявной (**НФ**) и параметрической (**ПФ**) формах имеют следующий вид:

$$\text{НФ:} \quad f(x, y) = 0 \text{ или } f(p) = 0; \quad (1.6, a)$$

$$\text{ПФ:} \quad p(t) = [x(t) \ y(t)]. \quad (1.6, б)$$

В любой регулярной (гладкой и некратной) точке на линии $p_0 = [x_0 \ y_0] = p(t_0)$ возможна *линеаризация* кривой (рис. 1.2), т. е. проведение к ней *касательной прямой*, уравнения которой имеют вид:

$$\text{НФ:} \quad \begin{cases} N_x(x - x_0) + N_y(y - y_0) = 0 \\ N_x = \left. \frac{df(x, y)}{dx} \right|_{x_0, y_0}, \quad N_y = \left. \frac{df(x, y)}{dy} \right|_{x_0, y_0} \end{cases} \quad \text{или} \quad \begin{cases} N \circ (p - p_0) = 0; \\ N = \nabla f(p) \Big|_{p_0}; \end{cases} \quad (1.7, a)$$

$$\text{ПФ:} \quad \begin{cases} x(t) = x_0 + V_x t, \quad y(t) = y_0 + V_y t \\ V_x = \left. \frac{dx(t)}{dt} \right|_{x_0, y_0}, \quad V_y = \left. \frac{dy(t)}{dt} \right|_{x_0, y_0} \end{cases} \quad \text{или} \quad \begin{cases} p(t) = p_0 + Vt; \\ V = \left. \frac{dp(t)}{dt} \right|_{p_0}. \end{cases} \quad (1.7, б)$$

Вектор нормали $N = [N_x \ N_y]$ ортогонален к линии и по свойству градиента направлен в ту сторону, где $f(p) > 0$. При замене уравнения линии на $-f(p) = 0$ направление нормали меняется на противоположное.

Направляющий вектор линии $V = [V_x \ V_y]$ начинается в точке p_0 и направлен по касательной к $p(t)$ в сторону *увеличения* t . Для перехода от одной формы (1.7) к

другой необходимо выбрать одно из двух ортогональных направлений нормали относительно направляющего вектора. Из условия ортогональности

$$N \circ V = 0 \Rightarrow N_x V_x + N_y V_y = 0$$

получаем способ быстрого преобразования $V \rightarrow N$ или $N \rightarrow V$:

- координаты исходного вектора переставляются местами;
- у одной из координат полученного вектора меняется знак.

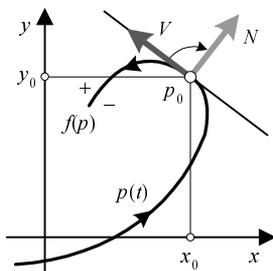


Рис. 1.2

Например, если дан вектор $V = [1 \ 2]$, то после перестановки координат он становится равным $[2 \ 1]$, а после изменения знака, например, у первой координаты получим вектор $N = [-2 \ 1]$, направленный влево от V . Проверка показывает, что векторы V и N ортогональны:

$$N \circ V = (-2) \cdot 1 + 1 \cdot 2 = 0.$$

Для единообразия изложения материала примем *правую* ориентацию нормали, означающую, что при движении точки по линии вектор N направлен перпендикулярно вправо от вектора V . Таким образом, связь векторов нормали и направления прямой будет следующей:

$$N = [V_y \ -V_x], \quad V = [-N_y \ N_x]. \tag{1.8}$$

1.1.1. Модели прямой линии на плоскости

Рассмотрим способы задания прямой на плоскости в различных формах.

1.1.1.1. Неявное уравнение прямой

Неявное уравнение прямой задается тремя коэффициентами A , B и D , составляющими вектор $F = [A \ B \ D]$:

$$\text{НФ:} \quad Ax + By + D = 0 \text{ или } [p \ 1] \circ F = \tilde{p} \circ F = 0. \tag{1.9}$$

Хотя бы одно из чисел A или B должно быть ненулевым. Если оба коэффициента $A \neq 0$ и $B \neq 0$, то прямая проходит наклонно к осям координат и пересекается с ними в следующих точках (рис. 1.3, а):

$$\left[\frac{-D}{A} \quad 0 \right], \left[0 \quad \frac{-D}{B} \right].$$

При $A = 0$ и $B \neq 0$ уравнение $Bu + D = 0$ описывает горизонтальную прямую, а уравнению $Ax + D = 0$ при $A \neq 0$ и $B = 0$ соответствует вертикальная прямая. Отметим ограниченность "школьного" способа приведения неявного координатного уравнения к явному виду

$$y(x) = -\frac{A}{B}x - \frac{D}{B}$$

и построения линии по точкам в желаемом интервале $x \in [x_{\min}, x_{\max}]$. Этот способ неприменим для вертикальных прямых с коэффициентом $B = 0$ и единственной координатой $x = -D/A$ всех точек таких прямых.

Прямая $f(p) \equiv \tilde{p} \circ F = 0$ проходит через начало координат при $f(O_2) \equiv D = 0$.

Благодаря ее свойству разделять пространство плоскости на два подпространства с противоположными знаками $f(p)$, неявное уравнение позволяет с минимальными вычислительными затратами (за два умножения и два сложения) определить по условию $f(q) = 0$ принадлежность произвольной точки q прямой линии, а также ориентацию двух точек a и b относительно прямой (рис. 1.3, а):

□ при $f(a) \cdot f(b) > 0$ точки лежат по одну сторону от прямой;

□ при $f(a) \cdot f(b) < 0$ — по разные стороны от прямой.

Для построения прямой по уравнению $Ax + By + D = 0$ необходимо и достаточно иметь либо две несовпадающие точки p_0 и p_1 , через которые она проходит, либо точку p_0 и направляющий вектор V , с помощью которого вторая точка вычисляется как $p_1 = p_0 + V$. Сравнивая формулы (1.9) и (1.7, а), заключаем, что координатами вектора нормали можно принять коэффициенты неявного уравнения A и B :

$$N = [A \quad B].$$

Тогда по (1.8) получаем направляющий вектор неявной прямой

$$V = [-B \quad A].$$

Рассмотрим два способа выбора точки $p_0 = [x_0 \quad y_0]$.

1. Первое, что приходит на ум — выбрать ее в том месте, где прямая пересекается с осью x при $A \neq 0$ или с осью y при $A = 0$ (только для горизонтальной прямой):

$$p_0 = \text{if} \left(A: \left[\frac{-D}{A} \quad 0 \right], \left[0 \quad \frac{-D}{B} \right] \right). \quad (1.10, a)$$

Несмотря на легкость понимания, этот способ не самый лучший для программирования из-за наличия альтернативы выбора точки в зависимости от коэффициента A .

2. Вычислим ближайшую к началу координат точку p_0 (рис. 1.3, а) методом наименьших квадратов (МНК) путем решения экстремальной задачи

$$|p|^2 = p \circ p \rightarrow \min$$

с ограничением $p \circ N + D = 0$. Для этого составим функцию Лагранжа

$$\mathfrak{J}(p, \lambda) = 0.5(p \circ p) + \lambda(p \circ N + D).$$

Из условий ее экстремума

$$\frac{\partial \mathfrak{J}(p, \lambda)}{\partial p} = p + \lambda N = O_2, \quad \frac{\partial \mathfrak{J}(p, \lambda)}{\partial \lambda} = p \circ N + D = 0$$

получим

$$p = -\lambda N, \quad \lambda = \frac{D}{N \circ N}$$

и оптимальное решение

$$p_0 = \frac{-D}{N \circ N} N = \frac{-D}{A^2 + B^2} [A \quad B]. \quad (1.10, б)$$

Данное решение безальтернативно по сравнению с (1.10, а), более удобно в программировании и эффективнее в вычислении. Поэтому далее будет использоваться преимущественно оно.

Таким образом, координатная и векторная параметрические формы прямой, заданной вектором $F = [A \quad B \quad D]$, принимают вид

$$\begin{cases} x(t) = \frac{-AD}{A^2 + B^2} - Bt \\ y(t) = \frac{-BD}{A^2 + B^2} + At \end{cases} \quad \text{или} \quad p(t) = \frac{-D}{A^2 + B^2} [A \quad B] + [-B \quad A]t. \quad (1.11)$$

1.1.1.2. Нормальное уравнение прямой

Нормальное уравнение прямой $\{p_0, N\}$, заданной точкой p_0 и вектором нормали N , выводится из условия ортогональности векторов N и $p - p_0$ для всех точек p , принадлежащих прямой (рис. 1.3, б):

НФ: $N_x(x - x_0) + N_y(y - y_0) = 0$ или $(p - p_0) \circ N = 0$; (1.12, а)

ПФ: $x(t) = x_0 - N_y t, y(t) = y_0 + N_x t$ или $p(t) = p_0 + Vt$, (1.12, б)

$$V = [-N_y \quad N_x].$$

Сравнивая (1.12, а) с (1.9), получим зависимость вектора неявного уравнения от параметров нормального уравнения:

$$F = [N \quad -p_0 \circ N]. \quad (1.13)$$

Неявная функция

$$f(p) = (p - p_0) \circ N$$

позволяет оценить положение точки p относительно прямой не только по знаку возвращаемого ей значения, но также и относительно направления нормали:

- при $f(a) > 0$ точка a лежит в том полупространстве, куда направлена нормаль, а угол $\angle(a - p_0, N)$ острый;
- при $f(b) < 0$ угол $\angle(b - p_0, N)$ тупой, а точка b и нормаль находятся по разные стороны от прямой (рис. 1.3, б).

1.1.1.3. Параметрическая функция прямой

Параметрическая функция прямой $\{p_0, V\}$, проходящей через точку p_0 в направлении вектора V (см. рис. 1.3, б), и эквивалентное ей уравнение в НФ будут иметь вид:

ПФ: $x(t) = x_0 + V_x t, y(t) = y_0 + V_y t$ или $p(t) = p_0 + Vt$; (1.14, а)

НФ: $V_y x - V_x y + D = 0, D = -V_y x_0 + V_x y_0 \Rightarrow F = [N - p_0 \circ N]$, (1.14, б)

где $N = [V_y - V_x]$. Параметрическая форма удобна для задания и построения частей прямой — отрезков и лучей. Для этого в (1.14, а) необходимо указать пределы изменения параметра t . Например:

- бесконечный интервал $-\infty < t < \infty$ не ограничивает протяженность бесконечной прямой;

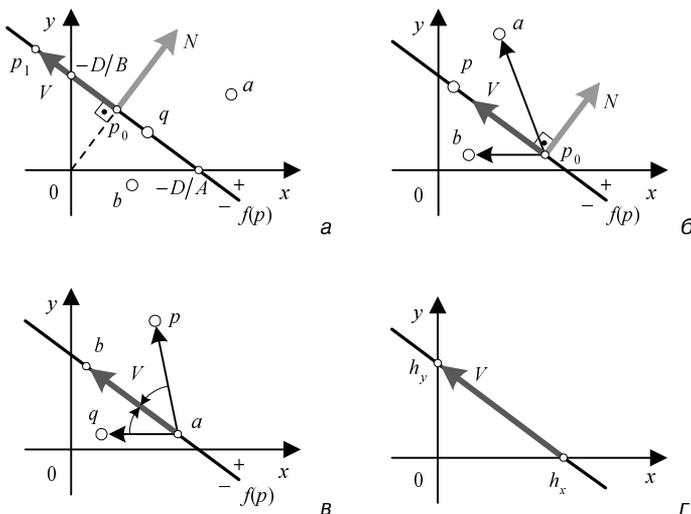


Рис. 1.3

- задание $t \geq 0$ дает луч, выходящий из точки p_0 в бесконечность в направлении вектора V ;
- конечный интервал $t_0 \leq t \leq t_1$ определяет отрезок прямой между точками $p_0 + Vt_0$ и $p_0 + Vt_1$.

Благодаря левой ориентации направляющего вектора V относительно вектора нормали N , эквивалентная нормальной форме функция

$$f(p) = (p - p_0) \circ \begin{bmatrix} V_y & -V_x \\ V \end{bmatrix} = \begin{bmatrix} p - p_0 \\ V \end{bmatrix}$$

позволяет определить положение точки относительно направления движения по прямой:

- при $f(a) > 0$ точка a лежит справа, так что угол $\angle(a - p_0, V)$ положительный;
- при $f(b) < 0$ угол $\angle(b - p_0, V)$ отрицательный, а точка b лежит слева от прямой.

1.1.1.4. Уравнения прямой, проходящей через две точки

Неявная форма уравнения прямой, проходящей через две точки $a = [a_x \ a_y]$ и $b = [b_x \ b_y]$ (рис. 1.3, в), выводится из условия принадлежности прямой этих точек и точки $p = [x \ y]$:

$$\text{НФ: } \begin{bmatrix} \tilde{p} \\ \tilde{b} \\ \tilde{a} \end{bmatrix} F^T = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{vmatrix} x & y & 1 \\ b_x & b_y & 1 \\ a_x & a_y & 1 \end{vmatrix} = 0 \Rightarrow \underbrace{\begin{bmatrix} b_y & 1 \\ a_y & 1 \end{bmatrix}}_{N_x} x + \underbrace{\begin{bmatrix} 1 & b_x \\ 1 & a_x \end{bmatrix}}_{N_y} y + \underbrace{\begin{bmatrix} b_x & b_y \\ a_x & a_y \end{bmatrix}}_D = 0. \quad (1.15, a)$$

Выбрав направление движения по прямой от точки a к точке b , получим направляющий вектор $V = b - a$ и параметрическую модель линии:

$$\text{ПФ: } x(t) = a_x + (b_x - a_x)t, y(t) = a_y + (b_y - a_y)t \text{ или } p(t) = a + (b - a)t. \quad (1.15, б)$$

Условие существования прямой очевидно: $|V| \neq 0$, т. е. $a \neq b$.

При изменении параметра от $t = 0$ до $t = 1$ движение точки происходит *внутри* отрезка ab от точки a до точки b . Ниже рассматриваются свойства плоских многоугольников (*полигонов*), определяемые ориентацией векторов нормалей к их сторонам. С учетом принятого соглашения (1.8) **внешняя ориентация нормалей к сторонам полигона обеспечивается при его обходе против часовой стрелки.**

Соответствующая данному способу задания прямой функция

$$f(p) = \begin{bmatrix} p - a \\ b - a \end{bmatrix}$$

позволяет определить нахождение точки p относительно движения по прямой от a до b . При $f(p) > 0$ точка p лежит справа от прямой, а при $f(q) < 0$ точка q находится от нее слева.

1.1.1.5. Уравнения прямой в отрезках

Неявное уравнение прямой в отрезках h_x и h_y , одновременно не равных нулю и отсекаемых прямой на осях x и y (рис. 1.3, ε), получим из уравнения (1.9) при $D = -1$ с учетом соотношений $h_x = -D / A$ и $h_y = -D / B$:

$$\text{НФ: } \frac{x}{h_x} + \frac{y}{h_y} - 1 = 0 \Rightarrow h_y x + h_x y - h_x h_y = 0 \Rightarrow F = [h_y \quad h_x \quad -h_x h_y]. \quad (1.16, a)$$

Выбрав на прямой две точки $a = [h_x \quad 0]$, $b = [0 \quad h_y]$, получим ее направляющий вектор

$$V = b - a = [-h_x \quad h_y]$$

и координатные параметрические функции прямой в отрезках:

$$\text{ПФ: } x(t) = h_x(1 - t), \quad y(t) = h_y t. \quad (1.16, б)$$

1.1.2. Взаимное положение графических элементов на плоскости

Рассмотрим некоторые полезные формулы, тесты и алгоритмы, описывающие свойства и взаимное расположение точек, прямых и полигонов на плоскости. Для получения необходимых результатов активно используются тесты ортогональности и параллельности векторов, приведенные в перечне основных обозначений и соотношений.

1.1.2.1. Коллинеарность точек

Три точки p_1 , p_2 и p_3 коллинеарны, т. е. лежат на одной прямой (рис. 1.4, a), если

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = 0 \quad \text{или} \quad \begin{vmatrix} p_2 - p_1 \\ p_3 - p_1 \end{vmatrix} = 0. \quad (1.17, a)$$

Обобщение (1.17, a) на произвольное число точек $p_1 \div p_n$:

$$\text{rang} \begin{bmatrix} \tilde{p}_1 \\ \tilde{p}_2 \\ \dots \\ \tilde{p}_n \end{bmatrix} = 2 \quad \text{или} \quad \text{rang} \begin{bmatrix} p_2 - p_1 \\ p_3 - p_1 \\ \dots \\ p_n - p_1 \end{bmatrix} = 1. \quad (1.17, б)$$

Точка p лежит на отрезке ab при нулевом угле между векторами $p - a$ и $b - p$:

$$(p - a) \circ (b - p) = |p - a| \cdot |b - p|. \quad (1.18)$$

Значение параметра t , соответствующее положению точки $p = a + (b - a)t$ на прямой относительно ее отрезка ab , вычисляется с помощью скалярного произведения:

$$(p - a) \circ (b - a) = (b - a)t \circ (b - a) = |b - a|^2 \cdot t,$$

откуда

$$t = \frac{(p - a) \circ (b - a)}{(b - a) \circ (b - a)}. \quad (1.19)$$

1.1.2.2. Взаимное расположение прямых

Две прямые совпадают (рис. 1.4, б), если

$$\text{НФ:} \quad F_1 \times F_2 = O_3; \quad \text{ПФ:} \quad \left\{ \left| \begin{array}{c} V_1 \\ V_2 \end{array} \right| = 0 \right\} \cap \left\{ \left| \begin{array}{c} V_1 \\ p_{20} - p_{10} \end{array} \right| = 0 \right\}. \quad (1.20)$$

Две прямые параллельны (рис. 1.4, в), если

$$\text{НФ:} \quad \left| \begin{array}{c} N_1 \\ N_2 \end{array} \right| = 0; \quad \text{ПФ:} \quad \left| \begin{array}{c} V_1 \\ V_2 \end{array} \right| = 0. \quad (1.21)$$

Расстояние между параллельными прямыми равно:

$$\text{НФ:} \quad d = \left| \text{pr}_{N_1}(p_{20} - p_{10}) \right| = \left| (p_{20} - p_{10}) \circ \bar{N}_1 \right|; \quad (1.22, a)$$

$$\begin{aligned} \text{ПФ:} \quad d &= \left| \left| \begin{array}{c} p_{20} - p_{10} \\ \bar{V}_1 \end{array} \right| \right| = \left| p_{20} - p_{10} - \text{pr}_{V_1}(p_{20} - p_{10}) \right| = \\ &= \left| p_{20} - p_{10} - \left((p_{20} - p_{10}) \circ \bar{V}_1 \right) \bar{V}_1 \right|. \end{aligned} \quad (1.22, б)$$

Две прямые ортогональны (рис. 1.4, г), если

$$\text{НФ:} \quad N_1 \circ N_2 = 0; \quad \text{ПФ:} \quad V_1 \circ V_2 = 0. \quad (1.23)$$

1.1.2.3. Взаимное расположение точки и прямой

Уравнения перпендикуляра, опущенного из точки $q = [q_x \quad q_y]$ на прямую, заданную в НФ или ПФ (рис. 1.4, д), выглядят следующим образом:

$$\text{НФ:} \quad N_y(x - q_x) - N_x(y - q_y) = 0 \Rightarrow N_{\perp} = [N_y \quad -N_x], \quad F_{\perp} = [N_{\perp} \quad -N_{\perp} \circ q]; \quad (1.24, a)$$

$$\text{ПФ:} \quad p_{\perp}(t) = q + N t \text{ или } p_{\perp}(t) = q + V_{\perp} t, \text{ где } V_{\perp} = [V_y \quad -V_x]. \quad (1.24, б)$$

Расстояние от точки q до прямой равно:

$$\text{НФ:} \quad d = \frac{|\tilde{q} \circ F|}{|N|} = |\text{pr}_N(q - p_0)| = |(q - p_0) \circ \bar{N}|; \quad (1.25, a)$$

$$\text{ПФ:} \quad d = \left| \left| \frac{q - p_0}{\bar{V}} \right| \right| = |q - p_0 - \text{pr}_V(q - p_0)| = |q - p_0 - ((q - p_0) \circ \bar{V})\bar{V}|. \quad (1.25, б)$$

Зеркальное отражение точки q от прямой лежит на перпендикуляре к прямой на расстоянии $2d$ от q в сторону, противоположную проекции вектора $q - p_0$ на нормаль N (рис. 1.4, д):

$$\text{НФ:} \quad q' = q - 2\text{pr}_N(q - p_0) = q - 2((q - p_0) \circ \bar{N})\bar{N}; \quad (1.26, a)$$

$$\text{ПФ:} \quad q' = q - 2(q - p_0 - \text{pr}_V(q - p_0)) = 2p_0 - q - 2((q - p_0) \circ \bar{V})\bar{V}. \quad (1.26, б)$$

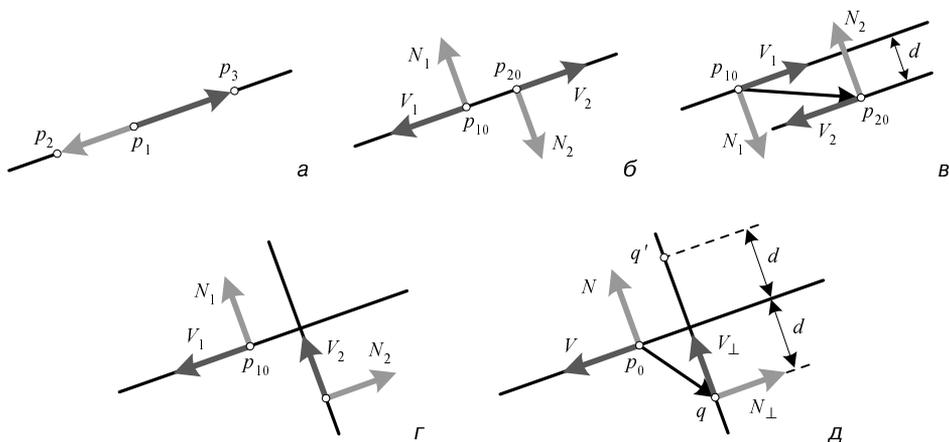


Рис. 1.4

📁 **Пример 1.1.** Отразить треугольник $p_1 p_2 p_3$ с вершинами $p_1 = [0 \ 0]$, $p_2 = [0 \ 1]$ и $p_3 = [2 \ 0]$ от прямой линии, проходящей через точки $a = [3 \ 0]$ и $b = [2 \ 2]$ (рис. 1.5).

Решение. Находим направляющий вектор прямой $V = b - a = [-1 \ 2]$, по формуле (1.8) нормаль $N = [2 \ 1]$ и квадрат ее длины $N \circ N = 5$. Пусть $p_0 = a$, тогда по (1.26, а) последовательно для точек p_1, p_2 и p_3 вычисляем отраженные точки

$$p'_i = p_i - \frac{2(p_i - a) \circ N}{N \circ N} N;$$

$$p'_1 = [0 \ 0] - 0.4([-3 \ 0] \circ [2 \ 1])[2 \ 1] = [4.8 \ 2.4],$$

$$p'_2 = [0 \ 1] - 0.4([-3 \ 1] \circ [2 \ 1])[2 \ 1] = [4 \ 3],$$

$$p'_3 = [2 \ 0] - 0.4([-1 \ 0] \circ [2 \ 1])[2 \ 1] = [3.6 \ 0.8]. \quad \square$$

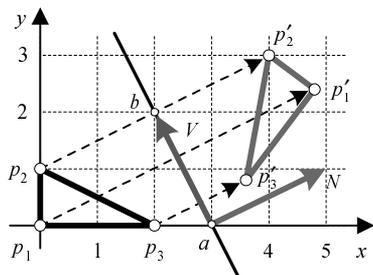


Рис. 1.5

1.1.2.4. Построение прямой, наименее удаленной от совокупности точек

Рассмотрим интересную задачу, связанную с понятием расстояния от точки до прямой линии. Допустим, в \mathbb{R}^n (здесь при решении задачи на плоскости xy будем полагать $n = 2$) задано $m \geq 2$ произвольных точек $p_i \div p_m$. Сформируем из них матрицу

$$P = \begin{bmatrix} p_1 \\ \dots \\ p_m \end{bmatrix} \in \mathbb{R}^{m \times n}. \quad (1.27)$$

Так как в общем случае точки не коллинеарны, то не существует ни одной прямой $f(p) = 0$, для которой все числа $d_i = f(p_i)$ обращаются в ноль. Другими словами, расстояния $|d_i| / |N|$ от точек p_i до прямой с нормалью N , вычисленные по (1.25, а) и пропорциональные d_i , не могут быть одновременно все нулевыми (рис. 1.6).

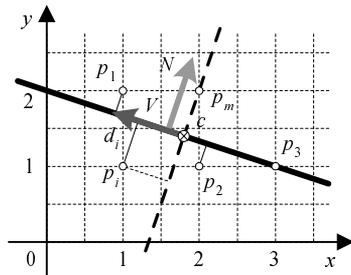


Рис. 1.6

Построим *аппроксимирующую прямую* $\{p_0, N\}$, наименее удаленную от совокупности точек p_i , путем решения экстремальной задачи

$$\sum_{i=1}^m d_i^2 \rightarrow \min .$$

С целью уменьшения неопределенности зададим ограничение $|N| = 1$, поскольку ориентация прямой линии зависит лишь от направления, а не от длины вектора нормали. При выполнении данного ограничения числа d_i являются *алгебраическими*, т. е. со знаками, расстояниями от точек до прямой. Из пяти рассмотренных выше способов описания прямой линии выберем неявную форму (1.9) с вектором $F = [N \ D]$, имеющим минимальное число скалярных параметров, равное трем при $n = 2$. При таком выборе вектор $d \in \mathbb{R}^m$ расстояний от точек до прямой принимает вид

$$d = F\tilde{P}^T = NP^T + D \underbrace{[1 \ \dots \ 1]}_m,$$

а квадрат его длины равен

$$\begin{aligned} |d|^2 &= dd^T = \left(NP^T + D [1 \ \dots \ 1] \right) \left(PN^T + D \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix} \right) = \\ &= NP^T PN^T + 2D \left(\sum_{i=1}^m p_i \right) N^T + mD^2 = NP^T PN^T + 2mD(N \circ c) + mD^2, \end{aligned}$$

где точка

$$c = \frac{1}{m} \sum_{i=1}^m p_i \quad (1.28)$$

есть *геометрический центр* массива точек p_i . Задача аппроксимации решается методом наименьших квадратов путем минимизации функции Лагранжа

$$\begin{aligned} \mathfrak{S}(N, D, \lambda) &= |d|^2 + \lambda \left(|N|^2 - 1 \right) = \\ &= NP^T PN^T + 2mD(N \circ c) + mD^2 + \lambda (NN^T - 1). \end{aligned} \quad (1.29, a)$$

Из условий ее экстремума

$$\begin{cases} \frac{1}{2} \frac{\partial \mathfrak{S}}{\partial N} = NP^T P + mDc + \lambda N = O_n; \\ \frac{1}{2} \frac{\partial \mathfrak{S}}{\partial D} = m(N \circ c) + mD = 0, \quad \frac{\partial \mathfrak{S}}{\partial \lambda} = NN^T - 1 = 0 \end{cases} \quad (1.29, б)$$

и свойства произведения скалярного произведения двух векторов на третий вектор

$$(a \circ b) \cdot c = a \cdot (b^T c)$$

следует, что

$$D = -N \circ c, N(P^T P - mc^T c + \lambda E_n) = O_n. \quad (1.29, \text{в})$$

Сравнение выражений для коэффициента D в (1.29, в) и (1.13) убеждает, что все прямые, доставляющие экстремум $\mathfrak{Z}(N, D, \lambda)$, проходят через точку $p_0 = c$.

Второе соотношение в (1.29, в) представляет собой однородную систему линейных алгебраических уравнений относительно координат вектора N . Она имеет нетривиальное решение $N \neq O_n$ при нулевом определителе матрицы $\lambda E_n - M$, где

$$M = mc^T c - P^T P \in \mathbb{R}^{n \times n}. \quad (1.29, \text{г})$$

Следовательно, множитель Лагранжа λ есть одно из n собственных чисел (значений), а N — соответствующий ему нормированный собственный вектор матрицы M . Количество собственных значений квадратной матрицы равно ее размерности. В случае $n = 2$ симметричная матрица M имеет две действительные собственные пары $\{\lambda_1, N_1\}$ и $\{\lambda_2, N_2\}$, причем $N_1 \perp N_2$ при $\lambda_1 \neq \lambda_2$ [см. 9, 13, 32]. Это означает, что прямые, доставляющие экстремум функции $\mathfrak{Z}(N, D, \lambda)$, ортогональны. Вычисление собственных чисел и векторов матрицы — хорошо отработанная процедура вычислительной математики, например, в [10, 27, 32]. Следует обратить внимание на *нормирование* собственных векторов в том случае, если алгоритм их вычисления не делает эту операцию сам.

Возникает закономерный вопрос: какой из полученных векторов N_1 или N_2 следует выбрать в качестве нормали искомой прямой? Ответ очевиден: тот, чья прямая доставляет *минимум* функции Лагранжа (1.29, а). Ведь условия (1.29, б) дают лишь экстремальные значения, которые при $D = -N \circ c$ и $|N| = 1$ равны

$$\mathfrak{Z} = NP^T P N^T - m(N \circ c)^2 = -NMN^T = -\lambda NN^T = -\lambda \geq 0.$$

Отсюда следует, во-первых, что все собственные числа матрицы M не положительны, во-вторых, что минимум функции Лагранжа доставляет собственная пара $\{\lambda, N\}$ с *максимальным* значением $\lambda = \max\{\lambda_1, \lambda_2\}$ и соответствующим ему вектором N — нормалью искомой прямой $\{c, N\}$.

Для удобства построения найденной прямой линии можно перейти от ее нормальной формы (1.12, а) к параметрической форме (1.14, а) с вектором $V = \begin{bmatrix} -N_y & N_x \end{bmatrix}$.

Оставшаяся собственная пара матрицы M с минимальным собственным числом дает максимум функции Лагранжа и ортогональную (при $\lambda_1 \neq \lambda_2$) прямую (на рис. 1.6 она изображена пунктиром), пересекающуюся с оптимальной прямой в точке c .

📁 **Пример 1.2.** Построить прямую линию, наименее отклоненную от точек $p_1 = [1 \ 2]$, $p_2 = [2 \ 1]$, $p_3 = [3 \ 1]$, $p_4 = [1 \ 1]$ и $p_5 = [2 \ 2]$ (см. рис. 1.6).

Решение. Из заданных точек составим матрицу (1.27) и найдем их центр (1.28):

$$P = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 1 \\ 1 & 1 \\ 2 & 2 \end{bmatrix}, \quad c = \frac{p_1 + p_2 + p_3 + p_4 + p_5}{5} = [1.8 \quad 1.4].$$

Вычислим по (1.29, z) матрицу M и ее собственные параметры:

$$M = 5 \begin{bmatrix} 1.8 \\ 1.4 \end{bmatrix} \begin{bmatrix} 1.8 & 1.4 \end{bmatrix} - \begin{bmatrix} 1 & 2 & 3 & 1 & 2 \\ 2 & 1 & 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 1 & 2 \\ 2 & 1 & 1 & 1 & 2 \end{bmatrix}^T = \begin{bmatrix} -2.8 & 0.6 \\ 0.6 & -1.2 \end{bmatrix},$$

$$\lambda_1 = -3, \lambda_2 = -1, N_1 = [-0.949 \quad 0.316], N_2 = [0.316 \quad 0.949].$$

Отделив максимальное значение $\lambda = \lambda_2$, получим параметры неявной формы искомой прямой:

$$N = [0.316 \quad 0.949], D = -N \circ c = -1.897.$$

Таким образом, аппроксимирующая прямая описывается неявным уравнением

$$0.316x + 0.949y - 1.897 = 0$$

или параметрическими функциями

$$x(t) = 1.8 - 0.949t, \quad y(t) = 1.4 + 0.316t.$$

Алгебраические расстояния от точек до этой прямой равны $d_1 = 0.316$, $d_2 = -0.316$, $d_3 = 0$, $d_4 = -0.632$ и $d_5 = 0.632$, а минимум функции Лагранжа составляет $\mathfrak{I}_{\min} = 1$.

Для сравнения на рис. 1.6 построена пунктирная прямая, соответствующая собственной паре $\{\lambda_1, N_1\}$. Расстояния от нее до заданных точек равны $d_1 = 0.949$, $d_2 = -0.316$, $d_3 = -1.265$, $d_4 = 0.632$ и $d_5 = 0$, а значение функции Лагранжа $\mathfrak{I}_{\max} = 3 > \mathfrak{I}_{\min}$, что подтверждает правильность решения задачи. \square

1.1.2.5. Пересечение двух прямых

Важнейшей задачей геометрии на плоскости является расчет точки q пересечения двух прямых (рис. 1.7, а), который в зависимости от формы их задания выполняется следующим образом:

$$\begin{aligned} \text{НФ:} \quad q \begin{bmatrix} N_1^T & N_2^T \end{bmatrix} &= \begin{bmatrix} p_{10} \circ N_1 & p_{20} \circ N_2 \end{bmatrix} \Rightarrow \\ \Rightarrow q &= \begin{bmatrix} p_{10} \circ N_1 & p_{20} \circ N_2 \end{bmatrix} \begin{bmatrix} N_1^T & N_2^T \end{bmatrix}^{-1}; \end{aligned} \quad (1.30, a)$$

$$\text{ПФ: } p_{10} + V_1 t_1 = p_{20} + V_2 t_2 \Rightarrow [t_1 \quad t_2] = (p_{20} - p_{10}) \begin{bmatrix} V_1 \\ -V_2 \end{bmatrix}^{-1} \Rightarrow \quad (1.30, \delta)$$

$$\Rightarrow q = p_{10} + V_1 t_1.$$

Условия существования решений в (1.30)

$$|N_1^T \quad N_2^T| \neq 0 \quad \text{или} \quad \begin{vmatrix} V_1 \\ -V_2 \end{vmatrix} \neq 0$$

равносильны *непараллельности прямых*.

Введем **функцию вычисления угла между векторами** V и W

$$\text{ang}(V, W) = \text{if} \left(\left(d = \begin{vmatrix} V \\ W \end{vmatrix} \right) \neq 0 : \text{sgn}(d), 1 \right) \cdot \arccos \left(\frac{V \circ W}{|V| \cdot |W|} \right) \quad (1.31)$$

как угла *кратчайшего* вращения от V к W . По свойству скалярного произведения векторов модуль угла равен $\arccos(\overline{V \circ W}) \in [0, \pi]$, а правильный знак определяется по *ненулевому* знаку определителя d матрицы, построенной на этих векторах.

При $d = 0$ векторы V и W линейно зависимы, а знак $\text{sgn}(d) = 0$ в (1.31) игнорируется. Это может быть в двух случаях:

- при *параллельности* векторов угол между ними нулевой, а $\arccos(1) = 0$;
- при *антипараллельности* векторов угол между ними развернутый, а по (1.31) получим $\arccos(-1) = \pi$.

Таким образом, *угол пересечения двух прямых* находится с помощью функции (1.31) как

$$\gamma = \text{ang}(N_1, N_2) \quad \text{или} \quad \gamma = \text{ang}(V_1, V_2). \quad (1.32)$$

1.1.2.6. Уравнения пучка прямых и биссектриса угла

Уравнения пучка прямых, заключенных между двумя прямыми (рис. 1.7, б), выводятся следующим образом:

$$\text{НФ: } \begin{cases} f_1(p) = \tilde{p} \circ F_1 \\ f_2(p) = \tilde{p} \circ F_2 \end{cases} \Rightarrow f_{\lambda+1}(p) = (1-\lambda)f_1(p) + \lambda f_2(p); \quad (1.33, a)$$

$$\text{ПФ: } \begin{cases} p_1(t) = p_{10} + V_1 t \\ p_2(t) = p_{20} + V_2 t \end{cases} \stackrel{(1.30, \delta)}{\Rightarrow} q \Rightarrow \begin{cases} p_{\lambda+1}(t) = q + V_{\lambda+1} t; \\ V_{\lambda+1} = (1-\lambda)V_1 + \lambda V_2. \end{cases} \quad (1.33, \delta)$$

Изменение параметра пучка в интервале $0 \leq \lambda \leq 1$ дает такие промежуточные прямые, что вращения векторов $N_{\lambda+1}$ и $V_{\lambda+1}$ происходит по *кратчайшим* углам от начальных положений N_1 и V_1 до конечных положений N_2 и V_2 . Уравнение (1.33, б) можно также рассматривать как параметрическую (с параметром λ) мо-