Матем.

М.С. Красс

МАТЕМАТИКА

для экономических специальностей

Учебник

М.С. Красс

МАТЕМАТИКА
ДЛЯ ЭКОНОМИЧЕСКИХ СПЕЦИАЛЬНОСТЕЙ

Рекомендовано
Министерством образования
Российской Федерации
в качестве учебника
для студентов высших учебных заведений,
обучающихся по экономическим
специальностям

2118

Академия народного хозяйства
при Правительстве Российской Федерации

Москва
Издательство "ДЕЛО"
2003

УДК 330.115(075.8)
ББК 22.11я73
К78

Рецензенты:
Д.Г. Орловский, доктор физико-математических наук, профессор;
кафедра высшей математики Академии легкой промышленности

Красс М.С.

К78 Математика для экономических специальностей: Учебник. — 4-е изд., испр. — М.: Дело, 2003. — 704 с.
ISBN 5-7749-0264-1

Изложены основы математического анализа, линейной алгебры, дифференциальных и разностных уравнений, теории вероятностей и математической статистики, экономико-математического моделирования. Именно такая совокупность знаний необходима тем, кто всерьез решил овладеть экономическими знаниями. По всем разделам, помимо решения соответствующих задач, приведены экономические приложения и модели.

Предназначена для студентов экономических и смежных технических специальностей вузов, экономистов, а также для тех, кто хочет самостоятельно углубить свои знания.

УДК 330.115(075.8)
ББК 22.11я73

ISBN 5-7749-0264-1

© Издательство "Дело", оформление, 2002

СОДЕРЖАНИЕ

Предисловие к третьему изданию 9
Введение 11

Часть 1. МАТЕМАТИЧЕСКИЙ АНАЛИЗ 15

Глава 1. МНОЖЕСТВА 15
1.1. Понятие множества и операции над множествами 15
1.2. Вещественные числа и их свойства 17
1.3. Числовая прямая (числовая ось) и множества на ней 20
1.4. Грани числовых множеств 21
1.5. Абсолютная величина числа 22

Глава 2. ПРЕДЕЛ ПОСЛЕДОВАТЕЛЬНОСТИ 24
2.1. Числовые последовательности 24
2.2. Сходящиеся последовательности 30
2.3. Монотонные последовательности 39

Глава 3. ФУНКЦИИ ОДНОЙ ПЕРЕМЕННОЙ 46
3.1. Понятие функции 46
3.2. Предел функции 58
3.3. Теоремы о пределах функций 64
3.4. Два замечательных предела 65
3.5. Бесконечно малые и бесконечно большие функции 69
3.6. Сравнение бесконечно малых и бесконечно больших функций 70
3.7. Понятие непрерывности функции 74
3.8. Непрерывность элементарных функций 76
3.9. Основные свойства непрерывных функций 81
3.10. Понятие сложной функции 88
3.11. Понятие обратной функции 89
3.12. Элементы аналитической геометрии на плоскости 91

Глава 4. ОСНОВЫ ДИФФЕРЕНЦИАЛЬНОГО ИСЧИСЛЕНИЯ ... 98
4.1. Понятие производной ... 98
4.2. Понятие дифференцируемости функции ... 101
4.3. Понятие дифференциала функции ... 104
4.4. Правила дифференцирования суммы, произведения и частного ... 106
4.5. Производные постоянной, степенной, логарифмической и тригонометрических функций ... 107
4.6. Теорема о производной обратной функции ... 110
4.7. Производные показательной и обратных тригонометрических функций ... 111
4.8. Дифференцирование сложной функции ... 113
4.9. Логарифмическая производная и производная степенной функции с любым вещественным показателем ... 115
4.10. Производные простейших элементарных функций ... 117
4.11. Производные и дифференциалы высших порядков ... 118

Глава 5. ПРИМЕНЕНИЕ ПРОИЗВОДНЫХ В ИССЛЕДОВАНИИ ФУНКЦИЙ ... 124
5.1. Основные теоремы дифференциального исчисления ... 124
5.2. Раскрытие неопределенностей ... 129
5.3. Формула Тейлора ... 133
5.4. Исследование функций и построение графиков ... 140
5.5. Применение в экономике ... 155

Глава 6. НЕОПРЕДЕЛЕННЫЙ ИНТЕГРАЛ ... 162
6.1. Первообразная и неопределенный интеграл ... 162
6.2. Основные свойства неопределенного интеграла ... 164
6.3. Основные неопределенные интегралы ... 164
6.4. Основные методы интегрирования ... 167
6.5. Интегрирование рациональных функций ... 173
6.6. Интегрирование иррациональных и трансцендентных функций ... 180

Глава 7. ОПРЕДЕЛЕННЫЙ ИНТЕГРАЛ ... 187
7.1. Условия существования определенного интеграла ... 187
7.2. Классы интегрируемых функций ... 195
7.3. Основные свойства определенного интеграла ... 198
7.4. Формулы оценки определенных интегралов ... 200
7.5. Основная формула интегрального исчисления ... 204
7.6. Основные правила интегрирования ... 207
7.7. Геометрические приложения определенного интеграла ... 211

7.8. Некоторые приложения определенного интеграла в экономике ... 229
7.9. Приближенное вычисление определенных интегралов ... 233
7.10. Несобственные интегралы ... 237

Глава 8. РЯДЫ ... 249
8.1. Понятие числового ряда ... 249
8.2. Числовые ряды с неотрицательными членами ... 252
8.3. Сходимость произвольных числовых рядов ... 259
8.4. Степенные ряды ... 262
8.5. Ряды Фурье ... 272

Глава 9. ФУНКЦИИ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ ... 282
9.1. Евклидово пространство E^m ... 282
9.2. Множества точек евклидова пространства E^m ... 283
9.3. Предел функции нескольких переменных ... 291
9.4. Непрерывные функции нескольких переменных ... 294
9.5. Элементы аналитической геометрии в пространстве ... 301

Глава 10. ДИФФЕРЕНЦИРОВАНИЕ ФУНКЦИЙ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ ... 318
10.1. Производные и дифференциалы ... 318
10.2. Производные и дифференциалы высших порядков ... 338
10.3. Локальный экстремум функции нескольких переменных ... 346
10.4. Функции нескольких переменных в задачах экономики ... 359

Глава 11. НЕЯВНЫЕ ФУНКЦИИ ... 371
11.1. Неявная функция ... 371
11.2. Зависимость функций ... 379
11.3. Условный экстремум ... 383

Глава 12. ДВОЙНЫЕ ИНТЕГРАЛЫ ... 390
12.1. Существование двойного интеграла ... 390
12.2. Вычисление двойных интегралов ... 394

Часть 2. ЭЛЕМЕНТЫ ЛИНЕЙНОЙ АЛГЕБРЫ ... 406

Глава 13. ВЕКТОРЫ ... 406
13.1. Векторное пространство ... 406
13.2. Линейная зависимость векторов ... 409
13.3. Разложение вектора по базису ... 412

Глава 14. МАТРИЦЫ ... 416
14.1. Матрицы и операции над ними ... 416
14.2. Обратная матрица ... 425

Глава 15. ОПРЕДЕЛИТЕЛИ ... 426
15.1. Операции над определителями и основные свойства ... 426
15.2. Применение определителей ... 431

Глава 16. СИСТЕМЫ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ ... 436
16.1. Основные понятия ... 436
16.2. Методы решения систем линейных уравнений ... 438
16.3. Однородные системы линейных уравнений ... 451

Глава 17. ПРИМЕНЕНИЕ ЭЛЕМЕНТОВ ЛИНЕЙНОЙ АЛГЕБРЫ В ЭКОНОМИКЕ ... 457
17.1. Использование алгебры матриц ... 457
17.2. Модель Леонтьева многоотраслевой экономики ... 466
17.3. Линейная модель торговли ... 474

ЧАСТЬ 3. ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ И РАЗНОСТНЫЕ УРАВНЕНИЯ ... 477

Глава 18. ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ ПЕРВОГО ПОРЯДКА ... 477
18.1. Основные понятия ... 477
18.2. Виды уравнений и методы решения ... 481

Глава 19. ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ ВЫСШИХ ПОРЯДКОВ ... 489
19.1. Дифференциальные уравнения второго порядка ... 489
19.2. Дифференциальные уравнения высших порядков ... 507

Глава 20. ПРИМЕНЕНИЕ АППАРАТА ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЭКОНОМИКЕ ... 515
20.1. Дифференциальные уравнения первого порядка ... 516
20.2. Дифференциальные уравнения второго порядка ... 523

Глава 21. ЭЛЕМЕНТЫ ТЕОРИИ ЛИНЕЙНЫХ РАЗНОСТНЫХ УРАВНЕНИЙ ... 527
21.1. Основные понятия ... 527
21.2. Применение разностных уравнений в экономике ... 538

Часть 4. ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ ... 545

Глава 22. ОСНОВНЫЕ ПОЛОЖЕНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ ... 545
22.1. Основные понятия теории вероятностей ... 546
22.2. Теорема сложения вероятностей ... 550
22.3. Теорема умножения вероятностей ... 552
22.4. Обобщения теорем сложения и умножения ... 556
22.5. Схема независимых испытаний ... 563

Глава 23. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ ... 569
23.1. Случайные величины и законы их распределения ... 569
23.2. Числовые характеристики дискретных случайных величин ... 574
23.3. Система двух случайных величин ... 584
23.4. Непрерывные случайные величины ... 590
23.5. Основные распределения непрерывных случайных величин ... 598
23.6. Некоторые элементы математической статистики ... 605

ЧАСТЬ 5. ОСНОВНЫЕ АСПЕКТЫ ЭКОНОМИКО-МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ... 619

Глава 24. ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МОДЕЛИ ... 619
24.1. Предназначение модели ... 620
24.2. Классификация экономико-математических моделей ... 625

Глава 25. АППАРАТ ПРОИЗВОДСТВЕННЫХ ФУНКЦИЙ ... 626
25.1. Производственные функции ... 626
25.2. Экономический смысл производственной функции ... 627
25.3. Основные виды производственных функций ... 628
25.4. Формальные свойства производственных функций ... 630
25.5. Средние и предельные значения производственной функции ... 631
25.6. Убывающая эффективность производства ... 634

Глава 26. МОДЕЛЬ ПОТРЕБИТЕЛЬСКОГО ВЫБОРА ... 636
26.1. Функция полезности ... 636
26.2. Линии безразличия ... 638
26.3. Бюджетное множество ... 640

26.4. Задача потребительского выбора 642
26.5. Решение задачи потребительского выбора 643
26.6. Функции спроса 645
26.7. Модель Р. Стоуна 646
26.8. Уравнение Слуцкого 647

Глава 27. ЭЛЕМЕНТЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ В ЭКОНОМИКЕ 651
27.1. Структура и классификация систем массового обслуживания 652
27.2. Основные показатели эффективности работы СМО 654
27.3. Случайный процесс в СМО 654
27.4. Одноканальная СМО с отказами 656
27.5. Многоканальная СМО с отказами 660
27.6. Многоканальная СМО с ожиданием и ограничением на длину очереди 665
27.7. Многоканальная СМО с ожиданием и неограниченной очередью 672

Глава 28. ЭЛЕМЕНТЫ УПРАВЛЕНИЯ РИСКАМИ В ЭКОНОМИКЕ 675
28.1. Оптимизация портфелей банка 676
28.2. Портфельный анализ 678
28.3. Методика VAR 683

Литература 686
Приложение 688
Предметный указатель 692

ПРЕДИСЛОВИЕ К ТРЕТЬЕМУ ИЗДАНИЮ

Этот учебник написан на основе лекций, прочитанных автором в течение ряда лет в экономических вузах Москвы. Быстрая математизация экономической науки и принятие нового государственного стандарта для подготовки специалистов с высшим образованием по экономическим специальностям создают предпосылки для ревизии содержания учебной литературы по математическим дисциплинам. Специфика преподавания математики в высшем экономическом образовании состоит прежде всего в том, что, во-первых, эта дисциплина должна быть математикой в экономике, а во-вторых, она должна состоять из ряда взаимосвязанных разделов с иллюстрацией применения в экономике.

В новом издании учебника содержатся необходимые основы математических дисциплин для вузов, осуществляющих высшее образование по экономическим специальностям: математический анализ функций одной и нескольких переменных, элементы линейной алгебры, основные положения теории обыкновенных дифференциальных уравнений, сведения из теории разностных уравнений, элементы теории вероятностей и математической статистики. Особое внимание уделено в учебнике математическому анализу, как фундаменту всех знаний в математике. Здесь изложение материала проведено особенно детально, с доказательством большого числа теорем. Именно на такой классической основе формируется строгость мышления будущего специалиста и определяется его индивидуальная направленность. Кроме того, поскольку существенное внимание при изучении естественнонаучных дисциплин в экономических вузах сегодня уделяется экономико-математическому модели-

рованию, в учебник включены соответствующие разделы, дающие представление об этом важном аспекте экономического образования. Эти разделы опираются на излагаемые в учебнике разделы математики, что обусловливает единую линию изложения материала.

Как и ранее, автор старался придерживаться классического изложения материала, хотя и с разной детальностью по разным разделам. Каждый раздел сопровождается решением большого числа характерных задач и соответствующих экономических приложений, сложность которых постепенно возрастает. Там, где приложения в экономике представляют самостоятельный интерес, они выделены в специальные разделы.

Памяти
Нины и Георгия Барабашевых
посвящаю

ВВЕДЕНИЕ

Математика — самая древняя наука, она развивается вместе с человечеством. Математика появилась, когда у человека возникла потребность в количественном отображении окружающего мира. Сначала это был простой подсчет расстояния от стойбища до мест охоты на диких зверей, числа членов племени и проч., затем стали рассчитывать площади обрабатываемых земель и т.д. С появлением цивилизации возникли начала геометрии как приемы проектирования и расчетов возводимых сооружений. Строители египетских пирамид обладали незаурядными по тем временам математическими знаниями.

Статус самостоятельной науки математика приобрела в древней Греции примерно в VI веке до н.э. Все философские школы того времени включали математику в круг вопросов миросозерцания; строгий язык формальной логики (именно он стал языком математики) формировал уровень и строй мышления. В III веке до н.э. математика выделилась из философии, что отражено в "Началах" — эпохальном труде, прославившем в веках имя Евклида, и заложившем фундамент классической геометрии. Более двух тысяч лет математику изучали по этой книге.

С древних веков развития цивилизации математика используется в строительстве и военном деле. Крепость Сиракузы долго оборонялась благодаря удивительным изобретениям великого математика и механика Архимеда. Аристотель использовал существовавшие в те времена геометрические построения для создания геоцентрической модели системы мира, где планеты, звезды и Солнце вращались на хрустальных сферах вокруг Земли. Усовершенствованная Птолемеем, эта

стройная и гармоничная система была в употреблении в течение почти двух тысячелетий.

Много веков после этого математика практически не эволюционировала; XVII век стал эпохой ее бурного развития. Применение математики Галилеем и Кеплером в исследовании движения небесных тел привело к поразительным по тому времени открытиям — законам движения планет вокруг Солнца. Труды Декарта, Ньютона и Лейбница ознаменовали новый этап развития математики и появление математики переменных величин. Начался период дифференциации единой науки на ряд самостоятельных математических наук: алгебру, математический анализ, аналитическую геометрию. В свою очередь, это способствовало интенсивному развитию физики и астрономии.

Новый этап развития математики в начале XIX века придал ей абстрактный характер, вне связи с наблюдаемыми явлениями окружающего мира. Имена Лобачевского и Римана стали достоянием истории. Созданная Кантором на рубеже XIX и XX столетий теория множеств стала фундаментом современной математики. Современный этап развития математики характеризуется возникновением внутри нее большого числа самостоятельных дисциплин.

Имена российских ученых занимают достойное место в истории развития математики: Н.И. Лобачевский (1792–1856), М.В. Остроградский (1801–1861), П.Л. Чебышев (1821–1894), А.А. Марков (1856–1922) и другие. Достижения современной математики во многом обусловлены трудами известных российских ученых В.И. Арнольда, С.Н. Бернштейна, Л.В. Канторовича, А.Н. Колмогорова, И.Г. Петровского, Ю.В. Прохорова, А.Н. Тихонова и многих других.

Математика интенсивно проникает в другие науки; это происходит благодаря дифференциации математики на ряд самостоятельных областей. Язык математики универсален, он объективно отражает универсальность законов окружающего нас многообразного мира.

Экономика как наука об объективных причинах функционирования и развития общества еще со времен Адама Смита пользуется разнообразными количественными характеристиками, а потому опирается на большое число математических

дисциплин. Сегодня на первый план выступает математическая модель как инструмент исследования и прогноза экономических явлений. Историческая цепочка развития математики "явления окружающего мира — математические построения и абстракции — использование в объяснении явлений окружающего мира" отразилась в развитии математических моделей.

Математическую модель можно определить как внутренне непротиворечивую замкнутую систему математических соотношений (объект конечной сложности), предназначенную для воспроизведения определенного качества (или группы определенных качеств) изучаемого реального явления или процесса. Математические модели представляют собой основу компьютерного моделирования и обработки информации. Они способствуют развитию наших представлений о закономерностях экономических процессов и, безусловно, обозначают новый образ мышления и анализа. В этом смысле математическая модель — это абстракция на определенном уровне реального процесса из окружающего нас мира. Как образ мышления математические модели могут эволюционировать и совершенствоваться.

Характерным свойством математических моделей является их общность или безотносительность к реальным явлениям: одна и та же модель может, например, описывать процессы прироста народонаселения и радиоактивного распада. Более того, наши представления о процессах, которые мы не сможем увидеть или измерить, например ядерные реакции в недрах звезд или распад оледенений Земли в далеком прошлом, во многом основаны на соответствующих математических моделях.

Активно использовать математический аппарат возможно только овладев необходимыми знаниями из области аксиоматической математики, т.е. той математики, что ведет свое начало еще со времен Евклида. В ее основу заложен принцип принятия без доказательства некоторых фундаментальных предположений — аксиом; затем на этой основе строится вся математическая теория при помощи некоторой совокупности определений и доказываемых предположений — теорем.

Математические доказательства представляют собой строгие логические рассуждения. В этом смысле математика явля-

ется наиболее простой наукой, нежели другие — скажем, науки об обществе: она не допускает множественного трактования; для опровержения какого-либо предположения достаточно привести всего лишь один противоречащий пример. Однако в такой простоте скрыта сила логических построений и умозаключений, которая позволяет оттачивать методику исследований сложных процессов, происходящих в природе и обществе.

Изучение математических дисциплин, составляющих основу современной математики, позволит будущему специалисту сформировать необходимые компоненты мышления: уровень, кругозор и культуру, которые понадобятся ему для успешной работы и ориентации в будущей профессиональной деятельности.

Часть 1. МАТЕМАТИЧЕСКИЙ АНАЛИЗ

Математический анализ представляет собой основу всей высшей математики. Его содержание составляют дифференциальное и интегральное исчисления одной и несколько переменных.

Глава 1
МНОЖЕСТВА

1.1. Понятие множества и операции над множествами

Понятие множества является одним из основных в математике. Система, семейство, совокупность — эти термины можно считать синонимами слова "множество". *Множество* можно определить как совокупность объектов, объединенных по определенному признаку. Например, множество зрителей в данном кинотеатре, множество студентов определенного учебного заведения, совокупность отличников и "хорошистов" в некоторой школе, совокупность коммерческих банков, имеющих уставный фонд не ниже 100 млн. рублей. Множество может содержать конечное или бесконечное число объектов.

Объекты, составляющие множество, называются его *элементами*, или *точками*. Обычно множества обозначаются большими буквами, а входящие в них элементы — малыми буквами. Элемент x из множества X соответствует записи $x \in X$ (x принадлежит X), если же элемент x не входит в

множество X, то это соответствует записи $x \notin X$ (x не принадлежит X).

Пусть X и Y — два множества. Тогда между ними можно определить следующие соотношения. Если оба множества состоят из одних и тех же элементов, то они совпадают, что соответствует записи $X = Y$. Если все элементы множества X содержатся в множестве Y, то X целиком содержится в Y, или $X \subset Y$ (X является *подмножеством* множества Y). Если ни один элемент множества X не содержится в Y, то, значит, и само множество X не содержится в Y, или $X \not\subset Y$.

В математике используется понятие *пустого множества*, обозначаемого символом $\varnothing$. Это множество, в котором не содержится ни одного элемента, и потому оно является подмножеством любого множества.

Введем также понятия суммы множеств и их пересечения. *Суммой*, или *объединением, множеств* X и Y называется совокупность элементов, входящих как в множество X, так и в множество Y (обладающих либо свойством множества X, либо свойством множества Y). Сумма этих множеств обозначается $X \cup Y$. Например, пусть X — множество государственных предприятий с годовым оборотом не ниже S денежных единиц, а Y множество негосударственных предприятий с тем же нижним порогом годового оборота; тогда $X \cup Y$ будет множеством всех предприятий с указанным нижним ограничением S. Отметим, что добавление пустого множества $\varnothing$ к любому множеству X не меняет этого множества, т.е.

$$X \cup \varnothing = X. \tag{1.1}$$

Пересечением множеств X и Y (или их общей частью) является совокупность элементов, входящих как в множество X, так и в множество Y; это множество обозначается $X \cap Y$ (заштрихованная область на рис. 1.1). Например, если X — это множество предприятий с годовым оборотом T не ниже s, а Y — совокупность предприятий с годовым оборотом не более S, причем $s < S$, то в пересечение $X \cap Y$ войдут объекты с годовым оборотом T, удовлетворяющим неравенству

$$s \leqslant T \leqslant S. \tag{1.2}$$

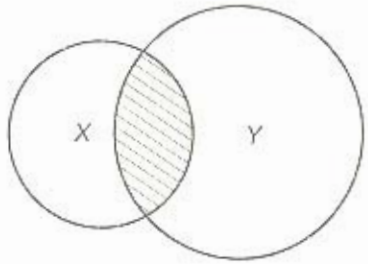

Рис. 1.1

Отсутствие элементов со свойствами множеств X и Y одновременно означает, что пересечение этих множеств представляет собой пустое множество $\varnothing$.

Разностью множеств X и Y называется множество Z, содержащее все элементы множества X, не содержащиеся в Y; эта разность обозначается $Z = X \setminus Y$.

В общем случае сложение и пересечение определяются для любого конечного числа множеств путем последовательного попарного проведения соответствующих операций.

При записи математических выражений целесообразно употреблять логическую символику. Вместо выражения "любое x из множества X" употребляют запись $\forall x \in X$, где перевернутая латинская буква $\forall$ (квантор общности) взята от начала английского слова *Any* — любой. Аналогично вместо выражения "существует элемент x из множества X" кратко пишут: $\exists x \in X$, где перевернутая латинская буква $\exists$ (квантор существования) является начальной в английском слове *Existence* — существование.

1.2. Вещественные числа и их свойства

Поскольку, как уже говорилось во Введении, мы изучаем аксиоматическую математику, необходимо строго определять объекты, которыми мы оперируем, а также их свойства.

Множество вещественных чисел является бесконечным. Оно состоит из рациональных и иррациональных чисел. *Рациональным* называется число вида p/q, где p и q — целые числа. Всякое вещественное число, не являющееся рациональным, называется *иррациональным*. Всякое рациональное число либо

является целым, либо представляет собой конечную или периодическую бесконечную десятичную дробь. Например рациональное число $1/9$ можно представить в виде $0{,}11111\ldots$. Иррациональное число представляет собой бесконечную непериодическую десятичную дробь, например, $\sqrt{2} = 1{,}41421356\ldots$, $\pi = 3{,}14159265\ldots$.

Перечислим основные свойства вещественных чисел.

I. Сложение и умножение вещественных чисел

Для любой пары вещественных чисел a и b определены единственным образом два вещественных числа $a + b$ и ab, называемые соответственно их *суммой* и *произведением*. Для любых чисел a, b и c имеют место следующие свойства:

1°. $a + b = b + a$, $ab = ba$ — переместительное свойство.

2°. $a + (b + c) = (a + b) + c$, $a(bc) = (ab)c$ — сочетательное свойство.

3°. $(a + b)c = ac + bc$ — распределительное свойство.

4°. *Существует единственное число* 0 *такое, что* $a + 0 = a$ *для любого числа* a.

5°. *Для любого числа* a *существует такое число* $(-a)$, *что* $a + (-a) = 0$.

6°. *Существует единственное число* $1 \neq 0$, *что для любого числа* a *имеет место равенство* $a \cdot 1 = a$.

7°. *Для любого числа* $a \neq 0$ *существует такое число* a^{-1}, *что* $aa^{-1} = 1$. *Число* a^{-1} *обозначается также символом* $\frac{1}{a}$.

II. Сравнение вещественных чисел

Для любых двух вещественных чисел установлено одно из трех соотношений: $a = b$ (a равно b), $a > b$ (a больше b) или $a < b$ (a меньше b). Отношение равенства обладает свойством транзитивности: если $a = b$ и $b = c$, то $a = c$.

Отношение "больше" обладает следующими свойствами. Для любых чисел a, b и c:

8°. *Если* $a > b$ *и* $b > c$, *то* $a > c$.

9°. *Если* $a > b$, *то* $a + c > b + c$.

10°. *Если* $a > 0$ *и* $b > 0$, *то* $ab > 0$.

Вместо соотношения $a > b$ употребляют также $b < a$. Запись $a \geqslant b$ ($b \leqslant a$) означает, что либо $a = b$, либо $a > b$. Соотношения со знаками $>$, $<$, $\geqslant$ и $\leqslant$ называются неравенствами, причем соотношения типа 8°–10° — строгими неравенствами.

11°. *Любое вещественное число можно приблизить рациональными числами с произвольной точностью.*

Для определенности будем считать это число неотрицательным и представим его в виде бесконечной дроби $a = a_0, a_1a_2\ldots a_n\ldots$. Обрывая эту дробь на n-м знаке после запятой, получим рациональное число $a_0, a_1a_2\ldots a_n$. Увеличим его на $1/10^n$. Тогда справедливо двойное неравенство

$$a_0, a_1a_2\ldots a_n \leqslant a \leqslant a_0, a_1a_2\ldots a_n + \frac{1}{10^n},$$

т.е. число a заключено между двумя рациональными числами. Поскольку n можно взять любым, мы доказали справедливость исходного утверждения.

III. Непрерывность вещественных чисел

12°. *Пусть* X *и* Y *— два множества вещественных чисел. Тогда, если для любых чисел* $x \in X$ *и* $y \in Y$ *выполняется неравенство* $x \leqslant y$, *существует хотя бы одно такое число* c, *что для всех* x *и* y *выполняются неравенства* $x \leqslant c \leqslant y$.

Свойством непрерывности обладает множество всех вещественных (действительных) чисел, но не обладает множество, состоящее только из рациональных чисел.

Из основных свойств I–III вещественных чисел вытекают все остальные их свойства, в том числе и известные из курса школьной математики. Можно определить, что вещественные числа представляют собой множество элементов, обладающих свойствами I–III. Такое определение, из которого выводятся остальные свойства, называется *аксиоматическим*, а сами свойства I–III — *аксиомами* вещественных чисел.

1.3. Числовая прямая (числовая ось) и множества на ней

Введем понятие о соответствии множеств; позже мы еще раз обратимся к нему в подпараграфе о функциональной зависимости (см. параграф 3.1). Будем говорить, что между множествами X и Y установлено соответствие, если по какому-либо закону или правилу каждому элементу $x \in X$ соответствует элемент $y \in Y$. Соответствие называется однозначным, если $\forall x \in X$ соответствует только один элемент из Y. Оказывается, что между множеством вещественных чисел и множеством точек на прямой может быть установлено взаимно однозначное соответствие. Это дает возможность наглядного геометрического изображения вещественных чисел на числовой оси. На прямой выберем точку O начало отсчета, укажем направление отсчета (обычно слева направо) и единицу измерения, или масштаб (рис. 1.2), тем самым полностью определив числовую (координатную) ось. На ней вещественные числа изображаются в виде точек. Пусть M — произвольная точка на этой оси. Поставим ей в соответствие число x, равное по величине длине отрезка OM со знаком "+", если точка M находится справа от начала отсчета, или со знаком "−", если M расположена слева от точки O. Тогда число x называется *координатой* точки x. Справедливо и обратное утверждение: каждому вещественному числу x соответствует определенная точка на координатной оси, координата которой равна x.

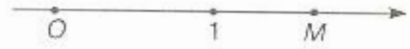

Рис. 1.2

Укажем некоторые наиболее употребительные числовые множества. Пусть a и b — два числа, причем $a < b$. Используем следующие обозначения:

а) множество всех чисел, удовлетворяющих неравенству $a \leqslant x \leqslant b$, называется *отрезком* (сегментом) $[a, b]$;

б) множество всех чисел, удовлетворяющих строгому неравенству $a < x < b$, называется *интервалом* (a, b);

в) множество всех вещественных (действительных) чисел будем обозначать

$$-\infty < x < \infty \quad \text{или} \quad (-\infty, \infty);$$

г) аналогично можно определить числовые множества типа $(a, b]$, $[a, b)$, $(a, +\infty)$, $(-\infty, b)$, $[a, +\infty)$ и $(-\infty, b]$. Первые два множества называются *полуинтервалами*.

Все перечисленные множества носят название *промежутков*: отрезки, интервалы и полуинтервалы называются *конечными* промежутками, а числа a и b их концами; остальные — *бесконечными*.

Числовым промежуткам соответствуют промежутки на координатной оси. Сегмент $[a, b]$ изображается отрезком M_1M_2, точка M_1 которого имеет координату a, а точка M_2 — координату b. Вся координатная прямая является изображением множества всех вещественных чисел, поэтому множество $(-\infty, \infty)$ называется *числовой прямой (осью)*, а любому числу соответствует точка на этой прямой.

1.4. Грани числовых множеств

Будем говорить, что множество X ограничено сверху (снизу), если существует такое число d, при котором для любого $x \in X$ выполняется неравенство $x \leqslant d$ ($x \geqslant d$). Число d тогда называется *верхней (нижней) гранью* множества X. Множества, ограниченные снизу и сверху, называются *ограниченными*. Любой конечный промежуток ограничен. Интервалы $(a, +\infty)$ и $(-\infty, b)$ представляют собой множества, ограниченные соответственно снизу (сверху), но не ограниченные сверху (снизу). Вся числовая прямая не ограничена ни снизу, ни сверху.

Любое ограниченное сверху (снизу) множество имеет бесконечное число верхних (нижних) граней. Действительно, если число d является верхней гранью множества X, то и любое число $d_1 > d$, согласно определению верхней грани, также будет верхней гранью этого множества. Наименьшая верхняя грань множества X, ограниченного сверху, называется *точной верхней гранью* этого множества; она обозначается символом $\sup X$. Наибольшая нижняя грань ограниченного снизу множества X

называется *точной нижней гранью* этого множества и обозначается символом $\inf X$. Эти символы образованы от латинских слов: *supremum* — наивысший и *infimum* — наинизший.

Приведем некоторые примеры. Пусть $X = (a, b)$. В таком случае числа a и b являются соответственно точными нижней и верхней гранями множества X, т.е. $a = \inf X$, $b = \sup X$.

Пусть $X = (-\infty, b)$. Тогда нижних граней (в том числе и точной нижней грани) множество X не имеет, а число b является его точной верхней гранью: $b = \sup X$.

Теперь докажем теорему существования точной верхней (нижней) грани множества, ограниченного сверху (снизу).

Теорема 1.1. *Если непустое числовое множество ограничено сверху (снизу), то оно имеет точную верхнюю (нижнюю) грань.*

Доказательство. Пусть X — непустое множество, ограниченное сверху. Тогда существует непустое множество Y чисел, ограничивающих множество X. Согласно определению верхней грани следует, что для любого $x \in X$ и любого $y \in Y$ имеет место неравенство $x \leqslant y$. В силу свойства III непрерывности вещественных чисел существует такое число c, что для любых x и y справедливы неравенства

$$x \leqslant c \leqslant y. \tag{1.3}$$

Из первого неравенства (1.3) следует, что число c является верхней гранью множества X. Из второго неравенства (1.3) следует, что число c — наименьшее из всех верхних граней, т.е. оно является точной верхней гранью множества X, что и требовалось доказать.

В случае ограниченного снизу множества доказательство существования точной нижней грани проводится аналогично.

1.5. Абсолютная величина числа

По определению абсолютная величина вещественного числа x (модуль числа)

$$|x| = \begin{cases} x, & \text{если } x \geqslant 0, \\ -x, & \text{если } x < 0. \end{cases} \tag{1.4}$$

Рассмотрим свойства абсолютной величины.

1°. $|x| \geqslant 0$.

Это свойство непосредственно вытекает из определения модуля числа: если число неотрицательно, его абсолютная величина равна самому числу; модуль отрицательного числа равен этому числу, взятому с обратным знаком.

2°. $|x| = |-x|$.

Также следует из определения модуля числа (1.4).

3°. $-|x| \leqslant x \leqslant |x|$.

Если $x \geqslant 0$, то $|x| = x$ и тогда $-x \leqslant 0$, отсюда $-x \leqslant |x|$, т.е. $-|x| \leqslant x = |x|$;

если $x < 0$, то по выражению (1.4) $|x| = -x$, откуда $-|x| = x$. Но так как $x < -x = |x|$, то и получаем $-|x| = x < |x|$.

4°. Пусть a — *положительное число, тогда неравенства* $|x| \leqslant a$ *и* $-a \leqslant x \leqslant a$ *равносильны.*

Пусть $|x| \leqslant a$, выведем отсюда второе неравенство.
Если $x \geqslant 0$, то $|x| = x$, т.е. $x \leqslant a$, или $0 \leqslant x \leqslant a$.
Если $x < 0$, то $|x| = -x$, откуда $-x \leqslant a$, или $-a \leqslant x < 0$.
Объединяя полученные неравенства, получаем

$$-a \leqslant x \leqslant a.$$

Пусть теперь справедливо неравенство $-a \leqslant x \leqslant a$; покажем, что отсюда следует первое неравенство. Действительно, одновременно выполняются неравенства $x \leqslant a$ и $x \geqslant -a$. Из второго неравенства имеем: $-x \leqslant a$. По определению модуля числа [см. формулу (1.4)] он равен либо x, либо $-x$, т.е. отсюда и следует, что $|x| \leqslant a$.

5°. *Для любых двух действительных чисел* x *и* y *справедливо неравенство*

$$|x + y| \leqslant |x| + |y|. \tag{1.5}$$

По свойству 3°

$$-|x| \leqslant x \leqslant |x|, \qquad -|y| \leqslant y \leqslant |y|.$$

Складывая эти неравенства почленно, получаем

$$-(|x| + |y|) \leqslant x + y \leqslant (|x| + |y|).$$

Но по свойству 4° эти неравенства равносильны двойному неравенству (1.5).

В это свойство можно включить также и двойное неравенство

$$|x - y| \leqslant |x| + |y|. \tag{1.6}$$

6°. *Для любых двух действительных чисел x и y справедливо неравенство*

$$|x - y| \geqslant |x| - |y|. \tag{1.7}$$

Очевидно тождество для любых двух чисел x и y

$$x = y + (x - y).$$

По свойству 5°

$$|x| = |y + (x - y)| \leqslant |y| + |x - y|,$$

откуда и получаем утверждение (1.7).

Глава 2
ПРЕДЕЛ ПОСЛЕДОВАТЕЛЬНОСТИ

2.1. Числовые последовательности

Числовые последовательности и операции над ними

Числовые последовательности представляют собой бесконечные множества чисел. Примерами последовательностей могут служить: последовательность всех членов арифметической и геометрической прогрессий, последовательность приближенных значений $\sqrt{2}$, т.е. $x_1 = 1$, $x_2 = 1{,}4$, $x_3 = 1{,}41$, ..., последовательность периметров правильных n-угольников, вписанных в данную окружность. Уточним понятие числовой последовательности.

Определение 1. Если каждому числу n из натурального ряда чисел $1, 2, 3, \ldots, n, \ldots$ поставлено в соответствие вещественное число x_n, то множество вещественных чисел

$$x_1, x_2, x_3, \ldots, x_n, \ldots \tag{2.1}$$

называется *числовой последовательностью* или просто *последовательностью*.

Числа $x_1, x_2, x_3, \ldots, x_n, \ldots$ будем называть *элементами*, или *членами*, последовательности (2.1), символ x_n — *общим элементом*, или членом, последовательности, а число n — его *номером*. Сокращенно последовательность (2.1) будем обозначать символом $\{x_n\}$. Например, символ $\{1/n\}$ обозначает последовательность чисел

$$1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots.$$

Иными словами, под последовательностью можно понимать бесконечное множество занумерованных элементов или множество пар чисел (n, x_n), в которых первое число последовательно принимает значения $1, 2, 3, \ldots$. Последовательность считается заданной, если указан способ получения любого ее элемента. Например, формула $x_n = 1 + (-1)^n$ определяет последовательность $0, 2, 0, 2, \ldots$.

Геометрически последовательность изображается на числовой оси в виде последовательности точек, координаты которых равны соответствующим членам последовательности. На рис. 2.1 изображена последовательность $\{x_n\} = \{1/n\}$ на числовой прямой.

Рис. 2.1

Рассмотрим арифметические действия над числовыми последовательностями. Пусть даны последовательности $\{x_n\}$ и $\{y_n\}$.

Суммой последовательностей назовем последовательность $x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n, \ldots$, или $\{x_n + y_n\}$, *разностью* — последовательность $\{x_n - y_n\}$, или $x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n, \ldots$.

Произведением последовательности $\{x_n\}$ *на число* m назовем последовательность $\{mx_n\} = mx_1, mx_2, mx_3, \ldots, mx_n, \ldots$.

Произведением последовательностей назовем последовательность $x_1y_1, x_2y_2, x_3y_3, \ldots, x_ny_n, \ldots$.

Частным назовем последовательность

$$\frac{x_1}{y_1}, \frac{x_2}{y_2}, \frac{x_3}{y_3}, \ldots, \frac{x_n}{y_n}, \ldots,$$

если все члены последовательности $\{y_n\}$ отличны от нуля.

Ограниченные и неограниченные последовательности

Поскольку последовательности являются числовыми множествами, то естественно ввести понятие их ограниченности, как это было сделано для множеств в параграфе 1.4.

Определение 2. Последовательность $\{x_n\}$ называется *ограниченной сверху (снизу)*, если существует такое число M (число m), что любой элемент x_n этой последовательности удовлетворяет неравенству $x_n \leqslant M$ ($x_n \geqslant m$).

Определение 3. Последовательность $\{x_n\}$ называется *ограниченной*, если она ограничена и сверху и снизу, т.е. существуют такие числа m и M, что любой элемент $\{x_n\}$ этой последовательности удовлетворяет неравенствам $m \leqslant x_n \leqslant M$.

Пусть $A = \max\{|m|, |M|\}$, тогда условие ограниченности последовательности можно записать в форме $|x_n| \leqslant A$.

Определение 4. Последовательность $\{x_n\}$ называется *неограниченной*, если для любого положительного числа A существует элемент x_n этой последовательности, удовлетворяющий неравенству $|x_n| > A$, (т.е. либо $x_n > A$, либо $x_n < -A$).

Из определений 2–4 следует, что все элементы ограниченной сверху последовательности принадлежат промежутку $(-\infty, M)$; а ограниченной снизу — лежат на промежутке $(m, +\infty)$. Неограниченная последовательность может быть ограничена сверху (снизу). Все элементы ограниченной последовательности принадлежат отрезку $[m, M]$.

Рассмотрим несколько примеров последовательностей.

1) Последовательность $-1, -2, -3, \ldots, -n, \ldots$ не ограничена снизу, но ограничена сверху.

2) Последовательность $1, 1/2, 1/3, \ldots, 1/n, \ldots$ ограничена, так как любой элемент x_n этой последовательности удовлетворяет неравенствам $0 \leqslant x_n \leqslant 1$.

3) Последовательность $-1, 2, -3, \ldots, (-1)^n n$ неограниченная. Действительно, каково бы ни было число $A > 0$, среди элементов x_n этой последовательности найдутся элементы, для которых будет выполняться неравенство $|x_n| > A$.

Бесконечно большие и бесконечно малые последовательности

Определение 5. Последовательность $\{x_n\}$ называется *бесконечно большой*, если для любого положительного числа A существует такой номер N, что при $n > N$ (для всех элементов последовательности с номерами $n > N$) выполняется неравенство $|x_n| > A$.

Очевидно, что любая бесконечно большая последовательность является неограниченной. Обратное утверждение неверно: неограниченная последовательность может и не быть бесконечно большой. Например, неограниченная последовательность $1, 2, 1, 3, 1, 4, \ldots, 1, n, \ldots$ не является бесконечно большой, так как при $A > 1$ неравенство $|x_n| > A$ не выполняется для всех элементов x_n с нечетными номерами.

Определение 6. Последовательность $\{\alpha_n\}$ называется *бесконечно малой*, если для любого положительного числа ε существует такой номер N, что при $n > N$ выполняется неравенство $|\alpha_n| < \varepsilon$.

Приведем примеры бесконечно больших и бесконечно малых последовательностей.

1. Докажем, используя определение 5, что последовательность $\{n\}$ является бесконечно большой. Возьмем любое число $A > 0$. Из неравенства $|x_n| = |n| > A$ получаем $n > A$. Если выбрать $N \geqslant A$, то для всех $n > N$ будет выполняться неравенство $|x_n| > A$.

2. Докажем, что последовательность $\{1/n\}$ является бесконечно малой. Возьмем произвольное число $\varepsilon > 0$. Из неравенства $|\alpha_n| < \varepsilon$ определения 6 получаем, что $1/n < \varepsilon$, т.е. $n > 1/\varepsilon$. Если принять $N = [1/\varepsilon]$[1], то для всех $n > N$ будет выполняться неравенство $n \geqslant [1/\varepsilon]+1 > 1/\varepsilon$, т.е. $1/n < \varepsilon$. Тогда, согласно определению 6, последовательность $\{1/n\}$ является бесконечно малой.

Теперь докажем теорему, устанавливающую связь между бесконечно малыми и бесконечно большими последовательностями.

Теорема 2.1. *Если $\{x_n\}$ — бесконечно большая последовательность и все ее члены отличны от нуля, то последовательность $\{1/x_n\}$ бесконечно малая. Если все элементы бесконечно малой последовательности $\{\alpha_n\}$ отличны от нуля, то последовательность $\{1/\alpha_n\}$ бесконечно большая.*

Доказательство. Пусть $\{x_n\}$ — бесконечно большая последовательность. Для произвольно взятого числа $\varepsilon > 0$ положим $A = 1/\varepsilon$. Согласно определению 5, для этого A существует такой номер N, что при $n > N$ выполняется неравенство $|x_n| > A$. Отсюда получаем: $|1/x_n| = 1/|x_n| < 1/A = \varepsilon$ для всех $n > N$, т.е. последовательность $\{1/x_n\}$ является бесконечно малой, согласно определению 6.

Доказательство второй части теоремы проводится аналогично.

Основные свойства бесконечно малых последовательностей

Теорема 2.2. *Сумма и разность двух бесконечно малых последовательностей являются бесконечно малыми последовательностями.*

[1]Символ $[a]$ означает целую часть числа a, т.е. наибольшее целое число, не превосходящее a. Например, $[2] = 2$, $[2,5] = 2$, $[0,8] = 0$, $[-0,5] = -1$, $[-23,7] = -24$.

Доказательство. Пусть $\{\alpha_n\}$ и $\{\beta_n\}$ — бесконечно малые последовательности. Докажем, что последовательность $\{\alpha_n \pm \beta_n\}$ бесконечно малая. Пусть ε — произвольное положительное число, N_1 — номер, начиная с которого $|\alpha_n| < \dfrac{\varepsilon}{2}$, а N_2 — номер, начиная с которого $|\beta_n| < \dfrac{\varepsilon}{2}$ (такие номера найдутся согласно определению бесконечно малой последовательности 6). Возьмем $N = \max(N_1, N_2)$. Тогда при $n > N$ одновременно выполняются два неравенства: $|\alpha_n| < \dfrac{\varepsilon}{2}$ и $|\beta_n| < \dfrac{\varepsilon}{2}$. Значит, при всех $n > N$

$$|\alpha_n \pm \beta_n| \leqslant |\alpha_n| + |\beta_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Но тогда, согласно определению 6, последовательность $\{\alpha_n \pm \pm \beta_n\}$ бесконечно малая, что и требовалось доказать.

Следствие. *Алгебраическая сумма любого конечного числа бесконечно малых последовательностей есть бесконечно малая последовательность.*

Теорема 2.3. *Произведение двух бесконечно малых последовательностей есть бесконечно малая последовательность.*

Доказательство. Пусть $\{\alpha_n\}$ и $\{\beta_n\}$ — бесконечно малые последовательности. Докажем, что последовательность $\{\alpha_n\beta_n\}$ бесконечно малая. Возьмем для первой последовательности произвольное положительное число ε, а для последовательности $\{\beta_n\}$ в качестве ε возьмем 1. Согласно определению бесконечно малой последовательности: для $\{\alpha_n\}$ найдется такой номер N_1, что $|\alpha_n| < \varepsilon$ при $n > N_1$; а для $\{\beta_n\}$ найдется такой номер N_2, что $|\beta_n| < 1$ при $n > N_2$. Примем $N = \max(N_1, N_2)$, тогда при $n > N$ будут выполнены оба неравенства. Значит, при $n > N$

$$|\alpha_n\beta_n| = |\alpha_n||\beta_n| < \varepsilon \cdot 1 = \varepsilon.$$

Это и означает, что последовательность $\{\alpha_n\beta_n\}$ бесконечно малая.

Следствие. *Произведение конечного числа бесконечно малых последовательностей есть бесконечно малая последовательность.*

Теорема 2.4. *Произведение бесконечно малой последовательности на ограниченную последовательность является бесконечно малой последовательностью.*

Доказательство. Пусть $\{x_n\}$ — ограниченная, а $\{\alpha_n\}$ — бесконечно малая последовательности. Докажем, что $\{x_n\alpha_n\}$ бесконечно малая последовательность. В силу ограниченности последовательности $\{x_n\}$ существует такое число A, что любой элемент ее удовлетворяет неравенству $|x_n| \leqslant A$. Поскольку последовательность $\{\alpha_n\}$ бесконечно малая, для положительного числа $\dfrac{\varepsilon}{A}$ существует такой номер N, что при всех $n > N$ выполняется неравенство $|\alpha_n| < \dfrac{\varepsilon}{A}$. Значит, при $n > N$

$$|x_n\alpha_n| = |x_n||\alpha_n| < A\frac{\varepsilon}{A} = \varepsilon.$$

Это и означает, что последовательность $|x_n\alpha_n|$ бесконечно малая.

Следствие. *Произведение бесконечно малой последовательности на постоянное число является бесконечно малой последовательностью.*

2.2. Сходящиеся последовательности

Понятие сходящейся последовательности

Определение 1. Число a называется *пределом последовательности* $\{x_n\}$, если для любого положительного числа ε существует такой номер N, что при всех $n > N$ выполняется неравенство

$$|x_n - a| < \varepsilon. \tag{2.2}$$

Последовательность, имеющая предел, называется *сходящейся*. Если предел последовательности равен числу a, это записывается так:

$$\lim_{n\to\infty} x_n = a, \quad \text{или} \quad x_n \to a \quad \text{при} \quad n \to \infty. \tag{2.3}$$

Последовательность, не имеющая предела, называется *расходящейся*.

Замечание 1. Пусть последовательность $\{x_n\}$ имеет своим пределом число a. Тогда последовательность $\{\alpha_n\} = \{x_n - a\}$ есть бесконечно малая, так как для любого $\varepsilon > 0$ существует такой номер N, что при $n > N$ выполняется неравенство $|\alpha_n| = |x_n - a| < \varepsilon$. Следовательно, любой элемент x_n сходящейся последовательности, имеющей предел a, можно представить в виде:

$$x_n = a + \alpha_n, \tag{2.4}$$

где α_n — элемент бесконечно малой последовательности $\{\alpha_n\}$.

Замечание 2. Неравенство (2.2) эквивалентно неравенствам (см. свойство 4° модуля числа из параграфа 1.5)

$$-\varepsilon < x_n - a < \varepsilon \quad \text{или} \quad a - \varepsilon < x_n < a + \varepsilon.$$

Это означает, что при $n > N$ все элементы последовательности $\{x_n\}$ находятся в ε-окрестности точки (рис. 2.2), причем номер N определяется величиной ε. Поэтому определение сходящейся последовательности можно сформулировать в следующем виде.

Определение 2. Последовательность $\{x_n\}$ называется *сходящейся*, если существует такое число a, что в любой его ε-окрестности находятся все элементы этой последовательности, начиная с некоторого номера, зависящего от ε.

Рис. 2.2

Интересно дать геометрическую интерпретацию этого определения. Поскольку последовательность представляет собой бесконечное множество чисел, то, если она сходится, в любой ε-окрестности точки a на числовой прямой находится бесконечное число точек — элементов этой последовательности, тогда как вне ε-окрестности остается конечное число элементов. Поэтому предел последовательности часто называют точкой сгущения.

Замечание 3. Бесконечно большая последовательность не имеет предела. Часто говорят, что она имеет *бесконечный предел*; записывается это в следующем виде:

$$\lim_{n\to\infty} x_n = \infty. \tag{2.5}$$

Если при этом, начиная с некоторого номера, все члены последовательности положительны (отрицательны), то пишут

$$\lim_{n\to\infty} x_n = +\infty \left(\lim_{n\to\infty} x_n = -\infty\right).$$

Определенный выше предел последовательности, в отличие от бесконечного предела, называют *конечным пределом*.

Замечание 4. Очевидно, что всякая бесконечно малая последовательность сходится и имеет своим пределом число $a = 0$.

Приведем примеры сходящихся и расходящихся последовательностей.

1. Покажем, используя определение предела последовательности, что $\lim\limits_{n\to\infty} \dfrac{n}{n+1} = 1$.

Возьмем любое число $\varepsilon > 0$. Так как $|x_n - 1| = \left|\dfrac{n}{n+1} - 1\right| = \dfrac{1}{n+1}$, то для удовлетворения неравенству (2.2) достаточно решить неравенство $1/(n+1) < \varepsilon$, откуда получаем $n > (1-\varepsilon)/\varepsilon$. Если принять $N = [(1-\varepsilon)/\varepsilon]$ (целая часть числа $(1-\varepsilon)/\varepsilon$), неравенство $|x_n - 1| < \varepsilon$ будет выполняться при всех $n > N$.

2. Последовательность $\{x_n\} = (-1)^n$ или $-1, 1, -1, 1, \ldots$ не имеет предела. Действительно, какое бы число мы ни предположили в качестве предела, 1 или -1, при $\varepsilon < 0,5$ неравенство (2.2), определяющее предел последовательности, не удовлетворяется: вне ε-окрестности этих чисел остается бесконечное число элементов x_n: все элементы с нечетными номерами равны -1, элементы с четными номерами равны 1.

3. Докажем, что последовательность $\{x_n\} = 0{,}3,\ 0{,}33,\ 0{,}333, \ldots, 0,\underbrace{33\ldots3}_{n\text{ раз}}, \ldots$ сходится и имеет своим пределом число $\dfrac{1}{3}$.

Поскольку число $\dfrac{1}{3}$ можно представить бесконечной десятичной дробью $0{,}333\ldots$, то справедливы неравенства:

$$0,\underbrace{33\ldots3}_{n\text{ раз}} \leqslant \frac{1}{3} \leqslant 0,\underbrace{33\ldots3}_{n\text{ раз}} + \frac{1}{10^n}.$$

Из этих неравенств получаем, что $\left|x_n - \dfrac{1}{3}\right| < \dfrac{1}{10^n}$. При $n \geqslant N$ имеем $1/10^n \leqslant 1/10^N$, так что, выбрав по любому $\varepsilon > 0$ номер N из условия $1/10^N \leqslant \varepsilon$, получим $\left|x_n - \dfrac{1}{3}\right| < \varepsilon$ при всех $n > N$.

Основные свойства сходящихся последовательностей

Теорема 2.5. *Если все элементы бесконечно малой последовательности $\{x_n\}$ равны одному и тому же числу c, то $c = 0$.*

Доказательство. Предположим противное, т.е. $c \neq 0$. Возьмем $\varepsilon = \dfrac{|c|}{2}$. По определению бесконечно малой последовательности существует такой номер N, что при $n > N$ выполняется неравенство $|x_n| < \varepsilon$. Так как $x_n = c$, то последнее неравенство можно переписать в виде $|c| < \dfrac{|c|}{2}$; это неравенство неверно при $c \neq 0$, что и доказывает теорему.

Теперь докажем теорему о единственности предела сходящейся последовательности.

Теорема 2.6. *Сходящаяся последовательность имеет только один предел.*

Доказательство. Проведем доказательство теоремы от противного, т.е. предположим, что сходящаяся последовательность $\{x_n\}$ имеет два предела a и b. Тогда по формуле (2.4) для представления элементов сходящейся последовательности имеем

$$x_n = a + \alpha_n \quad \text{и} \quad x_n = b + \beta_n,$$

где α_n и β_n — элементы бесконечно малых последовательностей $\{\alpha_n\}$ и $\{\beta_n\}$. Вычитая одно соотношение из другого, получаем, что $\alpha_n - \beta_n = b - a$. Но поскольку все элементы бесконечно

3 Зак. 6853

малой последовательности $\{\alpha_n - \beta_n\}$ равны постоянной величине $b - a$, то по теореме 2.5 следует, что $b - a = 0$, т.е. $b = a$, что и доказывает утверждение теоремы.

Теорема 2.7. *Сходящаяся последовательность ограничена.*

Доказательство. Пусть число a — предел сходящейся последовательности $\{x_n\}$. Тогда для произвольного положительного числа ε существует такой номер N, что при всех $n > N$ выполняется неравенство $|x_n - a| < \varepsilon$. Проведем оценку элементов x_n по модулю:

$$|x_n| = |(x_n - a) + a| \leqslant |x_n - a| + |a| < \varepsilon + a$$

для всех $n > N$. Поскольку вне ε-окрестности числа a осталось конечное количество элементов последовательности (не более N), то можно указать максимум из совокупности чисел $A = \max\{|a|+\varepsilon, |x_1|, |x_2|, \ldots, |x_N|\}$. Очевидно, что $|x_n| < A$ для всех номеров n, что и означает ограниченность последовательности $\{x_n\}$.

Замечание 5. Ограниченная последовательность может и не быть сходящейся. Так, последовательность $\{x_n\} = (-1)^n$, рассмотренная в примере 2, ограничена, но не имеет предела.

Далее рассмотрим предельные переходы при операциях с последовательностями.

Теорема 2.8. *Сумма (разность) сходящихся последовательностей $\{x_n\}$ и $\{y_n\}$ есть сходящаяся последовательность, предел которой равен сумме (разности) пределов последовательностей $\{x_n\}$ и $\{y_n\}$.*

Доказательство. Пусть a и b — соответственно пределы последовательностей $\{x_n\}$ и $\{y_n\}$. Тогда по формуле (2.4)

$$x_n = a + \alpha_n, \qquad y_n = b + \beta_n,$$

где α_n и β_n — элементы бесконечно малых последовательностей соответственно $\{\alpha_n\}$ и $\{\beta_n\}$.

Отсюда получаем

$$(x_n \pm y_n) - (a \pm b) = (x_n - a) \pm (y_n - b) = \alpha_n \pm \beta_n.$$

По теореме 2.2 последовательность $\{\alpha_n \pm \beta_n\}$ бесконечно малая, т.е. последовательность $\{(x_n \pm y_n) - (a \pm b)\}$ также бесконечно малая, откуда и следует, что число $(a \pm b)$ является пределом сходящейся последовательности $\{x_n \pm y_n\}$.

Теорема 2.9. *Произведение сходящихся последовательностей $\{x_n\}$ и $\{y_n\}$ есть сходящаяся последовательность, предел которой равен произведению пределов последовательностей $\{x_n\}$ и $\{y_n\}$.*

Доказательство. Пусть a и b — соответственно пределы последовательностей $\{x_n\}$ и $\{y_n\}$. Тогда, используя формулу (2.4) и перемножая соответствующие члены указанных последовательностей, получаем

$$x_n y_n - ab = a\beta_n + b\alpha_n + \alpha_n\beta_n.$$

В силу теорем 2.2–2.4 последовательность $\{a\beta_n + b\alpha_n + \alpha_n\beta_n\}$ бесконечно малая, т.е. последовательность $\{x_n y_n - ab\}$ также бесконечно малая, а значит, последовательность $\{x_n y_n\}$ сходится и имеет своим пределом число ab.

Теорема 2.10. *Частное двух сходящихся последовательностей $\{x_n\}$ и $\{y_n\}$ при условии, что предел последовательности $\{y_n\}$ отличен от нуля, есть сходящаяся последовательность, предел которой равен частному пределов последовательностей $\{x_n\}$ и $\{y_n\}$.*

Доказательство. Пусть a и b $(b \neq 0)$ — соответственно пределы последовательностей $\{x_n\}$ и $\{y_n\}$. В силу условия $b \neq 0$ элементы y_n не обращаются в нуль, начиная с некоторого номера N, и поэтому частное x_n/y_n имеет смысл для всех $n > N$. Используя формулу (2.4), получаем

$$\frac{x_n}{y_n} - \frac{a}{b} = \frac{b(a + \alpha_n) - a(b + \beta_n)}{by_n} = \frac{1}{y_n}\left(\alpha_n - \frac{a}{b}\beta_n\right).$$

В силу теорем 2.2–2.4 последовательность $\{\alpha_n - \frac{a}{b}\beta_n\}$ бесконечно малая.

Покажем теперь, что последовательность $\{1/y_n\}$ ограниченная. Поскольку $y_n \to b$ при $n \to \infty$, то для $\varepsilon = \frac{|b|}{2}$ найдется

такой номер N, что при всех $n > N$ выполняется неравенство $|y_n - b| < \frac{|b|}{2}$. Поэтому

$$|y_n| = |b - (b - y_n)| \geqslant |b| - |y_n - b| > |b| - \frac{|b|}{2} = \frac{|b|}{2}.$$

Так как $|y_n| > \frac{|b|}{2}$, то $|1/y_n| < \frac{2}{|b|}$ для всех $n > N$, что и означает ограниченность последовательности $\{1/y_n\}$. В силу теоремы 2.4 последовательность $\left\{\frac{x_n}{y_n} - \frac{a}{b}\right\}$ как произведение бесконечно малой последовательности на ограниченную последовательность есть бесконечно малая. Значит, последовательность $\{x_n/y_n\}$ сходится и имеет своим пределом число $\frac{a}{b}$.

Теоремы 2.5–2.10 имеют большое теоретическое и практическое значение и широко используются при вычислении пределов. Проиллюстрируем это на примерах

Пример 1. Найти предел $\lim\limits_{n\to\infty} \dfrac{3n^2 + 2n + 4}{4n^2 + n - 3}$.

При $n \to \infty$ числитель и знаменатель дроби стремятся к бесконечности, т.е. применить сразу теорему о пределе частного нельзя, так как она предполагает существование конечных пределов последовательностей. Преобразуем данную последовательность, разделив числитель и знаменатель на n^2. Применяя затем теоремы о пределе частного, пределе суммы и снова о пределе частного, последовательно найдем:

$$\lim_{n\to\infty} \frac{3n^2+2n+4}{4n^2+n-3} = \lim_{n\to\infty} \frac{3+2/n+4/n^2}{4+1/n-3/n^2} = \frac{\lim\limits_{n\to\infty}(3+2/n+4/n^2)}{\lim\limits_{n\to\infty}(4+1/n-3/n^2)} =$$

$$= \frac{\lim\limits_{n\to\infty} 3 + \lim\limits_{n\to\infty}(2/n) + \lim\limits_{n\to\infty}(4/n^2)}{\lim\limits_{n\to\infty} 4 + \lim\limits_{n\to\infty}(1/n) - \lim\limits_{n\to\infty}(3/n^2)} = \frac{3+0+0}{4+0-0} = \frac{3}{4}.$$

Пример 2. Найти предел последовательности $\{x_n\} = \dfrac{\sqrt{n}\cos n}{n+1}$ при $n \to \infty$.

Здесь, как и в предыдущем примере, числитель и знаменатель не имеют конечных пределов, и потому сначала необходимо выполнить соответствующие преобразования. Поделив

числитель и знаменатель на n, получаем:

$$\lim_{n\to\infty}\{x_n\} = \lim_{n\to\infty} \frac{(1/\sqrt{n})\cos n}{1+1/n} = \frac{\lim\limits_{n\to\infty}(1/\sqrt{n}) \lim\limits_{n\to\infty}\cos n}{\lim\limits_{n\to\infty} 1 + \lim\limits_{n\to\infty}(1/n)}.$$

Поскольку в числителе стоит произведение бесконечно малой последовательности на ограниченную последовательность, то в силу теоремы 2.4 окончательно получаем

$$\lim_{n\to\infty}\{x_n\} = \frac{0}{1+0} = 0.$$

Пример 3. Найти предел последовательности $\{x_n\} = \sqrt{n+1} - \sqrt{n}$ при $n \to \infty$.

Здесь применить непосредственно теорему о пределе суммы (разности) последовательностей нельзя, так как конечных пределов слагаемых в формуле для $\{x_n\}$ не существует. Сначала умножим и разделим формулу для $\{x_n\}$ на сопряженное выражение $\sqrt{n+1} + \sqrt{n}$:

$$\lim_{n\to\infty}\{x_n\} = \lim_{n\to\infty} \frac{\left(\sqrt{n+1}-\sqrt{n}\right)\left(\sqrt{n+1}+\sqrt{n}\right)}{\sqrt{n+1}+\sqrt{n}} = \lim_{n\to\infty} \frac{n+1-n}{\sqrt{n+1}+\sqrt{n}} =$$

$$= \lim_{n\to\infty} \frac{1}{\sqrt{n+1}+\sqrt{n}} = \frac{\lim\limits_{n\to\infty}(1/\sqrt{n})}{\lim\limits_{n\to\infty}\left(\sqrt{1+1/n}\right)+1} = \frac{0}{1+1} = 0.$$

Предельный переход в неравенствах

Итак, арифметические операции над сходящимися последовательностями приводят к таким же операциям над их пределами. Покажем, что неравенства, которым удовлетворяют элементы сходящихся последовательностей, переходят в пределе в соответствующие неравенства для пределов.

Теорема 2.11. *Если элементы сходящейся последовательности $\{x_n\}$ удовлетворяют неравенству $x_n \geqslant b$ ($x_n \leqslant b$), начиная с некоторого номера, то и предел a этой последовательности удовлетворяет неравенству $a \geqslant b$ ($a \leqslant b$).*

Доказательство. Проведем доказательство от противного, т.е. предположим, что $a < b$. Возьмем положительное число $\varepsilon = b - a$; для него существует такой номер N, что при всех $n > N$ выполняется неравенство $|x_n - a| < \varepsilon = b - a$, что, в свою очередь, равносильно двум неравенствам: $-(b-a) < x_n - a < b - a$. Отсюда следует, что $x_n < b$, что противоречит условию теоремы, т.е. неравенство $a < b$ невозможно.

Следствие 1. *Если элементы сходящихся последовательностей $\{x_n\}$ и $\{y_n\}$, начиная с некоторого номера, удовлетворяют неравенству $x_n \leqslant y_n$, то их пределы удовлетворяют неравенству* $\lim\limits_{n\to\infty} x_n \leqslant \lim\limits_{n\to\infty} y_n$.

Действительно, начиная с некоторого номера, элементы последовательности $\{y_n - x_n\}$ неотрицательны, а потому, согласно теореме 2.11, неотрицателен и ее предел: $\lim\limits_{n\to\infty}(y_n - x_n) = \lim\limits_{n\to\infty} y_n - \lim\limits_{n\to\infty} x_n \geqslant 0$. Отсюда и получаем, что $\lim\limits_{n\to\infty} x_n \leqslant \lim\limits_{n\to\infty} y_n$.

Следствие 2. *Если все элементы сходящейся последовательности $\{x_n\}$ находятся на отрезке $[a, b]$, то и ее предел находится на этом же отрезке.*

Пусть c — предел последовательности $\{x_n\}$. Так как $a \leqslant x_n \leqslant b$, то $a \leqslant c \leqslant b$.

Теорема 2.12. *Если $\{x_n\}$, $\{y_n\}$ и $\{z_n\}$ — сходящиеся последовательности, причем $x_n \leqslant y_n \leqslant z_n$ для всех n, а последовательности $\{x_n\}$ и $\{z_n\}$ имеют одинаковый предел a, тогда последовательность $\{y_n\}$ имеет тот же предел a.*

Доказательство. Пусть ε — произвольное положительное число. По этому ε можно указать такой номер N_1 для последовательности $\{x_n\}$, что при $n > N_1$ будет справедливо неравенство $|x_n - a| < \varepsilon$, или

$$a - \varepsilon < x_n < a + \varepsilon. \tag{2.6}$$

По тому же числу ε для последовательности $\{z_n\}$ найдется такой номер N_2, что $|z_n - a| < \varepsilon$ при всех $n > N_2$, или

$$a - \varepsilon < z_n < a + \varepsilon. \tag{2.7}$$

Возьмем за N номер $N = \max(N_1, N_2)$, тогда при $n > N$ будут одновременно выполняться неравенства (2.6) и (2.7). Используя левое неравенство из (2.6) и правое неравенство из (2.7), по условию теоремы получаем

$$a - \varepsilon < x_n \leqslant y_n \leqslant z_n < a + \varepsilon$$

при $n > N$. Отсюда при всех $n > N$ выполняются неравенства

$$a - \varepsilon < y_n < a + \varepsilon, \quad \text{или} \quad |y_n - a| < \varepsilon.$$

Последнее неравенство означает, что число a — предел последовательности $\{y_n\}$, что и требовалось доказать.

Эта теорема играет важную роль в различных приложениях.

2.3. Монотонные последовательности

Определение монотонных последовательностей

Монотонные последовательности классифицируются по соотношению между соседними элементами. Такие соотношения могут иметь один из следующих видов и выполняются для всех n:

$$x_n < x_{n+1}, \tag{2.8}$$

$$x_n \leqslant x_{n+1}, \tag{2.9}$$

$$x_n > x_{n+1}, \tag{2.10}$$

$$x_n \geqslant x_{n+1}. \tag{2.11}$$

Определение 1. Последовательность $\{x_n\}$ называется *возрастающей*, если для ее элементов выполняется условие (2.8); *неубывающей*, если выполняется условие (2.9); *убывающей*, если выполняется условие (2.10); *невозрастающей*, если выполняется условие (2.11).

Все эти последовательности называются *монотонными последовательностями*. Последовательности типов (2.8) и (2.10) называются *строго монотонными*. Монотонные последовательности ограничены либо сверху (невозрастающие и убывающие последовательности), либо снизу (возрастающие и неубывающие последовательности). К ограниченным монотонным последовательностям относятся возрастающие и неубывающие последовательности, ограниченные сверху, а также убывающие и невозрастающие последовательности, ограниченные снизу.

Приведем примеры монотонных последовательностей:

1) последовательность $1, 1/3, 1/5, \ldots, 1/(2n-1), \ldots$ — убывающая и ограниченная;

2) последовательность $1, 1, 1/2, 1/2, \ldots, 1/n, 1/n, \ldots$ — невозрастающая и ограниченная;

3) последовательность $1, 2, 3, \ldots, n, \ldots$ — возрастающая и неограниченная;

4) последовательность $1, 1, 2, 2, \ldots, n, n, \ldots$ — неубывающая и неограниченная;

5) последовательность $1/2, 2/3, 3/4, \ldots, n/(n+1)$ — возрастающая и ограниченная (снизу числом $1/2$, сверху — числом 1);

6) последовательность $-1, 1/2, -1/3, \ldots, (-1)^n/n, \ldots$ ограниченная, но не является монотонной, поскольку для всех n не выполняется ни одно из условий (2.8)–(2.11).

Признак сходимости монотонной последовательности

Теорема 2.13. *Монотонная ограниченная последовательность сходится.*

Доказательство. Ограничимся случаем неубывающей последовательности. Итак, по условию теоремы $x_n \leqslant x_{n+1}$ для всех n, а также существует такое число M, что для всех элементов x_n выполняется неравенство $x_n \leqslant M$. Данная последовательность представляет собой числовое множество, ограниченное сверху и непустое. В силу теоремы 1.1 это множество имеет точную верхнюю грань; обозначим это число через a и докажем, что оно является пределом данной последовательности.

Согласно свойству точной верхней грани числового множества $\{x_n\}$, для любого $\varepsilon > 0$ найдется такой номер N, что $x_N \geqslant a - \varepsilon$. Так как $\{x_n\}$ — неубывающая последовательность, то для всех $n > N$ будет $x_n > a - \varepsilon$. Но, по определению верхней грани, $x_n \leqslant a + \varepsilon$ при всех n. Следовательно, при $n > N$ справедливы неравенства $a - \varepsilon < x_n < a + \varepsilon$, т.е. $|x_n - a| < \varepsilon$. Это означает, что число a является пределом последовательности $\{x_n\}$, что и требовалось доказать.

Это основная теорема о свойстве монотонных последовательностей. Немонотонные ограниченные последовательности

этим свойством обладают не всегда. Например, последовательность $\{1 + (-1)^n\}$ ограничена, но немонотонна и не сходится.

Заметим, что *ограниченность* монотонной последовательности — это необходимое и достаточное условие ее сходимости. Действительно, если монотонная последовательность ограничена, то по теореме 2.13 она сходится; если же последовательность сходится, то в силу теоремы 2.7 она ограничена.

Число e

Применим теорему 2.13 для доказательства сходимости последовательности $\{x_n\}$, общий член которой выражается формулой

$$x_n = \left(1 + \frac{1}{n}\right)^n.$$

Докажем, что эта последовательность монотонно возрастает и ограничена сверху, а значит, имеет предел. Применив формулу бинома Ньютона, получаем

$$x_n = 1 + n\frac{1}{n} + \frac{n(n-1)}{2!}\frac{1}{n^2} + \frac{n(n-1)(n-2)}{3!}\frac{1}{n^3} + \cdots + \frac{n(n-1)(n-2)\ldots[n-(n-1)]}{n!}\frac{1}{n^n}.$$

Упростим это выражение и представим его в следующей форме:

$$x_n = 2 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \cdots + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{n-1}{n}\right). \tag{2.12}$$

Аналогичным образом запишем элемент x_{n+1}:

$$x_{n+1} = 2 + \frac{1}{2!}\left(1 - \frac{1}{n+1}\right) + \frac{1}{3!}\left(1 - \frac{1}{n+1}\right)\left(1 - \frac{2}{n+1}\right) + \cdots + \frac{1}{(n+1)!}\left(1 - \frac{1}{n+1}\right)\left(1 - \frac{2}{n+1}\right)\ldots\left(1 - \frac{n}{n+1}\right).$$

Поскольку $\left(1-\frac{k}{n}\right) < \left(1-\frac{k}{n+1}\right)$ для любого $0 < k < n$, то каждое слагаемое в выражении для x_{n+1} больше соответствующего слагаемого в выражении для x_n, начиная со второго (первые члены разложений у них равны 2). Значит, $x_n < x_{n+1}$, т.е. последовательность $\{x_n\}$ возрастающая.

Теперь докажем, что последовательность $\{x_n\}$ ограниченная (ограничена сверху). Во-первых, каждое выражение в скобках соотношения (2.12) меньше единицы. Во-вторых, при $n > 2$ имеем: $\frac{1}{n!} < \frac{1}{2^{n-1}}$. С учетом сказанного получаем неравенства

$$x_n < 2 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} < 1 + \left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{n-1}}\right).$$

Поскольку в скобках неравенства справа находится сумма геометрической прогрессии, то приходим к неравенству

$$x_n < 3 - \frac{1}{2^{n-1}} < 3.$$

Итак, последовательность $\{(1+1/n)^n\}$ монотонно возрастает и ограничена, т.е. по теореме 2.13 она имеет предел. Этот предел называют числом e. Следовательно, по определению,

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n. \qquad (2.13)$$

Число e играет большую роль в математике. Далее будет рассмотрен способ его вычисления с любой требуемой точностью. Отметим здесь, что число e является иррациональным и заключено в пределах $2 < e < 3$.

В заключение рассмотрим два примера из экономики на использование числа e как оценки сверху.

Пример 1. Известно, что *формула сложных процентов* имеет вид

$$Q = Q_0\left(1+\frac{p}{100}\right)^n, \qquad (2.14)$$

где Q — сумма вкладов по истечении n периодов; Q_0 — первоначальная сумма вклада в банк; p — процент начисления за определенный период (месяц, год); n — количество периодов хранения вклада.

Формулы типа (2.14) используются также в демографических расчетах (прирост народонаселения) и в экономических прогнозах (увеличение валового национального продукта).

Пусть первоначальный депозит Q_0 помещен в банк под $p = 100\%$ годовых, тогда через год сумма депозита составит $2Q_0$. Предположим, что через полгода счет закрыт с результатом $Q_1 = Q_0\left(1+\frac{1}{2}\right) = \frac{3}{2}Q_0$ и эта сумма вновь помещена в качестве депозита в том же банке. В конце года депозит будет составлять $Q_2 = Q_0\left(1+\frac{1}{2}\right)^2 = 2{,}25Q_0$. Будем уменьшать срок размещения депозита в банке при условии его последующего размещения после изъятия. При ежеквартальном повторении этих операций депозит в конце года составит $Q_3 = Q_0\left(1+\frac{1}{3}\right)^3 \approx \approx 2{,}37Q_0$.

Если банк настолько щедр, что позволяет повторять операцию изъятие-размещение в течение года сколько угодно раз, то при ежемесячном манипулировании сумма за год составит $Q_{12} = Q_0\left(1+\frac{1}{12}\right)^{12} \approx 2{,}61Q_0$; при ежедневном посещении банка $Q_{365} = Q_0\left(1+\frac{1}{365}\right)^{365} \approx 2{,}714Q_0$; при ежечасной активности $Q_{8720} = Q_0\left(1+\frac{1}{8720}\right)^{8720} \approx 2{,}718Q_0$ и т.д. Нетрудно заметить, что последовательность значений увеличения первоначального вклада $\{q_n\} = \{Q_n/Q_0\}$ совпадает с последовательностью, пределом которой является число e при $n \to \infty$, согласно формуле (2.13). Таким образом, доход, который можно получить при непрерывном использовании процентов на проценты, может составить за год не более чем

$$\lim_{n\to\infty}(Q_n - Q_0)\cdot 100\%/Q_0 = (e-1)\cdot 100\% \approx 172\%.$$

В общем случае, если p — процент начисления и год разбит на n частей, то через t лет сумма депозита достигнет величины

$$Q_n = Q_0\left(1+\frac{r}{n}\right)^{nt},$$

где $r = p/100$. Это выражение можно преобразовать

$$Q_n = Q_0 \left[\left(1 + \frac{r}{n}\right)^{\frac{n}{r}} \right]^{rt}.$$

Если ввести новую переменную $m = \dfrac{n}{r}$, то при $n \to \infty$ получим $m \to \infty$, или

$$\lim_{n\to\infty} Q_n = \lim_{m\to\infty} Q_0 \left[\left(1 + \frac{1}{m}\right)^{m} \right]^{rt} = Q_0 e^{rt}.$$

Расчеты, выполненные по этой формуле, называют *вычислениями по непрерывным процентам.*

Пример 2. Пусть темп инфляции составляет 1% в день. Насколько уменьшится первоначальная сумма через полгода?

Применение формулы сложных процентов дает

$$Q = Q_0 \left(1 - \frac{1}{100}\right)^{182},$$

где Q_0 — первоначальная сумма; 182 — число дней в полугодии.

Преобразуя это выражение, получаем

$$Q = Q_0 \left[\left(1 - \frac{1}{100}\right)^{-100} \right]^{-182/100} \approx Q_0/e^{1{,}82},$$

т.е. инфляция уменьшит первоначальную сумму примерно в 6 раз.

Последовательность вложенных отрезков

Возьмем последовательность таких отрезков $[a_1, b_1]$, $[a_2, b_2]$, $\ldots$, $[a_n, b_n], \ldots$, что каждый последующий отрезок содержится в предыдущем, т.е. $[a_1, b_1] \supset [a_2, b_2] \supset \cdots \supset [a_n, b_n] \supset \ldots$. Это можно записать в виде системы неравенств, справедливых для всех n:

$$a_n \leqslant a_{n+1} < b_{n+1} \leqslant b_n. \qquad (2.15)$$

Пусть при этом

$$\lim_{n\to\infty} (b_n - a_n) = 0, \qquad (2.16)$$

т.е. длина отрезков неограниченно уменьшается. Такая совокупность отрезков называется *последовательностью вложенных отрезков.*

Теорема 2.14. *Для последовательности вложенных отрезков существует единственная точка, принадлежащая всем отрезкам этой последовательности.*

Доказательство. Как следует из системы неравенств (2.15), левые концы отрезков образуют неубывающую последовательность

$$a_1 \leqslant a_2 \leqslant a_3 \leqslant \cdots \leqslant a_n \leqslant \ldots, \qquad (2.17)$$

тогда как правые концы — невозрастающую последовательность

$$b_1 \geqslant b_2 \geqslant b_3 \geqslant \cdots \geqslant b_n \geqslant \ldots. \qquad (2.18)$$

Последовательность (2.17) ограничена сверху, так как $a_n \leqslant b_1$ для всех n. Последовательность (2.18) ограничена снизу, так как $b_n \geqslant a_1$ для всех n. В силу теоремы 2.13 эти последовательности являются сходящимися. Пусть $\lim\limits_{n\to\infty} a_n = c_1$, а $\lim\limits_{n\to\infty} b_n = c_2$. Тогда по условию (2.16) получаем, что

$$\lim_{n\to\infty} b_n - \lim_{n\to\infty} a_n = c_1 - c_2 = 0,$$

откуда следует, что $c_1 = c_2$, т.е. последовательности (2.17) и (2.18) имеют общий предел (обозначим его буквой c). Таким образом, для любого n справедливы неравенства $a_n \leqslant c \leqslant b_n$. Это означает, что точка c принадлежит всем отрезкам последовательности (2.15).

Покажем, что общая точка c является единственной. Допустим обратное, т.е. что существует еще одна такая точка c' ($c \neq c'$). Но тогда для всех n должны выполняться неравенства $b_n - a_n \geqslant |c - c'|$, а значит, $\lim\limits_{n\to\infty} (b_n - a_n) \geqslant |c - c'| > 0$, что противоречит условию (2.16) теоремы. Доказательство теоремы 2.14 завершено.

Замечание. Теорема становится неверной, если вместо отрезков взять систему интервалов. Рассмотрим в качестве примера последовательность вложенных интервалов

$$(0,1) \supset (0,1/3) \supset (0,1/9) \supset \cdots \supset (0,1/3^n) \supset \ldots. \qquad (2.19)$$

Для этой последовательности выполнены условия теоремы 2.14, однако не существует точки, принадлежащей всем интервалам. Действительно, какую бы точку c на интервале $(0,1)$ ни взять, можно найти такой номер N, что $1/3^n < c$ при $n > N$ и, следовательно, точка c не принадлежит интервалам последовательности (2.19), начиная с интервала $(0,1/3^{N+1})$.

Глава 3
ФУНКЦИИ ОДНОЙ ПЕРЕМЕННОЙ

3.1. Понятие функции

Определение функциональной зависимости

Пусть X и Y — некоторые числовые множества и пусть каждому элементу $x \in X$ по какому-либо закону f поставлен в соответствие один элемент $y \in Y$. Тогда будем говорить, что определена функциональная зависимость y от x по закону $y = f(x)$. При этом x называют *независимой переменной* (или *аргументом*), y — *зависимой переменной*, множество X — *областью определения* (существования) функции, множество Y — *областью значений* (изменения) функции. Совокупность точек координатной плоскости xOy, удовлетворяющих уравнению $y = f(x)$, называется *графиком* этой функции.

Для обозначения функции и независимой переменной могут быть использованы и другие буквы. Примеры записи функций: $y = y(x)$, $y = F(x)$, $y = g(x)$.

Способы задания функций

Задать функцию — значит указать закон, по которому, согласно определению, каждому значению аргумента из области определения ставится в соответствие (вычисляется) значение

зависимой переменной из области значений функции. Существует три основных способа задания функций: *табличный, аналитический и графический.*

Табличный способ. Этот способ имеет широкое применение в разных отраслях знаний и приложениях: ряды экспериментальных измерений, социологические опросы, таблицы бухгалтерской отчетности и банковской деятельности и т.п. Как правило, в таких таблицах по крайней мере одну из переменных можно принять за независимую (например, время), тогда другие величины будут функциями от этого аргумента. По сути дела, базы данных основаны на табличном способе задания, хранения и обработки информации, а значит, и на табличной форме функциональной зависимости.

Табличный способ задания функций удобен еще по двум причинам. Во-первых, методами интерполяции можно вычислять не содержащиеся в таблице значения функции, соответствующие промежуточным значениям аргумента. Во-вторых, по данным таблицы можно приближенно установить аналитический способ задания функции.

Аналитический способ. Этот способ состоит в задании связи между аргументом и функцией в виде формул. Следует подчеркнуть, что функция может определяться и набором формул: разным участкам области определения функции соответствуют разные формулы.

Приведем примеры аналитического задания функций.

1. $y = x^3$. Эта функция задана на бесконечной прямой $-\infty < x < \infty$. Множество значений этой функции — тоже бесконечная числовая прямая $-\infty < \dot{x} < \infty$. Функция называется кубической параболой (рис. 3.1).

2. $y = \sqrt{1 - x^2}$. Функция задана на отрезке $[-1,1]$, множество ее значений — отрезок $[0,1]$. Это половина окружности, лежащая в верхней полуплоскости (рис. 3.2).

3. $y = \operatorname{sign} x = \begin{cases} +1, & \text{если} \quad x > 0, \\ 0, & \text{если} \quad x = 0, \\ -1, & \text{если} \quad x < 0. \end{cases}$

Термин sign происходит от латинского *signum* — знак. Функция задана на всем бесконечном промежутке $(-\infty, \infty)$, а область ее значений состоит из трех чисел: $-1, 0, 1$ (рис. 3.3).

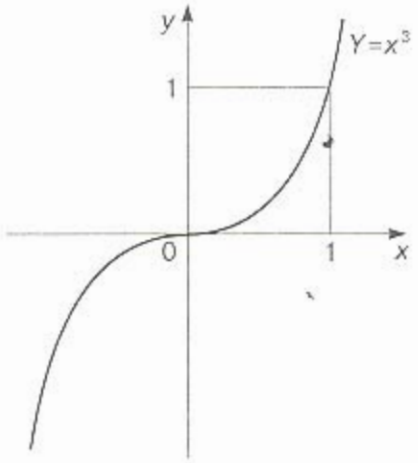

Рис. 3.1

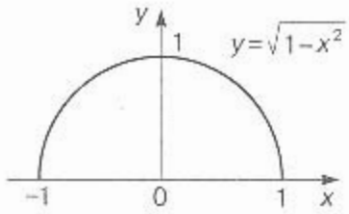

Рис. 3.2

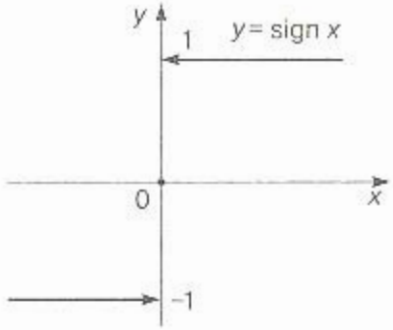

Рис. 3.3

Стрелки означают, что полупрямые не достигают точек на оси ординат, так как при $x = 0$ значение функции определено по другому соответствию.

4. $y = [x]$ — целая часть значения аргумента. Функция задана для всех вещественных значений x, а множество ее значений состоит из целых чисел (рис. 3.4).

5. $y = \begin{cases} x & \text{при} \quad x \leqslant 0, \\ \sqrt{x} & \text{при} \quad x > 0. \end{cases}$

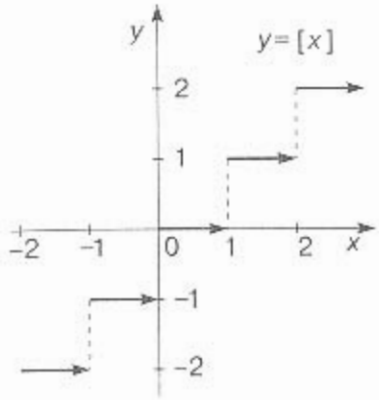

Рис. 3.4

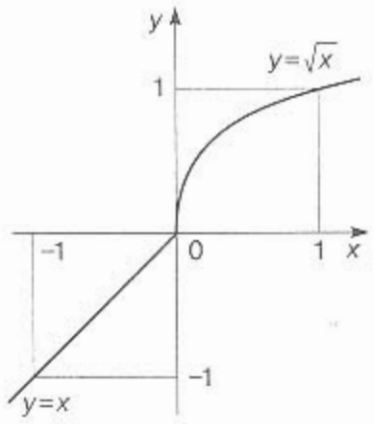

Рис. 3.5

Функция определена на бесконечном промежутке $(-\infty, \infty)$, область ее изменения — также бесконечный промежуток $(-\infty, \infty)$ (рис. 3.5).

6. Функция Дирихле[1]

$$y = \begin{cases} 0, & \text{если } x \text{ — иррациональное число,} \\ 1, & \text{если } x \text{ — рациональное число.} \end{cases}$$

Функция Дирихле определена на всей числовой прямой, а множество ее значений состоит из двух точек: 0 и 1. Изобразить ее графически невозможно.

Графический способ. Здесь соответствие между аргументом и функцией задается посредством графика. Этот способ обычно используется в экспериментальных измерениях с употреблением самопишущих приборов (осциллографы, сейсмографы и т.п.).

Область определения функции

Остановимся особо на процедуре нахождения области определения функции.

[1] Петер Густав Лежен Дирихле (1805–1859) — немецкий математик.

I. В том случае, когда функция задана аналитически (посредством формулы)

$$y = f(x) \tag{3.1}$$

и никаких ограничений или оговорок более не имеется, область ее определения находится только по соблюдению законности выполнения математических операций, входящих в формулу f. Эти ограничения хорошо известны: подкоренное выражение в корне четной степени не может быть отрицательным, знаменатель дроби не может быть равным нулю, выражение под знаком логарифма должно быть только положительным, а также некоторые другие.

Приведем несколько примеров.

Область определения функции $y = \log_2(x^2 - 5x + 6)$ находится из условия $x^2 - 5x + 6 > 0$. Поскольку $x = 2$ и $x = 3$ — корни квадратного трехчлена, стоящего под знаком логарифма, то это условие выполняется на двух полубесконечных интервалах: $(-\infty, 2)$ и $(3, \infty)$. На рис. 3.6 выделена заштрихованная полоса, в которой график функции отсутствует.

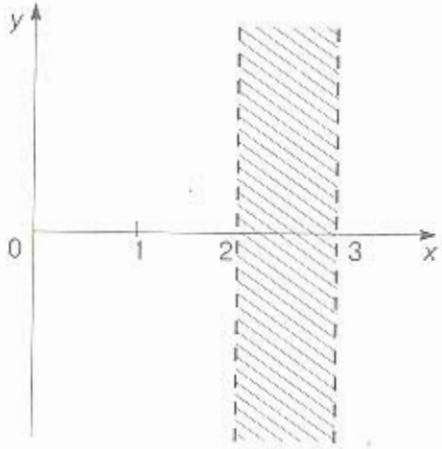

Рис. 3.6

Область определения функции $y = \arcsin \dfrac{1}{x+2}$ находится из совокупности двух условий: аргумент под знаком arcsin не

может быть по модулю больше единицы и знаменатель аргумента не должен равняться нулю, т.е.

$$-1 \leqslant \frac{1}{x+2} \leqslant 1, \quad x \neq -2.$$

Двойное неравенство эквивалентно двум более простым неравенствам: $x+2 \geqslant 1$ и $x+2 \leqslant -1$. Отсюда получаем, что область определения функции состоит из двух полубесконечных промежутков: $(-\infty, -3]$ и $[-1, \infty)$. Запретная точка $x = -2$ сюда не попадает. В отличие от предыдущего примера концы полуинтервалов входят в область определения функции.

2. Область определения функции задана вместе с функцией $f(x)$. Например, $y = 3x^{4/3} + 2$, $1 \leqslant x \leqslant 4$.

3. Функция имеет прикладной характер, и область ее существования определяется также и реальными значениями входящих параметров. Рассмотрим простой пример.

Как известно из курса физики, удлинение металлического стержня при нагреве описывается формулой

$$\Delta l = \alpha \Delta T, \tag{3.2}$$

где α — коэффициент линейного удлинения материала (м/°C); ΔT — изменение температуры (°C).

С точки зрения математики никаких запретных значений аргумента ΔT нет, и график зависимости увеличения длины стержня Δl имеет вид прямой (рис. 3.7). Однако существуют физические ограничения, например никакой металл не выдержит нагрев на 10 000°C. Следовательно, область существования функции (3.2) определяется как свойствами металла, из которого изготовлен стержень, так и условиями эксперимента. Иными словами, существует интервал значений ΔT, вне которого зависимость (3.2) не имеет места.

Аналогичный смысл вкладывается в решение любых прикладных задач, например: требование положительного знака величины скорости, ограничение снизу на скорость летящего самолета и т.п.

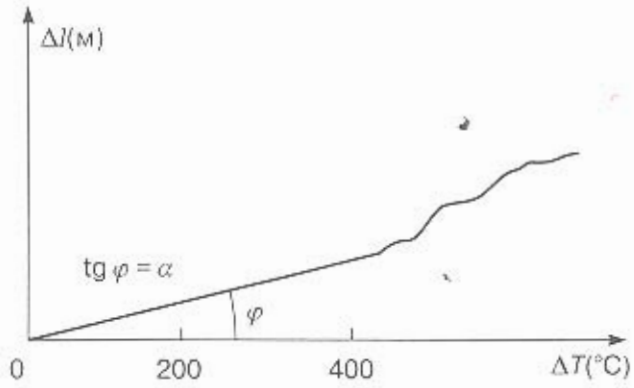

Рис. 3.7

Параллельный перенос и растяжение (сжатие) графиков функций

Напомним правила переноса и растяжения (сжатия) графиков функций. Эти приемы в значительной мере позволяют упростить анализ и построение графиков функций, заданных в аналитическом виде.

Пусть функция $y = f(x)$ известна.

1. График функции $y = f(x - a)$ получается из графика $y = f(x)$ переносом его вдоль координатной оси Ox на a единиц вправо (влево), если a — положительное (отрицательное) число.

2. График функции $y = f(x) + b$ можно получить путем переноса графика $y = f(x)$ вдоль координатной оси Oy вверх (вниз), если b — положительное (отрицательное) число.

3. График функции $y = f(kx)$ получается путем растяжения графика $y = f(x)$ вдоль оси Ox в $1/k$ раз при $k < 1$ (при $k > 1$ сжатие графика в k раз).

4. График функции $y = kf(x)$ получается путем растяжения графика $y = f(x)$ вдоль оси Oy в k раз при $k > 1$ (при $k < 1$ сжатие графика в $1/k$ раз).

Рассмотрим поясняющий пример. Построить график функции $y = \dfrac{1}{1-2x}$, если известен график гиперболы $y = \dfrac{1}{x}$. Перепишем формулу заданной функции в виде $y = -\dfrac{1}{2\left(x - \dfrac{1}{2}\right)}$.

Отсюда уже понятен цикл переносов и растяжений, преобразующих известную гиперболу $y = 1/x$ в заданную:

1) "переворачиваем" график гиперболы $y = 1/x$ — зеркально отображаем его относительно оси Ox (реализуем знак "минус" перед дробью в преобразованной функции), т.е. получаем функцию $y = -1/x$;

2) переносим полученный график на $1/2$ вправо вдоль оси Ox — получаем функцию $y = -1/(x - 1/2)$;

3) сжимая полученный график функции вдоль оси Oy в 2 раза, получаем заданную функцию.

На рис. 3.8 показан описанный выше процесс.

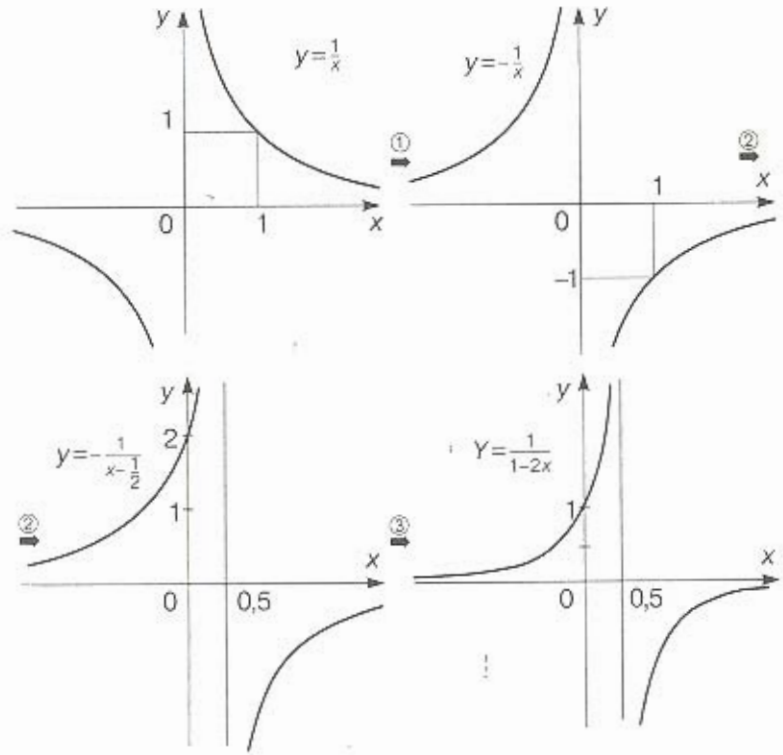

Рис. 3.8

Классификация функций

Простейшими элементарными функциями называются: постоянная функция $y = \text{const}$, степенная функция x^α (α — любое число), показательная функция $y = a^x (0 < a \neq 1)$, логарифмическая функция $\log_a x$ $(0 < a \neq 1)$, тригонометрические функции $\sin x$, $\cos x$, $\operatorname{tg} x$, $\operatorname{ctg} x$ и обратные тригонометрические функции $\arcsin x$, $\arccos x$, $\operatorname{arctg} x$, $\operatorname{arcctg} x$.

Функции, которые можно получить при помощи конечного числа арифметических операций над простейшими элементарными функциями, а также их суперпозицией (наложением), образуют класс *элементарных функций*.

Примеры элементарных функций:

$$f(x) = |x|/(1 - x^3), \quad f(x) = \log_2(\sin x - e^x), \quad f(x) = \operatorname{arctg} 3^x.$$

В свою очередь, элементарные функции классифицируются следующим образом:

1. Функция вида

$$P(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_{n-1}x + a_n,$$

где $n \geqslant 0$ — целое число, $a_0 = 0, a_1, a_2, \ldots, a_n$ — любые числа (коэффициенты), называется *целой рациональной функцией*, или *алгебраическим многочленом степени* n. Многочлен первой степени называется *линейной функцией* (поскольку его график изображается прямой линией).

2. Отношение двух целых рациональных функций

$$R(x) = \frac{a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n}{b_0x^m + b_1x^{m-1} + \cdots + b_{m-1}x + b_m}$$

называется *дробно-рациональной функцией*. Целые и дробно-рациональные функции образуют класс *рациональных функций*.

3. Функция, полученная путем конечного числа суперпозиций и арифметических действий над степенными функциями с целыми и дробными показателями и не являющаяся рациональной, называется *иррациональной функцией*. Примеры иррациональных функций:

$$f(x) = x^2 + \sqrt{x}, \quad f(x) = \frac{x^3 + 4}{x + \sqrt{x - 1}}, \quad f(x) = x^{3/2} + x^{1/3}.$$

4. Всякая функция, не являющаяся рациональной или иррациональной, называется *трансцендентной функцией*.

Примеры трансцендентных функций:

$$f(x) = \cos 2x, \quad f(x) = \operatorname{tg} x + 3x^2, \quad f(x) = e^x + 5x.$$

Приложения функций в экономике

Приведем примеры использования функций из области экономики.

Кривые спроса и предложения. Точка равновесия. Рассмотрим зависимости спроса D (*demand*) и предложения S (*supply*) от цены на товар P (*price*). Чем меньше цена, тем больше спрос при постоянной покупательной способности населения. Обычно зависимость D от P имеет вид ниспадающей кривой (рис. 3.9, а):

$$D = kP^a + c, \tag{3.3}$$

где $a < 0$. В свою очередь, предложение растет с увеличением цены на товар, и потому зависимость S от P имеет следующую характерную форму:

$$S = P^b + d, \tag{3.4}$$

где $b \geqslant 1$ (рис. 3.9, б). В формулах (3.3) и (3.4) c и d — так называемые экзогенные величины; они зависят от ряда других

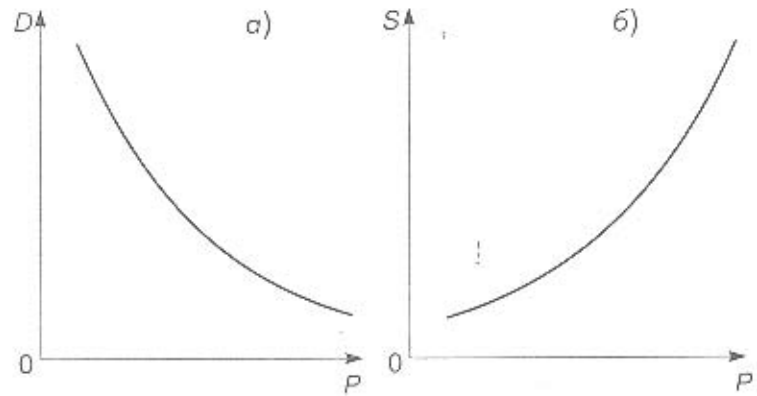

Рис. 3.9

причин (благосостояние общества, политическая обстановка и т.п.). Вполне понятно, что переменные, входящие в формулы (3.3) и (3.4), положительны, поэтому графики функций имеют смысл только в первой координатной четверти.

Для экономики представляет интерес условие равновесия, т.е. когда спрос равен предложению; это условие задается уравнением

$$D(P) = S(P) \tag{3.5}$$

и соответствует точке пересечения кривых D и S — это так называемая точка равновесия (точка M на рис. 3.10). Цена P_0, при которой выполняется условие (3.5), называется *равновесной*.

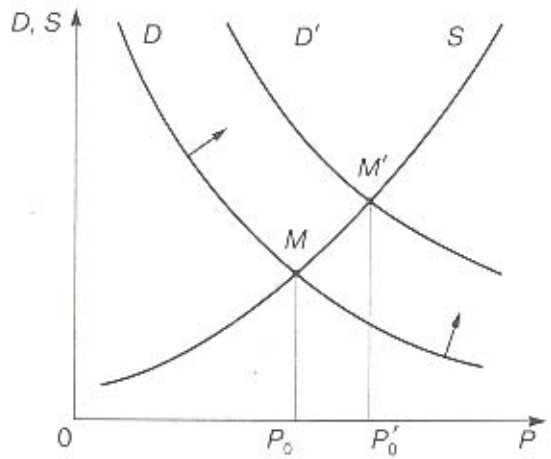

Рис. 3.10

При увеличении благосостояния населения, что соответствует росту величины c в формуле (3.3), точка равновесия M смещается вправо, так как кривая D поднимается вверх; при этом цена на товар растет при неизменной кривой предложения S.

Паутинная модель рынка. Рассмотрим простейшую задачу поиска равновесной цены. Это одна из основных проблем рынка, означающая фактически торг между производителем и покупателем.

Пусть сначала цену P_1 называет производитель (в простейшей схеме он же и продавец). Цена P_1 на самом деле выше равновесной (естественно, всякий производитель стремится получить максимум выгоды из своего производства). Покупатель оценивает спрос D_1 при этой цене, находит его низким и называет свою цену P_2, при которой спрос D_1 равен предложению. Цена P_2 ниже равновесной (всякий покупатель стремится купить подешевле). В свою очередь производитель оценивает спрос D_2, соответствующий цене P_2, и определяет свою цену P_3, при которой спрос равен предложению, эта цена выше равновесной (рис. 3.11, а). Процесс торга в итоге приводит к при-

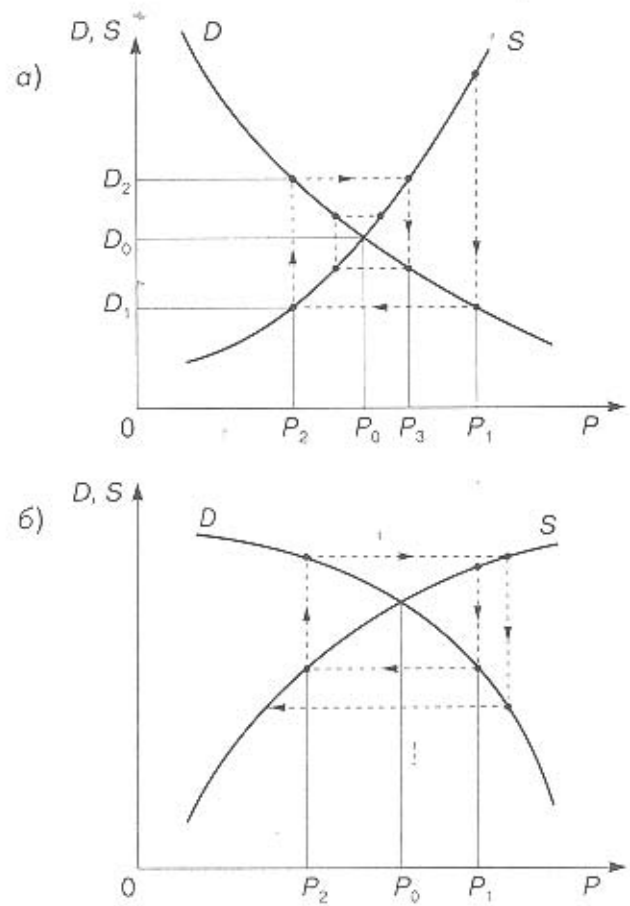

Рис. 3.11

ближению к равновесной цене, т.е. к "скручиванию" спирали. Если рассматривать последовательность чисел, состоящую из называемых в процессе торга цен, то она имеет своим пределом равновесную цену P_0.

Однако спираль установления равновесной цены не всегда "скручивается" к точке P_0. Если экономика "больна", то кривые спроса и предложения могут иметь другой вид, нежели кривые, описываемые формулами (3.3) и (3.4). Например, пусть предложение явно недостаточно и в формуле (3.4) $b < 1$, т.е.

$$S(P) = (P)^{1/m} + d, \quad m > 1,$$

а покупательная способность населения чрезвычайно быстро падает с ростом цены и в формуле (3.3) $a > 1$ при $k < 0$. В этом случае равновесная цена P_0 неустойчива в том смысле, что процесс торга "раскручивает" спираль цен и уводит от P_0: покупателю не удается сдерживать цену и производитель одерживает верх в этой гонке (рис. 3.11, б). Такое возможно, если, скажем, производитель является монополистом. Тогда равновесную цену можно удержать лишь внерыночными мерами (например, государственным вмешательством).

3.2. Предел функции

Предел функции в точке

Пусть функция $f(x)$ определена на некотором множестве X. Возьмем из X последовательность точек

$$x_1, x_2, x_3, \ldots, x_n, \ldots, \tag{3.6}$$

сходящуюся к точке a, причем $a \in X$ или $a \notin X$. Соответствующие значения функции в точках этой последовательности также образуют числовую последовательность

$$f(x_1), f(x_2), f(x_3), \ldots, f(x_n), \ldots, \tag{3.7}$$

и правомерно рассмотреть вопрос о ее сходимости. В математике существует два определения предела функции в точке.

Определение 1. Число A называется *пределом функции* $f(x)$ в точке a (или пределом функции при $x \to a$), если для любой

сходящейся к a последовательности (3.6) значений аргумента x, отличных от a, соответствующая последовательность значений функции (3.7) сходится к числу A.

Для обозначения предельного значения функции используется следующая символика: $\lim\limits_{n\to\infty} f(x) = A$.

Заметим, что функция $f(x)$ может иметь в точке a только одно предельное значение, поскольку последовательность $\{f(x_n)\}$ имеет только один предел.

Приведем несколько примеров.

1. Функция $f(x) - C = \text{const}$ имеет предел в каждой точке числовой прямой. Действительно, любой последовательности (3.6), сходящейся к точке a, соответствует последовательность (3.7), состоящая из одного и того же числа C, откуда следует, что $f(x_n) \to C$ при $n \to \infty$.

2. Функция $f(x) = x$ в любой точке a числовой прямой имеет предел, равный a. Действительно, последовательности значений аргумента (3.6) и функции (3.7) в этом случае тождественны, и если последовательность $\{x_n\}$ сходится к a, то и последовательность $\{f(x_n)\}$ также сходится к a.

3. Функция $f(x) = \dfrac{x^2 - 3x + 2}{2x - 1}$ имеет в точке $x = 0$ предел, равный -2. Действительно, пусть $\{x_n\}$ — любая последовательность значений аргумента, сходящаяся к нулю, т.е. $\lim\limits_{n\to\infty} x_n = 0$, тогда при $x_n \neq 0$ в силу теорем 2.8–2.10

$$\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} \frac{x_n^2 - 3x_n + 2}{2x_n - 1} =$$
$$= \frac{(\lim\limits_{n\to\infty} x_n)^2 - 3(\lim\limits_{n\to\infty} x_n) + 2}{2(\lim\limits_{n\to\infty} x_n) - 1} = -2.$$

4. Функция Дирихле, определенная в параграфе 3.1, не имеет предела ни в одной точке a числовой прямой. Действительно, для сходящейся к точке a последовательности рациональных значений аргумента предел соответствующей последовательности значений функции равен единице, а для сходящейся к точке a последовательности иррациональных значений аргумента соответствующая последовательность значений функции сходится к нулю.

5. Функция $f(x) = \sin(1/x)$ определена для всех $x \neq 0$. В точке $x = 0$ эта функция не имеет предела. Для доказательства возьмем две последовательности значений аргумента, сходящиеся к нулю: $1/\pi, 1/(2\pi), 1/(3\pi), \ldots, 1/(n\pi), \ldots$ и $2/\pi, 2/(5\pi), 2/(9\pi), \ldots, 2/[(4n-3)\pi], \ldots$. Соответствующие последовательности значений функций для них:

$$\{f(x_n)\} = \left\{ f\left(\frac{1}{n\pi}\right) \right\} = \sin n\pi = 0,$$

$$\{f(x_n)\} = \left\{ f\left(\frac{2}{(4n-3)\pi}\right) \right\} = \sin\frac{(4n-3)\pi}{2} = 1.$$

Таким образом, определение 1 не удовлетворяется, так как дли двух разных последовательностей значений аргумента, сходящихся к нулю, соответствующие последовательности значений функции имеют разные пределы.

Определение 2. Число A называется *пределом функции* $f(x)$ в точке a, если для любого числа $\varepsilon > 0$ существует такое число $\delta > 0$, что для всех $x \in X$, $x \neq a$, удовлетворяющих неравенству $|x - a| < \delta$, выполняется неравенство $|f(x) - A| < \varepsilon$.

Второе определение предела функции означает, что функция $f(x)$ имеет предел A в точке a, если для любой ε-окрестности точки A можно найти такую δ-окрестность точки a, что, как только значение аргумента попадет в эту δ-окрестность, соответствующее значение функции $f(x)$ будет находиться в ε-окрестности точки A. При этом по любому $\varepsilon > 0$ число $\delta > 0$ определяется по функции f.

Первое определение предела функции основано на понятии предела числовой последовательности, и его называют определением "на языке последовательностей" (предел функции по Гейне[1]). Второе определение носит название "на языке ε–δ" (предел функции по Коши[2]).

Теорема 3.1. *Первое и второе определения предела функции эквивалентны.*

Доказательство. 1. Пусть A — предел функции $f(x)$ в точке a по определению 1. Предположим, что число A не является пределом этой функции согласно определению 2. Следовательно,

[1] Генрих Эдуард Гейне (1821–1881) — немецкий математик.

[2] Огюстен Луи Коши (1789–1857) — французский математик.

существует такое $\varepsilon_0 > 0$, что из любой δ-окрестности точки a найдется хотя бы одна такая точка $x \neq a$, что $|f(x) - A| > \varepsilon$. В качестве δ будем выбирать последовательно числа:

$$1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots, \qquad \text{т.е. } \{\delta_n\} = \{1/n\}.$$

В каждой из этих δ-окрестностей, согласно нашему предположению, существует хотя бы по одному такому числу $x_n \neq a$, что $|f(x_n) - A| > \varepsilon$. В результате последовательность точек, не совпадающих с a, сходится к пределу a, так как при всех n выполняется неравенство $|x_n - a| < 1/n$. Но тогда, в силу предположения теоремы, соответствующая последовательность значений функции $\{f(x_n)\}$ сходится к числу A; иными словами, для $\varepsilon_0 > 0$ найдется такой номер N, что при всех $n > N$ будет выполняться неравенство $|f(x_n) - A| < \varepsilon_0$. Но это противоречит нашему предположению, по которому для всех x_n выполняется противоположное неравенство. Это противоречие доказывает, что число A является пределом функции $f(x)$ в точке a согласно второму определению.

2. Пусть теперь A — предел функции $f(x)$ согласно определению 2, т.е. для любого $\varepsilon > 0$ существует такое $\delta > 0$, что из неравенства $0 < |x - a| < \delta$ следует неравенство $|f(x) - A| < \varepsilon$. Докажем, что тогда A — предел функции $f(x)$ и согласно первому определению. Для этого возьмем любую последовательность значений аргумента $\{x_n\}$, сходящуюся к точке a ($x_n \neq a$). Тогда для данного значения $\delta > 0$, выбранного по $\varepsilon > 0$, найдется такое N, что при $n > N$ выполняются неравенства $0 < |x_n - a| < \delta$. Но, в силу второго определения предела функции, будет выполнено и неравенство $|f(x_n) - A| < \varepsilon$. Поскольку число ε было выбрано произвольно, то это означает, что последовательность $\{f(x_n)\}$ сходится к числу A для любой последовательности $\{x_n\}$, сходящейся к точке a согласно первому определению предела функции.

Теорема 3.1 полностью доказана. ∎

Левый и правый пределы функции

Введем понятия односторонних пределов функции.

Определение 3. Число a называется *правым (левым) пределом* функции в точке a, если для любой сходящейся к a последо-

вательности (3.6) такой, что $x_n > a$ ($x_n < a$), соответствующая последовательность (3.7) сходится к A.

Для правого (левого) предела функции используется символическая запись:

$$\lim_{x\to a+} f(x) = A \qquad \left(\lim_{x\to a-} f(x) = A\right).$$

Можно дать определение односторонних пределов функции "на языке ε–δ":

Число A называется *правым (левым) пределом* функции в точке a, если для любого $\varepsilon > 0$ существует такое $\delta > 0$, что для всех x из правой (левой) δ-окрестности точки a, т.е. $a < x < a+\delta$ ($a-\delta < x < a$), выполняется неравенство $|f(x)-A| < \varepsilon$.

Приведем в качестве примера функцию $f(x) = \operatorname{sign} x$ (см. параграф 3.1, пример 3). В точке $x = 0$ эта функция имеет левый и правый пределы: $\lim\limits_{x\to 0-} \operatorname{sign} x = -1$, $\lim\limits_{x\to 0+} \operatorname{sign} x = 1$. Действительно, для любой сходящейся к нулю последовательности $\{x_n\}$, у которой все элементы $x_n < 0$, соответствующая последовательность значений функции состоит только из одного числа -1, т.е. предел слева в точке $x = 0$ также равен этому числу. Аналогично устанавливается и предел справа.

Теорема 3.2. *Функция $f(x)$ имеет в точке a предел тогда и только тогда, если в этой точке существуют левый и правый пределы, причем они равны. В таком случае предел функции равен односторонним пределам.*

Доказательство. Пусть существует предел функции в точке a, т.е. $\lim\limits_{x\to a} f(x) = A$. Тогда, по второму определению предела функции в точке, для любого $\varepsilon > 0$ существует такое $\delta > 0$, что для всех x, удовлетворяющих неравенствам $|x-a| < \delta$, $x \neq a$, выполняется неравенство $|f(x) - A| < \varepsilon$. Стало быть, при $a - \delta < x < a$ и при $a < x < a + \delta$ выполняется равенство $|f(x) - A| < \varepsilon$; согласно определению 3, это и означает, что

$$\lim_{x\to a-} f(x) = \lim_{x\to a+} f(x) = A. \tag{3.8}$$

И наоборот, пусть выполняется условие (3.8). Тогда для любого $\varepsilon > 0$ существуют такие числа δ_1, и δ_2, что для всех x из левой δ_1-окрестности точки a ($a - \delta_1 < x < a$) и для всех x из правой δ_2-окрестности точки a ($a < x < a + \delta_2$) выполняется неравенство $|f(x) - A| < \varepsilon$. Возьмем $\delta = \min(\delta_1, \delta_2)$. Тогда для всех x, удовлетворяющих неравенствам $|x-a| < \delta$, $x \neq a$, будет выполняться неравенство $|f(x) - A| < \varepsilon$; это, согласно определению 2, и означает, что число A является пределом функции $f(x)$ в точке a.

Предел функции при $x \to \infty$, $x \to -\infty$, $x \to +\infty$

Кроме понятия предела функции в точке, существует также и понятие предела функции при стремлении аргумента к бесконечности.

Определение 4. Число A называется *пределом функции $f(x)$ при $x \to \infty$*, если для любой бесконечно большой последовательности значений аргумента (3.6) соответствующая последовательность значений функции (3.7) сходится к A.

Для обозначения предела функции при $x \to \infty$ используется следующая символика: $\lim\limits_{x\to\infty} f(x) = A$.

Определение 5. Число A называется *пределом функции $f(x)$ при $x \to -\infty$ (при $x \to +\infty$)*, если для любой бесконечно большой последовательности значений аргумента, элементы которой отрицательны (положительны), соответствующая последовательность значений функции сходится к A.

Определения 4 и 5 даны по Гейне ("на языке последовательностей"). Можно дать равносильные определения и по Коши ("на языке ε–δ"). Например:

Число A называется *пределом функции $f(x)$* при $x \to \infty$, если для любого числа $\varepsilon > 0$ существует такое число δ, что при всех $x > \delta$ справедливо неравенство $|f(x) - a| < \varepsilon$.

Приведем пример предела функции при $x \to \infty$. Пусть $f(x) = 1/x$. Эта функция имеет предел при $x \to \infty$, равный нулю. Действительно, если последовательность (3.6) — бесконечно большая последовательность значений аргумента, то соответствующая последовательность значений функции (3.7) имеет вид: $1/x_1, 1/x_2, \ldots, 1/x_n, \ldots$ и по теореме 2.1 она является бесконечно малой, т.е. ее предел равен нулю, или $\lim\limits_{x\to\infty} f(x) = 0$.

3.3. Теоремы о пределах функций

Покажем, что арифметические операции над функциями, имеющими предел в точке a, приводят к функциям, также имеющим предел в этой точке.

Теорема 3.3. *Пусть функции $f(x)$ и $g(x)$ имеют в точке a пределы соответственно A и B. Тогда функции $f(x) \pm g(x)$, $f(x)g(x)$ и $f(x)/g(x)$ (при $B \neq 0$) имеют в точке a пределы, равные соответственно $A \pm B$, AB и A/B.*

Доказательство. Воспользуемся определением предела функции по Гейне. Возьмем произвольную последовательность $\{x_n\}$, сходящуюся к a ($x_n \neq a$). Соответствующие последовательности значений функций $\{f(x_n)\}$ и $\{g(x_n)\}$, по условию теоремы, имеют в этой точке пределы A и B. В силу теорем 2.8–2.10 числовые последовательности $\{f(x_n) \pm g(x_n)\}$, $\{f(x_n)g(x_n)\}$ и $\{f(x_n)/g(x_n)\}$ имеют пределы, соответственно равные $A \pm B$, AB и A/B (при $g(x) \neq 0$ и $B \neq 0$). Согласно определению предела функции по Гейне, это и означает, что

$$\lim_{x\to a}[f(x) \pm g(x)] = A \pm B,$$

$$\lim_{x\to a}[f(x)g(x)] = AB,$$

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \frac{A}{B},$$

что и требовалось доказать.

Теорема 3.4. *Пусть функции $f(x)$, $g(x)$ и $h(x)$ определены в некоторой окрестности точки a за исключением, быть может, самой точки a, и функции $f(x)$ и $g(x)$ имеют в этой точке предел, равный A: $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = A$. Кроме того, пусть выполняются неравенства $f(x) \leqslant h(x) \leqslant g(x)$. Тогда*

$$\lim_{x\to a} h(x) = A.$$

Доказательство этой теоремы проводится на основе теоремы 2.12 с определением предела функции в точке по Гейне; здесь мы его не приводим.

Заметим, что теоремы 3.3 и 3.4 справедливы и в случаях, когда a является ∞, $+\infty$ или $-\infty$.

3.4. Два замечательных предела

Первый замечательный предел

Теорема 3.5. *Предел функции $\dfrac{\sin x}{x}$ в точке $x = 0$ существует и равен единице, т.е.*

$$\lim_{x\to 0}\frac{\sin x}{x} = 1. \tag{3.9}$$

Доказательство. Рассмотрим окружность радиуса 1. Пусть x — угол, отсчитываемый от единичного радиуса OA против часовой стрелки (рис. 3.12). Тогда при $0 \leqslant x < \pi/2$ имеем: $\sin x = MB$, $\operatorname{tg} x = NA$. Треугольник OMA содержится в секторе OMA, а этот сектор, в свою очередь, содержится в треугольнике ONA. Следовательно, для площадей этих фигур справедливы неравенства

$$OA \cdot MB/2 < OA \cdot \breve{MA}/2 < OA \cdot NA/2.$$

Отсюда, с учетом сказанного выше,

$$\sin x < x < \operatorname{tg} x.$$

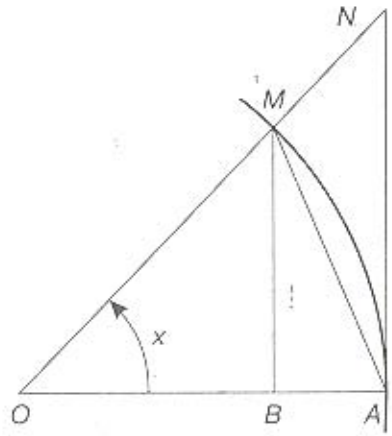

Рис. 3.12

Разделив эти неравенства на $\sin x$ ($\sin x > 0$), получаем $1 < x/\sin x < 1/\cos x$ или

$$\cos x < \frac{\sin x}{x} < 1. \qquad (3.10)$$

Нетрудно видеть, что последние неравенства справедливы и при $-\pi/2 < x < 0$, поскольку в них содержатся четные функции. Из неравенств (3.10) получаем $0 < 1 - \sin x/x < 1 - \cos x$. Так как $1 - \cos x = 2\sin^2(x/2) < 2|\sin(x/2)| < 2|x/2| = |x|$; поскольку $0 < x < \pi/2$, то $|x| = x$, т.е. окончательно получаем неравенства

$$0 < 1 - \frac{\sin x}{x} < x. \qquad (3.11)$$

Для любой последовательности $\{x_n\}$, сходящейся к нулю, пределы левой и правой частей неравенства (3.11) равны нулю. Следовательно, $\lim\limits_{x\to 0}\left(1 - \dfrac{\sin x}{x}\right) = 0$, откуда, с учетом теоремы 3.3, и следует утверждение теоремы 3.5. Случай при $x < 0$ доказывается аналогично.

Предел (3.9) называется *первым замечательным пределом*. Он применяется при вычислении ряда других пределов.

Рассмотрим несколько примеров на применение предела (3.9).

Пример 1. Найти предел функции $\sin(ax)/(bx)$ при $x \to 0$.

Преобразуем данную дробь так, чтобы в знаменателе был аргумент синуса; только тогда можно будет применить первый замечательный предел, поскольку при $x \to 0$ пределом ax также является нуль:

$$\lim_{x\to 0}\frac{\sin(ax)}{bx} = \lim_{x\to 0}\frac{\sin(ax)}{ax}\frac{a}{b} = \lim_{x\to 0}\frac{\sin(ax)}{ax}\lim_{x\to 0}\frac{a}{b} = 1\frac{a}{b} = \frac{a}{b}.$$

Пример 2. Найти $\lim\limits_{x\to 0}\dfrac{1-\cos x}{x^2}$.

Теорему 3.5 здесь непосредственно применить нельзя, так как при $x \to 0$ знаменатель дроби стремится к нулю. Для решения задачи необходимо преобразовать данную дробь:

$$\lim_{x\to 0}\frac{1-\cos x}{x^2} = \lim_{x\to 0}\frac{2\sin^2(x/2)}{x^2} = \lim_{x\to 0}\frac{1}{2}\left(\frac{\sin(x/2)}{x/2}\right)^2 =$$

$$= \frac{1}{2}\lim_{x\to 0}\frac{\sin(x/2)}{x/2}\lim_{x\to 0}\frac{\sin(x/2)}{x/2} = \frac{1}{2}\cdot 1\cdot 1 = \frac{1}{2}.$$

Пример 3. Найти $\lim\limits_{x\to 0}\dfrac{\cos x - \cos 3x}{x^2}$.

Как и в первых двух примерах, преобразуем данную дробь, чтобы "подогнать" ее под первый замечательный предел:

$$\lim_{x\to 0}\frac{\cos x - \cos 3x}{x^2} = \lim_{x\to 0}\frac{2\sin 2x\sin x}{x^2} = 2\cdot 2\cdot\lim_{x\to 0}\frac{\sin 2x}{2x}\times$$

$$\times\lim_{x\to 0}\frac{\sin x}{x} = 4\cdot 1\cdot 1 = 4.$$

Второй замечательный предел

Теорема 3.6. *Предел функции* $f(x) = \left(1 + \dfrac{1}{x}\right)^x$ *при* $x \to \infty$ *существует и равен* e.

Доказательство. Как известно, $\lim\limits_{n\to\infty}\left(1 + \dfrac{1}{n}\right)^n$ равен e (см. параграф 2.3). Сначала докажем теорему для $x > 1$. Возьмем $n = [x]$ — целая часть от числа x, тогда $x = n + \alpha$, $0 \leqslant \alpha < 1$. Поскольку $n \leqslant x < n + 1$, то $1/(n+1) < 1/x < 1/n$ и тогда: $[1 + 1/(n+1)]^n < (1 + 1/x)^x < (1 + 1/n)^{n-1}$. Мы получили ограничения на функцию сверху и снизу; для применения теоремы 3.4 нужно вычислить пределы этих ограничений. Так как при $x \to \infty$ целая переменная $n \to \infty$, то предел ограничения справа

$$\lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^{n+1} = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n\lim_{n\to\infty}\left(1 + \frac{1}{n}\right) = e\cdot 1 = e.$$

Аналогично вычисляется и предел ограничения слева:

$$\lim_{n\to\infty}\left(1 + \frac{1}{n+1}\right)^n = \frac{\lim\limits_{n\to\infty}[1 + 1/(n+1)]^{n+1}}{\lim\limits_{n\to\infty}[1 + 1/(n+1)]} = \frac{e}{1} = e.$$

Отсюда по теореме 3.4 получаем утверждение теоремы при $x \to +\infty$:

$$\lim_{x\to+\infty}(1+1/x)^x = e.$$

Для случая $x \to -\infty$ положим $x = -y$. Получаем:

$$\lim_{x\to-\infty}\left(1+\frac{1}{x}\right)^x = \lim_{y\to+\infty}\left(1-\frac{1}{y}\right)^{-y} = \lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right)^{y} =$$

$$= \lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right)^{y-1} \lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right) = e\cdot 1 = e.$$

Объединяя оба случая, получаем окончательно утверждение теоремы:

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = e. \qquad (3.12)$$

Число e — одна из фундаментальных величин в математике. Показательная функция типа e^{ax} называется *экспонентой*, логарифм с основанием e называется *натуральным* и обозначается символом ln, в теории вероятностей и статистике функция e^{-x^2} является основополагающей.

Второй замечательный предел (3.12) широко применяется для вычисления других пределов. Рассмотрим примеры его использования.

Пример 4. Найти $\lim\limits_{x\to0}(1+x)^{1/x}$.

Применим здесь замену переменной, полагая $1/x = y$. Тогда $y \to \infty$ при $x \to 0$, т.е.

$$\lim_{x\to0}(1+x)^{1/x} = \lim_{y\to\infty}(1+1/y)^y = e.$$

Пример 5. Найти $\lim\limits_{x\to\infty}(1+2/x)^x$.

Заменим переменную, положив $x = 2y$. При $x \to \infty$ и $y \to \infty$, откуда

$$\lim_{x\to\infty}(1+2/x)^x = \lim_{y\to\infty}(1+1/y)^{2y} = \lim_{y\to\infty}[(1+1/y)^y]^2 =$$

$$= \lim_{y\to\infty}(1+1/y)^y \lim_{y\to\infty}(1+1/y)^y = e\cdot e = e^2.$$

Пример 6. Найти $\lim\limits_{x\to0}\left[\frac{1}{x}\log_a(1+x)\right]$.

Преобразуем дробь под знаком предела:

$$\lim_{x\to0}\left[\frac{1}{x}\log_a(1+x)\right] = \lim_{x\to0}\log_a(1+x)^{1/x} =$$

$$= \log_a\left[\lim_{x\to0}(1+x)^{1/x}\right] = \log_a e = \frac{1}{\ln a} \qquad \text{см. пример 4.}$$

3.5. Бесконечно малые и бесконечно большие функции

Бесконечно малые функции

Определение 1. Функция $f(x)$ называется *бесконечно малой функцией* (или просто *бесконечно малой*) в точке $x = a$, если предел ее в этой точке равен нулю: $\lim\limits_{x\to a} f(x) = 0$.

Аналогично определяются бесконечно малые при $x \to \infty$, $x \to \pm\infty$, $x \to a+$ и $x \to a-$.

Теорема 3.7. *Если функция $f(x)$ имеет предел A в точке $x = a$, то функция $\alpha(x) = f(x) - A$ является бесконечно малой в точке a.*

Доказательство. Действительно, из теоремы 3.3 вытекает, что

$$\lim_{x\to a}\alpha(x) = \lim_{x\to a} f(x) - \lim_{x\to a} A = A - A = 0,$$

откуда, согласно определению 1, и следует, что $\alpha(x)$ — бесконечно малая функция в точке a.

Мы получаем специальное представление для функций, имеющих предел в точке $x = a$, через бесконечно малую функцию:

$$f(x) = A + \alpha(x). \qquad (3.13)$$

Это представление оказывается весьма удобным при доказательствах и будет использоваться в дальнейшем.

Теорема 3.8. *Алгебраическая сумма и произведение конечного числа бесконечно малых функций в точке a, как и произведение бесконечно малой на ограниченную функцию, являются бесконечно малыми функциями в точке a.*

Доказательство этой теоремы непосредственно вытекает из первого определения предела функции и теорем 2.2–2.4.

Бесконечно большие функции

Определение 2. Функция $f(x)$ называется *бесконечно большой функцией* в точке a (или просто *бесконечно большой*), если для любой сходящейся к a последовательности $\{x_n\}$ значений аргумента соответствующая последовательность $\{f(x_n)\}$ значений функции является бесконечно большой последовательностью.

В этом случае пишут $\lim\limits_{x\to a} f(x) = \infty$ $\left(\lim\limits_{x\to a} f(x) = +\infty \text{ или } \lim\limits_{x\to a} f(x) = -\infty\right)$ и говорят, что функция имеет в точке a бесконечный предел ($+\infty$ или $-\infty$). По аналогии с конечными односторонними пределами определены и односторонние бесконечные пределы:

$$\lim_{x\to a-} f(x) = +\infty, \quad \lim_{x\to a-} f(x) = -\infty,$$
$$\lim_{x\to a+} f(x) = +\infty, \quad \lim_{x\to a+} f(x) = -\infty.$$

Аналогично определяются бесконечно большие функции при $x \to \infty$, $x \to +\infty$ и $x \to -\infty$.

Между бесконечно малыми и бесконечно большими функциями существует та же связь, что между соответствующими последовательностями, т.е. если $\alpha(x)$ — бесконечно малая функция при $x \to a$, то $f(x) = 1/\alpha(x)$ — бесконечно большая функция, и наоборот. Это утверждение можно доказать, например, используя первое определение предела функции в точке и соответствующие теоремы о бесконечно малых и бесконечно больших последовательностях.

3.6. Сравнение бесконечно малых и бесконечно больших функций

Сравнение бесконечно малых функций

Как было показано ранее, сумма, разность и произведение бесконечно малых функций являются бесконечно малыми, чего

нельзя сказать о частном: деление одной бесконечно малой на другую может дать различные результаты.

Например, если $\alpha(x) = 2x$, $\beta(x) = 3x$, то

$$\lim_{x\to 0} \frac{\alpha(x)}{\beta(x)} = \lim_{x\to 0} \frac{2x}{3x} = \frac{2}{3}.$$

Если же $\alpha(x) = x^2$, $\beta(x) = x^3$, то

$$\lim_{x\to 0} \frac{\alpha(x)}{\beta(x)} = \lim_{x\to 0} \frac{x^2}{x^3} = \lim_{x\to 0} \frac{1}{x} = \infty, \quad \lim_{x\to 0} \frac{\beta(x)}{\alpha(x)} = \lim_{x\to 0} x = 0.$$

Целесообразно ввести *правила сравнения бесконечно малых функций* с использованием соответствующей терминологии.

Пусть при $x \to a$ функции $\alpha(x)$ и $\beta(x)$ бесконечно малые. Тогда различают следующие варианты их сравнения в зависимости от величины c предела их отношения в точке a:

$$c = \lim_{x\to a} \frac{\alpha(x)}{\beta(x)}. \tag{3.14}$$

1°. *Если $c = 1$, то $\alpha(x)$ и $\beta(x)$ — эквивалентные бесконечно малые: $\alpha(x) \sim \beta(x)$.*

2°. *Если $c = 0$, то $\alpha(x)$ — бесконечно малая более высокого порядка, чем $\beta(x)$ (или имеет более высокий порядок малости).*

3°. *Если $c = d \neq 0$ (d — число), то $\alpha(x)$ и $\beta(x)$ — бесконечно малые одного порядка.*

Часто бывает недостаточно знать, что одна бесконечно малая по отношению к другой является бесконечно малой более высокого порядка малости, нужно еще и оценить величину этого порядка. Поэтому используется следующее правило:

4°. *Если $\lim\limits_{x\to a} \dfrac{\alpha(x)}{\beta^n(x)} = d \neq 0$, то $\alpha(x)$ — бесконечно малая n-го порядка относительно $\beta(x)$. В этом случае используют символ o ("о малое"): $\alpha(x) = o(\beta(x))$.*

Заметим, что верны аналогичные правила сравнения бесконечно малых функций при $x \to \infty$, $x \to -\infty$, $x \to +\infty$, а также в случае односторонних пределов при $x \to a$ слева и справа.

Из правил сравнения вытекает одно важное свойство:

Если $\alpha(x) \sim \alpha_1(x)$ *и* $\beta(x) \sim \beta_1(x)$ *при* $x \to a$ *и существует предел* $\lim\limits_{x\to a} \dfrac{\alpha(x)}{\beta(x)}$, *то существует и предел* $\lim\limits_{x\to a} \dfrac{\alpha_1(x)}{\beta_1(x)}$, *причем оба эти предела равны.*

Действительно,

$$\lim_{x\to a} \frac{\alpha_1(x)}{\beta_1(x)} = \lim_{x\to a} \frac{\alpha_1(x)}{\alpha(x)} \frac{\alpha(x)}{\beta(x)} \frac{\beta(x)}{\beta_1(x)} =$$

$$= \lim_{x\to a} \frac{\alpha_1(x)}{\alpha(x)} \lim_{x\to a} \frac{\alpha(x)}{\beta(x)} \lim_{x\to a} \frac{\beta(x)}{\beta_1(x)} = 1 \cdot \lim_{x\to a} \frac{\alpha(x)}{\beta(x)} \cdot 1 = \lim_{x\to a} \frac{\alpha(x)}{\beta(x)}.$$

В ряде случаев применение доказанного свойства упрощает вычисление пределов и проведение оценок.

Рассмотрим несколько примеров.

Пример 1. Функции $\sin x$ и x при $x \to 0$ являются эквивалентными бесконечно малыми в силу предела (3.9), т.е. $\sin x \sim x$ при $x \to 0$.

Пример 2. Функции $\sin kx$ и $\sin x$ являются при $x \to 0$ бесконечно малыми одного порядка, поскольку

$$\lim_{x\to 0} \frac{\sin kx}{\sin x} = \lim_{x\to 0} \frac{kx}{x} = k.$$

Пример 3. Функция $\alpha(x) = \cos ax - \cos bx$ $(a \neq b)$ является при $x \to 0$ бесконечно малой второго порядка малости по отношению к бесконечно малой x, так как

$$\lim_{x\to 0} \frac{\cos ax - \cos bx}{x^2} = \lim_{x\to 0} \frac{2\sin[(a+b)x/2]\sin[(b-a)x/2]}{x^2} =$$

$$= \lim_{x\to 0} \frac{2[(a+b)x/2][(b-a)x/2]}{x^2} = \frac{b^2-a^2}{2}.$$

Пример 4. $\lim\limits_{x\to 0} \dfrac{\sin kx}{x+x^2}$.

Так как $\sin kx \sim kx$ и $x + x^2 \sim x$, то

$$\lim_{x\to 0} \frac{\sin kx}{x+x^2} = \lim_{x\to 0} \frac{kx}{x} = k.$$

Сравнение бесконечно больших функций

Для бесконечно больших функций также имеют место аналогичные правила сравнения с той лишь разницей, что для них вместо термина "порядок малости" употребляется термин "порядок роста".

Поясним сказанное на примерах.

Пример 5. Функции $f(x) = (2+x)/x$ и $g(x) = 2/x$ при $x \to 0$ являются эквивалентными бесконечно большими, поскольку

$$\lim_{x\to 0} \frac{f(x)}{g(x)} = \lim_{x\to 0} (1 + x/2) = 1.$$

Данные функции $f(x)$ и $g(x)$ имеют одинаковый порядок роста.

Пример 6. Сравним порядки роста функций $f(x) = x^2 + 2$ и $g(x) = x^3 + x$ при $x \to \infty$, для чего найдем предел их отношения:

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{x^2+2}{x^3+x} = \lim_{x\to\infty} \frac{1+2/x^2}{x+1/x^2} = \lim_{x\to\infty} \frac{1}{x} = 0.$$

Отсюда следует, что функция $g(x)$ имеет более высокий порядок роста, нежели функция $f(x)$.

Пример 7. Бесконечно большие при $x \to \infty$ функции $f(x) = 3x^3 + x$ и $g(x) = x^3 - 4x^2$ имеют одинаковый порядок роста, так как

$$\lim_{x\to\infty} \frac{3x^3+x}{x^3-4x^2} = \lim_{x\to\infty} \frac{3+1/x^2}{1-4/x} = 3.$$

Пример 8. Функция $f(x) = x^3 + 2x + 3$ является при $x \to \infty$ бесконечно большой третьего порядка роста по отношению к бесконечно большой функции $g(x) = x - 1$, поскольку

$$\lim_{x\to\infty} \frac{x^3+2x+3}{(x-1)^3} = \lim_{x\to\infty} \frac{x^3+2x+3}{x^3-3x^2+3x-1} =$$

$$= \lim_{x\to\infty} \frac{1+2/x^2=3/x^3}{1-3/x+3/x^2-1/x^3} = 1.$$

3.7. Понятие непрерывности функции

Определение непрерывности функции

Понятие непрерывности функции является фундаментальным в математическом анализе. Пусть функция $f(x)$ определена в некоторой окрестности точки a.

Определение 1. Функция $f(x)$ называется *непрерывной* в точке a, если предел этой функции и ее значение в этой точке равны, т.е.

$$\lim_{x\to a} f(x) = f(a). \tag{3.15}$$

Так как $\lim\limits_{x\to a} x = a$, то это равенство можно переписать в следующей форме:

$$\lim_{x\to a} f(x) = f\left(\lim_{x\to a} x\right).$$

Определение непрерывности функции можно сформулировать как "на языке последовательностей", так и "на языке ε–δ" в соответствии с двумя определениями предела функции в точке. Приведем здесь второе из них.

Определение 2. Функция $f(x)$ называется *непрерывной* в точке a, если для любого $\varepsilon > 0$ существует такое $\delta > 0$, что при всех x, удовлетворяющих неравенству $|x-a| < \delta$, выполняется неравенство $|f(x) - f(a)| < \varepsilon$.

Определение 3. Функция $f(x)$ называется *непрерывной справа* (*слева*) в точке a, если правый (левый) предел этой функции в точке a равен значению функции в этой точке.

Символическая запись непрерывности функции справа и соответственно слева:

$$\begin{aligned} &\lim_{x\to a+} f(x) = f(a), \quad \text{или} \quad f(a+0) = f(a), \\ &\lim_{x\to a-} f(x) = f(a), \quad \text{или} \quad f(a-0) = f(a). \end{aligned} \tag{3.16}$$

Если функция $f(x)$ непрерывна в точке a слева и справа, то она непрерывна в этой точке. В самом деле, по теореме 3.2 функция $f(x)$ имеет предел в точке a, который равен ее значению в этой точке, что и означает непрерывность функции при $x = a$.

Для практического использования полезно сформулировать еще одно определение непрерывности функции. Назовем разность $\Delta x = x - a$ *приращением аргумента* в точке a, а разность $\Delta y = f(x) - f(a)$ — *приращением функции* в точке a, обусловленным приращением аргумента Δx. Таким образом,

$$\Delta x = x - a, \quad \Delta y = f(a + \Delta x) - f(a).$$

Поскольку условия $x \to a$ и $x - a \to 0$ равносильны, то равенство (3.15) можно переписать в другой форме:

$$\lim_{\Delta x\to 0} \Delta y = 0. \tag{3.17}$$

Геометрическая интерпретация приращений Δx и Δy показана на рис. 3.13.

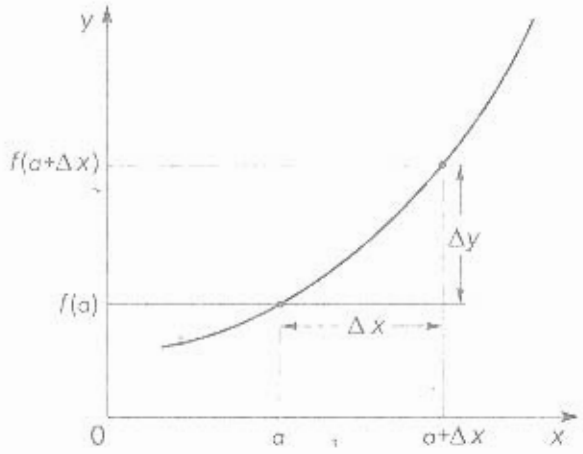

Рис. 3.13

Равенство (3.17) представляет собой еще одно определение непрерывности функции в точке.

Определение 4. Функция $f(x)$ называется *непрерывной* в точке a, если ее приращение в этой точке является бесконечно малой при $\Delta x \to 0$.

Точки, в которых функция не является непрерывной, называются *точками разрыва* функции.

Рассмотрим наиболее яркие примеры точек, в которых функция не является непрерывной.

1. Функция $f(x) = \operatorname{sign} x$ (см. параграф 3.1). Как было показано ранее (см. параграф 3.2), в точке $x = 0$ существуют левый и правый пределы этой функции, равные соответственно -1 и $+1$. Сама же точка $x = 0$ является точкой разрыва функции, поскольку пределы слева и справа не равны значению $f(0) = 0$.

2. Функция Дирихле (см. параграф 3.1). Эта функция не является непрерывной ни в одной точке области своего определения. Действительно, в любой окрестности произвольной точки a содержится бесконечно много рациональных и иррациональных значений аргумента, при которых функция принимает соответственно значения $+1$ и -1. Следовательно, условие непрерывности (3.17) не выполняется.

Арифметические действия над непрерывными функциями

Теорема 3.9. *Пусть функции $f(x)$ и $g(x)$ непрерывны в точке a. Тогда функции $f(x) \pm g(x)$, $f(x)g(x)$ и $f(x)/g(x)$ также непрерывны в точке a (частное при условии $g(a) \neq 0$).*

Доказательство. Так как непрерывные в точке a функции $f(x)$ и $g(x)$ имеют в этой точке пределы, равные соответственно $f(a)$ и $g(a)$, то по теореме 3.3 пределы функций $f(x) \pm g(x)$, $f(x)g(x)$ и $f(x)/g(x)$ также существуют и равны соответственно $f(a) \pm g(a)$, $f(a)g(a)$ и $f(a)/g(a)$. Но эти величины равны значениям соответствующих функций в точке a. Следовательно, в силу определения 1, функции $f(x) \pm g(x)$, $f(x)g(x)$ и $f(x)/g(x)$ непрерывны в точке a, что и требовалось доказать.

3.8. Непрерывность элементарных функций

Непрерывность рациональных функций

Постоянная функция $f(x) = C$ является непрерывной в любой точке числовой прямой. Действительно, $\lim\limits_{x \to a} f(x) = C = f(a)$, что соответствует определению непрерывности функции в точке.

Функция $f(x) = x$ непрерывна в каждой точке a числовой прямой, так как предел функции в точке a равен ее значению в этой точке: $\lim\limits_{x \to a} f(x) = a = f(a)$.

Из сказанного выше и теоремы 3.9 следует, что в любой точке числовой прямой функции $x^2 = x \cdot x$, $x^3 = x^2 \cdot x, \ldots, x^n = x^{n-1}x$ (n — натуральное число) непрерывны.

Алгебраический многочлен

$$P(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_n$$

также является непрерывной функцией в любой точке числовой прямой в силу теоремы 3.9, поскольку представляет собой сумму произведений непрерывных функций.

Дробно-рациональная функция

$$R(x) \frac{P(x)}{Q(x)},$$

где $P(x)$ и $Q(x)$ — алгебраические многочлены, в силу теоремы 3.9 непрерывна во всех точках числовой прямой за исключением корней знаменателя.

Непрерывность тригонометрических функций

Покажем, что функция $\sin x$ непрерывна в любой точке x, для чего воспользуемся определением 4 непрерывности функции в точке. Для произвольного значения аргумента x возьмем приращение Δx, тогда соответствующее приращение функции Δy будет иметь вид

$$\Delta y = \sin(x + \Delta x) - \sin x, \text{ или } \Delta y = 2\cos(x + \Delta x/2)\sin(\Delta x/2).$$

Переходя в обеих частях последнего равенства к пределу при $\Delta x \to 0$, получаем

$$\lim_{\Delta x \to 0} = 2 \lim_{\Delta x \to 0} [\cos(x + \Delta x/2)\sin(\Delta x/2)] = 0,$$

так как $\sin(\Delta x/2)$ — бесконечно малая при $\Delta x \to 0$ и под знаком предела находится произведение бесконечно малой функции на ограниченную функцию.

Непрерывность функции $\cos x$ доказывается аналогичным образом.

Непрерывность функций $\operatorname{tg} x = \sin x / \cos x$ и $\sec x = 1/\cos x$ во всех точках, $x \neq \pi/2 + n\pi$, следует из теоремы 3.9; аналогично непрерывность функций $\operatorname{ctg} x = \cos x / \sin x$ и $\operatorname{cosec} x = 1/\sin x$ обеспечена во всех точках $x \neq n\pi$ ($n = 0, \pm 1, \pm 2, \ldots$).

Непрерывность функции $f(x) = |x|$

График функции $f(x) = |x|$ показан на рис. 3.14. В точках полубесконечного интервала она непрерывна, так как при $x > 0$ функция $f(x) = x$. В точках полупрямой $(-\infty, 0)$ функция также непрерывна, поскольку $f(x) = -x$ при $x < 0$, т.е. ее можно представить в виде произведения непрерывных функций (-1) и x (см. теорему 3.9).

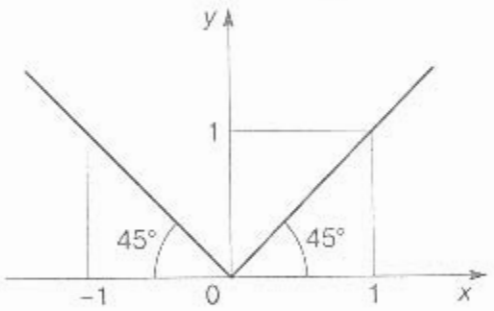

Рис. 3.14

Односторонние пределы функции $|x|$ в точке $x = 0$ равны:

$$\lim_{x\to 0-} |x| = \lim_{x\to 0-} (-x) = -\lim_{x\to 0-} x = 0;$$

$$\lim_{x\to 0+} |x| = \lim_{x\to 0+} x = 0.$$

Пределы функции в точке $x = 0$ слева и справа равны между собой и равны значению функции в этой точке. Следовательно, функция $|x|$ непрерывна в точке $x = 0$, а значит, и во всех точках числовой прямой.

Рассмотренные выше функции непрерывны в каждой точке, в окрестности которой они определены. В силу теоремы 3.9 функции, получаемые из них при использовании конечного числа арифметических операций, также непрерывны.

Говорят, что функция $f(x)$ непрерывна в интервале (a, b), если она непрерывна в каждой точке этого интервала. Функция $f(x)$ непрерывна на отрезке $[a, b]$, если она непрерывна в интервале (a, b) и непрерывна в точке a справа, а в точке b слева

$$\lim_{x\to a+} f(x) = f(a), \quad \lim_{x\to b-} f(x) = f(b).$$

Классификация точек разрыва функции

Точки разрыва, в которых функция не является непрерывной, классифицируются следующим образом.

Устранимый разрыв. Точка a называется *точкой устранимого разрыва* функции $f(x)$, если предел функции в этой точке существует, но в точке a функция $f(x)$ либо не определена, либо ее значение $f(a)$ не равно пределу в этой точке.

Например, функция $f(x) = \dfrac{\sin x}{x}$ в точке $x = 0$, как известно, имеет предел, равный единице (первый замечательный предел). Однако в самой точке $x = 0$ эта функция не определена, т.е. здесь разрыв 1-го рода. Этот разрыв можно устранить (поэтому он и называется устранимым), если доопределить функцию в этой точке значением предела в ней, т.е. ввести новую функцию:

$$f_1(x) = \begin{cases} \dfrac{\sin x}{x}, & x \neq 0, \\ 1, & x = 0. \end{cases}$$

Функция $f_1(x)$ является непрерывной на всей числовой прямой.

Разрыв 1-го рода. Точка a называется *точкой разрыва 1-го рода* функции $f(x)$, если в этой точке функция имеет конечные, но не равные друг другу левый и правый пределы:

$$\lim_{x\to a-} f(x) \neq \lim_{x\to a+} f(x).$$

Типичным примером является функция $f(x) = \operatorname{sign} x$: для нее точка $x = 0$ является точкой разрыва 1-го рода.

Разрыв 2-го рода. Точка a называется *точкой разрыва 2-го рода* функции $f(x)$, если в этой точке функция $f(x)$ не имеет по крайней мере одного из односторонних пределов или хотя бы один из односторонних пределов бесконечен.

Рассмотрим для этого случая несколько примеров.

1. Для функции $f(x) = 1/x$ точка $x = 0$ является точкой разрыва 2-го рода (см. рис. 3.8), поскольку $\lim\limits_{x\to 0+} (1/x) = +\infty$, $\lim\limits_{x\to 0} (1/x) = -\infty$.

2. Для функции $f(x) = \sin(1/x)$ (рис. 3.15) точка $x = 0$ является точкой разрыва 2-го рода, так как ни левого, ни правого предела функции в этой точке не существует (см. пример 5, параграф 3.2).

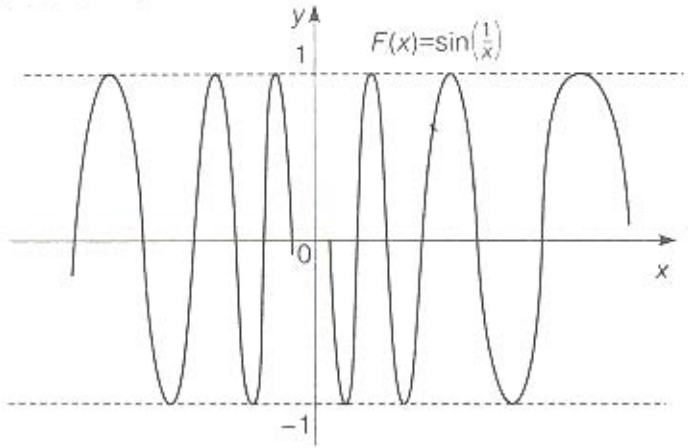

Рис. 3.15

3. Рассмотрим функцию $f(x) = e^{1/x} = \exp\left(\frac{1}{x}\right)$ (рис 3.16). Точка $x = 0$ является точкой разрыва 2-го рода для этой функции, так как предел слева равен нулю, а предел справа бесконечен:

$$\lim_{x\to 0-}\left(e^{1/x}\right) = 0, \qquad \lim_{x\to 0+}\left(e^{1/x}\right) = +\infty.$$

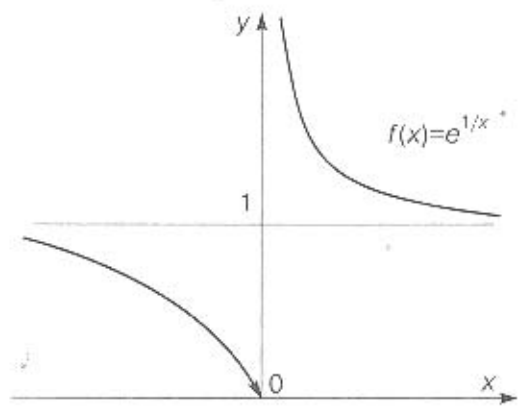

Рис. 3.16

Кусочно-непрерывные функции

Функция $f(x)$ называется *кусочно-непрерывной* на сегменте (отрезке) $[a, b]$, если она непрерывна во всех внутренних точках $[a, b]$, за исключением, быть может, конечного числа точек, в которых имеет разрыв 1-го рода и, кроме того, односторонние пределы в точках a и b.

В качестве примера рассмотрим функцию $f(x) = [x]$ (см. рис. 3.4). Она кусочно-непрерывна как на любом отрезке, так и на всей числовой оси. В точках $x = n (n = 0, \pm 1, \pm 2, \dots)$ функция непрерывна справа и разрывна слева. Во всех других точках она непрерывна.

3.9. Основные свойства непрерывных функций

Устойчивость знака непрерывной функции

Теорема 3.10. *Пусть функция $f(x)$ непрерывна в точке a, причем $f(a) \neq 0$. Тогда существует такая δ-окрестность точки a, что для всех $x \in (a - \delta, a + \delta)$ функция $f(x)$ имеет тот же знак, что и $f(a)$.*

Доказательство. Пусть для определенности $f(a) > 0$ (рис. 3.17). Тогда, согласно второму определению непрерывности функции в точке, для любого $\varepsilon > 0$ можно найти такое $\delta > 0$, что для всех x, удовлетворяющих неравенству $|x - a| < \delta$, вы-

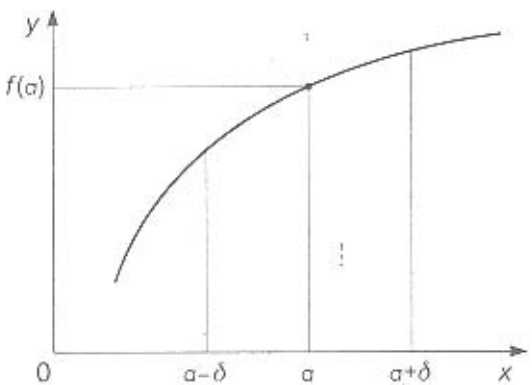

Рис. 3.17

6 Зак. 6857

полняется неравенство $|f(x) - f(a)| < \varepsilon$ или, что то же самое,

$$f(a) - \varepsilon < f(x) < f(a) + \varepsilon. \qquad (3.18)$$

Возьмем $\varepsilon < f(a)$, тогда из левого неравенства в (3.18) получаем, что $f(x) > 0$ для всех x из δ-окрестности точки a, что и требовалось доказать.

В случае $f(a) < 0$ рассмотрим функцию $-f(x)$; так как она положительна в точке a, то для нее существует δ-окрестность точки a, в которой $-f(x) > 0$, а значит, $f(x) < 0$.

Прохождение непрерывной функции через любое промежуточное значение

Теорема 3.11 (ПЕРВАЯ ТЕОРЕМА БОЛЬЦАНО[1]–КОШИ). *Пусть функция $f(x)$ непрерывна на отрезке $[a, b]$ и на его концах принимает значения разных знаков. Тогда внутри $[a, b]$ существует точка c, в которой $f(c) = 0$.*

Доказательство. Пусть для определенности $f(a) < 0$ и $f(b) > 0$ (рис. 3.18). Разделим отрезок $[a, b]$ пополам. Если значение функции в середине отрезка равно нулю, то теорема доказана. В противном случае из двух половинок отрезка возьмем тот сегмент, на концах которого $f(x)$ принимает значения разных знаков; обозначим его $[a_1, b_1]$. Разделим его пополам и повторим описанную выше процедуру, обозначив через $[a_2, b_2]$ отрезок вдвое меньше, на концах которого функция $f(x)$ принимает значения разных знаков. При неограниченном продолжении этого процесса возможны два варианта: либо на каком-то m-м шаге значение функции в середине отрезка $[a_m, b_m]$ равно нулю и тогда теорема доказана, либо мы получим последовательность вложенных отрезков

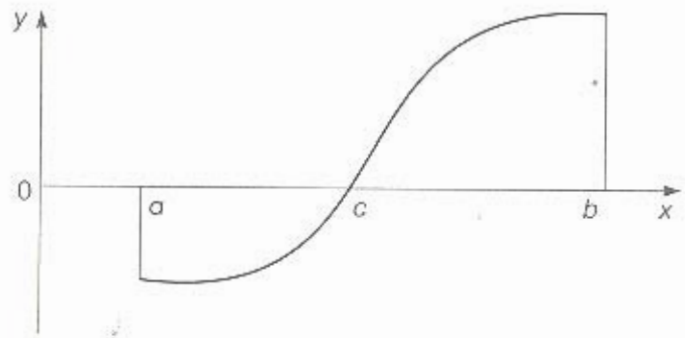

Рис. 3.18

[1] Бернард Больцано (1781–1848) — чешский математик.

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \cdots \supset [a_n, b_n] \ldots, \qquad (3.19)$$

причем $b_n - a_n = (b - a)/2^n \to 0$ при $n \to \infty$. По теореме о вложенных отрезках (см. теорему 2.14) существует точка c, принадлежащая всем отрезкам. Покажем, что $f(c) = 0$. Действительно, если $f(c) > 0$, то по теореме об устойчивости знака непрерывной функции существует некоторая окрестность точки c, в которой $f(x) > 0$. При достаточно большом n в эту окрестность попадет отрезок $[a_n, b_n]$ и, значит, на нем также будет выполняться неравенство $f(x) > 0$. Но это противоречит тому, что на концах всех вложенных отрезков функция $f(x)$ принимает значения разных знаков. Аналогично доказывается, что значение $f(c)$ не может быть отрицательным. Следовательно, $f(c) = 0$, причем точка $c \in (a, b)$, что и требовалось доказать.

Эта теорема имеет простой геометрический смысл: непрерывная кривая при переходе из нижней координатной плоскости в верхнюю обязательно пересекает ось абсцисс (см. рис. 3.18).

Теорема 3.12 (ВТОРАЯ ТЕОРЕМА БОЛЬЦАНО–КОШИ). *Пусть функция $f(x)$ непрерывна на отрезке $[a, b]$, причем $f(a) = A$, а $f(b) = B$. Пусть C — любое число, заключенное между A и B. Тогда на отрезке $[a, b]$ найдется такая точка c, что $f(c) = C$.*

Доказательство. Пусть для определенности $A < C < B$. Рассмотрим вспомогательную функцию

$$\varphi(x) = f(x) - C.$$

Новая функция непрерывна на отрезке $[a, b]$ и принимает на концах этого отрезка значения разных знаков:

$$\varphi(a) = f(a) - C = A - C < 0,$$

$$\varphi(b) = f(b) - C = B - C > 0.$$

По теореме 3.11 существует такая точка $c \in (a, b)$, что $\varphi(c) = f(c) - C = 0$, откуда $f(c) = C$, что и требовалось доказать.

Смысл этой теоремы заключается в утверждении, что непрерывная функция при переходе от одного значения к другому принимает все промежуточные значения.

Пусть функция $f(x)$ определена на множестве X, а Y — множество ее значений. Если множество Y ограничено сверху (снизу), то оно имеет точную верхнюю (нижнюю) грань. Точная верхняя (нижняя) грань множества Y называется *точной верхней (нижней) гранью функции* $f(x)$ на множестве X и обозначается $\sup\limits_X f(x)$ ($\inf\limits_X f(x)$).

Определение точной верхней (нижней) грани функции $f(x)$ можно сформулировать следующим образом.

Определение 1. Число M (m) называется *точной верхней (нижней) гранью функции* $f(x)$ на множестве X, если выполнены два условия:

1) $f(x) \leqslant M (f(x) \geqslant m)$ для любого $x \in X$;

2) для любого числа $M_1 < M$ ($m_1 > m$) найдется по крайней мере одна такая точка $x_1 \in X$, что $f(x_1) > M_1 (f(x_1) < m_1)$.

Если множество Y не ограничено сверху (снизу), то вместо числа M (m) пишут $+\infty$ ($-\infty$).

Ограниченность непрерывной функции на отрезке

Функция $f(x)$ называется *ограниченной на отрезке* $[a, b]$, если существует такое число M, что для всех $x \in [a, b]$ выполняется неравенство $|f(x)| \leqslant M$.

Теорема 3.13 (ПЕРВАЯ ТЕОРЕМА ВЕЙЕРШТРАССА[1]). *Если функция $f(x)$ определена и непрерывна на отрезке $[a, b]$, то она ограничена на этом отрезке.*

Докажем предварительно следующую лемму.

Лемма. *Функция $f(x)$, непрерывная в точке c, ограничена в некоторой ее окрестности.*

[1] Карл Вейерштрасс (1815–1897) — немецкий математик.

Доказательство леммы. Согласно второму определению непрерывности функции в точке, для любого $\varepsilon > 0$ существует такое $\delta > 0$, что для всех $x \in (c - \delta, c + \delta)$ выполняется неравенство $|f(x) - f(c)| < \varepsilon$. Из этого неравенства получаем $|f(x)| = |f(x) - f(c) + f(c)| \leqslant |f(c)| + |f(x) - f(c)| < \varepsilon + |f(c)|$. Отсюда $|f(x)| < M$, где $M = \varepsilon + |f(c)|$, т.е. функция $f(x)$ ограничена в δ-окрестности точки c.

Доказательство теоремы. Докажем теорему от противного, т.е. предположим, что функция $f(x)$ не ограничена на отрезке $[a, b]$. Разделим отрезок $[a, b]$ пополам; по крайней мере на одном из двух новых отрезков функция $f(x)$ не ограничена (иначе она была бы ограничена на $[a, b]$). Обозначим этот отрезок через $[a_1, b_1]$; разделим его пополам, ту его половину, на которой $f(x)$ не ограничена, обозначим через $[a_2, b_2]$ и т.д. Продолжая этот процесс неограниченно, получим последовательность вложенных отрезков (3.19), на каждом из которых функция $f(x)$ не ограничена, причем длина отрезков стремится к нулю: $b_n - a_n = (b - a)/2^n \to 0$ при $n \to \infty$.

По теореме 2.14 о вложенных отрезках существует точка c, принадлежащая всем отрезкам. Функция $f(x)$ определена и непрерывна в точке c, и, согласно доказанной лемме, она ограничена в некоторой ее окрестности. Однако при достаточно большом n в эту окрестность попадет отрезок $[a_n, b_n]$ из последовательности (3.19), на котором $f(x)$ не ограничена. Это противоречие снимает предположение о неограниченности функции $f(x)$ на отрезке $[a, b]$, что и доказывает теорему.

Заметим, что теорема неверна, если в ее формулировке отрезок $[a, b]$ заменить на интервал (a, b). Например, функция $f(x) = 1/x$ на интервале $(0, 1)$ непрерывна, но не ограничена: $\lim\limits_{x \to 0}(1/x) = +\infty$.

Достижение функцией своих точных граней

В том случае, если точные грани функции являются ее значениями, говорят, что функция достигает своих точных граней.

Теорема 3.14 (ВТОРАЯ ТЕОРЕМА ВЕЙЕРШТРАССА). *Если функция $f(x)$ непрерывна на отрезке $[a, b]$, то она достигает*

на этом отрезке своих точных граней, т.е. найдутся такие точки $x_1, x_2 \in [a,b]$, *что*

$$f(x_1) = M = \sup_{[a,b]} f(x), \quad f(x_2) = m = \inf_{[a,b]} f(x).$$

Доказательство. Поскольку функция $f(x)$ непрерывна на отрезке $[a,b]$, то в силу теоремы 3.13 она ограничена на этом отрезке. Но тогда, согласно теореме 1.1, существуют точная верхняя и точная нижняя грани функции $f(x)$ на отрезке $[a,b]$ (значения функции на отрезке образуют непустое числовое ограниченное множество).

Допустим, что функция $f(x)$ не принимает ни в одной точке $[a,b]$ значения, равного M. Тогда для всех $x \in [a,b]$ должно быть выполнено неравенство $f(x) < M$. Рассмотрим вспомогательную функцию

$$F(x) = \frac{1}{M - f(x)}.$$

Эта функция, в силу теоремы 3.9, положительна и непрерывна на $[a,b]$ как частное двух непрерывных функций. Тогда по теореме 3.13 функция $F(x)$ ограничена, т.е. существует такое положительное число μ, что для всех x на отрезке $[a,b]$

$$F(x) = \frac{1}{M - f(x)} \leqslant \mu, \text{ откуда } f(x) \leqslant M - \frac{1}{\mu}.$$

Последнее неравенство противоречит тому, что число M является точной верхней гранью функции $f(x)$ на отрезке $[a,b]$ (так как $M - 1/\mu < M$). Это противоречие доказывает, что существует точка $x_1 \in [a,b]$, в которой $f(x_1) = M$.

Достижение функцией $f(x)$ на отрезке $[a,b]$ своей точной нижней грани доказывается аналогично. Теорема доказана.

Заметим, что для непрерывной на отрезке функции ее точную верхнюю грань называют *максимальным значением*, а точную нижнюю грань — *минимальным значением функции* на этом отрезке.

Разность между максимальным и минимальным значениями непрерывной на отрезке $[a,b]$ функции $f(x)$ называется *колебанием* ее на этом отрезке:

$$\omega = M - m, \; M = \max_{[a,b]} f(x), \; m = \min_{[a,b]} f(x).$$

Понятие равномерной непрерывности функции

Пусть функция $f(x)$ непрерывна на некотором промежутке X. Если точка $x_0 \in X$, то $f(x)$ непрерывна в этой точке; тогда, согласно определению непрерывности функции в точке "на языке ε–δ", для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что $|f(x) - f(x_0)| < \varepsilon$ как только $|x - x_0| < \delta$. Вполне понятно, что для данного ε число δ зависит еще и от точки x_0 (рис. 3.19): чем "круче" график функции и окрестности точки x_0, тем меньше δ.

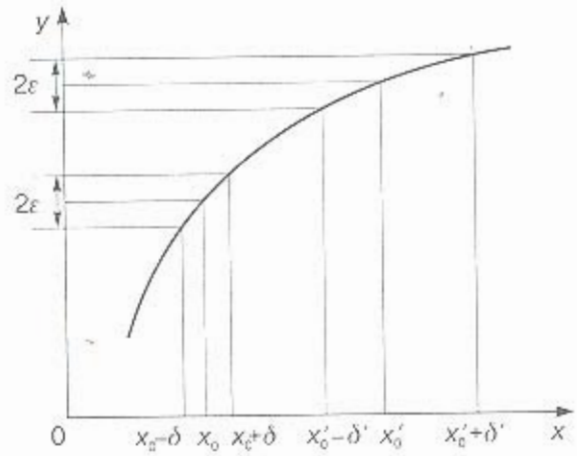

Рис. 3.19

Определение 2. Функция $f(x)$ называется *равномерно-непрерывной* на промежутке X, если для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что для любых двух точек $x_1, x_2 \in X$, удовлетворяющих неравенству $|x_1 - x_2| < \delta$ выполняется неравенство $|f(x_1) - f(x_2)| < \varepsilon$.

Для равномерно-непрерывной функции величина δ зависит только от ε и является общей для всего промежутка X.

Из определения следует, что если функция равномерно-непрерывна на X: а) она является непрерывной на этом промежутке, б) по заданному $\varepsilon > 0$ можно найти такое $\delta > 0$, что промежуток X может быть разбит на составные промежутки длиной меньше 2δ, на каждом из которых колебание функции составляет менее 2ε.

Приведем без доказательства следующую теорему.

Теорема 3.15 (ТЕОРЕМА КАНТОРА[1]). *Если функция $f(x)$ непрерывна на отрезке $[a,b]$, то она и равномерно-непрерывна на нем.*

Следует отметить, что теорема неверна для интервала или полуинтервала. Например, функция $f(x) = 1/x$ непрерывна на интервале $(0,a)$, где a — конечное число, однако она не является равномерно-непрерывной на нем; для любого $\varepsilon > 0$, какое бы $\delta > 0$ ни взять, найдутся достаточно близкие к нулю точки x_1 и x_2, что при $|x_2 - x_1| < \delta$ тем не менее $|f(x_2) - f(x_1)| > \varepsilon$.

Свойство равномерной непрерывности широко используется в математике.

3.10. Понятие сложной функции

Определение. Если на некотором промежутке X определена функция $z = \varphi(x)$ с множеством значений Z и на множестве Z определена функция $y = f(z)$, то функция $y = f[\varphi(x)]$ называется *сложной функцией* от x (или *суперпозицией функций*), а переменная z — промежуточной переменной сложной функции.

Примеры сложных функций.

1. $y = \cos\sqrt{1-x}$ — сложная функция, определенная на полубесконечном интервале $(-\infty, 1]$, так как $y = f(z) = \cos z$, $z = \varphi(x) = \sqrt{1-x}$.

2. $y = e^{-x^2}$ — сложная функция, определенная на всей числовой прямой, поскольку $y = f(z) = e^z$, $z = \varphi(x) = -x^2$.

3. $y = \left(\dfrac{1+x}{x}\right)^{3/2}$ — сложная функция, определенная на полубесконечных интервалах $(-\infty, 0)$ и $(0, +\infty)$, так как $y = f(z) = z^{3/2}$, $z = \varphi(x) = (1+x)/x$.

Теорема 3.16. *Пусть функция $z = \varphi(x)$ непрерывна в точке x_0, а функция $y = f(z)$ непрерывна в точке $z_0 = \varphi(x_0)$. Тогда сложная функция $y = f[\varphi(x)]$ непрерывна в точке x_0.*

[1] Георг Кантор (1845–1918) — немецкий математик, основатель современной теории множеств.

Доказательство. Возьмем из X любую последовательность $\{x_n\}$ точек, сходящуюся к точке x_0. Так как функция $z = \varphi(x)$ непрерывна в точке x_0, то соответствующая последовательность точек $\{z_n\}$ сходится к точке z_0, или $\lim\limits_{n\to\infty} z_n = \lim\limits_{n\to\infty} \varphi(x_n) = \varphi(x_0) = z_0$. Поскольку функция $f(z)$ непрерывна в точке z_0, имеем $\lim\limits_{n\to\infty} f[\varphi(x_n)] = f[\varphi(x_0)]$. Следовательно, предел функции $f[\varphi(x)]$ равен ее значению в этой точке, что и доказывает теорему.

Пример. Функция $y = \operatorname{tg}(x^2 + 2x)$ непрерывна в точке $x = 0$, так как функция $z = x^2 + 2x$ непрерывна в точке $x = 0$, а функция $y = \operatorname{tg} z$ непрерывна в точке $z = 0$.

3.11. Понятие обратной функции

Определение обратной функции

Говорят, что функция $f(x)$ не убывает (не возрастает) на множестве X, если для любых $x_1, x_2 \in X$, таких, что $x_1 < x_2$, справедливо неравенство $f(x_1) \leqslant f(x_2)$ $(f(x_1) \geqslant f(x_2))$.

Неубывающие и невозрастающие функции называются *монотонными* функциями.

Если для любых $x_1, x_2 \in X$, таких, что $x_1 < x_2$ выполняется неравенство $f(x_1) < f(x_2)$ $(f(x_1) > f(x_2))$, то функция $f(x)$ называется *возрастающей (убывающей)* на множестве X. Такие функции называются также *строго монотонными*.

Приведем несколько примеров.

1. Функция $f(x) = \operatorname{sign} x$ является неубывающей на всей числовой прямой.

2. Функция $f(x) = x^2$ является убывающей на промежутке $(-\infty, 0)$ и возрастающей на промежутке $[0, +\infty)$.

3. Функция $f(x) = x^3$ является возрастающей на всей числовой прямой.

Определение. Пусть на множестве X задана функция $f(x)$, причем Y — множество ее значений, т.е. задано множество пар чисел (x, y) $(x \in X, y \in Y)$, в котором каждое число x входит лишь в одну пару, а каждое число y — по крайней мере в одну пару. Если в каждой паре этого множества поменять местами числа x и y, то получим множество пар чисел (y, x), которое называется *обратной функцией* к функции $f(x)$.

Обозначим обратную функцию символом $x = \varphi(y)$.

Вообще говоря, обратная функция, определенная выше, не является функцией[1], поскольку теперь одному значению аргумента y может соответствовать несколько значений функции $x = \varphi(y)$. Например, для функции $y = x$ обратная функция $x = y$ однозначна, для функции $y = x^2$ обратная функция $x = \pm\sqrt{y}$ двузначна, для функции $y = \sin x$ обратная функция $x = \operatorname{Arcsin} y$ многозначна. Следовательно, для того чтобы обратная функция соответствовала определению функции, необходимо, чтобы исходная функция $f(x)$ была однозначной. В таком случае и обратная функция будет однозначной.

Пусть задана функция $y = f(x)$ на множестве X; рассмотрим процесс получения функции, обратной к функции $f(x)$, согласно определению, данному выше, поэтапно.

1. Сначала надо разбить область определения функции $y = f(x)$ на участки строгой монотонности. Без потери общности можно полагать далее, что функция $f(x)$ строго монотонна.

2. Из формулы, задающей функцию $y = f(x)$, нужно выразить x через y (из каждой пары (x, y) нужно получить пару (y, x), где y будет аргументом, а x — функцией). Теперь область значений Y функции $y = f(x)$ становится областью определения функции $x = \varphi(y)$, а область определения функции $y = f(x)$ становится областью значений функции $x = \varphi(y)$.

3. Функция $y = f(x)$ и обратная функция $x = \varphi(y)$ имеют один и тот же график. Например, взаимно обратные функции $y = 2x$ и $x = y/2$ изображаются графически одной прямой. Для получения графика обратной функции необходимо оси Ox и Oy поменять местами, для чего следует повернуть в пространстве координатную плоскость Oxy вокруг биссектрисы первого и третьего координатных углов на 180°.

4. Поскольку новое положение графика обратной функции $x = \varphi(y)$ является графиком обратной функции $y = \varphi(x)$, то в формуле $x = \varphi(y)$ следует формально поменять x и y местами (это следует из взаимной замены осей Ox и Oy).

Заметим, что:

а) графики взаимно обратных функций симметричны относительно биссектрисы первого и третьего координатных углов или прямой $y = x$ (рис. 3.20);

[1] Можно сказать, что она является многозначной функцией.

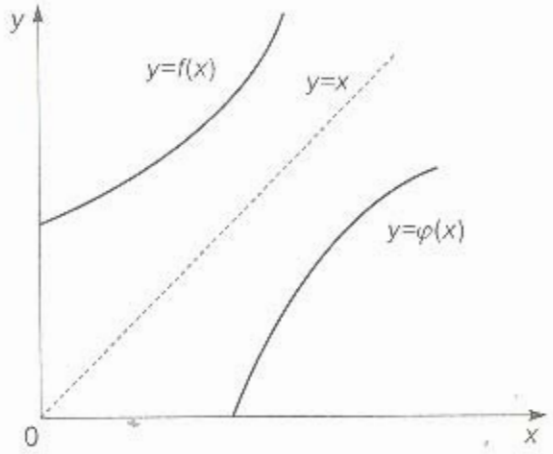

Рис. 3.20

б) взаимно обратные функции имеют одинаковый характер монотонности.

Непрерывность обратной функции

Приведем без доказательства теорему о непрерывности обратной функции.

Теорема 3.17. *Пусть функция $y = f(x)$ определена, строго монотонна и непрерывна на некотором промежутке X и пусть Y — множество ее значений. Тогда на множестве Y обратная функция $x = \varphi(y)$ однозначна, строго монотонна и непрерывна.*

3.12. Элементы аналитической геометрии на плоскости

Уравнение линии на плоскости

Пусть на плоскости задана система координат. Рассмотрим уравнение вида

$$F(x, y) = 0. \tag{3.20}$$

Это равенство, если оно соблюдается не для всех пар чисел x и y, называется *уравнением некоторой линии L в заданной*

системе координат. Говорят, что уравнение (3.20) определяет или задает линию L.

Вообще говоря, линии на координатной плоскости могут быть самыми различными.

Линии первого порядка

К линиям первого порядка относятся те линии, для которых задающее их уравнение (3.20) содержит переменные x и y только в первой степени. Иными словами, такие линии описываются уравнениями вида

$$Ax + By + C = 0, \tag{3.21}$$

где A, B и C — постоянные числа. Из этого уравнения можно выразить переменную y как функцию от аргумента x при $B \neq \neq 0$:

$$y = kx + b. \tag{3.22}$$

Уравнение (3.22) называют *уравнением прямой с угловым коэффициентом* $k = \operatorname{tg}\varphi$, где φ — угол наклона прямой к положительному направлению оси Ox (рис. 3.21). Если $k = 0$, то прямая параллельна оси Ox и отстоит от нее на b масштабных единиц.

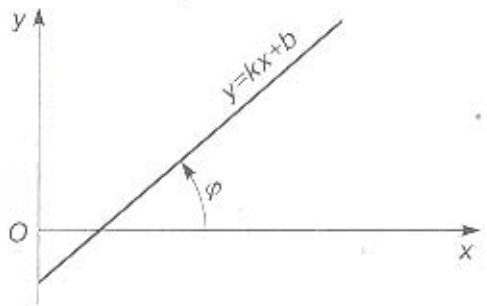

Рис. 3.21

Определим самые необходимые элементы знания о прямых на плоскости.

Кроме "классического" уравнения прямой (3.22) следует знать еще две его разновидности. Первая из них — это уравнение прямой с заданным угловым коэффициентом k и проходящей через заданную точку $M_0(x_0, y_0)$:

$$y - y_0 = k(x - x_0). \tag{3.23}$$

Другой вид — это уравнение прямой, проходящей через две заданные точки на плоскости $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$:

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1). \tag{3.24}$$

Уравнения (3.23) и (3.24) несложным образом следуют из уравнения (3.22).

Угол между прямыми. Рассмотрим две прямые, заданные уравнениями $y = k_1x + b_1$ и $y = k_2x + b_2$, где $k_1 = \operatorname{tg}\varphi_1$ и $k_2 = \operatorname{tg}\varphi_2$ (рис. 3.22). Пусть φ — угол между этими прямыми. Тогда $\varphi = \varphi_2 - \varphi_1$, и мы получаем, что

$$\operatorname{tg}\varphi = \operatorname{tg}(\varphi_2 - \varphi_1) = \frac{\operatorname{tg}\varphi_2 - \operatorname{tg}\varphi_1}{1 + \operatorname{tg}\varphi_1 \operatorname{tg}\varphi_2}, \quad \text{или} \quad \operatorname{tg}\varphi = \frac{k_2 - k_1}{1 + k_1k_2}. \tag{3.25}$$

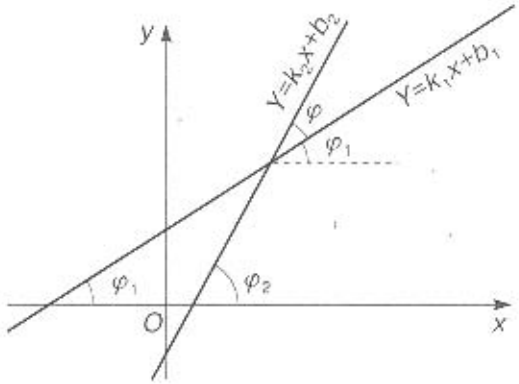

Рис. 3.22

Формула (3.25) определяет один из углов между пересекающимися прямыми; второй угол равен $\pi - \varphi$.

Из равенства (3.25) вытекают условия параллельности и перпендикулярности прямых. В самом деле, если прямые параллельны, то

$$k_1 = k_2.$$

Если прямые перпендикулярны, то $\varphi_2 = \pi/2 + \varphi_1$, откуда $\operatorname{tg}\varphi_2 = -\operatorname{ctg}\varphi_1 = -1/\operatorname{tg}\varphi_1$, или окончательно

$$k_2 = -\frac{1}{k_1}.$$

Пример. Найти угол между прямыми, заданными уравнениями

$$y = 2x - 5 \quad \text{и} \quad y = -3x + 4.$$

Подставляя в формулу (3.25) значения $k_1 = 2$ и $k_2 = -3$, имеем

$$\operatorname{tg}\varphi = \frac{-3-2}{1-6} = 1,$$

откуда один из углов равен $\varphi = \pi/4$.

Расстояние от точки до прямой. Пусть прямая задана уравнением общего вида (3.21). Тогда расстояние d от произвольной точки $M_0(x_0, y_0)$ до прямой (рис. 3.23) задается формулой

$$d = |Ax_0 + By_0 + C|/\sqrt{A^2 + B^2}. \qquad (3.26)$$

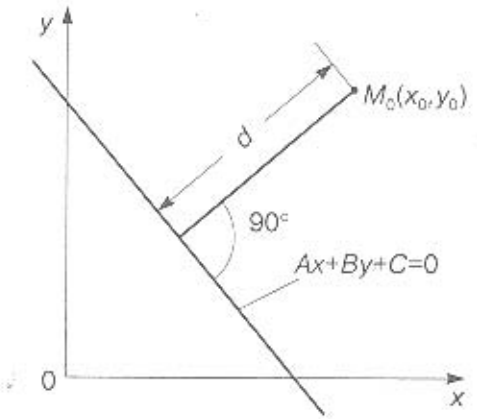

Рис. 3.23

Линии второго порядка

Рассмотрим три наиболее используемых вида линий: эллипс, гиперболу и параболу. Линия, для всех точек которой сумма расстояний от двух данных точек, называемых *фокусами*, есть величина постоянная и большая, чем расстояние между фокусами, называется **эллипсом.**

Согласно определению эллипса, сумма расстояний от произвольной точки M на этой линии до его фокусов F_1 и F_2 постоянна (рис. 3.24):

$$r_1 + r_2 = 2a.$$

Отсюда можно вывести *каноническое уравнение эллипса*:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad (3.27)$$

где a и b — полуоси эллипса (соответственно малая и большая), $b^2 = a^2 - c^2$. Точка $O(0,0)$ — центр эллипса, c — половина расстояния между фокусами эллипса. Из уравнения (3.27) следует, что оси эллипса являются его осями симметрии, а точка их пересечения является центром его симметрии.

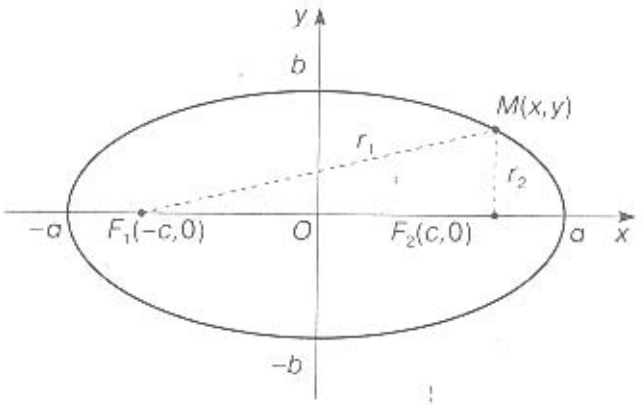

Рис. 3.24

В частном случае, когда $a = b$, фокусы эллипса сливаются, $c = 0$ и мы имеем окружность радиуса a с центром в начале координат.

Характеристикой эллипса, показывающей меру его вытянутости, является *эксцентриситет* — величина, определяемая отношением

$$\varepsilon = \frac{c}{a}.$$

Гиперболой называется линия, для всех точек которой модуль разности расстояний от двух данных точек, называемых *фокусами*, есть величина постоянная и меньшая, чем расстояние между фокусами.

На рис. 3.25 показаны все основные элементы гиперболы. Разность расстояний от произвольной точки M на гиперболе до фокусов F_1 и F_2, согласно определению, есть величина постоянная:

$$|r_1 - r_2| = 2a.$$

Из этой основной предпосылки выводится *каноническое уравнение гиперболы*, которое имеет вид

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad (3.28)$$

где $b^2 = c^2 - a^2$.

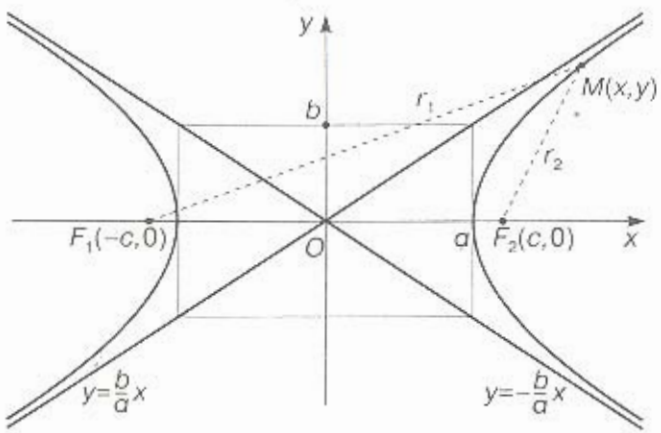

Рис. 3.25

Нетрудно видеть, что прямые $y = \pm\frac{b}{a}x$ являются наклонными асимптотами гиперболы. Эта линия имеет две оси симметрии, точка пересечения которых является центром симметрии гиперболы.

Параболой называется линия, все точки которой находятся на одинаковом расстоянии от данной точки, называемой *фокусом*, и от данной прямой, называемой *директрисой* и не проходящей через фокус.

Согласно определению, точка $M\ (x, y)$ лежит на параболе, если $r_1 = r_2$. Отсюда и выводится *каноническое уравнение параболы*, которое имеет вид

$$y = \sqrt{2px} \qquad (3.29)$$

График параболы показан на рис. 3.26. Нетрудно видеть, что перемена осей координат приводит к более привычным по курсу школьной математики уравнениям параболы вида $y = Ax^2$, где A — постоянное число.

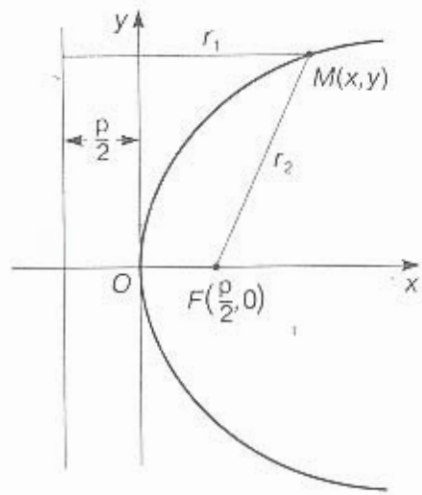

Рис. 3.26

Общее уравнение линии второго порядка

Одна из основных задач аналитической геометрии — исследование общего уравнения линии второго порядка и приведение его к простейшим каноническим формам.

Общее уравнение линии второго порядка имеет вид

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0, \qquad (3.30)$$

где A, B, C, D, E и F — некоторые числа, причем $A^2 + B^2 + C^2 \neq 0$. В аналитической геометрии существует теорема, согласно которой при $AC - B^2 \neq 0$ с помощью параллельного сдвига и последующего поворота осей координат уравнение (3.30) можно привести к виду

$$A'(x'')^2 + C'(y'')^2 + F' = 0, \qquad (3.31)$$

где A', C' и F' — некоторые числа, а (x'', y'') — координаты точки в новой системе координат.

Оказывается, что выражение $AC - B^2$ является *инвариантом* общего уравнения линии второго порядка, т.е. оно остается неизменным при параллельном переносе и повороте осей координат. В связи с этим в зависимости от знака указанного инварианта линии второго порядка классифицируются по следующим трем типам:

1) эллиптический, при $AC - B^2 > 0$;
2) гиперболический, при $AC - B^2 < 0$;
3) параболический, при $AC - B^2 = 0$.

В свою очередь, внутри каждого из указанных типов, в зависимости от знаков коэффициентов общего уравнения (3.31), возможны разные предельные случаи вырождения линий второго порядка в прямые или даже исчезновения с вещественной координатной плоскости.

Глава 4
ОСНОВЫ ДИФФЕРЕНЦИАЛЬНОГО ИСЧИСЛЕНИЯ

4.1. Понятие производной

Определение производной

Пусть функция $f(x)$ определена на некотором промежутке X. Придадим значению аргумента в точке $x_0 \in X$ произвольное приращение Δx так, чтобы точка $x_0 + \Delta x$ также принадлежала X. Тогда соответствующее приращение функции $f(x)$ составит $\Delta y = f(x_0 + \Delta x) - f(x_0)$.

Определение 1. *Производной функции* $f(x)$ в точке x_0 называется предел отношения приращения функции в этой точке к приращению аргумента при $\Delta x \to 0$ (если этот предел существует).

Для обозначения производной функции Δx применяются следующие символы $y'(x_0)$ или $f'(x_0)$:

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}. \qquad (4.1)$$

Если в некоторой точке x_0 предел (4.1) бесконечен:

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = +\infty \text{ или } \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = -\infty,$$

то говорят, что в точке x_0 функция $f(x)$ имеет *бесконечную производную*.

Если функция $f(x)$ имеет производную в каждой точке множества X, то производная $f'(x)$ также является функцией от аргумента x, определенной на X.

Геометрический смысл производной

Для выяснения геометрического смысла производной нам понадобится определение касательной к графику функции в данной точке.

Касательной к графику функции $y = f(x)$ в точке M называется предельное положение секущей MN, когда точка N стремится к точке M по кривой $f(x)$.

Пусть точка M на кривой $f(x)$ соответствует значению аргумента x_0, а точка N — значению аргумента $x_0 + \Delta x$ (рис. 4.1). Из определения касательной следует, что для ее существования в точке x_0 нужно, чтобы существовал предел $\lim\limits_{\Delta x \to 0} \varphi(x_0) = \varphi_0$, который равен углу наклона касательной к оси Ox. Из треугольника MNA следует, что

$$\operatorname{tg} \varphi(\Delta x) = \frac{\Delta y}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

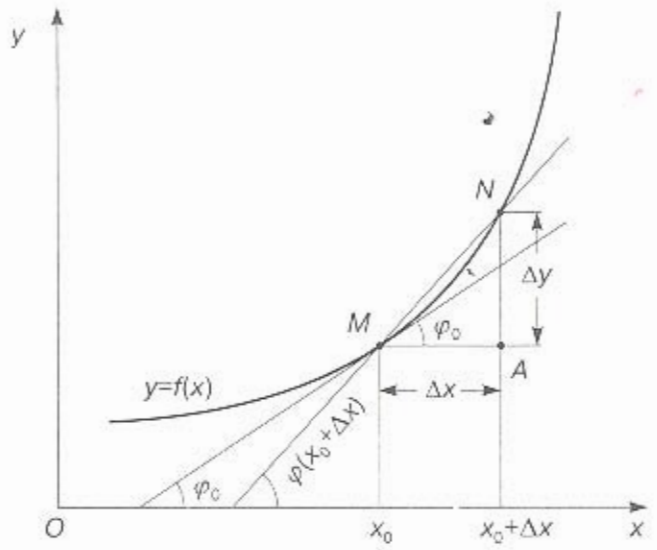

Рис. 4.1

Если производная функции $f(x)$ в точке x_0 существует, то, согласно определению производной (4.1), получаем

$$\operatorname{tg}\varphi_0 = f'(x_0). \tag{4.2}$$

Отсюда следует наглядный вывод о том, что производная $f'(x)$ равна угловому коэффициенту (тангенсу угла наклона к положительному направлению оси Ox) касательной к графику функции $y = f(x)$ в точке $M\ (x_0, f(x_0))$. При этом угол наклона касательной определяется из формулы (4.2):

$$\varphi_0 = \operatorname{arctg} f'(x_0). \tag{4.3}$$

Физический смысл производной

Предположим, что функция $l = f(t)$ описывает закон движения материальной точки по прямой, как зависимость пути l от времени t. Тогда разность $\Delta l = f(t+\Delta t) - f(t)$ — это путь, пройденный за интервал времени Δt, а отношение $\Delta l/\Delta t$ — средняя скорость за время Δt. Тогда предел $\lim\limits_{\Delta t\to 0}(\Delta l/\Delta t) = f'(t)$ определяет *мгновенную скорость* точки в момент времени t как производную пути по времени.

В определенном смысле производную функции $y = f(x)$ можно также трактовать как скорость изменения функции: чем больше величина $f'(x)$, тем больше угол наклона касательной к кривой, тем круче график $f(x)$ и быстрее растет функция.

Правая и левая производные

По аналогии с понятиями односторонних пределов функции вводятся понятия правой и левой производных функции в точке.

Определение 2. *Правой (левой) производной функции* $y = f(x)$ в точке x_0 называется правый (левый) предел отношения (4.1) при $\Delta x \to 0$, если этот предел существует.

Для обозначения односторонних производных используется следующая символика:

$$f'_+(x_0) = \lim_{\Delta x\to 0+}\frac{\Delta y}{\Delta x};\ f'_-(x_0) = \lim_{\Delta x\to 0-}\frac{\Delta y}{\Delta x}. \tag{4.4}$$

Если функция $f(x)$ имеет в точке x_0 производную, то она имеет левую и правую производные в этой точке, которые совпадают.

Приведем пример функции, у которой существуют односторонние производные в точке, не равные друг другу. Это $f(x) = |x|$. Действительно, в точке $x = 0$ имеем $f'_+(0) = 1$, $f'_-(x) = -1$ (см. рис. 3.14) и $f'_+(0) \neq f'_-(0)$, т.е. функция не имеет производной при $x = 0$.

Вообще, функции типа $y = |f(x)|$ не имеют производных в точках действительных корней уравнения $f(x) = 0$, поскольку, согласно определению модуля, для этих функций характерен излом в таких точках, а значит, и разрыв производной, так как $y'_+ \neq y'_-$ (рис. 4.2).

4.2. Понятие дифференцируемости функции

Дифференцируемость функции в точке

Определение 1. Функция $y = f(x)$ называется *дифференцируемой* в точке x_0, если ее приращение Δy в этой точке можно представить в виде

$$\Delta y = A\Delta x + \alpha(\Delta x)\Delta x, \tag{4.5}$$

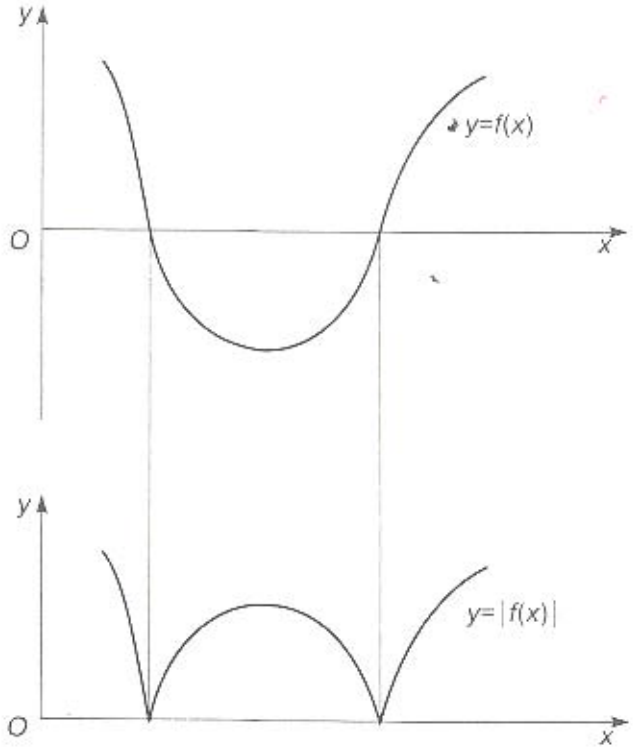

Рис. 4.2

где A — некоторое число, не зависящее от Δx, а $\alpha(\Delta x)$ — бесконечно малая функция при $\Delta x \to 0$.

Поскольку произведение двух бесконечно малых функций α и Δx является бесконечно малой более высокого порядка малости (см. параграф 3.5), то формулу (4.5) можно представить в виде

$$\Delta y = A\Delta x + o(\Delta x). \tag{4.6}$$

Теорема 4.1. *Для того чтобы функция $y = f(x)$ была дифференцируемой в точке x_0, необходимо и достаточно, чтобы она имела в этой точке конечную производную.*

Доказательство. Докажем сначала *необходимость*, т.е. пусть дано, что функция $f(x)$ дифференцируема в точке x_0, и нужно доказать, что она имеет производную в этой точке. Тогда приращение функции Δy, обусловленное приращением аргумента

Δx, можно представить в виде уравнения (4.5); поделив это равенство на $\Delta x \neq 0$, получим

$$\frac{\Delta y}{\Delta x} = A + \alpha(\Delta x).$$

В пределе при $\Delta x \to 0$

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = f'(x_0) = \lim_{\Delta x \to 0} (A + \alpha(\Delta x)) = A,$$

откуда и следует, что производная функции $f(x)$ в точке x_0 существует и равна A.

Теперь докажем *достаточность*: дано, что функция $y = f(x)$ имеет производную в точке x_0, надо доказать, что ее можно представить в виде выражения (4.5). Пусть $f'(x_0) = A$; из определения производной (4.1) следует, что функция $\alpha(\Delta x) = \Delta y/\Delta x - A$ бесконечно малая при $\Delta x \to 0$ (см. теорему 3.7). Умножив обе части последнего равенства на Δx, получаем формулу (4.5) для функции $f(x)$, дифференцируемой в точке x_0. Теорема доказана.

Теорема 4.1 позволяет в дальнейшем отождествлять дифференцируемость и существование производной для функций одной переменной. Операцию нахождения производной обычно называют *дифференцированием*.

Если функция $f(x)$ имеет производную в каждой точке некоторого промежутка, то говорят, что она дифференцируема на этом промежутке.

Связь между понятиями дифференцируемости и непрерывности функции

Теорема 4.2. *Если функция дифференцируема в точке x_0, то она и непрерывна в этой точке.*

Доказательство. В силу определения дифференцируемости функции $f(x)$ в точке x_0 ее приращение в этой точке можно представить в виде формулы (4.5). Переходя к пределу при $\Delta x \to 0$, получаем

$$\lim_{\Delta x \to 0} \Delta y = A \lim_{\Delta x \to 0} \Delta x + \lim_{\Delta x \to 0} \alpha(\Delta x) \lim_{\Delta x \to 0} \Delta x = 0,$$

что соответствует третьему определению непрерывности функции в форме записи (3.17). Теорема доказана.

Обратное утверждение неверно: функция $f(x)$, непрерывная в точке, может не иметь производную в этой точке. Например, функция $y = |x|$ непрерывна в точке $x = 0$, но не имеет производной в этой точке (см. рис. 3.14).

Таким образом, требование дифференцируемости функции более сильное, чем требование непрерывности, поскольку из первого автоматически вытекает второе.

4.3. Понятие дифференциала функции

Определение и геометрический смысл дифференциала

Поскольку для функции $y = f(x)$, дифференцируемой в точке x_0, приращение Δy в этой точке может быть представлено в виде формулы (4.6), то слагаемое $A\Delta x$ при $\Delta x \to 0$ является бесконечно малой одного порядка с Δx, тогда как второе слагаемое в уравнении (4.6) — бесконечно малая более высокого порядка, иными словами, величина $A\Delta x$ — главная часть приращения Δy, обусловленного приращением аргумента Δx.

Определение 1. *Дифференциалом функции* $y = f(x)$ в точке x_0 называется главная линейная относительно Δx часть приращения функции в этой точке:

$$dy = A\Delta x.$$

Поскольку из теоремы 4.1 следует, что $A = f'(x_0)$, то эту формулу можно переписать в виде

$$dy = f'(x_0)\Delta x. \qquad (4.7)$$

Дифференциалом dx независимой переменной x будем называть приращение этой переменной Δx, т.е. соотношение (4.7) принимает вид

$$dy = f'(x_0)dx. \qquad (4.8)$$

Из равенства (4.8) производную $f'(x)$ в любой точке x можно вычислить как отношение дифференциала функции dy к дифференциалу независимой переменной dx:

$$f'(x) = \frac{dy}{dx}. \qquad (4.9)$$

Равенство (4.6) с учетом теоремы 4.1 и формулы (4.7) можно переписать в виде

$$\Delta y = f'(x_0)\Delta x + o(\Delta x) = dy + o(\Delta x), \qquad (4.10)$$

что полностью соответствует определению дифференциала функции.

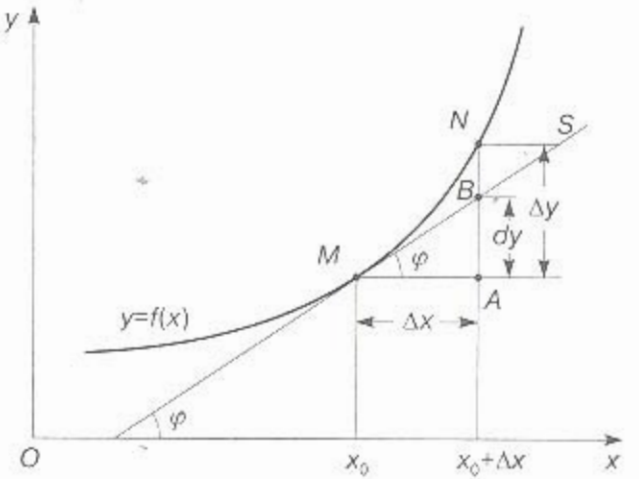

Рис. 4.3

Дифференциал функции имеет четкий геометрический смысл (рис. 4.3). Пусть точка M на графике функции $y = f(x)$ соответствует значению аргумента x_0, точка N — значению аргумента $x_0 + \Delta x$, MS — касательная к кривой $f(x)$ в точке M, φ — угол между касательной и осью Ox. Тогда MA — приращение аргумента, AN — соответствующее приращение функции. Рассматривая треугольник ABM, получаем $AB = \Delta x \operatorname{tg}\varphi = f'(x_0)\Delta x = dy$, т.е. это главная по порядку величины Δx и линейная относительно нее часть приращения функции Δy. Второе слагаемое в уравнении (4.10) более высокого порядка малости соответствует отрезку BN.

Приближенные вычисления с помощью дифференциала

Приближенные вычисления с применением дифференциала функции основаны на приближенной замене приращения функ-

ции в точке на ее дифференциал:

$$\Delta y = dy.$$

Абсолютная погрешность от такой замены является, как следует из формулы (4.10), при $x \to 0$ бесконечно малой более высокого порядка по сравнению с Δx. Подставляя в это приближенное соотношение формулу (4.7) и выражение для Δy, получаем

$$f(x_0 - \Delta x) \approx f(x_0) + f'(x_0)\Delta x. \tag{4.11}$$

Формула (4.11) является основной в приближенных вычислениях.

4.4. Правила дифференцирования суммы, произведения и частного

Теорема 4.3. *Если функции $u(x)$ и $v(x)$ дифференцируемы в точке x, то сумма (разность), произведение и частное этих функций (частное при условии $v(x) \neq 0$) также дифференцируемы в этой точке, причем справедливы следующие формулы:*

$$(u \pm v)' = u' \pm v',\ (uv)' = u'v + uv',\ \left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}. \tag{4.12}$$

Доказательство. Для вывода формул (4.12) будем использовать определение производной (4.1) и теорему 3.3 о пределах функций.

1) $$(u \pm v)' = \lim_{\Delta x \to 0} \frac{[u(x+\Delta x) - u(x)] \pm [v(x+\Delta x) - v(x)]}{\Delta x} =$$
$$= \lim_{\Delta x \to 0} \frac{u(x+\Delta x) - u(x)}{\Delta x} \pm \lim_{\Delta x \to 0} \frac{v(x+\Delta x) - v(x)}{\Delta x} =$$
$$= \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} \pm \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} = u' \pm v'.$$

Для вывода второй и третьей формул (4.12) используем также равенство $f(x + \Delta x) = f(x) + \Delta f(x)$.

2) В этом случае последовательно получаем

$$(uv)' = \lim_{\Delta x \to 0} \frac{u(x+\Delta x)v(x+\Delta x) - u(x)v(x)}{\Delta x} =$$

$$= \lim_{\Delta x \to 0} \frac{u(x)v(x) + \Delta u v(x) + u(x)\Delta v + \Delta u \Delta v - u(x)v(x)}{\Delta x} =$$
$$= \lim_{\Delta x \to 0} \left[v(x)\frac{\Delta u}{\Delta x} + u(x)\frac{\Delta v}{\Delta x} + \Delta u \frac{\Delta v}{\Delta x} \right] =$$
$$= v \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} + u \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} + \lim_{\Delta x \to 0} \Delta u \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} =$$
$$= u'v + uv' + 0 \cdot v' = u'v + uv'.$$

3) $$\left(\frac{u}{v}\right)' = \lim_{\Delta x \to 0} \left\{ \left[\frac{u(x+\Delta x)}{v(x+\Delta x)} - \frac{u}{v} \right] \Big/ \Delta x \right\} =$$
$$= \lim_{\Delta x \to 0} \frac{(u+\Delta u)v - u(v+\Delta v)}{\Delta x v(v+\Delta v)} =$$
$$= \lim_{\Delta x \to 0} \frac{uv + \Delta u v - uv - u\Delta v}{\Delta x v(v+\Delta v)} =$$
$$= \lim_{\Delta x \to 0} \left(v\frac{\Delta u}{\Delta x} - u\frac{\Delta v}{\Delta x} \right) \Big/ (v^2 + v\Delta u) =$$
$$= \left(v \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} - u \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} \right) \Big/ \left(v^2 + v \lim_{\Delta x \to 0} \Delta u \right) =$$
$$= \frac{u'v - uv'}{v^2}.$$

4.5. Производные постоянной, степенной, логарифмической и тригонометрических функций

Производная постоянной функции

Возьмем функцию $y = f(x) = C$, где C — постоянное число. Для любых x и $\Delta x \neq 0$ имеем $\Delta y = C - C = 0$, откуда $\Delta y / \Delta x = 0$, а значит,

$$y' = (C)' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 0. \tag{4.13}$$

Производная степенной функции

Выведем формулу для производной функции $y = x^n$, где n — целое положительное число. Здесь мы используем бином

Ньютона:

$$\Delta y = (x+\Delta x)^n - x^n =$$
$$= \left[x^n + nx^{n-1}\Delta x + \frac{n(n-1)}{2!}x^{n-2}(\Delta x)^2 + \cdots + (\Delta x)^n\right] - x^n =$$
$$= nx^{n-1}\Delta x + \frac{n(n-1)}{2!}x^{n-2}(\Delta x)^2 + \cdots + (\Delta x)^n.$$

Следовательно, при $\Delta x \neq 0$

$$\frac{\Delta y}{\Delta x} = nx^{n-1} + \frac{n(n-1)}{2!}x^{n-2}\Delta x + \cdots + (\Delta x)^{n-1},$$

но поскольку все слагаемые, содержащие Δx, в пределе при $\Delta x \to 0$ равны нулю, получаем

$$(x^n)' = nx^{n-1}. \tag{4.14}$$

Производная логарифмической функции

Для функции $y = \log_a x$ $(a \neq 1,\ a > 0,\ x > 0)$ при $\Delta x \neq 0$

$$\frac{\Delta y}{\Delta x} = \frac{1}{\Delta x}\left[\log_a(x+\Delta x) - \log_a x\right] = \frac{1}{\Delta x}\log_a\left(1+\frac{\Delta x}{x}\right) =$$
$$= \frac{1}{x}\frac{x}{\Delta x}\log_a\left(1+\frac{\Delta x}{x}\right) = \frac{1}{x}\log_a\left(1+\frac{\Delta x}{x}\right)^{\frac{x}{\Delta x}}.$$

Обозначим $x/\Delta x = t$, тогда при $\Delta x \to 0$ имеем, что $t \to \infty$. Отсюда, используя второй замечательный предел и непрерывность логарифмической функции, получаем

$$(\log_a x)' = \frac{1}{x}\lim_{t\to\infty}\log_a\left(1+\frac{1}{t}\right)^t = \frac{1}{x}\log_a\lim_{t\to\infty}\left(1+\frac{1}{t}\right)^t = \tag{4.15}$$
$$= \frac{1}{x}\log_a e = \frac{1}{x}\frac{1}{\ln a}.$$

Следствие. *Если $a = e$, то* $y' = (\ln x)' = \dfrac{1}{x}$.

Производные тригонометрических функций

Получим формулы для производных тригонометрических функций: $\sin x$, $\cos x$, $\operatorname{tg} x$ и $\operatorname{ctg} x$.

1) $y = \sin x$.

Приращение функции $\Delta y = \sin(x+\Delta x) - \sin x = 2\sin(\Delta x/2)\times$ $\times\cos(x+\Delta x/2)$, откуда при $\Delta x \neq 0$ получаем

$$\frac{\Delta y}{\Delta x} = \frac{2\sin(\Delta x/2)\cos(x+\Delta x/2)}{\Delta x} = \frac{\sin(\Delta x/2)}{\Delta x/2}\cos(x+\Delta x/2).$$

Так как $\lim\limits_{\Delta x\to 0}\dfrac{\sin(\Delta x/2)}{\Delta x/2} = 1$ (первый замечательный предел), а $\lim\limits_{\Delta x\to 0}\cos(x+\Delta x/2) = \cos x$ в силу непрерывности функции $\cos x$, то окончательно

$$y' = (\sin x)' = \cos x. \tag{4.16}$$

2) $y = \cos x$.

Приращение функции $\Delta y = \cos(x+\Delta x) - \cos x =$ $= -2\sin(\Delta x/2)\sin(x+\Delta x/2)$, откуда получаем при $\Delta x \neq 0$

$$\frac{\Delta y}{\Delta x} = -\frac{2\sin(\Delta x/2)\sin(x+\Delta x/2)}{\Delta x} =$$
$$= -\frac{\sin(\Delta x/2)}{\Delta x/2}\sin(x+\Delta x/2).$$

Поскольку $\lim\limits_{\Delta x\to 0}\dfrac{\sin(\Delta x/2)}{\Delta x/2} = 1$, а $\lim\limits_{\Delta x\to 0}\sin(x+\Delta x/2) = \sin x$ в силу непрерывности функции $\sin x$, то окончательно

$$y' = (\cos x)' = -\sin x. \tag{4.17}$$

3) $y = \operatorname{tg} x$.

Так как $\operatorname{tg} x = \sin x/\cos x$, то по теореме 4.3 получаем

$$y' = \frac{(\sin x)'\cos x - \sin x(\cos x)'}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x},$$

откуда

$$(\operatorname{tg} x)' = \frac{1}{\cos^2 x}, \left(x \neq \frac{\pi}{2} + n\pi,\ n = 0, \pm 1, \pm 2, \dots\right). \qquad (4.18)$$

4) $y = \operatorname{ctg} x$.

Так как $\operatorname{ctg} x = \cos x / \sin x$, то по аналогии с предыдущим

$$y' = \frac{(\cos x)' \sin x - \cos x (\sin x)'}{\sin^2 x} = -\frac{\sin^2 x + \cos^2 x}{\sin^2 x},$$

откуда окончательно получаем

$$y' = (\operatorname{ctg} x)' = -\frac{1}{\sin^2 x}, \ (x \neq \pi n, n = 0, \pm 1, \pm 2, \dots). \qquad (4.19)$$

4.6. Теорема о производной обратной функции

Пусть функция $y = f(x)$ удовлетворяет условиям теоремы 3.16 и функция $x = \varphi(y)$ является для нее обратной.

Теорема 4.4. *Пусть функция $y = f(x)$ является непрерывной и строго монотонной в некоторой окрестности точки x_0 и имеет в этой точке производную $f'(x_0) \neq 0$. Тогда обратная функция $x = \varphi(y)$ также имеет в соответствующей точке $y_0 = f(x_0)$ производную, причем*

$$\varphi'(y_0) = \frac{1}{f'(x_0)}. \qquad (4.20)$$

Доказательство. В силу теоремы 3.17 обратная функция $x = \varphi(y)$ определена и непрерывна в некоторой окрестности точки $y = f(x_0)$. Возьмем приращение аргумента $\Delta y \neq 0$ в точке y_0, тогда функция $x = \varphi(y)$ получит некоторое приращение Δx, причем в силу строгой монотонности этой функции $\Delta x \neq 0$. Мы можем записать:

$$\frac{\Delta x}{\Delta y} = \frac{1}{\Delta y / \Delta x}.$$

Перейдем в этом тождестве к пределу при $\Delta y \to 0$. Поскольку обратная функция $x = \varphi(y)$ непрерывна в точке y_0, то и $\Delta x \to 0$. Но предел правой части тождества существует и равен $1/f'(x_0)$. Значит, существует и предел левой части этого равенства, равный по определению $\varphi'(y_0)$, т.е. справедлива формула (4.20), что и требовалось доказать.

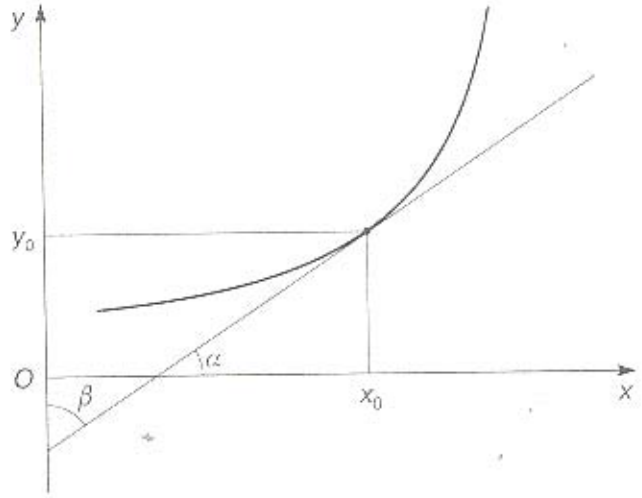

Рис. 4.4

Теорема 4.4 имеет ясный геометрический смысл (рис. 4.4). Производная $f'(x_0)$ равна тангенсу угла α наклона касательной к оси Ox в точке M. Производная обратной функции $\varphi'(y_0)$ равна тангенсу угла β наклона той же касательной к оси Oy. Так как $\alpha + \beta = \pi/2$, то формула (4.20) подтверждает очевидное равенство

$$\varphi'(y_0) = \operatorname{tg} \beta = \frac{1}{\operatorname{ctg} \beta} = \frac{1}{\operatorname{ctg}(\pi/2 - \alpha)} = \frac{1}{\operatorname{tg} \alpha} = \frac{1}{f'(x_0)}.$$

4.7. Производные показательной и обратных тригонометрических функций

Продолжим вычисление производных простейших элементарных функций, применяя доказанную теорему 4.4.

Производная показательной функции

Показательная функция $y = a^x$ является обратной для логарифмической функции $x = \log_a y$. Поскольку, согласно формуле (4.15),

$$x'(y) = \frac{1}{y} \frac{1}{\ln a},$$

то в силу теоремы о производной обратной функции (теорема 4.4)

$$y'(x) = (a^x)' = \frac{1}{x'(y)} = a^x \ln a. \tag{4.21}$$

Производные обратных тригонометрических функций

Выведем формулы для производных обратных тригонометрических функций: $\arcsin x$, $\arccos x$, $\operatorname{arctg} x$ и $\operatorname{arcctg} x$.

1) $y = \arcsin x$.

Эта функция является обратной для функции $x = \sin y$. Тогда по теореме о производной обратной функции

$$y'(x) = \frac{1}{x'(y)} = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - \sin^2 y}}.$$

Корень взят со знаком "плюс", так как $\cos y > 0$ при $-\frac{\pi}{2} < y < \frac{\pi}{2}$. Поскольку $x = \sin y$, окончательно получаем

$$(\arcsin x)' = \frac{1}{\sqrt{1 - x^2}}. \tag{4.22}$$

2) $y = \arccos x$.

Вывод формулы для производной этой функции аналогичен предыдущему:

$$(\arccos x)' = -\frac{1}{\sqrt{1 - x^2}}. \tag{4.23}$$

3) $y = \operatorname{arctg} x$.

Эта функция является обратной для функции $x = \operatorname{tg} y$. Поскольку $x'(y) = 1/\cos^2 y$, то по теореме о производной обратной функции получаем

$$y'(x) = \frac{1}{x'(y)} = \cos^2 y = \frac{\cos^2 y}{\sin^2 y + \cos^2 y} = \frac{1}{1 + \operatorname{tg}^2 y}.$$

Поскольку $\operatorname{tg} y = x$, то окончательно получаем

$$(\operatorname{arctg} x)' = \frac{1}{1 + x^2}. \tag{4.24}$$

4) $y = \operatorname{arcctg} x$.

Вывод формулы для производной этой функции аналогичен предыдущему, так что окончательно получаем

$$(\operatorname{arcctg} x)' = -\frac{1}{1 + x^2}. \tag{4.25}$$

4.8. Дифференцирование сложной функции

Правило дифференцирования сложной функции

Теорема 4.5. *Пусть функция $x = \varphi(t)$ имеет производную в точке t_0, а функция $y = f(x)$ имеет производную в соответствующей точке $x_0 = \varphi(t_0)$. Тогда сложная функция $f[\varphi(t)]$ имеет производную в точке t_0 и справедлива следующая формула:*

$$y'(t_0) = \left.\frac{dy}{dt}\right|_{t=t_0} = f'(x_0)\varphi'(t_0). \tag{4.26}$$

Доказательство. Согласно определению дифференциала функции в точке и формуле (4.8) при $x = x_0$

$$dy = f'(x_0)\,dx.$$

В свою очередь, дифференциал функции $x = \varphi(t)$ в точке (t_0)

$$dx = \varphi'(t_0)\,dt.$$

Подставляя это выражение в предыдущее равенство, получаем

$$dy = f'(x_0)\varphi'(t_0)\,dt.$$

Поделив обе части последнего равенства на dt, получаем формулу (4.26), что и завершает доказательство теоремы.

В теореме 4.5 рассмотрена суперпозиция двух функций, где y зависит от t через промежуточную переменную x. Возможна и более сложная зависимость с несколькими промежуточными переменными, однако правило дифференцирования сложной функции остается тем же. Например, если $y = f(x)$, $x = \varphi(u)$, $u = \psi(t)$, то производная $y'(t)$ вычисляется по формуле

$$y'(t) = y'(x)\varphi'(u)\psi'(t). \tag{4.27}$$

8 Зак. 6853

Рассмотрим несколько примеров на дифференцирование сложной функции.

Пример 1. Найти производную функции $y = \operatorname{tg}(x^3)$.

Эту функцию можно представить через промежуточную переменную u как $y = \operatorname{tg} u$, $u = x^3$. Тогда по формуле (4.26)

$$y'(x) = y'(u)u'(x) = \frac{1}{\cos^2 u} 3x^2 = \frac{3x^2}{\cos^2(x^3)}.$$

Пример 2. Найти производную функции $y = e^{\operatorname{tg}^2 4x}$.

Здесь функция представляется с помощью трех промежуточных переменных: $y = e^u$, $u = v^2$, $v = \operatorname{tg} w$, $w = 4x$.

Применяя правило (4.27) дифференцирования сложной функции, последовательно получаем

$$y'(x) = y'(u)u'(v)v'(w)w'(x) = e^u 2v \frac{1}{\cos^2 w} 4 = \frac{8 \operatorname{tg} 4x}{\cos^2 4x} e^{\operatorname{tg}^2 4x}.$$

Пример 3. Найти угол наклона к оси Ox касательной к графику функции $y = e^{-\sin 3x} + \operatorname{tg} 4x$ в точке $x = 0$.

Данная функция является суммой двух сложных функций, представляемых через промежуточные переменные как

$$y = e^u + \operatorname{tg} \alpha, \quad u = -w, \quad w = \sin z, \quad z = 3x, \quad \alpha = 4x.$$

Применяя правила дифференцирования суммы функций и сложных функций, получаем

$$y'(x) = e^u u'(w)w'(z)z'(x) + \frac{1}{\cos^2 \alpha} \alpha'(x) =$$

$$= e^u(-1)\cos z \cdot 3 + \frac{4}{\cos^2 \alpha} = -3\cos 3x e^{-\sin 3x} + \frac{4}{\cos^2 4x}.$$

Поскольку тангенс угла наклона касательной к оси Ox при $x = 0$ равен значению производной в этой точке, из последнего равенства получаем, подставляя в него $x = 0$:

$$\operatorname{tg} \varphi = y'(0) = -3 \cdot 1 \cdot 1 + \frac{4}{1} = -3 + 4 = 1,$$

откуда $\varphi = \operatorname{arctg} 1 = 45^\circ$.

Инвариантность формы первого дифференциала

В параграфе 4.3 было установлено, что в случае, когда аргумент x является независимой переменной, дифференциал функции определяется формулой (4.8)

$$dy = f'(x)dx.$$

Докажем инвариантность формы первого дифференциала, т.е. универсальность и применимость этой формулы и в том случае, когда аргумент x сам является функцией другой переменной t.

Пусть функция $y = f(x)$ дифференцируема в точке x, а сам аргумент x является дифференцируемой функцией аргумента t, т.е. $x = \varphi(t)$. Тогда y — сложная функция $y = f[\varphi(t)]$ аргумента t (т.е. x — промежуточная переменная). В силу теоремы 4.5

$$y'(t) = f'(x)\varphi'(t).$$

Поскольку t является независимой переменной, то по форме (4.8) записи дифференциала для функции y получаем

$$dy = y'(x)\varphi'(t)\,dt. \tag{4.28}$$

Аналогично для дифференциала функции $x = \varphi(t)$ имеем $dx = \varphi'(t)\,dt$. Подставляя это выражение в формулу (4.28), получаем $dy = y'(x)\,dx$, т.е. опять приходим к форме (4.8) дифференциала для функции y, как если бы аргумент x был независимой переменной.

4.9. Логарифмическая производная и производная степенной функции с любым вещественным показателем

Понятие логарифмической производной функции

Пусть функция $f(x)$ положительна и дифференцируема в точке x. Вычислим производную функции $y = \ln f(x)$. По правилу дифференцирования сложной функции получаем

$$y'(x) = [\ln f(x)]' = \frac{f'(x)}{f(x)}. \tag{4.29}$$

Это выражение называется *логарифмической производной* функции $f(x)$.

Найдем с помощью логарифмической производной производную показательно-степенной функции

$$y = u(x)^{v(x)}, \tag{4.30}$$

где $u > 0$ и v — некоторые функции от аргумента x, имеющие в точке x соответствующие производные. Поскольку $\ln y = v(x) \ln u(x)$, то использование формулы (4.29) приводит к равенству

$$\frac{y'}{y} = [v(x) \ln u(x)]' = v'(x) \ln u(x) + v(x) \frac{u'(x)}{u(x)}.$$

С учетом вида функции y получаем следующую формулу для производной показательно-степенной функции:

$$y' = u(x)^{v(x)} \left[v'(x) \ln u(x) + v(x) \frac{u'(x)}{u(x)} \right]. \tag{4.31}$$

Рассмотрим в качестве примера вычисление производной для функции $y = x^x$.

Эту функцию можно представить в виде (4.30), где $u(x) = x$, $v(x) = x$. Подставляя эти выражения в формулу (4.31), получаем

$$y' = x^x \left(1 \cdot \ln x + x \frac{1}{x} \right) = x^x (\ln x + 1).$$

Производная степенной функции с любым вещественным показателем

Выведем формулу производной для степенной функции $y = x^\alpha$, где α — любое вещественное число.

Прологарифмируем эту функцию: $\ln y = \alpha \ln x$. По формуле (4.29) находим

$$\frac{y'}{y} = [\alpha \ln x]' = \frac{\alpha}{x}.$$

Подставляя сюда $y = x^\alpha$, получаем искомую формулу:

$$y' = (x^\alpha)' = \alpha x^{\alpha - 1}. \tag{4.32}$$

4.10. Производные простейших элементарных функций

В предыдущих параграфах мы вычислили производные всех простейших элементарных функций; их можно свести в следующую таблицу.

I. $(C)' = 0$ где C — постоянное число.

II. $(x^\alpha)' = \alpha x^{\alpha - 1}$; в частности $\left(\frac{1}{x} \right)' = -\frac{1}{x^2}$, $(\sqrt{x})' = \frac{1}{2\sqrt{x}}$.

III. $(\log_a x)' = \frac{1}{x} \log_a e$; в частности $(\ln x)' = \frac{1}{x}$.

IV. $(a^x)' = a^x \ln a$, в частности $(e^x)' = e^x$.

V. $(\sin x)' = \cos x$.

VI. $(\cos x)' = -\sin x$.

VII. $(\operatorname{tg} x)' = \frac{1}{\cos^2 x}$.

VIII. $(\operatorname{ctg} x)' = -\frac{1}{\sin^2 x}$.

IX. $(\arcsin x)' = \frac{1}{\sqrt{1 - x^2}}$.

X. $(\arccos x)' = -\frac{1}{\sqrt{1 - x^2}}$.

XI. $(\operatorname{arctg} x)' = \frac{1}{1 + x^2}$.

XII. $(\operatorname{arcctg} x)' = -\frac{1}{1 + x^2}$.

Введем гиперболические функции, широко используемые в математике и прикладных науках:

$\operatorname{sh} x = \frac{e^x - e^{-x}}{2}$ — гиперболический синус;

$\operatorname{ch} x = \frac{e^x + e^{-x}}{2}$ — гиперболический косинус;

$\operatorname{th} x = \dfrac{\operatorname{sh} x}{\operatorname{ch} x}$ — гиперболический тангенс;

$\operatorname{cth} x = \dfrac{\operatorname{ch} x}{\operatorname{sh} x}$ — гиперболический котангенс.

Нетрудно показать, что эти функции непрерывны и дифференцируемы на всей числовой прямой ($x \neq 0$ для $\operatorname{cth} x$).

Из определения гиперболических функций элементарно вытекают следующие формулы для их производных, которыми мы дополним основную таблицу.

XIII. $(\operatorname{sh} x)' = \operatorname{ch} x$.

XIV. $(\operatorname{ch} x)' = \operatorname{sh} x$.

XV. $(\operatorname{th} x)' = \dfrac{1}{\operatorname{ch}^2 x}$.

XVI. $(\operatorname{cth} x)' = -\dfrac{1}{\operatorname{sh}^2 x}$.

Формулы, приведенные в таблице вместе с правилами дифференцирования (теорема 4.3), являются основными формулами дифференциального исчисления. Отсюда можно сделать важный вывод: поскольку производная любой элементарной функции также элементарная функция, то операция дифференцирования не выводит функцию из класса элементарных.

4.11. Производные и дифференциалы высших порядков

Понятие производной n-го порядка

Производная $f'(x)$ функции $f(x)$ сама является функцией аргумента x, и по отношению к ней также можно ставить вопрос о производной. Производная от первой производной некоторой функции $y = f(x)$ называется *второй производной*, или *производной второго порядка*, этой функции. Производная от второй производной называется *третьей производной*, или *производной третьего порядка*. Этот процесс можно продолжить. Производные начиная со второй называются *производными высших порядков*. Для их обозначения используют

символы: y'', y''', $y^{(4)}$, $y^{(5)}$, …, $y^{(n)}$ (для второй и третьей производных соответственно еще и $y^{(2)}$ и $y^{(3)}$) или вместо y пишут $f(x)$:

$$f''(x),\ f'''(x), \ldots, f^{(n)}(x).$$

Производная n-го порядка определяется, таким образом, как производная от производной $(n-1)$-го порядка: $y^{(n)} = \left(y^{(n-1)}\right)'$.

n-е производные некоторых функций

Укажем вид n-х производных для некоторых функций.

1) $y = x^\alpha$, где $x > 0$. Последовательное дифференцирование этой функции дает

$$y' = \alpha x^{\alpha-1}, y'' = \alpha(\alpha-1)x^{\alpha-2}, y''' = \alpha(\alpha-1)(\alpha-2)x^{\alpha-3}, \ldots,$$
$$y^{(n)} = \alpha(\alpha-1)(\alpha-2)\cdots(\alpha-n+1)x^{\alpha-n}.$$

Если $\alpha = m$, где m — натуральное число, получим

$$(x^m)^{(m)} = m!, (x^m)^{(n)} = 0 \text{ при } n > m.$$

2) $y = a^x$.

Последовательное дифференцирование приводит к формулам

$$y' = a^x \ln a, y'' = (a^x)' \ln a = a^x \ln^2 a, \ldots, y^{(n)} = a^x \ln^n a.$$

В частном случае, когда $a = e$, получаем $(e^x)^{(n)} = e^x$, т.е. дифференцирование любого порядка не меняет вида простой экспоненты.

3) $y = \sin x$.

Первая производная этой функции может быть представлена в следующем виде:

$$y' = \cos x = \sin\left(x + \frac{\pi}{2}\right).$$

Соответственно вторая производная

$$y'' = -\sin x = \sin(x + \pi).$$

Таким образом, каждое дифференцирование добавляет в аргумент синуса $\pi/2$, так что в общем случае получаем

$$y^{(n)} = \sin\left(x + n\frac{\pi}{2}\right).$$

4) $y = \cos x$.

Здесь налицо полная аналогия с предыдущей функцией, т.е.

$$y^{(n)} = \cos\left(x + n\frac{\pi}{2}\right).$$

Формула Лейбница для n-й производной произведения двух функций

Пусть $y = uv$, где u и v — функции, имеющие производные любого порядка. Последовательное дифференцирование функции y даст

$$y' = u'v + uv',\ y'' = u'' = u''v + u'v' + u'v' + uv'' =$$
$$= u''v + 2u'v' + uv'',\ y''' = u'''v + 3u''v' + 3u'v'' + uv'''.$$

Правые части этих равенств аналогичны разложению бинома Ньютона для случаев $n = 2$ и $n = 3$, только показатели степеней заменены на порядки производных, а сами функции u и v нужно рассматривать как производные нулевого порядка: $u^{(0)} = u$, $v^{(0)} = v$.

Методом математической индукции можно доказать справедливость формулы общего вида для n-й производной:

$$y^{(n)} = (uv)^{(n)} = u^{(n)}v + C_n^1 u^{(n-1)}v' + C_n^2 u^{(n-2)}v'' + \cdots + C_n^k u^{(n-k)}v^{(k)} + \cdots + uv^{(n)}, \tag{4.33}$$

где коэффициенты C_n^k — число сочетаний из n по k:

$$C_n^k = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!},$$
$$C_n^k = C_n^{(n-k)},$$

так что в правой части формулы (4.33) коэффициенты, равноотстоящие от левого и правого концов, равны.

Рассмотрим несколько примеров на применение формулы Лейбница.

Пример 1. Найти четвертую производную функции $y = x^3e^x$.

Положим $u = u^{(0)} = x^3$, $v = v^{(0)} = e^x$. Последовательно находим: $u' = 3x^2$, $u'' = 6x$, $u''' = 6$, $u^{(4)} = 0$; $v^{(k)} = e^x$. Подставляя эти выражения в формулу (4.33) при $n = 4$, окончательно получаем

$$y^{(4)} = e^x(24 + 36x + 24x^2 + x^3).$$

Пример 2. Найти пятую производную функции $y = e^{2x}\sin x$.

Полагая $u = u^{(0)} = e^{2x}$ и $v = v^{(0)} = \sin x$, находим: $u^{(n)} = 2^n e^{2x}$, $v^{(n)} = \sin\left(x + \frac{\pi}{2}\right)$. Подставляя эти выражения в разложение (4.33) при $n = 5$, получаем

$$y^{(5)} = e^{2x}(32\sin x + 16\cdot 5\cos x - 8\cdot 10\sin x - 4\cdot 10\cos x +$$
$$+2\cdot 5\sin x + \cos x) = e^{2x}(41\cos x - 38\sin x).$$

Пример 3. Найти третью производную функции $y = x^2\ln x$.

Последовательно выполняя ту же процедуру, что и в предыдущих примерах, получаем: $u = u^{(0)} = x^2$, $v = v^{(0)} = \ln x$; $u' = 2x$, $u'' = 2$, $u''' = 0$; $v' = \frac{1}{x}$, $v'' = -\frac{1}{x^2}$, $v''' = \frac{2}{x^3}$;

$$y''' = (x^2\ln x)''' = 3\cdot 2\frac{1}{x} + 3\cdot 2x\left(-\frac{1}{x^2}\right) + x^2\left(\frac{2}{x^3}\right) =$$
$$= \frac{6}{x} - \frac{6}{x} + \frac{2}{x} = \frac{2}{x}.$$

Дифференциалы высших порядков

Будем использовать для обозначения дифференциалов dy и dx также и символы δy и δx. Пусть функция $f(x)$ дифференцируема на некотором промежутке. Тогда, как было установлено в параграфе 4.3, ее дифференциал (4.8) $dy = f'(x)dx$, который называется *дифференциалом первого порядка*, или *первым дифференциалом*, является функцией двух переменных: аргумента x и дифференциала dx. Пусть функция $f'(x)$ также дифференцируема на указанном промежутке. Будем полагать dx

постоянным сомножителем в выражении для dy, тогда функция dy будет функцией только аргумента x и ее дифференциал в точке x выражается формулой

$$\delta(dy) = \delta[f'(x)\,dx] = [f'(x)\,dx]'\delta x = f''(x)\,dx\delta x.$$

Дифференциал $\delta(dy)$ в точке x при $\delta x = dx$ называется *дифференциалом второго порядка* функции $f(x)$ в точке x и обозначается d^2y:

$$d^2y = f''(x)(dx)^2.$$

Эту процедуру можно продолжить, если функция $f(x)$ многократно дифференцируема в каждой точке x некоторого интервала: дифференциал $\delta(d^2y)$ при $\delta x = dx$ называется *дифференциалом третьего порядка* в точке x и обозначается d^3y и т.д. Определим *дифференциал n-го порядка (n-й дифференциал)* функции $f(x)$ в точке x как дифференциал $\delta(d^{n-1}y)$ от дифференциала $(n-1)$-го порядка при $\delta x = dx$. Методом математической индукции можно доказать, что для n-го дифференциала верна формула

$$d^ny = y^{(n)}(x)(dx)^n. \tag{4.34}$$

Из формулы (4.34) следует выражение для производной n-го порядка функции $y = f(x)$:

$$y^n(x) = \frac{d^ny}{(dx)^n} = \frac{d^ny}{dx^n}. \tag{4.35}$$

Здесь следует заметить, что формулы (4.34) и (4.35) справедливы при $n > 1$ только для случая независимой переменной x, т.е. *дифференциалы высших порядков не обладают свойством инвариантности формы.*

Пример 4. Найти четвертый дифференциал функции $y = x^5 - x^3 - \ln x$.

Дифференцируя данную функцию последовательно, получаем: $y' = 5x^4 - 3x^2 - x^{-1}$, $y'' = 20x^3 - 6x + x^{-2}$, $y''' = 60x^2 - 6 - 2x^{-3}$, $y^{(4)} = 120x + 6x^{-4}$. Следовательно, $d^4y = (120x - 6x^{-4})\,dx^4$.

Дифференцирование функции, заданной параметрически

Пусть x и y заданы как функции одной независимой переменной t:

$$x = \varphi(t), \quad y = \psi(t), \tag{4.36}$$

причем они определены и непрерывны на одном и том же промежутке T. Если для функции x ее производная $\varphi'(t)$ отлична от нуля (значит, $\varphi(t)$ строго монотонна), то существует обратная функция $t = \gamma(x)$, которая также монотонна, однозначна и дифференцируема на множестве X значений функции $\psi(t)$ в силу теоремы об обратной функции. Поэтому y можно рассматривать как функцию, зависящую от аргумента x посредством параметра t:

$$y = \psi[\gamma(x)].$$

Рассмотрим пример. Пусть $x = r\cos t$, $y = r\sin t$, $0 \leqslant t \leqslant \pi$. Поскольку функция $\cos t$ убывает на указанном промежутке значений t, то эти равенства задают параметрически функцию y от x. Выразим t через x и подставим из первого уравнения во второе:

$$t = \arccos\frac{x}{r}, \quad y(x) = r\sin\left(\arccos\frac{x}{r}\right) =$$
$$= r\sqrt{1-(x/r)^2} = \sqrt{r^2-x^2}.$$

Последнее уравнение определяет верхнюю полуокружность радиуса r с центром в начале координат.

Теперь выведем формулу для производной функции, заданной параметрически. Пусть функция $\psi(t)$ также имеет производную на том же промежутке, что и $\varphi(t)$. Поскольку $\varphi'(t) \neq 0$, то в силу инвариантности формы первого дифференциала производную функции $y(x)$ можно записать в виде (4.9):

$$y'(x) = \frac{dy}{dx}, \quad dy = \psi'(t)\,dt, \quad dx = \varphi'(t)\,dt,$$

а из этих формул получаем уже выражение для первой производной:

$$y'(x) = \frac{\psi'(t)}{\varphi'(t)}. \tag{4.37}$$

Если для функций $\varphi(t)$ и $\psi(t)$ существуют вторые производные в некоторой точке t, то вторую производную функции $y(x)$, в свою очередь, можно рассматривать как производную по x первой производной (4.37), заданной параметрически. Следовательно, по аналогии с формулой (4.37)

$$y''(x) = \frac{\left(\dfrac{\psi'(t)}{\varphi'(t)}\right)'}{\varphi'(t)}.$$

После дифференцирования числителя как частного двух функций окончательно получаем

$$y''(x) = \frac{\psi''(t)\varphi'(t) - \varphi''(t)\psi'(t)}{[\varphi'(t)]^3}. \qquad (4.38)$$

Пример 5. Найти первую и вторую производные функции, заданной параметрически: $x = \cos t$, $y = \sin t$, $0 \leqslant t \leqslant \pi$.

Подстановка в формулу (4.37) дает первую производную:

$$y'(x) = -\frac{\cos t}{\sin t} = -\frac{\cos(\arccos x)}{\sin(\arccos x)} = -\frac{x}{\sqrt{1-x^2}}.$$

Подстановку в формулу (4.38) приводит к выражению

$$y''(x) = \frac{-(\operatorname{ctg} t)'}{\sin t} = -\frac{1}{\sin^3 t} = -\frac{1}{[\sin(\arccos x)]^3} = -\frac{1}{(1-x^2)^{3/2}}.$$

Глава 5
ПРИМЕНЕНИЕ ПРОИЗВОДНЫХ В ИССЛЕДОВАНИИ ФУНКЦИЙ

5.1. Основные теоремы дифференциального исчисления

Теорема 5.1 (ТЕОРЕМА ФЕРМА[1]). *Пусть функция $f(x)$ определена на интервале (a,b) и имеет наибольшее (наименьшее) значение в точке $x_0 \in (a,b)$. Тогда, если в точке x_0 существует производная этой функции, она равна нулю, т.е. $f'(x_0) = 0$.*

[1] Пьер Ферма (1601–1655) — французский математик.

Доказательство. Ограничимся случаем наибольшего значения функции в точке x_0 (для наименьшего значения доказательство аналогично). Тогда для всех $x \in (a,b)$ выполняется неравенство $f(x) \leqslant f(x_0)$, что означает $\Delta y = f(x_0 + \Delta x) - f(x_0) \leqslant 0$ для любой точки $x_0 + \Delta x \in (a,b)$. Если $x > x_0$ ($\Delta x > 0$), то $\Delta y/\Delta x \leqslant 0$, т.е.

$$\lim_{\Delta x \to 0+} \frac{\Delta y}{\Delta x} = f'_+(x_0) \leqslant 0.$$

если $x < x_0$ ($\Delta x < 0$), то $\Delta y/\Delta x \geqslant 0$, т.е.

$$\lim_{\Delta x \to 0-} \frac{\Delta y}{\Delta x} = f'_-(x_0) \geqslant 0.$$

Таким образом, производная в точке x_0 слева неотрицательна, а справа неположительна, но поскольку по условию теоремы в самой точке x_0 производная существует, то $f'(x_0) = f'_+(x_0) = f'_-(x_0)$. Последнее равенство возможно лишь в случае $f'_-(x_0) = f'_-(x_0) = 0$, но тогда и $f'(x_0) = 0$, что и требовалось доказать.

Геометрический смысл теоремы Ферма показан на рис. 5.1: если в точке x_0 дифференцируемая функция принимает наибольшее (наименьшее) значение, то в точке $M(x_0, f(x_0))$ касательная к графику этой функции параллельна оси Ox.

Заметим, что теорема неверна, если функция рассматривается на отрезке $[a,b]$: в этом случае она может принимать наибольшее и наименьшее значения на концах отрезка, где производная не равна нулю.

Теорема 5.2 (ТЕОРЕМА РОЛЛЯ[1]). *Пусть функция $f(x)$ непрерывна на отрезке $[a,b]$ и дифференцируема на интервале (a,b), причем $f(a) = f(b)$. Тогда существует точка $c \in (a,b)$, в которой $f'(c) = 0$.*

Доказательство. Поскольку функция $f(x)$ непрерывна на $[a,b]$, то в силу второй теоремы Вейерштрасса (см. теорему 3.14) она достигает на этом отрезке своего максимального M и минимального m значения, т.е. существуют такие точки x_1, $x_2 \in [a,b]$, что $f(x_1) = M$, $f(x_2) = m$. Тогда возможны два случая: $M = m$ и $M > m$.

[1] Мишель Ролль (1652–1719) — французский математик.

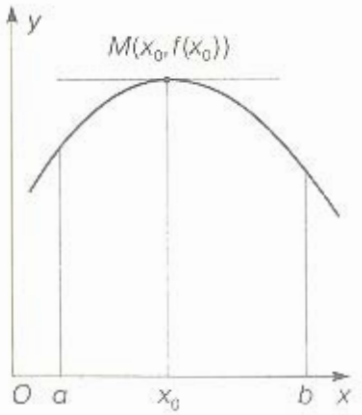

Рис. 5.1

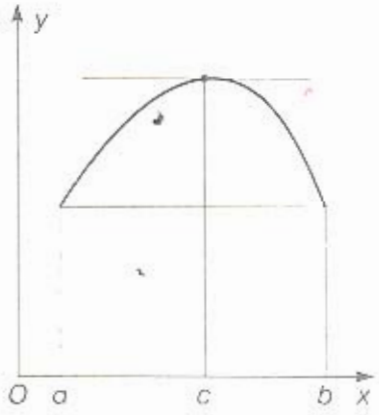

Рис. 5.2

В первом случае $f(x) = M = m = \text{const}$; тогда $f'(x) = 0$ везде на $[a, b]$ и теорема доказана.

Во втором случае хотя бы одно из двух значений M или m не принимается на концах отрезка $[a, b]$, поскольку $f(a) = f(b)$. Значит, существует точка $c \in (a, b)$, в которой функция $f(x)$ принимает наибольшее или наименьшее значение; в этом случае выполняются условия теоремы Ферма, следовательно, $f'(c) = 0$. Теорема доказана.

Геометрический смысл теоремы Ролля очевиден из рис. 5.2: если функция, непрерывная на отрезке и дифференцируемая внутри его, на концах этого отрезка принимает одинаковые значения, то хотя бы в одной внутренней точке этого отрезка касательная к графику функции параллельна оси Ox.

Заметим, что все условия теоремы Ролля существенны. Рассмотрим несколько примеров с нарушениями условий теоремы.

1. Функция $f(x) = x^2$ непрерывна на отрезке $[0, 1]$ и дифференцируема на интервале $(0, 1)$, однако на концах отрезка принимает разные значения 0 и 1; для нее не существует внутренней точки c, в которой $f'(c) = 0$ (рис. 5.3, *а*) и касательная была бы параллельна оси Ox.

2. Функция $f(x) = |x|$ непрерывна на отрезке $[-1, 1]$ и принимает на его концах равные значения (рис. 5.3, *б*); однако эта функция не имеет производной во всех точках интервала

$(-1, 1)$, в частности при $x = 0$, поэтому нет такой внутренней точки c, в которой производная обращалась бы в нуль.

3. Функция, определяемая формулой

$$f(x) = \begin{cases} 1 - x, & x \in (0, 1] \\ 0, & x = 0, \end{cases}$$

дифференцируема на интервале $(0, 1)$ и принимает на концах отрезка $[0, 1]$ одинаковые значения. Однако эта функция не является непрерывной на отрезке $[0, 1]$, поэтому не существует такой внутренней точки c, в которой $f'(c) = 0$ (рис. 5.3, *в*).

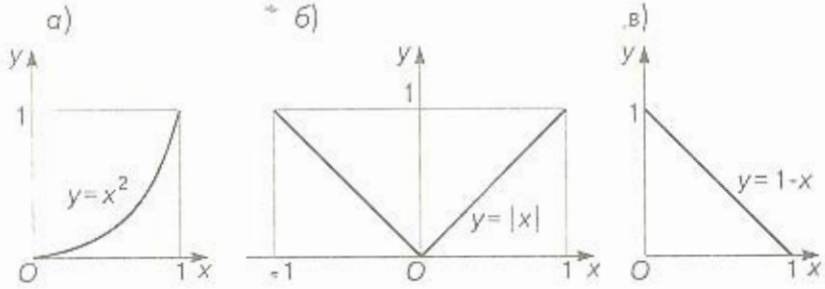

Рис. 5.3

Теорема 5.3 (ТЕОРЕМА ЛАГРАНЖА[1]). *Пусть функция $f(x)$ непрерывна на $[a, b]$ и дифференцируема на (a, b). Тогда существует такая точка $c \in (a, b)$, что справедлива формула*

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

Доказательство. Введем вспомогательную функцию

$$F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

Функция $F(x)$ удовлетворяет всем условиям теоремы Ролля: она непрерывна на $[a, b]$ (как разность непрерывных функций), дифференцируема на (a, b) $(F'(x) = f'(x) - [f(b) - f(a)]/(b - a))$ и принимает на концах отрезка $[a, b]$ одинаковые значения

[1] Жозеф Луи Лагранж (1736–1813) — французский математик.

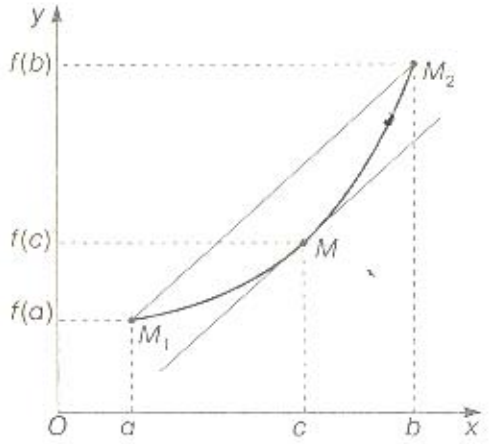

Рис. 5.4

$(F(a) = F(b) = 0)$. Тогда по теореме Ролля существует хотя бы одна такая точка $c \in (a, b)$, что $F'(c) = 0$. Но тогда $f'(c) - \dfrac{f(b) - f(a)}{b - a} = 0$, откуда и получаем утверждение теоремы.

Теорема Лагранжа имеет геометрический смысл (рис. 5.4). Секущая, проходящая через точки $M_1\ (a, f(a))$ и $M_2\ (b, f(b))$, имеет угловой коэффициент, равный $[f(b) - f(a)]/(b - a)$, а $f'(c)$ — угловой коэффициент касательной к графику функции в точке $M\ (c, f(c))$. Теорема Лагранжа утверждает, что существует хотя бы одна точка интервала (a, b), где касательная к графику функции параллельна секущей M_1M_2.

Отметим, что вытекающее из теоремы Лагранжа равенство

$$f(b) - f(a) = f'(c)(b - a) \tag{5.1}$$

называется *формулой Лагранжа*, или *формулой конечных приращений*.

Теорема 5.4 (ТЕОРЕМА КОШИ). *Пусть функции $f(x)$ и $g(x)$ непрерывны на $[a, b]$ и дифференцируемы на (a, b), причем $g'(x) \neq 0$. Тогда существует такая точка $c \in (a, b)$, что справедлива формула*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}. \tag{5.2}$$

Доказательство. Поскольку $g'(x) \neq 0$ на (a, b), то $g(b) \neq g(a)$; действительно, в противном случае для функции $g(x)$ были бы выполнены условия теоремы Ролля и тогда $g'(c) = 0$ в некоторой точке $c \in (a, b)$, что противоречит условию доказываемой теоремы.

Рассмотрим вспомогательную функцию на $[a, b]$:

$$F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}[g(x) - g(a)].$$

Эта функция на $[a, b]$ удовлетворяет условиям теоремы Ролля: она непрерывна на $[a, b]$, дифференцируема на (a, b) и $F(a) = F(b) = 0$. Стало быть, существует такая точка $c \in (a, b)$, что $F'(c) = 0$. Подставляя это условие в выражение для $F'(x)$ при $x = c$, получаем, что

$$f'(c) - \frac{f(b) - f(a)}{g(b) - g(a)} g'(c) = 0.$$

Поскольку $g'(c) \neq 0$, отсюда получаем формулу (5.2), что и доказывает теорему.

Формула (5.2) называется *формулой Коши*, или *обобщенной формулой конечных приращений*. Формула Лагранжа получается из формулы (5.2) как частный случай при $g(x) = x$.

5.2. Раскрытие неопределенностей

Правило Лопиталя

Будем говорить, что отношение двух функций $\dfrac{f(x)}{g(x)}$ при $x \to a$ есть *неопределенность вида* $\dfrac{0}{0}$, если

$$\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0.$$

Раскрыть эту неопределенность означает вычислить предел $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)}$, если он существует.

Теорема 5.5 (ТЕОРЕМА ЛОПИТАЛЯ[1]). *Пусть функции $f(x)$ и $g(x)$ определены и дифференцируемы в некоторой окрестности точки a, за исключением, быть может, самой точки a. Кроме того, пусть также $\lim\limits_{x \to a} f(x) = \lim\limits_{x \to a} g(x) = 0$, при-*

[1] Гийом Франсуа Лопиталь (1661–1704) — французский математик.

чем $g'(x) \neq 0$ в указанной окрестности точки a. Тогда, если существует предел отношения $\lim\limits_{x\to a}\dfrac{f'(x)}{g'(x)}$ — (конечный или бесконечный), существует и предел $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)}$, причем справедлива формула:

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \lim_{x\to a}\frac{f'(x)}{g'(x)}. \tag{5.3}$$

Доказательство. Возьмем произвольную последовательность $\{x_n\}$ значений аргумента, сходящуюся к точке a, причем $x_n \neq a$. Доопределим функции $f(x)$ и $g(x)$ в точке a: $f(a) = g(a) = 0$. Тогда для этих функций выполнены все условия теоремы Коши: они непрерывны на $[a, x_n]$, дифференцируемы на (a, x_n) и $g'(x) \neq 0$. Значит, внутри отрезка $[a, x_n]$ существует такая точка c_n, что

$$\frac{f(x_n) - f(a)}{g(x_n) - g(a)} = \frac{f'(c_n)}{g'(c_n)}, \quad c \in (a, x_n).$$

Так как по доопределению $f(a) = g(a) = 0$, то

$$\frac{f(x_n)}{g(x_n)} = \frac{f'(c_n)}{g'(c_n)}, \quad c_n \in (a, x_n). \tag{5.4}$$

Пусть теперь в этой формуле $n \to \infty$, тогда $c_n \to a$ в силу теоремы 2.12 о предельном переходе в неравенствах. Поскольку предел правой части в последней формуле существует, то существует и предел левой части при $n \to \infty$, а значит, по определению предела функции в точке, этот предел равен $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)}$. Отсюда следует выполнение равенства (5.3), что и требовалось доказать.

Эту теорему обычно называют *правилом Лопиталя*.

Замечание 1. Правило Лопиталя можно применять повторно, если $f'(x)$ и $g'(x)$ удовлетворяют тем же требованиям, что и исходные функции $f(x)$ и $g(x)$.

Замечание 2. Теорема остается верной и в случае, когда $x \to \infty$ $(x \to \pm\infty)$.

Теперь рассмотрим примеры раскрытия неопределенностей.

Пример 1. $\lim\limits_{x\to 0}\dfrac{1-\cos x}{2x^2} = \lim\limits_{x\to 0}\dfrac{\sin x}{4x} = \lim\limits_{x\to 0}\dfrac{\cos x}{4} = \dfrac{1}{4}$. Здесь мы дважды последовательно применили правило Лопиталя, поскольку два раза имели дело с неопределенностью типа $\dfrac{0}{0}$.

Пример 2. $\lim\limits_{x\to 0}\dfrac{e^{2x}-1}{x} = \lim\limits_{x\to 0}\dfrac{2e^{2x}}{1} = 2.$

Пример 3. $\lim\limits_{x\to a}\dfrac{x^a - a^x}{x-a} = \lim\limits_{x\to a}\dfrac{ax^{a-1} - a^x\ln a}{1} = a^a(1-\ln a).$

Неопределенности вида $\frac{\infty}{\infty}$

Будем называть отношение двух функций $\dfrac{f(x)}{g(x)}$ при $x \to a$ *неопределенностью вида* $\dfrac{\infty}{\infty}$, если $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = \infty$, $-\infty$ или $+\infty$. В этом случае правило Лопиталя остается справедливым при замене условия $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = 0$ на условие $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = \infty$.

Пример 4. $\lim\limits_{x\to\infty}\dfrac{\ln x}{x} = \lim\limits_{x\to\infty}\dfrac{1/x}{1} = \lim\limits_{x\to\infty}\dfrac{1}{x} = 0.$

Пример 5. $\lim\limits_{x\to 1}\dfrac{\ln(x-1)}{\operatorname{ctg}\pi x} = \lim\limits_{x\to 1}\dfrac{1/(x-1)}{-\pi\sin^{-2}\pi x} = -\dfrac{1}{\pi}\lim\limits_{x\to 1}\dfrac{\sin^2\pi x}{x-1} =$
$= -\lim\limits_{x\to 1}\dfrac{\sin 2\pi x}{1} = 0.$

Другие виды неопределенностей

Неопределенности вида $0\cdot\infty$ и $\infty - \infty$ можно свести к неопределенностям вида $\dfrac{0}{0}$ и $\dfrac{\infty}{\infty}$. Покажем это на примерах.

Пример 6. Найти предел $\lim\limits_{x\to 0-} x\ln x$.

Здесь неопределенность вида $0\cdot\infty$. Преобразуем функцию под знаком предела: $x\ln x = \dfrac{\ln x}{1/x}$, в результате имеем неопределенность вида $\dfrac{\infty}{\infty}$ при $x \to 0+$. Теперь, применяя правило

Лопиталя, получаем

$$\lim_{x\to 0+} x\ln x = \lim_{x\to 0+}\frac{(\ln x)'}{(1/x)'} = \lim_{x\to 0+}\frac{1/x}{-1/x^2} = -\lim_{x\to 0-} x = 0.$$

Пример 7. Найти $\lim\limits_{x\to 0}(\operatorname{cosec} x - \operatorname{ctg} x)$.

Это неопределенность вида $\infty - \infty$. Преобразуя функцию под знаком предела, получаем: $\operatorname{cosec} x - \operatorname{ctg} x = \dfrac{1}{\sin x} - \dfrac{\cos x}{\sin x} = \dfrac{1-\cos x}{\sin x}$. Теперь это неопределенность вида $\dfrac{0}{0}$ при $x \to 0$. Правило Лопиталя дает нам

$$\lim_{x\to 0}(\operatorname{cosec} x - \operatorname{ctg} x) = \lim_{x\to 0}\frac{1-\cos x}{\sin x} = \lim_{x\to 0}\frac{\sin x}{\cos x} = 0.$$

Неопределенности вида 0^0, 1^∞, ∞^0, имеющие место при рассмотрении пределов функций $y = u(x)^{v(x)}$, сводятся к неопределенности вида $0\cdot\infty$ с помощью тождественного преобразования

$$g(x)^{v(x)} = e^{v(x)\ln g(x)}. \tag{5.5}$$

Пример 8. Найти предел $\lim\limits_{x\to 0+} x^x$.

Это неопределенность вида 0^0; используя формулу (5.5), имеем с учетом решения первого примера

$$\lim_{x\to 0+} x^x = \lim_{x\to 0+} e^{x\ln x} = e^0 = 1.$$

Пример 9. Найти предел $\lim\limits_{x\to 0}(1+x)^{\operatorname{ctg} x}$.

Это неопределенность вида 1^∞. Найдем предел функции $y = \ln(1+x)\operatorname{ctg} x$ при $x \to 0$. В соответствии с представлением (5.5)

$$\lim_{x\to 0} y = \lim_{x\to 0}\ln(1+x)\operatorname{ctg} x = \lim_{x\to 0}\frac{\ln(1+x)}{\operatorname{tg} x} =$$
$$= \lim_{x\to 0}\frac{1/(1+x)}{1/\cos^2 x} = \lim_{x\to 0}\frac{\cos^2 x}{1+x} = 1.$$

Следовательно, искомый предел равен

$$\lim_{x\to 0} e^y = e^1 = e.$$

Пример 10. Найти предел $\lim\limits_{x\to 0}(\operatorname{ctg} x)^{\sin x}$.

Это неопределенность вида ∞^0. Используя формулу (5.5), представим функцию под знаком предела в виде

$$(\operatorname{ctg} x)^{\sin x} = e^{\sin x\ln(\operatorname{ctg} x)}$$

и найдем предел показателя степени экспоненты, предварительно преобразовав его:

$$\lim_{x\to 0}[\sin x\ln(\operatorname{ctg} x)] = \lim_{x\to 0}\frac{\ln(\operatorname{ctg} x)}{1/\sin x}.$$

Мы получили неопределенность вида $\dfrac{\infty}{\infty}$ и можем применить правило Лопиталя:

$$\lim_{x\to 0}[\sin x\ln(\operatorname{ctg} x)] = \lim_{x\to 0}\frac{\operatorname{tg} x(-1/\sin^2 x)}{(-1/\sin^2 x)\cos x} = \lim_{x\to 0}\frac{\sin x}{\cos^2 x} = 0.$$

Отсюда окончательно получаем

$$\lim_{x\to 0}(\operatorname{ctg} x)^{\sin x} = e^0 = 1.$$

5.3. Формула Тейлора

Выводимая в этом параграфе формула является одной из основных в математике и имеет широкое применение в смежных науках.

Разложение функций по формуле Тейлора

Теорема 5.6 (ТЕОРЕМА ТЕЙЛОРА[1]). *Пусть функция $f(x)$ имеет в точке $x = a$ и некоторой ее окрестности производные вплоть до порядка $n+1$. Тогда между точками a и*

[1] Брук Тейлор (1685–1731) — английский математик.

$x \neq a$ найдется такая точка ξ, что справедлива следующая формула:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}. \tag{5.6}$$

Доказательство. Обозначим через $\varphi(x,a)$ многочлен степени n относительно x в правой части формулы (5.6):

$$\varphi(x,a) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Он называется *многочленам Тейлора* степени n для функции $f(x)$.

Обозначим через $R_{n-1}(x)$ разность

$$R_{n+1}(x) = f(x) - \varphi(x,a).$$

Для доказательства теоремы следует получить формулу

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}.$$

Фиксируем любое значение x из указанной окрестности точки a. Пусть для определенности $x > a$. Обозначим через t переменную величину $a \leqslant t \leqslant x$ и рассмотрим на отрезке $[a,x]$ вспомогательную функцию

$$F(t) = f(x) - \varphi(x,t) - \frac{(x-t)^{n+1}R_{n+1}(x)}{(x-a)^{n+1}}. \tag{5.7}$$

Функция $F(t)$ удовлетворяет на $[a,x]$ всем условиям теоремы Ролля:

а) из формулы (5.7) и условий теоремы на функцию $f(x)$ следует, что $F(t)$ непрерывна и дифференцируема на $[a,x]$;

б) $F(a) = f(x) - \varphi(x,a) - R_{n+1}(x) = R_{n+1}(x) - R_{n+1}(x) = 0$,

$$F(x) = f(x) - \varphi(x,x) - \frac{(x-x)^{n+1}R_{n+1}(x)}{(x-a)^{n+1}} = 0,$$

так как $\varphi(x,x) = f(x)$. Таким образом, на концах отрезка $[a,x]$ функция $F(t)$ принимает одинаковые значения, и в силу теоремы Ролля существует такая точка $\xi (a < \xi < x)$, что

$$F'(\xi) = 0. \tag{5.8}$$

Найдем производную $F'(t)$, для чего дифференцируем равенство (5.7):

$$F'(t) = -f'(t) + \frac{f'(t)}{1!} - \frac{f''(t)}{1!}(x-t) + \frac{2f''(t)}{2!}(x-t)^2 - \cdots + \frac{f^{(n)}(t)}{n!}n(x-t)^{n-1} - \frac{f^{(n+1)}(t)}{n!}(x-t)^n + \frac{(n+1)(x-t)^n R_{n+1}(x)}{(x-a)^{n+1}}.$$

Легко видеть, что все слагаемые в правой части этого равенства, кроме двух последних, взаимно уничтожаются. Следовательно,

$$F'(t) = -\frac{f^{(n+1)}(t)}{n!}(x-t)^n + \frac{(n+1)(x-t)^n}{(x-a)^{n+1}}R_{n+1}(x). \tag{5.9}$$

Наконец, при $t = \xi$, согласно условию (5.8), получаем выражение для $R_{n+1}(x)$:

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}. \tag{5.10}$$

Теорема Тейлора доказана.

Формула (5.6) называется *формулой Тейлора*, а выражение (5.10) представляет *остаточный член в форме Лагранжа*. Так как точка $\xi \in (a,x)$, то $\xi = a + \theta(x-a)$, где $0 < \theta < 1$; тогда остаточный член можно переписать в виде

$$R_{n+1}(x) = \frac{f^{(n+1)}[a+\theta(x-a)]}{(n+1)!}(x-a)^{n+1}, \quad 0 < \theta < 1.$$

Предположим, что функция $f^{(n+1)}(x)$ ограничена в окрестности точки a, тогда остаточный член $R_{n-1}(x)$ является бесконечно малой при $x \to a$ более высокого порядка, чем $(x-a)^n$.

Действительно, в этом случае в силу теоремы 3.8 о произведении бесконечно малой функции на ограниченную функцию

$$\lim_{x\to a}\frac{R_{n-1}(x)}{(x-a)^n}=\lim_{x\to a}\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)=0,$$

так как первый сомножитель под знаком предела ограничен, а второй — бесконечно малая функция при $x\to a$. Таким образом, остаточный член $R_{n+1}(x)$ можно переписать в форме

$$R_{n+1}(x)=o[(x-a)^n] \text{ при } x\to a. \qquad (5.11)$$

Формула (5.11) представляет остаточный член $R_{n+1}(x)$ в форме Пеано[1].

Формула Маклорена

Формулой Маклорена[2] называется формула Тейлора (5.6) при $a=0$:

$$f(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\ldots+\frac{f^{(n)}(0)}{n!}x^n+R_{n+1}(x). \qquad (5.12)$$

Здесь остаточный член имеет вид:

1) в форме Лагранжа

$$R_{n+1}(x)=\frac{f^{n+1}(\theta x)}{(n+1)!}x^{n+1},\quad 0<\theta<1;$$

2) в форме Пеано

$$R_{n+1}(x)=o(x^n).$$

Разложение по формуле Маклорена некоторых элементарных функций

Один из основных принципов математики — представление сложного через более простое. Формулы Тейлора и Маклорена как раз и являются реализацией этого принципа. Любые функции, удовлетворяющие условиям теоремы Тейлора, с достаточной степенью точности могут быть представлены в виде многочлена n-й степени. Многочлены же являются наиболее простыми элементарными функциями, над которыми удобно выполнять арифметические действия, вычислять значения в любой точке и т.д.

Формулы (5.6) и (5.12) дают возможность разложить функцию $f(x)$ по формуле Тейлора (в окрестности точки a) и по формуле Маклорена (в окрестности нуля) или, что то же самое, представить $f(x)$ в виде многочленов, коэффициенты которых вычисляются достаточно просто. Эти формулы широко используются и для приближенных вычислений значений различных функций; при этом погрешность вычислений оценивается по остаточному члену в форме Лагранжа.

Выведем разложения некоторых элементарных функций по формуле Маклорена с остаточным членом в форме Пеано.

1. $f(x)=e^x$.

Поскольку $e^{(n)}(x)=e^x$ $f^{(n)}(0)=e^0=1$ для любого n, формула Маклорена (5.12) имеет вид

$$e^x=1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots+\frac{x^n}{n!}+o(x^n). \qquad (5.13)$$

Формула (5.13) используется для вычисления числа e с любой необходимой точностью. Запишем ее с остаточным членом в форме Лагранжа при $x=1$:

$$e=1+1+\frac{1}{2!}+\frac{1}{3!}+\frac{1}{4!}+\cdots+\frac{1}{n!}+\frac{\exp(\theta)}{(n+1)!},\quad 0<\theta<1.$$

Отбрасывая последнее слагаемое, получаем приближенное значение числа $e\approx 2{,}7182818\ldots$, причем погрешность расчета оценивается по формуле

$$R_{n+1}(1)=\frac{\exp(\theta)}{(n+1)!}<\frac{3}{(n+1)!},\quad 0<\theta<1.$$

Например, при $n=7$ погрешность вычисления числа e составит менее $7{,}44\cdot 10^{-5}$, или менее 0,01%.

[1] Джузеппе Пеано (1858–1932) — итальянский математик.

[2] Колин Маклорен (1698–1746) — шотландский математик.

2. $f(x) = \sin x$.

Так как $f^{(n)}(x) = \sin\left(x + n\frac{\pi}{2}\right)$, то

$$f^{(n)}(0) = \sin\left(n\frac{\pi}{2}\right) = \begin{cases} 0 & \text{при } n \text{ четном,} \\ (-1)^{(n-1)/2} & \text{при } n \text{ нечетном.} \end{cases}$$

Подстановка в формулу (5.12) приводит к выражению

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots + (-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!} + o(x^{2n}). \quad (5.14)$$

3. $f(x) = \cos x$.

Поскольку $f^{(n)}(x) = \cos\left(x + n\frac{\pi}{2}\right)$, то

$$f^{(n)}(0) = \cos\left(n\frac{\pi}{2}\right) = \begin{cases} 0 & \text{при } n \text{ нечетном,} \\ (-1)^{(n/2)} & \text{при } n \text{ четном.} \end{cases}$$

Подстановка в формулу (5.12) приводит к разложению по формуле Маклорена:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n\frac{x^{2n}}{(2n)!} + o(x^{2n}). \quad (5.15)$$

4. $f(x) = \ln(1+x)$.

Так как $f(0) = 0$, $f^{(n)}(x) = (-1)^{n-1}\dfrac{(n-1)!}{(1+x)^n}$, то $f^{(n)}(0) = (-1)^{n-1}(n-1)!$; подстановка в формулу (5.12) приводит к разложению функции $\ln(1+x)$ по формуле Маклорена (при этом полагается, что 0! равняется 1):

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n-1}\frac{x^n}{n} + o(x^n). \quad (5.16)$$

5. $f(x) = (1+x)^\alpha$, где α — вещественное число.

Так как $f^{(n)}(x) = \alpha(\alpha-1)(\alpha-2)\cdots(\alpha-n+1)(1+x)^{\alpha-n}$, т.е. $f^{(n)}(0) = \alpha(\alpha-1)\cdots(\alpha-n+1)$, формула Маклорена для данной функции имеет вид

$$(1+x)^\alpha = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots +$$

$$+ \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n + o(x^n). \quad (5.17)$$

В частном случае, когда $\alpha = n$ — целое число, $f^{(n+1)}(x) \equiv 0$ и формула (5.17) переходит в *бином Ньютона:*

$$(1+x)^n = 1 + \frac{n}{1!}x + \frac{n(n-1)}{2!}x^2 + \cdots + x^n, \quad (5.18)$$

т.е. бином Ньютона является частным случаем формулы Маклорена для этой функции.

6. $f(x) = \operatorname{arctg} x$.

Путем последовательного дифференцирования этой функции можно установить, что

$$f^{(n)}(0) = \begin{cases} 0 & \text{при } n \text{ четном,} \\ (-1)^{(n-1)/2}(n-1)! & \text{при } n \text{ нечетном.} \end{cases}$$

Тогда формула Маклорена с остаточным членом в форме Пеано имеет вид

$$\operatorname{arctg} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \ldots + (-1)^{n-1}\frac{x^{2n-1}}{2n-1} + o(x^n). \quad (5.19)$$

Формула Маклорена в асимптотических формулах и вычислениях пределов функций

Формулы (5.13)–(5.17) и (5.19) представляют собой асимптотические формулы (или оценки) соответственно для функций e^x, $\sin x$, $\cos x$, $\ln(1+x)$, $(1+x)^\alpha$ и $\operatorname{arctg} x$ при $x \to 0$. Указанные асимптотические формулы эффективно используются при вычислении пределов функций. Покажем это на примерах.

Пример 1. Найти $\lim\limits_{x\to 0}\dfrac{\sin x - x}{x^3}$.

Применяя формулу (5.14) при $n = 2$, получаем

$$\lim_{x\to 0}\frac{x - x^3/3! + o(x^4) - x}{x^3} = \lim_{x\to 0}\left(-\frac{1}{3!} + \frac{o(x^4)}{x^3}\right) = -\frac{1}{6}.$$

Пример 2. Найти $\lim\limits_{x\to 0}\dfrac{\cos x - \exp(-x^2/2)}{x^4}$.

Используем формулы (5.13) при $n = 4$ и (5.15) при $n = 2$. Получаем при подстановке в выражение под знаком предела

$$\lim_{x\to 0}\left[\frac{1}{x^4}\left(1-\frac{x^2}{2}+\frac{x^4}{24}-1+\frac{x^2}{2}-\frac{x^4}{8}+o(x^4)\right)\right]=$$

$$=\lim_{x\to 0}\left(-\frac{1}{12}+\frac{o(x^4)}{x^4}\right)=-\frac{1}{12}.$$

5.4. Исследование функций и построение графиков

Признак монотонности функции

Теорема 5.7. *Если функция $f(x)$ дифференцируема и $f'(x) \geqslant 0$ ($f'(x) \leqslant 0$) на интервале (a, b), то она не убывает (не возрастает) на этом интервале.*

Доказательство. Пусть для определенности $f'(x) \geqslant 0$. Возьмем две произвольные точки x_1 и x_2 на (a, b), причем $x_1 < x_2$. Тогда на отрезке $[x_1, x_2]$ для функции $f(x)$ выполняются все условия теоремы Лагранжа (см. теорему 5.3), в силу которой

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1),\ c \in (a, b).$$

По условию теоремы $f'(c) \geqslant 0$ и, кроме того, $x_2 - x_1 > 0$, значит, $f(x_2) \geqslant f(x_1)$ при $x_2 > x_1$, т.е. функция $f(x)$ не убывает на (a, b), что и требовалось доказать.

Случай $f'(x) \leqslant 0$ доказывается аналогично.

Признак строгой монотонности функции при $f'(x) > 0$ ($f'(x) < 0$) доказывается таким же образом.

Точки локального экстремума

Определение 1. Точка x_0 называется *точкой локального максимума* (*минимума*) функции $f(x)$, если для любого $x \neq x_0$ в некоторой окрестности точки x_0 выполнено неравенство $f(x_0) > f(x)$ ($f(x_0) < f(x)$).

Локальный минимум и локальный максимум объединены общим названием *локальный экстремум*.

Теорема 5.8 (НЕОБХОДИМОЕ УСЛОВИЕ СУЩЕСТВОВАНИЯ ЛОКАЛЬНОГО ЭКСТРЕМУМА). *Если функция $f(x)$ дифференцируема в точке x_0 и имеет в этой точке локальный экстремум, то $f'(x_0) = 0$.*

Доказательство. По определению локального экстремума существует такой интервал $(x_0 - \delta, x_0 + \delta)$, в котором значение функции $f(x_0)$ является наибольшим или наименьшим. Но тогда в силу теоремы Ферма (см. теорему 5.1) производная функции в этой точке равна нулю, т.е. $f'(x_0) = 0$, что и требовалось доказать.

Геометрический смысл теоремы 5.8 показан на рис. 5.5: если в точках локальных экстремумов существуют касательные, то они параллельны оси Ox.

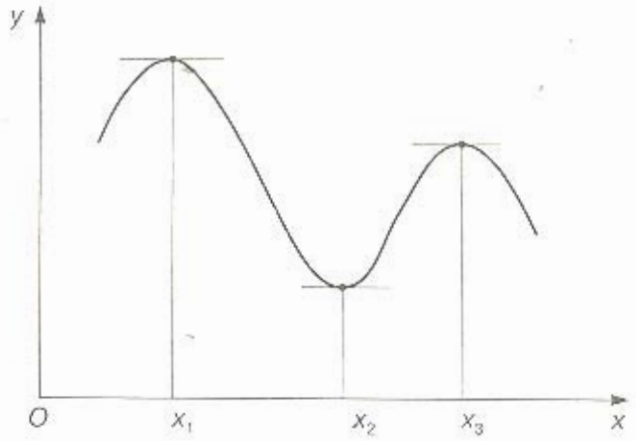

Рис. 5.5

Точки, в которых касательные параллельны оси Ox, а значит, производная равна нулю, называют *точками возможного экстремума*, или *стационарными точками*. Если x_0 — точка возможного экстремума, т.е. $f'(x_0) = 0$, то она может и не быть точкой локального экстремума. Например, для функции $f(x) = x^3$ (см. рис. 3.1) производная при $x = 0$ равна нулю, однако в этой точке нет локального экстремума. Таким образом, теорема 5.8 не является достаточным условием существования локального экстремума.

Теорема 5.9 (ДОСТАТОЧНОЕ УСЛОВИЕ СУЩЕСТВОВАНИЯ ЛОКАЛЬНОГО ЭКСТРЕМУМА). *Пусть функция $f(x)$ дифференцируема в некоторой окрестности точки x_0. Если при переходе через точку x_0 слева направо производная $f'(x)$ меняет знак с плюса на минус (с минуса на плюс), то в точке x_0 функция*

$f(x)$ имеет локальный максимум (минимум). Если же $f'(x)$ не меняет знака в δ-окрестности точки x_0, то данная функция не имеет локального экстремума в точке x_0.

Доказательство. Пусть $f(x)$ при переходе через точку x_0 меняет знак с "+" на "−" в интервале $(x_0-\delta, x_0+\delta)$. Возьмем произвольную точку $x \in (x_0-\delta, x_0)$, $x \neq x_0$; на отрезке $[x, x_0]$ выполняются условия теоремы Лагранжа (см. теорему 5.3) и, следовательно,

$$f(x_0) - f(x) = f'(c)(x_0 - x),\ c \in (x, x_0).$$

Так как $f'(x) > 0$ при $x \in (x_0-\delta, x_0)$ и $x_0 > x$, то из этого равенства получаем

$$f(x) < f(x_0). \tag{5.20}$$

Применяя теорему Лагранжа к функции $f(x)$ на отрезке $[x_0, x]$, когда $x \in (x_0, x_0+\delta)$, получаем

$$f(x) - f(x_0) = f'(c)(x - x_0),\ c \in (x_0, x).$$

Поскольку справа от точки x_0 выполняются неравенства $f'(x) < 0$ и $x_0 < x$, то из этого равенства следует, что

$$f(x_0) > f(x). \tag{5.21}$$

Из неравенств (5.20) и (5.21) вытекает, что в δ-окрестности точки x_0 выполняется условие локального максимума.

Случай локального минимума (перемены знака $f'(x)$ с "−" на "+") рассматривается аналогично.

Если же $f'(x)$ не меняет знака при переходе через точку x_0, то по теореме 5.7 она является монотонной на интервале $(x_0-\delta, x_0+\delta)$ и, значит, не имеет локального экстремума в точке x_0. Доказательство теоремы завершено.

Рассмотрим применение доказанных теорем на примерах нахождения точек локальных экстремумов функций.

Пример 1. Найти локальный экстремум и интервалы монотонности функции $f(x) = x^3 - 7{,}5x^2 + 18x$.

Сначала находим производную $f'(x) = 3x^2 - 15x + 18$. Приравняв ее к нулю и решая уравнение $x^2 - 5x + 6 = 0$, находим две точки возможного экстремума: $x_1 = 2$ и $x_2 = 3$. Нетрудно

видеть, что $f'(x)$ при переходе через точку $x_1 = 2$ меняет знак с "+" на "−", т.е. в этой точке имеет место локальный максимум; аналогично устанавливается, что в точке $x_2 = 3$ функция $f(x)$ имеет локальный минимум.

Рис. 5.6

Найдем теперь интервалы монотонности данной функции (рис. 5.6). Поскольку $f'(x) > 0$ при $x \in (-\infty, 2)$, то в силу теоремы 5.7 функция монотонно возрастает на этом интервале; $(2, 3)$ является интервалом монотонного убывания $f(x)$ $(f'(x) < 0)$, а на интервале $(3, +\infty)$ функция монотонно возрастает $(f'(x) > 0)$.

Пример 2. Найти размеры консервной банки, имеющей форму цилиндра (радиус r и высота h) заданного объема V, при которых полная поверхность сосуда будет минимальной. Эта задача имеет производственный смысл: найти оптимальные размеры банки, при которых затраты материала на ее изготовление будут минимальны.

Исходя из формулы объема цилиндра $V = \pi r^2 h$, выразим h:

$$h = V/(\pi r^2).$$

Как известно, формула полной поверхности цилиндра

$$S = 2\pi r h + 2\pi r^2.$$

Подставляя сюда формулу для h, получаем S как функцию от r:

$$S(r) = 2\pi\left(\frac{V}{\pi r} + r^2\right).$$

Минимум этой функции найдем из условия $S'(r) = 0$, откуда получаем уравнение $2r - V/(\pi r^2) = 0$. Из этого уравнения находим оптимальное значение r, подставляем его в формулу для h и окончательно вычисляем оптимальные размеры банки:

$$r = \sqrt[3]{V/(2\pi)},\quad h = \sqrt[3]{4V/\pi}.$$

Например, при $V = 0{,}33$ л оптимальные размеры банки составят: диаметр дна $\approx 7{,}5$ см и высота $\approx 7{,}5$ см.

Выпуклость и точки перегиба графика функции

Определение 2. Будем говорить, что график функции $y = f(x)$ имеет на (a, b) *выпуклость, направленную вниз (вверх)*, если он расположен не ниже (не выше) любой касательной к графику функции на (a, b) (рис. 5.7).

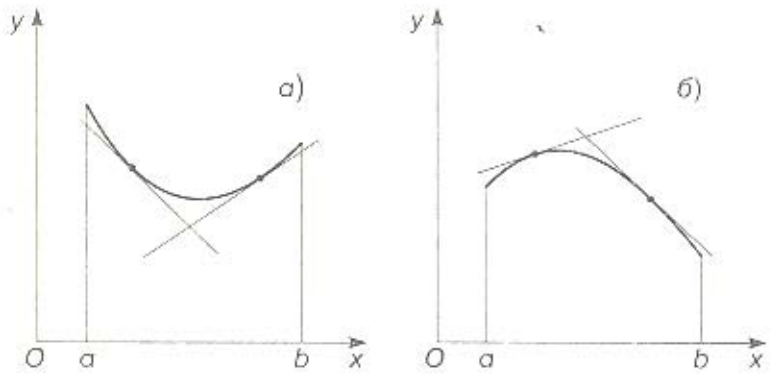

Рис. 5.7

Теорема 5.10. *Если функция $y = f(x)$ имеет на интервале (a, b) вторую производную и $f''(x) \geqslant 0$ ($f''(x) \leqslant 0$) на (a, b), то график функции имеет на (a, b) выпуклость, направленную вниз (вверх).*

Доказательство. Рассмотрим случай $f''(x) \geqslant 0$ на (a, b). Пусть c — произвольная точка интервала (a, b); уравнение касательной к графику функции $y = f(x)$, проходящей через точку $M(c, f(c))$, с угловым коэффициентом $f'(c)$ (рис. 5.8) имеет вид [см. формулу (3.23)]

$$Y - f(c) = f'(c)(x - c), \text{ или}$$
$$Y = f(c) + f'(c)(x - c), \tag{5.22}$$

где Y — текущая ордината касательной.

Разложим функцию $f(x)$ в окрестности точки c по формуле Тейлора при $n = 1$; для всех $x \in (a, b)$

$$y = f(x) = f(c) + \frac{f'(c)}{1!}(x - c) + \frac{f''(\xi)}{2!}(x - c)^2, \ \xi \in (a, b). \tag{5.23}$$

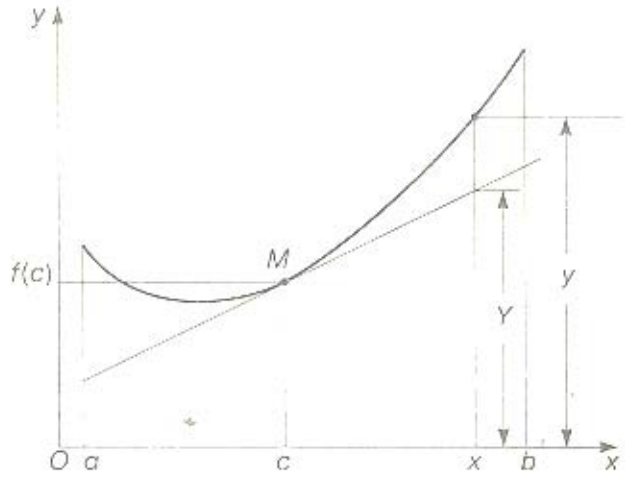

Рис. 5.8

Вычитая равенство (5.22) из равенства (5.23), получим

$$y - Y = \frac{f''(\xi)}{2!}(x - c)^2. \tag{5.24}$$

По условию $f''(x) \geqslant 0$ на (a, b), следовательно, правая часть уравнения (5.24) неотрицательна, т.е. $y \geqslant Y$ для всех x из (a, b). Это неравенство и доказывает теорему: график функции всюду на (a, b) лежит не ниже касательной в произвольной точке этого интервала.

Определение 3. Точка $M(x_0, f(x_0))$ называется *точкой перегиба* графика функции $y = f(x)$, если в точке M график имеет касательную и существует такая окрестность точки x_0, в пределах которой график функции $f(x)$ имеет разные направления выпуклости.

В точке перегиба касательная пересекает график функции, поскольку он переходит с одной стороны касательной на другую, т.е. "перегибается" через нее (рис. 5.9).

Теорема 5.11 (НЕОБХОДИМОЕ УСЛОВИЕ ТОЧКИ ПЕРЕГИБА). *Пусть график функции $y = f(x)$ имеет перегиб в точке $M(x_0, f(x_0))$ и функция $f(x)$ имеет в точке x_0 непрерывную вторую производную. Тогда*

$$f''(x_0) = 0. \tag{5.25}$$

10 Зак. 8853

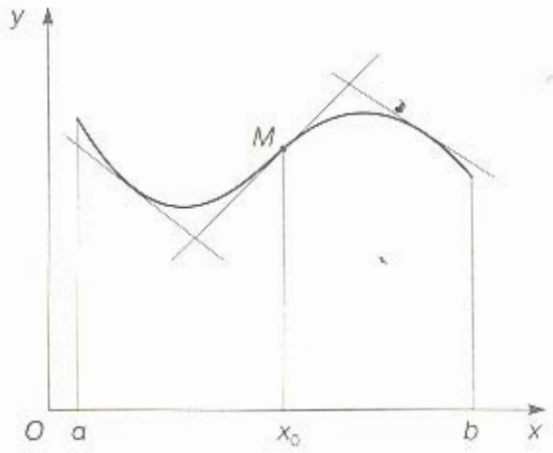

Рис. 5.9

Доказательство. Допустим обратное, т.е. что $f''(x_0) \neq 0$. Тогда в силу теоремы 3.10 об устойчивости знака непрерывной функции существует окрестность точки x_0, в которой либо $f''(x) > 0$, либо $f''(x) < 0$. Но, согласно теореме 5.10, график функции имеет определенное направление выпуклости в этой окрестности, что противоречит условию доказываемой теоремы о существовании перегиба в точке M. Полученное противоречие доказывает теорему.

Отметим, что не всегда условие $f''(x_0) = 0$ означает наличие точки перегиба на графике функции $y = f(x)$. Например, график функции $y = x^{2n}$ $(n = 2, 3, \dots)$ не имеет перегиба в точке $(0, 0)$, хотя при $x = 0$ вторая производная равна нулю (рис. 5.10). Потому равенство (5.25) является только необходимым условием перегиба.

Точки графика, для которых условие (5.25) выполняется, будем называть *критическими*. В каждой такой точке необходимо исследовать дополнительно вопрос о существовании перегиба.

Теорема 5.12 (ДОСТАТОЧНОЕ УСЛОВИЕ ТОЧКИ ПЕРЕГИБА). *Пусть функция $y = f(x)$ имеет вторую производную в некоторой окрестности точки x_0, которая имеет разные знаки слева и справа от x_0. Тогда график $y = f(x)$ имеет перегиб в точке $M(x_0, f(x_0))$.*

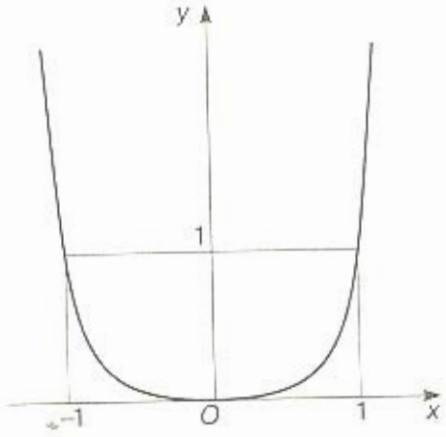

Рис. 5.10

Доказательство. В силу теоремы 5.10 направления выпуклости слева и справа от точки x_0 являются различными. Согласно определению 2 это и означает наличие перегиба графика функции $y = f(x)$ в точке x_0. Теорема доказана.

Теорема верна и для случая, когда $f''(x)$ существует в некоторой окрестности точки x_0, за исключением самой точки x_0, и существует касательная к графику функции в точке M. Например, функция $f(x) = x^{1/3}$ в точке $x = 0$ имеет бесконечные производные; в точке 0 касательная совпадает с осью Oy. Однако график этой функции имеет перегиб в начале координат, поскольку вторая производная $f''(x) = -2/(9x^{5/3})$ имеет разные знаки слева и справа от точки $x = 0$ (рис. 5.11).

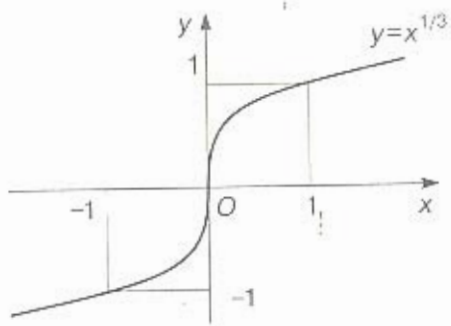

Рис. 5.11

Рассмотрим примеры: найти точки перегиба и направления выпуклости графиков функций.

Пример 3. $f(x) = \exp(-x^2)$.

Последовательно находим: $f'(x) = -2x\exp(-x^2)$, $f''(x) = 2\exp(-x^2)\cdot(2x^2-1)$. Приравнивая вторую производную к нулю, получаем критические точки: $x = \pm 1/\sqrt{2}$. Поскольку функция четная, достаточно исследовать точку $x = 1/\sqrt{2}$. Нетрудно видеть, что при переходе через эту точку слева направо $f''(x)$ меняет знак с минуса на плюс. Следовательно, на левой ветви функции точка $M_1(-1/\sqrt{2}, e^{-1/2})$ является точкой перегиба графика функции со сменой выпуклости вниз слева на выпуклость вверх справа (рис. 5.12). На правой ветви в точке перегиба $M_2(1/\sqrt{2}, e^{-1/2})$ графика функции имеет место смена выпуклости вверх слева на выпуклость вниз справа.

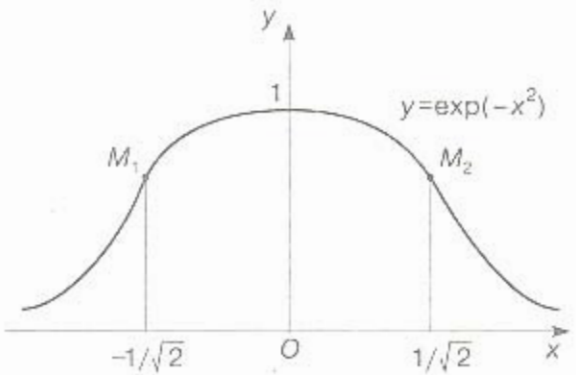

Рис. 5.12

Пример 4. $f(x) = \ln(x^2 - 2x + 2)$.

Вторая производная этой функции $f''(x) = 2\dfrac{2x - x^2}{(x^2-2x+2)^2}$.

Приравнивая ее к нулю, получаем критические точки $x_1 = 0$, $x_2 = 2$. Несложный анализ квадратного трехчлена $x(2-x)$, стоящего в числителе второй производной и определяющего ее знак, показывает, что точка перегиба M_1 $(0, \ln 2)$ графика функции меняет выпуклость вверх слева на выпуклость вниз справа; в другой точке перегиба M_2 $(2, \ln 2)$ выпуклость графика функции вниз слева меняется на выпуклость вверх справа (рис. 5.13).

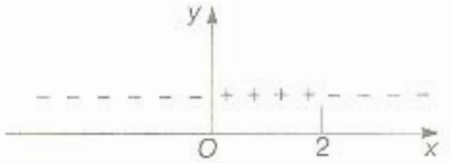

Рис. 5.13

Асимптоты графика функции

Часто оказывается, что график функции неограниченно близко приближается к некоторой прямой. Такого рода прямые называются *асимптотами*. Неограниченность приближения графика функции к асимптоте означает, что расстояние от графика до этой прямой (перпендикуляр, опущенный из произвольной точки графика на прямую) стремится к нулю. Различают три вида асимптот: *вертикальные, горизонтальные и наклонные*.

Определение 4. Прямая $x = a$ называется *вертикальной асимптотой* графика функции $y = f(x)$, если хотя бы одно из предельных значений $\lim\limits_{x\to a+} f(x)$ или $\lim\limits_{x\to a-} f(x)$ равно $+\infty$ или $-\infty$.

Вертикальные асимптоты обычно сопутствуют точкам разрыва второго рода. Например, график функции $y = c^{1/x}$ имеет вертикальную асимптоту $x = 0$, так как $f(x) \to \infty$ при $x \to 0+$ (см. рис. 3.16).

Определение 5. Прямая $y = kx + b$ называется *наклонной асимптотой* графика функции $y = f(x)$ при $x \to \pm\infty$, если $f(x)$ можно представить в виде

$$f(x) = kx + b + \alpha(x), \tag{5.26}$$

где $\alpha(x) \to 0$ при $x \to \pm\infty$.

Это определение относится как к наклонной, так и к горизонтальной асимптотам: в случае горизонтальной асимптоты угловой коэффициент k в выражении (5.26) равен нулю.

Укажем способ нахождения коэффициентов k и b в уравнении наклонной асимптоты. Разделив обе части равенства (5.26) на x и переходя к пределу при $x \to \infty$, получим

$$\lim_{x\to\infty}\frac{f(x)}{x} = \lim_{x\to\infty}\left[k + \frac{b}{x} + \frac{\alpha(x)}{x}\right] = k, \text{ т.е.}$$

$$k = \lim_{x\to\infty} \frac{f(x)}{x}. \tag{5.27}$$

Затем из равенства (5.26) получаем

$$\lim_{x\to\infty} [b + \alpha(x)] = b = \lim_{x\to\infty} [f(x) - kx]. \tag{5.28}$$

Рассмотрим примеры нахождения асимптот графиков функций.

Пример 5. $f(x) = \dfrac{2x^2 - 8}{x + 1}$.

Сначала находим вертикальную асимптоту. Точка $x = -1$ является точкой разрыва 2-го рода, причем

$$\lim_{x\to -1-} f(x) = +\infty, \quad \lim_{x\to -1+} f(x) = -\infty.$$

Затем находим наклонную асимптоту:

$$k = \lim_{x\to\infty} \frac{f(x)}{x} = \lim_{x\to\infty} \frac{2x^2 - 8}{x^2 + x} = \lim_{x\to\infty} \frac{2 - 8/x^2}{1 + 1/x} = 2,$$

$$b = \lim_{x\to\infty} [f(x) - kx] = \lim_{x\to\infty} \left[\frac{2x^2 - 8}{x + 1} - 2x\right] = -\lim_{x\to\infty} \frac{2x + 8}{x + 1} = -2.$$

Таким образом, получаем уравнение наклонной асимптоты:

$$y = 2x - 2.$$

Пример 6. $f(x) = x + e^{-x^2}$.

Вертикальных асимптот здесь нет, поскольку точки разрыва 2-го рода отсутствуют. Отыщем наклонную асимптоту:

$$k = \lim_{x\to\infty} \frac{x + e^{-x^2}}{x} = \lim_{x\to\infty} \frac{1 + e^{-x^2}/x}{1} = 1,$$

$$b = \lim_{x\to\infty} \left[x + e^{-x^2} - x\right] = \lim_{x\to\infty} e^{-x^2} = 0.$$

Таким образом, уравнение наклонной асимптоты имеет вид

$$y = x.$$

Схема исследования графика функции

Приведем схему исследования поведения функции и построения ее графика.

1. Найти область определения функции.

2. Определить возможный тип симметрии функции: четность или нечетность функции. Функция $f(x)$ называется *четной*, если выполнено условие симметрии ее графика относительно оси Oy:

$$f(-x) = f(x). \tag{5.29}$$

Функция $f(x)$ называется *нечетной*, если выполнено условие симметрии ее графика относительно начала координат:

$$f(-x) = -f(x). \tag{5.30}$$

При наличии симметрии достаточно построить график функции на правой координатной полуплоскости и затем отобразить его на левую половину: зеркально относительно оси Oy в случае выполнения равенства (5.29) (рис. 5.14, *а*) или с центральной симметрией, если справедливо равенство (5.30) (рис. 5.14, *б*).

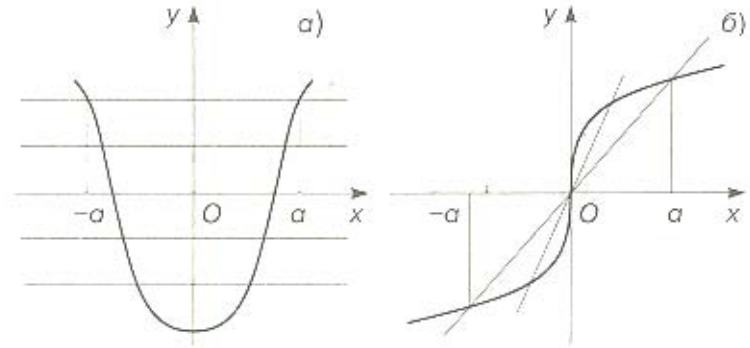

Рис. 5.14

3. Найти точки пересечения графика функции с осями координат Ox и Oy, т.е. решить соответственно уравнения $f(x) = 0$ и $y = f(0)$.

4. Найти асимптоты.

5. Найти точки возможного экстремума.

6. Найти критические точки.

7. Исследовать знаки первой и второй производных, определить участки монотонности функции, направление выпуклости графика, точки экстремума и перегиба.

8. Определить максимум и минимум функции на области ее определения. Если областью определения функции является отрезок $[a, b]$, необходимо вычислить значения функции в его концах и сопоставить их с внутренними экстремумами.

9. Построить график функции с учетом проведенного исследования. Приведем два примера на построение графиков функций по изложенной выше схеме.

Пример 7. $y = x^2 e^{-x^2}$. (5.31)

1. Областью определения данной функции является вся числовая прямая.

2. $f(-x) = f(x)$, т.е. функция четная; построим ее график на правой полуплоскости $x \geqslant 0$ и затем отобразим на левую полуплоскость зеркально относительно оси Oy.

3. $y = 0$ при $x = 0$ и из уравнения $f(x) = 0$ следует $x = 0$, т.е. график функции пересекается с осями координат в точке $O(0, 0)$.

4. Вертикальных асимптот нет, поскольку нет точек разрыва функции; наклонные (горизонтальные) асимптоты находим с использованием формул (5.27) и (5.28) и применением правила Лопиталя:

$$k = \lim_{x\to\infty} \frac{f(x)}{x} = \lim_{x\to\infty} \left(xe^{-x^2}\right) = \lim_{x\to\infty} \frac{x}{\exp(x^2)} =$$

$$= \lim_{x\to\infty} \frac{1}{2x\exp(x^2)} = 0,$$

$$b = \lim_{x\to\infty} [f(x) - kx] = \lim_{x\to\infty} [x^2 \exp(-x)] = \lim_{x\to\infty} \frac{x^2}{\exp(x^2)} =$$

$$= \lim_{x\to\infty} \frac{2x}{2x\exp(x^2)} = \lim_{x\to\infty} \frac{1}{\exp(x^2)} = 0.$$

Стало быть, уравнение асимптоты $y = 0$, т.е. ось Ox.

5. Находим точки возможного экстремума, для чего вычислим первую производную функции и приравняем ее к нулю: $f'(x) = 0$, или $2x(1 - x^2)\exp(-x^2) = 0$, откуда получаем $x_1 = 0$ и $x_2 = \pm 1$ (так как экспонента никогда не равна нулю).

6. Критические точки находятся путем приравнивания второй производной функции к нулю: $f''(x) = 0$, откуда $2(2x^4 - 5x^2 + 1)\exp(-x^2) = 0$. Следовательно, критические точки даются решениями биквадратного уравнения $2x^4 - 5x^2 + 1 = 0$, его положительные корни $x_1 = \left(\sqrt{5 - \sqrt{17}}\right)/2$, $x_2 = \left(\sqrt{5 + \sqrt{17}}\right)/2$.

7. Производная положительна при $x \in (0, 1)$ и отрицательна при $x > 1$, т.е. в точке $x = 0$ функция имеет локальный минимум, а в точках $x = \pm 1$ — локальный максимум. На интервале $(0, 1)$ функция возрастает от нуля до $1/e$, на промежутке $(1, -\infty)$ она убывает от $1/e$ до нуля.

Нетрудно проверить, что критические точки, определенные в п. 6, являются точками перегиба графика функции. Слева от точки x_1 вторая производная $f''(x) > 0$, справа от нее $f''(x) < 0$, т.е. на интервале $(0, x_1)$ выпуклость графика функции (5.31) направлена вниз, на (x_1, x_2) функция выпукла вверх и на $(x_2, -\infty)$ выпуклость направлена опять вниз.

8. Поскольку область определения функции (5.31) — вся числовая прямая, то минимум и максимум этой функции совпадают с ее соответствующими локальными экстремумами.

9. График функции с учетом ее полного исследования приведен на рис. 5.15. Функция всюду неотрицательна ($f(x) \geqslant 0$).

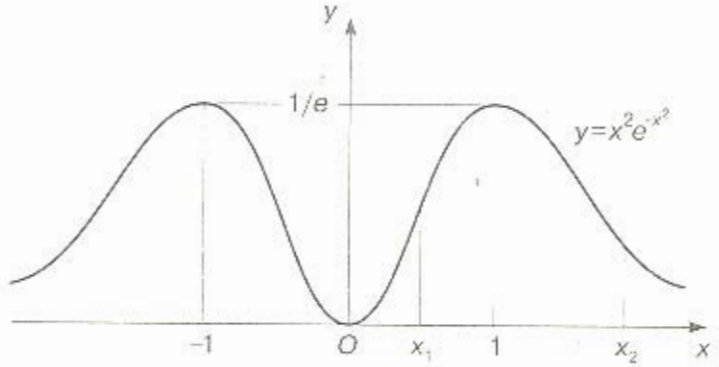

Рис. 5.15

Пример 8. $y = \dfrac{x^2 - 1}{x}$. (5.32)

Действуем по приведенной выше схеме.

1. Область определения функции: $x \neq 0$.

2. Функция (5.32) является нечетной, т.к. $f(-x) = -f(x)$.

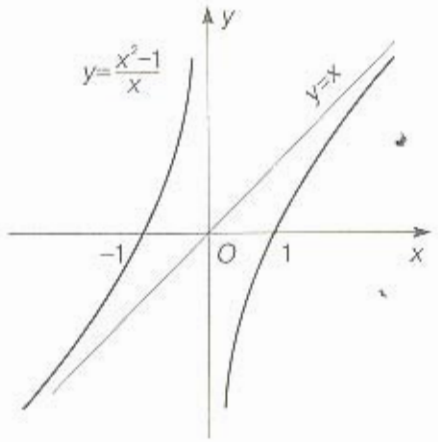

Рис. 5.16

3. Уравнение $f(x) = 0$ дает корни $x = \pm 1$ (точки пересечения с осью Ox). Пересечения с осью Oy нет в силу п. 1.

4. Имеется вертикальная асимптота — ось Oy, так как предел $f(x)$ при $x \to 0$ бесконечен: $f(x) \to -\infty$ при $x \to 0-$, $f(x) \to -\infty$ при $x \to 0+$.

Определяем наклонную асимптоту:

$$k = \lim_{x\to\infty} \frac{f(x)}{x} = \lim_{x\to\infty} \frac{x^2-1}{x^2} = 1,$$

$$b = \lim_{x\to\infty} [f(x) - k(x)] = \lim_{x\to\infty} \left[\frac{x^2-1}{x} - x\right] =$$

$$= \lim_{x\to\infty} \frac{-1}{x} = 0.$$

Итак, уравнение наклонной асимптоты: $y = x$.

5. $f'(x) = \dfrac{x^2+1}{x^2}$, т.е. производная нигде не равна нулю и точек возможного экстремума нет. В области определения везде $f'(x)$ положительна.

6. $f''(x) = -2/x^3$ — критических точек нет.

7. Функция (5.32) монотонно возрастает на всей области своего определения, так как ее производная всюду положительна.

В левой координатной полуплоскости выпуклость графика функции направлена вниз ($f''(x) > 0$), в правой полуплоскости выпуклость направлена вверх ($f''(x) < 0$).

8. Наибольшего и наименьшего значений функции не существует, поскольку область ее значений не ограничена.

9. График функции (5.32) приведен на рис. 5.16.

5.5. Применение в экономике

Предельные показатели в микроэкономике

Приведем пример двух предельных показателей в микроэкономике — себестоимости и эластичности.

1. Рассмотрим зависимость себестоимости C произведенной продукции от ее объема Q, т.е. функцию $C = f(Q)$. Так называемая *предельная себестоимость* характеризует себестоимость ΔC прироста продукции ΔQ:

$$MC = \frac{\Delta C}{\Delta Q}. \tag{5.33}$$

В предположении о непрерывной зависимости ΔC от ΔQ естественно напрашивается замена разностного отношения в выражении (5.33) его пределом:

$$MC \approx \lim_{\Delta Q\to 0} \frac{\Delta C}{\Delta Q} = C'(Q). \tag{5.34}$$

Обычно в приложениях с использованием аппарата математики под предельной себестоимостью понимают именно величину (5.34).

Например, пусть зависимость издержек производства от объема выпускаемой продукции выражается формулой

$$C = 40Q - 0{,}03Q^3.$$

Определим средние и предельные издержки при объеме продукции $Q = 15$ ден. ед.

Функция средних издержек на единицу продукции определяется по формуле $\overline{C} = C/Q$, или в нашем случае

$$\overline{C} = 40 - 0{,}03Q^2,$$

откуда $\overline{C}(15) = 40 - 0{,}03 \cdot 225 = 33{,}25$ ден. ед.

Предельные издержки определяются, согласно выражению (5.34), по формуле

$$C' = 40 - 0{,}09Q^2,$$

откуда при $Q = 15$ получаем $C'(15) = 19{,}75$ ден. ед.

Иными словами, при средних издержках на производство единицы продукции в 33,25 ден. ед. дополнительные затраты на производство единицы дополнительной продукции составят 19,75 ден. ед. и не превысят средних издержек.

2. В анализе и прогнозах ценовой политики применяется понятие *эластичности спроса*. Пусть $D = D(P)$ — функция спроса от цены товара P (см. параграф 3.1). Тогда под эластичностью спроса понимается относительное изменение спроса при изменении цены товара на один процент:

$$E = \frac{(\Delta D/D)100\%}{(\Delta P/P)100\%}. \tag{5.35}$$

Как и в предыдущем случае, при непрерывной зависимости ΔD от ΔP удобно перейти к пределу при $\Delta P \to 0$:

$$E(D) = P\frac{D'(P)}{D(P)}. \tag{5.36}$$

Аналогичное понятие можно ввести и для функции предложения $S(P)$. Напомним, что функция $D(P)$ убывает, а функция $S(P)$ возрастает с ростом P.

Укажем некоторые свойства эластичности спроса. Как следует из формулы (5.36), ее можно выразить так:

$$E(D) = P(\ln D(P))'. \tag{5.37}$$

Из равенства (5.37) следует, что $E(D)$ обладает свойствами логарифма, а значит:

$$\begin{aligned} E(D_1D_2) &= E(D_1) + E(D_2), \\ E(D_1/D_2) &= E(D_1) - E(D_2). \end{aligned} \tag{5.38}$$

Заметим, что поскольку функция $D(P)$ убывающая, то $D'(P) < 0$, а тогда, согласно формуле (5.36), и $E(D) < 0$.

Различают три вида спроса в зависимости от величины $|E(D)|$:

а) если $|E(D)| > 1$ $(E(D) < -1)$, то спрос считается эластичным;

б) если $|E(D)| = 1$ $(E(D) = -1)$, то спрос нейтрален;

в) если $|E(D)| < 1$ $(E(D) > -1)$, то спрос неэластичен.

Рассмотрим два примера из этой области.

Пример 1. Пусть функция спроса описывается формулой

$$D(P) = D_0 \exp(-kP^2), \tag{5.39}$$

где D_0 и k — известные величины. Найти, при каких значениях цены P спрос будет эластичным.

Согласно формуле (5.36) составляем выражение для $E(D)$:

$$E(D) = \frac{-2kPD_0 \exp(-kp^2)}{D_0 \exp(-kP^2)} P = -2kP^2.$$

Для того чтобы спрос был эластичным, необходимо, чтобы выполнялось неравенство

$$2kP^2 > 1,$$

откуда $P > 1/\sqrt{2k}$.

Пример 2. Найти изменение выручки с увеличением цены товара при разных вариантах эластичности спроса.

Выручка I равна произведению цены товара P на величину спроса D:

$$I(P) = D(P)P. \tag{5.40}$$

Найдем производную этой функции:

$$I'(P) = D(P) + PD'(P). \tag{5.41}$$

Теперь проанализируем все варианты эластичности спроса, приведенные выше, с учетом формулы (5.36):

а) $E(D) < -1$. Тогда, подставляя формулу (5.36) в это неравенство, получаем, что правая часть уравнения (5.41) отрицательна. Таким образом, при эластичном спросе повышение

цены P ведет к снижению выручки. Напротив, снижение цены на товар увеличивает выручку;

б) $E(D) = -1$. Из формулы (5.36) следует, что правая часть выражения (5.41) равна нулю, т.е. при нейтральном спросе изменение цены на товар не влияет на выручку;

в) $E(D) > -1$. Тогда $I'(P) > 0$, т.е. при неэластичном спросе повышение цены на товар P приводит к росту выручки.

Понятие эластичности распространяется и на другие области экономики. Рассмотрим один характерный пример.

Пример 3. Пусть зависимость между себестоимостью продукции C и объемом ее производства Q выражается формулой

$$C = 50 - 0{,}4Q.$$

Требуется определить эластичность себестоимости при выпуске продукции $Q = 30$ ден. ед. По формуле (5.36) получаем

$$E(C) = -\frac{0{,}4Q}{50 - 0{,}4Q},$$

откуда при $Q = 15$ искомая эластичность составит около $-0{,}14$, т.е. при данном объеме выпуска продукции увеличение его на 1% приведет к снижению его себестоимости примерно на 0,14%.

Максимизация прибыли

Пусть Q — количество реализованного товара, $R(Q)$ — функция дохода, $C(Q)$ — функция затрат на производство товара. В реальности вид этих функций зависит, в первую очередь, от способа производства, организации инфрастуктуры и т.п. Прибыль от реализации произведенного товара

$$\Pi(Q) = R(Q) - C(Q). \tag{5.42}$$

В микроэкономике известно утверждение: для того чтобы прибыль была максимальной, необходимо, чтобы предельный доход и предельные издержки были равны. Оба упомянутых предельных показателя определяются по аналогии с выражением (5.34), так что этот принцип можно записать в виде:

$R'(Q) = C'(Q)$. Действительно, из необходимого условия экстремума для функции (5.42) следует, что $\Pi'(Q) = 0$, откуда и получается основной принцип.

Пример 4. Пусть

$$\begin{aligned} R(Q) &= 100Q - Q^2, \\ C(Q) &= Q^3 - 37Q^2 + 169Q + 4000. \end{aligned} \tag{5.43}$$

Тогда прибыль определяется формулой: $\Pi(Q) = -Q^3 + 36Q^2 - 69Q - 4000$. Приравнивая производную функции прибыли к нулю, получаем уравнение

$$Q^2 - 24Q + 23 = 0.$$

Корни этого уравнения $Q_1 = 1$, $Q_2 = 23$. Проверка показывает, что максимальная прибыль достигается при $Q = 23$: $\Pi_{\max} = 1290$.

Оптимизация налогообложения предприятий

Пусть t — налог с единицы выпускаемой продукции. Тогда общий налог с Q единиц продукции составит $T = tQ$. В этом случае функция прибыли будет иметь вид

$$\Pi(Q) = R(Q) - C(Q) - tQ. \tag{5.44}$$

Возникает вопрос: каким должен быть налог t, чтобы величина суммарного налога T со всей продукции была наибольшей?

Рассмотрим этот вопрос на примере: пусть $R(Q) = 16Q - Q^2$, а $C(Q) = Q^2 + 1$. Тогда функция прибыли имеет вид

$$\Pi(Q) = 16Q - 2Q^2 - tQ - 1.$$

Как и в предыдущей задаче, условие максимума прибыли $\Pi'(Q) = 0$; отсюда получаем значение Q, максимизирующего прибыль с учетом пока еще неизвестного налога t:

$$Q_{opt} = 4 - \frac{t}{4}.$$

Полученное значение объема продукции следует подставить в величину суммарного налога T и, в свою очередь, найти условия, при которых величина T будет максимальной. Итак,

$$T = Qt = t\left(4 - \frac{t}{4}\right);\ T' = 0,\ \text{откуда получаем, что } t = 8.$$

Отсюда следует, что $Q_{opt} = 2$, и при этом значении максимальная величина прибыли составит $\Pi_{max} = 7$, а оптимальный (с точки зрения налогового законодательства) сбор налога $T_{opt} = 16$.

Интересно сопоставить эти цифры с цифрами при отсутствии налогообложения. При $t = 0$ решение задачи на максимизацию прибыли дает следующие результаты: $Q_{opt} = 4$, $\Pi_{max} = 31$. Следовательно, уменьшение налогообложения стимулирует рост выпуска продукции и приводит при этом к увеличению прибыли от ее реализации. Понятно, почему производители прикладывают столько усилий, чтобы снизить ставку налога.

Закон убывающей эффективности производства

Этот закон утверждает, что при увеличении одного из основных факторов производства, например капитальных затрат K, прирост производства, начиная с некоторого значения K, является убывающей функцией. Иными словами, объем произведенной продукции V как функция от K описывается графиком со сменой выпуклости вниз на выпуклость вверх.

Характерный вид этой функции дается уравнением:

$$V(K) = V_{lim}/(1 + ae^{-bK+c}), \tag{5.45}$$

где a, b и c — известные положительные числа (они определяются прежде всего структурой организации производства), а V_{lim} — предельно возможный объем выпускаемой продукции. Нетрудно подсчитать, что вторая производная функции (5.45) имеет вид

$$V''(K) = V_{lim}ab^2e^{-bK+c}\frac{ae^{-bK+c} - 1}{(1 + ae^{-bK+c})^3}. \tag{5.46}$$

Критическая точка находится по условию $V''(K) = 0$, откуда

$$K_{cr} = (c + \ln a)/b. \tag{5.47}$$

График функции (5.45) приведен на рис. 5.17. В точке перегиба выпуклость графика функции (5.47) вниз меняется на выпуклость вверх. До этой точки увеличение капитальных затрат приводит к интенсивному росту объема продукции: прирост объема продукции (аналог первой производной) возрастает, т.е. $V''(K) > 0$. При $K > K_{cr}$ прирост объема выпускаемой продукции снижается, т.е. $V''(K) < 0$, и эффективность увеличения капитальных затрат падает.

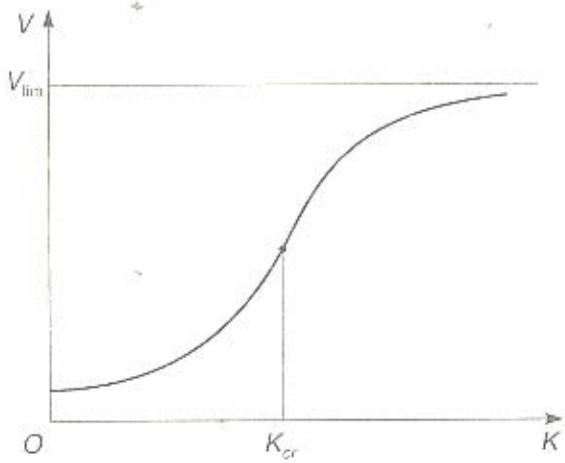

Рис. 5.17

Таким образом, в стратегии капиталовложений очень важным моментом оказывается определение критического объема затрат, сверх которого дополнительные затраты будут приводить ко все меньшей отдаче при данной структуре организации производства. Зная этот прогноз, можно пытаться совершенствовать и менять структуру организации производства: "улучшать" показатели a, b, c и V_{lim} в сторону повышения эффективности капиталовложений.

Глава 6

НЕОПРЕДЕЛЕННЫЙ ИНТЕГРАЛ

6.1. Первообразная и неопределенный интеграл

Понятие первообразной функции

Предыдущая глава была посвящена одной из основных задач дифференциального исчисления — нахождению производной заданной функции. Множество вопросов математического анализа и приложений в разнообразных науках приводят к другой задаче: по данной функции $f(x)$ найти такую функцию $F(x)$, производная которой равна функции $f(x)$.

Определение 1. Функция $F(x)$ называется *первообразной* для функции $f(x)$ на промежутке X, если для любого $x \in X$ функция $F(x)$ дифференцируема и выполняется равенство $F'(x) = f(x)$.

Приведем примеры. Функция $F(x) = \sin x$ является первообразной для функции $f(x) = \cos x$ на бесконечном промежутке $(-\infty, +\infty)$, так как при любых x выполняется равенство $(\sin x)' = \cos x$.

Функция $F(x) = \ln x$ первообразная для функции $f(x) = 1/x$ на промежутке $(0, +\infty)$, так как в каждой точке этого интервала выполняется равенство $(\ln x)' = 1/x$.

Функция $F(x) = x^2$ первообразная для функции $f(x) = 2x$ на всей числовой прямой, так как для любого x выполняется равенство $(x^2)' = 2x$.

Заметим, что задача отыскания по заданной функции $f(x)$ ее первообразной неоднозначна; если $F(x)$ первообразная, то и функция $F(x) + C$, где C — произвольное постоянное число, также первообразная для функции $f(x)$, так как $[F(x) + C]' = f(x)$.

Лемма. *Если производная функции $f(x)$ равна нулю на промежутке X, то эта функция постоянна на промежутке X.*

Доказательство. Для любых двух точек x_1, $x_2 \in X$ в силу теоремы Лагранжа

$$f(x_1) - f(x_2) = f'(c)(x_1 - x_2), \quad x_1 < c < x_2.$$

Поскольку $f'(x) = 0$ в любой точке промежутка X, то $f(x_1) = f(x_2)$; это и означает, что значения функции во всех точках X одинаковы, т.е. $f(x) = C$, где C — некоторое число. Лемма доказана.

Теорема 6.1. *Пусть $F(x)$ — первообразная для функции $f(x)$ на промежутке X; тогда любая другая первообразная для $f(x)$ на X может быть представлена в виде $F(x) + C$, где C — некоторое число.*

Доказательство. Пусть $\Phi(x)$ — любая другая первообразная для функции $f(x)$ на промежутке X, т.е. $\Phi'(x) = f(x)$. Тогда

$$[\Phi(x) - F(x)]' = \Phi'(x) - F'(x) = f(x) - f(x) = 0.$$

По доказанной выше лемме функция $\Phi(x) - F(x) = C$, где C — некоторое число, т.е. $\Phi(x) = F(x) + C$, что и требовалось доказать.

Из теоремы 6.1 вытекает, что множество функций $F(x) + C$, где $F(x)$ — одна из первообразных для функции $f(x)$, а C — произвольная постоянная, образует семейство первообразных для функции $f(x)$.

Неопределенный интеграл

Определение 2. Совокупность всех первообразных функций для функции $f(x)$ на промежутке X называется *неопределенным интегралом* от функции $f(x)$ на этом промежутке и обозначается символом

$$\int f(x)\,dx = F(x) + C. \tag{6.1}$$

В этом обозначении знак $\int$ называется *знаком интеграла* (это стилизованная латинская буква S, означающая суммирование), $f(x)$ — *подынтегральной функцией*, $f(x)\,dx$ — *подынтегральным выражением*, а переменная x — *переменной интегрирования*.

Операция нахождения первообразной по ее производной или неопределенного интеграла по заданной подынтегральной функции называется *интегрированием* этой функции. Интегрирование является операцией, обратной дифференцированию. Для проверки правильности выполнения интегрирования

нужно продифференцировать результат и получить при этом подынтегральную функцию.

Рассмотрим примеры.

1. $\int 2x\,dx = x^2 + C$; проверка: $(x^2 + C)' = 2x$.
2. $\int \sin x\,dx = -\cos x + C$; проверка: $(-\cos x + C)' = \sin x$.
3. $\int e^{3x}\,dx = \dfrac{1}{3}e^{3x} + C$; проверка: $\left(\dfrac{1}{3}e^{3x} + C\right)' = e^{3x}$.

6.2. Основные свойства неопределенного интеграла

Прежде всего отметим свойства, которые непосредственно вытекают из определения неопределенного интеграла:

1°. $d\left(\int f(x)\,dx\right) = f(x)\,dx$ и $\left(\int f(x)dx\right)' = f(x)$.

2°. $\int dF(x) = F(x) + C$.

Для доказательства свойства 1° достаточно взять дифференциал от обеих частей выражения (6.1) и записать, что $dF(x) = F'(x)\,dx$. Для установления свойства 2° нужно в левой части выражения (6.1) использовать запись дифференциала и определение первообразной: $dF(x) = f(x)\,dx$.

Следующие два свойства называются *линейными свойствами* неопределенного интеграла:

3°. $\int kf(x)\,dx = k\int f(x)\,dx$, где $k = \text{const} \neq 0$.

4°. $\int[f(x) \pm g(x)]\,dx = \int f(x)dx \pm \int g(x)\,dx$.

Свойство 3° вытекает из свойства для производной функции $F(x)$: $(kF(x))' = kF'(x) = kf(x)$. Свойство 4° вытекает из того, что если $F(x)$ и $G(x)$ — первообразные соответственно для функций $f(x)$ и $g(x)$, то производная их суммы равна сумме их производных.

Заметим, что последнее свойство справедливо для любого конечного числа слагаемых в подынтегральной функции.

6.3. Основные неопределенные интегралы

Ранее мы получили основные производные элементарных функций (см. параграф 4.10). Приводимые ниже основные неопределенные интегралы представляют собой вычислительный аппарат интегрального исчисления. Часть формул непосредственно следует из определения интегрирования как операции, обратной дифференцированию. Справедливость всех формул легко проверить дифференцированием.

I. $\displaystyle\int x^\alpha\,dx = \frac{x^{\alpha-1}}{\alpha+1} + C,\ (\alpha \neq -1)$.

II. $\displaystyle\int \frac{dx}{x} = \ln|x| + C,\ (x \neq 0)$.

III. $\displaystyle\int \frac{dx}{1+x^2} = \operatorname{arctg} x + C\ \ (-\operatorname{arcctg} x + C)$.

IV. $\displaystyle\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C\ \ (-\arccos x + C),\ (-1 < x < 1)$.

V. $\displaystyle\int a^x\,dx = \frac{a^x}{\ln a} + C,\ (0 < a \neq 1)$; $\int e^x\,dx = e^x + C$.

VI. $\displaystyle\int \sin x\,dx = -\cos x + C$.

VII. $\displaystyle\int \cos x\,dx = \sin x + C$.

VIII. $\displaystyle\int \frac{dx}{\cos^2 x} = \operatorname{tg} x + C,\ \left(x \neq \tfrac{\pi}{2} + n\pi,\ n = 0, \pm1, \pm2, \ldots\right)$.

IX. $\displaystyle\int \frac{dx}{\sin^2 x} = -\operatorname{ctg} x + C,\ (x \neq n\pi,\ n = 0,\ \pm1,\ \pm2,\ \ldots)$.

X. $\displaystyle\int \frac{dx}{x^2 - a^2} = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| + C,\ (|x| \neq a,\ a \neq 0)$.

XI. $\displaystyle\int \frac{dx}{\sqrt{x^2+k}} = \ln\left|x + \sqrt{x^2+k}\right| + C,\ (|x| > \sqrt{|k|}$ при $k < 0)$.

К этим формулам можно добавить и интегралы от гиперболических функций:

XII. $\displaystyle\int \operatorname{sh} x\,dx = \operatorname{ch} x + C$.

XIII. $\int \operatorname{ch} x\,dx = \operatorname{sh} x + C.$

XIV. $\displaystyle\int \frac{dx}{\operatorname{ch}^2 x} = \operatorname{th} x + C.$

XV. $\displaystyle\int \frac{dx}{\operatorname{sh}^2 x} = -\operatorname{cth} x + C.$

Приведенные выше интегралы принято называть *табличными*.

Как было установлено в параграфе 4.10, операция дифференцирования не выводит функцию из класса элементарных. С операцией интегрирования дело обстоит иначе: интегралы от некоторых элементарных функций уже не являются элементарными функциями. Укажем некоторые из них:

$\displaystyle\int e^{-x^2}\,dx$ — интеграл Пуассона (интеграл ошибок);

$\displaystyle\int \sin(x^2)\,dx$ — интеграл Френеля;

$\displaystyle\int \cos(x^2)\,dx$ — интеграл Френеля;

$\displaystyle\int \frac{dx}{\ln x}\,dx$ — интегральный логарифм;

$\displaystyle\int \frac{\sin x}{x}\,dx$ — интегральный синус;

$\displaystyle\int \frac{\cos x}{x}\,dx$ — интегральный косинус.

Каждый из этих интегралов есть функция, которая не является элементарной. Они играют большую роль в прикладных науках; так, интеграл Пуассона является одним из основных в теории вероятностей и статистике.

Как правило, интегралы, с которыми приходится иметь дело в различных приложениях, не выражаются элементарными функциями (или, как принято говорить, являются "неберущимися"). Тем не менее существуют достаточно хорошо разработанный аппарат приближенных формул с использованием элементарных функций, а также методы приближенных расчетов, позволяющие с любой степенью точности оценивать и вычислять "неберущиеся" интегралы.

6.4. Основные методы интегрирования

Непосредственное интегрирование

Вычисление интегралов с использованием основных свойств неопределенных интегралов и простейших интегралов называется *непосредственным интегрированием*.

Покажем это на примерах.

1. $\displaystyle\int (2\sin x + 6 - 3x^2)\,dx = 2\int \sin x\,dx + 6\int dx - 3\int x^2\,dx =$
$= -2\cos x + 6x - x^3 + C.$

2. $\displaystyle\int \frac{3x^4 - x^3 + x + 2}{x^2}\,dx = \int \left(3x^2 - x + \frac{1}{x} + \frac{2}{x^2}\right) dx =$
$\displaystyle= 3\int x^2 dx - \int x dx + \int \frac{1}{x}dx + 2\int \frac{1}{x^2}\,dx =$
$\displaystyle= x^3 - \frac{x^2}{2} + \ln x - \frac{2}{x} + C.$

3. $\displaystyle\int \left(\frac{2}{1+x^2} + \sin^2\frac{x}{2}\right) dx = 2\int \frac{1}{1+x^2}dx + \int \left(\frac{1}{2} - \frac{1}{2}\cos x\right) dx =$
$\displaystyle= 2\operatorname{arctg} x + \frac{1}{2}x - \frac{1}{2}\sin x + C.$

Метод подстановки

Замена переменной интегрирования является одним из самых эффективных приемов сведения неопределенного интеграла к табличному. Такой прием называется *методом подстановки*, или *методом замены переменной*. Он основан на следующей теореме.

Теорема 6.2. *Пусть функция $x = \varphi(t)$ определена и дифференцируема на некотором промежутке T, а X — множество значений этой функции, на котором определена функция $f(x)$. Тогда, если функция $f(x)$ имеет первообразную на множестве X, на множестве T справедлива формула*

$$\int f(x)\,dx = \int f[\varphi(t)]\varphi'(t)\,dt. \tag{6.2}$$

Доказательство. Пусть $F(x)$ — первообразная для $f(x)$ на множестве X. По правилу дифференцирования сложной функции $F[\varphi(t)]$ на множестве T, с учетом равенства $F'(x) = f(x)$, получаем

$$(F[\varphi(t)])' = F'_x[\varphi(t)]\varphi'(t) = f[\varphi(t)]\varphi'(t).$$

Таким образом, функция $f[\varphi(t)]\varphi'(t)$ имеет на множестве T первообразную $F[\varphi(t)]$, а значит,

$$\int f[\varphi(t)]\varphi'(t)\,dt = F[\varphi(t)] + C.$$

Поскольку $F[\varphi(t)] + C = F(x) + C = \int f(x)\,dx$, из последнего равенства и получаем формулу (6.2).

Выражение (6.2) называется *формулой замены переменной в неопределенном интеграле.*

Рассмотрим применение этого приема на примерах вычисления интегралов.

1. $\int x(x-1)^{12}\,dx$.

Решение этой задачи путем разложения бинома Ньютона представляется весьма громоздким. Введем новую переменную $t = x - 1$. Тогда $x = t + 1$, $dx = dt$ и исходный интеграл преобразуется следующим образом:

$$\int x(x-1)^{12}\,dx = \int (t+1)t^{12}\,dt = \int (t^{13}+t^{12})\,dt = \frac{t^{14}}{14} + \frac{t^{13}}{13} + C.$$

Сделав обратную замену переменной, получаем окончательный ответ:

$$\int x(x-1)^{12}\,dx = (x-1)^{13}\left(\frac{x}{14} + \frac{1}{182}\right) + C.$$

2. $\int \dfrac{x^3}{(2-x)^3}\,dx$.

Положим $t = 2 - x$; тогда $x = 2 - t$, $dx = -dt$. Отсюда по формуле (6.2) получаем

$$\int \frac{x^3}{(2-x)^3}\,dx = -\int \frac{(2-t)^3}{t^3}\,dt = -\int (8t^{-3} - 12t^{-2} + 6t^{-1} - 1)\,dt =$$

$$= 4t^{-2} - 12t^{-1} - 6\ln t - t + C =$$
$$= 4(2-x)^{-2} - 12(2-x)^{-1} - 6\ln(2-x) - x + C.$$

3. $\int \sin^4 x \cos x\,dx$.

Здесь нужна замена переменной $\sin x = t$, тогда $dt = \cos x\,dx$. При подстановке в исходный интеграл получим

$$\int \sin^4 x \cos x\,dx = \int t^4\,dt = \frac{1}{5}t^5 + C = \frac{1}{5}\sin^5 x + C.$$

4. $\int e^{\cos x}\sin x\,dx$.

Делаем замену переменной $t = \cos x$, тогда $dt = -\sin x\,dx$, $e^{\cos x} = e^t$; подставляя в интеграл, получаем

$$\int e^{\cos x}\sin x\,dx = -\int e^t\,dt = -e^t + C = -e^{\cos x} + C.$$

5. $\int \dfrac{dx}{\cos x}$.

Преобразуем этот интеграл, переписав его в виде

$$\int \frac{\cos x}{\cos^2 x}\,dx = \int \frac{\cos x}{1 - \sin^2 x}\,dx.$$

Из вида подынтегральной функции следует, что целесообразно ввести новую переменную $t = \sin x$. Тогда $1 - \sin^2 x = 1 - t^2$, $dt = \cos x\,dx$; подстановка в интеграл даст

$$\int \frac{dx}{\cos x} = \int \frac{dt}{1-t^2} = \frac{1}{2}\ln\left|\frac{1+t}{1-t}\right| + C =$$
$$= \frac{1}{2}\ln\left|\frac{1+\sin x}{1-\sin x}\right| + C = \ln\left|\operatorname{tg}\left(\frac{x}{2} + \frac{\pi}{4}\right)\right| + C.$$

Здесь использован табличный интеграл X.

6. $\int \dfrac{x^3}{x^8+1}\,dx$.

Введем новую переменную $t = x^4$ и выполним все необходимые операции: $x^8 + 1 = t^2 + 1$, $dt = 4x^3 dx$, откуда получаем

$$\int \frac{x^3}{x^8+1}\,dx = \frac{1}{4}\int \frac{dt}{t^2+1} = \frac{1}{4}\operatorname{arctg} t + C = \frac{1}{4}\operatorname{arctg} x^4 + C.$$

7. $\int \dfrac{x\,dx}{(x^2+1)^n}$ $(n \neq 1)$.

Положим $t = x^2 + 1$, тогда $dt = 2x\,dx$ или $x\,dx = \dfrac{dt}{2}$, и данный интеграл принимает вид табличного интеграла:

$$\int \frac{x\,dx}{(x^2+1)^n} = \frac{1}{2}\int \frac{dt}{t^n} = -\frac{1}{2(n-1)}t^{-(n-1)} + C = \\ = -\frac{1}{2(n-1)}\frac{1}{(x^2+1)^{n-1}} + C.$$

Интегрирование по частям

Теорема 6.3. *Пусть функции $u(x)$ и $v(x)$ определены и дифференцируемы на промежутке X и функция $u'(x)v(x)$ имеет первообразную на этом промежутке. Тогда функция $u(x)v'(x)$ также имеет первообразную на промежутке X, причем справедлива формула*

$$\int u(x)v'(x)\,dx = u(x)v(x) - \int v(x)u'(x)\,dx. \qquad (6.3)$$

Доказательство. По правилу дифференцирования произведения функций

$$[u(x)v(x)]' = u'(x)v(x) + u(x)v'(x),$$

откуда следует, что

$$u(x)v'(x) = [u(x)v(x)]' - u'(x)v(x).$$

В правой части последнего равенства первое слагаемое имеет первообразную функцию $u(x)v(x)$ на X, второе слагаемое имеет первообразную на X по условию. Следовательно, и левая часть имеет первообразную на X, которая находится путем интегрирования этого равенства. Формула (6.3), а значит и теорема, доказана.

С учетом вида дифференциала функции $v'(x)\,dx = dv$ и $u'(x)\,dx = du$ равенство (6.3) часто используют в форме

$$\int u\,dv = uv - \int v\,du. \qquad (6.4)$$

Равенство (6.3) или (6.4) называется *формулой интегрирования по частям.*

В интегрировании по частям самым сложным пунктом является выбор в подынтегральном выражении сомножителя $v'(x)\,dx = dv$. Под знак дифференциала d можно, в принципе, внести все что угодно; однако выбор должен быть таким, чтобы интеграл в правой части равенства (6.3) был проще исходного, а не сложнее. В этом смысле метод интегрирования по частям позволяет снести вычисление интеграла $\int u\,dv$ к интегралу $\int v\,du$, которое должно быть более простым.

Найдем следующие интегралы методом интегрирования по частям.

Пример 1. Вычислить $\int \ln x\,dx$.

Здесь берем $u(x) = \ln x$, $dv = dx$, т.е. $v = x$. По формуле (6.3) получаем

$$\int \ln x\,dx = x\ln x - \int x(\ln x)'dx = \\ = x\ln x - \int 1\,dx = x(\ln x - 1) + C.$$

В общем случае интегралы типа $\int x^n \ln x\,dx$, где $n \neq -1$ — целое число, берутся только интегрированием по частям: $u = \ln x$, $x^n dx = dv$, т.е. $v = x^{n+1}/(n+1)$.

Пример 2. Найти $\int xe^x dx$.

В этом случае $u = x$, $e^x dx = dv = d(e^x)$, тогда по формуле (6.4)

$$\int xe^x\,dx = xe^x - \int e^x x'\,dx = xe^x - \int e^x\,dx = \\ = e^x(x-1) + C.$$

Интегралы вида $\int x^n e^{kx} dx$, где $n > 0$ — целое число и $k \neq 0$ — любое число, берутся n-кратным интегрированием по частям до исчезновения степени x в подынтегральной функции; при этом каждый раз под знак d вносится e^{kx}, т.е. $e^{kx}\,dx = dv = \dfrac{1}{k}d(e^{kx})$.

Пример 3. Найти $\int x^2 \cos x\,dx$.

Интегралы вида $\int x^n \cos kx\,dx$ и $\int x^n \sin kx\,dx$, где k — любое число и $n > 0$ — целое число, вычисляются так же, как и интеграл общего вида, приведенный в примере 1. Под знак d каждый раз вносится тригонометрическая функция, и процедура интегрирования по частям повторяется n раз: $\cos kx\,dx = dv = \frac{1}{k} d(\sin kx)$, затем $\sin kx\,dx = \frac{1}{k} d(\cos kx)$ и т.д. В данном случае

$$\int x^2 \cos x\,dx = \int x^2\,d(\sin x) = x^2 \sin x - \int 2x \sin x\,dx =$$
$$= x^2 \sin x + 2\int x d(\cos x) = x^2 \sin x + 2x \cos x - 2\int \cos x\,dx =$$
$$= x^2 \sin x + 2x \cos x - 2\sin x + C.$$

Пример 4. Найти $\int \operatorname{arctg} \sqrt{3x-1}\,dx$.

Здесь ничего нельзя предложить, кроме как $u = \operatorname{arctg} \sqrt{3x-1}$, $dv = dx$. Интегрирование по частям дает

$$\int \operatorname{arctg} \sqrt{3x-1}\,dx =$$
$$= x \operatorname{arctg} \sqrt{3x-1} - \int \frac{x}{1+3x-1} \frac{3}{2\sqrt{3x-1}}\,dx =$$
$$= x \operatorname{arctg} \sqrt{3x-1} - \frac{1}{2}\int \frac{dx}{\sqrt{3x-1}} =$$
$$= x \operatorname{arctg} \sqrt{3x-1} - \frac{1}{3}\sqrt{3x-1} + C.$$

Заметим, что вычисление последнего интеграла проводилось с заменой переменной $t = 3x - 1$.

Пример 5. Найти $I_n = \int \frac{dx}{(x^2+1)^n}$, где n — целое положительное число.

При $n = 1$ получаем табличный интеграл III (см. параграф 6.3):

$$I_1 = \operatorname{arctg} x + C. \qquad (6.5)$$

Пусть $n > 1$. Выполним тождественное преобразование числителя подынтегральной функции: $1 = (x^2+1) - x^2$. Подставляя

в интеграл, получаем

$$I_n = \int \frac{dx}{(x^2+1)^{n-1}} - \int \frac{x^2\,dx}{(x^2+1)^n}.$$

Для нахождения второго интеграла применим метод интегрирования по частям, полагая

$$u = x,\ du = dx,\ dv = \frac{x\,dx}{(x^2+1)^n},$$

откуда $v = \int \frac{x\,dx}{(x^2+1)^n} = -\frac{1}{2(n-1)} \frac{1}{(x^2+1)^{n-1}}$ (см. п. 7); тогда

$$\int \frac{x^2\,dx}{(x^2+1)^n} = -\frac{1}{2(n-1)} \frac{x}{(x^2+1)^{n-1}} + \frac{1}{2(n-1)} \int \frac{dx}{(x^2+1)^{n-1}}.$$

Подставляя в исходный интеграл, получаем

$$I_n = I_{n-1} + \frac{1}{2(n-1)} \frac{x}{(x^2+1)^{n-1}} - \frac{1}{2(n-1)} I_{n-1},$$

и после приведения подобных следует окончательный ответ:

$$I_n = \frac{1}{2(n-1)} \frac{x}{(x^2+1)^{n-1}} + \frac{2n-3}{2(n-1)} I_{n-1}. \qquad (6.6)$$

Интеграл I_n оказался выраженным через I_{n-1}. Формулы вида (6.6) называются *рекуррентными*. Для вычисления I_n при любом n теперь нет нужды в проведении интегрирования: зная значение I_1 из выражения (6.5), по формуле (6.6) последовательно вычисляем $I_2, \ldots, I_{n-1}, I_n$. Примененное выше интегрирование по частям позволило установить саму формулу (6.6).

Покажем использование формулы (6.6): найдем интеграл $\int \frac{dx}{(x^2+1)^2}$. Получаем при $n = 2$: $I_2 = \frac{1}{2} \frac{x}{x^2+1} + \frac{1}{2} \operatorname{arctg} x + C$.

6.5. Интегрирование рациональных функций

Интегрирование простых дробей

При классификации элементарных функций был выделен важный класс рациональных функций, которые можно представить в виде дроби $P(x)/Q(x)$, где $P(x)$ и $Q(x)$ — многочлены. Если степень числителя не меньше степени знаменателя,

то такая дробь является *неправильной*; в таком случае, выполнив деление, получим

$$\frac{P(x)}{Q(x)} = W(x) + \frac{R(x)}{Q(x)}, \tag{6.7}$$

где $W(x)$ — некоторый многочлен, а второе слагаемое в правой части равенства (6.7) представляет собой *правильную дробь*, у которой степень числителя $R(x)$ меньше степени знаменателя.

Например, $\dfrac{x^5+2x^3-3x^2+x+2}{x^3-x^2+x+1} = x^2+x+2-\dfrac{3x^2+2x}{x^3-x^2+x-1}$.

Выделим из класса правильных дробей так называемые *основные простые* дроби. Это дроби следующих четырех типов:

I. $\dfrac{A}{x-a}$;

II. $\dfrac{A}{(x-a)^k}$;

III. $\dfrac{Mx+N}{x^2+2px+q}$;

IV. $\dfrac{Mx+N}{(x^2+2px+q)^k}$,

где a, p, q, A, M, N — вещественные числа, $k>1$ — целое число; кроме того, предполагается, что знаменатели дробей типов III и IV не имеют вещественных корней, т.е.

$$q-p^2>0. \tag{6.8}$$

Рассмотрим интегралы от четырех основных видов простых дробей. Дроби типов I и II интегрируются с помощью подстановки $t=x-a$:

$$\int \frac{A}{x-a}\,dx = A\ln|x-a|+C;$$

$$\int \frac{A}{(x-a)^k}\,dx = -\frac{1}{k-1}\frac{A}{(x-a)^{k-1}}+C.$$

При рассмотрении интеграла от дроби типа III выделим из трехчлена в знаменателе полный квадрат

$$x^2+2px+q = (x+p)^2+q-p^2.$$

Отсюда следует подстановка $t=x+p$, $x=t-p$, $dx=dt$, и тогда

$$\int \frac{Mx+N}{x^2+2px+q}\,dx = \int \frac{Mt+N-Mp}{t^2+a^2}\,dt =$$
$$= \frac{1}{2}M\int \frac{2t\,dt}{t^2+a^2} + (N-Mp)\int \frac{dt}{t^2+a^2},$$

где $a^2=q-p^2>0$ по условию (6.8). Первый интеграл в правой части берется непосредственно (или подстановкой $z=t^2+a^2$), второй интеграл с помощью подстановки $z=t/a$ сводится к табличному интегралу:

$$\int \frac{2t\,dt}{t^2+a^2} = \ln(t^2+a^2)+C = \ln(x^2+2px+q)+C,$$
$$\int \frac{dt}{t^2+a^2} = \frac{1}{a}\operatorname{arctg}\frac{t}{a}+C = \frac{1}{\sqrt{q-p^2}}\operatorname{arctg}\frac{x+p}{\sqrt{q-p^2}}+C.$$

Таким образом, окончательно получаем

$$\int \frac{Mx+N}{x^2+2px+q}\,dx =$$
$$= \frac{1}{2}M\ln(x^2+2px+q) + \frac{N-Mp}{\sqrt{q-p^2}}\operatorname{arctg}\frac{x+p}{\sqrt{q-p^2}}+C. \tag{6.9}$$

Пример 1. Найти $\displaystyle\int \frac{2x+1}{x^2+2x+5}\,dx$.

Можно повторить весь процесс нахождения интеграла типа III, однако здесь мы воспользуемся формулой (6.9), подставив в нее значения $M=2$, $N=1$, $p=1$, $q=5$. Получаем

$$\int \frac{2x+1}{x^2+2x+5}\,dx = \ln(x^2+2x+5) - \frac{1}{2}\operatorname{arctg}\frac{x+1}{2}+C.$$

Теперь найдем интеграл от дроби типа IV. Введем новую переменную

$t=(x+p)/a$, где $a=\sqrt{q-p^2}$, откуда $x=at-p$, $dx=a\,dt$.

В результате подстановки получаем

$$\int \frac{Mx+N}{(x^2+2px+q)^k}\,dx = \frac{1}{a^{2k-1}}\int \frac{Mat+N-Mp}{(t^2+1)^k}\,dt.$$

Разобьем интеграл в правой части последнего выражения на два слагаемых:

$$\frac{1}{a^{2k-1}}\int \frac{Mat+N-Mp}{(t^2+1)^k}\,dt = \\ = \frac{M}{a^{2k-2}}\int \frac{t\,dt}{(t^2+1)^k} + \frac{N-Mp}{a^{2k-1}}\int \frac{dt}{(t^2+1)^k}.$$

Первый интеграл сводится к табличному интегралу подстановкой $z = t^2+1$:

$$\int \frac{t\,dt}{(t^2+1)^k} = \frac{1}{2}\int \frac{dz}{z^k} = -\frac{1}{2(k-1)}\frac{1}{z^{k-1}} + C = \\ = -\frac{1}{2(k-1)}\frac{1}{(t^2+1)^{k-1}} + C.$$

Второй интеграл вычисляется по рекуррентной формуле (6.6). Сводя все подстановки вместе, окончательно получаем рекуррентную формулу

$$\int \frac{Mx+N}{(x^2+2px+q)^k}\,dx = -\frac{M}{2(k-1)}\frac{1}{(x^2+2px+q)^{k-1}} + \frac{N-Mp}{a^{2k-1}}I_k = \\ = \frac{1}{2(k-1)}\frac{(N-Mp)x+Np-Mq}{(q-p^2)(x^2+2px+q)^{k-1}} + \frac{N-MP}{q-p^2}\frac{2k-3}{2(k-1)}I_{k-1}, \tag{6.10}$$

где $I_{k-1} = \int \dfrac{dx}{(x^2+2px+q)^{k-1}}$.

Пример 2. Вычислить интеграл $\displaystyle\int \frac{2x+3}{(x^2+6x+13)^2}\,dx$.

Не повторяя вывода формулы (6.10), осуществим следующие подстановки:

$$M = 2,\ N = 3,\ p = 3,\ q = 13,\ k = 2,\ \sqrt{q-p^2} = 2.$$

Получаем

$$\int \frac{2x+3}{(x^2+6x+13)^2}dx = -\frac{1}{8}\frac{3x+17}{x^2+6x+13} - \frac{3}{8}\int \frac{dx}{x^2+6x+13}.$$

Второй интеграл в правой части этого равенства вычисляется по формуле (6.9) при $M = 0$, $N = 1$, т. е. исходный интеграл равен

$$-\frac{1}{8}\frac{3x+17}{x^2+6x+13} - \frac{3}{16}\operatorname{arctg}\frac{x+3}{2} + C.$$

Разложение правильных дробей на простые дроби

Из высшей алгебры известно, что каждый многочлен с вещественными коэффициентами может быть единственным образом представлен в виде произведения, содержащего сомножители вида $x-a$ и $x^2+2px+q$, причем квадратичные сомножители не имеют вещественных корней. Если среди сомножителей имеются совпадающие, то многочлен может быть представлен в следующем виде:

$$Q(x) = A(x-a)^r(x-b)^s\cdots\times \\ \times(x^2+2px+q)^t(x^2+2ux+v)^w\ldots. \tag{6.11}$$

Здесь a, $b,\ldots$ — вещественные корни многочлена $Q(x)$; r, $s,\ldots$ — целые числа, называемые кратностями корней; t, $w,\ldots$ — целые числа, называемые кратностями квадратных трехчленов; $p, q, u, v,\ldots$ — действительные коэффициенты квадратных трехчленов; A — коэффициент при старшей степени многочлена $Q(x)$. Поскольку квадратные трехчлены в разложении (6.11) не имеют действительных корней, то справедливы неравенства: $p^2-q<0$, $u^2-v<0,\ldots$. При этом $r+s+\cdots+2t+2w+\cdots = n$, где n — степень многочлена $Q(x)$.

Справедлива теорема высшей алгебры о разложении рациональной функции на простые дроби, которую мы приводим без доказательства.

Теорема 6.4. *Пусть* $\dfrac{P(x)}{Q(x)}$ — *правильная рациональная дробь, у которой знаменатель представлен в виде многочлена*

12 Зак. 6833

(6.11); *тогда эту дробь можно единственным образом представить в виде суммы простых дробей:*

$$\frac{P(x)}{Q(x)} = \frac{A_1}{x-a} + \frac{A_2}{(x-a)^2} + \dots + \frac{A_r}{(x-a)^r} + \dots + \\ + \frac{M_1x+N_1}{x^2+2px+q} + \frac{M_2x+N_2}{(x^2+2px+q)^2} + \dots + \\ + \frac{M_tx+N_t}{(x^2+2px+q)^t} + \dots, \tag{6.12}$$

где

$$A_1, A_2, \dots, A_r, M_1, N_1, \dots, M_t, N_t, \dots \tag{6.13}$$

— *некоторые вещественные числа.*

Выражение (6.12) называется *разложением рациональной дроби на простые дроби.*

Для того чтобы определить неизвестные пока коэффициенты $A_1, A_2, \dots, A_r, \dots, M_1, N_1, \dots$, умножим обе части выражения (6.12) на $Q(x)$. Так как равенство между многочленом $P(x)$ в левой части и многочленом в правой части должно соблюдаться для всех x, то коэффициенты при одинаковых степенях x в обеих частях должны быть равны между собой. Приравнивая их, получим систему линейных алгебраических уравнений относительно чисел (6.13). Этот метод нахождения коэффициентов (6.13) разложения (6.12) правильной рациональной дроби на сумму простых дробей называется *методом неопределенных коэффициентов.*

Рассмотрим этот метод на примерах разложения рациональных функций на простые дроби.

1. $\dfrac{3x-1}{x^2-3x+2}$.

Знаменатель имеет два корня $x_1 = 1$ и $x_2 = 2$. Разложение этой дроби на две простые имеет вид

$$\frac{3x-1}{(x-1)(x-2)} = \frac{A}{x-1} + \frac{B}{x-2}. \tag{6.14}$$

Умножая обе части этого равенства на $(x-1)(x-2)$, получаем $3x-1 = A(x-2) + B(x-1)$, или $3x-1 = (A+B)x - (2A+B)$. Приравнивая коэффициенты при одинаковых степенях x

в обеих частях последнего равенства, получаем систему двух линейных уравнений относительно A и B:

$$\begin{matrix} x^1 \\ x^0 \end{matrix} \begin{cases} A+B=3, \\ 2A+B=1, \end{cases}$$

откуда $A=-2$, $B=5$. Подставляя эти числа, получаем искомое разложение:

$$\frac{3x-1}{(x-1)(x-2)} = -\frac{2}{x-1} + \frac{5}{x-2}.$$

Для определения неизвестных коэффициентов A_r при старших степенях простых дробей со знаменателями $(x-a)^r$ в разложении дроби (6.12) существует эффективный метод (так называемый *метод вычеркивания*). Для этого следует убрать в знаменателе $Q(x)$ левой части множитель $(x-a)^r$ и в оставшемся выражении положить $x=a$.

Метод вычеркивания особенно эффективен в случае однократных вещественных корней. Так, в данном примере для определения коэффициента A следует в левой части равенства (6.14) убрать сомножитель $(x-1)$ в знаменателе и в оставшейся дроби принять $x=1$, т.е.

$$A = \left.\frac{3x-1}{x-2}\right|_{x=1} = \frac{2}{-1} = -2. \text{ Аналогично } B = \left.\frac{3x-1}{x-1}\right|_{x=2} = 5.$$

В общем случае разложения (6.12) методом вычеркивания можно найти часть неизвестных коэффициентов и существенно упростить общую систему линейных уравнений.

2. $\dfrac{2x^2+3x+6}{x(x-1)(x-3)(x^2+x+2)}$.

Согласно теореме 6.4 для этой дроби существует разложение:

$$\frac{2x^2+3x+6}{x(x-1)(x-3)(x^2+x+2)} = \frac{A}{x} + \frac{B}{x-1} + \frac{C}{x-3} + \frac{Mx+N}{x^2+x+2}. \tag{6.15}$$

Умножим обе части этого равенства на знаменатель левой части и получим основное уравнение для составления системы линейных уравнений относительно неизвестных коэффициентов:

$$0x^4 + 0x^3 + 2x^2 + 3x + 6 =$$

$$= A(x-1)(x-3)(x^2+x+2) + Bx(x-3)(x^2+x+2) +$$
$$+Cx(x-1)(x^2+x+2) + (Mx+N)x(x-1)(x-3).$$

Приравнивая коэффициенты при одинаковых степенях x слева и справа, получаем

$$\begin{array}{l} x^4 \\ x^3 \\ x^2 \\ x^1 \\ x^0 \end{array} \left\{ \begin{array}{l} A+B+C+M=0, \\ 3A+2B+4M-N=0, \\ A-B+C+3M-4N=2, \\ 5A+6B+2C-3N=-3, \\ 6A=6. \end{array} \right. \qquad (6.16)$$

Из последнего уравнения получаем $A=1$. Метод вычеркивания позволяет определить из (6.15): $B=-\dfrac{11}{8}$, $C=\dfrac{11}{28}$. Это упрощает решение системы линейных уравнений (6.16): из первого уравнения следует $M=-\dfrac{1}{56}$, из четвертого уравнения получаем $N=\dfrac{5}{28}$.

Таким образом, разложение (6.15) определено полностью:

$$\frac{2x^2+3x+6}{x(x-1)(x-3)(x^2+x+2)} =$$
$$= \frac{1}{x} - \frac{11}{8}\frac{1}{x-1} + \frac{11}{28}\frac{1}{x-3} + \frac{-x/56+5/28}{x^2+x+2}.$$

Интегрирование правильных рациональных дробей

Из изложенного выше следует, что при интегрировании правильных дробей необходимо сначала выполнить разложение вида (6.12) и методом неопределенных коэффициентов найти числа (6.13). Таким образом в разложении (6.12) получается сумма четырех видов основных простых дробей, интегралы от которых подробно рассмотрены выше.

6.6. Интегрирование иррациональных и трансцендентных функций

Уточним определение рациональной функции от двух переменных: это функция, полученная из переменных u и v путем проведения над ними арифметических операций. Например, функция

$$R(u,v) = \frac{2u^4v - uv^2}{u-v}.$$

Пусть переменные u и v являются, в свою очередь, функциями переменной x: $u=\varphi(x)$, $v=\psi(x)$; тогда функция $R[\varphi(x),\psi(x)]$ называется рациональной функцией от $\varphi(x)$ и $\psi(x)$. Например,

$$f(x) = R\left(x, \sqrt{x+2}\right) = \frac{x+\sqrt{x+2}}{1-\sqrt{x+2}},$$
$$f(x) = R(\sin x, \cos x) = \frac{\cos^2 x + 2\sin x}{\cos x - \sin x}.$$

Мы рассмотрим интегралы от некоторых иррациональных и трансцендентных функций. Покажем, что они могут быть сведены при помощи соответствующих подстановок к интегралам от рациональных функций (или рационализируются).

Дробно-линейные иррациональности

Рассмотрим интеграл от рациональной функции

$$\int R\left(x, \sqrt[m]{\frac{ax+b}{cx+k}}\right) dx, \qquad (6.17)$$

где m — натуральное число, a, b, c, k — вещественные числа, причем $\dfrac{a}{c} \neq \dfrac{b}{k}$. Здесь рационализирующей является подстановка

$$t = \sqrt[m]{\frac{ax+b}{cx+k}}. \qquad (6.18)$$

Действительно,

$$t^m = \frac{ax+b}{cx+k}, \quad x = \frac{b-kt^m}{ct^m-a}, \quad dx = \frac{mt^{m-1}(ak-bc)}{(ct^m-a)^2}\,dt.$$

Подстановка этих выражений в интеграл (6.17) дает

$$\int R\left(x,\ \sqrt[m]{\frac{ax+b}{cx+k}}\right)dx =$$

$$= \int R\left(\frac{b-kt^m}{ct^m-a},\ t\right)\frac{mt^{m-1}(ak-bc)}{(ct^m-a)^2}\,dt = \int R_1(t)\,dt,$$

где $R_1(t)$ — рациональная функция аргумента t.

Найдем интегралы при помощи рационализирующих подстановок.

Пример 1. Найти $\displaystyle\int \frac{1}{x}\sqrt{\frac{x-1}{x+1}}\,dx$.

Согласно формуле (6.18),

$$t=\sqrt{\frac{x-1}{x+1}},\quad t^2=\frac{x-1}{x+1},\quad x=-\frac{1+t^2}{t^2-1},\quad dx=\frac{4t}{(t^2-1)^2}\,dt.$$

Отсюда

$$\int \frac{1}{x}\sqrt{\frac{x-1}{x+1}}\,dx = -4\int \frac{t^2\,dt}{(t^2-1)(t^2+1)}.$$

Подынтегральная функция представляет собой правильную рациональную дробь и разлагается на сумму простых дробей:

$$\frac{t^2}{(t+1)(t-1)(t^2+1)} = \frac{A}{t+1}+\frac{B}{t-1}+\frac{Ct+D}{t^2+1},$$

где числа A, B, C, D находятся из уравнения

$$A(t-1)(t^2+1)+B(t+1)(t^2+1)+$$
$$+(Ct+D)(t^2-1)=t^2$$

методом неопределенных коэффициентов. Получаем

$$A=-\frac{1}{4},\ B=\frac{1}{4},\ C=0,\ D=\frac{1}{2}.$$

Подстановка этого разложения в интеграл в итоге дает

$$\int \frac{1}{x}\sqrt{\frac{x-1}{x+1}}dx = -\int\left(\frac{2}{t^2+1}-\frac{1}{t+1}+\frac{1}{t-1}\right)dt =$$
$$= -2\operatorname{arctg} t+\ln(t+1)-\ln(t-1)+C =$$
$$= -2\operatorname{arctg}\left(\sqrt{\frac{x-1}{x+1}}\right)+\ln\left[\left(\sqrt{\frac{x-1}{x+1}}+1\right)\Big/\left(\sqrt{\frac{x-1}{x+1}}-1\right)\right]+C.$$

Пример 2. Найти $\displaystyle\int \frac{dx}{\sqrt{x}+\sqrt[3]{x}}$.

Поскольку общее кратное показателей степеней x в знаменателе равно $1/6$, то нужна рационализирующая подстановка $t=x^{1/6}$, $t^6=x$, $dx=6t^5dt$, $\sqrt{x}=t^3$, $\sqrt[3]{x}=t^2$. Отсюда получаем

$$\int \frac{dx}{\sqrt{x}+\sqrt[3]{x}} = \int \frac{6t^5}{t^3+t^2}dt = 6\int\frac{t^3}{t+1}dt =$$
$$= 6\int\left(t^2-t+1-\frac{1}{t+1}\right)dt =$$
$$= 6\left[\frac{t^3}{3}-\frac{t^2}{2}+t-\ln(t+1)\right]+C =$$
$$= 2\sqrt{x}-3\sqrt[3]{x}+6\sqrt[6]{x}-6\ln(\sqrt[6]{x}+1)+C.$$

Квадратичные иррациональности

Рассмотрим интеграл от рациональной функции

$$\int R\left(x,\sqrt{ax^2+bx+c}\right)dx. \tag{6.19}$$

а) Если трехчлен ax^2+bx+c имеет вещественные корни x_1 и x_2, $a>0$, то при $x_1\neq x_2$:

$$\sqrt{ax^2+bx+c}=\sqrt{a(x-x_1)(x-x_2)}=\sqrt{a}|x-x_1|\sqrt{\frac{x-x_2}{x-x_1}},$$

откуда

$$R\left(x,\sqrt{ax^2+bx+c}\right)=R_1\left(x,\sqrt{a(x-x_2)/(x-x_1)}\right),$$

где R_1 — другая рациональная функция, т.е. интеграл (6.19) сводится к рассмотренному выше случаю.

б) Трехчлен ax^2+bx+c не имеет действительных корней и $a>0$. Тогда интеграл (6.19) рационализируется подстановкой Эйлера[1]:

$$t=\sqrt{ax^2+bx+c}+x\sqrt{a}. \tag{6.20}$$

Перенесем в левую часть второе слагаемое и возведем в квадрат обе части; получаем

$$x=\frac{t^2-c}{2\sqrt{a}t+b},\quad \sqrt{ax^2+bx+c}=\frac{\sqrt{a}t^2+bt+c\sqrt{a}}{2\sqrt{a}t+b},$$
$$dx=2\frac{\sqrt{a}t^2+bt+c\sqrt{a}}{(2\sqrt{a}t+b)^2}dt. \tag{6.21}$$

Нетрудно видеть, что при подстановке формул (6.21) в интеграл (6.19) получается интеграл от рациональной функции аргумента t:

$$\int R\left(x,\sqrt{ax^2+bx+c}\right)dx=\int R_1(t)\,dt.$$

При $a<0$ и $c>0$ в квадратном трехчлене ax^2+bx+c для рационализации интеграла (6.19) применяется другая подстановка Эйлера:

$$\sqrt{ax^2+bx+c}=xt\pm\sqrt{c}. \tag{6.22}$$

Следует заметить, что подстановки Эйлера представляют собой достаточно сильный метод вычисления интегралов; однако они приводят к громоздким выражениям и выкладкам, поэтому их целесообразно применять при отсутствии другой возможности.

Пример 3. Найти $\displaystyle\int\frac{dx}{\sqrt{x^2+k}}$.

Этот интеграл является табличным, тем не менее он вычисляется путем подстановки Эйлера (6.20):

$$t=\sqrt{x^2+k}+x,\quad dt=\left(\frac{x}{\sqrt{x^2+k}}+1\right)dx$$

[1]Леонард Эйлер (1707–1783) — российский математик.

или $\displaystyle dx=\frac{\sqrt{x^2+k}}{t}\,dt.$

Подставляя эти формулы в исходный интеграл, получаем

$$\int\frac{dx}{\sqrt{x^2+k}}=\int\frac{dt}{t}=\ln t+C=\ln\left(\sqrt{x^2+k}+x\right)+C.$$

Рациональная функция от $\sin x$ *и* $\cos x$

Рассмотрим интеграл вида

$$\int R(\sin x,\cos x)\,dx. \tag{6.23}$$

Этот интеграл рационализируется универсальной подстановкой

$$t=\operatorname{tg}\frac{x}{2},\quad -\pi<x<\pi. \tag{6.24}$$

Действительно,

$$\sin x=\frac{2t}{1+t^2},\ \cos x=\frac{1-t^2}{1+t^2},\ x=2\operatorname{arctg}t,\ dx=\frac{2\,dt}{1+t^2}. \tag{6.25}$$

Подстановка формул (6.25) в интеграл (6.23) дает

$$\int R(\sin x,\cos x)\,dx=\int R\left(\frac{2t}{1+t^2},\frac{1-t^2}{1+t^2}\right)\frac{2\,dt}{1+t^2}=\int R_1(t)\,dt,$$

где $R_1(t)$ — другая рациональная функция аргумента t.

Вычислим интегралы, содержащие рациональные функции от $\sin x$ и $\cos x$.

Пример 4. Найти $\displaystyle\int\frac{dx}{1+\cos x}$.

Используя формулы универсальной подстановки (6.25), после очевидных упрощений получаем

$$\int\frac{dx}{1+\cos x}=\int dt=t+C=\operatorname{tg}\frac{x}{2}+C.$$

Пример 5. Найти $\displaystyle\int\frac{dx}{2+\sin x}$.

Здесь снова используем универсальную подстановку (6.24) с вытекающими из нее формулами (6.25). После несложных преобразований получаем

$$\int \frac{dx}{2+\sin x} = \int \frac{dt}{t^2+t+1} = \frac{2}{\sqrt{3}} \operatorname{arctg} \frac{t+1/2}{\sqrt{3}/2} + C =$$
$$= \frac{2}{\sqrt{3}} \operatorname{arctg} \frac{2\operatorname{tg}(x/2)+1}{\sqrt{3}} + C.$$

Здесь использован интеграл от рациональной дроби типа III [см. формулу (6.10)].

Пример 6. Найти $\int \sin^5 x \cos^2 x\, dx$.

Для случая $\int \sin^m x \cos^n x\, dx$, где m и n — натуральные числа, универсальная подстановка приведет к громоздким выкладкам: гораздо удобнее применить метод замены переменной (см. параграф 6.4). В зависимости от четности m и n возможны четыре варианта:

а) m — четное, n — нечетное; подстановка $t = \sin x$;

б) m — нечетное, n — четное; подстановка $t = \cos x$;

в) m и n — нечетные; любая из двух подстановок "а" или "б";

г) m и n — четные; понизить степень тригонометрических функций и в полученной сумме проанализировать каждое слагаемое по этим пунктам.

В нашем примере m — нечетное, а n — четное, следовательно, согласно пункту "б", выполним подстановку $t = \cos x$; тогда $dt = -\sin x\, dx$, $\sin^4 x = (1-t^2)^2$; отсюда

$$\int \sin^5 x \cos^2 x\, dx = -\int (1-t^2)^2 t^2\, dt =$$
$$= -\int (t^2 - 2t^4 + t^6)\, dt = \frac{2}{5}t^5 - \frac{1}{3}t^3 - \frac{1}{7}t^7 + C =$$
$$= \cos^3 x \left(\frac{2}{5}\cos^2 x - \frac{1}{7}\cos^4 x - \frac{1}{3}\right) + C.$$

Рациональная функция от e^x

Интеграл вида

$$\int R(e^x)\, dx \tag{6.26}$$

рационализируется подстановкой

$$t = e^x, \quad \text{откуда} \quad x = \ln t, \quad dx = \frac{dt}{x}. \tag{6.27}$$

Действительно, подстановка формул (6.27) в интеграл (6.26) дает

$$\int R(e^x)\, dx = \int R(t) \frac{dt}{t},$$

где $R(t)$ — рациональная функция от t.

Пример 7. Найти $\int \dfrac{e^{5x}\, dx}{e^x + 1}$.

Применяя подстановку (6.27), получим

$$\int \frac{e^{5x}\, dx}{e^x+1} = \int \frac{t^5}{t+1}\frac{dt}{t} = \int \frac{t^4}{t+1}\, dt =$$
$$= \int \left(t^3 - t^2 + t - 1 + \frac{1}{t+1}\right) dt = \frac{t^4}{4} - \frac{t^3}{3} + \frac{t^2}{2} - t + \ln(t+1) + C =$$
$$= \frac{e^{4x}}{4} - \frac{e^{3x}}{3} + \frac{e^{2x}}{2} - e^x + \ln(e^x+1) + C.$$

Глава 7
ОПРЕДЕЛЕННЫЙ ИНТЕГРАЛ

7.1. Условия существования определенного интеграла

Определение определенного интеграла

Пусть функция $f(x)$ задана на отрезке $[a, b]$. Разобьем отрезок $[a, b]$ на n произвольных частей точками

$$a = x_0 < x_1 < x_2 < \cdots < x_i < x_{i+1} < \cdots < x_n = b. \tag{7.1}$$

Точки, разделяющие отрезок $[a, b]$ на частичные отрезки $[x_i, x_{i+1}]$ длиной $\Delta x_i = x_{i+1} - x_i$, будем называть *точками разбиения*. Выберем в каждом из частичных отрезков $[x_i, x_{i+1}]$ произвольную точку ξ_i

$$x_i \leqslant \xi_i \leqslant x_{i+1}, \quad 0 \leqslant i \leqslant n. \tag{7.2}$$

Образуем сумму произведений

$$\sigma = f(\xi_1)\Delta x_1 + f(\xi_2)\Delta x_2 + \ldots + f(\xi_n)\Delta x_n = \sum_{i=1}^{n} f(\xi_i)\Delta x_i, \quad (7.3)$$

которую будем называть *интегральной суммой* для функции $f(x)$ на отрезке $[a, b]$. Следует специально отметить, что интегральная сумма σ соответствует данному разбиению (7.1) и данному выбору промежуточных точек (7.2). Геометрический смысл величины σ показан на рис. 7.1: это сумма площадей прямоугольников с основаниями Δx_i и высотами $f(\xi)$, $i = 1, 2, \ldots, n$.

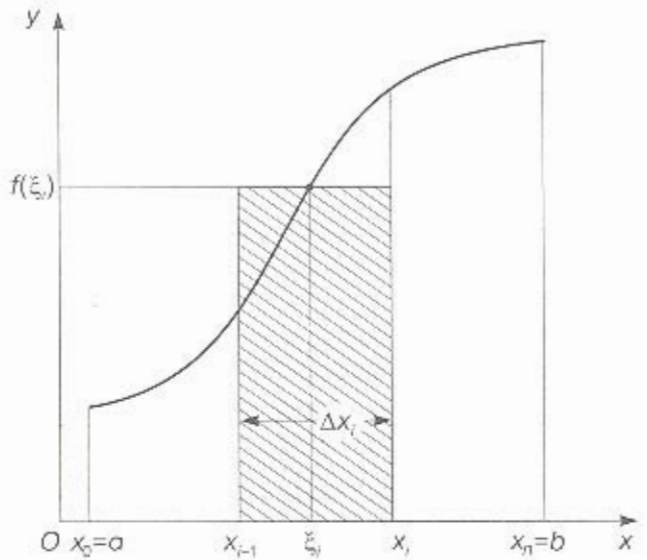

Рис. 7.1

Введем еще одну величину: обозначим через λ длину максимального частичного отрезка данного разбиения, т.е.

$$\lambda = \max_{1 \leqslant i \leqslant n} \{\Delta x_i\}. \quad (7.4)$$

Определение. Конечный предел I интегральной суммы σ при $\lambda \to 0$, если он существует, называется *определенным интегралом* от функции $f(x)$ по отрезку $[a, b]$:

$$I = \lim_{\lambda \to 0} \sum_{i=1}^{n} f(\xi_i)\Delta x_i. \quad (7.5)$$

Определенный интеграл обозначается символом

$$I = \int_a^b f(x)\,dx. \quad (7.6)$$

Если определенный интеграл (7.5) существует, то функция $f(x)$ называется *интегрируемой* на отрезке $[a, b]$, числа a и b — соответственно *нижним и верхним пределами интегрирования*, $f(x)$ — *подынтегральной функцией*, x — *переменной интегрирования*.

Вообще говоря, интегральная сумма (7.3) зависит от точек разбиения x_i и промежуточных точек ξ_i. Поскольку число тех и других стремится к бесконечности при $\lambda \to 0$, то определенный интеграл можно интерпретировать как бесконечную сумму бесконечно малых величин.

Пусть отрезок $[a, b]$ разбивается на частичные отрезки различными разбиениями T_k: сначала одним способом, потом другим и т.д. В каждом разбиении произвольно выбираются точки ξ_i, причем длина λ_k наибольшего частичного отрезка разбиения T_k стремится к нулю при $k \to \infty$; при этом интегральные суммы образуют числовую последовательность $\{\sigma_k\}$. Таким образом, возможно дать определение определенного интеграла "на языке последовательностей":

Функция $f(x)$ называется *интегрируемой* на $[a, b]$, если для любой последовательности разбиений $\{T_k\}$, для которой $\lim\limits_{k \to \infty} \lambda_k \to 0$, соответствующая последовательность интегральных сумм $\{\sigma_k\}$ стремится к одному и тому же числу I.

Определение определенного интеграла можно дать и "на языке ε–δ"; оба определения являются эквивалентными, что можно доказать по аналогии с доказательством эквивалентности двух определений предела функции (теорема 3.1).

Величина определенного интеграла, согласно данному выше определению, однозначно определяется видом функции $f(x)$ и числами a и b. Определенный интеграл не зависит от обозначения переменной интегрирования, т.е.

$$\int_a^b f(x)\,dx = \int_a^b f(t)\,dt = \int_a^b f(z)\,dz.$$

Ограниченность интегрируемой функции

Теорема 7.1 (НЕОБХОДИМОЕ УСЛОВИЕ ИНТЕГРИРУЕМОСТИ ФУНКЦИИ). *Интегрируемая на отрезке $[a,b]$ функция $f(x)$ ограничена на этом отрезке.*

Доказательство. Предположим обратное, т.е. что функция $f(x)$ не ограничена на $[a,b]$; тогда при любом разбиении отрезка $[a,b]$ она не ограничена хотя бы на одном частичном отрезке — пусть это будет отрезок $[x_0,x_1]$. Возьмем на остальных частичных отрезках произвольные точки ξ_i $(i=2,3,\ldots,n)$ и пусть

$$\sigma' = f(\xi_2)\Delta x_2 + f(\xi_3)\Delta x_3 + \cdots + f(\xi_n)\Delta x_n.$$

В силу неограниченности $f(x)$ на $[x_0,x_1]$ можно выбрать такую точку $\xi_1 \in [x_0,x_1]$, чтобы

$$|f(\xi_1)| > \frac{|\sigma'|+M}{\Delta x_1},$$

где M — произвольное положительное число. Тогда с учетом вида интегральной суммы σ [см. формулу (7.3)]

$$|\sigma| = |f(\xi_1)\Delta x_1 + \sigma'| \geqslant |f(\xi_1)|\Delta x_1 - |\sigma'| > M.$$

Оказалось, что интегральная сумма σ больше любого числа $M>0$, т.е. она не имеет конечного предела при $\lambda \to 0$ и, значит, $f(x)$ не интегрируема на отрезке $[a,b]$, что противоречит условию теоремы. Теорема доказана.

Заметим, что обратное утверждение неверно: ограниченная на отрезке функция может быть не интегрируемой на нем. Например, функция Дирихле

$$f(x) = \begin{cases} 1, & \text{если } x \text{ — рациональное число,} \\ 0, & \text{если } x \text{ — иррациональное число} \end{cases}$$

не интегрируема на отрезке $[0,1]$. Действительно, при любом разбиении этого отрезка можно выбрать ξ_i рациональными точками и тогда интегральная сумма (7.3) $\sigma = 1$ (так как $\Delta x_1 + \cdots + \Delta x_n = 1$; если же взять ξ_i иррациональными, то $\sigma = 0$. Следовательно, σ не имеет предела (7.5) при $\lambda \to 0$.

Суммы Дарбу[1]

Пусть функция $f(x)$ ограничена на отрезке $[a,b]$ и T — некоторое разбиение этого отрезка точками $a = x_0 < x_1 < \ldots < x_n = b$. Пусть m_i и M_i — соответственно точная нижняя и точная верхняя грани $f(x)$ на частичном отрезке $[x_i, x_{i+1}]$. Тогда суммы

$$S = M_1\Delta x_1 + M_2\Delta x_2 + \cdots + M_n\Delta x_n = \sum_{i=1}^{n} M_i\Delta x_i, \qquad (7.7)$$

$$s = m_1\Delta x_1 + m_2\Delta x_2 + \cdots + m_n\Delta x_n = \sum_{i=1}^{n} m_i\Delta x_i \qquad (7.8)$$

называются соответственно *верхней* и *нижней суммами Дарбу* функции $f(x)$ для данного разбиения T отрезка $[a,b]$.

Поскольку $m_i \leqslant f(\xi_i) \leqslant M_i$ при $\xi_i \in [x_i, x_{i+1}]$, то

$$s = \sum_{i=1}^{n} m_i\Delta x_i \leqslant \sum_{i=1}^{n} f(\xi_i)\Delta x_i \leqslant \sum_{i=1}^{n} M_i\Delta x_i = S,$$

т.е. для данного разбиения при любой интегральной сумме выполняется неравенство

$$s \leqslant \sigma \leqslant S. \qquad (7.9)$$

Геометрический смысл сумм Дарбу указан на рис. 7.2. Возьмем для простоты непрерывную положительную функцию $f(x)$ на $[a,b]$; тогда m_i и M_i — соответственно ее минимум и максимум на $[x_{i-1}, x_i]$, поскольку $f(x)$ непрерывна на этом частичном отрезке и достигает своих точных нижней и верхней граней. Поэтому сумма S равна площади ступенчатой фигуры, содержащей криволинейную трапецию, ограниченную графиком функции $f(x)$; аналогично сумма s равна площади ступенчатой фигуры, содержащейся в этой криволинейной трапеции.

Заметим, что суммы Дарбу зависят только от разбиения T, тогда как интегральная сумма σ зависит еще и от выбора точек ξ_i на частичных отрезках $[x_{i-1}, x_i]$. Для данного разбиения T отрезка $[a,b]$ сумма σ — переменная величина в силу произвольного выбора ξ_i, а s и S — некоторые числа.

[1] Гастон Дарбу (1842–1917) — французский математик.

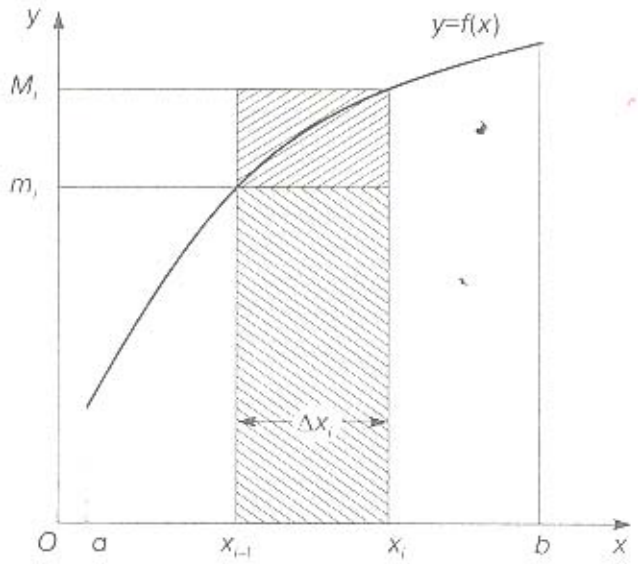

Рис. 7.2

Свойства сумм Дарбу

1°. *Для любого фиксированного разбиения T и для любого $\varepsilon > 0$ точки ξ_i на отрезках $[x_{i-1}, x_i]$ можно выбрать так, что интегральная сумма будет удовлетворять неравенствам $0 < S - \sigma < \varepsilon$. Точки ξ_i можно выбрать и таким образом, что σ будет удовлетворять неравенствам $0 < \sigma - s < \varepsilon$.*

Действительно, пусть T — некоторое фиксированное разбиение отрезка $[a, b]$. В силу свойства точной верхней грани M_i на частичном отрезке $[x_{i-1}, x_i]$ для данного $\varepsilon > 0$ можно указать такую точку ξ_i, что

$$0 < M_i - f(\xi_i) < \frac{\varepsilon}{b-a}, \quad i = 1, 2, \ldots, n.$$

Умножая эти неравенства на Δx_i и затем складывая, получаем $0 < S - \sigma < \varepsilon$. Аналогично можно установить и другие неравенства: $0 < \sigma - s < \varepsilon$.

2°. *При добавлении к данному разбиению T отрезка $[a, b]$ новых точек нижняя сумма Дарбу не убывает, а верхняя — не возрастает.*

Доказательство этого свойства достаточно провести для случая добавления одной точки, так как добавление нескольких точек можно рассматривать как последовательные добавления по одной точке. Пусть эта точка $x' \in [x_{i-1}, x_i]$; обозначим новое разбиение через T', а суммы Дарбу для разбиений T и T' — соответственно через s и s' (нижние) и S и S' (верхние).

Пусть M_i — точная верхняя грань функции $f(x)$ на $[x_{i-1}, x_i]$, а M_i' и M_i'' — точные верхние грани этой функции соответственно на отрезках $[x_{i-1}, x']$ и $[x', x_i]$. Поскольку точная верхняя грань функции на части отрезка не превосходит точной верхней грани этой же функции на всем отрезке, то $M_i' \leqslant M_i$ и $M_i'' \leqslant M_i$. Значит,

$$M_i' \Delta x' + M_i'' \Delta x'' \leqslant M_i \Delta x' + M_i \Delta x'' = M_i \Delta x,$$

откуда следует, что $S' \leqslant S$.

Соответствующие неравенства для нижних сумм Дарбу устанавливаются аналогично.

3°. *Нижняя сумма любого разбиения не превосходит верхней суммы любого другого разбиения.*

Пусть s', S' и s'', S'' — нижняя и верхняя суммы Дарбу, соответствующие разбиениям T' и T''. Возьмем разбиение T, состоящее из всех точек разбиений T' и T''; пусть суммы Дарбу для него будут s и S. Тогда, согласно свойству 2°, сопоставление сумм Дарбу для разбиений T и T' дает

$$s' \leqslant s \leqslant S \leqslant S'.$$

Аналогично сопоставление сумм Дарбу для разбиений T и T'' дает

$$s'' \leqslant s \leqslant S \leqslant S''.$$

Сравнение полученных неравенств приводит к неравенствам

$$s' \leqslant S'', \quad s'' \leqslant S',$$

что и доказывает сформулированное выше свойство.

4°. *Множество $\{S\}$ верхних сумм Дарбу данной функции $f(x)$ для любых разбиений отрезка $[a,b]$ ограничено снизу, а множество $\{s\}$ нижних сумм Дарбу ограничено сверху. При этом точная верхняя грань множества $\{s\}$ не превосходит точную нижнюю грань множества $\{S\}$.*

Действительно, множество $\{S\}$ ограничено снизу, как следует из свойства 3°, любой нижней суммой Дарбу, так же как множество $\{s\}$ ограничено сверху любой верхней суммой Дарбу. Тогда по теореме 1.1 множества $\{S\}$ и $\{s\}$ имеют точные грани — пусть это будут соответственно I^* и I_*:

$$I^* = \inf\{S\}, \quad I_* = \sup\{s\}.$$

Допустим, что $I_* > I^*$, тогда $I_* - I^* = \varepsilon > 0$. Из свойства точных граней следует, что для любого положительного числа, в том числе и для $\varepsilon/2$, существуют такие $S' \in \{S\}$ и $s'' \in \{s\}$, что $I^* + \varepsilon/2 > S'$ и $I_* - \varepsilon/2 < s''$. Вычитая первое неравенство из второго, получаем $s'' - S' > I_* - I^* - \varepsilon = 0$, т.е. $s'' > S'$, что противоречит свойству 3°. Это противоречие и доказывает свойство 4°, следовательно, $I_* \leqslant I^*$.

Необходимое и достаточное условие интегрируемости

Теорема 7.2. *Для того чтобы ограниченная на отрезке $[a,b]$ функция $f(x)$ была интегрируема на этом отрезке, необходимо и достаточно, чтобы*

$$\lim_{\lambda\to 0}(S - s) = 0. \tag{7.10}$$

Доказательство. Сначала докажем необходимость, т.е. пусть функция $f(x)$ интегрируема на $[a,b]$ и существует определенный интеграл (7.6) как предел интегральной суммы (7.3). Покажем, что тогда выполняется равенство (7.10). По второму определению предела для любого $\varepsilon > 0$ существует такое $\sigma > 0$, что для всех разбиений с $\lambda < \sigma$ выполняется неравенство

$$|\sigma - I| < \varepsilon \text{ или } I - \varepsilon < \sigma < I + \varepsilon$$

при любом выборе ξ_i. Поскольку при фиксированном разбиении суммы Дарбу s и S являются точными нижней и верхней гранями для σ, то для них

$$I - \varepsilon \leqslant s \leqslant S \leqslant I + \varepsilon,$$

откуда следует, что

$$\lim_{\lambda\to 0} s = I, \ \lim_{\lambda\to 0} S = I.$$

Таким образом, получаем равенство (7.10).

Теперь докажем достаточность, т.е. пусть выполняется условие (7.10) и нужно показать, что тогда функция $f(x)$ интегрируема на отрезке $[a,b]$, т.е. существует предел (7.6) интегральной суммы (7.3). В этом случае, согласно свойству 4°, $I_* = I^*$. Пусть $I = I_* = I^*$, тогда для любого разбиения

$$s \leqslant I \leqslant S.$$

Если интегральная сумма σ и суммы Дарбу s и S соответствуют одному и тому же разбиению $[a,b]$, то

$$s \leqslant \sigma \leqslant S.$$

Из этих неравенств следует (в силу предельного перехода в неравенствах), что $|\sigma - I| \leqslant S - s$. Поскольку $S - s \to 0$ при $\lambda \to 0$ по условию (7.10), то и $\sigma - I \to 0$ при $\lambda \to 0$ или

$$\lim_{\lambda\to 0}\sigma = I,$$

что и требовалось доказать. Доказательство теоремы завершено.

7.2. Классы интегрируемых функций

Обозначим колебание $M_i - m_i$ функции $f(x)$ на отрезке $[x_{i-1}, x_i]$ через ω_i; тогда

$$S - s = \sum_{i=1}^{n} M_i\Delta x_i - \sum_{i=1}^{n} m_i\Delta x_i = \sum_{i=1}^{n}\omega_i\Delta x_i. \tag{7.11}$$

Необходимое и достаточное условие существования определенного интеграла может быть переписано в виде

$$\lim_{\lambda\to 0}\sum_{i=1}^{n}\omega_i\Delta x_i = 0. \tag{7.12}$$

Непрерывные функции

Теорема 7.3. *Если функция $f(x)$ непрерывна на отрезке $[a,b]$, то она интегрируема на нем.*

Доказательство. По теореме Кантора непрерывная на $[a,b]$ функция $f(x)$ равномерно непрерывна на нем, т.е. для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что для любых точек $x_1, x_2 \in [a,b]$, удовлетворяющих условию $|x_2 - x_1| < \delta$, выполняется неравенство $|f(x_2) - f(x_1)| < \varepsilon/(b-a)$. Разобьем отрезок $[a,b]$ произвольно на конечное число отрезков с длинами $\Delta x_i < \delta$. В силу непрерывности $f(x)$ на $[a,b]$ на каждом из частичных отрезков она достигает своих точных верхних и нижних граней. Следовательно, колебание функции $\omega_i = M_i - m_i$ на данном частичном отрезке удовлетворяет условию $\omega_i < \varepsilon/(b-a)$. Отсюда при $\lambda < \delta$

$$S - s = \sum_{i=1}^{n} \omega_i \Delta x_i < \frac{\varepsilon}{b-a} \sum_{i=1}^{n} \Delta x_i = \varepsilon.$$

Это означает, что $S - s \to 0$ при $\lambda \to 0$, т.е. на $[a,b]$ выполняется достаточное условие интегрируемости функции $f(x)$, что и требовалось доказать.

Разрывные функции

Теорема 7.4. *Если определенная и ограниченная на отрезке $[a,b]$ функция $f(x)$ имеет конечное число точек разрыва, то она интегрируема на этом отрезке.*

Доказательство. Пусть функция $f(x)$ имеет k точек разрыва на $[a,b]$: $x_1, x_2, \ldots, x_k$. Возьмем произвольное разбиение T отрезка $[a,b]$ с условием, чтобы внутри каждого частичного отрезка Δx_i находилось не более одной точки разрыва функции $f(x)$, т.е. таких частичных отрезков будет k (рис. 7.3). Обозначим частичные отрезки, содержащие точки разрыва $f(x)$ через $\Delta x_1', \Delta x_2', \ldots, \Delta x_k'$; частичные отрезки, не содержащие точек разрыва этой функции, будем обозначать через $\Delta x_i''$, M_i'' и m_i'' — соответственно максимум и минимум функции $f(x)$ на каждом из этих частичных отрезков.

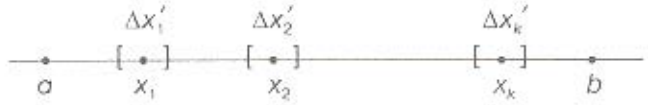

Рис. 7.3

Возьмем произвольное число $\varepsilon > 0$, тогда можно найти такое $\delta_1 > 0$, что при $\lambda = \max\limits_{1 \leqslant i \leqslant n} \Delta x_i < \delta_1$ колебания функции $f(x)$ на каждом из частичных отрезков ее непрерывности, согласно теореме Кантора, будут удовлетворять неравенству $M_i'' - m_i'' = \omega_i'' < \dfrac{\varepsilon}{2(b-a)}$.

Разделим разность сумм Дарбу для функции $f(x)$ на $[a,b]$ на два слагаемых, соответствующих суммированию по совокупности частичных отрезков непрерывности $f(x)$ и по k частичным отрезкам с точками разрыва $f(x)$: $S' - s'$ и $S'' - s''$, т.е.

$$S - s = (S' - s') + (S'' - s'') = \sum_{i=1}^{k} \omega_i' \Delta x_i' + \sum_{i=1}^{n-k} \omega_i'' \Delta x_i''. \quad (7.13)$$

Поскольку $f(x)$ ограничена на $[a,b]$, то ее колебания на частичных отрезках с точками разрыва $\omega_i \leqslant M - m$, где M и m — точные верхняя и нижняя грани функции $f(x)$ на $[a,b]$ ($M \neq m$, иначе $f(x)$ будет постоянной величиной без точек разрыва), $1 \leqslant i \leqslant k$. Возьмем $0 < \delta_2 < \dfrac{\varepsilon}{2k(M-m)}$ и выберем $\delta = \min(\delta_1, \delta_2)$, тогда из выражения (7.13) при $\lambda < \delta$ получаем неравенство

$$S - s < (M-m) \sum_{i=1}^{k} \Delta x_i' + \frac{\varepsilon}{2(b-a)} \sum_{i=1}^{n-k} \Delta x_i'' < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Это неравенство и означает, что $S - s \to 0$ при $\lambda \to 0$, т.е. функция $f(x)$ интегрируема на $[a,b]$, что и требовалось доказать.

Теорема 7.5. *Монотонная на отрезке $[a,b]$ функция $f(x)$ интегрируема на этом отрезке.*

Доказательство. В силу монотонности $f(x)$ на $[a,b]$ ее колебание на этом отрезке равно $f(b) - f(a)$ ($f(a) \neq f(b)$, иначе

$f(x) = \text{const}$). Возьмем произвольное $\varepsilon > 0$ и в любом разбиении T отрезка $[a, b]$ возьмем $\lambda < \dfrac{\varepsilon}{f(b) - f(a)}$. Поскольку для неубывающей функции $\sum\limits_{i=1}^{n} \omega_i = f(b) - f(a)$, то разность сумм Дарбу удовлетворяет неравенству

$$S - s = \sum_{i=1}^{n} \omega_i \Delta x_i < \frac{\varepsilon}{f(b) - f(a)} \sum_{i=1}^{n} \omega_i = \varepsilon,$$

что означает $S - s \to 0$ при $\lambda \to 0$ или интегрируемость функции $f(x)$ на отрезке $[a, b]$. Теорема доказана.

Заметим, что требование монотонности функции здесь весьма существенно. Так, функция Дирихле не является монотонной на любом отрезке $[a, b]$, хотя и ограничена на нем; как было показано ранее, она не интегрируема на $[a, b]$. По сути дела, требование монотонности функции дополняет условие ее ограниченности, которое, согласно теореме 7.1, является необходимым, но не достаточным для интегрируемости функции на отрезке.

7.3. Основные свойства определенного интеграла

В этом параграфе предполагается, что интегралы, входящие в формулы, существуют.

Интеграл $\int\limits_a^b f(x)dx$ был определен для случая, когда $a < b$. Обобщим понятие определенного интеграла и на другие случаи.

1°. *По определению полагаем*

$$\int_a^a f(x)\,dx = 0 \tag{7.14}$$

как определенный интеграл от функции на отрезке нулевой длины, а также

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx. \tag{7.15}$$

Равенство (7.15) верно, поскольку при движении от b к a все длины частичных отрезков $\Delta x_i = x_{i-1} - x_i$ имеют отрицательный знак в интегральной сумме (7.3).

2°. *Для любых чисел a, b и c имеет место равенство*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx. \tag{7.16}$$

Положим сначала, что $a < c < b$. Будем выбирать разбиения отрезка $[a, b]$ таким образом, чтобы точка c была одной из точек разбиения. Пусть $c = x_m$, тогда интегральную сумму σ можно разбить на две суммы:

$$\sigma = \sum_{i=1}^{n} f(\xi_i)\Delta x_i = \sum_{i=1}^{m} f(\xi_i)\Delta x_i + \sum_{i=m+1}^{n} f(\xi_i)\Delta x_i.$$

Переходя к пределу в этом равенстве при $\lambda \to 0$, получим равенство (7.16).

При другом расположении точек a, b и c доказательство сводится к рассмотренному случаю 1°: нужно только выбрать наибольший отрезок интегрирования, разбить интегральную сумму на две суммы, перейти к пределу при $\lambda \to 0$ и переставить пределы интегрирования.

3°. *Постоянный множитель можно выносить за знак определенного интеграла:*

$$\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx. \tag{7.17}$$

4°. *Определенный интеграл от алгебраической суммы функций равен алгебраической сумме их определенных интегралов:*

$$\int_a^b [f(x) \pm g(x)]\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx. \tag{7.18}$$

Свойства 3° и 4° доказываются единообразно исходя из интегральных сумм. Приведем только вывод свойства 4°.

Для любого разбиения отрезка $[a, b]$ и выбора точек ξ_i

$$\sum_{i=1}^{n}[f(\xi_i) \pm g(\xi_i)]\Delta x_i = \sum_{i=1}^{n} f(\xi_i)\Delta x_i \pm \sum_{i=1}^{n} g(\xi_i)\Delta x_i.$$

Перейдем в этом равенстве к пределу при $\lambda \to 0$; поскольку определенные интегралы от всех функций существуют, то в пределе получаем справедливость равенства (7.18).

Заметим, что свойство 4° имеет место для любого конечного числа слагаемых.

5°. $\int\limits_a^b dx = b - a.$

Действительно, по определению $\int\limits_a^b dx = \lim\limits_{\lambda \to 0} \sum\limits_{i=1}^{n} 1 \cdot \Delta x_i = b - a.$

7.4. Формулы оценки определенных интегралов

Оценки интегралов

В данном параграфе будем полагать, что $a < b$.

1°. *Если функция* $f(x) \geqslant 0$ *всюду на отрезке* $[a, b]$, *то*

$$\int\limits_a^b f(x)\,dx \geqslant 0.$$

Действительно, для любого разбиения отрезка $[a, b]$ и при любом выборе точек ξ_i интегральная сумма $\sum\limits_{i=1}^{n} f(\xi_i)\Delta x_i \geqslant 0$ при $b > a$; переход к пределу в этом неравенстве при $\lambda \to 0$ приводит к указанному свойству.

2°. *Если* $f(x) \leqslant g(x)$ *всюду на отрезке* $[a, b]$, *то*

$$\int\limits_a^b f(x)\,dx \leqslant \int\limits_a^b g(x)\,dx. \tag{7.19}$$

Это свойство следует из оценки 1° и свойства определенного интеграла 4° (см. параграф 7.3) применительно к функции $g(x) - f(x) \geqslant 0$.

3°. *Если функция* $f(x)$ *интегрируема на* $[a, b]$, *то*

$$\left|\int\limits_a^b f(x)\,dx\right| \leqslant \int\limits_a^b |f(x)|\,dx. \tag{7.20}$$

Докажем сначала, что функция $|f(x)|$ интегрируема на $[a, b]$. Так как для любых двух точек x_1 и x_2 промежутка $[x_{i-1}, x_i]$ имеем $|f(x_2)| - |f(x_1)| \leqslant |f(x_2) - f(x_1)|$, то и колебание ω_i^* функции $|f(x)|$ не превосходит колебания функции $f(x)$ на этом промежутке. Отсюда получаем, что $0 \leqslant \sum\limits_{i=1}^{n} \omega_i^* \Delta x_i \leqslant \sum\limits_{i=1}^{n} \omega_i \Delta x_i$, и поскольку предел правой части при $\lambda \to 0$ существует, то существует и предел левой части, т.е. $|f(x)|$ интегрируема на отрезке $[a, b]$.

Исходя из неравенств $-|f(x)| \leqslant f(x) \leqslant |f(x)|$, получаем по оценке 2°

$$-\int\limits_a^b |f(x)|\,dx \leqslant \int\limits_a^b f(x)\,dx \leqslant \int\limits_a^b |f(x)|\,dx,$$

а это равносильно неравенству (7.20).

4°. *Если* M *и* m — *соответственно максимум и минимум функции* $f(x)$ *на отрезке* $[a, b]$, *то*

$$m(b - a) \leqslant \int\limits_a^b f(x)\,dx \leqslant M(b - a). \tag{7.21}$$

Действительно, так как по условию $m \leqslant f(x) \leqslant M$ для всех значений x на $[a, b]$, то применение оценки 2° и свойства 5° из параграфа 7.3 к этим неравенствам доказывает двустороннюю оценку интеграла (7.21).

Формулы среднего значения

Теорема 7.6 (ТЕОРЕМА О СРЕДНЕМ). *Пусть функция $f(x)$ интегрируема на $[a,b]$ и числа m и M — соответственно ее точные нижняя и верхняя грани на этом отрезке. Тогда*

$$\int_a^b f(x)\,dx = \mu(b-a), \tag{7.22}$$

где $m \leqslant \mu \leqslant M$.

Доказательство. Для функции $f(x)$ выполняется оценка 4°. Поделив обе части неравенства (7.21) на $b-a$, имеем

$$m \leqslant \frac{1}{b-a}\int_a^b f(x)\,dx \leqslant M.$$

Обозначая через μ число $\dfrac{1}{b-a}\displaystyle\int_a^b f(x)\,dx$, получаем утверждение теоремы (7.22).

Примечание. Равенство (7.22) принимает особенно простой вид, если функция $f(x)$ непрерывна на отрезке $[a,b]$. Тогда по второй теореме Вейерштрасса (см. теорему 3.14) m и M — соответственно наименьшее и наибольшее значения функции; в силу второй теоремы Больцано–Коши о промежуточном значении (см. теорему 3.12) существует такая точка $c \in [a,b]$, что $f(c) = \mu$, откуда

$$\int_a^b f(x)\,dx = f(c)(b-a). \tag{7.23}$$

Равенство (7.23) называется *формулой среднего значения*, а величина $f(c)$ — средним значением функции $f(x)$ на отрезке $[a,b]$.

Формула (7.23) имеет ясный геометрический смысл (рис. 7.4): при $f(x) \geqslant 0$ величина определенного интеграла равна площади прямоугольника с высотой $f(c)$ и основанием $b-a$.

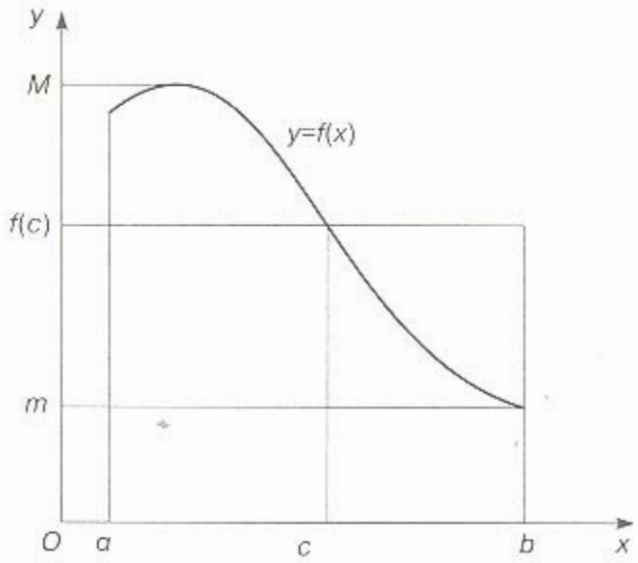

Рис. 7.4

Теорема 7.7 (ОБОБЩЕННАЯ ТЕОРЕМА О СРЕДНЕМ). *Пусть $f(x)$ и $g(x)$ интегрируемы на $[a,b]$, m и M — точные нижняя и верхняя грани функции $f(x)$, причем $g(x)$ не меняет знака на $[a,b]$. Тогда*

$$\int_a^b f(x)g(x)\,dx = \mu\int_a^b g(x)\,dx, \tag{7.24}$$

где $m \leqslant \mu \leqslant M$.

Доказательство. Пусть $g(x) \geqslant 0$ и $a < b$. Так как $m \leqslant f(x) \leqslant M$, то $mg(x) \leqslant f(x)g(x) \leqslant Mg(x)$. Отсюда на основании оценки 2° и свойства 3° получаем

$$m\int_a^b g(x)\,dx \leqslant \int_a^b f(x)g(x)\,dx \leqslant M\int_a^b g(x)\,dx.$$

В силу оценки 1° имеем $\displaystyle\int_a^b g(x)\,dx \geqslant 0$. Если этот интеграл равен нулю, то из предыдущих неравенств следует, что

одновременно также $\int_a^b f(x)g(x)\,dx = 0$ и утверждение теоремы очевидно. Если же $\int_a^b g(x)\,dx > 0$, то, поделив на него обе части двойного неравенства и положив

$$\mu = \left(\int_a^b g(x)\,dx\right)^{-1} \int_a^b f(x)g(x)\,dx,$$

получим утверждение теоремы.

В случае непрерывной на отрезке $[a,b]$ функции $f(x)$ формула (7.24) записывается в виде

$$\int_a^b f(x)g(x)\,dx = f(c)\int_a^b g(x)\,dx, \qquad (7.25)$$

где $c \in [a,b]$.

Формулы (7.23) и (7.25) используются при приближенных оценках определенных интегралов.

7.5. Основная формула интегрального исчисления

Интеграл с переменным верхним пределом

Если функция $f(x)$ интегрируема на отрезке $[a,b]$, то она интегрируема и на отрезке $[a,x]$, где $x \in [a,b]$. Рассмотрим функцию аргумента x

$$F(x) = \int_a^x f(t)\,dt. \qquad (7.26)$$

Назовем функцию $F(x)$ *интегралом с переменным верхним пределом*. В формуле (7.26) переменная интегрирования обозначена буквой t, чтобы избежать путаницы с верхним переменным пределом x.

Теорема 7.8. *Непрерывная на отрезке $[a,b]$ функция $f(x)$ имеет на этом отрезке первообразную. Одной из первообразных является функция*

$$F(x) = \int_a^x f(t)\,dt.$$

Доказательство. Из определения первообразной следует, что нам нужно для доказательства теоремы вывести формулу

$$F'(x) = \left(\int_a^x f(t)\,dt\right)' = f(x) \qquad (7.27)$$

Возьмем любое значение $x \in [a,b)$ и значение $x + \Delta x \in [a,b]$, $\Delta x \neq 0$. Тогда приращение функции $F(x)$ в точке x, с учетом свойства 2 определенного интеграла (см. параграф 7.3),

$$F(x-\Delta x) - F(x) = \int_a^{x-\Delta x} f(t)\,dt - \int_a^x f(t)\,dt =$$

$$= -\int_a^x f(t)\,dt + \int_x^{x-\Delta x} f(t)\,dt = -\int_a^x f(t)\,dt - \int_{x+\Delta x}^x f(t)\,dt.$$

Отсюда, используя формулу среднего значения (7.23), получаем

$$\frac{F(x+\Delta x) - F(x)}{\Delta x} = f(c),$$

где $c \in [x, x+\Delta x]$. При $\Delta x \to 0$ точка $c \to x$ и поскольку функция $f(x)$ непрерывна в точке x, то $f(c) \to f(x)$. Следовательно, переходя в последнем равенстве к пределу при $\Delta x \to 0$, имеем

$$\lim_{\Delta x \to 0} \frac{F(x-\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} f(c), \quad \lim_{c \to x} f(c) = f(x),$$

или $F'(x) = f(x)$. Теорема доказана.

Итак, установлено, что любая непрерывная на отрезке $[a,b]$ функция $f(x)$ имеет первообразную в форме определенного интеграла $F(x)$ с переменным верхним пределом. Поскольку всякая другая первообразная отличается от $F(x)$ на постоянную величину, то связь между неопределенным и определенным интегралами имеет вид

$$\int f(x)\,dx = \int_a^x f(t)\,dt - C,$$

где C — произвольная постоянная.

Как было отмечено в параграфе 6.3, первообразные некоторых элементарных функций не являются элементарными функциями. Поэтому интеграл с переменным верхним пределом используется для определения неэлементарных функций. К их числу относятся, например, функции

$$\int_0^x e^{-t^2}\,dt, \quad \int_0^x \cos t^2\,dt, \quad \int_0^x \frac{\sin t}{t}\,dt.$$

Формула Ньютона-Лейбница

Согласно теореме 7.8, непрерывная на отрезке $[a,b]$ функция $f(x)$ имеет первообразную, которая определяется формулой

$$\int_a^x f(t)\,dt = F(x) + C, \tag{7.28}$$

где C — некоторая постоянная. Подставляя $x = a$ в формулу (7.28), получаем, с учетом свойства 1^c определенного интеграла (см. параграф 7.3),

$$\int_a^a f(t)\,dt = F(a) + C, \quad 0 = F(a) + C, \text{ откуда } C = -F(a).$$

Тогда из выражения (7.28) имеем

$$\int_a^x f(t)\,dt = F(x) - F(a).$$

Полагая теперь $x = b$, получаем формулу

$$\int_a^b f(x)\,dx = F(b) - F(a). \tag{7.29}$$

Равенство (7.29) называется *основной формулой интегрального исчисления*, или *формулой Ньютона–Лейбница*.

Разность $F(b)-F(a)$ условно записывают символом $F(x)\Big|_a^b$, т.е.

$$\int_a^b f(x)dx = F(x)\Big|_a^b. \tag{7.30}$$

Формула (7.30) дает широкие возможности вычисления определенных интегралов. Нужно вычислить неопределенный интеграл и затем найти разность значений первообразной, согласно формуле (7.30).

Рассмотрим примеры.

1. $\int_1^e \frac{dx}{x} = \ln x\Big|_1^e = \ln e - \ln 1 = 1.$

2. $\int_{-1}^1 \frac{dx}{1+x^2} = \operatorname{arctg} x\Big|_{-1}^1 = \operatorname{arctg} 1 - \operatorname{arctg}(-1) = \frac{\pi}{4} - \left(-\frac{\pi}{4}\right) = \frac{\pi}{2}.$

3. $\int_{-1}^2 \frac{dx}{\sqrt{1+x^2}} = \ln\left(x+\sqrt{1+x^2}\right)\Big|_{-1}^2 = \ln\left(2+\sqrt{5}\right) - \ln\left(\sqrt{2}-1\right) =$
$= \ln\frac{2+\sqrt{5}}{\sqrt{2}-1}.$

4. $\int_1^2 \frac{5x^8+1}{x^4}\,dx = \int_1^2 \left(5x^4 + \frac{1}{x^4}\right)dx = \left(x^5 - \frac{1}{3}x^{-3}\right)\Big|_1^2 = \frac{751}{24}.$

5. $\int_0^{\pi/2} \cos^2\frac{x}{2}\,dx = \frac{1}{2}\int_0^{\pi/2}(1+\cos x)\,dx = \frac{1}{2}(x+\sin x)\Big|_0^{\pi/2} = \frac{1}{2}\left(\frac{\pi}{2}+1\right).$

7.6. Основные правила интегрирования

Замена переменной в определенном интеграле

Теорема 7.9. *Пусть:* 1) $f(x)$ — *непрерывная функция на отрезке* $[a,b]$; 2) *функция* $\varphi(t)$ *дифференцируема на* $[\alpha,\beta]$, *причем* $\varphi'(t)$ *непрерывна на* $[\alpha,\beta]$ *и множеством значений функции* $\varphi(t)$ *является отрезок* $[a,b]$; 3) $\varphi(\alpha) = a$, $\varphi(\beta) = b$. *Тогда справедлива формула*

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f[\varphi(t)]\varphi'(t)\,dt. \tag{7.31}$$

Доказательство. Пусть $F(x)$ — какая-либо первообразная для функции $f(x)$ на $[a,b]$. Рассмотрим сложную функцию $\Phi(t) = F[\varphi(t)]$; так как $F(x)$ и $\varphi(t)$ дифференцируемы на соответствующих отрезках, то по правилу дифференцирования сложной функции находим

$$\Phi'(t) = F'[\varphi(t)]\varphi'(t) = f[\varphi(t)]\varphi'(t).$$

Из этого равенства следует, что функция $\Phi(t)$ является первообразной для функции $f[\varphi(t)]\varphi'(t)$, непрерывной на отрезке $[\alpha, \beta]$. По формуле Ньютона–Лейбница

$$\int_\alpha^\beta f[\varphi(t)]\varphi'(t)\,dt = \Phi(\beta) - \Phi(\alpha) =$$

$$= F[\varphi(\beta)] - F[\varphi(\alpha)] = F(b) - F(a) = \int_a^b f(x)\,dx.$$

Из этого равенства следует справедливость формулы (7.31).

Формула (7.31) называется *формулой замены переменной* или *подстановки в определенном интеграле*.

Заметим, что при вычислении определенного интеграла с помощью замены переменной нет нужды возвращаться к прежней переменной, как при вычислении неопределенного интеграла, так как определенный интеграл представляет собой число, которое, согласно формуле (7.31), равно значению каждого из рассматриваемых интегралов. Теперь при подстановке следует сначала найти новые пределы интегрирования и выполнить необходимые преобразования подынтегральной функции.

Заметим также, что при замене переменной в определенном интеграле необходимо обращать внимание на соблюдение условий теоремы 7.9, иначе можно получить неверный результат (особое внимание следует уделять выполнению условия 2 теоремы).

Вычислим определенные интегралы методом подстановки.

Пример 1. Найти $\int_0^1 \frac{2x\,dx}{1+x^2}$.

Выполним подстановку $t = 1 + x^2$. Тогда $dt = 2x\,dx$, $t = 1$ при $x = 0$ и $t = 2$ при $x = 1$. Поскольку функция $x = \sqrt{1-t^2}$ непрерывна на $[1, 2]$, то и новая подынтегральная функция также непрерывна, и для нее в силу теоремы 7.9 существует первообразная на этом отрезке. Получаем

$$\int_0^1 \frac{2x\,dx}{1+x^2} = \int_1^2 \frac{dt}{t} = \ln t\Big|_1^2 = \ln 2.$$

Пример 2. Найти $\int_0^a \sqrt{a^2 - x^2}\,dx$.

Применим здесь подстановку $x = a\sin t$. Тогда $dx = a\cos t\,dt$, $\sqrt{a^2 - x^2} = a\cos t$, $t = \arcsin\frac{x}{a}$, $t = 0$ при $x = 0$, $t = \frac{\pi}{2}$ при $x = a$. Подставляя все это в исходный интеграл, получаем

$$\int_0^a \sqrt{a^2 - x^2}\,dx = a^2 \int_0^{\pi/2} \cos^2 t\,dt = a^2 \int_0^{\pi/2} \frac{1}{2}(1 + \cos 2t)\,dt =$$

$$= \frac{a^2}{2}\left(t + \frac{\sin 2t}{2}\right)\Big|_0^{\pi/2} = \frac{\pi a^2}{4}.$$

Пример 3. Найти $\int_0^\pi dx$.

По формуле Ньютона–Лейбница

$$\int_0^\pi dx = x\Big|_0^\pi = \pi.$$

Вычислим этот интеграл при помощи замены переменной $t = \operatorname{tg} x$. Тогда $t = 0$ при $x = 0$ и $t = 0$ при $x = \pi$, $x = \operatorname{arctg} t$, т.е. $dx = dt/(1 + t^2)$. Подстановка в исходный интеграл дает

$$\int_0^\pi dx = \int_0^0 \frac{dt}{1+t^2} = 0.$$

Полученное противоречие объясняется тем, что функция замены переменной $t = \operatorname{tg} x$ имеет разрыв при $x = \pi/2$ и не удовлетворяет условию 2 теоремы 7.9.

Интегрирование по частям в определенном интеграле

Теорема 7.10. *Пусть функции $u(x)$ и $v(x)$ имеют непрерывные производные на отрезке $[a, b]$; тогда справедлива формула*

$$\int_a^b u\,dv = uv\Big|_a^b - \int_a^b v\,du. \qquad (7.32)$$

14 Зак. 6853

Доказательство. Поскольку функция $u(x)v(x)$ является первообразной для функции $[u(x)v(x)]' = u'(x)v(x) + u(x)v'(x)$, то по формуле Ньютона–Лейбница

$$\int_a^b [u'(x)v(x) + u(x)v'(x)]\,dx = [u(x)v(x)]\Big|_a^b.$$

С учетом свойства 4° определенных интегралов (см. параграф 7.3) и вида дифференциалов функций

$$du = u'(x)\,dx, \quad dv = v'(x)\,dx$$

получаем, после переноса первого слагаемого в правую часть указанного равенства, формулу (7.32). Теорема доказана.

Равенство (7.32) называется *формулой интегрирования по частям в определенном интеграле.*

Вычислим определенные интегралы методом интегрирования по частям.

Пример 4. Найти $\int_0^1 xe^{-x}\,dx$.

Положим здесь $u = x$, $v = e^{-x}$ или $dv = -e^{-x}\,dx$; получаем

$$\int_0^1 xe^{-x}\,dx = -\int_0^1 x\,d(e^{-x}) = -xe^{-x}\Big|_0^1 + \int_0^1 e^{-x}dx =$$

$$= -e^{-1} - e^{-x}\Big|_0^1 = -e^{-1} - e^{-1} + 1 = 1 - 2e^{-1}.$$

Пример 5. Найти $\int_0^\pi x\sin x\,dx$.

Здесь $u = x$, $\sin x\,dx = dv$ или $v = -\cos x$; далее по формуле (7.32)

$$\int_0^\pi x\sin x\,dx = -\int_0^\pi x\,d(\cos x) = -x\cos x\Big|_0^\pi + \int_0^\pi \cos x\,dx =$$

$$= -\pi(-1) + \sin x\Big|_0^\pi = \pi.$$

Пример 6. Найти $\int_0^{1/2} \operatorname{arctg} 2x\,dx$.

Положим здесь $u = \operatorname{arctg} 2x$, $v = x$; тогда $du = 2x\,dx/(1+4x^2)$, и по формуле (7.32), с последующей подстановкой $t = 1 + 4x^2$, имеем

$$\int_0^{1/2} \operatorname{arctg} 2x\,dx = x\operatorname{arctg} 2x\Big|_0^{1/2} - \int_0^{1/2} \frac{2x\,dx}{1+4x^2} =$$

$$= \frac{\pi}{8} - \frac{1}{4}\ln(1+4x^2)\Big|_0^{1/2} = \frac{1}{8}(\pi - \ln 4).$$

7.7. Геометрические приложения определенного интеграла

Площадь плоской фигуры

Рассмотрим на плоскости Oxy фигуру, ограниченную графиком непрерывной и положительной функции $f(x)$ на отрезке $[a, b]$, отрезком $[a, b]$ и вертикальными прямыми $x = a$ и $x = b$ (рис. 7.5).

Разобьем отрезок $[a, b]$ на n частей произвольно выбранными точками $a = x_0 < x_1 < x_2 < \cdots < x_{i-1} < x_i \cdots < x_n = b$. На

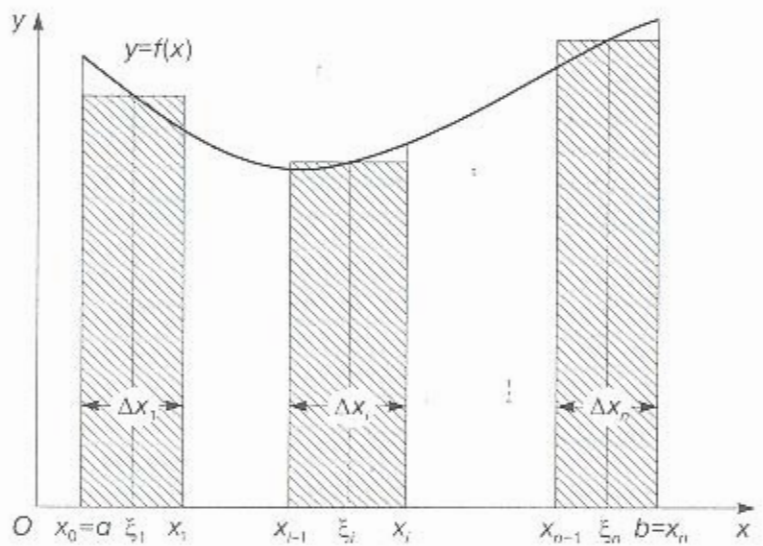

Рис. 7.5

каждом частичном отрезке $[x_{i-1}, x_i]$ возьмем произвольно точку ξ_i. Тогда площадь S криволинейной трапеции, ограниченной сверху графиком функции $f(x)$, приближенно равна площади ступенчатой фигуры, составленной из прямоугольников с основаниями Δx_i и высотами $f(\xi_i)$ (на рис. 7.5 показаны такие фрагменты):

$$S \approx \sum_{i=1}^{n} f(\xi_i)\Delta x_i. \qquad (7.33)$$

Очевидно, что при $\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\Delta x_i\} \to 0$ площадь ступенчатой фигуры стремится к площади криволинейной трапеции; однако величина площади ступенчатой фигуры, выражаемая правой частью формулы (7.33), является интегральной суммой (7.3) для функции $f(x)$ на отрезке $[a, b]$. Поскольку $f(x)$ непрерывна на $[a, b]$, то в силу теоремы 7.3 предел правой части формулы (7.33) существует и равен определенному интегралу от функции $f(x)$ на $[a, b]$. Отсюда следует, что *величина площади криволинейной трапеции равна определенному интегралу от функции $f(x)$ на отрезке $[a, b]$:*

$$S = \lim_{\lambda \to 0} \sigma = \lim_{\lambda \to 0} \sum_{i=1}^{n} f(\xi_i)\Delta x_i = \int_a^b f(x)\,dx. \qquad (7.34)$$

Если фигура ограничена сверху и снизу неотрицательными функциями $f(x)$ и $g(x)$ соответственно, непрерывными на отрезке $[a, b]$ (рис. 7.6), то площадь S криволинейной фигуры равна разности площадей криволинейных трапеций, ограниченных сверху графиками $f(x)$ и $g(x)$:

$$S = \int_a^b f(x)\,dx - \int_a^b g(x)\,dx = \int_a^b [f(x) - g(x)]\,dx. \qquad (7.35)$$

В том случае, когда верхняя граница криволинейной трапеции задана параметрическими уравнениями

$$x = \varphi(t),\ y = \psi(t),\ \alpha \leqslant t \leqslant \beta,\ \varphi(\alpha) = a,\ \psi(\beta) = b,$$

в определенном интеграле формулы (7.34) следует сделать замену переменной $x = \varphi(t)$, $dx = \varphi'(t)\,dt$. Тогда, согласно фор-

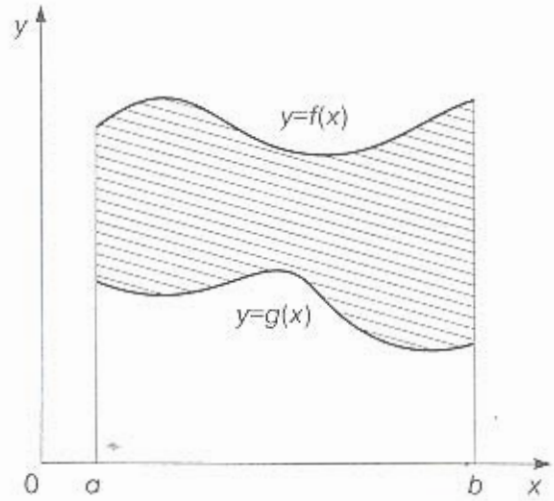

Рис. 7.6

муле (7.31), получим соответствующее выражение для площади фигуры:

$$S = \int_\alpha^\beta \psi(t)\varphi'(t)\,dt. \qquad (7.36)$$

Рассмотрим несколько примеров.

Пример 1. Найти площадь фигуры, ограниченной графиком функции $y = \ln x \geqslant 0$, осью Ox и прямой $x = 2$.

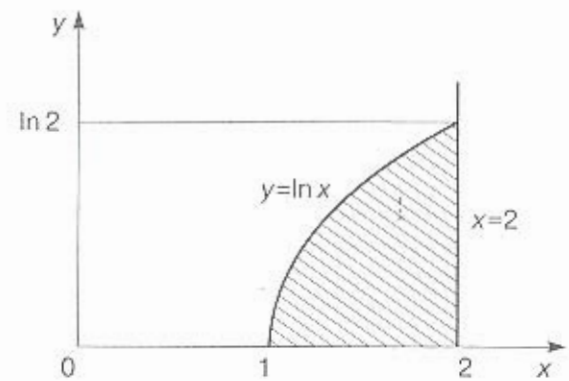

Рис. 7.7

Отрезок интегрирования $1 \leqslant x \leqslant 2$ (рис. 7.7), так что искомая площадь, согласно формуле (7.34), равна

$$S = \int_1^2 \ln x\, dx = x \ln x \Big|_1^2 - \int_1^2 1\, dx = 2\ln 2 - x \Big|_1^2 = 2\ln 2 - 1.$$

Пример 2. Найти площадь фигуры, ограниченной линиями $y = \sqrt{x}$, $y = x^2$.

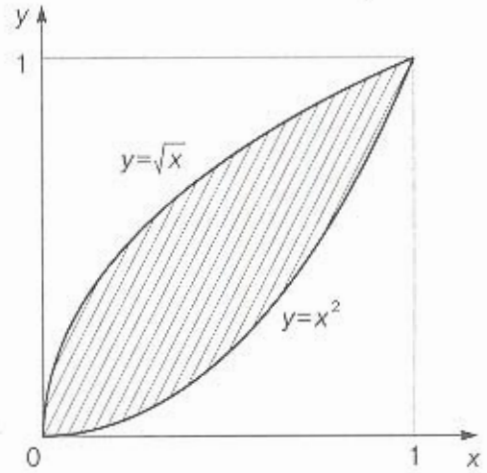

Рис. 7.8

Вычислим абсциссы точек пересечения указанных кривых, для чего приравняем правые части этих уравнений: $x^2 = \sqrt{x}$. Корни этого уравнения $x_1 = 0$, $x^2 = 1$. Следовательно, площадь фигуры, ограниченной сверху функцией $y = \sqrt{x}$ и снизу функцией $y = x^2$ (рис. 7.8), дается определенным интегралом на отрезке $[0, 1]$:

$$S = \int_0^1 \left(\sqrt{x} - x^2\right) dx = \left(\frac{2}{3}x^{3/2} - \frac{1}{3}x^3\right)\Big|_0^1 = \frac{1}{3}.$$

Пример 3. Найти площадь фигуры, ограниченной одной аркой циклоиды [линия, описываемая фиксированной точкой M на окружности радиуса a при качении окружности по оси Ox (рис. 7.9)]:

$$x = a(t - \sin t), \quad y = a(1 - \cos t), \qquad 0 \leqslant t \leqslant 2\pi,$$

и осью Ox.

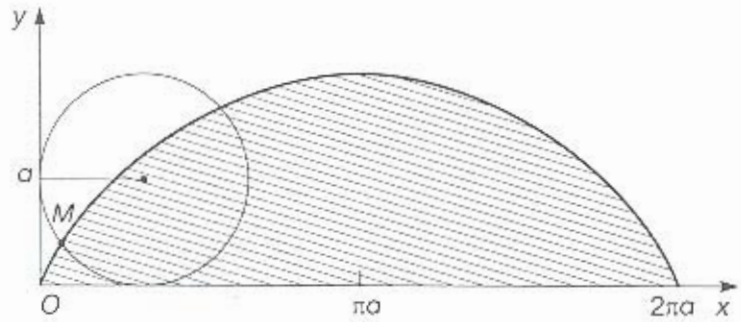

Рис. 7.9

Так как $dx = a(1 - \cos t)\, dt$, то по формуле (7.36) получаем

$$S = a^2 \int_0^{2\pi} (1 - \cos t)^2\, dt = a^2 \int_0^{2\pi} \left(\frac{3}{2} - 2\cos t + \frac{1}{2}\cos 2t\right) dt =$$

$$= a^2 \left(\frac{3}{2}t - 2\sin t + \frac{1}{4}\sin 2t\right)\Big|_0^{2\pi} = 3\pi a^2.$$

Площадь криволинейного сектора

Чаще всего (после прямоугольной системы координат) используют полярную систему координат. Она состоит из некоторой точки O, называемой *полюсом*, и *полярной оси* — фиксированного луча OA, исходящего из полюса (рис. 7.10). В этой системе *полярными координатами* произвольной точки M на плоскости называются числа ρ и φ — соответственно расстояние точки M до полюса, или *полярный радиус*, и угол между полярной осью и лучом OM, или *полярный угол*, отсчитываемый против часовой стрелки.

Точка M с полярными координатами ρ и φ обозначается: $M(\rho, \varphi)$. Полярный радиус может быть любым неотрицательным числом:

$$0 \leqslant \rho < +\infty.$$

Полярный угол чаще всего полагают изменяющимся в пределах:

$$0 \leqslant \varphi \leqslant 2\pi,$$

однако в ряде случаев рассматривают углы и вне указанного промежутка.

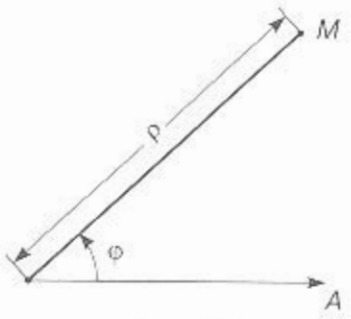

Рис. 7.10

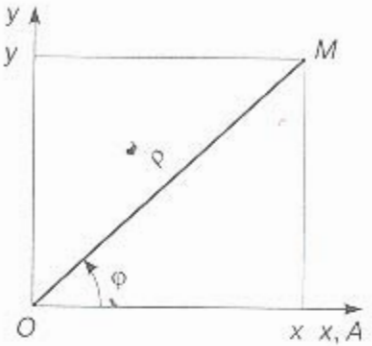

Рис. 7.11

Связь между полярными координатами точки на плоскости и ее прямоугольными координатами устанавливается следующим образом. Совместим начало прямоугольной системы координат с полюсом, а ось Ox — с полярной осью (рис. 7.11). Тогда прямоугольные координаты x и y точки M выражаются через ее полярные координаты ρ и φ по формулам:

$$x = \rho\cos\varphi, \quad y = \rho\sin\varphi. \tag{7.37}$$

Из формул (7.37) можно выразить полярные координаты через прямоугольные:

$$\rho = \sqrt{x^2 + y^2}, \quad \operatorname{tg}\varphi = y/x. \tag{7.38}$$

Вторая из формул определяет два значения полярного угла φ, так как он изменяется в пределах от нуля до 2π. Из них выбирают то, которое удовлетворяет формулам (7.37). В полярной системе координат уравнение линии имеет вид $\rho = \rho(\varphi)$. Полярные координаты удобны тем, что многие сложные кривые описываются в них простыми уравнениями; например, уравнение окружности радиуса a имеет вид $\rho = a$.

Пусть кривая MN задана в полярной системе координат уравнением $\rho = \rho(\varphi)$, где $\alpha \leqslant \varphi \leqslant \beta$, причем функция $\rho(\varphi)$ неотрицательна на $[\alpha, \beta]$. *Криволинейным сектором* называется плоская фигура, ограниченная кривой $\rho(\varphi)$ и лучами, направленными от полярной оси под углами α и β (рис. 7.12).

Путем разбиения отрезка $[\alpha, \beta]$ на частичные отрезки и приближения отдельных "долек" круговыми секторами можно

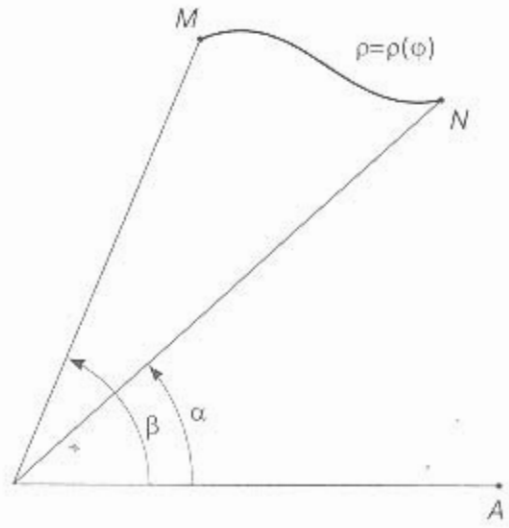

Рис. 7.12

показать, что площадь S криволинейного сектора выражается определенным интегралом

$$S = \frac{1}{2}\int_{\alpha}^{\beta} \rho^2(\varphi)\, d\varphi. \tag{7.39}$$

Рассмотрим примеры.

Пример 4. Найти площадь фигуры, ограниченной полярной осью и первым витком спирали Архимеда: $\rho = a\varphi$, где $a > 0$.

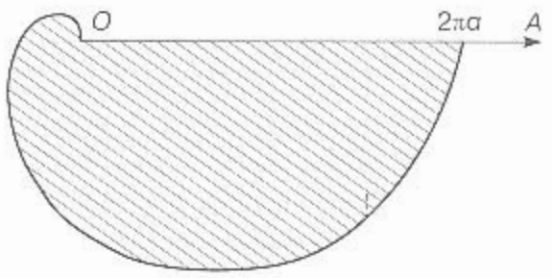

Рис. 7.13

Первый виток спирали Архимеда описывается полярным радиусом при изменении φ от 0 до 2π. Тогда, согласно форму-

ле (7.39), получаем для площади заштрихованной фигуры на рис. 7.13

$$S = \frac{a^2}{2}\int_0^{2\pi} \varphi^2\, d\varphi = \frac{a^2}{2}\frac{\varphi^3}{3}\Big|_0^{2\pi} = \frac{4}{3}\pi^3 a^2.$$

Пример 5. Вычислить площадь кардиоиды (рис. 7.14), описываемой уравнением

$$\rho = a(1+\cos\varphi), \quad 0 \leqslant \varphi \leqslant 2\pi,\ a > 0.$$

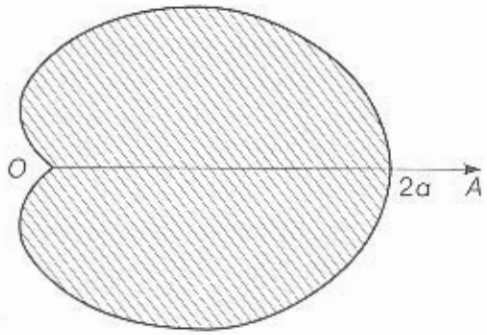

Рис. 7.14

По формуле (7.39) получаем

$$S = \frac{a^2}{2}\int_0^{2\pi}(1+\cos\varphi)^2\, d\varphi = \frac{a^2}{2}\int_0^{2\pi}(1+2\cos\varphi+\cos^2\varphi)\, d\varphi =$$

$$= \frac{a^2}{2}\int_0^{2\pi}\left(\frac{3}{2}+2\cos\varphi+\frac{1}{2}\cos 2\varphi\right) d\varphi =$$

$$= \frac{a^2}{2}\left(\frac{3}{2}\varphi+2\sin\varphi+\frac{1}{4}\sin 2\varphi\right)\Big|_0^{2\pi} = \frac{3}{2}\pi a^2.$$

Длина дуги плоской кривой

Пусть незамкнутая кривая AB на плоскости Oxy задана уравнением $y = f(x)$, где $f(x)$ — функция с непрерывной производной на отрезке $[a, b]$. Разделим кривую AB на n произвольных частей; соединив соседние точки хордами, получим

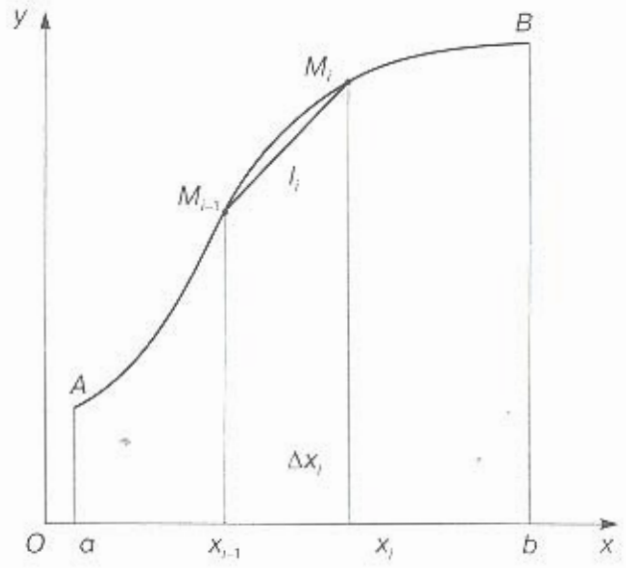

Рис. 7.15

вписанную в AB ломаную, длину которой обозначим через l. Пусть l_i — длина одной хорды ломаной (рис. 7.15), а $\mu = \max\limits_{1\leqslant i\leqslant n}\{l_i\}$.

Определение. Число L, равное пределу при $\mu \to 0$ длины ломаной, вписанной в дугу AB, называется *длиной* этой *дуги плоской кривой:*

$$L = \lim_{\mu\to 0} l = \lim_{\mu\to 0}\sum_{i=1}^{n}\Delta l_i. \tag{7.40}$$

Теперь исходя из этого определения выведем формулу для длины дуги плоской кривой. Координаты точек M_i разбиения дуги на многозвенную ломаную равны $(x_i, f(x_i))$, где $a = x_0 < x_1 < x_2 < \cdots < x_n = b$. Длина одного звена

$$l_i = \sqrt{(\Delta x_i)^2 + [f(x_i) - f(x_{i-1})]^2} \quad \text{где} \quad \Delta x_i = x_i - x_{i-1},$$

по формуле Лагранжа

$$f(x_i) - f(x_{i-1}) = f'(\xi_i)\Delta x_i, \quad \xi_i \in [x_{i-1}, x_i],$$

откуда

$$l_i = \sqrt{1 + [f'(\xi_i)]^2}\Delta x_i.$$

Следовательно, длина всей ломаной, вписанной в дугу AB,

$$l = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} \sqrt{1 + [f'(\xi_i)]^2}\Delta x_i.$$

В правой части этого равенства стоит интегральная сумма для непрерывной на отрезке $[a, b]$ функции $\sqrt{1 + [f'(x)]^2}$. Так как $\lambda \leqslant \mu \left(\lambda = \max\limits_{1 \leqslant i \leqslant n} \{\Delta x_i\}\right)$, то $\lambda \to 0$ при $\mu \to 0$, и, значит, предел l при $\lambda \to 0$ существует и равен определенному интегралу, который и представляет собой длину дуги AB

$$L = \lim_{\lambda \to 0} l = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx. \qquad (7.41)$$

Если кривая AB задана параметрическими уравнениями $x = \varphi(t)$, $y = \psi(t)$, $\alpha \leqslant t \leqslant \beta$, $\varphi(\alpha) = a$, $\varphi(\beta) = b$, то в интеграле (7.41) надо сделать замену переменной $x = \varphi(t)$, $dx = \varphi'(t)\, dt$. С учетом вида производной от функции, заданной параметрически,

$$f'(x) = \frac{\psi'(t)}{\varphi'(t)}$$

получаем формулу для длины дуги:

$$L = \int_\alpha^\beta \sqrt{1 + \left[\frac{\psi'(t)}{\varphi'(t)}\right]^2} \varphi'(t)\, dt = \int_\alpha^\beta \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\, dt. \qquad (7.42)$$

В случае когда кривая AB задана в полярных координатах уравнением $\rho = \rho(\varphi)$, $\alpha \leqslant \varphi \leqslant \beta$, при условии существования непрерывной производной $\rho'(\varphi)$ на отрезке $[\alpha, \beta]$, нужно перейти от полярных координат к прямоугольным координатам по формулам (7.37). Тогда получим параметрические уравнения кривой AB с параметром φ:

$$x = \rho(\varphi)\cos\varphi, \quad y = \rho(\varphi)\sin\varphi, \quad \alpha \leqslant \varphi \leqslant \beta.$$

Так как

$$x'(\varphi) = \rho'(\varphi)\cos\varphi - \rho(\varphi)\sin\varphi,$$
$$y'(\varphi) = \rho'(\varphi)\sin\varphi + \rho(\varphi)\cos\varphi,$$

то после подстановки в формулу (7.42) и приведения подобных получаем

$$L = \int_\alpha^\beta \sqrt{\rho^2(\varphi) + [\rho'(\varphi)]^2}\, d\varphi. \qquad (7.43)$$

Рассмотрим примеры на вычисление длин дуг.

Пример 6. Найти длину дуги кривой $y = \ln\cos x$ при $0 \leqslant x \leqslant \pi/6$.

Вычисляем производную $y' = -\operatorname{tg} x$ и подставляем ее в формулу (7.41):

$$L = \int_0^{\pi/6} \sqrt{1 + \operatorname{tg}^2 x}\, dx = \int_0^{\pi/6} \frac{dx}{\cos x} =$$
$$= \int_0^{\pi/6} \frac{\cos x\, dx}{1 - \sin^2 x} = \frac{1}{2}\ln\left(\frac{1 + \sin x}{1 - \sin x}\right)\Big|_0^{\pi/6} = \frac{1}{2}\ln 3.$$

Пример 7. Вычислить длину дуга одной арки циклоиды: $x = a(t - \sin t)$, $y = a(1 - \cos t)$, $0 \leqslant t \leqslant 2\pi$ (см. рис. 7.9).

Из параметрических уравнений находим: $\varphi'(t) = a(1 - \cos t)$, $\psi'(t) = a\sin t$; используя формулу (7.42), получаем

$$L = \int_0^{2\pi} a\sqrt{(1 - \cos t)^2 + \sin^2 t}\, dt = a\int_0^{2\pi} \sqrt{2 - 2\cos t}\, dt =$$
$$= 2a\int_0^{2\pi} \sin\frac{t}{2} dt = -4a\cos\frac{t}{2}\Big|_0^{2\pi} = 8a.$$

Пример 8. Вычислить длину дуги кардиоиды $\rho = a(1 + \cos\varphi)$, $0 \leqslant \varphi \leqslant 2\pi$ (см. рис. 7.14).

Поскольку кривая является симметричной относительно полярной оси, найдем длину половины дуги, т.е. при $0 \leqslant \varphi \leqslant \pi$. Используя формулу (7.43), получаем

$$L = 2\int_0^{\pi} \sqrt{\rho^2(\varphi)+[\rho'(\varphi)]^2}\,d\varphi = 2a\int_0^{\pi} \sqrt{(1+\cos\varphi)^2+\sin^2\varphi}\,d\varphi =$$

$$= 2a\int_0^{\pi} \sqrt{2+2\cos\varphi}\,d\varphi = 4a\int_0^{\pi} \cos\frac{\varphi}{2}\,d\varphi = 8a\sin\frac{\varphi}{2}\Big|_0^{\pi} = 8a.$$

Объем тела вращения

Рассмотрим тело, которое образуется при вращении вокруг оси Ox криволинейной трапеции, ограниченной сверху непрерывной и положительной на отрезке $[a, b]$ функции $f(x)$ (рис. 7.16). Разобьем отрезок $[a, b]$ произвольным образом на n частей и на каждом i-м частичном отрезке произвольно выберем точку ξ_i. При вращении кривой $f(x)$ вокруг оси Ox каждый прямоугольник с высотой $f(\xi_i)$ и с основанием $\Delta x_i = x_i - x_{i-1}$ описывает цилиндр с радиусом $f(\xi_i)$ и высотой Δx_i. Сумма объемов всех цилиндров приближенно равна объему тела вращения и легко подсчитывается:

$$V \approx \sum_{i=1}^{n} \pi\,[f(\xi_i)]^2\,\Delta x_i.$$

Нетрудно видеть, что это выражение является интегральной суммой для непрерывной на $[a, b]$ функции $\pi f^2(x)$, т.е. предел ее при $\lambda = \max\limits_{1\leqslant i\leqslant n}\{\Delta x_i\} \to 0$ существует и равен определенному интегралу. С другой стороны, при $\lambda \to 0$ объем ступенчатого тела вращения стремится к объему V тела вращения. Таким образом, получаем формулу

$$V = \lim_{\lambda\to 0}\sum_{i=1}^{n} \pi\,[f(\xi_i)]^2\,\Delta x_i = \pi\int_a^b f^2(x)\,dx. \qquad (7.44)$$

Если тело образовано вращением криволинейной трапеции вокруг оси Oy, то, выражая x через y как обратную функцию,

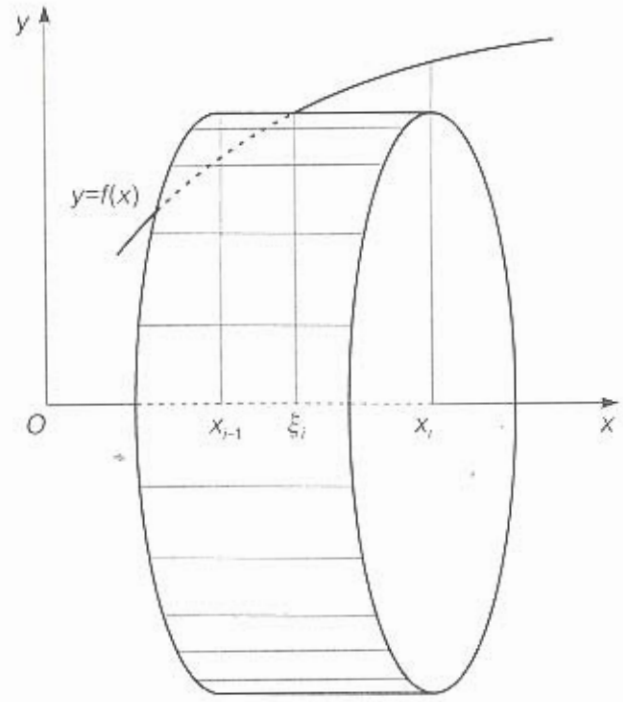

Рис. 7.16

мы можем получить аналогичным образом формулу для объема тела вращения:

$$V = \pi\int_c^d x^2(y)\,dy, \qquad (7.45)$$

где $[c, d]$ — область изменения функции $y = f(x)$.

Вычислим объемы тел, образованных вращением фигур.

Пример 9. Найти объем тела вращения вокруг оси Ox фигуры, ограниченной следующими линиями: $y = x^2$, $y = \sqrt{x}$ (см. рис. 7.8).

Искомый объем вращения равен разности объемов, образованных вращением криволинейных трапеций с верхними границами соответственно $y = \sqrt{x}$ и $y = x^2$. Пределы интегрирования определяются по точкам пересечения этих кривых: $a = 0$ и $b = 1$. По формуле (7.44) получаем

$$V = \pi \int_0^1 \left(\sqrt{x}\right)^2 dx - \pi \int_0^1 \left(x^2\right)^2 dx = \pi \int_0^1 x dx - \pi \int_0^1 x^4 dx =$$

$$= \pi \left(\frac{x}{2} - \frac{x^5}{5} \right) \Big|_0^1 = 0,3\pi.$$

Пример 10. Найти объем тела вращения вокруг оси Oy фигуры, ограниченной следующими линиями: $y = e^x$, $x = 0$, $x = 1$, $y = 0$.

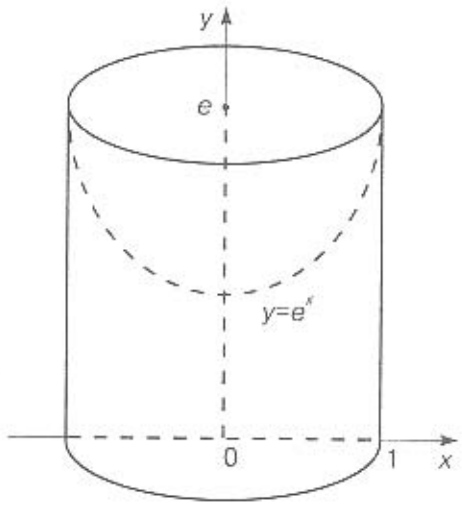

Рис. 7.17

Выражаем x через y: $x = \ln y$; промежуток интегрирования $[1, e]$ определяется очевидным образом. Объем тела вращения (рис. 7.17) равен разности объемов соответственно цилиндра с радиусом 1 и высотой e и тела вращения вокруг оси Oy криволинейной трапеции, ограниченной сверху кривой $x = \ln y$. Согласно формуле (7.45), получаем

$$V = \pi e - \pi \int_1^e \ln^2 y \, dy = \pi e - \pi y \ln^2 y \Big|_1^e + \pi \int_1^e 2 \ln y \, dy =$$

$$= 2\pi y \ln y \Big|_1^e - \pi \int_1^e 2 \, dy = 2\pi e - 2\pi e + 2\pi = 2\pi.$$

Площадь поверхности вращения

Пусть функция $f(x)$ неотрицательна и имеет непрерывную производную на отрезке $[a, b]$. Докажем, что площадь поверхности, образованной вращением графика функции вокруг оси Ox, вычисляется по формуле

$$P = 2\pi \int_a^b f(x) \sqrt{1 + [f'(x)]^2} \, dx. \qquad (7.46)$$

Возьмем произвольное разбиение отрезка $[a, b]$ на частичные отрезки $\Delta x_i = x_i - x_{i-1} \neq 0$, $i = 1, 2, \ldots, n$; $x_0 = a$, $x_n = b$. Пусть $M_0, M_1, M_2, \ldots, M_n$ — соответствующие точки на графике функции $f(x)$. Проведем хорды, соединяющие соседние точки; при вращении полученной ломаной вокруг оси Ox образуется поверхность, составленная из боковых поверхностей усеченных конусов. Площадь боковой поверхности i-го усечен-

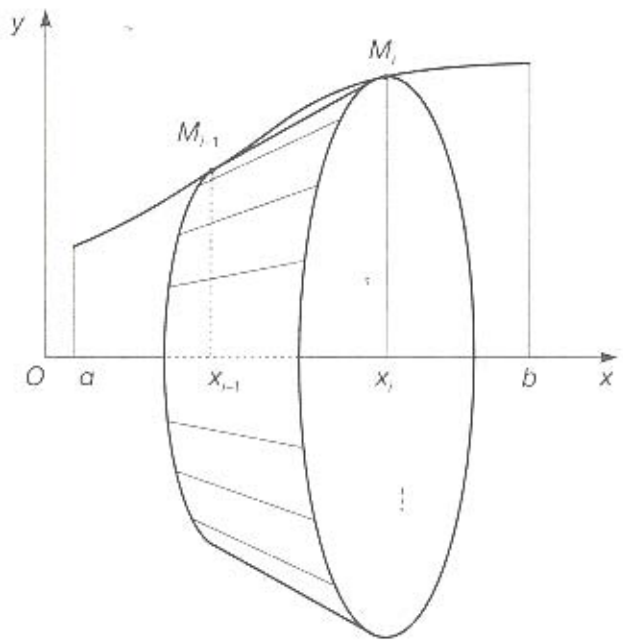

Рис. 7.18

ного конуса (рис. 7.18) равна

$$2\pi[f(x_{i-1}) + f(x_i)]l_i/2,$$

где l_i — длина хорды $M_{i-1}M_i$, определяемая формулой расстояния между двумя точками:

$$l_i = \sqrt{(x_i - x_{i-1})^2 + [f(x_i) - f(x_{i-1})]^2}.$$

Но по формуле Лагранжа

$$f(x_i) - f(x_{i-1}) = f'(\xi_i)\Delta x_i, \quad \xi_i \in [x_{i-1}, x_i],$$

откуда получаем, что

$$l_i = \sqrt{1 + [f'(\xi_i)]^2}\,\Delta x_i.$$

Таким образом, площадь поверхности вращения P приближенно равна площади поверхности, образованной от вращения ломаной,

$$P \approx 2\sum_{i=1}^{n} \frac{\pi}{2}[f(x_{i-1}) + f(x_i)]\sqrt{1 + [f'(\xi_i)]^2}\,\Delta x_i.$$

Поскольку функция $f(x)$ дифференцируема на отрезке $[a, b]$, значит, на каждом частичном отрезке $[x_{i-1}, x_i]$ она может быть разложена по формуле Тейлора в окрестности точки ξ_i с остаточным членом первого порядка. Представим такие разложения для $f(x_{i-1})$ и $f(x_i)$:

$$f(x_{i-1}) = f(\xi_i) + f'(\theta_{i1})(\xi_i - x_{i-1}), \quad \theta_{i1} \in [x_{i-1}, x_i],$$
$$f(x_i) = f(\xi_i) + f'(\theta_{i2})(\xi_i - x_i), \quad \theta_{i1} \in [x_{i-1}, x_i].$$

Подставим эти разложения в приближенную формулу для P и представим сумму в правой части в виде двух сумм:

$$P \approx 2\pi\sum_{i=1}^{n} f(\xi_i)\sqrt{1 + [f'(\xi_i)]^2}\,\Delta x_i + 2\pi\sum\nolimits_r. \qquad (7.47)$$

Первое слагаемое является интегральной суммой для непрерывной на отрезке $[a, b]$ функции $f(x)\sqrt{1 + [f'(x)]^2}$ и потому имеет предел в виде определенного интеграла при $\lambda = \max\limits_{1 \leqslant i \leqslant n}\{\Delta x_i\} \to 0$. Оценим второе слагаемое $\sum_r$:

$$\sum\nolimits_r = \frac{1}{2}\sum_{i=1}^{n}[f'(\theta_{i1})(\xi_i - x_{i-1}) + f'(\theta_{i2})(\xi_i - x_i)]\sqrt{1 + [f'(\xi_i)]^2}\,\Delta x_i.$$

Поскольку $f'(x)$ непрерывна на $[a, b]$, то она ограничена всюду на этом отрезке, т.е. можно указать такое число $A > 0$, что $|f'(x)| < A$ для любого $x \in [a, b]$. Так как $\xi_i \in [x_{i-1}, x_i]$, то справедливы оценки

$$|\xi_i - x_{i-1}| \leqslant \Delta x_i \leqslant \lambda, \quad |\xi_i - x_i| \leqslant \Delta x_i \leqslant \lambda.$$

Отсюда получаем общую оценку $\sum_r$:

$$\left|\sum\nolimits_r\right| < A\left|\sum_{i=1}^{n}\sqrt{1 + [f'(\xi_i)]^2}\,\Delta x_i\right|\lambda.$$

Под знаком модуля стоит интегральная сумма для непрерывной на $[a, b]$ функции, т.е. существует ее конечный предел при $\lambda \to 0$. Значит, $\sum_r \to 0$ при $\lambda \to 0$. Стало быть, переход к пределу в формуле (7.47) приводит к формуле (7.46), что и требовалось доказать.

Если кривая, вращением которой образуется поверхность, задана параметрически, т.е. $x = \varphi(t)$, $y = \psi(t)$, $\alpha \leqslant t \leqslant \beta$, и при изменении t от α до β функция $\varphi(t)$ изменяется от a до b, $\varphi(\alpha) = a$, $\varphi(\beta) = b$, то при замене переменной $x = \varphi(t)$ в формуле (7.46) получаем

$$P = 2\pi\int_{\alpha}^{\beta}\psi(t)\sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\,dt. \qquad (7.48)$$

Наконец, если кривая задана в полярных координатах со всеми необходимыми требованиями, то формула (7.46) будет иметь вид

$$P = 2\pi\int_{\alpha}^{\beta}\rho(\varphi)\sin\varphi\sqrt{\rho^2(\varphi) + [\rho'(\varphi)]^2}\,d\varphi. \qquad (7.49)$$

Вычислим площади поверхности, образованной вращением линий вокруг оси Ox.

Пример 11. Найти площадь поверхности вращения вокруг оси Ox дуги кривой $y = x^3/3$ от $x = 0$ до $x = 1$.

Подставляя в формулу (7.46) $f(x) = x^3/3$ и $f'(x) = x^2$, получаем

$$P = 2\pi \int_0^1 \frac{x^3}{3}\sqrt{1+x^4}\,dx = \frac{\pi}{9}(1+x^4)^{3/2}\Big|_0^1 = \frac{\pi}{9}(2\sqrt{2}-1).$$

Пример 12. Найти площадь поверхности вращения вокруг оси Ox циклоиды $x = a(t - \sin t)$, $y = a(1 - \cos t)$, $0 \leqslant t \leqslant 2\pi$.

Применение формулы (7.48) дает

$$P = 2\pi a^2 \int_0^{2\pi} (1-\cos t)\sqrt{2-2\cos t}\,dt = 2\sqrt{2}a^2 \int_0^{2\pi} (1-\cos t)^{3/2}\,dt =$$

$$= 2\sqrt{2}a^2 \int_0^{2\pi} \left(2\sin^2\frac{t}{2}\right)^{3/2} dt = 16\pi a^2 \left[\frac{1}{3}\cos^3\frac{t}{2} - \cos\frac{t}{2}\right]\Big|_0^{2\pi} = \frac{64\pi a^2}{3}.$$

Пример 13. Найти площадь поверхности вращения вокруг оси Ox кардиоиды $\rho = a(1 + \cos\varphi)$.

Предварительно вычисляем

$$\rho'(\varphi) = -a\sin\varphi, \quad \rho^2 + (\rho')^2 = 2a^2(1+\cos\varphi) = 4a^2\cos^2\frac{\varphi}{2}.$$

Подставляя в формулу (7.49), получаем

$$P = 4\pi a^2 \int_0^{\pi} (1+\cos\varphi)\sin\varphi\cos\frac{\varphi}{2}\,d\varphi =$$

$$= 16\pi a^2 \int_0^{\pi} \cos^4\frac{\varphi}{2}\sin\frac{\varphi}{2}\,d\varphi = -32\pi a^2 \frac{\cos^5(\varphi/2)}{5}\Big|_0^{\pi} = \frac{64}{5}\pi a^2.$$

7.8. Некоторые приложения определенного интеграла в экономике

Вообще говоря, в экономических задачах переменные меняются дискретно. Для использования определенного интеграла нужно составить некоторую идеализированную модель, предполагающую непрерывное изменение зависимых переменных (функций) и независимых переменных (аргумента). Рассмотрим соответствующие примеры.

Пример 1. Найти дневную выработку P за рабочий день продолжительностью 8 часов, если производительность труда в течение дня меняется по эмпирической формуле

$$p = f(t) = p_0(-0,2t^2/t_0^2 + 1,6t/t_0 + 3),$$

где t — время (ч), p_0 — размерность производительности (единица продукции/ч), t_0 — размерность времени (ч). Эта формула вполне отражает реальный процесс работы (рис 7.19): производительность сначала растет, достигая максимума в середине рабочего дня при $t = 4$ ч, а затем падает.

Решение. Полагая, что производительность меняется в течение дня непрерывно, т.е. что p является непрерывной функ-

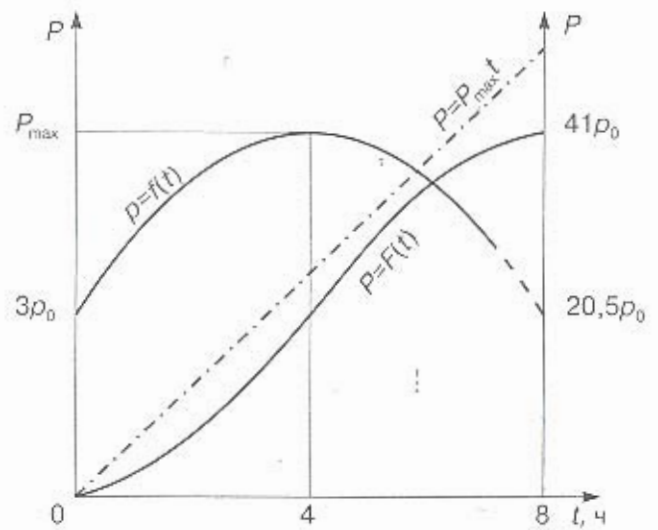

Рис. 7.19

цией аргумента t на отрезке $[0, 8]$, дневную выработку P можно выразить определенным интегралом:

$$p = p_0 t_0 \int_0^8 (-0{,}2t^2/t_0^2 + 1{,}6t/t_0 + 3)\, dt = F(t)\Big|_0^8 =$$

$$= p_0 t_0 \left(-0{,}2\frac{t^3}{3}/t_0^2 + 0{,}8t^2/t_0 + 3t\right)\Big|_0^8 \approx$$

$$\approx 41{,}07 p_0 t_0 = 41{,}07 a_0,$$

где a_0 — множитель, имеющий размерность единицы продукции.

Если бы в течение всего дня работа велась ритмично и с максимальной производительностью $p_{\max} = 6{,}2p_0$, то дневная выработка составила бы $P_{\max} = 49{,}6a_0$, или примерно на 21% больше. Рисунок 7.19 иллюстрирует решение задачи: дневная выработка численно равна площади криволинейной трапеции, ограниченной сверху кривой $f(t)$; вторая кривая показывает рост выпуска продукции во времени (график первообразной $F(t)$ соответствует правой оси ординат P). Значение $t = 4$ ч соответствует точке перегиба кривой $F(t)$: в первой половине рабочего дня интенсивность выработки продукции выше, чем во второй. Штрихпунктирная прямая $P = p_{\max} t$ соответствует выпуску продукции с равномерной производительностью $p_{\max}$.

Пример 2. Производство оборудования некоторого вида характеризуется темпом роста его выпуска

$$K = \frac{\Delta y}{\Delta t}\frac{1}{y},$$

где Δy — прирост выпуска этого оборудования за промежуток времени Δt, а y — уровень его производства за единицу времени на момент времени t. Найти общее количество оборудования, произведенного к моменту времени t, полагая, что K — известная постоянная величина, единицей времени является год, а в начальный момент времени $t = 0$ уровень ежегодного производства оборудования составлял y_0.

Решение. Перейдем к пределу при $\Delta t \to 0$, полагая, что он существует. Будем также полагать, что y является непрерывной функцией от времени t. Получаем

$$K = \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t}\frac{1}{y} = \frac{y'}{y} = (\ln y)'.$$

Интегрируя это равенство в пределах от 0 до t, получаем

$$\ln y\Big|_{y_0}^{y} = Kt\Big|_0^t, \quad \text{или} \quad \ln\frac{y}{y_0} = Kt, \quad \text{откуда} \quad y = y_0 e^{Kt}.$$

Суммарное количество оборудования, выпущенного за время t, дается определенным интегралом

$$Y(t) = \int_0^t y(t)dt = \int_0^t y_0 e^{Kt} dt = \frac{1}{K} y_0 e^{Kt}\Big|_0^t = \frac{1}{K} y_0 (e^{Kt} - 1).$$

Например, при $K = 0{,}05$ (5% ежегодного темпа роста) общее количество оборудования, выпущенного за 10 лет, составит

$$Y(10) = 20 y_0 (e^{0.5} - 1) \approx 13 y_0,$$

причем уровень производства за указанный период увеличится почти на 65%.

При определении экономической эффективности капитальных вложений встречаются так называемые *задачи дисконтирования:* определение начальной суммы S через время t по ее конечной величине S_t при процентной ставке ρ.

Как было рассмотрено ранее (см. параграф 2.3), при непрерывном начислении процента конечная сумма вычисляется по формуле $S_t = Se^{rt}$, где $r = 0{,}01p$. Если сумма S_t также является функцией времени $f(t)$, то дисконтированная сумма к моменту времени t составит $S = f(t)e^{-rt}$.

Полная дисконтированная сумма за время t вычисляется по формуле

$$S_d = \int_0^t f(t)e^{-rt} dt.$$

Пример 3. Определим дисконтированную сумму S_d при $f(t) = S_0(1 + kt)$, где S_0 — начальные капиталовложения, k — ежегодная доля их увеличения. Иными словами, при заданных

величинах p и k требуется оценить, что выгоднее: наращивать капиталовложения или вложить их одновременно при непрерывно начисляемой процентной ставке.

Решение. Вычисляя указанный интеграл методом интегрирования по частям, получаем

$$S_d = \int_0^t S_0(1+kt)e^{-rt}\,dt = S_0\left\{-\frac{1}{r}e^{-rt}\Big|_0^t - \frac{k}{r}\int_0^t t\,d(e^{-rt})\right\} =$$

$$= S_0\left\{\frac{1}{r}(1-e^{-rt}) - \frac{k}{r}te^{-rt}\Big|_0^t + \frac{k}{r}\int_0^t e^{-rt}dt\right\} =$$

$$= S_0\left\{\frac{1}{r}\left(1+\frac{k}{r}\right) - \frac{1}{r}\left(1+kt+\frac{k}{r}\right)e^{-rt}\right\}.$$

Из полученной формулы можно сделать некоторые выводы:

1. Чем выше процентная ставка p (а значит, и r), тем меньше дисконтная сумма S_d и, следовательно, выше доход, вычисляемый как разность между суммой ежегодно растущих капиталовложений за t лет и величиной S_d. Если рассматривать S_d как дисконтный доход, то увеличение процентной ставки p снижает рентабельность помещения капитала.

2. Увеличение интенсивности ежегодных капиталовложений (k) приводит к увеличению S_d.

3. При неизменных p и k дисконтный доход растет с увеличением промежутка времени t (количества лет).

Например, при процентной ставке $p = 5\%$ ($r = 0{,}05$) и при ежегодном увеличении капиталовложений на 5% ($k = 0{,}05$) получаем, что за 5 лет дисконтная сумма $S_d \approx 5S_0$, в то время как сумма ежегодных капиталовложений за этот период составит $S_a = 6{,}25S_0$; при $k = 0{,}1$ (10%-ное увеличение ежегодных капиталовложений) и тех же самых p и t соответственно имеем $S_d \approx 5{,}5S_0$ и $S_a = 6{,}5S_0$, т.е. в первом случае разность $S_a - S_d \approx 1{,}25S_0$, тогда как во втором случае она снижается до $\approx S_0$.

При аналогичных оценочных расчетах в реальных условиях следует учитывать существенную роль темпа инфляции, который, в первую очередь, определяет приемлемую величину промежутка времени t. Очевидно, что при высоком уровне инфляции выгодны только краткосрочные капиталовложения.

7.9. Приближенное вычисление определенных интегралов

При решении ряда прикладных задач чаще всего приходится сталкиваться с вычислением "неберущихся" определенных интегралов (когда первообразные не выражаются через элементарные функции). Существует много формул приближенного вычисления определенных интегралов; здесь мы приведем две наиболее простые из них: *формулу трапеций* и *формулу Симпсона*[1].

Формула трапеций

Пусть в интеграле $\int_a^b f(x)\,dx$ функция $f(x)$ непрерывна на отрезке $[a,b]$. Разобьем отрезок $[a,b]$ на n равных частей точками (узлами) $x = x_i$ $(0 \leqslant i \leqslant n)$, как показано на рис. 7.20. Проведем прямые $x = x_i$ и соединим соседние точки их пересечения с кривой $f(x)$ хордами, т.е. построим n трапеций с основаниями $y_{i-1} = f(x_{i-1})$ и $y_i = f(x_i)$ и высотой $\dfrac{b-a}{n}$ каждая $(0 \leqslant i \leqslant n)$. Сумма площадей этих прямоугольных трапеций приближенно равна площади криволинейной трапеции или искомому интегралу:

$$\int_a^b f(x)\,dx \approx \left[\frac{f(x_0)+f(x_1)}{2} + \frac{f(x_1)+f(x_2)}{2} + \cdots + \right.$$
$$\vdots \qquad (7.50)$$
$$\left. + \frac{f(x_{n-1})+f(x_n)}{2}\right]\frac{b-a}{n} = \frac{b-a}{n}\left\{\frac{f(a)+f(b)}{2} + \sum_{i=1}^{n-1} f(x_i)\right\}.$$

[1] Томас Симпсон (1710–1761) — английский математик.

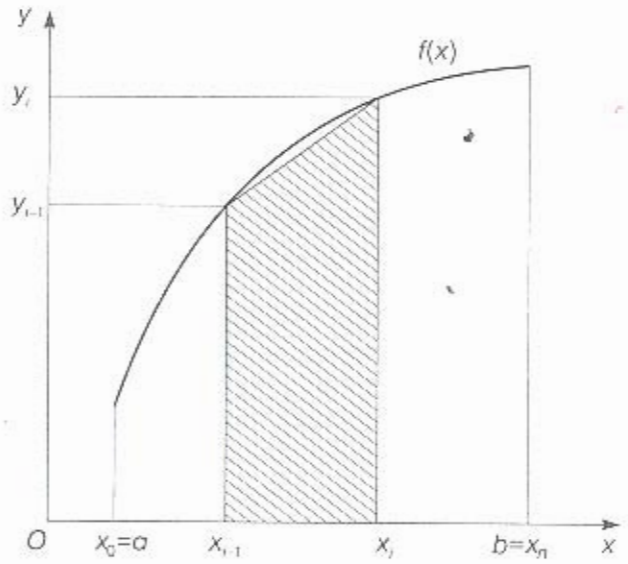

Рис. 7.20

Рассмотрим в качестве примера интеграл, точное значение которого вычисляется достаточно просто:

$$\int_0^1 \frac{dx}{1+x^2} = \operatorname{arctg} x\Big|_0^1 = \frac{\pi}{4} = 0{,}7853981\ldots .$$

Применим к вычислению этого интеграла формулу трапеций при $n = 5$. Получаем: $x_0 = 0$, $x_1 = 0{,}2$, $x_2 = 0{,}4$, $x_3 = 0{,}6$, $x_4 = 0{,}8$, $x_5 = 1{,}0$. Тогда соответственно значения функции в этих узлах: $f(x_0) = 1$, $f(x_5) = 0{,}5$; $f(x_1) = 0{,}9615$, $f(x_2) = 0{,}8621$, $f(x_3) = 0{,}7353$ и $f(x_4) = 0{,}6098$ (с точностью до четырех знаков). Подставляя эти значения в формулу (7.50), получаем

$$\int_0^1 \frac{dx}{1+x^2} \approx \frac{1}{5}\left[\frac{1{,}5}{2} + 0{,}9615 + 0{,}8621 + 0{,}7353 + 0{,}6098\right] =$$

$$= 0{,}7837.$$

Абсолютная ошибка полученного результата как модуль разности между точным и приближенным значениями интеграла составила примерно 0,00156. Если увеличить число n узлов, то формула (7.50) дает большую точность; так, при $n = 10$

$$\int_0^1 \frac{dx}{1+x^2} \approx 0{,}78498,$$

что отличается от точного значения уже менее чем на 0,0005. Вообще, абсолютная величина погрешности вычислений определенного интеграла по формуле трапеции оценивается по формуле

$$\delta \leqslant \max_{a \leqslant x \leqslant b} |f''(x)| \frac{(b-a)^3}{12n^2}, \qquad (7.51)$$

если функция $f(x)$ имеет на отрезке $[a, b]$ непрерывную вторую производную.

Формула Симпсона

Основной принцип построения формул приближенного вычисления определенных интегралов состоит в замене частичных криволинейных трапеций, образующихся при разбиении отрезка интегрирования, на более простые фигуры. В основе формулы Симпсона лежит замена двух соседних частичных криволинейных трапеций, ограниченных сверху функцией $f(x)$ (рис. 7.21), на криволинейную трапецию, ограниченную сверху параболой вида

$$y = Ax^2 + Bx + C. \qquad (7.52)$$

Именно поэтому формулу Симпсона часто называют *формулой парабол*.

Разобьем отрезок $[a, b]$ на $2n$ равных частей, тогда кривая $f(x)$ разобьется прямыми $x = x_i$ $(i = 0, 1, \ldots, 2n)$ на $2n$ соответствующих частей точками $M_0, M_1, M_2, \ldots, M_{2i}, M_{2i+1}, M_{2i+2}, \ldots, M_{2n}$. Через каждую тройку точек начиная с M_0:

$$M_0, M_1, M_2;\ M_2, M_3, M_4;\ \ldots;\ M_{2i}, M_{2i+1}, M_{2i+2};\ldots$$
$$\ldots;\ M_{2n-2}, M_{2n-1}, M_{2n}$$

проведем параболу вида (7.52). Коэффициенты A, B и C для каждой параболы находятся из условия ее прохождения че-

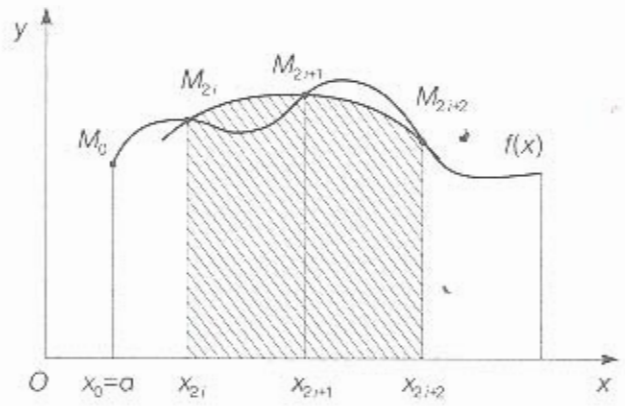

Рис. 7.21

рез соответствующую тройку точек. Таким образом, криволинейная трапеция, ограниченная сверху функцией $f(x)$, заменяется составной фигурой, ограниченной сверху n параболами; на рис. 7.21 показан фрагмент такой замены: заштрихованная площадь ограничена сверху параболой. В полном курсе высшей математики выводится формула Симпсона (парабол) приближенного вычисления определенного интеграла как приближенного значения площади криволинейной трапеции, ограниченной сверху функцией $f(x)$:

$$\int_a^b f(x)dx \approx \frac{b-a}{6n}\sum_{i=0}^{n-1}(y_{2i}+4y_{2i+1}+y_{2i+2}) = \\ = \frac{b-a}{6n}[y_0+y_{2n}+2(y_2+y_4+\cdots+y_{2n-2})+ \\ +4(y_1+y_3+\cdots+y_{2n-1})], \tag{7.53}$$

где y_i — значение функции $y=f(x)$ в i-м узле.

Погрешность вычислений по формуле Симпсона оценивается как

$$\delta \leqslant \max_{a\leqslant x\leqslant b}\left|f^{(4)}(x)\right|\frac{(b-a)^5}{2880n^4}. \tag{7.54}$$

Здесь предполагается, что функция $f(x)$ имеет на отрезке $[a,b]$ непрерывную четвертую производную. Сравнивая оценку точности вычислений (7.51) определенного интеграла по формуле

трапеций с аналогичной оценкой (7.54) по формуле Симпсона, можно увидеть, что при одном и том же числе n формула Симпсона гораздо более точная.

В качестве примера применения формулы Симпсона рассмотрим тот же интеграл, который вычислялся по формуле трапеций:

$$\int_0^1 \frac{dx}{1+x^2} = 0{,}7853981\ldots.$$

Возьмем $n=2$, так что полное число узлов, включая концы отрезка, равно 5. Выпишем таблицу значений абсцисс узлов и соответствующих ординат с точностью до шести знаков:

i	0	1	2	3	4
x_i	0	0,25	0,5	0,75	1,0
y_i	1	0,941176	0,8	0,64	0,5

Подстановка в формулу (7.53) дает

$$\int_0^1 \frac{dx}{1+x^2} \approx \frac{1}{12}[1+0{,}5+2\cdot 0{,}8+4(0{,}941176+0{,}64)] = \\ = 0{,}785392,$$

что с точностью до шестого знака совпадает со значением $\pi/4$.

7.10. Несобственные интегралы

При рассмотрении определенного интеграла как предела интегральных сумм предполагалось, что подынтегральная функция ограничена на конечном отрезке интегрирования. Данное выше определение определенного интеграла не имеет смысла, если не выполняется хотя бы одно из этих условий. Нельзя разбить бесконечный интервал на конечное число отрезков конечной длины; при неограниченной функции интегральная сумма не имеет предела. Тем не менее возможно обобщить понятие определенного интеграла и на эти случаи, с чем и связано понятие *несобственного интеграла*.

Интегралы с бесконечными пределами интегрирования

Определение 1. Пусть функция $f(x)$ определена на промежутке $[a, +\infty)$ и интегрируема на любом отрезке $[a, R]$, $R > 0$, так что интеграл

$$\int_a^R f(x)\,dx$$

имеет смысл. Предел этого интеграла при $R \to \infty$ называется *несобственным интегралом первого рода* и обозначается

$$\int_a^{\infty} f(x)\,dx = \lim_{R\to+\infty} \int_a^R f(x)\,dx. \qquad (7.55)$$

В случае если этот предел конечен, говорят, что несобственный интеграл (7.55) *сходится*, а функцию $f(x)$ называют *интегрируемой на бесконечном промежутке* $[a, \infty)$; если же предел в формуле (7.55) бесконечен или не существует, то говорят, что несобственный интеграл *расходится*.

Аналогичным образом вводится понятие несобственного интеграла по промежутку $(-\infty, b]$:

$$\int_{-\infty}^b f(x)\,dx = \lim_{R\to-\infty} \int_R^b f(x)\,dx. \qquad (7.56)$$

Наконец, несобственный интеграл с двумя бесконечными пределами можно определить как сумму несобственных интегралов (7.55) и (7.56):

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^c f(x)\,dx + \int_c^{\infty} f(x)\,dx, \qquad (7.57)$$

где c — любое число.

Геометрический смысл несобственного интеграла первого рода основан на геометрической интерпретации определенного

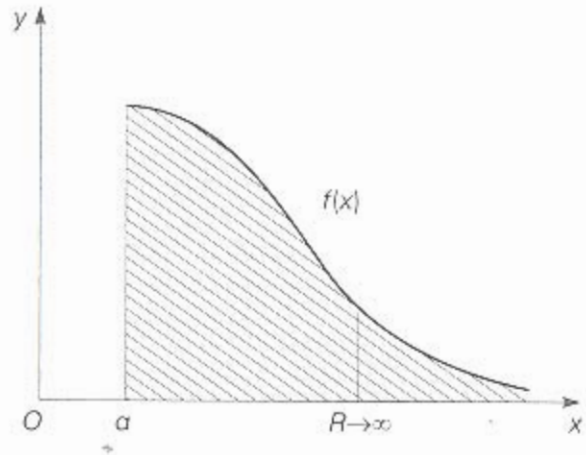

Рис. 7.22

интеграла на отрезке $[a, R]$: это площадь бесконечной области, ограниченной сверху неотрицательной функцией $f(x)$, снизу — осью Ox, слева — прямой $x = a$ (рис. 7.22). Такая же интерпретация имеет место и для несобственных интегралов (7.56) и (7.57).

Рассмотрим несколько примеров несобственных интегралов первого рода.

Пример 1. $\int_0^{\infty} \cos x\,dx = \lim_{R\to\infty} \int_0^R \cos x\,dx = \lim_{R\to\infty} \sin x\Big|_0^R =$ $= \lim_{R\to\infty} \sin R$, но предела функции $\sin R$ при $R \to \infty$ не существует, т.е. данный интеграл расходится.

Пример 2. $\int_0^{\infty} e^{-x}\,dx = \lim_{R\to\infty} \int_0^R e^{-x}\,dx = -\lim_{R\to\infty} e^{-x}\Big|_0^{-R} =$ $= -\lim_{R\to\infty} e^{-R} + 1 = 1$, т.е. данный интеграл сходится.

Пример 3. $\int_0^{\infty} xe^{-x^2}\,dx = \lim_{R\to\infty} \int_0^R xe^{-x^2}\,dx =$

$= -\frac{1}{2} \lim_{R\to\infty} \int_0^R e^{-x^2}\,d(-x^2) = -\frac{1}{2} \lim_{R\to\infty} e^{-x^2}\Big|_0^R = -\frac{1}{2} \lim_{R\to\infty} e^{-R^2} +$ $+\frac{1}{2} = 0 + \frac{1}{2} = \frac{1}{2}$. Конечный предел существует — значит, исходный интеграл сходится.

Пример 4. $\int\limits_{-\infty}^{\infty}\frac{dx}{1+x^2}=\int\limits_{-\infty}^{0}\frac{dx}{1+x^2}+\int\limits_{0}^{\infty}\frac{dx}{1+x^2}=\lim\limits_{A\to-\infty}\int\limits_{A}^{0}\frac{dx}{1+x^2}+$

$+\lim\limits_{B\to\infty}\int\limits_{0}^{B}\frac{dx}{1+x^2}=\lim\limits_{A\to-\infty}\operatorname{arctg}x\Big|_{A}^{0}+\lim\limits_{B\to\infty}\operatorname{arctg}x\Big|_{0}^{B}=$

$=-\lim\limits_{A\to-\infty}\operatorname{arctg}A+\lim\limits_{B\to\infty}\operatorname{arctg}B=-\left(-\frac{\pi}{2}\right)+\frac{\pi}{2}=\pi$. Здесь пришлось разделить исходный интеграл на два и к каждом из них применить определение несобственного интеграла первого рода.

Пример 5. Найти $\int\limits_{1}^{\infty}\frac{dx}{x^\alpha}$, где α — некоторое положительное число.

Рассмотрим разные случаи для числа α:

а) при $\alpha=1$ для любого $R>0$

$$\int\limits_{1}^{\infty}\frac{dx}{x}=\lim_{R\to\infty}\int\limits_{1}^{R}\frac{dx}{x}=\lim_{R\to\infty}\ln x\Big|_{1}^{R}=\lim_{R\to\infty}\ln R-0=+\infty,$$

т.е. конечного предела не существует и несобственный интеграл расходится;

б) при $\alpha\neq1$ для любого $R>0$

$$\int\limits_{1}^{\infty}\frac{dx}{x^\alpha}=\lim_{R\to\infty}\int\limits_{1}^{R}\frac{dx}{x^\alpha}=\lim_{R\to\infty}\frac{x^{1-\alpha}}{1-\alpha}\Big|_{1}^{R}=$$

$$=\frac{1}{1-\alpha}\left(\lim_{R\to\infty}R^{1-\alpha}-1\right)=\begin{cases}\dfrac{1}{\alpha-1} & \text{при}\ \alpha>1,\\ \infty & \text{при}\ \alpha<1.\end{cases}$$

Следовательно, данный интеграл сходится при $\alpha>1$ и расходится при $\alpha\leqslant1$.

В приведенных выше примерах сначала с помощью первообразной вычислялся интеграл по конечному промежутку, а затем осуществлялся переход к пределу. Между тем, если для функции $f(x)$ существует первообразная $F(x)$ на всем промежутке интегрирования $[a,\infty)$, по формуле Ньютона–Лейбница

$$\int\limits_{a}^{R}f(x)\,dx=F(R)-F(a)=F(x)\Big|_{a}^{R}.$$

Отсюда следует, что несобственный интеграл существует (сходится) в том и только в том случае, если существует конечный предел

$$\lim_{R\to\infty}F(R)=F(\infty),$$

и тогда можно записать:

$$\int\limits_{a}^{\infty}f(x)\,dx=F(x)\Big|_{a}^{\infty}=F(\infty)-F(a).\qquad(7.58)$$

Аналогичный вывод справедлив и для несобственных интегралов вида (7.56) и (7.57):

$$\int\limits_{-\infty}^{b}f(x)\,dx=F(x)\Big|_{-\infty}^{b},\quad\int\limits_{-\infty}^{\infty}f(x)\,dx=F(x)\Big|_{-\infty}^{+\infty}.\qquad(7.59)$$

Пример 6. Вычислить несобственный интеграл $\int\limits_{2}^{\infty}\frac{dx}{x^2-1}$.

Подынтегральная функция представляет собой правильную рациональную дробь; разлагая ее на элементарные дроби, найдем первообразную:

$$\int\limits_{2}^{\infty}\frac{dx}{x^2-1}=\frac{1}{2}\int\limits_{2}^{\infty}\left(\frac{1}{x-1}-\frac{1}{x+1}\right)dx=$$

$$=\frac{1}{2}\ln\left(\frac{x-1}{x+1}\right)\Big|_{2}^{\infty}=0-\frac{1}{2}\ln\frac{1}{3}=\ln\sqrt{3}.$$

Интегралы от неограниченных функций

Рассмотрим функцию $f(x)$, определенную на промежутке $[a,b)$, но неограниченную на нем. Для определенности положим, что $f(x)$ ограничена и интегрируема на любом отрезке $[a,b-\varepsilon]$, $0<\varepsilon<b-a$, но неограничена в любой окрестности точки b или на промежутке $[b-\varepsilon,b]$. В таком случае точка b называется *особой точкой*.

16 Зак. 6833

Определение 2. Предел интеграла $\int\limits_a^{b-\varepsilon} f(x)\,dx$ при $\varepsilon \to 0$ называется *несобственным интегралом второго рода* и обозначается

$$\int\limits_a^b f(x)\,dx = \lim_{\varepsilon\to 0} \int\limits_a^{b-\varepsilon} f(x)\,dx. \tag{7.60}$$

Если этот предел конечный, то говорят, что интеграл (7.60) *существует* или *сходится*, а функцию $f(x)$ называют *интегрируемой на промежутке* $[a,b)$; если же предела нет или он бесконечен, то говорят, что интеграл (7.60) *расходится*.

Аналогично, если особой является точка $x=a$, то несобственный интеграл второго рода определяется как

$$\int\limits_a^b f(x)\,dx = \lim_{\varepsilon\to 0} \int\limits_{a-\varepsilon}^b f(x)\,dx. \tag{7.61}$$

Если функция $f(x)$ не ограничена в окрестности некоторой внутренней точки $c\in[a,b]$, то по определению полагают

$$\int\limits_a^b f(x)\,dx = \int\limits_a^c f(x)\,dx + \int\limits_c^b f(x)\,dx,$$

где несобственные интегралы второго рода в правой части этого равенства определяются соответственно по формулам (7.60) и (7.61).

Наконец, если a и b — особые точки, т.е. функция $f(x)$ ограничена и интегрируема на интервале (a,b), то несобственный интеграл второго рода определяется в виде суммы

$$\int\limits_a^b f(x)\,dx = \int\limits_a^c f(x)\,dx + \int\limits_c^b f(x)\,dx,$$

где c — произвольная точка на (a,b), а несобственные интегралы второго рода в правой части этого равенства определяются по формулам (7.61) и (7.60) соответственно.

Рассмотрим несколько примеров на вычисление несобственных интегралов второго рода с применением предельного перехода.

Пример 7. Найти $\int\limits_0^1 \dfrac{dx}{\sqrt{1-x}}$, особая точка $x=1$.

По формуле (7.60) получаем

$$\int\limits_0^1 \frac{dx}{\sqrt{1-x}} = \lim_{\varepsilon\to 0}\int\limits_0^{1-\varepsilon} \frac{dx}{\sqrt{1-x}} =$$

$$= -2\lim_{\varepsilon\to 0}\sqrt{1-x}\Big|_0^{1-\varepsilon} = -2\lim_{\varepsilon\to 0}\sqrt{\varepsilon}+2 = 2,$$

т.е. данный интеграл сходится.

Пример 8. Найти $\int\limits_{-1}^1 \dfrac{dx}{\sqrt{1-x^2}}$, две особые точки: -1 и 1.

Разобьем этот интеграл на два, используя внутреннюю точку $x=0$ интервала $(-1,1)$:

$$\int\limits_{-1}^1 \frac{dx}{\sqrt{1-x^2}} = \lim_{\varepsilon\to 0}\int\limits_{-1+\varepsilon}^0 \frac{dx}{\sqrt{1-x^2}} + \lim_{\varepsilon\to 0}\int\limits_0^{1-\varepsilon} \frac{dx}{\sqrt{1-x^2}} =$$

$$= \lim_{\varepsilon\to 0}\arcsin x\Big|_{-1+\varepsilon}^0 + \lim_{\varepsilon\to 0}\arcsin x\Big|_0^{1-\varepsilon} =$$

$$= -\lim_{\varepsilon\to 0}\arcsin(-1+\varepsilon) + \lim_{\varepsilon\to 0}\arcsin(1-\varepsilon) =$$

$$= -\arcsin(-1) + \arcsin(1) = \frac{\pi}{2}+\frac{\pi}{2} = \pi,$$

Следовательно, данный несобственный интеграл второго рода сходится.

Пример 9. Найти $\int\limits_0^1 \dfrac{dx}{x^\alpha}$, где $\alpha>0$ — некоторое число; $x=0$ — особая точка.

Рассмотрим разные случаи для числа α:

а) при $\alpha = 1$

$$\int\limits_0^1 \frac{dx}{x} = \lim_{\varepsilon\to 0}\int\limits_\varepsilon^1 \frac{dx}{x} = \lim_{\varepsilon\to 0}\ln x\Big|_\varepsilon^1 = 0 - \lim_{\varepsilon\to 0}\ln\varepsilon = \infty,$$

т.е. при $\alpha = 1$ несобственный интеграл расходится;

б) при $\alpha \neq 1$

$$\int_0^1 \frac{dx}{x^\alpha} = \lim_{\varepsilon \to 0} \int_\varepsilon^1 \frac{dx}{x^\alpha} = \lim_{\varepsilon \to 0} \frac{x^{1-\alpha}}{1-\alpha}\Big|_\varepsilon^1 = \frac{1}{1-\alpha} - \left(1 - \lim_{\varepsilon \to 0} \varepsilon^{1-\alpha}\right) =$$

$$= \begin{cases} \dfrac{1}{1+\alpha} & \text{при } 0 < \alpha < 1, \\ \infty & \text{при } \alpha > 1. \end{cases}$$

Следовательно, данный интеграл расходится при $\alpha \geqslant 1$ и сходится при $0 < \alpha < 1$.

По аналогии с несобственным интегралом первого рода можно обосновать применение основной формулы интегрального исчисления для несобственного интеграла второго рода. Пусть функция $f(x)$ интегрируема на промежутке $[a, b)$, причем b является для нее особой точкой. Если для $f(x)$ существует первообразная $F(x)$, имеющая предел в особой точке $x = b$ или непрерывная на отрезке $[a, b]$, то для вычисления несобственного интеграла второго рода имеет место обычная формула Ньютона–Лейбница:

$$\int_a^b f(x)\,dx = F(b) - F(a) = F(x)\Big|_a^b. \tag{7.62}$$

Та же формула справедлива и в случае, когда особая точка находится внутри промежутка интегрирования, однако при этом непременным является условие *непрерывности первообразной функции в особых точках*.

Рассмотрим примеры на применение формулы Ньютона–Лейбница.

Пример 10. Найти $\int_{-1}^{1} x^{-2/3}\,dx$.

Здесь особая точка $x = 0$, однако, так как первообразная функция $3x^{1/3}$ непрерывна в этой точке, данный интеграл существует:

$$\int_{-1}^{1} x^{-2/3}\,dx = 3x^{1/3}\Big|_{-1}^{1} = 3[1 - (-1)] = 6.$$

Пример 11. $\int_{-1}^{1} \frac{2x\,dx}{x^2 - 1}$ не существует, поскольку первообразная $\ln|x^2 - 1|$ обращается в бесконечность в особых точках $x = \pm 1$.

Пример 12. Найти $\int_0^1 \ln x\,dx$.

Здесь особая точка $x = 0$. Первообразная функция, определяемая с помощью интегрирования по частям, имеет вид $x \ln x - x$. При $x \to 0$ ее предел равен нулю, и потому данный интеграл сходится:

$$\int_0^1 \ln x\,dx = [x \ln x - x]\Big|_0^1 = -1.$$

Признаки сходимости несобственных интегралов

Теорема 7.11 (ПРИЗНАК СРАВНЕНИЯ НЕСОБСТВЕННЫХ ИНТЕГРАЛОВ). *Пусть функции $f(x)$ и $g(x)$ непрерывны на промежутке $[a, \infty)$ и удовлетворяют на нем условию $0 \leqslant f(x) \leqslant g(x)$. Тогда из сходимости интеграла*

$$\int_a^\infty g(x)\,dx \tag{7.63}$$

следует сходимость интеграла

$$\int_a^\infty f(x)\,dx; \tag{7.64}$$

и наоборот, из расходимости интеграла (7.64) *следует расходимость интеграла* (7.63).

Доказательство. Введем обозначения

$$F(R) = \int_a^R f(x)\,dx, \quad G(R) = \int_a^R g(x)\,dx,$$

причем $0 \leqslant F(R) \leqslant G(R)$.

Функции $F(R)$ и $G(R)$ являются неубывающими; в самом деле, если $a \leqslant R_1 \leqslant R_2$, то $\int\limits_{R_1}^{R_2} f(x)\,dx \geqslant 0$ и тогда

$$F(R_2) = \int\limits_a^{R_2} f(x)\,dx = \int\limits_a^{R_1} f(x)dx + \int\limits_{R_1}^{R_2} f(x)\,dx \geqslant$$

$$\geqslant \int\limits_a^{R_1} f(x)\,dx = F(R_1).$$

Возьмем последовательность значений $\{R_n\} \to \infty$; тогда соответствующая последовательность значений функции $\{F(R_n)\}$ является монотонной и неубывающей. Пусть интеграл (7.63) сходится, тогда последовательность $\{G(R_n)\}$ ограничена, но тогда ограничена и последовательность $\{F(R_n)\}$, а значит, в силу теоремы 2.13 она сходится. Следовательно, существует предел $F(R)$ при $R \to \infty$, т.е. интеграл (7.64) сходится.

Теперь докажем вторую часть теоремы: пусть интеграл (7.64) расходится. Если предположить, что интеграл (7.63) сходится, то по доказанному выше интеграл (7.64) также должен сходиться, что противоречит условию. Теорема доказана.

Замечание. Аналогичный признак сравнения справедлив и для несобственных интегралов второго рода. Если функции $f(x)$ и $g(x)$ непрерывны на полуинтервале $[a, b)$ и для всех точек в некоторой окрестности особой точки b выполняются условия $0 \leqslant f(x) \leqslant g(x)$, то из сходимости интеграла $\int\limits_a^b g(x)\,dx$ следует сходимость интеграла $\int\limits_a^b f(x)\,dx$, а из расходимости интеграла $\int\limits_a^b f(x)\,dx$ следует расходимость интеграла $\int\limits_a^b g(x)\,dx$.

Рассмотрим примеры на исследование сходимости несобственных интегралов.

Пример 13. $\int\limits_1^\infty \dfrac{dx}{x^3(1+e^x)}$.

Сравним подынтегральную функцию в этом интеграле с функцией $\dfrac{1}{x^3}$. Очевидно, что $\dfrac{1}{x^3(1+e^x)} < \dfrac{1}{x^3}$. Интеграл $\int\limits_1^\infty \dfrac{1}{x^3}\,dx$

сходится (см. пример 5, случай «б»); следовательно, в силу признака сравнения сходится и данный интеграл.

Пример 14. $\int\limits_1^\infty \dfrac{1+\ln x}{x}\,dx$.

Сравнивая подынтегральную функцию данного интеграла с функцией $1/x$, видим, что $(1+\ln x)/x > 1/x$ на промежутке $1 \leqslant x \leqslant \infty$. Интеграл $\int\limits_1^\infty \dfrac{1}{x}dx$ расходится, следовательно, по признаку сравнения расходится и данный интеграл.

В заключение приведем критерий Коши сходимости несобственного интеграла первого рода.

Теорема 7.12 (критерий Коши). *Для сходимости несобственного интеграла* (7.64) *необходимо и достаточно, чтобы для любого $\varepsilon > 0$ можно было найти такое число $A > 0$, что для любых R' и R'', больших, чем A, выполняется неравенство*

$$|F(R'') - F(R')| = \left| \int\limits_{R'}^{R''} f(x)\,dx \right| < \varepsilon. \qquad (7.65)$$

Абсолютная и условная сходимость несобственных интегралов

Определение 3. Несобственный интеграл $\int\limits_a^\infty f(x)\,dx$ называется *абсолютно сходящимся*, если сходится интеграл $\int\limits_a^\infty |f(x)|\,dx$.

Определение 4. Несобственный интеграл $\int\limits_a^\infty f(x)\,dx$ называется *условно сходящимся*, если он сходится, а интеграл $\int\limits_a^\infty |f(x)|\,dx$ расходится.

Заметим, что из абсолютной сходимости интеграла следует и его сходимость в силу оценки 3^0 определенного интеграла (см. параграф 7.4) и критерия Коши.

Теорема 7.13 (признак Дирихле–Абеля[1]). *Пусть функция $f(x)$ непрерывна и имеет ограниченную первообразную*

[1] Нильс Хенрик Абель (1802–1829) — норвежский математик.

$F(x)$ на промежутке $[a,\infty)$, а функция $g(x)$ имеет непрерывную производную на этом промежутке, не возрастает и стремится к нулю при $x\to\infty$. Тогда несобственный интеграл

$$\int_a^\infty f(x)g(x)\,dx \tag{7.66}$$

сходится.

Доказательство. Применим интегрирование по частям к интегралу $\int_{R'}^{R''} f(x)g(x)\,dx$ на произвольном отрезке $[R',R'']\subset$ $\subset[a,\infty)$. Имеем:

$$\int_{R'}^{R''} f(x)g(x)\,dx = F(x)g(x)\Big|_{R'}^{R''} - \int_{R'}^{R''} F(x)g'(x)\,dx. \tag{7.67}$$

По условию теоремы $F(x)$ ограничена, т.е. $|F(x)|\leqslant K$. Так как функция $g(x)$ не возрастает и стремится к нулю при $x\to\infty$, то $g(x)\geqslant 0$, а $g'(x)\leqslant 0$. Таким образом, получаем

$$\begin{aligned}\left|\int_{R'}^{R''} f(x)g(x)\,dx\right| &\leqslant K[g(R')+g(R'')]+\\ &+K\int_{R'}^{R''}[-g'(x)]\,dx = 2Kg(R').\end{aligned} \tag{7.68}$$

Так как по условию теоремы $g(x)\to 0$ при $x\to\infty$, то для произвольного числа $\varepsilon>0$ можно найти такое число A, что при $R'>A$ будет выполнено неравенство $g(R')<\varepsilon/2K$. Подставляя это в оценку (7.68), получаем

$$\left|\int_{R'}^{R''} f(x)g(x)dx\right|<\varepsilon,$$

что соответствует критерию Коши сходимости интеграла (7.66). Теорема доказана.

Рассмотрим примеры использования признака Дирихле–Абеля сходимости несобственных интегралов.

Пример 15. $\int_1^\infty \dfrac{\sin x}{x^\alpha}\,dx$, $\alpha>0$.

Положим $f(x)=\sin x$, $g(x)=1/x^\alpha$; легко убедиться, что все условия теоремы выполнены, т.е. данный интеграл сходится.

При $\alpha>1$ данный интеграл сходится абсолютно. Действительно, $|\sin x|/x^\alpha\leqslant 1/x^\alpha$, интеграл $\int_1^\infty (1/x^\alpha)dx$ сходится, т.е. по признаку сравнения (см. теорему 7.11) сходится абсолютно и данный интеграл.

Пример 16. Рассмотрим интеграл Френеля $\int_0^\infty \sin x^2\,dx$.

Представим этот интеграл в виде суммы:

$$\int_0^\infty \sin x^2\,dx = \int_0^1 \sin x^2\,dx + \int_1^\infty \sin x^2\,dx. \tag{7.69}$$

Поскольку $\sin x^2$ — непрерывная функция на отрезке $[0,1]$, то первый интеграл в формуле (7.69) существует. Для выяснения сходимости несобственного интеграла в правой части формулы (7.69) положим $f(x)=x\sin x^2$, $g(x)=1/x$. Тогда для функции $f(x)$ первообразная $F(x)=-\cos x^2/2$ ограничена на промежутке $[1,\infty)$, а $g(x)$ — положительна, стремится к нулю при $x\to\infty$ и имеет непрерывную производную на $[1,\infty)$. Значит, по признаку Дирихле–Абеля второй интеграл в формуле (7.69) сходится, т.е. интеграл Френеля также сходится.

Глава 8
РЯДЫ

8.1. Понятие числового ряда

Основные определения

Рассмотрим числовую последовательность $a_1,a_2,a_3,\dots,a_n,\dots$. Образуем из элементов этой последовательности выражение вида

$$a_1+a_2+a_3+\dots+a_n+\dots=\sum_{n=1}^{\infty}a_n, \tag{8.1}$$

которое называется *числовым рядом*, или просто *рядом*. Слагаемые в формуле (8.1) называются *членами ряда*. Суммы пер-

вых n членов ряда называются *частичными суммами ряда:*

$$S_1 = a_1, \quad S_2 = a_1 + a_2, \ldots,$$
$$S_n = a_1 + a_2 + a_3 + \cdots + a_n = \sum_{k=1}^{n} a_k. \tag{8.2}$$

Поскольку число членов ряда бесконечно, то частичные суммы ряда образуют числовую последовательность $\{S_n\}$:

$$S_1,\ S_2,\ S_3, \ldots, S_n, \ldots . \tag{8.3}$$

Ряд (8.1) называется *сходящимся*, если существует предел S последовательности частичных сумм (8.3); в таком случае число S называется *суммой ряда:*

$$S = \lim_{n\to\infty} S_n = \sum_{n=1}^{\infty} a_n. \tag{8.4}$$

Если же последовательность частичных сумм (8.3) не имеет предела, числовой ряд (8.1) называется *расходящимся.*

Рассмотрим примеры числовых рядов.

1. Дан ряд $1 - 1 + 1 - 1 + \cdots = \sum\limits_{n=1}^{\infty} (-1)^{n-1}$.

Последовательность частичных сумм этого ряда $S_1 = 1$, $S_2 = 0, \ldots, S_{2n-1} = 1$, $S_{2n} = 0, \ldots$ не имеет предела, т.е. ряд расходится.

2. Дан ряд, составленный из элементов геометрической прогрессии с первым членом $b \neq 0$ и знаменателем $q \neq 1$:

$$b + bq + bq^2 + \cdots + bq^{n-1} + \cdots = \sum_{n=1}^{\infty} bq^{n-1}.$$

Частичная сумма общего вида этого ряда выражается формулой

$$S_n = b + bq + bq^2 + \cdots + bq^{n-1} = b\frac{1-q^n}{1-q}.$$

Очевидно, что при $|q| < 1$ пределом S_n при $n \to \infty$ является число $S = b/(1-q)$, которое также будет и суммой данного ряда. При $|q| > 1$ предел S_n при $n \to \infty$ бесконечен и ряд расходится. Если же $|q| = 1$, то данный ряд также расходится: при $q = 1$ расходимость очевидна, при $q = -1$ имеем числовой ряд, рассмотренный в первом примере.

Свойства сходящихся рядов

1°. *Отбрасывание конечного числа членов ряда не влияет на сходимость (расходимость) ряда.*

Действительно, все частичные суммы ряда, начиная с некоторой, изменятся на одно и то же постоянное число, равное сумме отбрасываемых членов.

2°. *Если ряд* (8.1) *сходится и имеет сумму* S, *то сходится также и ряд* $\sum\limits_{n=1}^{\infty} ka_n$, *где* $k \neq 0$ — *постоянное число, причем сумма этого ряда равна* kS.

Это утверждение следует из соответствующего свойства сходящихся последовательностей.

3°. *Если ряды* $\sum\limits_{n=1}^{\infty} a_n$ *и* $\sum\limits_{n=1}^{\infty} b_n$ *сходятся и суммы их соответственно равны* S' *и* S'', *то и ряд* $\sum\limits_{n=1}^{\infty} (a_n \pm b_n)$ *также сходится, причем его сумма* $S = S' \pm S''$.

Справедливость этого утверждения также следует из свойств сходящихся последовательностей частичных сумм $\{S'_n\}$ и $\{S''_n\}$.

4°. *Общий член* a_n *сходящегося ряда стремится к нулю при* $n \to \infty$.

Действительно, $a_n = S_n - S_{n-1}$ и, поскольку ряд сходится, $S_n \to S$ и $S_{n-1} \to S$ при $n \to \infty$, где S — сумма ряда; отсюда и следует справедливость данного свойства, которое часто называют *необходимым условием сходимости ряда.* Если оно не соблюдается, то ряд заведомо расходится.

Например, ряд $\sum\limits_{n-1}^{\infty} \dfrac{n}{4n+1}$ — заведомо расходится, так как

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{n}{4n+1} = \frac{1}{4} \neq 0.$$

Следует заметить, что свойство 4° не является достаточным. Рассмотрим ряд $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}}$. Необходимое условие сходимос-

ти выполняется ($1/\sqrt{n} \to 0$ при $n \to \infty$), однако его частичная сумма

$$S_n = 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}} > n\frac{1}{\sqrt{n}} = \sqrt{n}$$

стремится к бесконечности вместе с n, т.е. ряд расходится.

8.2. Числовые ряды с неотрицательными членами

Ряды с неотрицательными членами часто встречаются в различных приложениях. Сразу отметим основное свойство таких рядов: *последовательность частичных сумм ряда с неотрицательными членами является неубывающей.*

Необходимое и достаточное условие сходимости ряда

Теорема 8.1. *Для того чтобы ряд с неотрицательными членами сходился, необходимо и достаточно, чтобы последовательность его частичных сумм была ограничена.*

Доказательство. По определению сходящегося ряда последовательность его частичных сумм сходится, значит, она необходимо ограничена. Что касается достаточности, то ограниченная монотонная неубывающая последовательность частичных сумм сходится в силу теоремы 2.13.

Признаки сравнения

Определим несколько признаков, позволяющих устанавливать сходимость или расходимость числового ряда.

Теорема 8.2. *Пусть для двух рядов с неотрицательными членами*

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots + a_n + \ldots \tag{8.5}$$

и

$$\sum_{n=1}^{\infty} b_n = b_1 + b_2 + b_3 + \cdots + b_n + \ldots \tag{8.6}$$

выполняется неравенство $a_n \leqslant b_n$ для всех n. Тогда из сходимости ряда (8.6) следует сходимость ряда (8.5), а расходимость ряда (8.5) влечет за собой расходимость ряда (8.6).

Доказательство. Пусть S_n и S'_n — частичные суммы соответственно рядов (8.5) и (8.6). Из условия теоремы $a_n \leqslant b_n$ следует, что $S_n \leqslant S'_n$. Если ряд (8.6) сходится, то по теореме 8.1 последовательность $\{S'_n\}$ ограничена; значит, монотонная последовательность $\{S_n\}$ также ограничена, т.е. она также сходится.

Если ряд (8.5) расходится, то ряд (8.6) не может сходиться, иначе в силу первой части доказательства будет сходиться и ряд (8.5). Доказательство теоремы завершено.

Рассмотрим примеры применения теоремы 8.2 по установлению сходимости (расходимости) неотрицательных рядов.

Пример 1. $\sum\limits_{n=1}^{\infty} \frac{1}{n^n}$.

Поскольку $1/n^n < 1/2^n$ при $n \geqslant 2$, а ряд $\sum\limits_{n=1}^{\infty} (1/2^n)$ сходится (сумма бесконечной убывающей геометрической прогрессии), то сходится и данный ряд.

Пример 2. $\sum\limits_{n=1}^{\infty} \frac{1}{\ln n}$.

Так как $\ln n < \sqrt{n}$ для достаточно больших n (это легко проверить по правилу Лопиталя для функции $\ln x/x^{\alpha}$ при $x \to \infty$, где $\alpha > 0$), то $1/\ln n \geqslant 1/\sqrt{n}$; но ряд $\sum\limits_{n=1}^{\infty} (1/\sqrt{n})$ расходится (см. параграф 8.1) — значит, расходится и данный ряд.

Часто более удобным и употребимым бывает другой признак сравнения рядов как следствие из теоремы 8.2.

Теорема 8.3. *Пусть ряд (8.5) — ряд с неотрицательными членами, а (8.6) — с положительными членами. Если существует предел*

$$\lim_{n\to\infty} \frac{a_n}{b_n} = K, \quad 0 < K < \infty, \tag{8.7}$$

то оба ряда сходятся или расходятся одновременно.

Доказательство. По определению предела для любого числа $\varepsilon > 0$ для достаточно больших n

$$\frac{a_n}{b_n} < K + \varepsilon, \quad \text{или} \quad a_n < (K + \varepsilon) b_n.$$

Значит, из сходимости ряда (8.6) следует сходимость ряда

$$\sum_{n=1}^{\infty}(K+\varepsilon)b_n \qquad (8.8)$$

по свойству 2° сходящегося ряда (см. параграф 8.1), а по теореме 8.2 вытекает сходимость ряда (8.5).

Если же ряд (8.5) расходится, то в этом случае расходится и ряд (8.8); но тогда расходится и ряд (8.6), поскольку в противном случае сходился бы ряд (8.8) в силу свойства 2° сходящегося ряда.

Теорема доказана.

Другие признаки сходимости

Теорема 8.4 (ПРИЗНАК ДАЛАМБЕРА[1]). *Пусть для ряда* (8.5) *с положительными членами существует предел*

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n}=\rho. \qquad (8.9)$$

Тогда этот ряд сходится при $\rho<1$ *и расходится при* $\rho>1$.

Доказательство. По определению предела последовательности для любого $\varepsilon>0$ найдется такой номер N, что при всех $n>N$ выполняются неравенства

$$\rho-\varepsilon<\frac{a_{n+1}}{a_n}<\rho+\varepsilon \qquad (8.10)$$

а) Пусть $\rho<1$, тогда ε можно взять таким, что $q=\rho+\varepsilon<1$; из неравенств (8.10)

$$\frac{a_{n+1}}{a_n}<q, \quad \text{или} \quad a_{n+1}<a_nq$$

для всех $n>N$, т.е. мы получаем систему неравенств

$$a_{N+1}<a_Nq,\ a_{N+2}<a_{N+1}q$$

[1]Жан Лерон Даламбер (1717–1783) — французский математик и философ-просветитель.

или

$$a_{N+2}<a_Nq^2,\ldots,a_{N+k}<a_Nq^k,\ldots.$$

Таким образом, члены ряда (8.5), начиная с a_{N-1} меньше соответствующих членов сходящегося ряда — бесконечной геометрической прогрессии со знаменателем $q<1$; следовательно, в силу свойства 1° (см. параграф 8.1) и теоремы сравнения 8.2 ряд (8.5) сходится.

б) Пусть $\rho>1$; положим ε таким малым, чтобы $\rho-\varepsilon>1$. Тогда из левого неравенства (8.10) следует, что $a_{n+1}>a_n$ при $n>N+1$. Значит, члены ряда, начиная с $(N+1)$-го, возрастают, т.е. свойство 4° сходящегося ряда не выполняется, следовательно, ряд (8.5) расходится. Доказательство теоремы завершено.

ЗАМЕЧАНИЕ. При $\rho=1$ необходимо дополнительное исследование ряда с использованием других признаков, так как в этом случае ряд (8.5) может как сходиться, так и расходиться.

В качестве примеров применения признака Даламбера исследуем сходимость следующих рядов.

Пример 3. $\sum\limits_{n=1}^{\infty}na^n$, $a>0$.

Составим отношение $(n+1)$-го члена к n-му и перейдем к пределу при $n\to\infty$:

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n}=\lim_{n\to\infty}\frac{(n+1)a^{n+1}}{na^n}=\lim_{n\to\infty}\frac{n+1}{n}a=a.$$

По признаку Даламбера имеем: если $a<1$, то данный ряд сходится, если $a>1$ — ряд расходится.

Пример 4. $\sum\limits_{n=1}^{\infty}\dfrac{n^n}{n!}$.

Вычисляем предел отношения a_{n+1}/a_n:

$$\lim_{n\to\infty}\frac{a_{n-1}}{a_n}=\lim_{n\to\infty}\frac{(n+1)^{n+1}}{(n+1)!}\frac{n!}{n^n}=$$
$$=\lim_{n\to\infty}\frac{(n+1)^n}{n^n}=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n=e>1,$$

т.е. данный ряд расходится.

Пример 5. $\sum\limits_{n=1}^{\infty} n!\left(\frac{a}{n}\right)^n$, $a>0$.

Находим предел отношения

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n}=\lim_{n\to\infty}\frac{(n+1)![a/(n+1)]^{n+1}}{n!(a/n)^n}=$$

$$=\lim_{n\to\infty} a\left(1+\frac{1}{n}\right)^{-n}=a/e,$$

т.е. при $a<e$ данный ряд сходится, при $a>e$ он расходится.

Теорема 8.5 (ПРИЗНАК КОШИ). *Если существует предел*

$$\lim_{n\to\infty}\sqrt[n]{a_n}=L, \tag{8.11}$$

то ряд (8.5) *сходится при* $L<1$ *и расходится при* $L>1$.

Доказательство. Из условия (8.11) следует, что для произвольного $\varepsilon>0$ справедливы неравенства

$$L-\varepsilon<\sqrt[n]{a_n}<L+\varepsilon, \quad \text{или} \quad (L-\varepsilon)^n<a_n<(L+\varepsilon)^n.$$

а) Пусть $L<1$, тогда выберем ε таким образом, чтобы величина $q=L+\varepsilon<1$. В этом случае все члены ряда (8.5) меньше соответствующих степеней бесконечной сходящейся геометрической прогрессии, т.е. в силу теоремы 8.2 ряд (8.5) сходится.

б) Пусть $L>1$, тогда выберем ε таким, чтобы $L-\varepsilon>1$. В этом случае свойство 4° сходящегося ряда (см. параграф 8.1) не выполняется (a_n не стремится к нулю при $n\to\infty$), т.е. ряд (8.5) расходится.

Теорема полностью доказана.

Пример 6. Исследовать сходимость ряда $\sum\limits_{n=1}^{\infty}(b/c_n)^n$, где $\lim\limits_{n\to\infty} c_n=c$.

Применяя признак Коши, получаем

$$\lim_{n\to\infty}\sqrt[n]{a_n}=\lim_{n\to\infty}(b/c_n)=\frac{b}{c}.$$

Отсюда, при $b<c$ данный ряд сходится, а при $b>c$ он расходится.

Теорема 8.6 (ИНТЕГРАЛЬНЫЙ ПРИЗНАК СХОДИМОСТИ). *Пусть функция* $f(x)$ *непрерывная, положительная и убывающая всюду на промежутке* $[a,\infty)$. *Тогда числовой ряд*

$$\sum_{n=1}^{\infty} f(n)=f(1)+f(2)+f(3)+\cdots+f(n)+\ldots \tag{8.12}$$

сходится вместе с несобственным интегралом

$$\int_1^{\infty} f(x)dx \tag{8.13}$$

Доказательство. По геометрическому смыслу определенный интеграл (8.13) представляет собой площадь криволинейной трапеции бесконечной протяженности, ограниченной сверху кривой $f(x)$, снизу осью Ox и слева прямой $x=1$ (рис. 8.1). Ограничим эту трапецию справа прямой $x=n$, после чего выполним два построения: впишем в нее ступенчатую фигуру, состоящую из прямоугольников с основаниями, равными единице, и высотами $f(2), f(3), \ldots, f(n)$; опишем около нее ступенчатую фигуру, составленную из прямоугольников с единичными основаниями и высотами $f(1), f(2), \ldots, f(n-1)$. Тогда мы

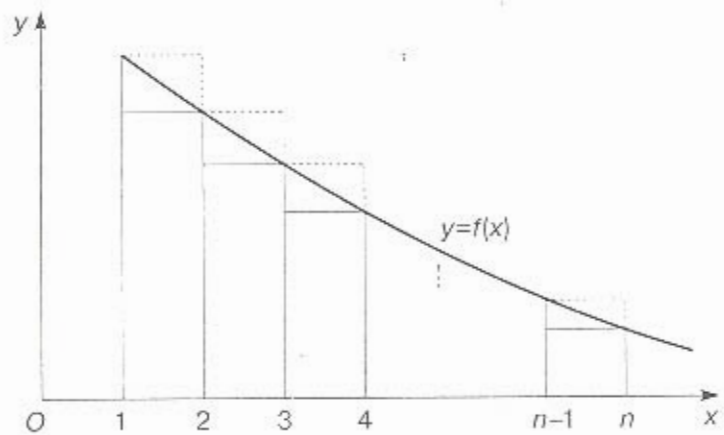

Рис. 8.1

получим, исходя из вида ряда (8.12),

$$S_n - f(1) < \int_1^n f(x)dx < S_n - f(n),$$

где S_n — частичные суммы этого ряда. Отсюда получаем неравенства

$$S_n < f(1) + \int_1^n f(x)dx, \quad S_n > f(n) + \int_1^n f(x)dx. \tag{8.14}$$

Пусть интеграл (8.13) сходится, т.е. существует предел интеграла I в неравенствах (8.14) при $n \to \infty$. Поскольку $f(x) > 0$, то последовательность $\left\{\int_1^n f(x)dx\right\}$ является монотонно возрастающей и ограничена сверху своим пределом I. Но тогда последовательность $\{S_n\}$ ограничена сверху и в силу теоремы 8.1 ряд (8.12) сходится.

Если же интеграл (8.13) расходится, то из второго неравенства (8.14) следует, что конечного предела частичных сумм ряда (8.12) при $n \to \infty$ не существует и он расходится. Теорема доказана.

Рассмотрим примеры применения интегрального признака.

Пример 7. $\sum_{n=1}^{\infty}(1/n^\alpha)$, $\alpha > 0$.

Членами этого ряда являются значения функции $f(x) = 1/x^\alpha$ в целочисленных точках. Ранее было исследовано (см. параграф 7.10, пример 5), что соответствующий несобственный интеграл первого рода $\int_1^\infty \frac{dx}{x^\alpha}$ сходится при $\alpha > 1$ и расходится при $\alpha \leqslant 1$. Следовательно, данный ряд также сходится при $\alpha > 1$ и расходится при $\alpha \leqslant 1$. В частности, отсюда следует расходимость так называемого *гармонического ряда:*

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} + \cdots = \sum_{n=1}^{\infty}\frac{1}{n}. \tag{8.15}$$

Пример 8. $\sum_{n=1}^{\infty}\frac{1}{n\ln n(\ln\ln n)}$.

Здесь $f(x) = \frac{1}{x\ln x(\ln\ln x)}$; первообразная для этой функции $F(x) = \ln\ln\ln x \to \infty$ при $x \to +\infty$, и, значит, данный ряд расходится.

8.3. Сходимость произвольных числовых рядов

Знакопеременные ряды

Рассмотрим распространенный тип числового ряда, так называемые *знакопеременные ряды*, элементы которых имеют чередующиеся знаки. Полагая первый член положительным, знакопеременный ряд можно записать в виде

$$\sum_{n=1}^{\infty}(-1)^{n+1}a_n = a_1 - a_2 + a_3 - \cdots + (-1)^{n+1}a_n + \ldots, \tag{8.16}$$

где $a_n > 0$.

Для знакопеременных рядов имеет место достаточный признак сходимости.

Теорема 8.7 (ПРИЗНАК ЛЕЙБНИЦА). *Если последовательность абсолютных величин членов $\{a_n\}$ знакопеременного ряда* (8.16) *является монотонно убывающей и стремящейся к нулю при $n \to \infty$, то этот ряд сходится.*

Доказательство. Запишем частичную сумму четного порядка S_{2n} в виде

$$S_{2n} = (a_1 - a_2) + (a_3 - a_4) + \cdots + (a_{2n-1} - a_{2n}).$$

В силу первого условия теоремы каждая разность в скобках является положительным числом, и потому последовательность частичных сумм $\{S_{2n}\}$ монотонно возрастающая. Теперь представим ее в виде

$$S_{2n} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \cdots - a_{2n},$$

откуда следует, что $S_{2n} < a_1$, т.е. последовательность $\{S_{2n}\}$ ограничена, а потому она имеет предел $\lim\limits_{n\to\infty} S_{2n} = S$.

Поскольку частичные суммы нечетного числа членов ряда (8.16) $S_{2n+1} = S_{2n} + a_{2n+1}$, то по второму условию теоремы

$$\lim_{n\to\infty} S_{2n+1} = \lim_{n\to\infty}(S_{2n} + a_{2n+1}) = S + 0 = S.$$

Любая последовательность частичных сумм $\{S_n\}$ ряда (8.16) сходится к пределу S, что и означает сходимость этого ряда. Теорема доказана.

Приведем примеры знакопеременных рядов.

Пример 1. $\sum\limits_{n=1}^{\infty} \dfrac{(-1)^{n+1}}{n} = 1 - \dfrac{1}{2} + \dfrac{1}{3} + \cdots + \dfrac{(-1)^{n+1}}{n} + \ldots$.

Этот ряд сходится по признаку Лейбница, так как его члены убывают по абсолютной величине и $(-1)^{n+1}/n \to 0$ при $n \to \infty$.

Пример 2. $\sum\limits_{n=2}^{\infty} \dfrac{(-1)^n}{\ln n}$.

Нетрудно убедиться, что данный ряд удовлетворяет условиям теоремы 8.7 и потому сходится.

Абсолютная и условная сходимость ряда

Рассмотрим случай ряда с членами, имеющими произвольные знаки:

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots + a_n + \ldots . \tag{8.17}$$

Одновременно рассмотрим ряд

$$\sum_{n=1}^{\infty} |a_n| = |a_1| + |a_2| + |a_3| + \cdots + |a_n| + \ldots , \tag{8.18}$$

где a_n — члены ряда (8.17).

Теорема 8.8. *Из сходимости ряда* (8.18) *следует сходимость ряда* (8.17).

Доказательство. Пусть S_n и $\sum_n$ — соответственно частичные суммы рядов (8.17) и (8.18). Тогда по условию теоремы существует предел $\lim\limits_{n\to\infty} \sum_n = \sum$, причем для любого n имеет место неравенство

$$\sum_n \leqslant \sum , \tag{8.19}$$

поскольку последовательность $\{\sum_n\}$ является неубывающей. Обозначим через S'_n и S''_n соответственно сумму положительных членов и сумму абсолютных величин отрицательных членов частичной суммы S_n. Тогда

$$S_n = S'_n - S''_n, \quad \sum_n = S'_n + S''_n. \tag{8.20}$$

Последовательности $\{S'_n\}$ и $\{S''_n\}$ не убывают и ограничены сверху, как это следует из сходимости ряда (8.18) и второго равенства (8.20), т.е. они имеют пределы при $n \to \infty$: $\lim\limits_{n\to\infty} S'_n = S'$, $\lim\limits_{n\to\infty} S''_n = S''$. Тогда из первого равенства (8.20) получаем, что существует и предел последовательности частичных сумм ряда (8.17), т.е. этот ряд сходится, что и требовалось доказать.

Если сходится ряд (8.18), то ряд (8.17) называется *абсолютно сходящимся*. Если ряд (8.17) сходится, но ряд (8.18) расходится, то ряд (8.17) называется *условно сходящимся*.

Приведем примеры условной и абсолютной сходимости числовых рядов.

Пример 3. Ряд $\sum\limits_{n=1}^{\infty} \dfrac{(-1)^{n+1}}{n}$ сходится по признаку Лейбница, однако гармонический ряд $\sum\limits_{n=1}^{\infty} \dfrac{1}{n}$ расходится, следовательно, сходимость исходного ряда является условной.

Пример 4. $\sum\limits_{n=1}^{\infty} (-1)^{n+1} \dfrac{1}{n^{\alpha}}$, $\alpha > 0$.

При $\alpha > 1$ этот ряд сходится абсолютно (см. параграф 8.2, пример 7), при $\alpha \leqslant 1$ данный ряд сходится условно.

В заключение приведем без доказательства один важный признак сходимости числового ряда.

Теорема 8.9 (ПРИЗНАК ДИРИХЛЕ–АБЕЛЯ). *Пусть дан ряд*

$$\sum_{n=1}^{\infty} a_n b_n . \tag{8.21}$$

Если: а) *последовательность частичных сумм ряда* $\sum\limits_{n=1}^{\infty} a_n$ *ограничена,* б) *последовательность* $\{b_n\}$ *является невозрастающей и бесконечно малой, то ряд* (8.21) *сходится.*

Заметим, что признак Лейбница является частным случаем этой теоремы при $a_n = (-1)^{n+1}$.

Применим сформулированный выше признак Дирихле–Абеля к установлению сходимости ряда

$$\sum_{n=1}^{\infty} \frac{\sin n\alpha}{n}, \quad 0 < \alpha < 2\pi. \tag{8.22}$$

Положим $a_n = \sin n\alpha$, $b_n = 1/n$. Вычислим частичные суммы ряда

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \sin n\alpha, \tag{8.23}$$

для чего умножим и поделим каждое слагаемое этой суммы на постоянную величину $2\sin(\alpha/2)$:

$$S_n = \sin\alpha + \sin 2\alpha + \ldots + \sin n\alpha = \frac{1}{2\sin(\alpha/2)} \left\{ \left[\cos\frac{\alpha}{2} - \cos\frac{3\alpha}{2} \right] + \right.$$

$$\left. + \left[\cos\frac{3\alpha}{2} - \cos\frac{5\alpha}{2} \right] + \ldots + \left[\cos\left(n - \frac{1}{2}\right)\alpha - \cos\left(n + \frac{1}{2}\right)\alpha \right] \right\} =$$

$$= \frac{1}{2\sin(\alpha/2)} \left[\cos\frac{\alpha}{2} - \cos\left(n + \frac{1}{2}\right)\alpha \right].$$

Таким образом, частичные суммы ряда (8.23) ограничены, а последовательность $\{b_n\} = \{1/n\}$ является невозрастающей и бесконечно малой; условия признака Дирихле–Абеля выполняются, т.е. ряд (8.22) сходится.

8.4. Степенные ряды

В курсе математического анализа изучаются последовательности и ряды, членами которых являются не числа, а функции, определенные на некотором множестве. Такие функциональные последовательности и ряды широко применяются в различных приложениях для анализа и приближенных вычислений. Мы ограничимся рассмотрением степенных рядов.

Определение и терминология

Определение. Функциональный ряд вида

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n + \cdots = \sum_{n=0}^{\infty} a_n x^n \tag{8.24}$$

называется *степенным рядом*. Постоянные числа $a_0, a_1, \ldots, a_n, \ldots$ называются *коэффициентами степенного ряда* (8.24).

При разных значениях переменной x мы получим разные числовые ряды, которые могут быть сходящимися или расходящимися. Особый интерес представляет множество значений x, при которых ряд (8.24) сходится, оно называется *областью сходимости степенного ряда*.

Понятно, что частичная сумма степенного ряда $S_n = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ представляет собой функцию переменной x. Стало быть, последовательность частичных сумм является функциональной последовательностью $\{S_n(x)\}$ и сумма ряда (8.24) является функцией переменной x: $S = S(x)$; они определены в области сходимости ряда.

Область сходимости степенного ряда

Теорема 8.10 (ТЕОРЕМА АБЕЛЯ). *Если степенной ряд* (8.24) *сходится при* $x = x_0$ $(x_0 \neq 0)$, *то он абсолютно сходится для всех* x, *таких, что* $|x| < |x_0|$; *если ряд* (8.24) *расходится при* $x = x_1$ *то он расходится и при всех* x, *удовлетворяющих неравенству* $|x| > |x_1|$.

Доказательство. Пусть степенной ряд (8.24) сходится при $x = x_0$. Тогда в силу свойства 4° (см. параграф 8.1) общий член числового ряда $a_n x_0^n \to 0$ при $n \to \infty$; отсюда следует, что последовательность $\{a_n x_0^n\}$ ограничена, т.е. существует такое число $M > 0$, что

$$|a_n x_0^n| < M, \quad n = 0, 1, 2, \ldots. \tag{8.25}$$

Перепишем ряд (8.24) с помощью тождественного преобразования в виде

$$\sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_n x_0^n (x/x_0)^n. \tag{8.26}$$

Составим теперь ряд из абсолютных величин членов ряда (8.26)

$$\sum_{n=0}^{\infty} |a_n x_0^n| |x/x_0|^n. \tag{8.27}$$

Все члены этого ряда в силу неравенства (8.25) меньше соответствующих членов ряда

$$\sum_{n=0}^{\infty} M|x/x_0|^n, \tag{8.28}$$

который является суммой геометрической прогрессии со знаменателем $|x/x_0| < 1$ и потому сходится. Тогда в силу признака сравнения (см. теорему 8.2) сходится и ряд (8.27), т.е. ряд (8.24) сходится абсолютно.

Пусть теперь ряд (8.24) расходится при $x = x_1$. Предположим обратное утверждению теоремы, т.е. что при некотором $x > x_1$ данный ряд сходится. Тогда по уже доказанной части теоремы этот ряд должен сходиться в точке $x = x_1$, что противоречит условию теоремы.

Теорема полностью доказана.

Теорема Абеля примечательна утверждением, что если степенной ряд (8.24) сходится при $x = x_0$, то он сходится абсолютно всюду на отрезке $[-|x_0|, |x_0|]$. Если же x_1 — точка расходимости ряда, то он расходится везде вне интервала $(-|x_1|, |x_1|)$.

Отсюда следует основополагающая теорема.

Теорема 8.11. *Если степенной ряд сходится не только при $x = 0$, то существует такое положительное число R (возможно, и бесконечное), что ряд абсолютно сходится в интервале $(-R, R)$ и расходится везде вне этого интервала.*

Доказательство. Пусть X — множество точек x, в которых ряд (8.24) сходится. Оно по условию теоремы не пустое; покажем, что оно ограничено. Действительно, пусть x_1 — точка, где ряд расходится; тогда по теореме Абеля для любого $x \in X$ выполняется условие $|x| < |x_1|$. Стало быть, у этого множества существует точная верхняя грань $R = \sup\limits_{x \in X} |x|$, причем $R > 0$, поскольку ряд сходится не только при $x = 0$.

Если не существует такой точки x_1, в которой степенной ряд расходится, то $R = \infty$ и тогда множество X не ограничено. Пусть x — любое число, удовлетворяющее условию $|x| < R$ или $-R < x < R$; тогда по теореме Абеля при этих значениях x имеет место абсолютная сходимость ряда (8.24).

Возьмем теперь любое x, для которого $|x| > R$, если $R < \infty$. Такие значения x находятся вне промежутка сходимости X и в этих точках ряд расходится.

Теорема полностью доказана.

Число R и интервал $(-R, R)$ называются соответственно *радиусом сходимости и интервалом сходимости степенного ряда*. Всякий степенной ряд имеет свой радиус сходимости $R \geqslant 0$. При $x = \pm R$ вопрос о сходимости должен рассматриваться конкретно для каждого ряда.

Способ определения радиуса сходимости степенного ряда указывает следующая теорема.

Теорема 8.12. *Если для степенного ряда* (8.24) *существует предел*

$$L = \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| \neq 0, \tag{8.29}$$

то радиус сходимости этого ряда определяется формулой $R = 1/L$.

Доказательство. Возьмем ряд $\sum\limits_{n=0}^{\infty} |a_n x^n|$ и применим к нему признак Даламбера (теорема 8.4 для рядов с неотрицательными членами). Этот ряд сходится, если

$$\lim_{n \to \infty} \left| \frac{a_{n+1} x^{n+1}}{a_n x^n} \right| = \lim_{n \to \infty} |x| \left| \frac{a_{n+1}}{a_n} \right| = |x| L < 1.$$

Отсюда в силу теоремы 8.11 следует, что степенной ряд (8.24) сходится абсолютно при $|x| < 1/L$, и, значит, $R = \dfrac{1}{L} = \lim\limits_{n \to \infty} \left| \dfrac{a_n}{a_{n+1}} \right|$, что и требовалось доказать.

Заметим, что если предел L в (8.29) равен нулю, то степенной ряд сходится на всей числовой прямой, т.е. $R = \infty$.

Рассмотрим примеры на определение радиуса сходимости степенных рядов.

Пример 1. $\sum\limits_{n=0}^{\infty} \dfrac{x^n}{n!}$.

Согласно теореме 8.12, радиус сходимости этого ряда определяется по формуле

$$R = \lim_{n \to \infty} \frac{(n+1)!}{n!} = \lim_{n \to \infty} (n+1) = \infty,$$

т.е. данный ряд абсолютно сходится на всей числовой прямой.

Пример 2. $\sum\limits_{n=1}^{\infty} \dfrac{x^n}{n^\alpha}$, $\alpha > 0$.

Радиус сходимости

$$R = \lim_{n\to\infty} \frac{(n+1)^\alpha}{n^\alpha} = \lim_{n\to\infty} (1+1/n)^\alpha = 1.$$

Радиус сходимости данного ряда $R = 1$.

Выясним вопрос о сходимости ряда в точке $|x| = R = 1$. При подстановке в степенной ряд значения $|x| = 1$ получаем числовой ряд $\sum\limits_{n=1}^{\infty} \dfrac{1}{n^\alpha}$, который имеет различный характер сходимости в зависимости от α (см. параграф 8.2, пример 7).

а) При $\alpha \leqslant 1$ ряд x^n/n^α сходится условно на отрезке $[-1, 0]$ (как знакопеременный ряд), а на интервале $(-1, 1)$ он сходится абсолютно (так как $|x^n| > |x^n|/n^\alpha$ и ряд $\sum\limits_{n=1}^{\infty} |x^n|$ сходится как сумма геометрической прогрессии со знаменателем $|x| < 1$).

б) При $\alpha > 1$ ряд x^n/n^α сходится абсолютно на отрезке $[-1, 1]$.

Пример 3. $\sum\limits_{n=0}^{\infty} \dfrac{x^n}{a^n}$, $a > 0$.

Получаем: $R = \lim\limits_{n\to\infty} \dfrac{a^{n+1}}{a^n} = a$. При $x = \pm a$ получаем, что необходимое условие сходимости числового ряда не соблюдается (см. параграф 8.1, условие 4°). Следовательно, данный ряд сходится абсолютно на интервале $(-a, a)$ как сумма геометрической прогрессии со знаменателем меньше единицы.

Пример 4. $\sum\limits_{n=1}^{\infty} n^n x^n$.

Радиус сходимости

$$R = \lim_{n\to\infty} \frac{n^n}{(n+1)^{n+1}} = \lim_{n\to\infty} \left[\frac{1}{(1+1/n)^n} \frac{1}{n+1}\right] = \frac{1}{e} \cdot 0 = 0.$$

Следовательно, данный ряд сходится лишь в точке $x = 0$.

Свойства степенных рядов

Вообще говоря, сумма степенного ряда является функцией от переменной x:

$$f(x) = \sum_{n=0}^{\infty} a_n x^n. \tag{8.30}$$

Пусть интервал сходимости этого ряда $(-R, R)$. Тогда говорят, что функция $f(x)$ *разлагается в степенной ряд на интервале* $(-R, R)$.

Степенные ряды обладают рядом свойств; два из них мы приведем без доказательства.

1°. *Степенной ряд можно дифференцировать почленно на промежутке его сходимости, так что*

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1} = \tag{8.31}$$
$$= a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1} + \ldots.$$

При этом интервал сходимости ряда (8.31) тот же, что и для ряда (8.30). Таким же образом можно вычислить производные любого порядка.

2°. *Степенной ряд можно интегрировать почленно в интервале его сходимости* $-R < x < R$, *т.е. при* $x \in \in (-R, R)$,

$$\int_a^x f(x)\,dx = \int_a^x \left(\sum_{n=0}^{\infty} a_n x^n\right) dx = \sum_{n=0}^{\infty} \int_a^x a_n x^n\,dx. \tag{8.32}$$

Особый интерес предствляет интегрирование степенного ряда (8.30) по отрезку $[0, x]$, где $|x| < R$:

$$\int_0^x f(x)\,dx = a_0 x + \frac{1}{2} a_1 x^2 + \frac{1}{3} a_2 x^3 + \cdots + \tag{8.33}$$
$$+ \frac{1}{n+1} a_n x^{n+1} + \cdots = \sum_{n=0}^{\infty} \frac{1}{n+1} a_n x^{n+1}.$$

Замечание. Из свойства 1° следует, что сумма степенного ряда непрерывна на интервале его сходимости.

Разложение функций в степенные ряды

Теорема 8.13. *Если функция $f(x)$ может быть разложена на интервале $(-R, R)$ в степенной ряд, то это разложение единственно.*

Доказательство. Так как по условию ряд (8.30) сходится на интервале $(-R, R)$ и $f(x)$ — его сумма, то в силу свойства 1° этот степенной ряд можно дифференцировать почленно на указанном интервале сколько угодно раз. Тогда, дифференцируя n раз равенство (8.30), получаем

$$f^{(n)}(x) = a_n n! + 2 \cdot 3 \ldots (n+1) a_{n+1} x + \ldots, \quad n = 0, 1, 2, \ldots,$$

откуда при $x = 0$ находим $f^{(n)}(0) = a_n n!$, или

$$a_n = \frac{f^{(n)}(0)}{n!}, \; n = 0, 1, 2, \ldots. \tag{8.34}$$

Таким образом, коэффициенты степенного ряда (8.30) однозначно определяются формулами (8.34), что и требовалось доказать.

Подстановка полученных коэффициентов (8.34) в формулу (8.30) дает вид разложения функции $f(x)$ в степенной ряд:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \cdots = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n. \tag{8.35}$$

Ряд (8.35) называется *рядом Маклорена* для функции $f(x)$.

Для любой бесконечно дифференцируемой функции можно составить ряд Маклорена (8.35). Установим теперь связь между формулой Маклорена и рядом (8.35). Как известно, для любой $(n+1)$ раз дифференцируемой функции справедлива формула Маклорена [см. формулу (5.12)]:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + R_n(x),$$

где $R_n(x)$ — остаточный член в форме Лагранжа:

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}x^{n+1}, \qquad \xi = \theta x, \quad 0 < \theta < 1. \tag{8.36}$$

Пусть S_n — частичная сумма ряда Маклорена (8.35); нетрудно видеть, что формула Маклорена (8.36) может быть представлена в виде

$$f(x) = S_n(x) + R_n(x). \tag{8.37}$$

Из представления (8.37) почти очевидным образом следует теорема о сходимости ряда Маклорена.

Теорема 8.14. *Для того чтобы для бесконечно дифференцируемой функции $f(x)$ имело место разложение* (8.35) *в ряд Маклорена на интервале $(-R, R)$, необходимо и достаточно, чтобы остаточный член формулы Маклорена для этой функции стремился к нулю на указанном интервале при $n \to \infty$:*

$$\lim_{n\to\infty} R_n(x) = 0. \tag{8.38}$$

Доказательство. Необходимость вытекает из формулы (8.37): если $S_n(x) \to f(x)$ при $n \to \infty$, то для всех x на интервале $(-R, R)$

$$\lim_{n\to\infty} R_n(x) = f(x) - \lim_{n\to\infty} S_n(x) = f(x) - f(x) = 0.$$

Достаточность: пусть $R_n(x) \to 0$ при $n \to \infty$ для всех $x \in (-R, R)$. Тогда из формулы (8.37) следует, что

$$\lim_{n\to\infty} S_n(x) = \lim_{n\to\infty} [f(x) - R_n(x)] = \lim_{n\to\infty} f(x) - 0 = f(x).$$

Разложение некоторых элементарных функций в ряд Маклорена

Рассмотрим разложения известных элементарных функций в ряд Маклорена вида (8.35). Как следует из изложенного ранее, для того чтобы разложить функцию в степенной ряд, следует: найти коэффициенты (8.34) ряда, определить радиус R его сходимости и проверить выполнение условия (8.38) на интервале $(-R, R)$.

1. $f(x) = e^x$.

При $x = 0$ получаем $f^{(n)}(0) = e^0 = 1$, откуда по формулам (8.34) $a_n = 1/n!$.

Далее определяем радиус сходимости степенного ряда с найденными коэффициентами

$$R = \lim_{n\to\infty} \frac{a_n}{a_{n+1}} = \lim_{n\to\infty} \frac{(n+1)!}{n!} = \lim_{n\to\infty} (n+1) = \infty,$$

т.е. степенной ряд сходится на всей числовой прямой.

Теперь докажем, что e^x — это сумма ряда $\sum\limits_{n=0}^{\infty} \frac{1}{n!}x^n$, для чего проведем оценку остаточного члена R_n согласно формуле (8.36):

$$|R_n(x)| = |x|^{n+1} \frac{e^\xi}{(n+1)!} < |x|^{n-1} \frac{\exp|x|}{(n+1)!},$$

так как $\xi < |x|$. При фиксированном x правая часть этого неравенства стремится к нулю при $n \to \infty$, а значит, и $R_n \to 0$ при $n \to \infty$ для любого x, т.е. функция e^x является суммой ряда:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}. \qquad (8.39)$$

2. $f(x) = \sin x$.

Используя формулу для производных $f^{(n)}(x) = \sin\left(x + \frac{n\pi}{2}\right)$ и последовательно выполнив все этапы по аналогии с предыдущим случаем, получим, что на всей числовой прямой имеет место разложение

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}. \qquad (8.40)$$

3. $f(x) = \cos x$.

На всей числовой прямой справедливо разложение

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}. \qquad (8.41)$$

Согласно теореме 8.12 ряды (8.40) и (8.41) сходятся абсолютно на всей числовой прямой.

Это разложение можно легко получить почленным дифференцированием (8.40) в силу свойства 1° степенного ряда.

4. $f(x) = \ln(1+x)$.

Здесь будем исходить из суммы бесконечной геометрической прогрессии

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots = \sum_{n=0}^{\infty} (-1)^n x^n, \qquad (8.42)$$

этот ряд сходится при $|x| < 1$. Воспользуемся свойством 2° сходящегося степенного ряда и проинтегрируем его почленно:

$$\int_0^x \frac{dx}{1+x} = \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}. \qquad (8.43)$$

Радиус сходимости этого ряда такой же, как и для ряда (8.42). При $x = 1$ получаем знакопеременный числовой ряд, рассмотренный в примере 1 параграфа 8.3; он сходится по признаку Лейбница. При $x = -1$ имеем гармонический ряд (см. пример 7, параграф 8.2) с обратным знаком, этот ряд расходится. Следовательно, разложение (8.42) имеет место на полуинтервале $(-1, 1]$.

5. $f(x) = \operatorname{arctg} x$.

Здесь также применим почленное интегрирование степенного ряда. Заменив в формуле (8.42) всюду x на x^2, получим разложение функции $1/(1+x^2)$ в ряд, сходящийся на интервале $(-1, 1)$:

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \cdots = \sum_{n=0}^{\infty} (-1)^n x^{2n}.$$

Проинтегрируем этот ряд почленно в пределах от 0 до x:

$$\int_0^x \frac{dx}{1+x^2} = \operatorname{arctg} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}. \qquad (8.44)$$

Нетрудно убедиться, что разложение (8.44) имеет место на промежутке $(-1, 1)$.

8.5. Ряды Фурье

Периодические величины и гармонический анализ

В различных приложениях науки и техники часто приходится иметь дело с периодическими явлениями, повторяющимися или воспроизводящимися через определенный промежуток времени. Такие явления описываются периодическими функциями.

Определение 1. Функция $f(x)$ называется *периодической* на промежутке X, если существует такое наименьшее число T, называемое *периодом* этой функции, что для любых x, $x+T \in X$ выполняется равенство

$$f(x+T) = f(x). \tag{8.45}$$

Заметим, что если T — период функции $f(x)$, то: при $x + nT \in X$, где n — целое число, также выполняется равенство типа (8.45):

$$f(x+nT) = f(x).$$

Поэтому часто любое число nT называют периодом функции $f(x)$ в широком понимании этого термина.

Простейшими периодическими функциями являются $\sin x$ и $\cos x$ с периодом 2π. Из простых периодических функций можно составить более сложные. Например, рассмотрим функции

$$f_0(x) = A_0,\ f_1(x) = A_1 \sin\left(\frac{2\pi}{T}x + \alpha_1\right),$$
$$f_2(x) = A_2 \sin\left(\frac{4\pi}{T}x + \alpha_2\right),\ f_3(x) = A_3 \sin\left(\frac{6\pi}{T}x + \alpha_3\right), \ldots, \tag{8.46}$$

где A_n и α_n — постоянные величины, называемые соответственно *амплитудами* и *сдвигами фаз*. Периоды этих функций соответственно равны: любое число для постоянной величины

A_0, T, $T/2$, $T/3$, ... Если сложить эти функции, то снова получим периодическую функцию с периодом T, но гораздо более сложного вида, чем каждая из составляющих. Геометрически это означает, что график сложной периодической функции является результатом наложения ряда простых синусоид.

Естественно поставить обратный вопрос: можно ли данную периодическую функцию $f(x)$ с периодом T представить в виде конечной или бесконечной суммы функций вида (8.46)? Оказывается, такой процесс осуществим для широкого класса функций, и он называется *гармоническим анализом;* при этом отдельные функции вида (8.46) называются *гармоническими составляющими* или просто *гармониками* функции $f(x)$.

Тригонометрический ряд

Определение 2. Ряд вида

$$\frac{a_0}{2} + a_1\cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \\ + a_3 \cos 3x + b_3 \sin 3x + \cdots = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx) \tag{8.47}$$

называется *тригонометрическим рядом*, а числа a_0, a_1, b_1, a_2, b_2, ... — его коэффициентами.

В тригонометрический ряд (8.47) входят функции

$$\frac{1}{2},\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x, \ldots, \cos nx,\ \sin nx,\ \ldots, \tag{8.48}$$

которые являются периодическими с общим периодом 2π. Следовательно, любая частичная сумма тригонометрического ряда (8.47) будет так же периодической функцией с периодом 2π. Отсюда следует, что если этот ряд сходится на отрезке $[-\pi, \pi]$, то он, будучи также периодической функцией с периодом 2π (как предел частичных сумм), сходится на всей числовой прямой.

Ортогональность системы тригонометрических функций

Важным свойством тригонометрических функций системы (8.48), широко используемым в математике, является их *ортогональность* на отрезке $[-\pi, \pi]$. Она понимается в следующем

18 Зак. 6853

смысле: интеграл по этому отрезку от произведения двух любых функций системы (8.48) равен:

нулю, если в подынтегральную функцию входят различные функции системы (8.48);

положительному числу, если подынтегральная функция состоит из квадрата любой функции системы (8.48).

Проверим указанное свойство для всех видов функций системы.

а) При $k \neq n$; $k, n = 0, 1, 2, \ldots$,

$$\int_{-\pi}^{\pi} \cos kx \cos nx\, dx = \frac{1}{2}\int_{-\pi}^{\pi} [\cos(k+n)x + \cos(k-n)x]\, dx =$$

$$= \frac{1}{2}\left[\frac{\sin(k+n)x}{k+n} + \frac{\sin(k-n)x}{k-n}\right]\Bigg|_{-\pi}^{\pi} = 0;$$

$$\int_{-\pi}^{\pi} \sin kx \sin nx\, dx = \frac{1}{2}\int_{-\pi}^{\pi} [\cos(k-n)x - \cos(k+n)x]\, dx = 0;$$

$$\int_{-\pi}^{\pi} \sin kx \cos nx\, dx = \frac{1}{2}\int_{-\pi}^{\pi} [\sin(k+n)x + \sin(k-n)x] dx =$$

$$= \frac{1}{2}\left[\frac{\cos(n-k)x}{n-k} - \frac{\cos(n+k)x}{n+k}\right]\Bigg|_{-\pi}^{\pi} = 0. \qquad (8.49)$$

б) При $k = n$; $k, n = 1, 2, 3, \ldots$,

$$\int_{-\pi}^{\pi} \cos^2 nx\, dx = \frac{1}{2}\int_{-\pi}^{\pi} (1 + \cos 2nx)\, dx = \frac{1}{2}\left[x + \frac{1}{2n}\sin 2nx\right]\Bigg|_{-\pi}^{\pi} = \pi; \qquad (8.50)$$

$$\int_{-\pi}^{\pi} \sin^2 nx\, dx = \frac{1}{2}\int_{-\pi}^{\pi} (1 - \cos 2nx)\, dx = \frac{1}{2}\left[x - \frac{1}{2n}\sin 2nx\right]\Bigg|_{-\pi}^{\pi} = \pi.$$

в) $$\int_{-\pi}^{\pi} dx = x\Big|_{-\pi}^{\pi} = 2\pi. \qquad (8.51)$$

Равенства (8.49) и (8.50) являются свойствами ортогональности системы тригонометрических функций на отрезке $[-\pi, \pi]$.

Ряд Фурье[1]

Для тригонометрического ряда, как и для степенного ряда, можно установить условия разложения функций.

Теорема 8.15. *Если функция $f(x)$ интегрируема на отрезке $[-\pi, \pi]$ и разлагается в тригонометрический ряд*

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx), \qquad (8.52)$$

который можно интегрировать почленно при умножении его на ограниченную функцию, то такое разложение единственно.

Доказательство. Для определения коэффициентов разложения воспользуемся формулами (8.49)–(8.51). Сначала проинтегрируем (8.52) по отрезку $[-\pi, \pi]$; с учетом указанных формул ортогональности получим, что все интегралы в правой части, кроме первого, обращаются в нуль, т.е.

$$\int_{-\pi}^{\pi} f(x)\, dx = \frac{1}{2}a_0 \int_{-\pi}^{\pi} dx = a_0\pi,$$

откуда находим нулевой коэффициент разложения

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\, dx. \qquad (8.53)$$

Для определения коэффициентов a_n и b_n последовательно умножаем обе части разложения (8.52) на $\cos nx$ и $\sin nx$ и интегрируем по отрезку $[-\pi, \pi]$. Согласно соотношениям (8.49) и (8.50), все интегралы в правой части, кроме содержащих квадраты указанных функций, равны нулю, т.е.

$$\int_{-\pi}^{\pi} f(x)\cos nx\, dx = a_n \int_{-\pi}^{\pi} \cos^2 nx\, dx = a_n\pi,$$

[1] Жан Батист Жозеф Фурье (1768–1830) — французский математик.

откуда

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx dx, \qquad (8.54)$$

а также

$$\int_{-\pi}^{\pi} f(x)\sin nx dx = b_n \int_{-\pi}^{\pi} \sin^2 nx dx = b_n \pi,$$

откуда

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx dx. \qquad (8.55)$$

Формулы (8.53)–(8.55) определяют единственным образом все коэффициенты разложения, что и требовалось доказать.

Для функции $f(x)$, интегрируемой на отрезке $[-\pi, \pi]$, числа a_0, a_n, b_n, определяемые формулами (8.53)–(8.55), называются *коэффициентами Фурье*, ряд (8.52) с этими коэффициентами называется *рядом Фурье функции* $f(x)$.

Сходимость ряда Фурье

Для определения сходимости ряда Фурье вводится понятие периодического продолжения функции, заданной на отрезке $[-T, T]$. Будем говорить, что функция $F(x)$, определенная на всей числовой прямой и периодическая с периодом T, является *периодическим продолжением* функции $f(x)$, если $F(x) = f(x)$ на отрезке $[-T, T]$.

Если ряд Фурье сходится к функции $f(x)$ на отрезке $[-\pi, \pi]$, то он сходится на всей числовой прямой к ее периодическому продолжению. Ответ на вопрос, при каких условиях ряд Фурье функции $f(x)$ сходится к этой функции, дает следующая теорема, которая приводится здесь без доказательства.

Теорема 8.16. *Пусть функция $f(x)$ непрерывна вместе со своей производной на отрезке $[-\pi, \pi]$ или они имеют на этом отрезке конечное число точек разрыва. Тогда:*

1) *ряд Фурье функции $f(x)$ сходится на всей числовой прямой; в каждой точке непрерывности $f(x)$ в интервале $(-\pi, \pi)$ сумма ряда равна значению $f(x)$ в этой точке;*

2) *в каждой точке x_b разрыва функции сумма ряда равна полусумме односторонних пределов $f(x)$ в этой точке:*

$$\frac{1}{2}\left[\lim_{x\to x_b-} f(x) + \lim_{x\to x_b+} f(x)\right];$$

3) *на концах отрезка $[-\pi, \pi]$ сумма ряда равна*

$$\frac{1}{2}[f(-\pi) + f(\pi)];$$

4) *для любой точки $x \notin [-\pi, \pi]$ утверждения пп. 1–3 справедливы для периодического продолжения $F(x)$ функции $f(x)$.*

Эта теорема фактически устанавливает классы функций, которые могут быть разложены в ряд Фурье.

Ряды Фурье для четных и нечетных функций

Разложение (8.52) с коэффициентами (8.53)–(8.55) имеет место для произвольной функции $f(x)$, удовлетворяющей условиям теоремы 8.16. Однако особый интерес представляют разложения для четных и нечетных функций, поскольку вид соответствующих рядов Фурье для них существенно упрощается.

Пусть функция $f(x)$ определена и является нечетной на отрезке $[-\pi, \pi]$, т.е. $f(-x) = -f(x)$. Тогда коэффициенты Фурье a_n для нее равны нулю. В самом деле,

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\, dx =$$

$$= \frac{1}{\pi}\left[\int_{-\pi}^{0} f(x)\cos nx\, dx + \int_{0}^{\pi} f(x)\cos nx\, dx\right].$$

В первом интеграле правой части этого равенства сделаем замену переменной: $x = -t$, отсюда $dx = -dt$, $f(x) = -f(t)$, $t = 0$ при $x = 0$, $t = \pi$ при $x = -\pi$. Следовательно,

$$\int_{-\pi}^{0} f(x)\cos nx\, dx = \int_{\pi}^{0} f(t)\cos nt\, dt = -\int_{0}^{\pi} f(x)\cos nx\, dx,$$

откуда и получаем $a_n = 0$; $n = 0, 1, 2, \ldots$.

Используя аналогичный прием, можно показать, что в этом случае коэффициенты b_n вычисляются по формулам

$$b_n = \frac{2}{\pi}\int_0^{\pi} f(x)\sin nx\,dx, \quad n = 1, 2, 3, \ldots. \tag{8.56}$$

Таким образом, нечетная на отрезке $[-\pi, \pi]$ функция $f(x)$ разлагается в ряд Фурье вида

$$f(x) = \sum_{n=1}^{\infty} b_n \sin nx, \tag{8.57}$$

где b_n определяются по формулам (8.56). Аналогично, определенная и четная на отрезке $[-\pi, \pi]$ функция $f(x)$ может быть разложена в ряд Фурье

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx, \tag{8.58}$$

где коэффициенты Фурье a_n определяются по формулам

$$a_n = \frac{2}{\pi}\int_0^{\pi} f(x)\cos nx\,dx, \quad n = 0, 1, 2, \ldots. \tag{8.59}$$

Рассмотрим примеры разложения функций в ряды Фурье на отрезке $[-\pi, \pi]$.

Пример 1. $f(x) = x$.

Эта функция удовлетворяет условиям теоремы 8.16. Поскольку она нечетная на отрезке $[-\pi, \pi]$, то коэффициенты ее разложения, соответствующие четным слагаемым ряда Фурье $a_n = 0$, а b_n находятся по формулам (8.56):

$$b_n = \frac{2}{\pi}\int_0^{\pi} x\sin nx\,dx = \frac{2}{\pi n}\left[-x\cos nx\Big|_0^{\pi} + \int_0^{\pi}\cos nx\,dx\right] =$$

$$= \frac{2}{\pi n}\left[\pi(-1)^{n+1} + \frac{1}{n}\sin nx\Big|_0^{\pi}\right] = (-1)^{n+1}\frac{2}{n}, \quad n = 1, 2, \ldots.$$

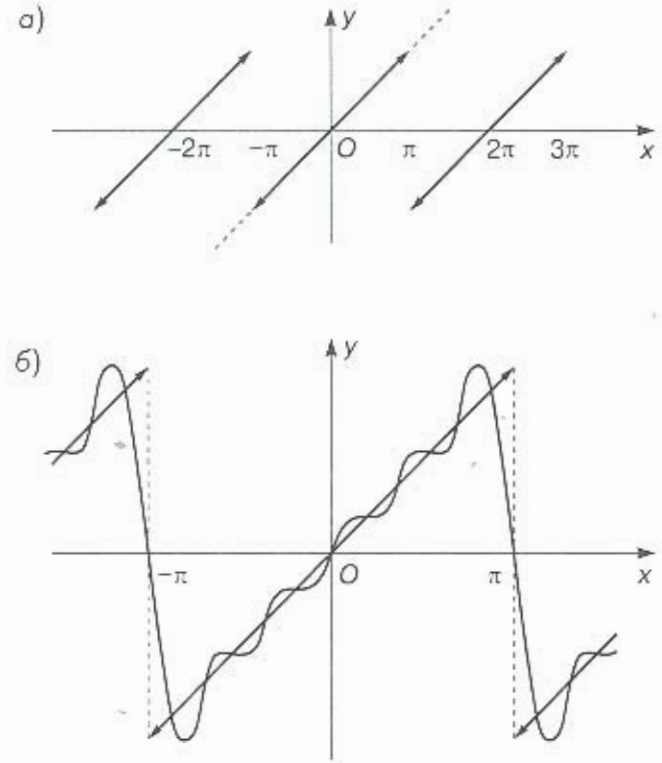

Рис. 8.2

Ряд Фурье для данной функции имеет вид

$$x = 2\left(\frac{\sin x}{1} - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \ldots\right) = 2\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\sin nx}{n}.$$

Это разложение справедливо для интервала $(-\pi, \pi)$; при $x = \pm\pi$, согласно теореме (8.16), сумма ряда равна $\frac{f(-\pi)+f(\pi)}{2} = \frac{-\pi+\pi}{2} = 0$. График периодического продолжения функции $y = x$ на всю числовую прямую и приближение этой функции на отрезке $[-\pi, \pi]$ частичной суммой S_5 ряда Фурье показаны на рис. 8.2, *а* и *б* соответственно.

Пример 2. $f(x) = x^2$.

Данная функция удовлетворяет условиям теоремы 8.16, и потому может быть разложена в ряд Фурье на интервале $(-\pi, \pi)$. Поскольку эта функция четная, то все коэффициенты

Фурье b_n для нее равны нулю, а коэффициенты a_n определяются по формулам (8.59). Получаем:

а) при $n = 0$

$$a_0 = \frac{2}{\pi}\int_0^{\pi} x^2\,dx = \frac{2}{\pi}\frac{x^3}{3}\Big|_0^{\pi} = \frac{2\pi^2}{3};$$

б) при $n = 1, 2, 3, \ldots$ (см. предыдущий пример)

$$a_n = \frac{2}{\pi}\int_0^{\pi} x^2\cos nx\,dx = \frac{2}{\pi n}\left[x^2\sin nx\Big|_0^{\pi} - 2\int_0^{\pi} x\sin nx\,dx\right] =$$

$$= \frac{4}{n^2\pi}\left[x\cos nx\Big|_0^{\pi} - \int_0^{\pi}\cos nx\,dx\right] = (-1)^n\frac{4}{n^2}.$$

Следовательно, для данной функции имеет место разложение в ряд Фурье:

$$x^2 = \frac{\pi^2}{3} - 4\left(\frac{\cos x}{1} - \frac{\cos 2x}{4} + \frac{\cos 3x}{9} - \ldots\right) =$$

$$= \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos nx.$$

Пример 3. $f(x) = e^{ax}$, $a \neq 0$.

Эта функция удовлетворяет условиям теоремы 8.16, и для нее существует разложение в ряд Фурье общего вида с коэффициентами, определяемыми по формулам (8.53)–(8.55). Имеем:

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} e^{ax}\,dx = \frac{e^{a\pi} - e^{-a\pi}}{a\pi} = 2\frac{\operatorname{sh} a\pi}{a\pi},$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} e^{ax}\cos nx\,dx = \frac{1}{\pi}\frac{a\cos nx + n\sin nx}{a^2+n^2}e^{ax}\Big|_{-\pi}^{\pi} =$$

$$= \frac{(-1)^n}{a^2+n^2}\frac{2a\operatorname{sh} a\pi}{\pi}, \quad n = 1, 2, \ldots,$$

$$b_n = \frac{1}{\pi}\int_0^{\pi} e^{ax}\sin nx\,dx = \frac{1}{\pi}\frac{a\sin nx - n\cos nx}{a^2+n^2}e^{ax}\Big|_{-\pi}^{\pi} =$$

$$= \frac{(-1)^{n+1}n}{a^2+n^2}\frac{2\operatorname{sh} a\pi}{\pi}, \quad n = 1, 2, \ldots.$$

Итак, получаем ряд Фурье, сходящийся на интервале $(-\pi, \pi)$ к функции

$$e^{ax} = \frac{2}{\pi}\operatorname{sh} a\pi\left\{\frac{1}{2a} + \sum_{n=1}^{\infty}\frac{(-1)^n}{a^2+n^2}[a\cos nx - n\sin nx]\right\}.$$

Ряд Фурье с произвольным периодом

Рассмотрим функцию $f(x)$, определенную на отрезке $[-l, l]$, где l — произвольное число. Пусть эта функция удовлетворяет условиям теоремы 8.16 на данном отрезке. Введем новую переменную $x = yl/\pi$ и рассмотрим функцию $f(yl/\pi)$. Эта функция определена на отрезке $[-\pi, \pi]$ и удовлетворяет условиям теоремы 8.16, т.е. может быть разложена в ряд Фурье на этом отрезке:

$$f(yl/\pi) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n\cos ny + b_n\sin ny), \tag{8.60}$$

где

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(yl/\pi)dy, \quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(yl/\pi)\cos nydy,$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(yl/\pi)\sin nydy, \quad n = 1, 2, 3, \ldots.$$

Вернемся к старой переменной x: $y = \dfrac{\pi x}{l}$, $dy = \dfrac{\pi}{l}dx$, $x = \pm l$ при $y = \pm\pi$, тогда формула (8.60) перепишется в виде

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n\cos\frac{n\pi x}{l} + b_n\sin\frac{n\pi x}{l}\right), \tag{8.61}$$

где коэффициенты Фурье определяются формулами

$$a_0 = \frac{1}{l}\int_{-l}^{l} f(x)\,dx, \quad a_n = \frac{1}{l}\int_{-l}^{l} f(x)\cos\frac{\pi n x}{l}\,dx,$$
$$b_n = \frac{1}{l}\int_{-l}^{l} f(x)\sin\frac{\pi n x}{l}\,dx, \quad n = 1, 2, 3, \ldots. \tag{8.62}$$

Формула (8.61) представляет собой ряд Фурье для функции $f(x)$ с периодом $2l$.

Глава 9
ФУНКЦИИ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

9.1. Евклидово пространство E^m

Евклидова плоскость и евклидово пространство

Как мы знаем, множество всех упорядоченных пар вещественных чисел (x, y) называется координатной плоскостью и каждая точка на ней характеризуется парой своих координат: $M(x, y)$.

Определение 1. Координатная плоскость называется *евклидовой плоскостью*, если расстояние между двумя любыми точками $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$ определено по формуле

$$\rho(M_1, M_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \tag{9.1}$$

Аналогично вводится и понятие *евклидова пространства*. В этом случае каждая точка координатной плоскости характеризуется тройкой чисел, и тогда расстояние между двумя любыми точками пространства $M_1(x_1, y_1, z_1)$ и $M_2(x_2, y_2, z_2)$ определяется формулой

$$\rho(M_1, M_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}. \tag{9.2}$$

Стало быть, евклидова плоскость и евклидово пространство определяются путем измерения расстояния между двумя любыми своими точками.

Понятия m-мерного координатного пространства и m-мерного евклидова пространства

Определение 2. Множество всевозможных упорядоченных совокупностей m действительных чисел $(x_1, x_2, x_3, \ldots, x_m)$ называется *m-мерным координатным пространством* A^m.

Каждую упорядоченную совокупность $(x_1, x_2, \ldots, x_m)$ называют *точкой* этого пространства и обозначают одной буквой M. При этом числа $x_1, x_2, \ldots, x_m$ называются *координатами* точки M, что символически записывается следующим образом: $M(x_1, x_2, \ldots, x_m)$.

Определение 3. Координатное пространство A^m называется *m-мерным евклидовым пространством* E^m, если между двумя любыми точками $M'(x_1', x_2', \ldots, x_m')$ и $M''(x_1'', x_2'', \ldots, x_m'')$ пространства A^m расстояние $\rho(M', M'')$ определено по формуле

$$\rho(M', M'') = \sqrt{(x_1' - x_1'')^2 + (x_2' - x_2'')^2 + \cdots + (x_m' - x_m'')^2}. \tag{9.3}$$

Очевидно, что введенные понятия m-мерного координатного пространства A^m и m-мерного евклидова пространства E^m являются обобщениями понятий соответственно координатных плоскости и пространства и евклидовых плоскости и пространства.

9.2. Множества точек евклидова пространства E^m

Примеры множеств евклидова пространства E^m

Будем обозначать символом $\{M\}$ некоторое множество точек m-мерного пространства E^m. Рассмотрим некоторые примеры множеств в этом пространстве.

1. Множество $\{M\}$ всевозможных точек, координаты $x_1, x_2, \ldots, x_m$ которых удовлетворяют неравенству

$$(x_1 - x_1^0)^2 + (x_2 - x_2^0)^2 + \cdots + (x_m - x_m^0)^2 \leqslant R^2, \tag{9.4}$$

называется *m-мерным шаром* радиуса R с центром в точке $M_0(x_1^0, x_2^0, \ldots, x_m^0)$.

Этот пример является m-мерным обобщением соответственно круга на евклидовой плоскости и шара в трехмерном

евклидовом пространстве, которые задаются следующими неравенствами:

$$x^2 + y^2 \leqslant R^2, \quad x^2 + y^2 + z^2 \leqslant R^2.$$

Неравенство (9.4) можно переписать, с учетом формулы (9.3), в виде:

$$\rho(M, M_0) \leqslant R. \tag{9.5}$$

В случае строгого неравенства $\rho(M, M_0) < R$ множество $\{M\}$ называется *открытым m-мерным шаром.*

2. Множество $\{M\}$ таких точек, что расстояние от каждой из них до некоторой точки M_0 удовлетворяет равенству $\rho(M, M_0) = R$, называется *m-мерной сферой* радиуса R с центром в точке M_0.

Аналогия: для плоскости — окружность $(x - x_0)^2 + (y - y_0)^2 = R^2$ радиуса R с центром в точке $M_0(x_0, y_0)$, для пространства — сфера $(x-x_0)^2+(y-y_0)^2+(z-z_0)^2 = R^2$ радиуса R с центром в точке $M_0(x_0, y_0, z_0)$.

3. Множество $\{M\}$ точек, координаты которых удовлетворяют неравенствам

$$|x_1 - x_1^0| \leqslant d_1, \ |x_2 - x_2^0| \leqslant d_2, \ \ldots, \ |x_m - x_m^0| \leqslant d_m, \tag{9.6}$$

называется *m-мерным параллелепипедом* с центром в точке $M_0(x_1^0, x_2^0, \ldots, x_m^0)$; при этом числа $d_1, d_2, \ldots, d_m$ называются *полуразмерами параллелепипеда.*

Если координаты точек множества $\{M\}$ удовлетворяют строгим неравенствам $|x_i - x_i^0| < d_i$, $i = 1, 2, \ldots, m$, то это множество называется *открытым m-мерным параллелепипедом.*

Виды множеств пространства E^m

Нам понадобится понятие ε-окрестности точки M_0 евклидова m-мерного пространства E^m.

Определение 1. Будем называть *ε-окрестностью точки* $M_0(x_1^0, x_2^0, \ldots, x_m^0)$ открытый m-мерный шар радиуса ε с центром в точке M_0.

Определение 2. Любой открытый параллелепипед с центром в точке M_0 $(x_1^0, x_2^0, \ldots, x_m^0)$ называется *прямоугольной окрестностью точки* M_0.

Очевидно, что любая ε-окрестность точки M_0 евклидова m-мерного пространства содержит некоторую прямоугольную окрестность этой точки и наоборот, любая прямоугольная окрестность точки M_0 содержит некоторую ε-окрестность этой точки.

Введем следующие фундаментальные понятия.

Определение 3. Точка M называется *внутренней точкой* множества $\{M\}$, если для этой точки существует некоторая ε-окрестность, все точки которой принадлежат $\{M\}$.

Определение 4. Точка M называется *граничной точкой* множества $\{M\}$, если любая ее ε-окрестность содержит как точки множества $\{M\}$, так и точки, не принадлежащие ему.

Определение 5. Множество $\{M\}$ пространства E^m называется *открытым,* если все точки этого множества внутренние.

Определение 6. Множество $\{M\}$ пространства E^m называется *замкнутым,* если все граничные точки этого множества принадлежат ему.

Заметим, что если все точки множества $\{M\}$ находятся внутри некоторого шара, то это множество называется *ограниченным.*

Существует еще один важный вид множества, широко используемый в математике и в различных приложениях. В связи с этим введем соответствующие определения.

Определение 7. Будем называть *непрерывной кривой* L в пространстве E^m множество точек $\{M\}$, координаты которых являются непрерывными функциями параметра t:

$$x_1 = \varphi_1(t), \ x_2(t) = \varphi_2(t), \ldots, x_m(t) = \varphi_m(t), \quad \alpha \leqslant t \leqslant \beta. \tag{9.7}$$

Говорят, что точки M' и M'' пространства E^m *можно соединить непрерывной кривой* L, если существует такая непрерывная кривая вида (9.7), что координаты этих точек удовлетворяют системе равенств

$$x_i' = \varphi_i(\alpha), \quad x_i'' = \varphi_i(\beta), \qquad i = 1, 2, \ldots, m. \tag{9.8}$$

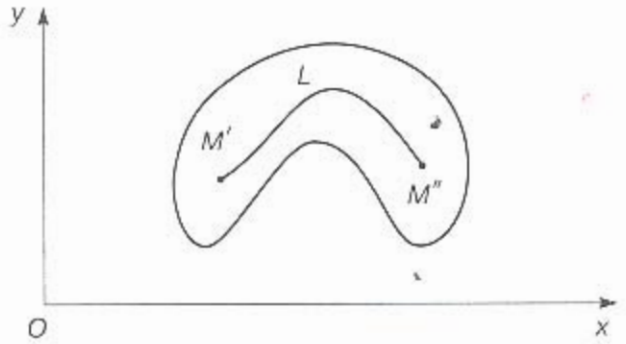

Рис. 9.1

Определение 8. Множество $\{M\}$ пространства E^m называемся *связным*, если две его любые точки можно соединить непрерывной кривой, все точки которой принадлежат множеству $\{M\}$.

Пример связного множества на евклидовой плоскости показан на рис. 9.1.

Последовательности точек в пространстве E^m

Пусть каждому числу n натурального ряда чисел $0, 1, 2, \ldots$ ставится в соответствие точка M_n евклидова пространства E^m; пронумерованный таким образом ряд точек M_1, M_2, $M_3, \ldots$ называется *последовательностью* точек евклидова пространства E^m и обозначается $\{M_n\}$.

Определение 9. Последовательность точек $\{M_n\}$ евклидова пространства E^m называется *сходящейся*, если существует такая точка A, что для любого числа $\varepsilon > 0$ можно указать номер N, начиная с которого (при $n > N$) все точки этой последовательности будут находиться в ε-окрестности точки A, т.е. $\rho(M_n, A) < \varepsilon$. Тогда число A называется *пределом последовательности* $\{M_n\}$.

Для обозначения предела A последовательности $\{M_n\}$ используют символы:

$$\lim_{n\to\infty} M_n = A, \quad \text{или} \quad M_n \to A \quad \text{при} \quad n \to \infty.$$

Для сходящейся последовательности $\{M_n\}$ точек пространства E^m справедлива следующая очевидная теорема.

Теорема 9.1. *Для того, чтобы последовательность точек $\{M_n\}$ евклидова пространства E^m сходилась к точке $A(a_1, a_2, a_3, \ldots, a_m)$, необходимо и достаточно, чтобы последовательности координат $\{x_i^{(n)}\}, \{x_2^{(n)}\}, \ldots, \{x_m^{(n)}\}$ сходились к соответствующим координатам a_i $(i = 1, 2, \ldots, m)$ точки A.*

Определим, что началом координат евклидова пространства E^m является точка, все координаты которой равны нулю: $O(0, 0, \ldots, 0)$.

Определение 10. Последовательность точек $\{M_n\}$ m-мерного евклидова пространства называется *ограниченной*, если существует такое число a, что для всех точек этой последовательности выполняется неравенство $\rho(M_n, O) \leqslant a$, т.е. все ее точки содержатся внутри или на границе шара конечного радиуса a с центром в начале координат.

Понятие функции нескольких переменных

Введем понятие функции нескольких переменных.

Определение 11. Пусть каждой точке M из множества точек $\{M\}$ евклидова пространства E^m по какому-либо закону ставится в соответствие некоторое число u из числового множества U. Тогда будем говорить, что на множестве $\{M\}$ задана *функция* $u = f(M)$. При этом множества $\{M\}$ и U называются соответственно *областью определения* (задания) и *областью изменения функции* $f(M)$.

Как и в случае функции одной переменной, область определения функции нескольких переменных либо задается априори (может быть, в силу каких-либо ограничений конкретной задачи), либо определяется из формулы функциональной зависимости f путем соблюдения корректности выполнения соответствующих математических операций.

Как известно, функция одной переменной $y = f(x)$ изображается на плоскости в виде линии. В случае двух переменных область определения $\{M\}$ функции $z = f(x, y)$ представляет собой некоторое множество точек на координатной плоскости Oxy (рис. 9.2) и тогда сама функция изображается в виде некоторой поверхности.

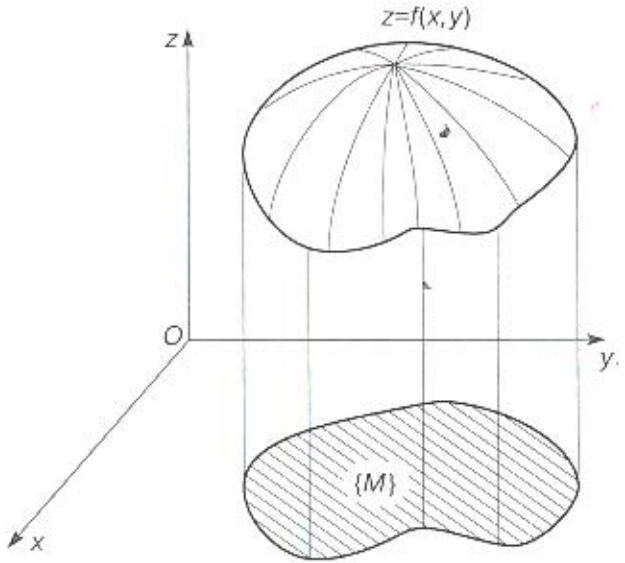

Рис. 9.2

Рассмотрим примеры функций нескольких переменных и найдем их области определения.

1) $z = x^2 + y^2$ — поверхность в евклидовом пространстве E^3.

Областью определения этой функции является все множество точек плоскости Oxy. Область значений этой функции — промежуток $[0, \infty)$. Данная функция представляет собой параболоид (рис. 9.3): в вертикальных сечениях этой поверхности плоскостями Oxz и Oyz получаются соответственно параболы $z = x^2$ и $z = y^2$.

2) $z^2 = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$ — поверхность в евклидовом пространстве E^3.

Область определения данной функции — все множество точек евклидова пространства E^2, или плоскости Oxy. Эта функция является так называемым эллиптическим конусом с вершиной на начале координат $O(0, 0, 0)$; приведенная формула суммирует две функции, задающие две его симметричные относительно плоскости Oxy части:

$$z = \pm\sqrt{\frac{x^2}{a^2} + \frac{y^2}{b^2}}.$$

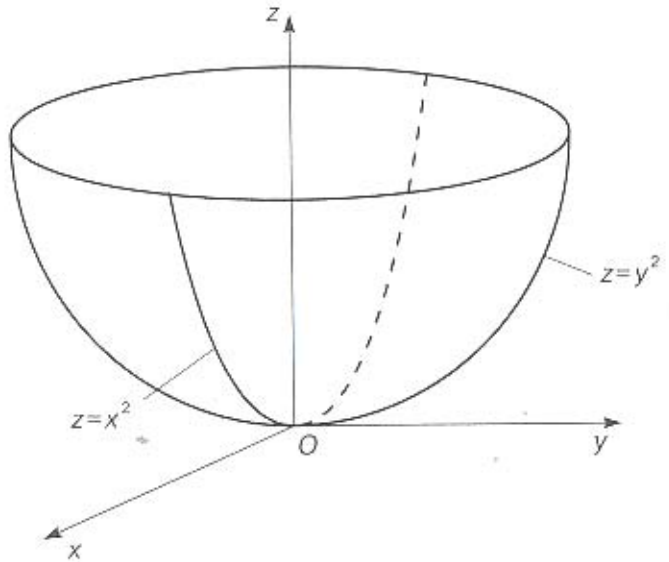

Рис. 9.3

3) $u - f(M) = \sqrt{1 - \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2}}$.

Поскольку выражение под знаком квадратного корня должно быть неотрицательным, то областью определения этой функции является замкнутый эллипсоид с полуосями a, b и c:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leqslant 1.$$

4) $u = 1/\sqrt{1 - 2x + 2y - z}$.

В область определения этой функции входят все точки евклидова пространства E^3, удовлетворяющие неравенству $1 - 2x + 2y - z > 0$, т.е. открытое полупространство под плоскостью $z = -2x + 2y + 1$.

5) $u = f(M) = 1/\sqrt{1 - x_1^2 - x_2^2 - x_3^2 - \cdots - x_m^2}$.

Данная функция определена на множестве точек евклидова пространства E^m. Так как квадратный корень формулы задания функции стоит в знаменателе, то областью определения этой функции является открытый m-мерный шар радиуса $R = 1$.

Некоторые виды функций нескольких переменных

Приведем наиболее часто встречающиеся в различных приложениях виды функций нескольких переменных.

1. Функция вида

$$u = a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_mx_m + b, \tag{9.9}$$

где $a_1, a_2, \ldots, a_m, b$ — постоянные числа, называется *линейной функцией*, или *гиперплоскостью* в евклидовом пространстве E^{m+1}.

2. Функция вида

$$u = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2a_{13}x_1x_3 + \cdots + a_{mm}x_m^2 = \frac{1}{2}\sum_{i,j=1}^{m} a_{ij}x_ix_j, \tag{9.10}$$

где a_{ij} — постоянные числа, называется *квадратичной формой от переменных* $x_1, x_2, \ldots, x_m$.

3. Одним из базовых понятий экономической теории является *функция полезности*, выражающая полезность приобретения m разновидностей приобретенных товаров. Чаще всего она употребляется в следующих формах:

$$u = \sum_{i=1}^{m} a_i \ln(x_i - c_i), \qquad a_i > 0, \quad x_i - c_i > 0, \quad c_i \geqslant 0, \tag{9.11}$$

— *логарифмическая функция;*

$$u = \sum_{i=1}^{m} a_i(x_i - c_i)^{1-b_i}/(1 - b_i), \tag{9.12}$$

$$a_i > 0, \quad 0 < b_i < 1, \quad x_i - c_i > 0, \quad c_i \geqslant 0,$$

— *функция постоянной эластичности.*

4. *Функция Кобба–Дугласа* — производственная функция, показывающая объем выпуска продукции Q при затратах капитала K и трудовых ресурсов L. Для случая двух переменных она имеет вид

$$Q = AK^{\alpha}L^{1-\alpha}, \tag{9.13}$$

где $A > 0$ — параметр производительности конкретно взятой технологии, $0 < \alpha < 1$ — доля капитала в доходе.

9.3. Предел функции нескольких переменных

Понятие предела функции нескольких переменных

Рассмотрим функцию $f(M)$, определенную на множестве $\{M\}$ евклидова пространства E^m. Пусть $A(a_1, a_2, \ldots, a_m)$ — точка этого пространства, в любой ε-окрестности которой содержится хотя бы одна точка множества $\{M\}$, отличная от A.

Определение 1. Число b называется *пределом функции* $f(M)$ в точке A (пределом функции при $M \to A$), если для любой последовательности точек $\{M_n\}$ из множества $\{M\}$, сходящейся к точке $A(M_n \neq A)$, соответствующая последовательность значений функции $\{f(M_n)\}$ сходится к b.

Другое определение предела функции, с использованием терминологии "ε–δ", равносильно приведенному выше.

Определение 2. Число b называется *пределом функции* $f(M)$ в точке A, если для любого числа $\varepsilon > 0$ можно найти такое число $\delta > 0$, что для всех точек M множества $\{M\}$ из δ-окрестности точки A (удовлетворяющих неравенству $\rho(M, A) < \delta$) выполняется неравенство $|f(M) - b| < \varepsilon$.

Для обозначения предела функции в точке используется символика

$$\lim_{M \to A} f(M) = b \text{ или } \lim_{\substack{x_1 \to a_1 \\ x_2 \to a_2 \\ \cdots\cdots\cdots \\ x_m \to a_m}} f(x_1, x_2, \ldots, x_m) = b.$$

Для функций нескольких переменных, имеющих предел в данной точке, справедлива следующая теорема об арифметических операциях над ними.

Теорема 9.2. *Пусть функции $f(M)$ и $g(M)$, определенные на одном множестве $\{M\}$ имеют соответственно пределы b и c в точке A. Тогда функции $f(M) \pm g(M)$, $f(M)g(M)$ и $f(M)/g(M)$ (при $c \neq 0$) имеют пределы в точке A, равные соответственно $b \pm c$, bc и b/c.*

Доказательство этой теоремы следует из теоремы 9.1 и соответствующих теорем главы 2.

Данное выше определение предела функции нескольких переменных легко распространяется на случай, когда точка M стремится к бесконечности.

Определение 3. Число b называется *пределом функции* $f(M)$ при $M \to \infty$, если для любого числа $\varepsilon > 0$ можно найти такое число $R > 0$, что для всех точек M из множества $\{M\}$, удовлетворяющих условию $\rho(O, M) > R$ (находящихся вне R-окрестности точки O), выполняется неравенство $|f(M) - b| < \varepsilon$.

Задача о нахождении предела функции нескольких переменных является более сложной, чем для функции одной переменной, в особенности в случае неопределенности типа $\frac{0}{0}$. Рассмотрим примеры.

Пример 1. Найти

$$\lim_{\substack{x\to 0\\ y\to 0}} \frac{\ln(1+x^2+y^2)}{\sqrt{x^2+y^2}}.$$

Положим $\rho = \sqrt{x^2+y^2}$, тогда $\rho \to 0$ при $x \to 0$ и $y \to 0$. Подстановка в исходный предел дает

$$\lim_{\substack{x\to 0\\ y\to 0}} \frac{\ln(1+x^2+y^2)}{\sqrt{x^2+y^2}} = \lim_{\rho\to 0} \frac{\ln(1+\rho^2)}{\rho} = \lim_{\rho\to 0} \frac{\rho^2}{\rho} = \lim_{\rho\to 0} \rho = 0.$$

Здесь мы использовали эквивалентность $\ln(1+\rho^2) \sim \rho^2$ при $\rho \to 0$.

Пример 2. Найти

$$\lim_{\substack{x\to 0\\ y\to 0}} \frac{x+y}{x-y}.$$

Функция под знаком предела определена всюду, кроме точек прямой $x - y = 0$. Покажем, что она не имеет предела в точке $O(0,0)$.

Возьмем две сходящиеся к этой точке последовательности точек M: $\{M_n(1/n, 0)\}$ и $\{M'_n(0, 1/n)\}$. Пределы соответствующих последовательностей данной функции

$$\lim_{n\to\infty} f(M_n) = \lim_{n\to\infty} \frac{1/n+0}{1/n-0} = 1, \quad \lim_{n\to\infty} f(M'_n) = \lim_{n\to\infty} \frac{0+1/n}{0-1/n} = -1.$$

Двум разным последовательностям точек из множества $\{M\}$ определения данной функции, сходящимся к началу координат, соответствуют две последовательности ее значений,

имеющие разные пределы в этой точке, что противоречит определению 1 предела функции нескольких переменных.

Пример 3. Найти предел функции $f(x, y) = \dfrac{x^2 y}{x^4 + y^2}$ в точке $O(0,0)$.

Данная функция определена всюду на координатной плоскости Oxy, кроме точки $O(0,0)$.

Рассмотрим сначала предел этой функции при стремлении точки M к началу координат по любой прямой, проходящей через точку O, т.е. вдоль линии $y = kx (k \neq 0)$:

$$\lim_{M\to 0} \frac{x^2 y}{x^4 + y^2} = \lim_{x\to 0} \frac{kx^3}{x^4 + k^2 x^2} = \lim_{x\to 0} \frac{kx}{x^2 + k^2} = 0$$

Теперь рассмотрим предел данной функции при стремлении точки M к точке O вдоль параболы $y = ax^2$ $(a \neq 0)$:

$$\lim_{M\to 0} \frac{x^2 y}{x^4 + y^2} = \lim_{x\to 0} \frac{ax^4}{x^4 + a^2 x^4} = \frac{a}{1+a^2}.$$

Здесь, как и в предыдущем случае, не соблюдено определение 1 предела функции нескольких переменных, следовательно, предела данной функции в начале координат $O(0,0)$ не существует.

Рассмотренные примеры позволяют говорить в случае функции нескольких переменных о пределах функции в точке *вдоль определенных линий*.

Бесконечно малые функции

Определение 4. Функция $u = f(M)$ называется *бесконечно малой* в точке A (при $M \to A$), если предел функции в этой точке равен нулю:

$$\lim_{M\to A} f(M) = 0.$$

Все, что ранее было сказано о бесконечно малых функциях (см. параграф 3.5), относится и к случаю функции нескольких переменных. Здесь мы отметим лишь два основных момента.

Если функция $u = f(M)$ имеет в точке A предел, равный b, то функция $\alpha(M) = f(M) - A$ является бесконечно малой

в точке A. Отсюда получаем специальное представление для функции, имеющей в точке A предел b:

$$f(M) = b + \alpha(M), \quad \text{где} \quad \lim_{M \to A} \alpha(M) = 0.$$

Сравнение бесконечно малых функций нескольких переменных производится точно так же, как и для бесконечно малых функций одной переменной. Под символом $o(\alpha)$ будем понимать любую бесконечно малую в данной точке A функцию более высокого порядка малости, чем $\alpha(M)$, т.е.

$$\lim_{M \to A} \frac{o(\alpha(M))}{\alpha(M)} = 0.$$

9.4. Непрерывные функции нескольких переменных

Понятие непрерывности функции нескольких переменных

Пусть функция $u = f(M)$ определена на множестве $\{M\}$. Возьмем точку $A \in \{M\}$, любая ε-окрестность которой содержит точки множества $\{M\}$.

Определение 1. Функция $u = f(M)$ называется *непрерывной в точке A*, если предел функции в этой точке существует и равен значению функции в этой точке:

$$\lim_{M \to A} f(M) = f(A).$$

Данное определение можно сформулировать как на "языке последовательностей", так и на "языке ε–δ" в соответствии с определениями 1 и 2 предела функции нескольких переменных, данными в параграфе 9.3; следует лишь заменить в упомянутых формулировках число b на значение функции в точке A: $f(A)$.

Отсюда, в частности, следует свойство устойчивости знака функции, непрерывной в точке M евклидова пространства E^m.

Точки, в которых функция $f(M)$ не обладает свойством непрерывности, называются *точками разрыва* этой функции.

Определение 2. Функция $u = f(M)$, называется *непрерывной на множестве* $\{M\}$, если она непрерывна в любой точке этого множества.

Введем понятие полного приращения функции нескольких переменных. *Полным приращением* функции $u = f(M)$ в точке $A\,(a_1, a_2, \ldots, a_m)$ назовем функцию Δu, определяемую формулой

$$\Delta u = f(M) - f(A),$$

где $M(x_1, x_2, \ldots, x_m)$ — любая точка множества $\{M\}$. Введем приращения переменных

$$\Delta x_1 = x_1 - a_1, \ \Delta x_2 = x_2 - a_2, \ \ldots, \ \Delta x_m = x_m - a_m,$$

тогда приращение функции Δu можно переписать в виде

$$\Delta u = f(a_1 + \Delta x_1, a_2 + \Delta x_2, \ldots, a_m + \Delta x_m) - f(a_1, a_2, \ldots, a_m). \tag{9.14}$$

Теперь определение непрерывности функции нескольких переменных в точке можно сформулировать на "языке приращений".

Определение 3. Функция $u = f(M)$ называется *непрерывной в точке A*, если ее полное приращение в этой точке является бесконечно малой при $M \to A$:

$$\lim_{M \to A} \Delta u = \lim_{M \to A} [f(M) - f(A)] = 0, \quad \text{или} \quad \lim_{\substack{x_1 \to a_1 \\ x_2 \to a_2 \\ \ldots\ldots \\ x_m \to a_m}} \Delta u = 0. \tag{9.15}$$

Геометрический смысл непрерывности функции многих переменных особенно очевиден при $m = 2$: график функции $z = f(x, y)$ представляет собой в точке непрерывности (x_0, y_0) сплошную поверхность в некоторой окрестности этой точки.

Рассмотрим примеры на исследование непрерывности функций нескольких переменных.

Пример 1. $z = f(x, y) = \dfrac{xy + 3}{x^2 + y^2}$.

Эта функция непрерывна на всей координатной плоскости Oxy, исключая точку $O(0, 0)$.

Пример 2.

$$z = f(M) = \begin{cases} \sin(x^2 + y^2)/(x^2 + y^2) & \text{если } x^2 + y^2 \neq 0, \\ 1, & \text{если } x^2 + y^2 = 0. \end{cases}$$

Введем расстояние от начала координат $\rho^2 = x^2 + y^2$, тогда данную функцию можно переписать в виде

$$f(M) = \begin{cases} \sin \rho^2/\rho^2, & \rho \neq 0, \\ 1, & \rho = 0. \end{cases}$$

Отсюда видно, что эта функция непрерывна на всем евклидовом пространстве E^2, поскольку функция $\sin \rho^2/\rho^2$ в точке устранимого разрыва $\rho = 0$ доопределена ее предельным значением.

Пример 3. $u = f(M) = [\ln(1 - x_1^2 - x_2^2 - \cdots - x_m^2)]^{-1}$.

Поскольку формула, выражающая данную функциональную зависимость, имеет смысл при условии $1 - x_1^2 - x_2^2 - \cdots - x_m^2 > 0$, то областью определения этой функции является m-мерный незамкнутый шар евклидова пространства E^m радиуса $R = 1$, не включающий центр — точку O. На этом множестве функция логарифма непрерывна и не обращается в нуль, и, следовательно, данная функция также непрерывна.

Пример 4. $u = \ln|1 - x_1^2 - x_2^2 - x_3^2 - \cdots - x_m^2|$.

Очевидно, что данная функция разрывна на гиперповерхности, которой является m-мерная сфера с центром в точке O радиуса $R = 1$ евклидова пространства E^m.

Непрерывность функции по каждой переменной

Для функции многих переменных можно определить понятие непрерывности по одной переменной при фиксированных значениях остальных. Придадим первому аргументу x_1 приращение Δx_1, при фиксированных значениях других переменных так, чтобы точка $M_1(x_1 + \Delta x_1, x_2, \ldots, x_m)$ принадлежала множеству $\{M\}$ определения функции $f(M)$. Соответствующее приращение функции называется *частным приращением функции* в точке $M(x_1, x_2, \ldots, x_m)$:

$$\Delta_{x_1} u = f(M_1) - f(M).$$

Аналогично можно определить частные приращения функции $f(M)$ в точке M, соответствующие приращению Δx_k любого из m аргументов x_k. Пусть точка $M_k(x_1, x_2, \ldots, x_{k-1}, x_k + \Delta x_k, x_{k+1}, \ldots, x_m)$ находится в области определения $\{M\}$ функции $u = f(M)$. Тогда частное приращение этой функции по аргументу x_k определяется формулой

$$\Delta_{x_k} u = f(M_k) - f(M), \quad k = 1, 2, \ldots, m. \tag{9.16}$$

Определение 4. Функция $f(M)$ называется *непрерывной в точке* M евклидова пространства E^m по переменной x_k, если соответствующее частное приращение (9.16) этой функции в точке M является бесконечно малой при $\Delta x_k \to 0$:

$$\lim_{\Delta x_k \to 0} \Delta_{x_k} u = 0. \tag{9.17}$$

Отметим, что при фиксированных значениях всех переменных, кроме x_k, функция $u = f(M)$ является функцией одной этой переменной.

Из непрерывности функции $f(M)$ в точке M евклидова пространства E^m очевидным образом следует ее непрерывность по каждой из переменных $x_1, x_2, \ldots, x_m$ (покоординатная непрерывность). Однако из покоординатной непрерывности функции $f(M)$ в точке M не следует, вообще говоря, непрерывности данной функции в этой точке.

Рассмотрим соответствующие примеры.

Пример 5. $u = \begin{cases} \dfrac{xy}{x^2 + y^2}, & x^2 + y^2 \neq 0, \\ 0, & x^2 + y^2 = 0. \end{cases}$

Нетрудно видеть, что данная функция непрерывна в точке $O\,(0, 0)$ по каждой из переменных x и y (на каждой из координатных осей). Рассмотрим теперь ее непрерывность на других прямых, проходящих через точку O, вида $y = kx$, $k \neq 0$. На таких прямых функция u постоянна и равна $k/(1 + k^2)$, что не совпадает со значением функции в точке O.

Пример 6. $u = \begin{cases} \dfrac{x^2 y}{x^4 + y^2}, & x^4 + y^2 \neq 0, \\ 0, & x^4 + y^4 = 0. \end{cases}$

Как мы уже убедились (см. пример 3 параграфа 9.3), на каждой прямой $y = kx$ данная функция имеет вид $u = kx/(x^2 + k^2)$, т.е. она непрерывна на координатных осях (по каждой из переменных x и y). Однако значения функции на параболе вида $y = ax^2$ постоянны и равны $a/(1 + a^2)$, что не совпадает с ее частным значением в точке O.

Значит, данная функция, будучи непрерывной в начале координат по каждой из переменных, не является непрерывной в этой точке.

Свойства непрерывных функций

Приведем без доказательства основные свойства непрерывных функций нескольких переменных, поскольку они аналогичны основным свойствам непрерывных функций одной переменной.

1°. *Если функция $u = f(M)$ определена и непрерывна на замкнутом ограниченном множестве $\{M\}$ точек евклидова пространства E^m, то она ограничена на этом множестве, т.е. существует такое число $C > 0$, что для всех точек множества $\{M\}$ выполняется неравенство*

$$|f(M)| < C.$$

2°. *Если функция $u = f(M)$ непрерывна на замкнутом ограниченном множестве $\{M\}$, то она достигает на этом множестве своих точных верхней и нижней граней.*

3°. *Если функция $u = f(M)$ непрерывна на замкнутом ограниченном множестве $\{M\}$, то она принимает все промежуточные значения между своими минимальным и максимальным значениями.*

4°. *Если функция $u = f(M)$ непрерывна на замкнутом ограниченном множестве $\{M\}$, то она равномерно-непрерывна на этом множестве, т.е. для любого числа $\varepsilon > 0$ существует такое число $\delta > 0$, что для любых двух точек M' и M'' множества $\{M\}$, удовлетворяющих условию $\rho(M', M'') < \delta$ (находящихся внутри m-мерной сферы радиуса δ), выполняется неравенство $|f(M'') - f(M')| < \varepsilon$.*

5°. *Пусть $f(M)$ и $g(M)$ — непрерывные функции на замкнутом ограниченном множестве точек $\{M\}$ евклидова пространства E^m. Тогда функции $f(M) \pm g(M)$, $f(M)g(M)$ и $f(M)/g(M)$ (частное при $g(M) \neq 0$) также являются непрерывными функциями на множестве $\{M\}$.*

Непрерывность сложной функции нескольких переменных

Введем понятие сложной функции нескольких переменных. Пусть функции

$$x_i = \varphi_i(T) = \varphi_i(t_1, t_2 < \ldots, t_k), \quad i = 1, 2, \ldots, m, \qquad (9.18)$$

заданы на множестве $\{T\}$ точек $T(t_1, t_2, \ldots, t_k)$ евклидова пространства E^k. Тогда каждой точке T из множества $\{T\}$ посредством формул (9.18) ставится в соответствие точка $M\ (x_1, x_2, \ldots, x_m)$ евклидова пространства E^m; обозначим через $\{M\}$ множество таких точек. Пусть на этом множестве $\{M\}$ задана функция $u = f(M)$. Будем говорить, что на множестве $\{T\}$ евклидова пространства E^k определена *сложная функция* $u = f(x_1, x_2, \ldots, x_m)$, где переменные $x_1, x_2, \ldots, x_m$ являются функциями переменных $t_1, t_2, \ldots, t_k$, определяемыми соотношениями (9.18).

Справедлива следующая теорема о непрерывности сложной функции нескольких переменных.

Теорема 9.3. *Пусть функции (9.18) непрерывны в точке $A\ (a_1, a_2, \ldots, a_k)$, а функция $u = f(M)$ непрерывна в точке $B\ (b_1, b_2, \ldots, b_m)$, где $b_i = \varphi_i(a_1, a_2, \ldots, a_k)$, $i = 1, 2, \ldots, m$. Тогда сложная функция $u = f[\varphi_1(T), \varphi_2(T), \ldots, \varphi_m(T)]$ непрерывна в точке $A\ (a_1, a_2, \ldots, a_k)$.*

Доказательство этой теоремы в известной мере аналогично доказательству теоремы о непрерывности сложной функции одной переменной (см. теорему 3.14) и проводится также на "языке последовательностей".

Линии уровня

Понятие линии уровня широко используется прежде всего в геодезии, картографии, составлении синоптических карт, а также в описании различных физических полей (температура, давление и проч.).

Определение 5. *Линией уровня* функции двух переменных $z = f(x, y)$ называется плоская кривая, получаемая при пересечении графика этой функции плоскостью, параллельной координатной плоскости Oxy, $z = C$, где C — постоянная величина.

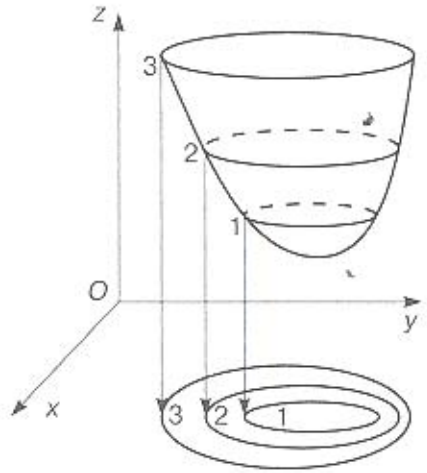

Рис. 9.4

Обычно линии уровня, соответствующие различным значениям постоянной величины C, проецируются на одну плоскость, например на координатную плоскость Oxy; тогда их удобно анализировать и с их помощью исследовать сложный характер поверхности, описываемой функцией $z = f(x, y)$.

Таким образом, можно сказать, что линии уровня функции $z = f(x, y)$ — это семейство кривых на координатной плоскости Oxy, описываемое уравнениями вида

$$f(x, y) = C.$$

Обычно берут арифметическую прогрессию чисел C_i с постоянной разностью h; тогда по взаимному расположению линий уровня можно получить представление о форме поверхности, описываемой функцией $z = f(x, y)$. Там, где функция изменяется быстрее, линии уровня сгущаются, а там, где поверхность пологая, линии уровня располагаются реже (рис. 9.4).

Пример 7. Найти линии уровня функции $z = x^2 + y^2 - 2x - 2y$.

Линии уровня данной функции - это семейство кривых на плоскости Oxy, описываемое уравнением

$$x^2 + y^2 - 2x - 2y = C, \text{ или } (x-1)^2 + (y-1)^2 = 2 + C.$$

Последнее уравнение описывает семейство окружностей с центром в точке O_1, $(1, 1)$ радиуса $r = \sqrt{2 + C}$. На рис. 9.5

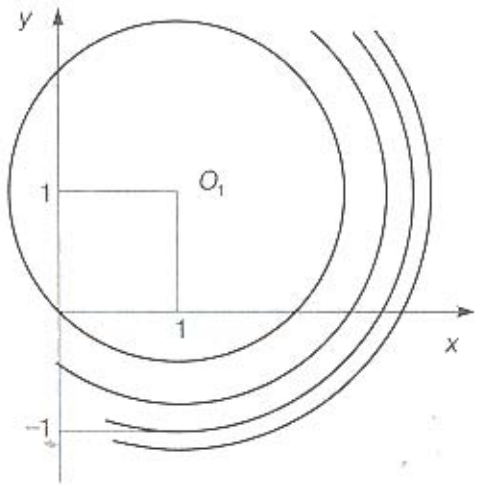

Рис. 9.5

показаны эти линии уровня для равноотстоящих значений C, начиная с нуля. Видно, что поверхность вращения, описываемая данной функцией, становится "круче" по мере ее удаления от оси, которая задается уравнениями $x = 1$, $y = 1$.

9.5. Элементы аналитической геометрии в пространстве

В этом параграфе мы изложим основные сведения аналитической геометрии в пространстве: векторы и операции над ними, плоскости и прямые, поверхности второго порядка.

Скаляры и векторы

Многие характеристики окружающих нас явлений описываются числами, например вес товара и его стоимость, температура тела, количество мест в салоне авиалайнера и др. Такие величины называются *скалярными*, или просто *скалярами*. Однако имеются и такие величины, которые требуют для своего описания еще и указания направления, например скорость и ускорение при движении тела. Такие величины называются *векторными*, или просто *векторами*.

Определение 1. Направленный отрезок, на котором заданы начало, конец и направление, называется *вектором* (рис. 9.6).

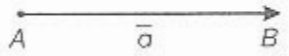

Рис. 9.6

Обозначается вектор либо символом, куда входят его начало и конец $\overline{AB}$, либо одной буквой с чертой наверху $(\bar{a})$. Расстояние между началом и концом вектора называется его *длиной*.

Векторы $\bar{a}$ и $\bar{b}$; $\bar{a}$ и $\bar{c}$ называются *коллинеарными*, если они лежат соответственно на одной прямой или на параллельных прямых (рис. 9.7).

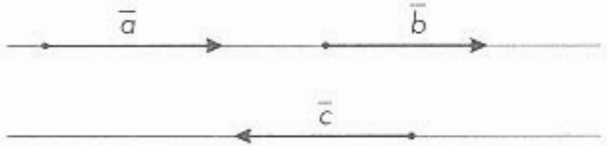

Рис. 9.7

Определение 2. Векторы $\bar{a}$ и $\bar{b}$ называются *равными*, если они коллинеарны, одинаково направлены и их длины равны.

В любой системе координат вектор характеризуется своими координатами: $\overline{AB} = (X, Y, Z)$. Пусть в системе координат $Oxyz$ координаты начала и конца вектора соответственно $A(x_1, y_1, z_1)$ и $B(x_2, y_2, z_2)$ (рис. 9.8). Тогда координаты этого вектора определяются формулами

$$X = x_2 - x_1,\ Y = y_2 - y_1,\ Z = z_2 - z_1. \qquad (9.19)$$

Длина вектора $\overline{AB}$ очевидным образом определяется по формуле

$$|\overline{AB}| = \sqrt{X^2 + Y^2 + Z^2}.$$

Важным моментом является определение нулевого вектора $\bar{0}(0, 0, 0)$ нулевой длины.

Операции над векторами

Над векторами определены две линейные операции. Пусть даны два вектора: $\bar{a} = (a_1, a_2, a_3)$ и $\bar{b} = (b_1, b_2, b_3)$.

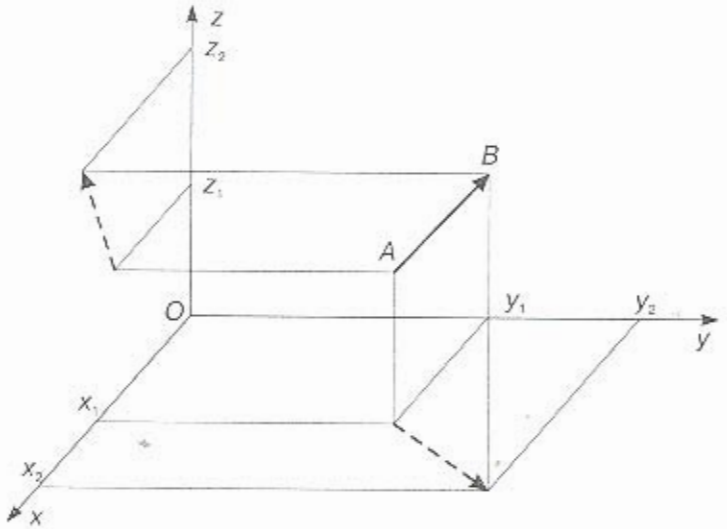

Рис. 9.8

Определение 3. *Суммой векторов* $\bar{a} + \bar{b}$ называется вектор $\bar{c} = (c_1, c_2, c_3)$, координаты которого равны суммам соответствующих координат $\bar{a}$ и $\bar{b}$:

$$c_1 = a_1 + b_1,\ c_2 = a_2 + b_2,\ c_3 = a_3 + b_3. \qquad (9.20)$$

Определение 4. *Произведением вектора* $\bar{a} \neq \bar{0}$ *на число* $\alpha \neq 0$ называется вектор $\alpha\bar{a}$, координаты которого соответственно равны αa_1, αa_2, αa_3.

Можно показать, что для получения суммы векторов нужно совместить конец вектора $\bar{a}$ с началом вектора $\bar{b}$, тогда вектор $\bar{a} + \bar{b}$ будет направлен от начала первого вектора к концу второго (рис. 9.9, а). Геометрический смысл умножения вектора на число состоит в увеличении его длины в α раз при $|\alpha| > 1$ или ее сокращении в α раз при $|\alpha| < 1$ (рис. 9.9, б). При $\alpha = -1$ вектор $\alpha\bar{a}$ имеет направление, противоположное вектору $\bar{a}$; векторы $\alpha\bar{a}$ и $\bar{a}$ коллинеарны.

Теперь приведем основные свойства линейных; операций. Пусть $\bar{a}$, $\bar{b}$ и $\bar{c}$ — любые векторы, а α и μ — произвольные числа.

1°. $\bar{a} + \bar{b} = \bar{b} + \bar{a}$ — переместительное свойство.

2°. $(\bar{a} + \bar{b}) + \bar{c} = \bar{a} + (\bar{b} + \bar{c})$ — сочетательное свойство сложения.

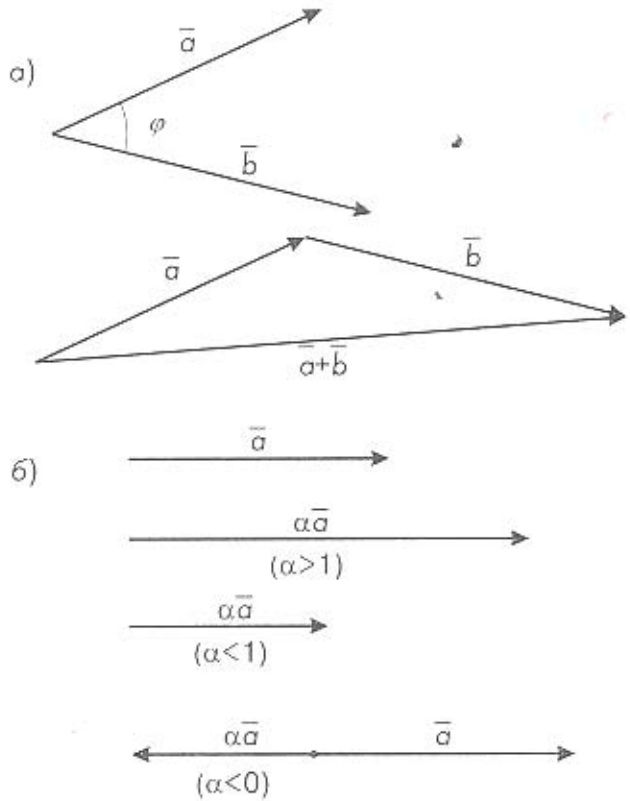

Рис. 9.9

3°. $\alpha(\mu\bar{a}) = (\alpha\mu)\bar{a}$ — сочетательное свойство умножения.

4°. $(\alpha+\mu)\bar{a} = \alpha\bar{a} + \mu\bar{a}$ — распределительное свойство относительно суммы чисел.

5°. $\alpha(\bar{a}+\bar{b}) - \alpha\bar{a} + \alpha\bar{b}$ — распределительное свойство относительно суммы векторов.

Пусть даны два вектора: $\bar{a} = (a_1, a_2, a_3)$ и $\bar{b} = (b_1, b_2, b_3)$. Из определения коллинеарности векторов и определения произведения вектора на число вытекает, что векторы $\bar{a}$ и $\bar{b}$ коллинеарны в том и только в том случае, если их координаты пропорциональны:

$$\frac{a_1}{b_1} = \frac{a_2}{b_2} = \frac{a_3}{b_3}. \tag{9.21}$$

Соотношения (9.21) носят название *условия коллинеарности двух векторов*.

Скалярное произведение векторов

Пусть $\bar{a}$ и $\bar{b}$ — произвольные векторы, а φ — угол между ними, как показано на рис. 9.9, а.

Определение 5. *Скалярным произведением двух векторов* $\bar{a}$ *и* $\bar{b}$ называется число, равное

$$\bar{a}\bar{b} = |\bar{a}||\bar{b}|\cos\varphi. \tag{9.22}$$

Скалярное произведение векторов обладает следующими свойствами, которые нетрудно вывести из его определения.

1°. $\bar{a}\bar{b} = \bar{b}\bar{a}$ — переместительность сомножителей.

2°. $(\alpha\bar{a})\bar{b} = \alpha(\bar{a}\bar{b})$ — сочетательность относительно умножения на число.

3°. $\bar{a}(\bar{b}+\bar{c}) = \bar{a}\bar{b} + \bar{a}\bar{c}$ — распределительность относительно суммы векторов.

4°. $\bar{a}\bar{a} = |\bar{a}|^2$ — формула скалярного квадрата.

5°. $\bar{a}\bar{b} = 0$, *если вектор* $\bar{a}$ *перпендикулярен вектору* $\bar{b}$, *и наоборот, если* $\bar{a}\bar{b} = 0$, *то при* $\bar{a} \neq \bar{0}$ *и* $\bar{b} \neq \bar{0}$ *векторы* $\bar{a}$ *и* $\bar{b}$ *взаимно перпендикулярны.*

Скалярное произведение векторов выражается через их координаты следующим образом: пусть даны векторы $\bar{a}(a_1, a_2, a_3)$ и $\bar{b}(b_1, b_2, b_3)$, тогда

$$\bar{a}\bar{b} = a_1b_1 + a_2b_2 + a_3b_3. \tag{9.23}$$

Из формул (9.22) и (9.23) получаем формулу для определения угла между векторами:

$$\cos\varphi = \frac{\bar{a}\bar{b}}{|\bar{a}||\bar{b}|} = \frac{a_1b_1 + a_2b_2 + a_3b_3}{|\bar{a}||\bar{b}|}. \tag{9.24}$$

Уравнения поверхности и линии

Пусть в прямоугольной системе координат $Oxyz$ координаты x, y и z связаны некоторым уравнением

$$F(x, y, z) = 0. \tag{9.25}$$

Будем говорить, что соотношение (9.25) является уравнением поверхности S в заданной системе координат, если ему удовлетворяют координаты любой точки $M(x, y, z)$, принадлежащей S, и не удовлетворяют координаты никакой точки, не лежащей на этой поверхности.

Линию в пространстве можно рассматривать как пересечение двух поверхностей, или множество точек, лежащих на двух поверхностях одновременно. Поэтому систему двух уравнений вида (9.25)

$$\begin{cases} F(x, y, z) = 0, \\ \Phi(x, y, z) = 0 \end{cases} \tag{9.26}$$

называют *уравнением линии L в пространстве*, если этим уравнениям удовлетворяют только координаты точек, лежащих на линии L.

Плоскость в пространстве

Уравнение вида

$$Ax + By + Cz + d = 0 \tag{9.27}$$

называется *общим уравнением плоскости* в системе координат $Oxyz$. Вектор $\overline{N} = (A, B, C)$ перпендикуляреп плоскости (9.27); он называется *нормальным вектором* этой плоскости. Если известно, что плоскость проходит через некоторую точку $M_0(x_0, y_0, z_0)$, то она может быть задана уравнением

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0. \tag{9.28}$$

Например, составить уравнение плоскости с перпендикулярным вектором $\overline{N} = (1, 2, -1)$, проходящей через точку $M_0(2, 1, 1)$.

Согласно формуле (9.28),

$$1(x - 2) + 2(y - 1) - 1(z - 1) = 0, \quad \text{или} \quad x + 2y - z - 3 = 0.$$

Теперь рассмотрим вопрос о взаимном расположении плоскостей в пространстве. Пусть две плоскости заданы уравнениями

$$\begin{aligned} A_1x + B_1y + C_1z + D_1 = 0 \text{ и} \\ A_2x + B_2y + C_2z + D_2 = 0. \end{aligned} \tag{9.29}$$

Угол между плоскостями в пространстве. При любом положении двух плоскостей угол между ними равен углу φ между их нормальными векторами $\overline{N_1} = (A_1, B_1, C_1)$ и $\overline{N_2} = (A_2, B_2, C_2)$. Следовательно, по формуле (9.24) получаем, что угол φ между плоскостями (9.29) определяется из формулы

$$\cos\varphi = \frac{\bar{N}_1\bar{N}_2}{|\bar{N}_1||\bar{N}_2|} = \frac{A_1A_2 + B_1B_2 + C_1C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2}\sqrt{A_2^2 + B_2^2 + C_2^2}}. \tag{9.30}$$

Второй (смежный) угол равен $\pi - \varphi$.

Условие параллельности плоскостей. Если плоскости параллельны, то их нормальные векторы коллинеарны, и наоборот. В этом случае из формулы (9.21) имеем условие параллельности плоскостей, заданных уравнениями (9.29):

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2}. \tag{9.31}$$

Условие перпендикулярности плоскостей. В случае если плоскости (9.29) перпендикулярны, то перпендикулярны между собой и их нормальные векторы. Так как $\cos\varphi = 0$, то из формулы (9.30) получаем условие перпендикулярности двух плоскостей

$$A_1A_2 + B_1B_2 + C_1C_2 = 0. \tag{9.32}$$

Расстояние от точки до плоскости. Пусть плоскость задана уравнением (9.27). Тогда расстояние от точки $M_0(x_0, y_0, z_0)$ до этой плоскости определяется по формуле

$$d = \frac{|Ax_0 + By_0 + Cz_0 + D|}{|\bar{N}|} = \frac{|Ax_0 + By_0 + Cz_0 + D|}{\sqrt{A^2 + B^2 + C^2}}. \tag{9.33}$$

Прямая в пространстве

По аналогии с линией в пространстве прямую в пространстве можно рассматривать как линию пересечения двух плоскостей и задавать ее двумя уравнениями первой степени вида (9.29). Эти два уравнения определяют прямую в пространстве, только если соответствующие плоскости не параллельны или не совпадают друг с другом, т.е. их нормальные векторы

не коллинеарны. Иными словами, система линейных уравнений (9.29) определяет прямую, если условия (9.31) не выполняются; тогда эти уравнения называются *общими уравнениями прямой*.

Уравнения прямой в пространстве. При решении задач чаще употребляется другой вид уравнении прямой. Пусть заданы прямая, проходящая через точки $M(x, y, z)$ и $M_0(x_0, y_0, z_0)$, и ненулевой вектор $\bar{a} = (l, m, n)$, параллельный данной прямой или лежащий на ней (рис. 9.10). Этот вектор

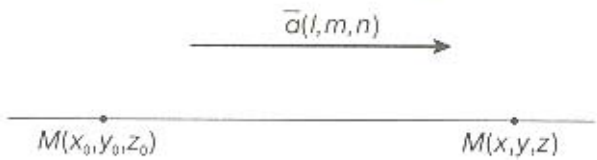

Рис. 9.10

называется *направляющим вектором прямой*. Поскольку векторы $\overline{M_0M}$ и $\bar{a}$ коллинеарны, то из условия коллинеарности двух векторов (9.21) получаем искомые уравнения прямой

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} = \frac{z - z_0}{n}. \tag{9.34}$$

Уравнения (9.34) называются *каноническими уравнениями прямой в пространстве*.

В случае когда прямая задана уравнениями общего вида (9.29), переход к каноническим уравнениям осуществляется путем вычисления координат направляющего вектора $\bar{a}$ из коэффициентов уравнений (9.29) как соответствующих определителей второго порядка:

$$\bar{a} = (l, m, n) = \left\{ \begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix}, \begin{vmatrix} C_1 & A_1 \\ C_2 & A_2 \end{vmatrix}, \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} \right\}. \tag{9.35}$$

Пример. Найти канонические уравнения прямой, заданной уравнениями общего вида

$$\begin{cases} x + 2y + 3z - 8 = 0, \\ 2x + y - 2z - 2 = 0. \end{cases}$$

Сначала определим какую-либо точку M_0, через которую проходит данная прямая. Положим $x_0 = 1$, тогда из данных уравнений получаем систему для определения y_0 и z_0:

$$\begin{cases} 2y_0 + 3z_0 = 7 \\ y_0 - 2z_0 = 0, \end{cases}$$

откуда $y_0 = 2$, $z_0 = 1$. Координаты направляющего вектора находим по формуле (9.35):

$$l = \begin{vmatrix} 2 & 3 \\ 1 & -2 \end{vmatrix} = -7, \quad m = \begin{vmatrix} 3 & 1 \\ -2 & 2 \end{vmatrix} = 8, \quad n = \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = -3.$$

По формулам (9.34) находим канонические уравнения прямой:

$$\frac{x - 1}{-7} = \frac{y - 2}{8} = \frac{z - 1}{-3}.$$

Угол между прямыми. Угол между прямыми в пространстве равен углу φ между их направляющими векторами: $\bar{a}_1 = (l_1, m_1, n_1)$ и $\bar{a}_2 = (l_2, m_2, n_2)$, т.е. вычисляется по формуле (9.30):

$$\cos\varphi = \frac{l_1 l_2 + m_1 m_2 + n_1 n_2}{\sqrt{l_1^2 + m_1^2 + n_1^2}\sqrt{l_2^2 + m_2^2 + n_2^2}}. \tag{9.36}$$

Условие параллельности прямых. Две прямые параллельны тогда и только тогда, когда их направляющие векторы коллинеарны. Отсюда из (9.21) следует:

$$\frac{l_1}{l_2} = \frac{m_1}{m_2} = \frac{n_1}{n_2}. \tag{9.37}$$

Условие перпендикулярности прямых. Прямые перпендикулярны тогда и только тогда, когда их направляющие векторы перпендикулярны. Тогда из (9.36) следует искомое условие:

$$l_1 l_2 + m_1 m_2 + n_1 n_2 = 0. \tag{9.38}$$

Расстояние от точки до прямой. Пусть дана прямая, проходящая через точку $M_0(x_0, y_0, z_0)$ и заданная каноническим уравнением (9.34), а также точка $M_1(x_1, y_1, z_1)$ находящаяся вне этой прямой. Тогда расстояние d от точки M_1 до заданной прямой вычисляется по формуле

$$d = \frac{\sqrt{\begin{vmatrix} y_1-y_0 & z_1-z_0 \\ m & n \end{vmatrix}^2 + \begin{vmatrix} z_1-z_0 & x_1-x_0 \\ n & l \end{vmatrix}^2 + \begin{vmatrix} x_1-x_0 & y_1-y_0 \\ l & m \end{vmatrix}^2}}{\sqrt{l^2+m^2+n^2}}. \tag{9.39}$$

Взаимное расположение прямой и плоскости

Пусть прямая и плоскость в пространстве заданы соответственно уравнениями

$$\frac{x-x_0}{l} = \frac{y-y_0}{m} = \frac{z-z_0}{n},$$
$$Ax + By + Cz + D = 0.$$

Параллельность прямой и плоскости. Прямая параллельна плоскости только в том случае, если ее направляющий вектор перпендикулярен нормальному вектору плоскости. С учетом (9.24) имеем условие параллельности прямой и плоскости:

$$Al + Bm + Cn = 0. \tag{9.40}$$

Перпендикулярность прямой и плоскости. Очевидно, что прямая перпендикулярна плоскости лишь в том случае, когда ее направляющий вектор коллинеарен нормальному вектору плоскости. Тогда, с учетом (9.21), получаем условие перпендикулярности прямой и плоскости:

$$\frac{A}{l} = \frac{B}{m} = \frac{C}{n}. \tag{9.41}$$

Угол между прямой и плоскостью. Под углом между прямой и плоскостью понимают острый угол α между этой прямой и ее проекцией на плоскость (рис. 9.11). Поскольку этот угол

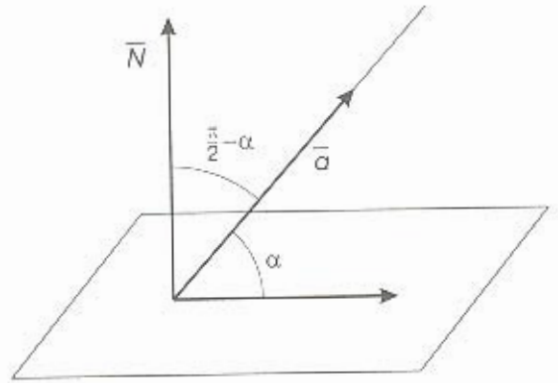

Рис. 9.11

дополняет до $\pi/2$ угол между нормальным вектором плоскости N и направляющим вектором прямой, то, согласно формуле (9.24),

$$\sin\alpha = \cos\left(\frac{\pi}{2} - \alpha\right) = \frac{|Al + Bm + Cn|}{\sqrt{A^2+B^2+C^2}\sqrt{l^2+m^2+n^2}}. \tag{9.42}$$

Поверхности второго порядка

Поверхности, определяемые в прямоугольной системе координат алгебраическими уравнениями второго порядка, называются *поверхностями второго порядка*. Исследование поверхностей второго порядка проводится по их уравнениям с использованием метода параллельных сечений.

Эллипсоид. *Каноническое уравнение эллипсоида:*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{9.43}$$

В сечениях эллипсоида в плоскостях Oxy и Oxz получаются эллипсы с парами полуосей соответственно a, b и a, c (рис. 9.12), уравнения которых имеют вид

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad \frac{x^2}{a^2} + \frac{z^2}{c^2} = 1.$$

Числа a, b и c называются *полуосями эллипсоида*. Нетрудно видеть, что самые большие эллипсы получаются в сечениях координатными плоскостями; по мере удаления от них эллипсы

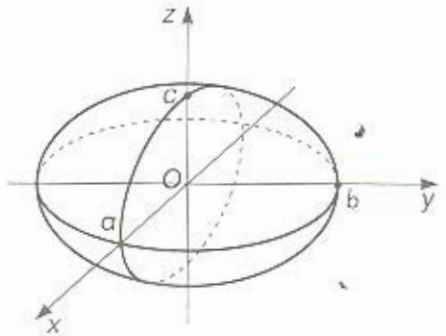

Рис. 9.12

в сечениях плоскостями, параллельными координатным плоскостям, уменьшаются до вырождения в точку и полного исчезновения.

Однополостный гиперболоид. *Каноническое уравнение однополостного гиперболоида:*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1. \tag{9.44}$$

Вид этой поверхности показан на рис. 9.13. В сечениях однополостного гиперболоида координатными плоскостями Oxz и Oyz получаются гиперболы, уравнения которых соответственно имеют вид

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1, \quad \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

В сечениях данной поверхности плоскостями $z = h$, параллельными плоскости Oxy, получаются эллипсы, полуоси которых увеличиваются по мере удаления от плоскости Oxy. Наименьший эллипс лежит в координатной плоскости Oxy; его уравнение имеет вид

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Положительные числа a, b и c называются *полуосями однополостного гиперболоида*. Сама поверхность имеет вид полости бесконечного протяжения по оси Oz, расширяющейся по мере удаления от координатной плоскости Oxy.

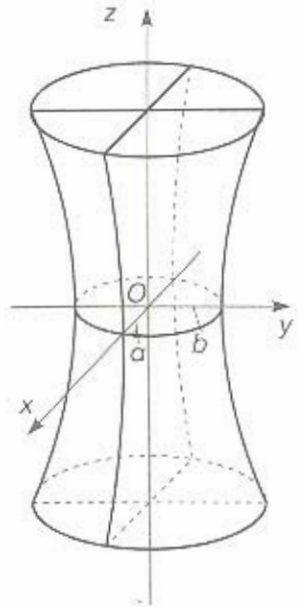

Рис. 9.13

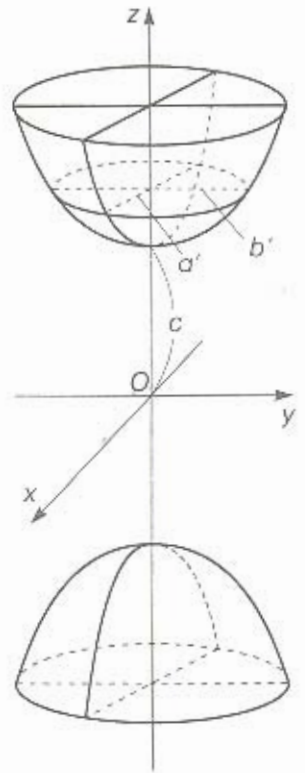

Рис. 9.14

Двуполостный гиперболоид. *Каноническое уравнение двуполостного гиперболоида:*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1. \tag{9.45}$$

Положительные числа a, b и c называются *полуосями двуполостного гиперболоида*.

В сечениях этой поверхности координатными плоскостями Oxz и Oyz получаются гиперболы, уравнения которых соответственно имеют вид

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = -1, \quad \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1.$$

Сечения этого гиперболоида плоскостями $z = h$, параллельными координатной плоскости Oxy, приводят к линиям пересече-

ния, которые определяются уравнениями

$$z = h,\ \frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{h^2}{c^2} - 1.$$

Нетрудно видеть, что эти уравнения имеют смысл при $|h| > c$. Следовательно, в таких сечениях получаются эллипсы с полуосями

$$a' = a\sqrt{h^2/c^2 - 1} \ \text{ и } \ b' = b\sqrt{h^2/c^2 - 1},$$

которые увеличиваются по мере удаления от плоскости Oxy.

Таким образом, двуполостный гиперболоид представляет собой поверхность, состоящую из двух отдельных полостей типа бесконечных выпуклых чаш, которые расположены симметрично относительно плоскости Oxy на расстоянии c от их вершин до этой плоскости (рис. 9.14).

Эллиптический параболоид. *Каноническое уравнение эллиптического параболоида:*

$$\frac{x^2}{p} + \frac{y^2}{q} = 2z, \tag{9.46}$$

где $p > 0$, $q > 0$ — параметры эллиптического параболоида.

Из вида уравнения следует, что данная поверхность второго порядка расположена вдоль положительного направления оси Oz.

Как и в предыдущих случаях, исследуем данную поверхность с помощью сечений. Сечения в координатных плоскостях Oxz и Oyz имеют форму парабол с осью симметрии Oz; их уравнения соответственно имеют вид

$$x^2 = 2pz,\ y^2 = 2qz.$$

Вид этой поверхности показан на рис. 9.15.

Сечения данной поверхности плоскостями $z = h$ приводят к эллипсам

$$z = h,\ \frac{x^2}{p} + \frac{y^2}{q} = 2h,$$

которые увеличиваются по мере удаления от плоскости Oxy, с полуосями $a' = \sqrt{2ph}$ и $b' = \sqrt{2qh}$.

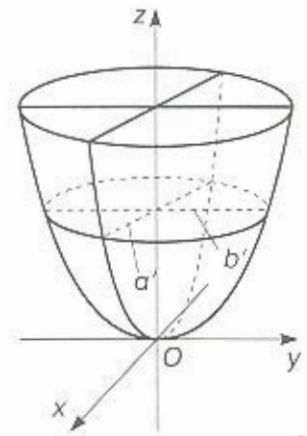

Рис. 9.15

Точка $O(0,0,0)$ называется *вершиной эллиптического параболоида*. При $p = q$ уравнение (9.46) определяет *параболоид вращения*, т.е. поверхность, образованную вращением параболы $x^2 = 2pz$ вокруг оси Oz.

Гиперболический параболоид. *Каноническое уравнение гиперболического параболоида:*

$$\frac{x^2}{p} - \frac{y^2}{q} = 2z, \tag{9.47}$$

где $p > 0$, $q > 0$ — параметры гиперболического параболоида.

В сечении этой поверхности координатной плоскостью Oxz получается парабола

$$x^2 = 2pz,$$

осью симметрии которой является положительная полуось Oz. В сечениях $y = h$, параллельных плоскости Oxz, также получаются направленные ветвями вверх параболы

$$y = h,\ x^2 = 2p\left(z + \frac{h^2}{2q}\right),$$

вершины которых опускаются вниз вдоль оси Oz по мере удаления сечений от плоскости Oxz.

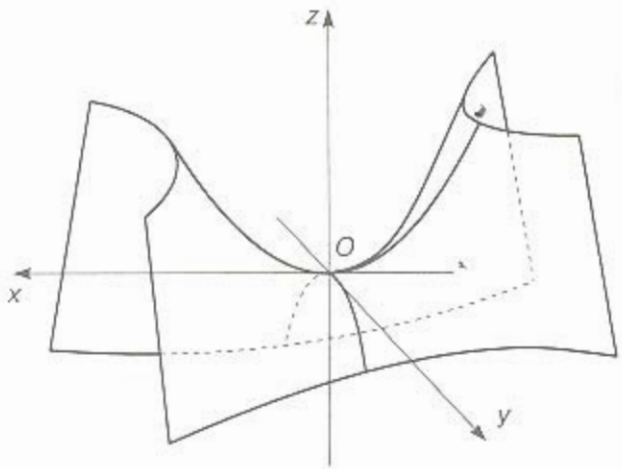

Рис. 9.16

В сечении данной поверхности плоскостью Oyz также получается парабола

$$y^2 = -2qz,$$

однако ее ось симметрии — отрицательная полуось Oz, т.е. ветви этой параболы направлены вниз. В сечениях плоскостями, параллельными плоскости Oyz, также получаются параболы, направленные ветвями вниз, вершины которых опускаются вдоль оси Oz по мере удаления сечений от плоскости Oyz.

Наконец, рассмотрим сечения этого параболоида плоскостями, параллельными плоскости Oxy ($z = h$). Из уравнений

$$z = h,\ \frac{x^2}{p} - \frac{y^2}{q} = 2h$$

следует, что при $h > 0$ в сечениях получаются гиперболы, пересекающие плоскость Oxz; при $h < 0$ (вдоль отрицательного направления оси Oz) в сечениях получаются гиперболы, пересекающие плоскость Oyz. В самой плоскости Oxy (при $h = 0$) гиперболы вырождаются в пару пересекающихся прямых

$$y = \pm\sqrt{\frac{q}{p}}x.$$

Таким образом, гиперболический параболоид (рис. 9.16) представляет собой седлообразную поверхность, вершина которого находится в точке $O(0,0,0)$.

Конус второго порядка. *Каноническое уравнение конуса второго порядка:*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0. \tag{9.48}$$

В сечениях этой поверхности плоскостями $z = h$, параллельными плоскости Oxy, получаем эллипсы

$$z = h,\ \frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{h^2}{c^2},$$

полуоси которых $a' = a|h|/c$ и $b' = b|h|/c$ увеличиваются по мере удаления сечений от плоскости Oxy. При $h = 0$ линия пересечения конуса с плоскостью Oxy вырождается в точку $O(0,0,0)$.

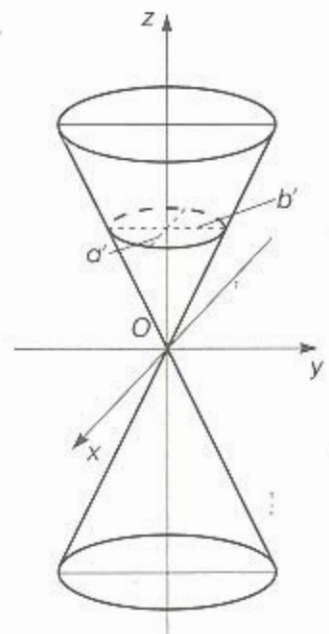

Рис. 9.17

В сечениях данной поверхности плоскостями Oxz и Oyz получаем по паре пересекающихся прямых, соответствующие уравнения которых имеют вид

$$y = 0,\ \frac{x}{a} \pm \frac{z}{c} = 0;\quad x = 0,\ \frac{y}{b} \pm \frac{z}{c} = 0.$$

Поверхность конуса второго порядка показана на рис. 9.17.

Глава 10
ДИФФЕРЕНЦИРОВАНИЕ ФУНКЦИЙ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

10.1. Производные и дифференциалы

Частные производные функции нескольких переменных

Пусть функция $F(M)$ определена в некоторой окрестности точки M евклидова пространства E^m. Рассмотрим в данной точке M отношение частного приращения $\Delta_{x_k} u$ [см. формулу (9.16)] к соответствующему приращению Δx_k аргумента x_k при условии, что точка M_k, полученная изменением координаты x_k точки M, также находится в указанной окрестности точки M:

$$\frac{\Delta_{x_k} u}{\Delta x_k} = \frac{f(x_1, x_2, \ldots, x_{k-1}, x_k + \Delta x_k, \ldots, x_m) - f(x_1, x_2, \ldots, x_m)}{\Delta x_k}. \tag{10.1}$$

Определение 1. Если существует предел отношения частного приращения функции $\Delta_{x_k} u$ в точке $M\ (x_1, x_2, \ldots, x_m)$ к соответствующему приращению аргумента Δx_k при $\Delta x_k \to 0$, то он называется *частной производной* функции $f(M)$ в точке M по аргументу x_k и обозначается одним из символов:

$$\frac{\partial u}{\partial x_k},\ \frac{\partial f}{\partial x_k},\ u'_{x_k},\ f'_{x_k}.$$

Следовательно, согласно данному определению,

$$\frac{\partial u}{\partial x_k} = \lim_{\Delta x_k \to 0} \frac{\Delta_{x_k} u}{\Delta x_k}. \tag{10.2}$$

Таким образом, частная производная функции $u = f(x_1, x_2, \ldots, x_m)$ по аргументу x_k является обыкновенной производной функции одной переменной x_k при фиксированных значениях других переменных. Частная производная функции нескольких переменных характеризует скорость ее изменения по данной координате при фиксированных значениях других координат.

Пример 1. Найти частные производные функции $u = x^2 - 2xy + 2y^2$.

Дифференцируем функцию $u = f(x, y)$ сначала по x, полагая y фиксированной величиной, потом повторяем эту же процедуру, меняя роли x и y. Получаем:

$$\frac{\partial u}{\partial x} = 2x - 2y,\ \frac{\partial u}{\partial y} = 4y - 2x.$$

Пример 2. Найти частные производные функции $u = \operatorname{arctg} xy$.

В результате дифференцирования получаем $\dfrac{\partial u}{\partial x} = \dfrac{y}{1 + (xy)^2}$, $\dfrac{\partial u}{\partial y} = \dfrac{x}{1 + (xy)^2}$.

Пример 3. Найти частные производные функции $u = ye^{yz} + \ln(x^2 - 2y + z)$.

Частные производные этой функции трех переменных имеют следующие формулы:

$$\frac{\partial u}{\partial x} = \frac{2x}{x^2 - 2y + z},\quad \frac{\partial u}{\partial y} = (1 + yz)e^{yz} - \frac{2}{x^2 - 2y + z},$$
$$\frac{\partial u}{\partial z} = y^2 e^{yz} + \frac{1}{x^2 - 2y + z}.$$

Пример 4. Найти частные производные функции $u = \sin(e^{2x} - yz)$.

Дифференцируя по различным аргументам, получаем

$$\frac{\partial u}{\partial x} = 2e^{2x}\cos(e^{2x} - yz),\quad \frac{\partial u}{\partial y} = -z\cos(e^{2x} - yz),$$

$$\frac{\partial u}{\partial z} = -y\cos(e^{2x} - yz).$$

Пример 5. Найти скорости изменения объема продукции Q при изменениях одного из факторов: затрат капитала K или величины трудовых ресурсов L по функции Кобба–Дугласа [см. формулу (9.13)]

$$Q = AK^{\alpha}L^{1-\alpha}.$$

Частные производные этой функции

$$Q'_k = A\alpha K^{\alpha-1}L^{1-\alpha},\ Q'_L = A(1-\alpha)K^{\alpha}L^{-\alpha}$$

дают решение сформулированной выше задачи. Очевидно, что в функции Кобба–Дугласа показатели степеней α и $1-\alpha$ представляют собой соответственно коэффициенты эластичности $E_k(Q)$ и $E_L(Q)$ по каждому из входящих в нее аргументов.

Отметим два весьма важных момента для функций нескольких переменных.

1. Из существования в данной точке всех частных производных функции $f(M)$, вообще говоря, не следует, что функция непрерывна в этой точке. Мы уже знаем, что функция

$$u = \begin{cases} \dfrac{xy}{x^2+y^2}, & x^2+y^2 \neq 0, \\ 0, & x^2+y^2 = 0 \end{cases}$$

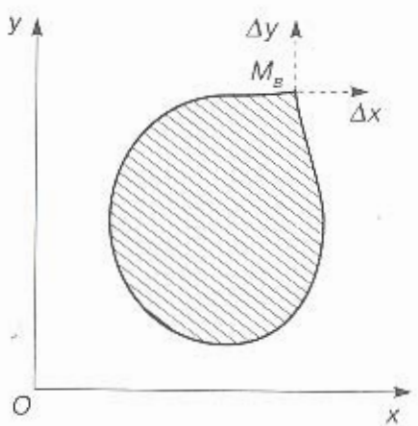

Рис. 10.1

не является непрерывной в точке $O(0,0)$ (см. параграф 9.4, пример 5). Тем не менее в этой точке указанная функция имеет частные производные по x и y, равные нулю.

2. Понятие частных производных определено для внутренних точек множества задания $\{M\}$ функции $f(M)$. Для граничных точек оно в общем случае непригодно, так как для этих точек не всегда можно определить точку M_k, соответствующую приращению аргумента x_k (рис. 10.1). В таком случае по определению полагают, что для граничной точки M_B частные производные определяются как предел:

$$u'_{x_k}(M_B) = \lim_{M \to M_B} u'_{x_k}(M), \tag{10.3}$$

где M — внутренняя точка множества $\{M\}$ задания функции $u = f(M)$ и существования ее частных производных.

Понятие дифференцируемости функции

Как уже говорилось ранее, полным приращением функции $f(M)$ в точке $M\ (x_1, x_2, \ldots, x_m)$, соответствующим приращениям $\Delta x_1, \Delta x_2, \ldots, \Delta x_m$, является разность

$$\Delta u = f(x_1 + \Delta x_1, x_2 + \Delta x_2, \ldots, x_m + \Delta x_m) - f(x_1, x_2, \ldots, x_m). \tag{10.4}$$

Определение 2. Функция $u = f(M)$ называется *дифференцируемой* в точке $M\ (x_1, x_2, \ldots, x_m)$, если ее полное приращение (10.4) в этой точке можно представить в виде

$$\Delta u = A_1\Delta x_1 + A_2\Delta x_2 + \cdots + A_m\Delta x_m + \alpha_1\Delta x_1 + \alpha_2\Delta x_2 + \cdots + \alpha_m\Delta x_m, \tag{10.5}$$

где $A_1, A_2, \ldots, A_m$ — некоторые, не зависящие от $\Delta x_1, \Delta x_2, \ldots, \Delta x_m$ числа, а $\alpha_1, \alpha_2, \ldots, \alpha_m$ — бесконечно малые функции соответственно при $\Delta x_1 \to 0, \Delta x_2 \to 0, \ldots, \Delta x_m \to 0$.

Условие (10.5) называется *условием дифференцируемости* функции в данной точке M евклидова пространства E^m. Если хотя бы одно из чисел $A_1, A_2, \ldots, A_m$ отлично от нуля, то сумма

$$L = A_1\Delta x_1 + A_2\Delta x_2 + \cdots + A_m\Delta x_m \tag{10.6}$$

является *главной линейной частью* полного приращения дифференцируемой функции относительно приращений аргументов.

Теперь приведем необходимые условия дифференцируемости функции нескольких переменных.

Теорема 10.1. *Если функция $u = f(M)$ дифференцируема в точке $M(x_1, x_2, \ldots, x_m)$, то она имеет в этой точке частные производные по всем аргументам, причем $\dfrac{\partial u}{\partial x_i} = A_i$, где A_i определены из условия* (10.5).

Доказательство. Из условия (10.5) следует, что частное приращение функции в точке M

$$\Delta_{x_i} = A_i \Delta x_i + \alpha_i \Delta x_i,$$

откуда

$$\frac{\Delta_{x_i} u}{\Delta x_i} = A_i + \alpha_i.$$

Так как $\alpha_i \to 0$ при $\Delta x_i \to 0$, то отсюда следует, что предел левой части этого равенства равен A_i, что и требовалось доказать.

Рассмотрим функцию $\rho = \sqrt{\Delta x_1^2 + \Delta x_2^2 + \cdots + \Delta x_m^2}$, которая является бесконечно малой при $\Delta x_1 \to 0$, $\Delta x_2 \to 0$, $\ldots$, $\Delta x_m \to 0$. Поскольку $|\Delta x_i|/\rho \leqslant 1$, то

$$\Big|\sum_{i=1}^{m} \alpha_i \Delta x_i\Big| \leqslant \rho \sum_{i=1}^{m} |\alpha_i||\Delta x_i|/\rho \leqslant \rho \sum_{i=1}^{m} |\alpha_i| = o(\rho).$$

Следовательно, условие дифференцируемости функции нескольких переменных в данной точке M, в силу теоремы 10.1, может быть записано в форме:

$$\Delta u = \frac{\partial u}{\partial x_1}\Delta x_1 + \frac{\partial u}{\partial x_2}\Delta x_2 + \cdots + \frac{\partial u}{\partial x_m}\Delta x_m + o(\rho). \qquad (10.7)$$

Далее следует важная теорема о связи между дифференцируемостью функции нескольких переменных и ее непрерывностью.

Теорема 10.2. *Если функция $u = f(M)$ дифференцируема в некоторой точке M, то она непрерывна в этой точке.*

Доказательство. Из условия дифференцируемости (10.5) функции $f(M)$ следует, что выполняется условие (9.15) определения непрерывности функции в точке, т.е. $\Delta u \to 0$ при $\Delta x_i \to 0$, $i = 1, 2, \ldots, m$, что и требовалось доказать.

Достаточные условия дифференцируемости функции нескольких переменных определяются следующей теоремой.

Теорема 10.3. *Если функция $u = f(M)$ имеет частные производные по всем аргументам в некоторой окрестности точки M, причем эти производные непрерывны в самой точке M, то данная функция дифференцируема в точке M.*

Доказательство этой теоремы мы опускаем, скажем только, что оно основано на использовании формулы Лагранжа для частных приращений функции по каждому аргументу.

Теорема 10.3 имеет важное следствие: *непрерывность функции вытекает из непрерывности ее частных производных.* Непрерывность функции нескольких переменных проверить напрямую бывает довольно трудно, но, оказывается, это важное свойство функции можно установить более простым способом, проверив непрерывность ее частных производных.

Условие дифференцируемости функции двух переменных $z = f(x, y)$ имеет геометрический смысл. Определим касательную плоскость к некоторой поверхности.

Определение 3. Плоскость, проходящая через точку N_0 поверхности, называется *касательной плоскостью* в данной точке, если угол между этой плоскостью и секущей, проходящей через точку N_0 и любую точку N на поверхности, стремится к нулю, когда точка N стремится по этой поверхности к точке N_0.

Если в точке N_0 поверхности существует касательная плоскость (рис. 10.2), то касательная в точке N_0 к любой кривой, расположенной на этой поверхности и проходящей через точку N_0, находится в касательной плоскости в данной точке поверхности.

Дифференциал функции нескольких переменных

Определение 4. *Дифференциалом* функции $u = f(M)$, дифференцируемой в точке $M(x_1, x_2, \ldots, x_m)$, называется главная

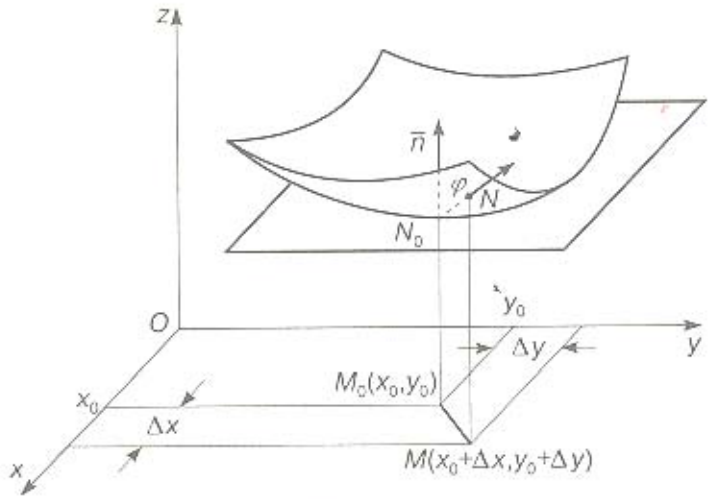

Рис. 10.2

линейная относительно приращений Δx_i $(i = 1, 2, \ldots, m)$ часть (10.6) приращения этой функции в точке M.

Таким образом, с учетом теоремы 10.1, дифференциалом функции $u = f(M)$ в точке M называется выражение

$$du = \frac{\partial u}{\partial x_1}\Delta x_1 + \frac{\partial u}{\partial x_2}\Delta x_2 + \cdots + \frac{\partial u}{\partial x_m}\Delta x_m. \qquad (10.8)$$

Назовем дифференциалами dx независимых переменных x_i соответствующие приращения Δx_i. Тогда формула для дифференциала функции нескольких переменных перепишется в виде

$$du = \frac{\partial u}{\partial x_1}dx_1 + \frac{\partial u}{\partial x_2}dx_2 + \cdots + \frac{\partial u}{\partial x_m}dx_m = \sum_{i=1}^{m}\frac{\partial u}{\partial x_i}dx_i. \qquad (10.9)$$

Отметим особо случай функции двух переменных $u = f(x, y)$; дифференциал для нее в точке $M(x, y)$ имеет вид

$$du = f'_x(x, y)dx + f'_y(x, y)dy. \qquad (10.10)$$

Аналогично тому, как дифференциал функции одной переменной представляет собой в геометрическом смысле приращение "ординаты касательной", дифференциал функции двух

переменных в этом плане является приращением "аппликаты касательной плоскости". Докажем этот геометрический смысл дифференциала функции двух переменных.

Покажем сначала, что касательная плоскость к поверхности в точке $N_0\,(x_0, y_0, z_0)$, где $z_0 = f(x_0, y_0)$, определяется уравнением

$$z - z_0 = f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0). \qquad (10.11)$$

Как известно из аналитической геометрии, уравнение (10.11) определяет плоскость с нормальным вектором $\bar{n} = \{f'_x, f'_y, -1\}$, проходящую через точку $N_0\,(x_0, y_0, z_0)$. Пусть $\Delta x = x - x_0$, $\Delta y = y - y_0$ и $\Delta z = z - z_0$. Докажем, что угол φ между вектором $\bar{n}$ и вектором $\overline{N_0N}$ любой секущей, проходящей через точки N $(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z)$ и N_0, стремится к $\pi/2$ при $N \to N_0$ (см. рис. 10.2), где Δx и Δy — приращения соответственно аргументов x и y. Так как координаты вектора $\overline{N_0N}$ равны Δx, Δy и Δz, то угол между векторами $\bar{n}$ и $\overline{N_0N}$ определяется по формуле

$$\cos\varphi = \frac{f'_x\Delta x + f'_y\Delta y - \Delta z}{[(f'_x)^2 + (f'_y)^2 + 1]^{1/2}[(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2]^{1/2}}.$$

Поскольку в силу соотношения (10.7) для случая функции двух переменных

$$\Delta z = f'_x\Delta x + f'_y\Delta y + o(\rho), \text{ где } \rho = \sqrt{(\Delta x)^2 + (\Delta y)^2}, \qquad (10.12)$$

то $|\cos\varphi| \leqslant \dfrac{|o(\rho)|}{[\rho^2 + (\Delta z)^2]^{1/2}} \leqslant \dfrac{|o(\rho)|}{\rho}$,

т.е. $\cos\varphi \to 0$ при $\rho \to 0$. Значит, при $N \to N_0$, согласно определению касательной плоскости, $\varphi \to 0$, что и требовалось доказать.

Поскольку приращение Δz "аппликаты касательной плоскости" определяется формулой

$$\Delta z = f'_x(x_0, y_0)\Delta x + f'_y(x_0, y_0)\Delta y,$$

то из равенства (10.11) получаем, что эта форма записи приращения совпадает дифференциалом dz функции $z = f(x, y)$

[см. формулу (10.10)], что и определяет геометрический смысл дифференциала функции двух переменных.

Следовательно, из условия дифференцируемости функции двух переменных $z = f(x, y)$ в точке $M_0\ (x_0, y_0)$ следует существование касательной плоскости к поверхности, соответствующей этой функции, в точке $N_0\ (x_0, y_0, z_0)$.

Пример 6. Замена факторов по функции Кобба–Дугласа: установить, на какую величину следует изменить объем вложенного капитала K, чтобы при изменении трудовых ресурсов на Δl выпуск продукции оставался неизменным.

Функция Кобба–Дугласа имеет вид (см. пример 5)

$$Q = AK^{\alpha}L^{1-\alpha}.$$

Полагая Q постоянным, получаем, что дифференциал этой функции равен нулю, т.е.

$$dQ = 0 \text{ или } \frac{\partial Q}{\partial K}\Delta K + \frac{\partial Q}{\partial L}\Delta L = 0, \text{ откуда}$$
$$\Delta K = -\Delta l \frac{\partial Q}{\partial L} \Big/ \frac{\partial Q}{\partial K},$$

или

$$\Delta K = -\frac{1-\alpha}{\alpha}\frac{K}{L}\Delta L.$$

В относительных величинах мы имеем отношение соответствующих эластичностей:

$$\frac{\Delta K}{K} = -\frac{1-\alpha}{\alpha}\frac{\Delta L}{L},$$

так что для компенсации изменения ресурса труда на 1% следует изменить ресурс капитала на $-\dfrac{1-\alpha}{\alpha}$ процентов. Формула для ΔK содержит важное экономическое понятие — *предельная норма* замены трудовых ресурсов L капиталом K:

$$\frac{dK}{dL} = -\frac{1-\alpha}{\alpha}\frac{K}{L}.$$

Производные сложной функции

Рассмотрим вопрос о дифференцировании сложной функции нескольких переменных вида

$$u = f(M) = f(x_1, x_2, \ldots, x_m), \tag{10.13}$$

где

$$x_i = \varphi_i(t_1, t_2, \ldots, t_k), \quad i = 1, 2, \ldots, m. \tag{10.14}$$

Теорема 10.4. *Пусть функции* (10.14) *дифференцируемы в некоторой точке* $T\ (t_1, t_2, \ldots, t_k)$, *а функция* (10.13) *дифференцируема в соответствующей точке* M *с координатами* $x_i = \varphi_i(T)$. *Тогда сложная функция* $u = f[\varphi_1(T), \varphi_2(T), \ldots, \varphi_m(T)]$ *дифференцируема в точке* T. *При этом частные производные сложной функции определяются по формулам*

$$\frac{\partial u}{\partial t_j} = \frac{\partial u}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \frac{\partial u}{\partial x_2}\frac{\partial x_2}{\partial t_j} + \cdots + \frac{\partial u}{\partial x_m}\frac{\partial x_m}{\partial t_j}, \tag{10.15}$$
$$j = 1, 2, \ldots, k,$$

где частные производные функции $\partial u/\partial x_1$, $\partial u/\partial x_2$, $\ldots$, $\partial u/\partial x_m$ *берутся в точке* M, *а частные производные функций* x_i *по аргументам* t_j *берутся в точке* T.

Доказательство. Придадим переменным t_j в точке T приращения Δt_j, не равные одновременно нулю. Тогда, в силу дифференцируемости функций (10.14) в точке T, соответствующие приращения Δx_i имеют вид, согласно теореме 10.1 и формуле (10.7),

$$\Delta x_i = \frac{\partial x_i}{\partial t_1}\Delta t_1 + \frac{\partial x_i}{\partial t_2}\Delta t_2 + \cdots + \frac{\partial x_i}{\partial t_k}\Delta t_k + o_i(\rho_T) =$$
$$\vdots$$
$$= \sum_{j=1}^{k}\frac{\partial x_i}{\partial t_j}\Delta t_j + o_i(\rho_T),$$
$$\rho_T = \sqrt{(\Delta t_1)^2 + (\Delta t_2)^2 + \cdots + (\Delta t_k)^2},\ i = 1, 2, \ldots, m, \tag{10.16}$$

где $o_i(\rho_T)$ — бесконечно малые при $\Delta t_j \to 0$ в точке T.

В свою очередь, приращениям Δx_i соответствует приращение функции $u = f(M)$ в точке M, которое, в силу теоремы 10.1 и формулы (10.7), равно

$$\Delta u = \frac{\partial u}{\partial x_1}\Delta x_1 + \frac{\partial u}{\partial x_2}\Delta x_2 + \cdots + \frac{\partial u}{\partial x_m}\Delta x_m + o(\rho_M) = \sum_{i=1}^{m} \frac{\partial u}{\partial x_i}\Delta x_i + o(\rho_M), \tag{10.17}$$

$$\rho_M = \sqrt{(\Delta x_1)^2 + (\Delta x_2)^2 + \cdots + (\Delta x_m)^2},$$

где $o(\rho_M)$ — бесконечно малая при $\Delta x_i \to 0$ в точке M.

Подставляя в (10.17) формулы (10.16) для приращений Δx_i, меняя порядок суммирования и отделяя слагаемые первого порядка малости от слагаемых, содержащих $o_i(\rho_T)$, имеем

$$\Delta u \sum_{i=1}^{m} \frac{\partial u}{\partial x_i}\left(\sum_{j=1}^{k} \frac{\partial x_i}{\partial t_j}\Delta t_j + o_i(\rho_T)\right) + o(\rho_M) = \sum_{j=1}^{k} A_j \Delta t_j + \sum_{i=1}^{m} \frac{\partial u}{\partial x_i} o_i(\rho_T) + o(\rho_M), \tag{10.18}$$

где

$$A_j = \sum_{i=1}^{m} \frac{\partial u}{\partial x_i}\frac{\partial x_i}{\partial t_j} = \frac{\partial u}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \frac{\partial u}{\partial x_2}\frac{\partial x_2}{\partial t_j} + \cdots + \frac{\partial u}{\partial x_m}\frac{\partial x_m}{\partial t_j}, \quad j = 1, 2, \ldots, k. \tag{10.19}$$

Второе слагаемое в (10.18) является бесконечно малым относительно ρ_T (как произведение конечного числа ограниченных величин на бесконечно малые); последнее слагаемое в силу дифференцируемости x_i (а значит, и их непрерывности) также является бесконечно малым относительно ρ_T ($\rho_M \to 0$ при $\rho_T \to 0$). Следовательно, по теореме 10.1 сложная функция

$$u = f[\varphi_1(T), \varphi_2(T), \ldots, \varphi_m(T)]$$

дифференцируема в точке T, причем ее частные производные, согласно (10.19), определяются формулами (10.15), что и требовалось доказать.

Отметим важный частный случай, когда $k = 1$, т.е. переменные x_i зависят от одного аргумента t: $u = f[\varphi_1(t), \varphi_2(t), \ldots, \varphi_m(t)]$. Тогда производная этой сложной функции определяется формулой

$$\frac{du}{dt} = \frac{\partial u}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial u}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial u}{\partial x_m}\frac{dx_m}{dt}. \tag{10.20}$$

В качестве упражнения советуем повторить доказательство теоремы 10.4 для этого случая, приняв $m = 2$, т.е. рассмотреть функцию $u = f(x, y)$, $x = \varphi(t)$, $y = \psi(t)$.

Рассмотрим примеры дифференцирования сложных функций нескольких переменных.

Пример 7. $z = x\cos xy$, где $x = t^2$, $y = \sin t$.

По формуле (10.20)

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = (\cos xy - xy\sin xy)2t - x^2 \sin xy \cos t.$$

Пример 8. $u = xy^2 + \sin 3z$, $x = 2t$, $y = t^2$, $z = t^3$.

Согласно формуле (10.20),

$$\frac{du}{dt} = \frac{\partial u}{\partial x}\frac{dx}{dt} + \frac{\partial u}{\partial y}\frac{dy}{dt} + \frac{\partial u}{\partial z}\frac{dz}{dt} = 2y^2 + 2xy \cdot 2t + 3\cos 3z 3t^2 = 10t^4 + 9t^2 \cos 3t^3.$$

Пример 9. $u = xy^2$, $x = 2s + t$, $y = s\sin t$.

Применяем формулы (10.15), а именно $x_1 = x$, $x_2 = y$, $t_1 = s$, $t_2 = t$; получаем

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} = y^2 \cdot 2 + 2xy\sin t = 2s^2 \sin^2 t + 2s(2s + t)\sin^2 t;$$

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t} = y^2 + 2xys\cos t = s^2 \sin^2 t + (2s + t)s^2 \sin 2t.$$

Однородные функции

Определение 5. Функция $u = f(M)$, определенная на множестве $\{M\}$, называется *однородной функцией степени* p на этом множестве, если для любой точки $M\,(x_1, x_2, \ldots, x_m)$ этого множества и для каждого числа t, для которого точка $N\,(tx_1, tx_2, \ldots, tx_m)$ также принадлежит $\{M\}$, выполняется равенство

$$f(tx_1, tx_2, \ldots, tx_m) = t^p f(x_1, x_2, \ldots, x_m). \tag{10.21}$$

Рассмотрим некоторые примеры однородных функций.

1. Функция $u = xy^2 + xz^2 + yz^2$ является однородной функцией третьей степени, определенной на всем евклидовом пространстве E^3.

2. Производственная функция Кобба–Дугласа (см. пример 5)

$$Q = f(K, L) = AK^{\alpha}L^{1-\alpha},$$

где K и L — соответственно затраты капитала и труда, Q — выпуск продукции, α и $1-\alpha$ — эластичности выпуска продукции по капиталу и труду. Очевидным образом функция Кобба–Дугласа — однородная функция первой степени.

Однородные функции играют важную роль в различных приложениях, в том числе и в экономике. Для них справедлива следующая теорема.

Теорема 10.5 (ТЕОРЕМА ЭЙЛЕРА). *Пусть $u = f(M)$ — дифференцируемая однородная функция степени p на некотором множестве $\{M\}$. Тогда для любой точки $M(x_1, x_2, \ldots, x_m)$ этого множества справедливо равенство*

$$\frac{\partial u}{\partial x_1}x_1 + \frac{\partial u}{\partial x_2}x_2 + \cdots + \frac{\partial u}{\partial x_m}x_m = pu. \tag{10.22}$$

Доказательство. Рассмотрим сложную функцию $u = f(x_1, x_2, \ldots, x_m)$, где $x_i = tx_i^0$, причем точки $M(x_1, x_2, \ldots, x_m)$ и $M_0(x_1^0, x_2^0, \ldots, x_m^0)$ принадлежат множеству $\{M\}$. Производная этой функции определяется по

формуле (10.20). С учетом того, что $dx_i/dt = x_i^0$, при $t = 1$ получаем

$$\left.\frac{du}{dt}\right|_{t=1} = \frac{\partial u}{\partial x_1}x_1^0 + \frac{\partial u}{\partial x_2}x_2^0 + \cdots + \frac{\partial u}{\partial x_m}x_m^0, \tag{10.23}$$

где производные $\partial u/\partial x_i$ вычислены в точке M_0.

С другой стороны, поскольку данная функция, согласно условию теоремы, является однородной степени p, то

$$f(tx_1^0, tx_2^0, \ldots, tx_m^0) = t^p f(x_1^0, x_2^0, \ldots, x_m^0),$$

а значит, $\dfrac{du}{dt} = pt^{p-1}f(x_1^0, x_2^0, \ldots, x_m^0)$. Отсюда при $t = 1$ получаем

$$\left.\frac{du}{dt}\right|_{t=1} = pf(x_1^0, x_2^0, \ldots, x_m^0) = pu.$$

Сравнивая это равенство с (10.23), в силу произвольности точки M_0 получаем утверждение теоремы (10.22).

Инвариантность формы первого дифференциала

Согласно формуле (10.9), дифференциал du функции m переменных можно представить в виде

$$du = \sum_{i=1}^{m}\frac{\partial u}{\partial x_i}dx_i. \tag{10.24}$$

Оказывается, эта форма записи первого дифференциала универсальна, т.е. она не меняется и в случае, когда переменные x_i являются функциями других переменных $t_1, t_2, \ldots, t_k$, определенными на множестве $\{T\}$ евклидова пространства E^k. Как и для случая функции одной переменной, это свойство называется *инвариантностью формы первого дифференциала*.

Действительно, дифференциал сложной функции можно представить в виде

$$du = \sum_{j=1}^{k}\frac{\partial u}{\partial t_j}dt_j. \tag{10.25}$$

Подставляя сюда выражения для частных производных $\partial u/\partial t_j$ из (10.15), получаем, приводя подобные при $\partial u/\partial x_i$ (или меняя порядок суммирования),

$$du = \frac{\partial u}{\partial x_1}\left(\sum_{j=1}^{k}\frac{\partial x_1}{\partial t_j}dt_j\right) + \frac{\partial u}{\partial x_2}\left(\sum_{j=1}^{k}\frac{\partial x_2}{\partial t_j}dt_j\right) + \cdots + \frac{\partial u}{\partial x_m}\left(\sum_{j=1}^{k}\frac{\partial x_m}{\partial t_j}dt_j\right).$$

Все суммы, стоящие в круглых скобках, представляют собой соответствующие дифференциалы dx_i функций зависимых переменных $x_i = \varphi_i(t_1, t_2, \ldots, t_k)$. Таким образом, из формы для дифференциала сложной функции (10.25) мы получили формулу (10.9), в которой дифференциалы dx_i являются дифференциалами функций зависимых переменных, что и доказывает инвариантность формы первого дифференциала функции нескольких переменных.

Установленное свойство инвариантности формы первого дифференциала позволяет вывести основные правила дифференцирования для функций u и v каких-либо переменных:

I. $d(cu) = cdu (c = \text{const})$.

II. $d(u \pm v) = du \pm dv$.

III. $d(uv) = udv + vdu$.

IV. $d\left(\frac{u}{v}\right) = \frac{vdu - udv}{v^2}$.

Приведем, например, вывод последней формулы. Пусть $w = \frac{u}{v}$. Тогда

$$dw = \frac{\partial w}{\partial u}du + \frac{\partial w}{\partial v}dv = \frac{1}{v}du - \frac{u}{v^2}dv = \frac{vdu - udv}{v^2}.$$

Форма этого дифференциала (как и остальных) останется неизменной и в том случае, если u и v сами являются функциями каких-либо переменных.

Производная по направлению

Пусть функция двух переменных $u = f(M)$ задана в некоторой окрестности точки $M(x, y)$. Рассмотрим некоторое направление, определяемое единичным вектором $\bar{l} = \{\cos\alpha, \cos\beta\}$, где $\cos^2\alpha + \cos^2\beta = 1$ (рис. 10.3). На прямой, проходящей по этому направлению через точку M, возьмем точку $M_1(x + \Delta x, y + \Delta y)$, так что длина Δl отрезка MM_1 равна $\sqrt{(\Delta x)^2 + (\Delta y)^2}$. Приращение функции $f(M)$ определяется формулой

$$\Delta u = f(x + \Delta x, y + \Delta y) - f(x, y),$$

где Δx и Δy связаны соотношениями $\Delta x = \Delta l\cos\alpha$, $\Delta y = \Delta l\cos\beta$.

Определение 6. Предел отношения $\frac{\Delta u}{\Delta l}$ при $\Delta l \to 0$ называется *производной* функции $u = f(M)$ в точке $M(x, y)$ *по направлению* $\bar{l}$ и обозначается символом $\frac{\partial u}{\partial l}$:

$$\frac{\partial u}{\partial l} = \lim_{\Delta l \to 0}\frac{\Delta u}{\Delta l}.$$

Если функция $f(M)$ дифференцируема в точке $M(x, y)$, то ее приращение в этой точке, с учетом соотношений для Δx и

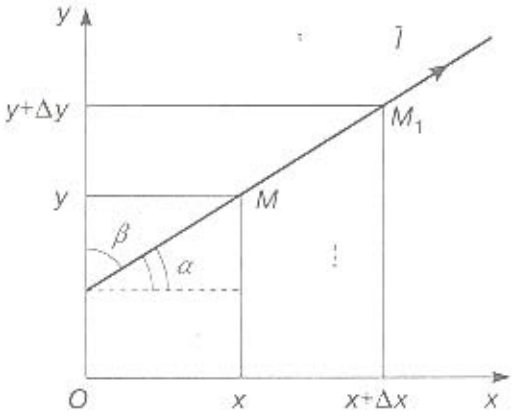

Рис. 10.3

Δy, может быть записано в форме (10.12):

$$\Delta u = \frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y + o(\Delta l) =$$
$$= \frac{\partial u}{\partial x}\Delta l\cos\alpha + \frac{\partial u}{\partial y}\Delta l\cos\beta + o(\Delta l).$$

Поделив обе части этого равенства на Δl и переходя к пределу при $\Delta l \to 0$, получаем формулу для производной функции $u = f(x, y)$ по направлению:

$$\frac{\partial u}{\partial l} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\cos\beta. \qquad (10.26)$$

Аналогичным образом для функции, определенной и дифференцируемой в некоторой окрестности точки M евклидова пространства E^m, определяется производная по направлению единичного m-мерного вектора $\bar{l} = \{\cos\alpha_1, \cos\alpha_2, \ldots, \cos\alpha_m\}$:

$$\frac{\partial u}{\partial l} = \frac{\partial u}{\partial x_1}\cos\alpha_1 + \frac{\partial u}{\partial x_2}\cos\alpha_2 + \cdots + \frac{\partial u}{\partial x_m}\cos\alpha_m, \qquad (10.27)$$

где

$$\cos^2\alpha_1 + \cos^2\alpha_2 + \cdots + \cos^2\alpha_m = 1.$$

Рассмотрим примеры нахождения производных функций по направлению.

Пример 10. $z = x^2 + y^2$ в точке $M(1, 1)$, направление составляет с осью Ox угол $\alpha = \pi/3$.

Следовательно, угол $\beta = \pi/6$; по формуле (10.26)

$$\frac{\partial z}{\partial l} = \frac{\partial z}{\partial x}\cos\frac{\pi}{3} + \frac{\partial z}{\partial y}\cos\frac{\pi}{6} = 2x\frac{1}{2} + 2y\frac{\sqrt{3}}{2} = x + \sqrt{3}y.$$

В точке $M\ (1, 1)$ получаем:

$$\left.\frac{\partial z}{\partial l}\right|_{(1,1)} = 1 + \sqrt{3}.$$

Пример 11. $u = \dfrac{x^2}{2} + \dfrac{y^2}{9} - z^2$ в точке $M\ (2, 3, 1)$, направление совпадает с радиус-вектором этой точки.

Сначала находим координаты единичного вектора данного направления: радиус-вектор $\bar{a} = \{2, 3, 1\}$, его длина $a = \sqrt{14}$, значит, $\cos\alpha = 2/\sqrt{14}$, $\cos\beta = 3/\sqrt{14}$, $\cos\gamma = 1/\sqrt{14}$. Теперь по формуле (10.27) при $m = 3$

$$\frac{\partial u}{\partial l} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\cos\beta + \frac{\partial u}{\partial z}\cos\gamma = \left(2x + 3\frac{2}{9}y - 2z\right)\Big/\sqrt{14}.$$

В точке $M(2, 3, 1)$ производная по направлению

$$\left.\frac{\partial u}{\partial l}\right|_{(2,3,1)} = (4 + 2 - 2)/\sqrt{14} = 4/\sqrt{14}.$$

Градиент

Рассмотрим функцию трех переменных $u = f(M)$, дифференцируемую в некоторой точке $M\ (x, y, z)$.

Определение 7. *Градиентом* функции $u = f(x, y, z)$ в точке M называется вектор, координаты которого равны соответственно частным производным $\dfrac{\partial u}{\partial x}$, $\dfrac{\partial u}{\partial y}$, $\dfrac{\partial u}{\partial z}$ в этой точке.

Для обозначения градиента функции используется символ $\operatorname{grad} u$:

$$\operatorname{grad} u = \left\{\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}\right\}. \qquad (10.28)$$

Поскольку единичный вектор $\bar{l}$ имеет координаты $\cos\alpha$, $\cos\beta$ и $\cos\gamma$, то производная по направлению для случая функции трех переменных запишется в виде

$$\frac{\partial u}{\partial l} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\cos\beta + \frac{\partial u}{\partial z}\cos\gamma = \bar{l}\operatorname{grad} u, \qquad (10.29)$$

т.е. $\dfrac{\partial u}{\partial l}$ имеет форму скалярного произведения векторов $\bar{l}$ и $\operatorname{grad} u$.

Перепишем формулу (10.29) в другом виде:

$$\frac{\partial u}{\partial l} = \bar{l}\operatorname{grad} u = |\bar{l}||\operatorname{grad} u|\cos\varphi,$$

где φ — угол между векторами $\bar{l}$ и $\operatorname{grad} u$. Поскольку $|\bar{l}| = 1$, то из этого равенства следует, что производная функции по направлению принимает максимальное значение при $\varphi = 0$, т.е. когда направления векторов $\bar{l}$ и $\operatorname{grad} u$ совпадают; при этом $\left(\dfrac{\partial u}{\partial l}\right)_{\max} = |\operatorname{grad} u|$.

Таким образом, градиент функции характеризует направление и величину максимальной скорости возрастания этой функции в точке.

Для определения геометрического смысла градиента функции введем понятие поверхности уровня. Это понятие аналогично понятию линии уровня, рассмотренному в параграфе 9.4.

Определение 8. *Поверхностью уровня* функции $u=f(x,y,z)$ называется поверхность, на которой эта функция сохраняет постоянное значение

$$f(x,y,z) = c = \text{const}. \tag{10.30}$$

Рассмотрим дифференциал поверхности уровня (10.30)

$$du = \frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial y}dy + \frac{\partial u}{\partial z}dz = 0.$$

Поскольку вектор $\{dx, dy, dz\}$ лежит в касательной плоскости в точке $M\ (x,y,z)$ поверхности уровня, то градиент в данной точке ортогонален к этой поверхности.

В случае функции двух переменных все сказанное выше остается в силе, только вместо поверхности уровня будет фигурировать линия уровня.

Понятие градиента функции $u = f(M)$ в точке $M(x_1, x_2, \ldots, x_m)$ обобщается на случай любого числа переменных m:

$$\operatorname{grad} u = \left\{\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \ldots, \frac{\partial u}{\partial x_m}\right\}. \tag{10.31}$$

При этом для производной по направлению (10.27) также справедлива формула

$$\frac{\partial u}{\partial l} = \bar{l}\operatorname{grad} u.$$

Рассмотрим некоторые примеры.

Пример 12. Найти градиент и его модуль для функции $z = \dfrac{xy}{x+y+1}$ в точке $M(0,1)$.

По формуле (10.31) при $m=2$, $x_1 = x$, $x_2 = y$

$$\operatorname{grad} z = \left\{\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right\} = \left\{\frac{y(y+1)}{(x+y+1)^2}, \frac{x(x+1)}{(x+y+1)^2}\right\}.$$

При $x=0$ и $y=1$ получаем

$$\operatorname{grad} z\Big|_{(0,1)} = \{1,0\}, \quad |\operatorname{grad} z| = 1.$$

Пример 13. Найти градиент и его модуль для функции $u = x^2 + y^2 - z^2$ в точке $M(1,1,-2)$.

По формуле (10.28) имеем

$$\operatorname{grad} u = \left\{\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}\right\} = \{2x, 2y, -2z\}.$$

Подставляя в это выражение координаты точки M, получаем

$$\operatorname{grad} u\big|_{(1,1,-2)} = \{2,2,-4\},$$
$$|\operatorname{grad} u| = \sqrt{4+4+16} = 2\sqrt{6}.$$

Пример 14. Найти поверхности уровня функции $u = x^2 - 2x + y^2 + 2y - z$.

Согласно определению поверхности уровня (10.30),

$$x^2 - 2x + y^2 + 2y - z = c,$$

откуда

$$z = (x-1)^2 + (y+1)^2 - C,$$

где $C = c + 2$. Следовательно, поверхностями уровня данной функции являются параболоиды вращения с осью $x=1$, $y=-1$, параллельной оси Oz, вершины которых лежат в точках с координатами $(1,-1,-C)$.

22 Зак. 6853

10.2. Производные и дифференциалы высших порядков

Частные производные высших порядков

В том случае, когда частная производная $\dfrac{\partial}{\partial x_i}$ функции нескольких переменных $u = f(x_1, x_2, \ldots, x_m)$ существует в каждой точке области $\{M\}$, она является функцией этих переменных в указанной области. Тогда по аналогии с определением частных производных функции $u(M)$ (см. параграф 10.1) можно определить частную производную от функции $\dfrac{\partial u}{\partial x_i}$ по аргументу x_k в некоторой точке M области $\{M\}$. Такая частная производная называется *второй частной производной*, или *частной производной второго порядка*, функции $u = f(x_1, x_2, \ldots, x_m)$ в точке M; сначала она берется по аргументу x_i, а затем по аргументу x_k и обозначается одним из следующих символов:

$$\frac{\partial^2 u}{\partial x_i \partial x_k}, \quad f''_{x_i x_k}, \quad u''_{x_i x_k}. \tag{10.32}$$

При $i \neq k$ частная производная (10.32) называется *смешанной частной производной* второго порядка.

После того как определено понятие частной производной второго порядка, можно последовательно ввести понятие частной производной третьего порядка, затем четвертой производной и т.д. Пусть уже введено понятие $(n-1)$-й частной производной функции $u = f(x_1, x_2, \ldots, x_m)$ по аргументам $x_{i_1}, x_{i_2}, \ldots, x_{i_{n-1}}$, причем отдельные или даже все номера аргументов могут совпадать. Пусть эта $(n-1)$-я производная имеет в точке M частную производную, тогда она называется n-й частной производной, или частной производной n-го порядка, функции $u = f(x_1, x_2, \ldots, x_m)$ по аргументам $x_{i_1}, x_{i_2}, \ldots, x_{i_n}$. Такое определение понятия n-й частной производной называется *индуктивным*, с переходом от первой частной производной к последующим. Таким образом, соотношение, определяющее n-ю частную производную, имеет следующий вид:

$$\frac{\partial^n u}{\partial x_{i_n} \partial x_{i_{n-1}} \ldots \partial x_{i_2} \partial x_{i_1}} = \frac{\partial}{\partial x_{i_n}} \left(\frac{\partial^{n-1} u}{\partial x_{i_{n-1}} \ldots \partial x_{i_2} \partial x_{i_1}} \right). \tag{10.33}$$

Если не все индексы аргументов, по которым берется частная производная (10.33), совпадают, то она называется *смешанной частной производной* n-го порядка. Поскольку частная производная функции по аргументу x_i определяется как обыкновенная производная функции одной переменной x_i при фиксированных значениях других переменных, то и вычисление частных производных высших порядков производится уже известными приемами вычисления обыкновенных производных первого порядка.

Теперь введем понятие n раз дифференцируемой функции нескольких переменных. *Функция $u = f(x_1, x_2, \ldots, x_m)$ называется n раз дифференцируемой в точке M_0 $(x_1^0, x_2^0, \ldots, x_m^0)$, если все ее частные производные $(n-1)$-го порядка являются дифференцируемыми функциями в этой точке.*

Справедливо следующее утверждение, вытекающее из определения дифференцируемости функции нескольких переменных и теоремы 10.3.

Для того чтобы функция $u = f(M)$ была n раз дифференцируема в точке M_0 $(x_1^0, x_2^0, \ldots, x_m^0)$, достаточно, чтобы все ее частные производные n-го порядка были непрерывными функциями в точке M_0.

Рассмотрим два примера нахождения частных производных второго порядка для функции двух переменных.

Пример 1. $u = x^3 - xy^2 + x + y + y^4$.

Последовательно дифференцируя, имеем

$$\frac{\partial u}{\partial x} = 3x^2 - y^2 + 1, \quad \frac{\partial u}{\partial y} = -2xy + 1 + 4y^3,$$

$$\frac{\partial^2 u}{\partial x^2} = 6x, \quad \frac{\partial^2 u}{\partial x \partial y} = -2y, \quad \frac{\partial^2 u}{\partial y \partial x} = -2y, \quad \frac{\partial^2 u}{\partial y^2} = 12y^2 - 2x.$$

Пример 2. $u = e^x \cos 2y$.

Дифференцируя, имеем

$$\frac{\partial u}{\partial x} = e^x \cos 2y, \quad \frac{\partial u}{\partial y} = -2e^x \sin 2y, \quad \frac{\partial^2 u}{\partial x^2} = e^x \cos 2y,$$

$$\frac{\partial^2 u}{\partial x \partial y} = -2e^x \sin 2y, \quad \frac{\partial^2 u}{\partial y \partial x} = -2e^x \sin 2y, \quad \frac{\partial^2 u}{\partial y^2} = -4e^x \cos 2y.$$

В рассмотренных примерах смешанные производные оказались равными, хотя это бывает и не всегда. Ответ на вопрос о независимости смешанных вторых производных от порядка дифференцирования функции двух переменных дает следующая теорема.

Теорема 10.6. *Если функция $u = f(x, y)$ дважды дифференцируема в точке M (x, y), то ее смешанные производные в этой точке равны.*

Доказательство. Поскольку данная функция дважды дифференцируема в точке M, то ее частные производные f'_x и f'_y определены в некоторой окрестности точки M и дифференцируемы в этой точке. Возьмем приращения аргументов Δx и Δy настолько малые, чтобы точка M_1 $(x+\Delta x, y = \Delta y)$ находилась в указанной окрестности точки M.

Рассмотрим выражение

$$\Phi = [f(x + \Delta x, y + \Delta y) - f(x + \Delta x, y)] - \\ -[f(x, y + \Delta y) - f(x, y)].$$

Если ввести вспомогательную функцию $\varphi(x) = f(x, y + \Delta y) - -f(x, y)$, то выражение Φ будет являться приращением дифференцируемой на отрезке $[x, x + \Delta x]$ функции одной переменной $\varphi(x)$:

$$\Phi = \Delta\varphi = \varphi(x + \Delta x) - \varphi(x).$$

Применяя здесь теорему Лагранжа (см. теорему 5.3), получаем

$$\Phi = \varphi'(x + \alpha_1\Delta x)\Delta x = \\ = [f'_x(x + \alpha_1\Delta x, y + \Delta y) - f'_x(x + \alpha_1\Delta x, y)],$$

где $0 < \alpha_1 < 1$. В свою очередь, выражение в квадратных скобках можно рассматривать как приращение на отрезке $[y, y+\Delta y]$ дифференцируемой функции одной переменной y. Повторное применение теоремы Лагранжа по переменной y дает

$$\Phi = f''_{yx}(x + \alpha_1\Delta x, y + \alpha_2\Delta y),\ 0 < \alpha_1 < 1,\ 0 < \alpha_2 < 1. \quad (10.34)$$

Введем другую вспомогательную функцию $\psi(y) = = f(x+\Delta x, y) - f(x, y)$; проводя по отношению к ней те же операции, что и для функции $\varphi(x)$, получим

$$\Phi = \Delta\psi = \psi(y + \Delta y) - \psi(y) = \\ = f''_{xy}(x + \beta_1\Delta x, y + \beta_2\Delta y), \quad (10.35)$$

где $0 < \beta_1 < 1$, $0 < \beta_2 < 1$. Сравнивая выражения (10.34) и (10.35), получаем

$$f''_{yx}(x + \alpha_1\Delta x, y + \alpha_2\Delta y) = f''_{xy}(x + \beta_1\Delta x, \beta_2\Delta y).$$

Так как по условию теоремы обе смешанные частные производные непрерывны в точке M, то, переходя к пределу в последнем равенстве при $\Delta x \to 0$ и $\Delta y \to 0$ (или, что то же самое, при $M_1 \to M$), получаем утверждение теоремы:

$$f''_{xy}(x, y) = f''_{yx}(x, y).$$

Относительно смешанных частных производных n-го порядка функций нескольких переменных справедлива более общая теорема, которую мы приводим здесь без доказательства.

Теорема 10.7. *Если функция $u = f(M)$ дифференцируема n раз в точке $M(x_1, x_2, \ldots, x_m)$, то значение любой смешанной производной n-го порядка в этой точке не зависит от порядка последовательных дифференцирований данной функции.*

Дифференциалы высших порядков

Функция двух переменных. Пусть функция $u = f(x, y)$ дифференцируема в точке M (x, y) и в некоторой ее окрестности. Тогда, как мы знаем, дифференциал этой функции в точке M имеет вид [см. формулу (10.10)]

$$du = f'_x(x, y)dx + f'_y(x, y)dy.$$

Этот дифференциал называют *дифференциалом первого порядка*. Пусть функция $u = f(x, y)$ дважды дифференцируема в точке M. Будем полагать в этой формуле dx и dy постоянными, тогда du является функцией переменных x и y, дифференци-

руемой в точке M; ее дифференциал в точке $M\ (x, y)$, согласно формуле (10.10), имеет вид

$$\delta(du) = \delta(f'_x dx + f'_y dy) = (f'_x dx + f'_y dy)'_x \delta x + (f'_x dx + f'_y dy)'_y \delta y.$$

Это выражение при $\delta x = dx$ и $\delta y = dy$ называется *вторым дифференциалом*, или *дифференциалом второго порядка*, функции $u = f(M)$ в точке $M\ (x, y)$ и обозначается d^2u. Выполняя дифференцирование в последней формуле, в силу равенства смешанных производных, получаем выражение для второго дифференциала

$$\begin{aligned} d^2u &= f''_{xx} dx^2 + 2f''_{xy} dxdy + f''_{yy} dy^2 = \\ &= \frac{\partial^2 u}{\partial x^2} dx^2 + 2\frac{\partial^2 u}{\partial x \partial y} dxdy + \frac{\partial^2 u}{\partial y^2} dy^2. \end{aligned} \tag{10.36}$$

Выражения для первого и второго дифференциалов удобно записать в символической форме, введя символ дифференциала соотношением

$$d = \frac{\partial}{\partial x} dx + \frac{\partial}{\partial y} dy. \tag{10.37}$$

Определим операцию возведения этого символа в степень как обычное возведение в степень двучлена в правой части (10.37). Тогда формулу (10.36) можно переписать в виде

$$\begin{aligned} d^2u &= \left(\frac{\partial}{\partial x} dx + \frac{\partial}{\partial y} dy\right)^2 u = \\ &= \left(\frac{\partial^2}{\partial x^2} dx^2 + 2\frac{\partial^2}{\partial x \partial y} dxdy + \frac{\partial^2}{\partial y^2} dy^2\right) u. \end{aligned} \tag{10.38}$$

Аналогичным образом можно определить *дифференциал третьего порядка* $\delta(d^2u)$ функции $u = f(M)$, трижды дифференцируемой в точке $M\ (x, y)$, который имеет вид

$$\begin{aligned} d^3u &= \left(\frac{\partial}{\partial x} dx + \frac{\partial}{\partial y} dy\right)^3 u = \\ &= \left(\frac{\partial^3}{\partial x^3} dx^3 + 3\frac{\partial^3}{\partial x^2 \partial y} dx^2 dy + 3\frac{\partial^3}{\partial x \partial y^2} dxdy^2 + \frac{\partial^3}{\partial y^3} dy^3\right) u. \end{aligned}$$

В формулах (10.38) и далее после формального возведения в степень достаточно тоже формально перемножить функцию u на все слагаемые в скобках, чтобы получить выражения для соответствующего дифференциала. Нетрудно видеть, что формальное возведение в степень символа (10.37) напоминает разложение по формуле бинома Ньютона. Выражение для *дифференциала n-го порядка* d^nu также запишем в этом удобном виде:

$$\delta(d^{n-1}u) = d^n u = \left(\frac{\partial}{\partial x} dx + \frac{\partial}{\partial y} dy\right)^n u. \tag{10.39}$$

Пример 3. Найти d^4u для функции $u = x^2y + 2xy^2 + 3x + y^4$.

Согласно формуле (10.39), нам нужно получить выражения для всех частных производных четвертого порядка; исходя из вида данной функции

$$\frac{\partial^4 u}{\partial x^4} = 0,\ \frac{\partial^4 u}{\partial x^3 \partial y} = 0,\ \frac{\partial^4 u}{\partial x^2 \partial y^2} = 0,\ \frac{\partial^4 u}{\partial x \partial y^3} = 0,\ \frac{\partial^4 u}{\partial y^4} = 24.$$

Выполняя разложение в (10.39) при $n = 4$ и подставляя в него найденные значения для частных производных, окончательно получаем

$$d^4u = 24dy^4.$$

Функция нескольких переменных. По аналогии с рассмотренным уже случаем функции двух переменных также индуктивно вводится понятие дифференциала n-го порядка для функции $u = f(M)$ в точке $M(x_1, x_2, \ldots, x_m)$. Для удобства записи и запоминания введем символ d по аналогии с (10.37), полагая

$$d = \frac{\partial}{\partial x_1} dx_1 + \frac{\partial}{\partial x_2} dx_2 + \cdots + \frac{\partial}{\partial x_m} dx_m. \tag{10.40}$$

Определим операцию возведения символа d в степень как обычное возведение в степень многочлена в правой части (10.40). Тогда для n-го дифференциала функции $u = f(M)$ получим выражение

$$d^n u = \left(\frac{\partial}{\partial x_1} dx_1 + \frac{\partial}{\partial x_2} dx_2 + \cdots + \frac{\partial}{\partial x_m} dx_m\right)^n u. \tag{10.41}$$

Неинвариантность формы дифференциалов высших порядков. Выше мы полагали, что значения дифференциалов dx_i фиксированы, что выполняется в случае независимых переменных $x_1, x_2, \ldots, x_m$. Если же аргументы x_i являются дифференцируемыми функциями других независимых переменных $t_1, t_2, \ldots, t_k$, то выражение дифференциала n-го порядка при $n \geqslant 2$ отличается от его формы (10.41). Убедимся в этом для случая функции двух переменных $u=f(x, y)$. Пусть $x=x(t_1, t_2)$, $y = y(t_1, t_2)$. По определению второго дифференциала имеем, полагая постоянными dt_1 и dt_2 и используя свойство инвариантности первого дифференциала,

$$d^2u = d\left(\frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial y}dy\right) = d\left(\frac{\partial u}{\partial x}\right)dx + d\left(\frac{\partial u}{\partial y}\right)dy + \frac{\partial u}{\partial x}d(dx) +$$
$$+\frac{\partial u}{\partial y}d(dy) = \frac{\partial^2 u}{\partial x^2}dx^2 + 2\frac{\partial^2 u}{\partial x \partial y}dxdy + \frac{\partial^2 u}{\partial y^2}dy^2 + \frac{\partial u}{\partial x}d^2x + \frac{\partial u}{\partial y}d^2y =$$
$$= \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^2 u + \frac{\partial u}{\partial x}d^2x + \frac{\partial u}{\partial y}d^2y.$$

Полученное выражение очевидным образом отличается от формы второго дифференциала (10.38), что и требовалось показать.

Формула Тейлора для функции нескольких переменных

Теорема 10.8. *Пусть функция $u = f(M)$ дифференцируема $(n+1)$ раз в некоторой ε-окрестности точки $M(x_1, x_2, \ldots, x_m)$. Тогда приращение этой функции $\Delta u = f(M_1) - f(M)$ в точке M для любой точки M_1 $(x_1 + \Delta x_1, x_2 + \Delta x_2, \ldots, x_m + \Delta x_m)$ из указанной окрестности можно представить в следующей форме:*

$$\Delta u = df(M) + \frac{1}{2!}d^2 f(M) + \cdots + \frac{1}{n!}d^n f(M) + \frac{1}{(n+1)!}d^{n+1} f(M'), \tag{10.42}$$

где $M'(x_1 + \theta \cdot \Delta x_1, x_2 + \theta \cdot \Delta x_2, \ldots, x_m + \theta \cdot \Delta x_m)$ — некоторая точка указанной ε-окрестности точки M $(0 < \theta < 1)$.

Формула (10.42) называется *формулой Тейлора для функции $u = f(M)$ с центром разложения в точке M.*

Доказательство. Проведем доказательство для случая функции двух переменных, т.е. $u = f(x, y)$.

Введем вспомогательную функцию

$$F(t) = f(x + t\Delta x, y + t\Delta y) = f(M_t),$$

которая является функцией одной переменной t. Нетрудно видеть, что последовательное дифференцирование функции $F(t)$ дает

$$F'(t) = u'_x(M_t)\Delta x + u'_y(M_t)\Delta y = \left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right) f(M_t),$$
$$F''(t) = u''_{xx}(M_t)(\Delta x)^2 + 2u''_{xy}(M_t)\Delta x \Delta y + u''_{yy}(M_t)(\Delta y)^2 =$$
$$= \left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^2 f(M_t).$$

По индукции находим, что для любого n справедлива формула

$$F^{(n)}(t) = \left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^n f(M_t). \tag{10.43}$$

Функция одной переменной $F(t)$ разлагается по формуле Маклорена (см. параграф 5.3); полагая в этом разложении $t = 1$, получаем

$$F(1) = F(0) + \frac{F'(0)}{1!} + \frac{F''(0)}{2!} + \cdots + \frac{F^{(n)}(0)}{n!} + \frac{F^{(n+1)}(\theta)}{(n+1)!}, \quad 0 < \theta < 1. \tag{10.44}$$

Поскольку справедливы следующие равенства [в силу определения функции $F(t)$ и формулы (10.43)]:

$$F(1) = f(x + \Delta x, y + \Delta y) = f(M_1),$$
$$F(0) = f(x, y) = f(M),$$
$$F'(0) = \left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right) f(M) = df(M),$$

. . . . . .

$$F^{(n)}(0) = \left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^n f(M) = d^n f(M),$$

$$F^{(n+1)}(\theta) = \left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^{n+1} f(x+\theta\cdot\Delta x, y+\theta\cdot\Delta y) =$$

$$= d^{n+1} f(M'),$$

то из разложения (10.44) получаем

$$F(1) - F(0) = f(M_1) - f(M) = \Delta u = df(M) + \frac{d^2 f(M)}{2!} + \cdots +$$

$$+\frac{d^n f(M)}{n!} + \frac{d^{n+1} f(M')}{(n+1)!},$$

что и требовалось доказать.

10.3. Локальный экстремум функции нескольких переменных

Определение и необходимые условия локального экстремума

Пусть функция $u = f(M)$ определена на множестве $\{M\}$, а $M_0\ (x_1^0, x_2^0, \ldots, x_m^0)$ — некоторая точка этого множества.

Определение 1. Функция $u = f(M)$ имеет в точке M_0 *локальный максимум (минимум)*, если существует такая окрестность точки M_0, принадлежащая множеству $\{M\}$, что для любой точки $M\ (x_1, x_2, \ldots, x_m)$ из этой окрестности выполняется неравенство $f(M) \leqslant f(M_0)$ $(f(M) \geqslant f(M_0))$.

Согласно данному определению локального экстремума (минимума или максимума), полное приращение функции $\Delta u = f(M) - f(M_0)$ удовлетворяет одному из условий в окрестности точки M_0:

$\Delta u \leqslant 0$, если M_0 — точка локального максимума,

$\Delta u \geqslant 0$, если M_0 — точка локального минимума.

Теперь установим необходимые условия локального экстремума.

Теорема 10.9. *Если функция $u = f(x_1, x_2, \ldots, x_m)$ имеет в точке $M_0(x_1^0, x_2^0, \ldots, x_m^0)$ локальный экстремум и частные производные первого порядка, то все эти частные производные равны нулю в точке M_0:*

$$\left.\frac{\partial u}{\partial x_i}\right|_{M_0} = 0, \quad i = 1, 2, \ldots, m. \tag{10.45}$$

Доказательство. Зафиксируем все значения переменных x_i, кроме переменной x_k, положив их равными x_i^0. Тогда получим, что функция $u = f(x_1^0, x_2^0, \ldots, x_{k-1}^0, x_k^0, x_{k+1}^0, \ldots, x_m^0)$ зависит от одной переменной x_k и в точке $x_k = x_k^0$ имеет локальный экстремум. Но тогда производная этой функции, которая и является частной производной $\dfrac{\partial u}{\partial x_k}$ в точке M_0, равна нулю, что и требовалось доказать.

Заметим, что необходимые условия локального экстремума (10.45) определяют, что в точке локального экстремума

$$\left. du \right|_{M_0} = 0. \tag{10.46}$$

Обратное утверждение также верно: если в точке M_0 первый дифференциал функции тождественно равен нулю (как функция относительно dx_i), то все частные производные этой функции в указанной точке равны нулю в силу произвольности dx_i.

Следует особо отметить, что условия (10.45) или (10.46) не являются достаточными условиями экстремума. Например, для функции $z = x^2 - y^2$ частные производные равны нулю в точке $O(0,0)$, однако в этой точке функция, которая является гиперболическим параболоидом, не имеет экстремума: $f(0,0) = 0$, но в любой окрестности точки O есть значения функции как положительные, так и отрицательные (см. рис. 9.16).

Точки, в которых выполняются условия (10.45), называются *точками возможного экстремума*, или *стационарными*.

Пример 1. Найти точки возможного экстремума функции $u = x^2 + y^2 + xy - 4x - 5y$.

Согласно условиям (10.45), $\dfrac{\partial u}{\partial x} = 0$ и $\dfrac{\partial u}{\partial y} = 0$, откуда получаем систему двух алгебраических уравнений с двумя неизвестными:

$$\begin{cases} 2x + y = 4 \\ 2y + x = 5. \end{cases}$$

Решение этой системы: $x = 1$, $y = 2$, т.е. точка с координатами $(1, 2)$ является стационарной для данной функции двух переменных.

Пример 2. Найти точки возможного экстремума функции $u = x^2 + 2x + y^2 + 2xy + z^2 + zy$.

Согласно необходимым условиям экстремума, все три первые частные производные функции равны в этой точке нулю, откуда получаем систему трех линейных алгебраических уравнений с тремя неизвестными

$$\begin{cases} x + y = -1 \\ 2x + 2y + z = 0 \\ y + 2z = 0. \end{cases}$$

Решение этой системы дает единственную стационарную точку возможного экстремума: $(3, -4, 2)$.

Некоторые сведения о квадратичных формах

Функция

$$\Phi(x_1, x_2, \ldots, x_m; y_1, y_2, \ldots, y_m) = \sum_{i,k=1}^{m} a_{ik} x_i y_k, \qquad (10.47)$$

зависящая от переменных x_i и y_k, называется *билинейной формой* от этих переменных; коэффициенты формы a_{ik} не зависят от x_i и y_k. Это название объясняется тем, что при фиксированных значениях переменных x_i функция Φ будет линейной относительно переменных y_k и наоборот.

Функция

$$\Phi(x_1, x_2, \ldots, x_m; x_1, x_2, \ldots, x_m) = \sum_{i,k=1}^{m} a_{ik} x_i x_k \qquad (10.48)$$

называется *квадратичной формой*, соответствующей билинейной форме (10.47). Если $a_{ik} = a_{ki}$, то билинейная и соответствующая ей квадратичная формы называются *симметричными*.

Квадратичная симметричная форма

$$\Phi = \sum_{i,k=1}^{m} a_{ik} x_i x_k, \quad a_{ik} = a_{ki}, \qquad (10.49)$$

называется *положительно определенной (отрицательно определенной)*, если для любых значений переменных $x_1, x_2, \ldots, x_m$, не равных одновременно нулю, указанная форма имеет положительные (отрицательные) значения. Оба этих случая объединены под названием *знакоопределенных форм*. Если квадратичная форма (10.49) имеет как положительные, так и отрицательные значения, то она называется *знакопеременной*.

Сформулируем *критерий знакоопределенности квадратичной формы*. Будем называть симметрическую матрицу

$$A = \| a_{ik} \| = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mm} \end{pmatrix}, \quad a_{ik} = a_{ki}, \qquad (10.50)$$

матрицей квадратичной формы (10.49). Определители

$$\Delta_1 = a_{11}, \ \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \ \Delta_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix},$$
$$\ldots, \ \Delta_m = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mm} \end{vmatrix} \qquad (10.51)$$

называются *главными минорами* матрицы A квадратичной формы (10.49).

Критерий Сильвестра[1]. Для того чтобы квадратичная форма (10.49) была положительно определенной, необходимо и достаточно, чтобы выполнялись условия

$$\Delta_1 > 0, \Delta_2 > 0, \ldots, \Delta_m > 0. \qquad (10.52)$$

Для того чтобы квадратичная форма (10.49) была отрицательно определенной, необходимо и достаточно, чтобы знаки главных миноров $\Delta_1, \Delta_2, \ldots, \Delta_m$ чередовались, причем $\Delta_1 < 0$.

[1] Джеймс Джозеф Сильвестр (1814–1897) — английский математик.

Достаточные условия экстремума

Нетрудно видеть, что второй дифференциал функции нескольких переменных $u = f(x_1, x_2, \ldots, x_m)$

$$d^2u = \left(\frac{\partial}{\partial x_1}dx_1 + \frac{\partial}{\partial x_2}dx_2 + \cdots + \frac{\partial}{\partial x_m}dx_m\right)^2 u$$

представляет собой симметричную квадратичную форму относительно дифференциалов независимых переменных $dx_1, dx_2, \ldots, dx_m$. Матрица этой квадратичной формы, элементы которой являются вторыми частными производными функции $f(x_1, x_2, \ldots, x_m)$, т.е. $a_{ik} = \dfrac{\partial^2 f}{\partial x_i \partial x_k}$, называется *матрицей Гессе*

$$\boldsymbol{H} = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_m} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_m} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial^2 f}{\partial x_m \partial x_1} & \dfrac{\partial^2 f}{\partial x_m \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_m^2} \end{pmatrix}. \qquad (10.53)$$

Определитель этой матрицы, или минор, Δ_m называется *гессианом*.

Понятия знакоопределенной и знакопеременной квадратичной формы используются в достаточных признаках локального экстремума функции нескольких переменных.

Теорема 10.10. *Пусть в точке M_0 $(x_1^0, x_2^0, \ldots, x_m^0)$ возможного экстремума и в некоторой ее окрестности функция $u = f(x_1, x_2, \ldots, x_m)$ имеет все непрерывные вторые частные производные. Тогда, если в этой точке второй дифференциал d^2u является знакоопределенной квадратичной формой от дифференциалов $dx_1, dx_2, \ldots, dx_m$ независимых переменных, данная функция имеет в точке M_0 локальный экстремум. При этом:*

а) *если $d^2u > 0$, то это локальный минимум;*

б) *если $d^2u < 0$, то это локальный максимум.*

Если же d^2u является знакопеременной квадратичной формой в точке M_0, то функция $u = f(x_1, x_2, \ldots, x_m)$ не имеет локального экстремума.

Доказательство. Разложим функцию $u = f(M)$ по формуле Тейлора (10.42) при $n = 1$. Тогда полное приращение функции в точке M_0, поскольку в точке локального экстремума первый дифференциал равен нулю, имеет вид

$$\Delta u = \frac{1}{2}d^2u\Big|_{M'}, \qquad (10.54)$$

где M' — точка из окрестности точки M_0.

Как уже говорилось, второй дифференциал представляет собой квадратичную форму от приращений независимых аргументов Δx_i; с учетом формулы (10.53) равенство (10.54) можно переписать в виде

$$\Delta u = \frac{1}{2}\sum_{i,k=1}^{m} \frac{\partial^2 u}{\partial x_1 \partial x_k}\Big|_{M'} \Delta x_i \Delta x_k. \qquad (10.55)$$

Поскольку все вторые частные производные в точке M_0 (обозначим их $a_{ik} = \dfrac{\partial^2 u}{\partial x_i \partial x_k}$) являются непрерывными функциями в этой точке, то

$$\frac{\partial^2 u}{\partial x_i \partial x_k}\Big|_{M'} = a_{ik} + \alpha_{ik}, \qquad (10.56)$$

где α_{ik} — бесконечно малые при $\Delta x_1 \to 0, \Delta x_2 \to 0, \ldots, \Delta x_m \to 0$. Следовательно, равенство (10.55) запишется в виде

$$\Delta u = \frac{1}{2}\left(\sum_{i,k=1}^{m} a_{ik}\Delta x_i \Delta x_k + \sum_{i,k=1}^{m} \alpha_{ik}\Delta x_i \Delta x_k\right) = \frac{1}{2}d^2u\Big|_{M_0} + \frac{1}{2}\sum_{i,k=1}^{m} \alpha_{ik}\Delta x_i \Delta x_k.$$

В правой части этого равенства вторая сумма в силу (10.56) является бесконечно малой по отношению к первой сумме. Следовательно, мы можем выбрать столь малую окрестность точки M_0, что знак Δu будет определяться знаком d^2u в этой точке, или знаком квадратичной формы

$$\Phi = \sum_{i,k=1}^{m} a_{ik}dx_i dx_k, \quad a_{ik} = \frac{\partial^2 u}{\partial x_i \partial x_k}\Big|_{M_0}. \qquad (10.57)$$

Отсюда получаем утверждения "а" и "б" теоремы относительно вида локального экстремума функции $u = f(M)$ в точке M_0: если $d^2u > 0$, то это минимум, если $d^2u < 0$, то это максимум.

В случае если d^2u в точке M_0 представляет собой знакопеременную квадратичную форму, всегда можно указать достаточно малую окрестность точки M_0, в которой приращение Δu будет принимать как положительные значения, так и отрицательные. Иными словами, функция $u = f(M)$ не имеет локального экстремума в этой точке.

Теорема доказана полностью.

Рассмотрим функцию двух переменных $u = f(x, y)$, часто используемую на практике. Обозначим вторые частные производные этой функции $\dfrac{\partial^2 u}{\partial x^2}$, $\dfrac{\partial^2 u}{\partial x \partial y}$, $\dfrac{\partial^2 u}{\partial y^2}$ в некоторой точке M_0 через a_{11}, a_{12}, a_{22} соответственно. Тогда, с учетом критерия Сильвестра при $m = 2$, теорема 10.10 формулируется следующим образом.

Пусть в точке $M_0(x_0, y_0)$ возможного экстремума функции $u = f(x, y)$ и в некоторой ее окрестности все вторые частные производные этой функции непрерывны. Тогда, если $a_{11}a_{22} - a_{12}^2 > 0$, то функция $u = f(x, y)$ имеет в точке M_0 локальный экстремум: минимум при $a_{11} > 0$ и максимум при $a_{11} < 0$. Если же $a_{11}a_{22} - a_{12}^2 < 0$, то данная функция не имеет локального экстремума в точке M_0.

Нетрудно увидеть, что выражение $a_{11}a_{22} - a_{12}^2$ является гессианом функции двух переменных.

Пример 3. Найти точки локального экстремума и значения в них функции $u = x^3 - y^3 - 3xy$.

Сначала находим стационарную точку из условий $\dfrac{\partial u}{\partial x} = \dfrac{\partial u}{\partial y} = 0$. Получаем систему двух алгебраических уравнений с двумя неизвестными

$$\begin{cases} x^2 - y = 0 \\ x + y^2 = 0, \end{cases}$$

решения которой дают координаты двух точек $(0, 0)$ и $(-1, 1)$. Теперь составим гессиан данной функции, для чего найдем вторые производные

$$a_{11} = \frac{\partial^2 u}{\partial x^2} = 6x,\ a_{12} = \frac{\partial^2 u}{\partial x \partial y} = -3,\ a_{22} = \frac{\partial^2 u}{\partial y^2} = -6y,$$

откуда $a_{11}a_{22} - a_{12}^2 = -36xy - 9$. В точке $(0, 0)$ гессиан отрицателен, и, значит, в ней нет локального экстремума. В точке $(-1, 1)$ гессиан $a_{11}a_{22} - a_{12}^2 = 27 > 0$, т.е. в этой точке данная функция имеет локальный экстремум; поскольку $a_{11} < 0$, то это точка максимума. Значение функции в ней: $u_{\max} = f(-1, 1) = 1$.

Максимальное и минимальное значения функции в замкнутой области

При нахождении экстремума функции нескольких переменных, непрерывной на некотором замкнутом множестве, возникает далеко не простая задача отыскания наименьшего и наибольшего ее значений или глобального минимума и максимума. Дело в том, что эти значения могут достигаться на границе области задания функции, и эти случаи наиболее трудны для исследования при достаточно большом числе независимых переменных. Существуют специальные разделы математики и методы, позволяющие решать подобные задачи. Здесь мы остановимся лишь на случае функции двух переменных.

Пример 4. Найти максимум и минимум функции $u = xy$, определенной в круге $x^2 + y \leqslant 1$.

Сначала найдем стационарные точки; используя условие $\dfrac{\partial u}{\partial x} = \dfrac{\partial u}{\partial y} = 0$, получаем единственную точку $(0, 0)$. Однако в этой точке гессиан $a_{11}a_{22} - a_{12}^2 = -1$, т.е. экстремума в ней нет. Теперь исследуем поведение функции на границе области ее задания — на окружности $x^2 + y^2 = 1$, для чего выразим явно y через x и подставим в формулу для данной функции. Получаем, что на верхней и нижней полуокружностях границы области (рис. 10.4) функция $u = xy$ имеет соответственно выражения

$$u_1(x) = x\sqrt{1 - x^2},\ u_2(x) = -x\sqrt{1 - x^2}.$$

23 Зак. 6853

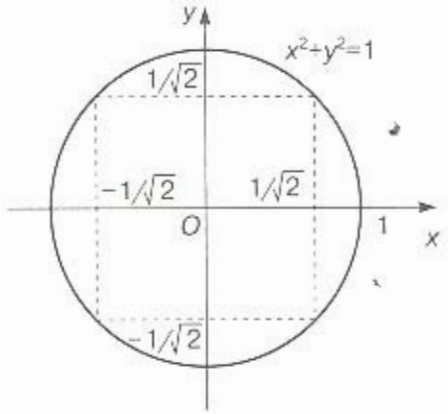

Рис. 10.4

Производные этих функций выражаются соответственно формулами

$$u_1' = \frac{1-2x^2}{\sqrt{1-x^2}}, \quad u_2' = -\frac{1-2x^2}{\sqrt{1-x^2}}.$$

Приравнивая их нулю, получаем критические точки $x = \pm 1/\sqrt{2}$. Нетрудно видеть, что при переходе через эти точки производные u_1' и u_2' меняют знак, т.е. функции $u_1(x)$ и $u_2(x)$ имеют экстремумы. Следовательно, функция $u = xy$ принимает в точках $(1/\sqrt{2}, 1/\sqrt{2})$ и $(-1/\sqrt{2}, -1/\sqrt{2})$ максимальные значения $u_{\max} = 1/2$, а в точках $(-1/\sqrt{2}, 1/\sqrt{2})$ и $(1/\sqrt{2}, -1/\sqrt{2})$ — минимальные значения $u_{\min} = -1/2$.

Понятие вектор-функции и операции над ней

Вектор-функция. Пусть на некотором множестве $\{M\}$ евклидова пространства E^m заданы n функций от m переменных

$$f_1(M), f_2(M), \ldots, f_n(M). \tag{10.58}$$

Каждую из этих функций можно представить как соответствующую координату некоторого n-мерного вектора. Тогда совокупность функций (10.58) определяет n-мерный функциональный вектор и мы будем говорить, что на множестве $\{M\}$ задана *вектор-функция*

$$\bar{f}(M) = (f_1(M), f_2(M), \ldots, f_n(M)).$$

Для удобства совокупность m координат точек M евклидова пространства E^m также представляют в виде вектор-аргумента $\bar{x}$:

$$\bar{x} = (x_1, x_2, \ldots, x_m). \tag{10.59}$$

Тогда, если заданы n функций (10.58), будем говорить, что задана *n-мерная вектор-функция $\bar{f}$ от m-мерного вектор-аргумента $\bar{x}$*:

$$\bar{f}(\bar{x}) = (f_1(\bar{x}), f_2(\bar{x}), \ldots, f_n(\bar{x})). \tag{10.60}$$

Производная вектор-функции. Пусть все функции набора (10.58) дифференцируемы в некоторой области $\{M\} \subset E^m$; тогда будем говорить, что в этой области задана *дифференцируемая вектор-функция* $\bar{f}(\bar{x})$.

Определим операцию дифференцирования вектор-функции (10.60) по вектор-аргументу. *Производной вектор-функции $\bar{f}$ по вектор-аргументу $\bar{x}$* будем называть матрицу, составленную из частных производных функций (10.58); будем обозначать производную вектор-функции следующими символами:

$$\frac{\partial \bar{f}}{\partial \bar{x}} = \frac{\partial(f_1, f_2, \ldots, f_n)}{\partial(x_1, x_2, \ldots, x_m)} = \\ = A_x = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_m} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_m} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial f_n}{\partial x_1} & \dfrac{\partial f_n}{\partial x_2} & \cdots & \dfrac{\partial f_n}{\partial x_m} \end{pmatrix}. \tag{10.61}$$

Матрица вида (10.61) размерностью $n \times m$ называется *матрицей Якоби*[1]. В случае квадратной матрицы, т.е. когда $m = n$, определитель матрицы Якоби называется *якобианом*. Сокращенная форма его записи (при сокращенной форме (10.50) записи матрицы) имеет вид

$$J = \det \left\| \frac{\partial(f_1, f_2, \ldots, f_m)}{\partial(x_1, x_2, \ldots, x_m)} \right\| = \det \left\| \frac{\partial f_i}{\partial x_k} \right\|; \\ i, k = 1, 2, \ldots, m. \tag{10.62}$$

[1] Карл Густав Якоб Якоби (1804–1851) — немецкий математик.

Дифференцирование сложной вектор-функции. Пусть координаты вектор-аргумента $\bar{x}$, в свою очередь, являются дифференцируемыми функциями вектор-аргумента $\bar{t}$:

$$\bar{x} = \bar{\varphi}(\bar{t}) = (\varphi_1(\bar{t}), \varphi_2(\bar{t}), \ldots, \varphi_m(\bar{t})), \quad \bar{t} = (t_1, t_2, \ldots, t_s). \tag{10.63}$$

Тогда, в силу теоремы 10.4 (см. параграф 10.1), производная вектор-функции $\bar{u} = \bar{f}(\bar{x})$ по вектор-аргументу $\bar{t}$ имеет вид

$$\frac{\partial \bar{u}}{\partial \bar{t}} = \frac{\partial \bar{f}}{\partial \bar{x}} \frac{\partial \bar{\varphi}}{\partial \bar{t}}. \tag{10.64}$$

Равенство (10.64) имеет вид, аналогичный производной сложной функции одной переменной, однако здесь его следует понимать в смысле *произведения матриц* Якоби, имеющих размерности соответственно $n \times m$ и $m \times s$. Следовательно, произведение этих матриц является матрицей размерности $n \times s$:

$$\frac{\partial \bar{u}}{\partial \bar{t}} = \frac{\partial(f_1, f_2, \ldots, f_n)}{\partial(x_1, x_2, \ldots, x_m)} \frac{\partial(x_1, x_2, \ldots, x_m)}{\partial(t_1, t_2, \ldots, t_s)} = A_x A_t = \left\| \frac{\partial f_i}{\partial x_j} \right\| \left\| \frac{\partial x_j}{\partial t_k} \right\|, \tag{10.65}$$

$$i = 1, 2, \ldots, n; \; j = 1, 2, \ldots, m; \; k = 1, 2, \ldots, s.$$

Рассмотрим примеры составления матриц Якоби для следующих вектор-функций.

Пример 5. $\bar{f}(\bar{x}) = (xy, y^2, x + y)$.

В данном случае мы имеем вектор-функцию трех координат, каждая из которых зависит от двух независимых переменных x и y. Применяя формулу (10.61), получаем

$$\frac{\partial \bar{f}}{\partial \bar{x}} = \frac{\partial(xy, y^2, x + y)}{\partial(x, y)} = A_x = \begin{pmatrix} y & x \\ 0 & 2y \\ 1 & 1 \end{pmatrix}.$$

Пример 6. $\bar{u} = \bar{f}(\bar{x}) = (xy, y^2, x+y)$, $\bar{x} = \bar{\varphi}(\bar{t}) = (2t_1 + t_2, t_1 t_2)$.

Здесь мы имеем сложную вектор-функцию, где зависимые переменные x и y являются функциями независимых переменных t_1 и t_2. Применяя формулу дифференцирования (10.65), получаем

$$\frac{\partial \bar{u}}{\partial \bar{t}} = \frac{\partial(xy, y^2, x + y)}{\partial(x, y)} \frac{\partial(2t_1 + t_2, t_1 t_2)}{\partial(t_1, t_2)} = A_x A_t = \begin{pmatrix} y & x \\ 0 & 2y \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ t_2 & t_1 \end{pmatrix} = \begin{pmatrix} 4t_1 t_2 + t_2^2 & 2t_1 t_2 + 2t_1^2 \\ 2t_1 t_2^2 & 2t_1^2 t_2 \\ 2 + t_2 & 1 + t_1 \end{pmatrix}.$$

Понятие о выпуклых множествах

Для упрощения мы не будем в дальнейшем делать различие между точками евклидова пространства E^m и векторами их координат. Тогда на точки пространства E^m переносятся все операции над векторами и их свойства. Пусть $\bar{x}$ и $\bar{y}$ — две точки евклидова пространства E^m.

Выпуклое множество евклидова пространства. Введем следующие определения.

Определение 2. *Отрезком* $\overline{xy}$, соединяющим точки $\bar{x}$ и $\bar{y}$ евклидова пространства E^m, называется множество точек $\bar{x} + t(\bar{y} - \bar{x})$ этого пространства, где $t \in [0, 1]$.

Определение 3. Множество R точек евклидова пространства называется *выпуклым*, если вместе с его любыми точками $\bar{x}$ и $\bar{y}$ все точки отрезка $\overline{xy}$ также принадлежат этому множеству.

Примеры двумерных выпуклых множеств показаны на рис. 10.5, *а*; множества, показанные на рис. 10.5, *б*, являются невыпуклыми.

Выпуклая функция. Понятие выпуклости можно распространить и на случай функции нескольких переменных.

Определение 4. Функция $f(\bar{x})$, заданная на выпуклом множестве R, называется *выпуклой вниз* на этом множестве, если для любых двух точек $\bar{x}$ и $\bar{y}$ из R выполняется неравенство

$$f(\bar{x} + t(\bar{y} - \bar{x})) \leqslant f(\bar{x}) + t(f(\bar{y}) - f(\bar{x})), \quad t \in [0, 1]. \tag{10.66}$$

Понятие функции, *выпуклой вверх*, определяется аналогично, с переменой знака неравенства в (10.66) на противоположный. Если в соотношении (10.66) для всех точек $\bar{x}$ и $\bar{y}$ из множества R при всех $t \in [0, 1]$ имеет место знак строгого не-

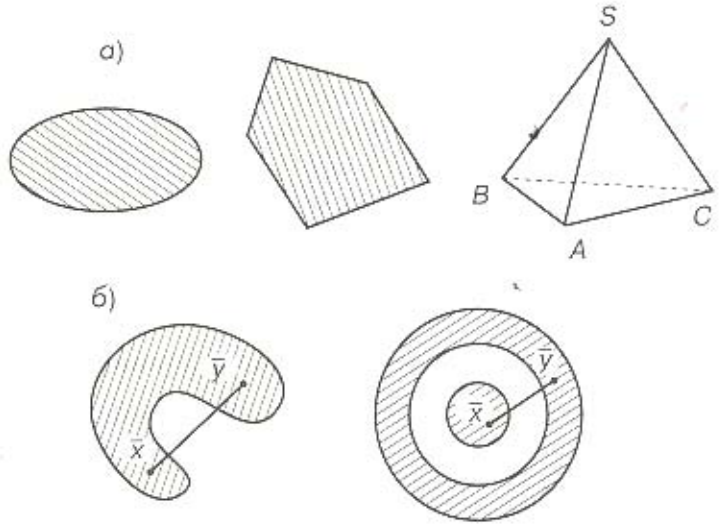

Рис. 10.5

равенства, то функцию $f(\bar{x})$ называют *строго выпуклой вниз* (*вверх*) на множестве R.

Геометрический смысл выпуклой вниз функции двух переменных показан на рис. 10.6. Все точки поверхности, описываемой функцией $f(\bar{x})$, соответствующие отрезку $\overline{xy}$ множества R, лежат не выше хорды, соединяющей точки $(\bar{x}, f(\bar{x}))$ и $(\bar{y}, f(\bar{y}))$.

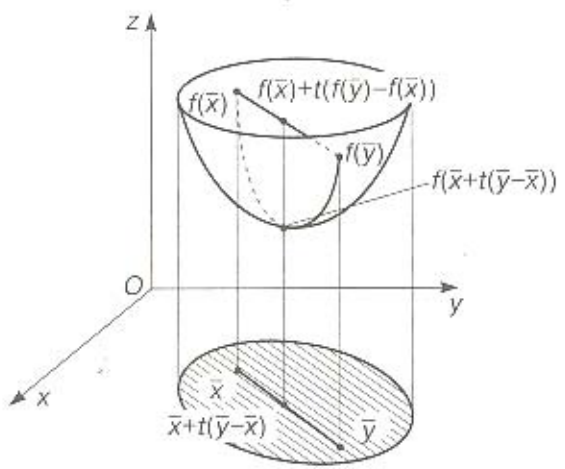

Рис. 10.6

Достаточный признак выпуклости функции. Приведем без доказательства следующую теорему.

Теорема 10.11. *Пусть функция $f(\bar{x})$ задана и дважды дифференцируема на выпуклом множестве R. Если во всех точках этого множества второй дифференциал d^2f является положительно определенной формой, то $f(\bar{x})$ — выпуклая вниз функция на множестве R.*

Важным моментом в поиске минимума выпуклой вниз функции, что является одним из ключевых вопросов в прикладных задачах, представляется теорема, формулировка которой приводится ниже.

Теорема 10.12. *Если функция $f(\bar{x})$ дифференцируема и строго выпукла на выпуклом множестве R, то она имеет локальный экстремум лишь в одной точке этого множества.*

10.4. Функции нескольких переменных в задачах экономики

Экстремум функции нескольких переменных

Рассмотрим некоторые типичные задачи нахождения экстремума функции нескольких переменных, возникающие в экономике.

Прибыль от производства товаров разных видов. Пусть $x_1, x_2, \ldots, x_m$ — количество производимых m разновидностей товара, а их цены — соответственно $P_1, P_2, \ldots, P_m$ (все P_i — постоянные величины). Пусть затраты на производство этих товаров задаются функцией издержек

$$C = S(x_1, x_2, \ldots, x_m).$$

Тогда функция прибыли имеет вид

$$\Pi = P_1x_1 + P_2x_2 + \cdots + P_mx_m - S(x_1, x_2, \ldots, x_m). \qquad (10.67)$$

Максимум прибыли естественно искать как условие локального экстремума функции многих переменных (10.67) при $x_i \geqslant 0$ (при отсутствии других ограничений):

$$\frac{\partial \Pi}{\partial x_i} = 0, \ i = 1, 2, \ldots, m.$$

Это условие приводит к системе алгебраических уравнений относительно переменных x_i

$$P_i - \frac{\partial S}{\partial x_i} = 0,\ i = 1, 2, \ldots, m. \tag{10.68}$$

Система уравнений (10.68) реализует известное правило экономики: предельная стоимость (цена) товара равна предельным издержкам на производство этого товара. Решениями этой системы уравнений являются наборы, состоящие из m значений каждый. Нужно заметить, что сам процесс нахождения решения системы уравнений (10.68) зависит от вида функции издержек и может быть достаточно сложным.

Найденные наборы значений $(x_1^0, x_2^0, \ldots, x_m^0)$ необходимо проверить по достаточному признаку локального максимума (см. теорему 10.10 и признак Сильвестра для отрицательно определенной квадратичной формы). Иными словами, нужно выполнить следующие операции:

а) составить гессиан для функции вида (10.67):

$$\boldsymbol{H}_m = \begin{vmatrix} -S''_{x_1x_1} & -S''_{x_1x_2} & \ldots & -S''_{x_1x_m} \\ -S''_{x_2x_1} & -S''_{x_2x_2} & \ldots & -S''_{x_2x_m} \\ \ldots & \ldots & \ldots & \ldots \\ -S''_{x_mx_1} & -S''_{x_mx_2} & \ldots & -S''_{x_mx_m} \end{vmatrix}. \tag{10.69}$$

Если знаки главных миноров этого определителя чередуются, начиная со знака "минус", то проверяемое решение $(x_1^0, x_2^0, \ldots, x_m^0)$ системы уравнений (10.68) является m-мерной точкой локального максимума функции прибыли (10.67);

б) далее, необходимо выбрать из этих точек локальных максимумов ту, в которой функция (10.67) принимает наибольшее значение. Найденные таким образом значения $(x_1^*, x_2^*, \ldots, x_m^*)$ определяют набор количества товаров x_i, которые необходимо выпускать, чтобы при имеющейся функции затрат на их производство и спектре сложившихся цен обеспечить максимальную прибыль.

Пример 1. Пусть производится два вида товаров, обозначим их количества через x и y. Пусть цены на эти товары соответственно $P_1 = 8$ и $P_2 = 10$, а функция затрат $C = x^2 + xy + y^2$. Найти максимум прибыли. Тогда, согласно формуле (10.67),

прибыль выражается функцией

$$\Pi(x, y) = 8x + 10y - x^2 - xy - y^2.$$

Условия локального экстремума приводят к системе линейных алгебраических уравнений

$$\begin{cases} 2x + y = 8 \\ x + 2y = 10, \end{cases}$$

решение которой определяет точку $(2, 4)$. Поскольку

$$a_{11} = -2 < 0,\ \boldsymbol{H}_2 = a_{11}a_{22} - a_{12}^2 = 3 > 0,$$

то найденная точка определяет локальный максимум функции прибыли, который равен $\Pi_{\max} = 28$.

Задача ценовой дискриминации. Эта задача связана с распределением товара одного вида по разным рынкам с разными спросами с тем, чтобы максимизировать общую прибыль. Поскольку эластичность спроса на рынках неодинакова, на товар устанавливаются разные цены, что ведет к так называемой ценовой дискриминации.

Сформулируем общую задачу. Пусть $x_1, x_2, \ldots, x_m$ — количество одного и того же вида товара, продаваемого на m рынках по ценам $P_i(x_i)$, т.е. цена на каждом рынке зависит от количества продаваемого товара. Допустим, что функция затрат зависит от общего количества продаваемого товара, т.е. $C = S(x_1 + x_2 + \cdots + x_m)$. Тогда функция общей прибыли имеет вид

$$\begin{aligned} \Pi = x_1P_1(x_1) + x_2P_2(x_2) + \cdots + x_mP_m(x_m) - \\ -S(x_1 + x_2 + \cdots + x_m). \end{aligned} \tag{10.70}$$

Необходимое условие экстремума

$$\frac{\partial \Pi}{\partial x_i} = 0,\ i = 1, 2, \ldots, m,$$

приводит к системе алгебраических уравнений для определения стационарных точек функции (10.70) в области $x_1 \geqslant 0$, $x_2 \geqslant 0, \ldots, x_m \geqslant 0$ m-мерного евклидова пространства E^m:

$$\begin{aligned} P_i(x_i) + x_iP'(x_i) - S'(x_1 + x_2 + \cdots + x_m) = 0, \\ i = 1, 2, \ldots, m. \end{aligned} \tag{10.71}$$

Проанализируем доход R_i на каждом рынке в зависимости от цены на товар x_i. Так как $R_i = x_i P_i$, предельный доход выражается суммой первых двух слагаемых в левой части (10.71)

$$R_i' = P_i + x_i P_i' = P_i(1 + P_i' x_i / P_i) = P_i(1 + 1/E_i),$$

где E_i — эластичность спроса на i-м рынке. Так как E_i — обычно отрицательная величина, последнее равенство можно переписать в более удобной форме:

$$R_i' = P_i\,(1 - 1/|E_i|)\,. \tag{10.72}$$

Если $|E_i| < 1$, то $R_i' < 0$, т.е. рынок неэластичный. Если $S' > 0$, то условие (10.71) требует выбора рынка с положительным предельным доходом, или с эластичным спросом, т.е. с условием $|E_i| > 1$. Из уравнений (10.71)

$$P_1(1 - 1/|E_1|) = P_2(1 - 1/|E_2|) = \cdots = P_m(1 - 1/|E_m|) = S',$$

откуда и выводится условие "ценовой дискриминации": чем меньше по абсолютной величине эластичность данного рынка при данном количестве продаваемого товара, тем выше должна быть цена на товар на этом рынке при условии максимизации прибыли.

Гессиан для функции прибыли (10.70) в сокращенной форме записи имеет вид

$$\boldsymbol{H}_m = \det \|h_{ij}\|,\ h_{ij} = \begin{cases} P_i'' x_i + 2P_i' - S'', & i = j, \\ -S'', & i \neq j, \end{cases} \tag{10.73}$$
$$i, j = 1, 2, \ldots, m.$$

Пример 2. Пусть имеется три рынка с количеством продаваемого товара x_1, x_2 и x_3 и с ценами на этот товар соответственно $P_i = a_i - b_i x_i$, так что доход $R_i = x_i(a_i - b_i x_i)$. Пусть функция затрат имеет вид $C = A + B(x_1 + x_2 + x_3)$. Найти максимум прибыли.

В указанных соотношениях все постоянные величины положительны. Тогда функция прибыли выражается формулой

$$\Pi = x_1(a_1 - b_1 x_1) + x_2(a_2 - b_2 x_2) + x_3(a_3 - b_3 x_3) - \\ - A - B(x_1 + x_2 + x_3).$$

Условие локального экстремума приводит к системе уравнений

$$a_i - 2b_i x_i - B = 0,\ i = 1, 2, 3.$$

откуда находим единственную стационарную точку, определяемую тремя равенствами:

$$x_i = \frac{a_i - B}{2b_i},\ i = 1, 2, 3.$$

Гессиан функции прибыли, согласно (10.73), имеет вид

$$\boldsymbol{H}_3 = \begin{vmatrix} -2b_1 & 0 & 0 \\ 0 & -2b_2 & 0 \\ 0 & 0 & -2b_3 \end{vmatrix},$$

откуда при $b_i > 0$ получаем по критерию Сильвестра точку максимума, так как знаки главных миноров чередуются, начиная со знака "минус".

Для конкретизации возьмем числа $a_1 = 25$, $b_1 = 5$; $a_2 = 45$, $b_2 = 4$; $a_3 = 85$, $b_3 = 10$; $A = 10$, $B = 5$. Тогда получаем распределение товара между рынками $x_1 = 2$, $x_2 = 5$, $x_3 = 4$ при ценах на них соответственно $P_1 = 15$, $P_2 = 25$, $P_3 = 45$. Нетрудно подсчитать, что соответствующие эластичности $|E_1| = 1{,}5$, $|E_2| = 1{,}25$, $|E_3| = 1{,}125$ удовлетворяют принципу "ценовой дискриминации": чем меньше величина $|E_i|$ i-го рынка, тем выше должна быть цена товара на нем. Максимальная прибыль $\Pi_{\max} = 270$.

Оптимальное распределение ресурсов

Рассмотрим типичную задачу оптимального распределения ресурсов на примере функции выпуска $u = a_0 x y^2$ при допущении, что функция затрат на ресурсы x и y линейна, т.е. имеет вид $u = P_1 x + P_2 y$, где P_1 и P_2 — соответствующие цены на эти факторы.

В точке (x_0, y_0) оптимального определения ресурсов линии уровня функций выпуска и затрат касаются (рис. 10.7). Эти линии определяются соответственно уравнениями:

$$a_0 x y^2 = C,\ P_1 x + P_2 y = A,\ \text{или } y = (b/x)^{1/2},$$
$$y = -(P_1/P_2)x + A/P_2,$$

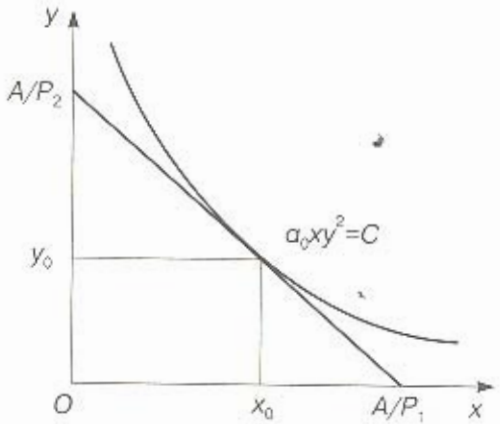

Рис. 10.7

где $C > 0$ и $A > 0$ — постоянные числа, $b = C/a_0$. Условие касания этих линий определяется уравнением

$$[(b/x)^{1/2}]'|_{x_0} = -P_1/P_2.$$

Из этого уравнения находится значение $x_0 = b^{1/3}(P_2/2P_1)^{2/3}$. Тогда из уравнения линии уровня функции выпуска определяется значение $y_0 = (b/x_0)^{1/2} = b^{1/3}(2P_1/P_2)^{1/3}$. Отсюда получаем, что оптимальное распределение ресурсов x_0/y_0 должно быть произведено в отношении $P_2 : 2P_1$.

Максимизация прибыли производства продукции

Функция прибыли обычно вычисляется по формуле

$$\Pi(K, L) = P \cdot F(K, L) - WL - RK, \qquad (10.74)$$

где P — цена продукции, $F(K, L)$ — производственная функция, W и R — соответственно факторные цены на труд и капитальные затраты, L и K — соответственно затраты трудовых ресурсов и капитала. Мы рассмотрим две задачи, связанные с определением максимума прибыли.

Оптимальный план. Точка (K_0, L_0) называется *оптимальным планом*, если в ней функция прибыли (10.74) принимает максимальное значение. Найти предельную норму замещения производственной функции F при оптимальном плане.

В точке локального экстремума первые производные функции прибыли $\Pi(K, L)$ равны нулю, откуда имеем систему двух уравнений:

$$P \cdot F'_K(K_0, L_0) - R = 0, \;\; P \cdot F'_L(K_0, L_0) - W = 0.$$

Как известно, предельная норма замещения вычисляется по формуле $\mu = -F'_L/F'_K$, откуда при оптимальном плане получаем $\mu = -W/R$.

Максимизация функции прибыли. Найти оптимальный план и максимум функции прибыли (10.74), если $F(K, L) = 2(KL)^{1/3}$.

Таким образом, функция прибыли в данном случае имеет вид.

$$\Pi(K, L) = 2P(KL)^{1/3} - WL - RK.$$

Условия локального экстремума приводят к системе двух линейных алгебраических уравнений относительно координат K_0 и L_0 оптимального плана:

$$\begin{cases} \dfrac{2}{3} P L_0^{1/3} K_0^{-2/3} = R \\ \dfrac{2}{3} P K_0^{1/3} L_0^{-2/3} = W. \end{cases}$$

Отсюда получаем координаты оптимального плана:

$$K_0 = (2P/3)^3/(R^2 W), \;\; L_0 = (2P/3)^3/(RW^2).$$

Подстановка этих величин в функцию прибыли дает ее максимум:

$$\Pi_{\max} = (2P/3)^3/(PW).$$

Оптимизация спроса

Типичной задачей исследования спроса является оптимизация функции полезности при ограничениях на бюджет покупателя. Рассмотрим такой пример: найти величины спроса x и y на две разновидности товара при ценах на них соответственно

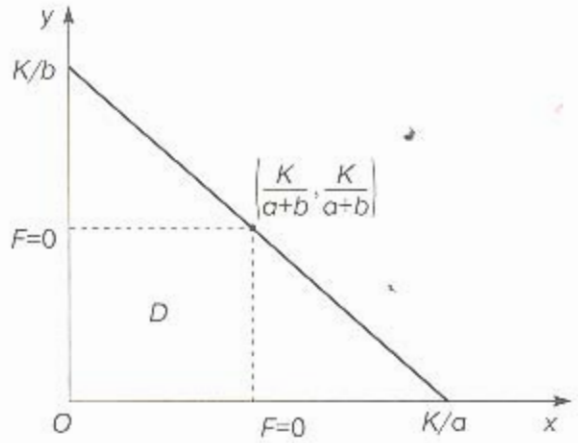

Рис. 10.8

a и b, если потребитель при бюджете K стремится максимизировать функцию полезности, которая имеет вид

$$F(x,y) = x^{\frac{a}{a+b+1}} y^{\frac{b}{a+b+1}}. \qquad (10.75)$$

Из условия задачи следует, что на покупку, стоимость которой $ax + by$, потребитель может израсходовать сумму, не превышающую K. Следовательно, необходимо найти точку (x, y), в которой функция (10.75) достигает максимума при ограничениях:

$$ax + by \leqslant K,\ x \geqslant 0,\ y \geqslant 0. \qquad (10.76)$$

Ограничения (10.76) задают на плоскости Oxy замкнутую область D в виде треугольника (рис. 10.8), в которой следует искать точку максимума функции (10.76).

Вычислив частные производные функции полезности (10.75), находим, что единственная критическая точка $(0,0)$ находится на границе области D; и в ней, как и на граничных линиях $x = 0$ и $y = 0$, функция $F(x,y) = 0$, что является минимальным ее значением. Следовательно, нужно искать точку максимума этой функции на границе

$$ax + by = K$$

области D. Подставляя из этого уравнения выражение для y в функцию (10.75), получаем функцию одной переменной

$$f(x) = b^{-\frac{b}{a+b+1}} x^{\frac{a}{a+b+1}} (K - ax)^{\frac{b}{a+b+1}}.$$

Критическую точку этой функции найдем из условия обращения в нуль ее первой производной; получаем, что $x = K/(a+b)$, откуда $y = K/(a+b)$. Таким образом, в данной модели оптимальный спрос на оба вида товара одинаков: он пропорционален бюджету и обратно пропорционален суммарной цене товара.

Метод наименьших квадратов

Метод наименьших квадратов относится к так называемым методам аппроксимации, или приближенного восстановления функции по известным ее значениям в ряде точек.

На практике часто возникает задача о наилучшем подборе так называемых *эмпирических формул*, которые позволяют аналитически представить данные измерений, статистической обработки экспериментов, наблюдений и т.д. Обычно задача формулируется следующим образом: имеются данные наблюдений в n точках $M_1, M_2, \ldots, M_n$ определенной величины u, и получены соответствующие значения этой величины $u_1, u_2, \ldots, u_n$; нужно подобрать такую функцию $u = f(M)$, чтобы она по возможности наиболее точно отражала общую зависимость измеряемой величины u от параметров точек измерений $\{M_i\}$.

Задача нахождения эмпирических формул состоит из двух этапов:

на первом определяется общий вид зависимости $f(M)$, т.е. по ряду наблюдений $\{u_i\}$ устанавливается вид, который может иметь функция $u = f(M)$, с точностью до постоянных коэффициентов или неизвестных параметров, входящих в нее;

на втором этапе эти неизвестные параметры подбираются таким образом, чтобы в точках наблюдений $\{M_i\}$ подобранная функция наилучшим способом отвечала данным измерений $\{u_i\}$.

Итак, пусть в результате первого этапа определено, что измерения в n точках наблюдения

$$M_1, M_2, \ldots, M_n, \tag{10.77}$$

давшие соответственно ряд данных

$$u_1, u_2, \ldots, u_n, \tag{10.78}$$

желательно приблизить совокупностью функций

$$\varphi_1(M), \varphi_2(M), \ldots, \varphi_m(M) \tag{10.79}$$

или эмпирической формулой вида

$$u = f(M) = a_1\varphi_1(M) + a_2\varphi_2(M) + \cdots + a_m\varphi_m(M), \tag{10.80}$$

где

$$a_1, a_2, \ldots, a_m \tag{10.81}$$

— неизвестные параметры функции $u = f(M)$. Следует особо подчеркнуть роль первого этапа подбора вида эмпирической функции, например можно подобрать функцию, которая точно проходит через измеренные значения $\{u_i\}$, однако ее колебания столь велики, что вряд ли можно говорить об удовлетворительном установлении функциональной зависимости (рис. 10.9).

Второй этап состоит как раз в определении неизвестных параметров (10.81). Их следует выбрать такими, чтобы значения эмпирической функции (10.80) по возможности давали

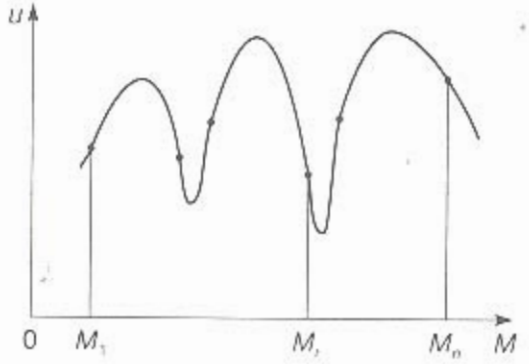

Рис. 10.9

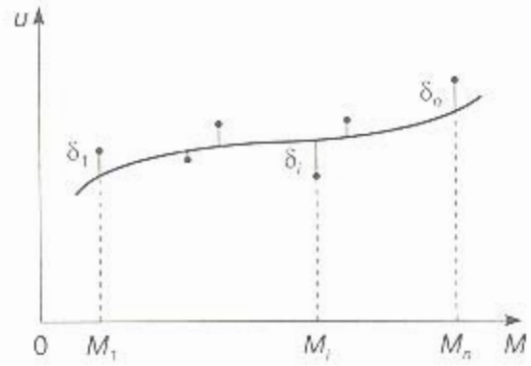

Рис. 10.10

бы в точках (10.77) минимальные отклонения от измеренных значений (10.78) (рис. 10.10).

Метод наименьших квадратов состоит в минимизации суммы квадратов *погрешностей* функции (10.80) в точках (10.77) как функции от m аргументов — неизвестных параметров (10.81):

$$S(a_1, a_2, \ldots, a_m) = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} [u_1 - f(M_i)]^2 = \\ = \sum_{i=1}^{n} \left[u_i - \sum_{k=1}^{m} a_k \varphi_k(M_i) \right]^2. \tag{10.82}$$

Для установления точки минимума функции (10.82) найдем частные производные этой функции и приравняем их к нулю:

$$\frac{\partial S}{\partial a_j} = -2 \sum_{i=1}^{n} \left[u_i - \sum_{k=1}^{m} a_k \varphi_k(M_i) \right] \varphi_j(M_i) = 0, \\ j = 1, 2, \ldots, m.$$

Переменив порядок суммирования, получаем отсюда систему, состоящую из m линейных алгебраических уравнений относительно m неизвестных параметров a_k:

$$A_{j1}a_1 + A_{j2}a_2 + \cdots + A_{jm}a_m = B_j, \; j = 1, 2, \ldots, m, \tag{10.83}$$

24 Зак. 6833

Коэффициенты и свободные члены уравнений этой системы определяются по формулам

$$A_{jk} = A_{kj} = \sum_{i=1}^{n} \varphi_j(M_i)\varphi_k(M_i),\ B_j = \sum_{i=1}^{n} u_i\varphi_j(M_i);$$
$$j,k = 1,2,\ldots,m. \tag{10.84}$$

Как правило, при случайном выборе точек наблюдения (10.77) система уравнений (10.83) с коэффициентами (10.84) имеет единственное решение, дающее набор значений для неизвестных параметров (10.81), соответствующий стационарной точке функции (10.82). Так как эта функция, состоящая из суммы квадратов *невязок* δ_i, является положительной, выпуклой вниз (см. параграф 10.3) и неограниченной в евклидовом пространстве E^m по аргументам (10.81), то найденная стационарная точка может быть только точкой ее локального минимума.

При обработке данных наблюдений наиболее распространенным является приближение в виде линейной функции одной переменной. В этом случае набор точек измерения (10.77) представляет собой набор значений аргумента $x_1, x_2, \ldots, x_n$, а набор функций (10.79) состоит из двух функций: x и 1. Эмпирическая формула (10.80) имеет вид

$$u = a_1 x + a_2, \tag{10.85}$$

где неизвестные параметры a_1 и a_2 определяются из системы двух линейных уравнений

$$\begin{cases} A_{11}a_1 + A_{12}a_2 = B_1, \\ A_{21}a_1 + A_{22}a_2 = B_2, \end{cases} \tag{10.86}$$

в которой коэффициенты и свободные члены выражаются формулами

$$A_{11} = \sum_{i=1}^{n} x_i^2,\ A_{12} = A_{21} = \sum_{i=1}^{n} x_i,$$
$$A_{22} = n,\ B_1 = \sum_{i=1}^{n} u_i x_i,\ B_2 = \sum_{i=1}^{n} u_i. \tag{10.87}$$

Степенные многочлены двух переменных. При обработке данных наблюдений, когда точки, в которых проводятся измерения, имеют две координаты $M_i(x_i, y_i)$, иногда возникает надобность приближения в виде многочлена степени N от двух переменных, который имеет вид

$$P_N(x,y) = \sum_{r=0}^{N}\sum_{s=0}^{r} b_{rs}x^{r-s}y^s. \tag{10.88}$$

Количество слагаемых в этой формуле, а значит, и число функций в наборе (10.79) определяются по формуле

$$m = (N+1)(N+2)/2, \tag{10.89}$$

а функции φ_k и неизвестные параметры a_k в (10.79), (10.83) и (10.84) определены формулами:

$$\varphi_k = x^{r-s}y^s,\ a_k = b_{rs},\ k = \frac{r+1}{2}r + s + 1,$$
$$r = 0,1,2,\ldots,N,\ s = 0,1,\ldots,r. \tag{10.90}$$

Глава 11
НЕЯВНЫЕ ФУНКЦИИ

11.1. Неявная функция

Понятие неявной функции

До сих пор мы имели дело с функциями, задаваемыми явно, т.е. $u = f(M)$, где M — точка евклидова пространства определенной размерности. Однако часто встречается случай, когда переменная u, являющаяся функцией аргументов $x, y, \ldots$, определена посредством уравнения

$$F(u,x,y,\ldots) = 0. \tag{11.1}$$

Тогда говорят, что функция u *задана неявно*, а уравнение (11.1) называется *функциональным уравнением*.

Возникает вопрос об однозначной разрешимости уравнения (11.1), т.е. при каких условиях это уравнение *однозначно* определяет u как функция аргументов $x, y, \ldots$. Приведем теоремы

о существовании неявно заданных функций; соответствующие доказательства мы опускаем.

Рассмотрим сначала случай неявной функции двух переменных:

$$F(x,y)=0. \tag{11.2}$$

Здесь справедлива следующая теорема о существовании y как функции аргумента x.

Теорема 11.1. *Пусть функция $F(x,y)$ обращается в нуль в точке $M_0(x_0,y_0)$ и дифференцируема в некоторой ее окрестности, причем частная производная $\dfrac{\partial F}{\partial y}$ непрерывна и не равна нулю в этой точке. Тогда можно указать такую ε-окрестность точки x_0 на числовой прямой Ox, что для всех точек из этой окрестности существует единственная функция $y=f(x)$, которая является решением уравнения (11.2), причем эта функция дифференцируема в указанной окрестности точки x_0 и удовлетворяет условию $f(x_0)=y_0$.*

Теперь обратимся к общему случаю неявного задания функции u от аргументов $x_1,x_2,\ldots,x_m$, определяемой функциональным уравнением

$$F(u,\bar{x})=F(u,x_1,x_2,\ldots,x_m)=0. \tag{11.3}$$

Здесь справедлива следующая теорема о существовании неявной функции $u=f(\bar{x})$.

Теорема 11.2. *Пусть функция $F(u,\bar{x})$ дифференцируема в точке $M_0(u_0,\bar{x}_0)$ и в некоторой ее окрестности, причем $F(u_0,\bar{x}_0)=0$, а частная производная $\dfrac{\partial F}{\partial u}$ непрерывна и не равна нулю в этой точке. Тогда можно указать такую ε-окрестность точки $\bar{x}_0$, что для всех точек из этой окрестности существует единственная дифференцируемая функция $u=f(\bar{x})$, которая является решением уравнения (11.3), причем $u_0=f(\bar{x}_0)$.*

Наконец, рассмотрим вопрос о существовании совокупности n неявных функций, определяемых системой функциональных уравнений.

Пусть n функций m переменных

$$\begin{cases} u_1=f_1(x_1,x_2,\ldots,x_m)=f_1(\bar{x}),\\ u_2=f_2(x_1,x_2,\ldots,x_m)=f_2(\bar{x}),\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ u_n=f_n(x_1,x_2,\ldots,x_m)=f_n(\bar{x}) \end{cases} \tag{11.4}$$

находятся как решение системы n функциональных уравнений

$$\begin{cases} F_1(u_1,u_2,\ldots,u_n;x_1,x_2,\ldots,x_m)=F_1(\bar{f},\bar{x})=0,\\ F_2(u_1,u_2,\ldots,u_n;x_1,x_2,\ldots,x_m)=F_2(\bar{f},\bar{x})=0,\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ F_n(u_1,u_2,\ldots,u_n;x_1,x_2,\ldots,x_m)=F_n(\bar{f},\bar{x})=0, \end{cases} \tag{11.5}$$

где $\bar{f}=(u_1,u_2,\ldots,u_n)$ — вектор-функция, компонентами которой являются функции (11.4).

Заметим, что более компактная форма систем (11.4) и (11.5) с использованием вектор-функций имеет вид

$$\bar{u}=\bar{f}(\bar{x}), \tag{11.6}$$

$$\bar{F}=\bar{F}(\bar{f},\bar{x})=\bar{0}, \tag{11.7}$$

где размерность векторов $\bar{u}$ и $\bar{F}$ равна n, а вектора $\bar{x}$ — равна m, $\bar{0}$ — нулевой вектор.

Составим из частных производных функций F_i якобиан, который нам понадобится в дальнейшем:

$$\frac{\partial\bar{F}}{\partial\bar{u}}=\left|\frac{\partial(F_1,F_2,\ldots,F_n)}{\partial(u_1,u_2,\ldots,u_n)}\right|=\begin{vmatrix} \dfrac{\partial F_1}{\partial u_1} & \dfrac{\partial F_1}{\partial u_2} & \ldots & \dfrac{\partial F_1}{\partial u_n}\\ \dfrac{\partial F_2}{\partial u_1} & \dfrac{\partial F_2}{\partial u_2} & \ldots & \dfrac{\partial F_2}{\partial u_n}\\ \ldots & \ldots & \ldots & \ldots\\ \dfrac{\partial F_n}{\partial u_1} & \dfrac{\partial F_n}{\partial u_2} & \ldots & \dfrac{\partial F_n}{\partial u_n} \end{vmatrix}. \tag{11.8}$$

Ответ на вопрос о разрешимости системы функциональных уравнений (11.5) относительно совокупности функций (11.4) дает следующая обобщенная теорема.

Теорема 11.3. *Пусть функции (11.5) дифференцируемы в некоторой окрестности точки $M_0(\bar{u}_0,\bar{x}_0)$, причем в точке M_0 выполняются следующие условия:*

а) *частные производные этих функций по переменным* $u_1, u_2, \ldots, u_n$ *непрерывны,*

б) *все функции* (11.5) *обращаются в нуль;*

в) *якобиан* (11.8) *отличен от нуля.*

Тогда найдется такая окрестность точки $\bar{x}_0 = (x_1^0, x_2^0, \ldots, x_m^0)$, *что в ее пределах существуют единственным образом определенные* n *функций* (11.4), *которые являются решением системы функциональных уравнений* (11.5), *причем это решение* $\bar{u} = \bar{f}(\bar{x})$ *дифференцируемо в указанной окрестности точки* $\bar{x}_0$ *и удовлетворяет условию* $\bar{f}(\bar{x}_0) = \bar{u}_0$.

Дифференцирование неявных функций

Правила дифференцирования неявных функций вытекают непосредственно из способа их задания при соблюдении условий, сформулированных в теоремах 11.1–11.3. Для случая функции двух переменных (11.2) получаем, дифференцируя $F(x, y)$ как сложную функцию аргумента x,

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} y' = 0,$$

откуда

$$y' = \frac{dy}{dx} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}. \tag{11.9}$$

В случае если функция m аргументов задана неявно одним функциональным уравнением (11.3), также используем правило дифференцирования сложной функции F, когда переменная u зависит от независимых переменных $x_1, x_2, \ldots, x_m$. Тогда, дифференцируя равенство (11.3) по x_i, имеем:

$$\frac{\partial F}{\partial x_i} + \frac{\partial F}{\partial u}\frac{\partial u}{\partial x_i} = 0, \qquad i = 1, 2, \ldots, m.$$

откуда получаем выражения для частных производных функции u

$$\frac{\partial u}{\partial x_i} = -\frac{\frac{\partial F}{\partial x_i}}{\frac{\partial F}{\partial u}}, \qquad i = 1, 2, \ldots, m. \tag{11.10}$$

Заметим, что соотношения (11.9) и (11.10) имеют смысл, когда соблюдено условие теорем 11.1 и 11.2, а именно $\dfrac{\partial F}{\partial u} \neq 0$.

В общем случае неявного задания системы n функций (11.4) m переменных посредством системы n функциональных уравнений (11.5), дифференцируя каждое уравнение (11.5), имеем:

$$\begin{cases} \dfrac{\partial F_1}{\partial u_1}\dfrac{\partial u_1}{\partial x_i} + \dfrac{\partial F_1}{\partial u_2}\dfrac{\partial u_2}{\partial x_i} + \cdots + \dfrac{\partial F_1}{\partial u_n}\dfrac{\partial u_n}{\partial x_i} + \dfrac{\partial F_1}{\partial x_i} = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial F_n}{\partial u_1}\dfrac{\partial u_1}{\partial x_i} + \dfrac{\partial F_n}{\partial u_2}\dfrac{\partial u_2}{\partial x_i} + \cdots + \dfrac{\partial F_n}{\partial u_n}\dfrac{\partial u_n}{\partial x_i} + \dfrac{\partial F_n}{\partial x_i} = 0, \end{cases} \tag{11.11}$$
$$i = 1, 2, \ldots, m.$$

Таким образом, мы имеем m систем алгебраических уравнений, линейных относительно неизвестных частных производных функций u_k по независимой переменной x_i. Все эти системы имеют один и тот же определитель — якобиан (11.8), который по условию теоремы 11.3 не равен нулю в фиксированной точке $(\bar{u}, \bar{x})$; следовательно, система (11.11) имеет единственное решение — набор $\{\partial u_k / \partial x_i\}$.

Неявная вектор-функция. Согласно параграфу 10.3, уравнения (11.11) можно представить в матричной форме, используя понятие производной вектор-функции по вектор-аргументу

$$\frac{\partial \bar{F}}{\partial \bar{u}}\frac{\partial \bar{u}}{\partial \bar{x}} + \frac{\partial \bar{F}}{\partial \bar{x}} = \mathbf{0}, \tag{11.12}$$

где $\mathbf{0}$ — нулевая матрица. Соответствующие матрицы Якоби имеют вид

$$\frac{\partial \bar{F}}{\partial \bar{u}} = \left\| \frac{\partial F_k}{\partial u_j} \right\|, \qquad k, j = 1, 2, \ldots, n; \tag{11.13}$$

$$\frac{\partial \bar{u}}{\partial \bar{x}} = \left\| \frac{\partial u_j}{\partial x_i} \right\|, \qquad j = 1, 2, \ldots, n; \quad i = 1, 2, \ldots, m; \tag{11.14}$$

$$\frac{\partial \bar{F}}{\partial \bar{x}} = \left\| \frac{\partial F_k}{\partial x_i} \right\|, \qquad k = 1, 2, \ldots, n; \quad i = 1, 2, \ldots, m. \tag{11.15}$$

Поскольку якобиан матрицы (11.13) в силу теоремы 11.3 не равен нулю, то существует обратная матрица, и тогда решение

матричного уравнения (11.12) можно представить в виде

$$\frac{\partial \bar{u}}{\partial \bar{x}} = \left\| \frac{\partial u_j}{\partial x_i} \right\| = -\left\| \frac{\partial F_k}{\partial x_i} \right\| \left\| \frac{\partial F_k}{\partial u_j} \right\|^{-1}. \qquad (11.16)$$

Отсюда следует, что формулировка теоремы 11.3 целиком переносится на случай неявной n-мерной вектор-функции $\bar{u}$ от m-мерного вектор-аргумента $\bar{x}$ с требованием, чтобы матрица (11.13) была невырожденной (чтобы существовала обратная по отношению к ней матрица).

Правило Крамера. Решение системы уравнений можно также представить, используя правило Крамера, в форме

$$\frac{\partial u_k}{\partial x_i} = -\left| \frac{\partial (F_1, F_2, \ldots, F_n)}{\partial (u_1, \ldots, u_{k-1}, x_i, u_{k+1}, \ldots, u_n)} \right| \left| \frac{\partial (F_1, F_2, \ldots, F_n)}{\partial (u_1, u_2, \ldots, u_n)} \right|^{-1},$$

$$k = 1, 2, \ldots, n; \quad i = 1, 2, \ldots, m. \qquad (11.17)$$

Рассмотрим примеры дифференцирования неявно заданных функций.

Пример 1. $F(x, y) = xy^2 - 1 = 0$.

Это уравнение задает в неявной форме y как функцию аргумента x. По формуле (11.9) получаем

$$y' = -\frac{y^2}{2xy} = -\frac{y}{2x}.$$

Пример 2. $F(x, y, z) = ax^2 + by^2 + cz^2 - 1 = 0$.

Это уравнение задает z как функцию аргументов x и y. По формулам (11.10) получаем выражения для частных производных функции z:

$$\frac{\partial z}{\partial x} = -\frac{ax}{cz}, \qquad \frac{\partial z}{\partial y} = -\frac{by}{cz}.$$

Пример 3. $\begin{cases} F_1 = u_1 + u_2^2 + x^2 + y^2 - 2 = 0, \\ F_2 = u_1^2 + u_2 - x^2 - y^2 + 1 = 0. \end{cases}$

Эта система двух функциональных уравнений определяет неявные функции u_1 и u_2 от аргументов x и y. Применим здесь правило Крамера (11.17), для чего вычислим якобиан

$$J = \frac{\partial (F_1, F_2)}{\partial (u_1, u_2)} = \begin{vmatrix} 1 & 2u_2 \\ 2u_1 & 1 \end{vmatrix} = 1 - 4u_1 u_2.$$

В силу теоремы 11.3 для разрешимости данной системы функциональных уравнений должно выполняться условие $1 - 4u_1 u_2 \neq 0$. Теперь последовательно находим формулы для частных производных функций u_1 и u_2 по формулам Крамера (11.17):

$$\frac{\partial u_1}{\partial x} = -\frac{\partial (F_1, E_2)}{\partial (x, u_2)} J^{-1} = -\begin{vmatrix} 2x & 2u_2 \\ -2x & 1 \end{vmatrix} J^{-1} = -2x \frac{1 + 2u_2}{1 - 4u_1 u_2},$$

$$\frac{\partial u_1}{\partial y} = -\frac{\partial (F_1, F_2)}{\partial (y, u_2)} J^{-1} = -\begin{vmatrix} 2y & 2u_2 \\ -2y & 1 \end{vmatrix} J^{-1} = -2y \frac{1 + 2u_2}{1 - 4u_1 u_2},$$

$$\frac{\partial u_2}{\partial x} = -\frac{\partial (F_1, F_2)}{\partial (u_1, x)} J^{-1} = -\begin{vmatrix} 1 & 2x \\ 2u_1 & -2x \end{vmatrix} J^{-1} = 2x \frac{1 + 2u_1}{1 - 4u_1 u_2},$$

$$\frac{\partial u_2}{\partial y} = -\frac{\partial (F_1, F_2)}{\partial (u_1, y)} J^{-1} = -\begin{vmatrix} 1 & 2y \\ 2u_1 & -2y \end{vmatrix} J^{-1} = 2y \frac{1 + 2u_1}{1 - 4u_1 u_2}.$$

Существование обратной функции

Теорема 11.1 может быть применена для установления условий, при которых функция $y = f(x)$ имеет в некоторой окрестности точки x_0 обратную функцию $x = f^{-1}(y)$, определенную в некоторой окрестности точки $y_0 = f(x_0)$. Представим функцию $y = f(x)$ в виде функционального уравнения $F(x, y) = f(x) - y = 0$. Если в этом выражении x является функцией от аргумента y, то по условию теоремы 11.1 мы получим следующее утверждение: *если функция $y = f(x)$ имеет в некоторой окрестности точки x_0 отличную от нуля производную, то существует функция $x = f^{-1}(y)$, определенная и дифференцируемая в некоторой окрестности точки y_0, где $y_0 = f(x_0)$. Производная этой обратной функции, согласно формуле (11.9), в которой x и y нужно поменять местами, равна $1/f'(x_0)$.*

Теперь определим обратную вектор-функцию по отношению к вектор-функции $\bar{u} = \bar{f}(\bar{x})$. Рассмотрим случай, когда размерности векторов $\bar{u}$ и $\bar{x}$ равны (см. параграф 10.3). Образуем неявную вектор-функцию

$$\bar{F}(\bar{x}, \bar{u}) = \bar{f}(\bar{x}) - \bar{u} = 0. \tag{11.18}$$

Тогда вопрос о существовании обратной вектор-функции $\bar{x} = \bar{f}^{-1}(\bar{u})$ связан с вопросом о разрешимости векторного функционального уравнения (11.18). Найдем векторную производную вектор-функции $\bar{F}$ по вектору $\bar{u}$; из уравнения (11.18) имеем

$$\frac{\partial \bar{F}}{\partial \bar{u}} = \frac{\partial \bar{f}}{\partial \bar{x}} \frac{\partial \bar{x}}{\partial \bar{u}} - \boldsymbol{E} = \boldsymbol{0},$$

где $\partial \bar{f}/\partial \bar{u}$ — матрица Якоби (10.61), $\boldsymbol{E}$ и $\boldsymbol{0}$ — соответственно единичная и нулевая матрицы. Умножая обе части этого равенства на матрицу, обратную матрице (10.61), получаем, что матрица Якоби для производной вектор-функции $\bar{x}$ по вектор-аргументу $\bar{u}$ выражается соотношением

$$\frac{\partial \bar{x}}{\partial \bar{u}} = \left(\frac{\partial \bar{f}}{\partial \bar{x}} \right)^{-1} = \left\| \frac{\partial f_i}{\partial x_j} \right\|^{-1}; \tag{11.19}$$

$$i, j = 1, 2, \ldots, m.$$

Как следствие теоремы 11.3 справедливо следующее утверждение, которое можно трактовать как теорему об обратной вектор-функции.

Если вектор-функция $\bar{u} = \bar{f}(\bar{x})$ имеет невырожденную матрицу Якоби (10.61) *в некоторой окрестности точки $\bar{x}_0$, то для этой функции существует обратная функция $\bar{x} = \bar{f}^{-1}(\bar{u})$, которая определена и дифференцируема в некоторой окрестности точки $\bar{u}_0$, где $\bar{u}_0 = \bar{f}(\bar{x}_0)$. Матрица Якоби для этой обратной функции равна обратной матрице для* (10.61).

11.2. Зависимость функций

Понятие зависимости функций

Пусть даны n функций одних и тех же m переменных

$$\begin{cases} u_1 = f_1(x_1, x_2, \ldots, x_m), \\ u_2 = f_2(x_1, x_2, \ldots, x_m), \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ u_n = f_n(x_1, x_2, \ldots, x_m). \end{cases} \tag{11.20}$$

Пусть эти функции определены и дифференцируемы в некоторой открытой области D из m-мерного пространства.

Пусть для всех точек из области D для одной из функций (11.20), например для u_k может быть установлена функциональная зависимость от остальных функций

$$u_k = \Phi(u_1, u_2, \ldots, u_{k-1}, u_{k+1}, \ldots, u_n), \tag{11.21}$$

где Φ — некоторая функция, которая определена и дифференцируема в соответствующей области изменения своих аргументов. В таком случае будем говорить, что *функция u_k зависит в области D от остальных функций* (11.20). Функции (11.20) называются *зависимыми в области D*, если хотя бы одна из них зависит от остальных в области D.

Если для набора функций (11.20) не существует такой дифференцируемой функции Φ, что для всех точек некоторой области D выполняется тождество (11.21) для какой-либо u_k, то функции (11.20) будем называть независимыми в области D.

Например, четыре функции двух переменных

$$\begin{aligned} u_1 &= \sin x + \cos y, \\ u_2 &= \cos x + \sin y, \\ u_3 &= \sin(x + y), \\ u_4 &= 1 \end{aligned}$$

являются зависимыми в любой области D координатной плоскости Oxy, поскольку для любой точки (x, y) функция u_4 зависит от трех других:

$$u_4 = (u_1^2 + u_2^2)/2 - u_3.$$

В курсе линейной алгебры вводится понятие *линейной зависимости n функций в некоторой области D*, если для всех точек из D одна из функций выражается в виде линейной функции от остальных. Определенная выше зависимость функций включает в себя как частный случай и их линейную зависимость.

Достаточное условие независимости функций

Теорема 11.4. *Пусть n функций от m переменных* (11.20) *определены и дифференцируемы в окрестности точки M_0, причем $n \leqslant m$. Тогда, если якобиан из этих функций по каким-либо n переменным отличен от нуля в точке M_0, эти функции независимы в некоторой окрестности точки M_0.*

Доказательство. Пусть в точке M_0 отличен от нуля якобиан по первым n переменным:

$$J = \left|\frac{\partial(u_1, u_2, \ldots, u_n)}{\partial(x_1, x_2, \ldots, x_n)}\right| \neq 0. \tag{11.22}$$

Доказательство проведем от противного, т.е. предположим, что функции (11.20) зависимы в некоторой окрестности точки M_0. Это означает, согласно определению, данному ранее, что одна из них, например u_k, является функцией от остальных как от аргументов, т.е. выполняется равенство (11.21), где Φ — некоторая функция, дифференцируемая в окрестности точки M_0. Тогда производная функции u_k в точке M_0 по любой из переменных x_i, согласно правилу дифференцирования сложной функции, имеет вид

$$\begin{gathered}\frac{\partial u_k}{\partial x_i} = \frac{\partial \Phi}{\partial u_1}\frac{\partial u_1}{\partial x_i} + \cdots + \frac{\partial \Phi}{\partial u_{k-1}}\frac{\partial u_{k-1}}{\partial x_i} \\ + \frac{\partial \Phi}{\partial u_{k+1}}\frac{\partial u_{k+1}}{\partial x_i} + \cdots + \frac{\partial \Phi}{\partial u_n}\frac{\partial u_n}{\partial x_i}, \\ i = 1, 2, \ldots, n.\end{gathered} \tag{11.23}$$

Индекс i пробегает значения от 1 до n, следовательно, k-я строка определителя матрицы Якоби является линейной комбинацией остальных $(n-1)$ его строк с коэффициентами соответственно $\dfrac{\partial \Phi}{\partial u_j}$, где $j = 1, 2, \ldots, k-1, k+1, \ldots, n$. Но тогда якобиан J равен нулю, т.е. условие (11.22) не выполняется, что противоречит условию теоремы. Доказательство завершено.

Рассмотрим примеры на определение зависимости наборов функций.

Пример 1. $u_1 = 2x + y$, $u_2 = x - 2y$.

Составим якобиан этих функций и вычислим его:

$$J = \left|\frac{\partial(u_1, u_2)}{\partial(x, y)}\right| = \begin{vmatrix} 2 & 1 \\ 1 & -2 \end{vmatrix} = -5 \neq 0.$$

Якобиан отличен от нуля независимо от точки (x, y), следовательно, эти функции независимы на всей плоскости Oxy.

Пример 2. $u_1 = x^2 + y^2 - 2$, $u_2 = 2x + y$.

Якобиан этой системы функций

$$J(x, y) = \begin{vmatrix} 2x & 2y \\ 2 & 1 \end{vmatrix} = 2x - 4y.$$

Якобиан равен нулю на прямой $y = x/2$, следовательно, данные функции независимы на плоскости Oxy, за исключением точек на этой прямой.

Пример 3. $u_1 = x^3 + 2y^3 + z^2$, $u_2 = x + y + z$, $u_3 = 2x - y + z$.

Найдем якобиан

$$J(x, y, z) = \begin{vmatrix} 3x^2 & 6y^2 & 2z \\ 1 & 1 & 1 \\ 2 & -1 & 1 \end{vmatrix} = 6x^2 + 6y^2 - 6z.$$

Данные функции независимы в пространстве $Oxyz$, исключая поверхность параболоида вращения $z = x^2 + y^2$, на которой якобиан обращается в нуль.

Функциональные матрицы

Пусть функции (11.20) определены и дифференцируемы в некоторой окрестности точки $M(x_1, x_2, \ldots, x_m)$, причем все

первые частные производные этих функций непрерывны в самой точке M. *Функциональная матрица* Якоби, составленная из частных производных функций (11.20), имеет вид

$$A = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_m} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_m} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial f_n}{\partial x_1} & \dfrac{\partial f_n}{\partial x_2} & \cdots & \dfrac{\partial f_n}{\partial x_m} \end{pmatrix}. \qquad (11.24)$$

Эта матрица содержит n строк и m столбцов; напомним, что $n \leqslant m$.

Имеет место следующая теорема, которую мы приводим без доказательства.

Теорема 11.5. *Пусть ранг функциональной матрицы* (11.24) *равен r в точке $M(x_1, x_2, \ldots, x_m)$, т.е. у данной матрицы существует некоторый минор r-го порядка, составленный с участием r функций из* (11.20), *который не равен нулю в точке M:*

$$\left| \frac{\partial(f_{i_1}, f_{i_2}, \ldots, f_{i_r})}{\partial(x_{j_1}, x_{j_2}, \ldots, x_{j_r})} \right| \neq 0. \qquad (11.25)$$

Тогда эти r функций независимы в окрестности точки M, а каждая из остальных $n-r$ функций в (11.20) *зависит в этой окрестности от r указанных функций.*

Эту теорему называют *достаточным условием зависимости функций*. Покажем применение ее на примере: исследовать зависимость функций

$$u_1 = x_1^2 + x_2^2 + x_3^2 + x_4^2,$$
$$u_2 = x_1 + x_2 + x_3 + x_4,$$
$$u_3 = x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4.$$

Функциональная матрица (11.24) в данном случае имеет вид

$$A = \begin{pmatrix} 2x_1 & 2x_2 & 2x_3 & 2x_4 \\ 1 & 1 & 1 & 1 \\ x_2+x_3+x_4 & x_1+x_3+x_4 & x_1+x_2+x_4 & x_1+x_2+x_3 \end{pmatrix}.$$

Все определители третьего порядка тождественно равны нулю при любых значениях неизвестных x_1, x_2, x_3, x_4, так как линейная комбинация первой и последней строк матрицы равна второй строке. Нетрудно убедиться, что при несовпадающих значениях четырех координат хотя бы один из определителей второго порядка, включающих в себя элементы первой и второй строк матрицы, отличен от нуля. Следовательно, в евклидовом пространстве E^4 исключая гиперпрямую $x_1 = x_2 = x_3 = x_4$ функции u_1 и u_2 независимы, а функция u_3 зависит от u_1 и u_2. На гиперпрямой $x_1 = x_2 = x_3 = x_4$, все данные функции являются зависимыми.

11.3. Условный экстремум

Понятие и постановка задачи условного экстремума

В различных приложениях часто встречается задача нахождения экстремумов функции нескольких переменных при наличии *дополнительных условий связи* на ее аргументы. Такие экстремумы называют *условными*.

Например, пусть требуется найти экстремум функции $u = x^2 + y^2$, если ее аргументы удовлетворяют условию связи $x + y - 2 = 0$. Экстремумы данной функции ищутся не на всей плоскости Oxy, а только на прямой $x+y-2 = 0$ (рис. 11.1). Подставим в уравнение функции выражение для y из условия связи и тем самым сведем исходную задачу об условном экстремуме функции $u = x^2 + y^2$ к задаче отыскания безусловного экстремума функции $u = 2x^2 - 4x + 4$. Эта функция имеет экстремум в точке $x = 1$, т.е. функция $u = x^2 + y^2$ при наличии указанной связи имеет условный минимум $u = 2$ в точке (1,1), который не совпадает с безусловным минимумом параболоида вращения в точке $O(0,0)$.

В частности, задачи на условный экстремум возникают при отыскании экстремума функции в замкнутой ограниченной области, когда условием связи на аргументы функции является уравнение границы области ее задания (см. параграфы 10.3 и 10.4).

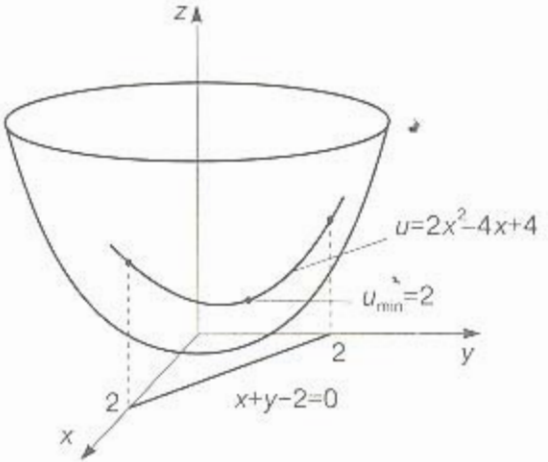

Рис. 11.1

Рассмотрим общую постановку задачи нахождения условного экстремума. Пусть задана функция $m + n$ переменных

$$u = f(x_1, x_2, \ldots, x_m, y_1, \ldots, y_n), \tag{11.26}$$

для которой нужно найти условный экстремум при наличии n условий связи:

$$\begin{cases} F_1(x_1, x_2, \ldots, x_m, y_1, \ldots, y_n) = 0, \\ F_2(x_1, x_2, \ldots, x_m, y_1, \ldots, y_n) = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ F_n(x_1, x_2, \ldots, x_m, y_1, \ldots, y_n) = 0, \end{cases} \tag{11.27}$$

Будем говорить, что функция (11.26) имеет условный экстремум при наличии связей (11.27) в точке $M_0(x_1^0, x_2^0, \ldots, x_m^0, y_1^0, \ldots, y_n^0)$, координаты которой удовлетворяют условиям связи (11.27), если существует такая окрестность точки M_0, для которой значение функции (11.26) в точке M_0 является наибольшим (наименьшим) для всех точек, координаты которых удовлетворяют условиям связи (11.27).

Сведение к задаче о безусловном экстремуме

Пусть неявные функции связей (11.27) дифференцируемы в некоторой окрестности точки M_0, причем в самой точке M_0

частные производные этих функций по $y_1, \ldots, y_n$ непрерывны и якобиан

$$J = \left| \frac{\partial(F_1, F_2, \ldots, F_n)}{\partial(y_1, y_2, \ldots, y_n)} \right| \tag{11.28}$$

не равен нулю. Тогда, в силу обобщенной теоремы 11.3, существует такая окрестность точки $M_0'(x_1^0, x_2^0, \ldots, x_m^0)$ пространства E^m, что для нее определены n функций

$$\begin{cases} y_1 = \varphi_1(x_1, x_2, \ldots, x_m), \\ y_2 = \varphi_2(x_1, x_2, \ldots, x_m), \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = \varphi_n(x_1, x_2, \ldots, x_m), \end{cases} \tag{11.29}$$

которые являются единственным и дифференцируемым решением системы уравнений (11.27).

Подставляя функции (11.29) в выражение для функции u (11.26), мы получим, что задача об условном экстремуме в точке M_0 для функции (11.26) при наличии связей (11.27) свелась к задаче о безусловном экстремуме в точке M_0' для сложной функции независимых аргументов $x_1, x_2, \ldots, x_m$:

$$\begin{aligned} u = f\big[&x_1, x_2, \ldots, x_m, \\ &\varphi_1(x_1, \ldots, x_m), \ldots, \varphi_n(x_1, \ldots, x_m)\big]. \end{aligned} \tag{11.30}$$

Задача нахождения безусловного экстремума функции нескольких переменных была изучена ранее (см. параграф 10.3).

Метод неопределенных множителей Лагранжа

Поскольку в точке M_0 сложная функция (11.30) имеет экстремум, то его необходимым условием является равенство нулю дифференциала данной функции в этой точке (см. параграф 10.3). При этом будем полагать, что все необходимые условия дифференцируемости функций (11.30) и (11.29) соблюдены. В силу инвариантности формы первого дифференциала в точке M_0 имеем

$$\begin{aligned} du = &\frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_m} dx_m + \\ &+ \frac{\partial f}{\partial y_1} dy_1 + \cdots + \frac{\partial f}{\partial y_n} dy_n = 0, \end{aligned} \tag{11.31}$$

где все частные производные берутся в точке M_0. Заметим, что в (11.31) dy_i являются дифференциалами функций (11.29), так что это равенство не представляет собой тождества относительно dy_i.

Подставим функции (11.29) в уравнения связи (11.27); так как эти функции являются решением системы функциональных уравнений связи, то уравнения (11.27) превратятся в тождества. Дифференцируя эти тождества, получаем

$$\begin{cases} \dfrac{\partial F_1}{\partial x_1}dx_1+\cdots+\dfrac{\partial F_1}{\partial x_m}dx_m+\dfrac{\partial F_1}{\partial y_1}dy_1+\cdots+\dfrac{\partial F_1}{\partial y_n}dy_n=0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial F_n}{\partial x_1}dx_1+\cdots+\dfrac{\partial F_n}{\partial x_m}dx_m+\dfrac{\partial F_n}{\partial y_1}dy_1+\cdots+\dfrac{\partial F_n}{\partial y_n}dy_n=0. \end{cases} \tag{11.32}$$

Умножим равенства (11.32) соответственно на произвольные и еще не определенные *множители Лагранжа* $\lambda_1, \lambda_2, \ldots, \lambda_n$ и сложим их почленно с равенством (11.31). В результате имеем равенство

$$\frac{\partial \Phi}{\partial x_1}dx_1+\cdots+\frac{\partial \Phi}{\partial x_m}dx_m+\frac{\partial \Phi}{\partial y_1}dy_1+\cdots+\frac{\partial \Phi}{\partial y_n}dy_n=0, \tag{11.33}$$

где $\Phi(x_1, x_2, \ldots, x_m, y_1, \ldots, y_n)$ — *функция Лагранжа*

$$\Phi = f+\lambda_1 F_1+\lambda_2 F_2+\cdots+\lambda_n F_n. \tag{11.34}$$

Выберем теперь множители Лагранжа такими, чтобы выполнялись равенства

$$\frac{\partial \Phi}{\partial y_1}=0, \ldots, \frac{\partial \Phi}{\partial y_n}=0. \tag{11.35}$$

Это можно сделать, так как эти равенства, с учетом вида функции Лагранжа (11.34), приводят к системе n линейных уравнений относительно чисел $\lambda_1, \lambda_2, \ldots, \lambda_n$:

$$\begin{cases} \dfrac{\partial f}{\partial y_1}+\lambda_1\dfrac{\partial F_1}{\partial y_1}+\cdots+\lambda_n\dfrac{\partial F_n}{\partial y_1}=0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial f}{\partial y_n}+\lambda_1\dfrac{\partial F_1}{\partial y_n}+\cdots+\lambda_n\dfrac{\partial F_n}{\partial y_n}=0. \end{cases} \tag{11.36}$$

Определитель этой системы [якобиан (11.28)] отличен от нуля, и потому множители Лагранжа определяются из нее в точке экстремума M_0 однозначно.

При выполнении равенств (11.35) из равенства (11.31) получаем в силу произвольности дифференциалов независимых переменных $x_1, x_2, \ldots, x_m$

$$\frac{\partial \Phi}{\partial x_1}=0, \ \frac{\partial \Phi}{\partial x_2}=0, \ldots, \frac{\partial \Phi}{\partial x_m}=0. \tag{11.37}$$

Присоединяя к уравнениям (11.35) и (11.37) уравнения связи (11.27), получаем систему, которая состоит из $m+2n$ уравнений

$$\begin{cases} \dfrac{\partial \Phi}{\partial x_1}=0, \ldots, \dfrac{\partial \Phi}{\partial x_m}=0, \dfrac{\partial \Phi}{\partial y_1}=0, \ldots, \dfrac{\partial \Phi}{\partial y_n}=0, \\ F_1=0, \ldots, F_n=0. \end{cases} \tag{11.38}$$

Из этой системы определяют $m+n$ координат точек возможного условного экстремума и n множителей $\lambda_1, \ldots, \lambda_n$.

Практическая реализация метода неопределенных множителей Лагранжа выглядит следующим образом:

1) составляется функция Лагранжа (11.34);
2) для этой функции как функции $m+2n$ координат $x_1, \ldots, x_m, y_1, \ldots, y_n, \lambda_1, \ldots, \lambda_n$ находят точки безусловного экстремума;
3) исключают множители $\lambda_1, \ldots, \lambda_n$ посредством использования условий связи (11.27).

Рассмотрим пример на условный экстремум.

Пример 1. Пусть дана производственная функция Кобба–Дугласа, зависящая от двух переменных,

$$Q(x, y) = Ax^{\alpha}y^{1-\alpha},$$

где A — постоянная конкретного производства, x — затраты капитала, y — суммарные затраты ресурсов. Найти значения величин x и y при ценах используемых ресурсов соответственно P_1 и P_2 чтобы при фиксированном объеме выпускаемой продукции Q_0 обеспечивался минимум затрат.

Заметим прежде всего, что функция затрат выражается формулой

$$C = P_1 x + P_2 y.$$

Математическая формулировка задачи на условный экстремум такова: найти минимум функции C при условии, что $Q(x, y) = Q_0$. Составим функцию Лагранжа

$$\Phi = P_1 x + P_2 y + \lambda(Q_0 - A x^{\alpha} y^{1-\alpha}).$$

Необходимое условие экстремума этой функции трех переменных:

$$\frac{\partial \Phi}{\partial x} = 0, \quad \frac{\partial \Phi}{\partial y} = 0, \quad \frac{\partial \Phi}{\partial \lambda} = 0.$$

Отсюда получаем систему трех уравнений относительно трех неизвестных x, y и λ:

$$\begin{cases} Q_0 - A x^{\alpha} y^{1-\alpha} = 0, \\ P_1 - \lambda \alpha A x^{\alpha-1} y^{1-\alpha} = 0, \\ P_2 - \lambda (1-\alpha) A x^{\alpha} y^{-\alpha} = 0. \end{cases}$$

Решение этой системы дает точку условного минимума функции затрат C:

$$x = \frac{Q_0}{A}\left(\frac{1-\alpha}{\alpha}\frac{P_1}{P_2}\right)^{\alpha-1}, \quad y = \frac{Q_0}{A}\left(\frac{1-\alpha}{\alpha}\frac{P_1}{P_2}\right)^{\alpha}.$$

Достаточные условия условного экстремума

Приведенные выше условия (11.38) являются необходимыми для существования экстремума функции (11.26) при наличии условий связи (11.27). Но тогда для получения достаточного условия экстремума функции (11.26) в точке $M_0(x_1^0, x_2^0, \ldots, x_m^0, y_1^0, \ldots, y_n^0)$ при наличии связей (11.27) нужно присоединить к условиям (11.38) требование знакоопределенности в этой точке второго дифференциала функции Лагранжа (11.34). Если $d^2\Phi\big|_{M_0} > 0$, то в точке M_0 имеет место

условный минимум, если же $d^2\Phi\big|_{M_0} < 0$, имеет место условный максимум. В общем случае второй дифференциал не обладает свойством инвариантности, и, поскольку $y_1, y_2, \ldots, y_n$ являются функциями независимых переменных $x_1, x_2, \ldots, x_m$, его запись имеет вид

$$d^2\Phi = \left(\frac{\partial}{\partial x_1}dx_1 + \cdots + \frac{\partial}{\partial x_m}dx_m + \right.$$
$$\left. + \frac{\partial}{\partial y_1}dy_1 + \cdots + \frac{\partial}{\partial y_n}dy_n\right)^2 \Phi +$$
$$+ \frac{\partial \Phi}{\partial y_1}d^2 y_1 + \cdots + \frac{\partial \Phi}{\partial y_n}d^2 y_n.$$

Но поскольку в точке возможного условного экстремума M_0 справедливо равенство (11.35), второй дифференциал функции Лагранжа в этой точке имеет такую же форму записи, как если бы все переменные $x_1, x_2, \ldots, x_m$, $y_1, \ldots, y_n$ были независимыми:

$$d^2\Phi = \left(\frac{\partial}{\partial x_1}dx_1 + \cdots + \frac{\partial}{\partial x_m}dx_m + \right.$$
$$\left. + \frac{\partial}{\partial y_1}dy_1 + \cdots + \frac{\partial}{\partial y_n}dy_n\right)^2 \Phi. \qquad (11.39)$$

Поскольку требуется установить знакоопределенность $d^2\Phi$ лишь при наличии условий связи (11.27), то при проведении вычислений в формулу (11.39) следует подставить вместо $dy_1, \ldots, dy_n$ их выражения, определяемые единственным образом из системы линейных уравнений (11.32). Только после этого можно изучить уже известными методами с применением матрицы Гессе вопрос о знакоопределенности $d^2\Phi$ в данной точке M_0.

Глава 12
ДВОЙНЫЕ ИНТЕГРАЛЫ

В этой главе рассмотрены вопросы интегрирования функций двух переменных. Полученные результаты могут быть обобщены на случай функций трех и более переменных.

12.1. Существование двойного интеграла

Понятие двойного интеграла

Пусть D — некоторая замкнутая и ограниченная область на координатной плоскости Oxy и в ней определена произвольная ограниченная функция $z = f(x, y)$. Разобьем область D сетью кривых на n произвольных частей D_i, не имеющих общих внутренних точек, площади которых ΔS_i (рис. 12.1). В каждой из частичных областей D_i возьмем произвольную внутреннюю точку (ξ_i, η_i) и составим сумму произведений

$$\sigma = \sum_{i=1}^{n} f(\xi_i, \eta_i)\Delta S_i. \qquad (12.1)$$

Выражение (12.1) называется *интегральной суммой* для функции $f(x, y)$ в области D.

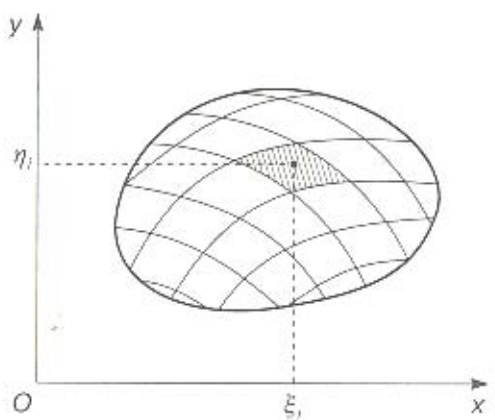

Рис. 12.1

Введем понятие диаметра области: *диаметром* $d(D)$ области D будем называть наибольшее расстояние между граничными точками этой области. Пусть λ — наибольший из всех диаметров частичных областей:

$$\lambda = \max_{1 \leqslant i \leqslant n} \{d(D_i)\}. \qquad (12.2)$$

Определение. Если интегральная сумма (12.1) имеет предел при $\lambda \to 0$, то этот предел называется *двойным интегралом* от функции $f(x, y)$ по области D.

Двойной интеграл обозначается следующими символами:

$$\lim_{\lambda \to 0} \sum_{i=1}^{n} f(\xi_i, \eta_i)\Delta S_i = \iint\limits_D f(x, y)\, dS = \iint\limits_D f(x, y)\, dx\, dy. \qquad (12.3)$$

Функция $f(x, y)$ называется *интегрируемой* в области D, D — областью интегрирования, x и y — *переменными интегрирования*, а dS или $dx\, dy$ — *элементом площади*.

По аналогии с определенным интегралом для случая двойного интеграла действуют следующие теоремы о классах интегрируемых функций.

Теорема 12.1. *Если функция $f(x, y)$ непрерывна в замкнутой ограниченной области, то она интегрируема в этой области.*

Как и для случая функций одной переменной, класс интегрируемых функций оказывается более широким, чем класс непрерывных функций.

Теорема 12.2. *Пусть функция $f(x, y)$ ограничена в замкнутой ограниченной области D и имеет разрывы лишь на конечном числе кривых этой области, которые являются графиками непрерывных функций вида $y = g(x)$ или $x = \varphi(y)$. Тогда функция $f(x, y)$ интегрируема в области D.*

Геометрический смысл двойного интеграла

Рассмотрим тело P, ограниченное сверху поверхностью, описываемой непрерывной неотрицательной в области D функцией $z = f(x, y)$, снизу — областью D, а с боков — цилиндрической поверхностью, у которой направляющей является граница области D и образующие параллельны оси Oz (рис. 12.2). Такое тело называется *криволинейным цилиндром*.

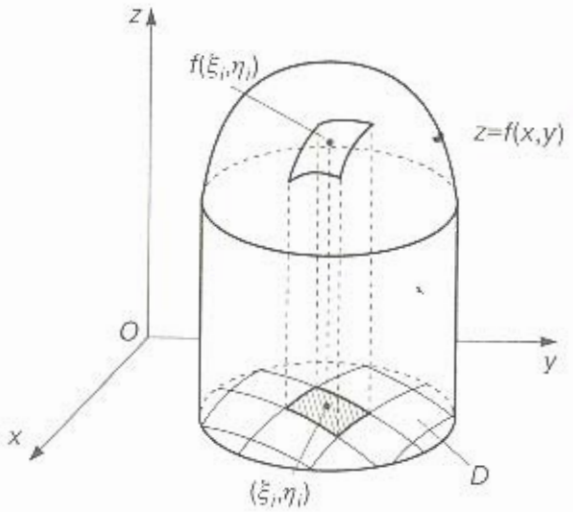

Рис. 12.2

Интегральная сумма (12.1) представляет собой сумму объемов прямых цилиндров с площадями оснований ΔS_i и высотами $f(\xi_i, \eta_i)$. Эта сумма является приближенным значением объема тела P:

$$V_P \approx \sum_{i=1}^{n} f(\xi_i, \eta_i)\Delta S_i.$$

Поскольку функция $f(x, y)$ интегрируема, то предел этой интегральной суммы существует и равен двойному интегралу от этой функции по области D, т.е.

$$V_P = \iint\limits_D f(x, y)\, dx\, dy.$$

Отсюда ясен геометрический смысл двойного интеграла: *двойной интеграл по области D от непрерывной неотрицательной функции равен объему криволинейного цилиндра с основанием D.*

Свойства двойного интеграла

Основные свойства двойного интеграла аналогичны свойствам однократного определенного интеграла. Здесь мы ограничимся их формулировкой.

Пусть функции $f(x, y)$ и $g(x, y)$ интегрируемы в области D.

1°. *Для любых вещественных чисел α и β функция $[\alpha f(x, y) + \beta g(x, y)]$ интегрируема в области D, причем*

$$\iint\limits_D [\alpha f(x, y) + \beta g(x, y)]\, dx\, dy = \alpha \iint\limits_D f(x, y)\, dx\, dy + \beta \iint\limits_D g(x, y)\, dx\, dy$$

— линейное свойство.

2°. *Если область D является объединением областей D_1 и D_2, не имеющих общих внутренних точек, и в каждой из этих областей функция $f(x, y)$ интегрируема, то эта функция интегрируема также и в области D, причем*

$$\iint\limits_D f(x, y)\, dx\, dy = \iint\limits_{D_1} f(x, y)\, dx\, dy + \iint\limits_{D_2} f(x, y)\, dx\, dy$$

— свойство аддитивности.

3°. *Произведение функций fg интегрируемо в D.*

4°. *Если $f(x, y) \leqslant g(x, y)$ всюду в D, то*

$$\iint\limits_D f(x, y)\, dx\, dy \leqslant \iint\limits_D g(x, y)\, dx\, dy.$$

5°. *Функция $|f(x, y)|$ интегрируема в D, причем*

$$\left| \iint\limits_D f(x, y)\, dx\, dy \right| \leqslant \iint\limits_D |f(x, y)|\, dx\, dy.$$

6°. *(теорема о среднем). Если функция $f(x, y)$ непрерывна в области D, то найдется такая точка $(\xi, \eta) \in D$, что*

$$\iint\limits_D f(x, y)\, dx\, dy = f(\xi, \eta)S,$$

где S — площадь области D.

7°. $\iint\limits_D 1\,dx\,dy$ *равен площади области* D — геометрическое свойство.

12.2. Вычисление двойных интегралов

Сведение двойного интеграла к повторному

В вычислении двойных интегралов основную роль играет теорема о сведении двойного интеграла к повторному, т.е. о возможности дважды применить процесс обычного интегрирования.

Теорема 12.3. *Пусть функция* $f(x,y)$ *определена и интегрируема в области* D, *ограниченной снизу и сверху двумя непрерывными кривыми* $y = y_1(x)$ *и* $y = y_2(x)$, *причем* $y_1(x) \leqslant y \leqslant y_2(x)$, $a \leqslant x \leqslant b$ (*рис.* 12.3). *Пусть также для каждого* $x \in [a,b]$ *существует определенный интеграл*

$$I(x) = \int\limits_{y_1(x)}^{y_2(x)} f(x,y)\,dy.$$

Тогда существует повторный интеграл

$$\int\limits_b^a I(x)dx = \int\limits_a^b dx \int\limits_{y_1(x)}^{y_2(x)} f(x,y)\,dy$$

и справедливо равенство

$$\iint\limits_D f(x,y)\,dx\,dy = \int\limits_b^a dx \int\limits_{y_1(x)}^{y_2(x)} f(x,y)\,dy. \qquad (12.4)$$

Доказательство этой теоремы мы опускаем.

Отметим частный случай прямоугольной области D, т.е. $y_1(x) \equiv c$, $y_2(x) \equiv d$; тогда сведение двойного интеграла к повторному имеет наиболее простой вид:

$$\iint\limits_D f(x,y)\,dx\,dy = \int\limits_a^b dx \int\limits_c^d f(x,y)\,dy. \qquad (12.5)$$

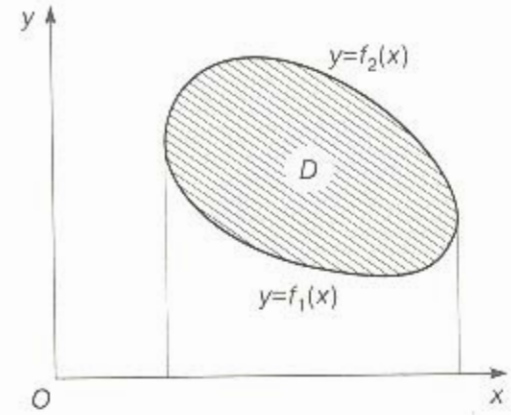

Рис. 12.3

Рассмотрим примеры вычисления двойных интегралов.

Пример 1. Найти $\iint\limits_D (3x^2y - 2x^3)\,dx\,dy$; $D = \{0 \leqslant x \leqslant 1, 1 \leqslant y \leqslant 2\}$.

Данный интеграл нужно вычислить по прямоугольной области. Согласно формуле (12.5),

$$\iint\limits_D (3x^2y - 2x^3)\,dx\,dy = \int\limits_0^1 dx \int\limits_1^2 (3x^2y - 2x^3)\,dy =$$

$$= \int\limits_0^1 (3x^2y^2/2 - 2x^3y)\Big|_1^2\,dx = \int\limits_0^1 (9x^2/2 - 2x^3)\,dx =$$

$$= (3x^3 - x^4)/2\Big|_0^1 = 1.$$

Пример 2. Найти $\iint\limits_D y^2x\,dx\,dy$, если область D ограничена линиями $x^2 + y^2 = 4$, $x + y - 2 = 0$.

Область D заключена между прямой $y_1 = 2 - x$ и частью верхней полуокружности радиуса 2: $y_2 = \sqrt{4 - x^2}$ (рис. 12.4); точки пересечения этих линий $(0,2)$ и $(2,0)$. Следовательно, область интегрирования $D = \{0 \leqslant x \leqslant 2, 2 - x \leqslant y \leqslant \sqrt{4 - x^2}\}$. По формуле (12.4) данный двойной интеграл сводится к по-

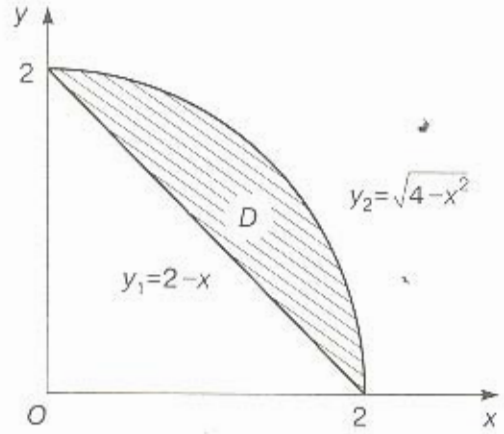

Рис. 12.4

вторному интегралу

$$\iint\limits_D y^2x\,dx\,dy = \int\limits_0^2 dx \int\limits_{2-x}^{\sqrt{4-x^2}} y^2x\,dy = \frac{1}{3}\int\limits_0^2 xy^3\Big|_{2-x}^{\sqrt{4-x^2}}\,dx =$$

$$= \frac{1}{3}\int\limits_0^2 [x(4-x^2)^{3/2} - x(2-x)^3]\,dx = -\frac{1}{15}(4-x^2)^{5/2}\Big|_0^2 +$$

$$+\frac{1}{3}\int\limits_0^2 t^3(t-2)\,dt = \frac{32}{15} + \frac{1}{3}(t^2/5 - t^4/2)\Big|_0^2 = \frac{8}{5}.$$

Здесь при интегрировании второго слагаемого в квадратных скобках была применена замена переменной $t = 2 - x$.

Пример 3. Найти $\iint\limits_D \sin(x+y)\,dx\,dy$, если область D ограничена линиями $x = y$, $x + y = \dfrac{\pi}{2}$, $y = 0$.

Здесь область интегрирования представляет собой треугольник, одна сторона которого длиной $\frac{\pi}{2}$ лежит на оси Ox, а две другие, уравнения которых соответственно $y = x$ и $y = \dfrac{\pi}{2} - x$, пересекаются при $x = \dfrac{\pi}{4}$ (рис. 12.5). Стало быть, область интегрирования разбивается на два треугольника D_1 и D_2, и, согласно свойству аддитивности 2° (см. параграф 12.1),

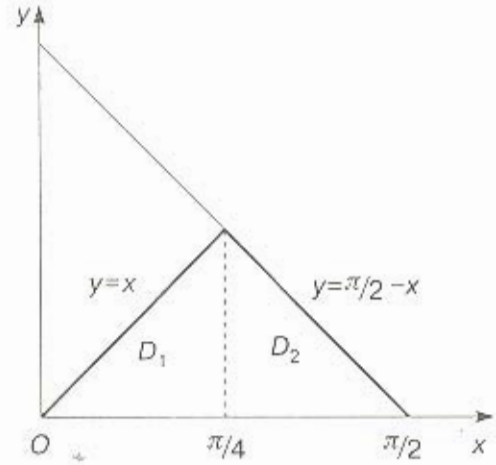

Рис. 12.5

данный интеграл является суммой двойных интегралов с областями интегрирования соответственно D_1 и D_2:

$$\iint\limits_D \sin(x+y)\,dx\,dy =$$

$$= \iint\limits_{D_1} \sin(x+y)\,dx\,dy + \iint\limits_{D_2} \sin(x+y)\,dx\,dy =$$

$$= \int\limits_0^{\pi/4} dx \int\limits_0^x \sin(x+y)\,dy + \int\limits_{\pi/4}^{\pi/2} dx \int\limits_0^{\pi/2-x} \sin(x+y)\,dy =$$

$$= \int\limits_0^{\pi/4} \cos(x+y)\Big|_0^x\,dx - \int\limits_{\pi/4}^{\pi/2} \cos(x+y)\Big|_0^{\pi/2-x}\,dx =$$

$$= \int\limits_0^{\pi/4} (\cos x - \cos 2x)\,dx + \int\limits_{\pi/4}^{\pi/2} \left(\cos x - \cos\frac{\pi}{2}\right)dx =$$

$$= (\sin x - \frac{1}{2}\sin 2x)\Big|_0^{\pi/4} + \sin x\Big|_{\pi/4}^{\pi/2} = \frac{1}{2}.$$

Замена переменных в двойном интеграле

Пусть функция $f(x,y)$ непрерывна в некоторой замкнутой ограниченной области D, так что существует двойной интеграл

$$\iint\limits_D f(x,y)\,dx\,dy. \tag{12.6}$$

Предположим, что возможен переход от переменных x и y к новым переменным u и v:

$$x = x(u,v), \quad y = y(u,v). \tag{12.7}$$

Будем полагать, что преобразование координат (12.7) является взаимно однозначным, т.е. u и v определяются из этих формул единственным образом при помощи *формул обратного преобразования*:

$$u = u(x,y), \quad v = v(x,y). \tag{12.8}$$

Справедлива следующая теорема, определяющая правило замены переменных в двойном интеграле; доказательство теоремы мы не приводим ввиду его сложности.

Теорема 12.4. *Пусть преобразование* (12.7) *переводит замкнутую ограниченную область D в замкнутую ограниченную область D_* и является взаимно однозначным. Если функции* (12.7) *имеют в D_* непрерывные частные производные первого порядка и отличный от нуля якобиан*

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}, \tag{12.9}$$

то при условии существования интеграла (12.6) *имеет место формула замены переменных:*

$$\iint\limits_D f(x,y)\,dx\,dy = \iint\limits_{D_*} f[x(u,v),y(u,v)]\left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv. \tag{12.10}$$

Замена переменных в двойном интеграле является одним из основных способов приведения интеграла к более удобному для вычисления виду.

Рассмотрим несколько примеров на замену переменных в двойном интеграле.

Пример 4. Обратимся еще раз к примеру 3. Как мы видели, непосредственное вычисление этого интеграла привело к разбиению области интегрирования на две и довольно сложным выкладкам. Более простой метод заключается в замене переменных, которая вводится следующим образом:

$$x + y = u, \quad x - y = v. \tag{12.11}$$

Выразим из этой системы уравнений x и y через u и v:

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

Теперь проведем все необходимые предварительные вычисления. Нетрудно видеть, что линии, ограничивающие область D, переходят: прямая $x + y = \pi/2$ — в прямую $u = \pi/2$, прямая $x - y = 0$ — в прямую $v = 0$, прямая $y = 0$ — в прямую $v = u$. Иными словами, исходная треугольная область D переходит в треугольную область D_*, которая более удобна для интегрирования (рис. 12.6). Модуль якобиана равен $1/2$, так

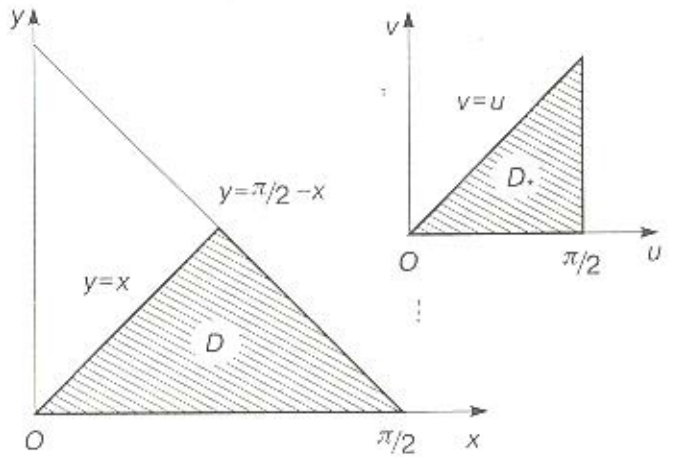

Рис. 12.6

что окончательно по формуле (12.10) получаем

$$\iint\limits_D \sin(x+y)\,dx\,dy = \frac{1}{2}\iint\limits_{D_*} \sin u\,du\,dv = \frac{1}{2}\int\limits_0^{\pi/2} du \int\limits_0^u \sin u\,dv =$$

$$= \frac{1}{2}\int\limits_0^{\pi/2} u\sin u\,du = -\frac{1}{2}u\cos u\,\Big|_0^{\pi/2} + \frac{1}{2}\int\limits_0^{\pi/2} \cos u\,du =$$

$$= \frac{1}{2}\sin u\,\Big|_0^{\pi/2} = \frac{1}{2}.$$

Пример 5. Найти $\iint\limits_D e^{x^2+y^2}\,dx\,dy$, где область D ограничена осью Ox и верхней полуокружностью $x^2+y^2=1$.

Перейдем к полярным координатам (см. параграф 7.7) по формулам

$$x = \rho\cos\varphi, y = \rho\sin\varphi, 0 \leqslant \varphi \leqslant \pi. \qquad (12.12)$$

Тогда область D переходит в прямоугольную область D_* (рис. 12.7), где $0 \leqslant \rho \leqslant 1$, $0 \leqslant \varphi \leqslant \pi$. При этом якобиан перехода от декартовых координат к полярным координатам равен

$$\frac{\partial(x,y)}{\partial(\rho,\varphi)} = \begin{vmatrix} \cos\varphi & -\rho\sin\varphi \\ \sin\varphi & \rho\cos\varphi \end{vmatrix} = \rho. \qquad (12.13)$$

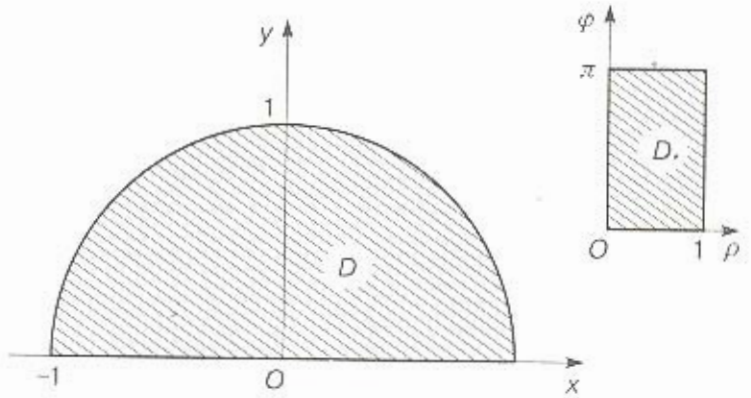

Рис. 12.7

Так как $x^2+y^2=\rho^2$, то из формулы (12.10)

$$\iint\limits_D e^{x^2+y^2}\,dx\,dy = \iint\limits_{D_*} \rho e^{\rho^2}\,d\rho\,d\varphi = \int\limits_0^{\pi} d\varphi \int\limits_0^1 \rho e^{\rho^2}\,d\rho =$$

$$= \int\limits_0^{\pi} \frac{1}{2}e^{\rho^2}\Big|_0^1 d\varphi = \frac{\pi}{2}(e-1).$$

Пример 6. Найти $\iint\limits_D (x^2+y^2)\,dx\,dy$, где область D ограничена линиями $x^2+y^2+2x-1=0$ и $x^2+y^2+2x=0$.

Переписав уравнения этих линий в форме

$$(x+1)^2+y^2 = 2 \text{ и } (x+1)^2+y^2 = 1,$$

нетрудно увидеть, что границами области D являются окружности радиусов $\sqrt{2}$ и 1 с центром в точке $(-1,0)$, т.е. область D представляет собой кольцо (рис. 12.8). Для вычисления данного интеграла перейдем к полярным координатам по формулам

$$x+1 = \rho\cos\varphi, \quad y = \rho\sin\varphi, \quad 0 \leqslant \varphi \leqslant 2\pi. \qquad (12.14)$$

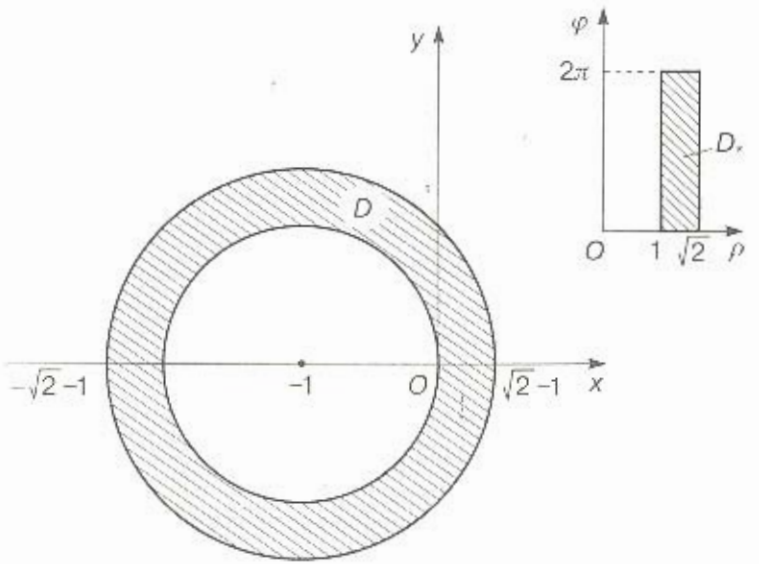

Рис. 12.8

Тогда кольцо D преобразуется в новых координатах в прямоугольник D_*, где $1 \leqslant \rho \leqslant \sqrt{2}$, $0 \leqslant \varphi \leqslant 2\pi$. Якобиан, как и в предыдущем примере, равен ρ; подставляя формулы (12.14) в подынтегральную функцию, получаем:

$$\iint\limits_D (x^2+y^2)\,dx\,dy = \int\limits_0^{2\pi} d\varphi \int\limits_1^{\sqrt{2}} (\rho^2 - 2\rho\cos\varphi + 1)\rho\,d\rho =$$

$$= \int\limits_0^{2\pi} (\rho^4/4 - 2p^3\cos\varphi/3 + \rho^2/2)\Big|_1^{\sqrt{2}}\, d\varphi =$$

$$= \int\limits_0^{2\pi} [5/4 - 2(2\sqrt{2}-1)\cos\varphi/3]\,d\varphi = \frac{5\pi}{2}.$$

Некоторые приложения двойного интеграла

Как и в случае определенного интеграла, двойной интеграл имеет различные приложения, прежде всего в геометрии и физике. Здесь мы приведем без вывода его геометрические приложения.

Вычисление объема. Как уже говорилось в параграфе 12.1, двойной интеграл равен объему криволинейного цилиндра, ограниченного сверху поверхностью $z = f(x,y) \geqslant 0$, снизу плоскостью Oxy, а также боковой цилиндрической поверхностью с образующей, параллельной оси Oz, и с направляющей в виде контура области интегрирования D, Так, все примеры вычисления двойных интегралов (примеры 4–6) можно рассматривать как вычисление объемов тел, ограниченных соответствующими поверхностями: сверху — графиками подынтегральных функций, снизу — областями интегрирования D, сбоку — цилиндрическими вертикальными поверхностями с направляющими по границам D.

Пример 7. Найти объем тела, ограниченного поверхностями

$$z = xy, x + y = a (a > 0), z = 0.$$

Поскольку подынтегральная функция должна быть неотрицательной, то область интегрирования ограничена на плоскости Oxy прямоугольным равнобедренным треугольником с

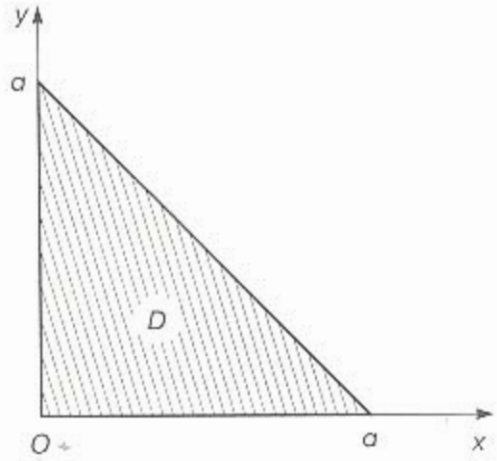

Рис. 12.9

катетами, равными a (рис. 12.9). Тогда искомый объем тела, ограниченного данными поверхностями, вычисляется как двойной интеграл

$$V = \iint\limits_D xy\,dx\,dy = \int\limits_0^a dx \int\limits_0^{a-x} xy\,dy = \frac{1}{2}\int\limits_0^a xy^2\Big|_0^{a-x} dx =$$

$$= \frac{1}{2}\int\limits_0^a (x^3 - 2ax^2 + a^2x)\,dx = \frac{1}{2}\left(\frac{x^4}{4} - \frac{2ax^3}{3} + \frac{a^2x^2}{2}\right)\Big|_0^a = \frac{a^4}{24}.$$

Вычисление площади. По геометрическому свойству 7° (см. параграф 12.1) площадь области D может быть вычислена двойным интегралом по формуле

$$S = \iint\limits_D dx\,dy \qquad (12.15)$$

Пример 8. Вычислить площадь фигуры, ограниченной линиями $y = \ln x$, $x - y = 1$, $y = -1$.

Нетрудно видеть, что первые две линии касаются друг друга в точке $(1,0)$, так что искомая площадь представляет собой разность площадей прямоугольного треугольника с катетами,

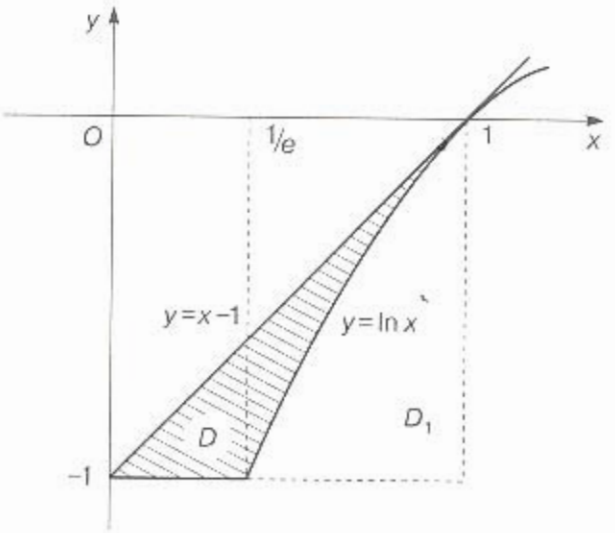

Рис. 12.10

равными единице, и фигуры D_1, ограниченной снизу прямой $y = -1$ при $1/e \leqslant x \leqslant 1$, сверху кривой $y = \ln x$ и справа — вертикальной прямой $x = 1$ (рис. 12.10). Отсюда

$$S = \frac{1}{2} - \iint\limits_{D_1} dx\,dy = \frac{1}{2} - \int\limits_{1/e}^{1} dx \int\limits_{-1}^{\ln x} dy = \frac{1}{2} - \int\limits_{1/e}^{1} (\ln x + 1)\,dx =$$

$$= \frac{1}{2} - x \ln x \Big|_{1/e}^{1} = \frac{1}{2} - \frac{1}{e}.$$

Площадь поверхности. Пусть поверхность P задана уравнением $z = f(x, y)$, причем проекцией P на плоскость Oxy является область D, в которой функция $f(x, y)$ непрерывна вместе со своими частными производными первого порядка. Тогда площадь S поверхности P равна двойному интегралу:

$$S = \iint\limits_D \sqrt{1 + [f'_x(x, y)]^2 + [f'_y(x, y)]^2}\,dx\,dy. \qquad (12.16)$$

Пример 9. Найти площадь части параболоида вращения $z = x^2 + y^2$, вырезаемой прямым цилиндром $x^2 + y^2 = 4$ (рис. 12.11).

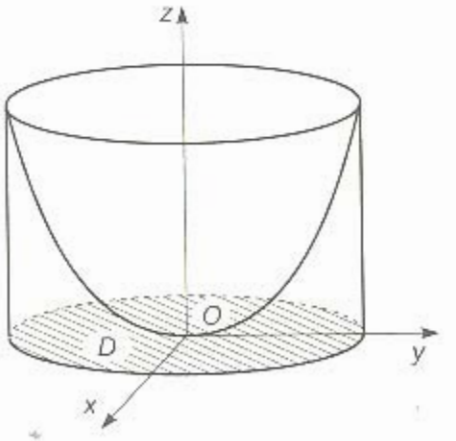

Рис. 12.11

Применяя формулу (12.16), имеем

$$S = \iint\limits_D \sqrt{1 + 4x^2 + 4y^2}\,dx\,dy,$$

где D — круг радиуса 2. Переходя к полярным координатам по формулам (12.12) при $0 \leqslant \varphi \leqslant 2\pi$, получим

$$S = \int\limits_0^{2\pi} d\varphi \int\limits_0^2 \sqrt{1 + 4\rho^2}\,\rho\,d\rho = \frac{1}{12} \int\limits_0^{2\pi} (1 + 4\rho^2)^{3/2} \Big|_0^2 d\varphi =$$

$$= \frac{\pi}{6}(17\sqrt{17} - 1).$$

Часть 2. ЭЛЕМЕНТЫ ЛИНЕЙНОЙ АЛГЕБРЫ

Глава 13
ВЕКТОРЫ

13.1. Векторное пространство

Понятие и основные свойства вектора

Ранее мы ввели в рассмотрение векторы в трехмерном евклидовом пространстве (см. параграф 9.5). Здесь мы приведем обобщение этих понятий на n-мерный случай.

Определение 1. Любой упорядоченный набор из n действительных чисел $a_1, a_2, \cdots, a_n$ называется *n-мерным вектором* $\bar{a}$; при этом числа, составляющие упомянутый набор, называются *координатами* (компонентами) вектора $\bar{a}$.

Определение 2. Совокупность всех n-мерных векторов называется *n-мерным векторным пространством* R^n.

Координаты n-мерного вектора $\bar{a}$ можно расположить либо в строку

$$\bar{a} = (a_1, a_2, \cdots, a_n), \tag{13.1}$$

либо в столбец

$$\bar{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}. \tag{13.2}$$

Запись вида (13.1) называется вектором-строкой, а вида (13.2) — вектором-столбцом.

Определение 3. Два вектора с одним и тем же числом координат $\bar{a} = (a_1, a_2, \cdots, a_n)$ и $\bar{b} = (b_1, b_2, \cdots, b_n)$ называются *равными*, если их соответствующие координаты равны, т.е. $a_1 = b_1, a_2 = b_2, \cdots, a_n = b_n$.

Определение 4. Вектор, все координаты которого равны нулю, называется *нулевым*

$$\bar{0} = (0, 0, \cdots, 0).$$

Операции над векторами

Пусть векторы $\bar{a}$ и $\bar{b}$ принадлежат n-мерному векторному пространству

$$\bar{a} = (a_1, a_2, \cdots, a_n), \qquad \bar{b} = (b_1, b_2, \cdots, b_n). \tag{13.3}$$

Будем называть *суммой векторов* $\bar{a}$ и $\bar{b}$ вектор $\bar{c}$, координаты которого равны суммам соответствующих координат этих векторов:

$$\bar{c} = \bar{a} + \bar{b} = (a_1 + b_1, a_2 + b_2, \cdots, a_n + b_n). \tag{13.4}$$

Пусть λ — любое действительное число. *Произведением вектора $\bar{a}$ на число λ* будем называть вектор, координаты которого получаются умножением соответствующих координат вектора $\bar{a}$ на это число:

$$\bar{c} = \lambda\bar{a} = (\lambda a_1, \lambda a_2, \cdots, \lambda a_n). \tag{13.5}$$

Из введенных таким образом операций над векторами вытекают следующие свойства этих операций. Пусть $\bar{a}$, $\bar{b}$ и $\bar{c}$ — произвольные векторы n-мерного векторного пространства. Тогда:

1°. $\bar{a} + \bar{b} = \bar{b} + \bar{a}$ — *переместительное свойство.*

2°. $(\bar{a} + \bar{b}) + \bar{c} = \bar{a} + (\bar{b} + \bar{c})$ — *сочетательное свойство.*

3°. $\lambda(\bar{a} + \bar{b}) = \lambda\bar{a} + \lambda\bar{b}$, где λ — *действительное число.*

4°. $(\lambda + \mu)\bar{a} = \lambda\bar{a} + \mu\bar{a}$, где λ и μ — *действительные числа.*

5°. $\lambda(\mu\bar{a}) = (\lambda\mu)\bar{a}$, где λ и μ — *действительные числа.*

6°. $\bar{a} + \bar{0} = \bar{a}$.

7°. *Для любого вектора $\bar{a}$ существует такой вектор $-\bar{a}$, что $-\bar{a} = (-1)\bar{a}$, $\bar{a} + (-\bar{a}) = \bar{0}$.*

8°. $0\bar{a} = \bar{0}$ *для любого вектора* $\bar{a}$.

Скалярное произведение векторов

Определение 5. *Скалярным произведением векторов* $\bar{a}$ и $\bar{b}$ (13.3) называется число, состоящее из суммы произведений соответствующих координат этих векторов:

$$\bar{a}\bar{b} = a_1b_1 + a_2b_2 + \cdots + a_nb_n. \tag{13.6}$$

Как мы видим, формально такое определение скалярного произведения двух векторов согласуется с аналогичной величиной для двух- и трехмерных векторов. Из данного определения следуют основные свойства скалярного произведения векторов:

1°. $\bar{a}\bar{b} = \bar{b}\bar{a}$.

2°. $(\lambda\bar{a})\bar{b} = \bar{a}(\lambda\bar{b}) = \lambda(\bar{a}\bar{b})$, *где* λ — *действительное число.*

3°. $\bar{a}(\bar{b} + \bar{c}) = \bar{a}\bar{b} + \bar{a}\bar{c}$.

4°. $\bar{a}\bar{a} > 0$, если $\bar{a} \neq \bar{0}$, и $\bar{a}\bar{a} = \bar{0}$, если $\bar{a} = \bar{0}$.

Из параграфа 9.5 становится ясно, как ввести понятие модуля вектора (его длины) и угла между векторами в виде обобщения на случай при $n > 3$.

Определение 6. Для векторов из n-мерного векторного пространства *модуль вектора* $\bar{a}$ и *угол* φ *между двумя ненулевыми векторами* $\bar{a}$ и $\bar{b}$ определяются по формулам:

$$|\bar{a}| = \sqrt{\bar{a}\bar{a}} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} = \sqrt{\sum_{i=1}^{n} a_i^2}, \tag{13.7}$$

$$\cos\varphi = \frac{\bar{a}\bar{b}}{|\bar{a}||\bar{b}|}. \tag{13.8}$$

Необходимое условие для формулы (13.8), что $|\cos\varphi| \leqslant 1$, гарантируется неравенством Коши–Буняковского, справедливого для любых двух векторов $\bar{a}$ и $\bar{b}$:

$$(\bar{a}\bar{b})^2 \leqslant (\bar{a}\bar{a})(\bar{b}\bar{b}). \tag{13.9}$$

Докажем это неравенство. Возьмем вектор $\bar{c} = x\bar{a} + \bar{b}$, где x — любое число. Согласно свойству 4°, имеем $\bar{c}\bar{c} > 0$, откуда в силу свойств 1°–3°

$$(x\bar{a} + \bar{b})(x\bar{a} + \bar{b}) = (\bar{a}\bar{a})x^2 + 2(\bar{a}\bar{b})x + (\bar{b}\bar{b}) > 0.$$

Поскольку квадратный трехчлен в правой части этого равенства всегда неотрицателен, то его дискриминант должен быть неположительным, т.е. $(\bar{a}\bar{b})^2 - (\bar{a}\bar{a})(\bar{b}\bar{b}) \leqslant 0$, откуда и получаем требуемое неравенство.

Введем одно важное свойство векторов. Векторы $\bar{a}$ и $\bar{b}$ будем называть *ортогональными*, если их скалярное произведение равно нулю:

$$\bar{a}\bar{b} = 0. \tag{13.10}$$

Равенство (13.10) является аналогом условия перпендикулярности векторов в двух- и трехмерном случаях, когда в равенстве (13.8) $\cos\varphi = 0$.

13.2. Линейная зависимость векторов

Линейно зависимые и линейно независимые системы векторов

При решении различных задач, как правило, приходится иметь дело не с одним вектором, а с некоторой совокупностью векторов одной и той же размерности. Такие совокупности называют *системой векторов* и обозначают одной буквой и порядковым номером:

$$\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_k. \tag{13.11}$$

Определение 1. *Линейной комбинацией векторов* (13.11) называется вектор вида

$$\bar{b} = \lambda_1\bar{a}_1 + \lambda_2\bar{a}_2 + \cdots + \lambda_k\bar{a}_k, \tag{13.12}$$

где $\lambda_1, \lambda_2, \cdots, \lambda_k$ — любые действительные числа.

Например, пусть даны три вектора: $\bar{a}_1 = (1, 2, 0)$, $\bar{a}_2 = (2, 1, 1)$ и $\bar{a}_3 = (-1, 1, -2)$. Их линейной комбинацией с коэффициентами соответственно 2, 3 и 4 является вектор $\bar{b} = (4, 11, -5)$.

В случае равенства (13.12) говорят также, что вектор $\bar{b}$ *линейно выражается* через векторы (13.11) или *разлагается* по этим векторам.

Определение 2. Система ненулевых векторов (13.11) называется *линейно зависимой*, если существуют такие числа $\lambda_1, \lambda_2, \cdots, \lambda_k$, не равные одновременно нулю, что линейная комбинация данной системы с указанными числами равна нулевому вектору:

$$\lambda_1 \bar{a}_1 + \lambda_2 \bar{a}_2 + \cdots + \lambda \bar{a}_k = \bar{0}. \qquad (13.13)$$

Если же равенство (13.13) для данной системы векторов (13.11) возможно лишь при $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$, то эта система векторов называется *линейно независимой*.

Например, система двух векторов $\bar{a}_1 = (1, 0)$ и $\bar{a}_2 = (0, 2)$ линейно независима; система двух векторов $\bar{b}_1 = (1, 2, 1)$ и $\bar{b}_2 = (2, 4, 2)$ линейно зависима, так как $\bar{b}_2 - 2\bar{b}_1 = \bar{0}$.

Пусть система векторов (13.11) линейно зависима. Выберем в сумме (13.13) слагаемое, в котором коэффициент $\lambda_s \neq 0$, и выразим его через остальные слагаемые:

$$\bar{a}_s = -\frac{\lambda_1}{\lambda_s}\bar{a}_1 - \frac{\lambda_2}{\lambda_s}\bar{a}_2 - \cdots - \frac{\lambda_{s-1}}{\lambda_s}\bar{a}_{s-1} - \frac{\lambda_{s+1}}{\lambda_s}\bar{a}_{s+1} - \cdots - \frac{\lambda_k}{\lambda_s}\bar{a}_k.$$

Как видно из этого равенства, один из векторов линейно зависимой системы (13.11) оказался выраженным через другие векторы этой системы (или разлагается по остальным ее векторам).

Укажем свойства линейно зависимой системы векторов.

1°. *Система, состоящая из одного ненулевого вектора, линейно независима.*

2°. *Система, содержащая нулевой вектор, всегда линейно зависима.*

3°. *Система, содержащая более одного вектора, линейно зависима тогда и только тогда, когда среди ее векторов содержится по крайней мере один вектор, который линейно выражается через остальные.*

Геометрический смысл линейной зависимости векторов очевиден для случаев двумерных векторов на плоскости и трехмерных векторов в пространстве. В случае двух векторов, когда один вектор выражается через другой,

$$\bar{a}_1 = \lambda \bar{a}_2,$$

т.е. эти векторы *коллинеарны* или, что то же самое, находятся на параллельных прямых (см. рис. 9.7). В пространственном случае линейной зависимости трех векторов они параллельны одной плоскости, т.е. *компланарны* (рис. 13.1, а); достаточно "подправить" соответствующими сомножителями длины этих векторов, чтобы один из них стал суммой двух других или выражался через них (рис. 13.1, б).

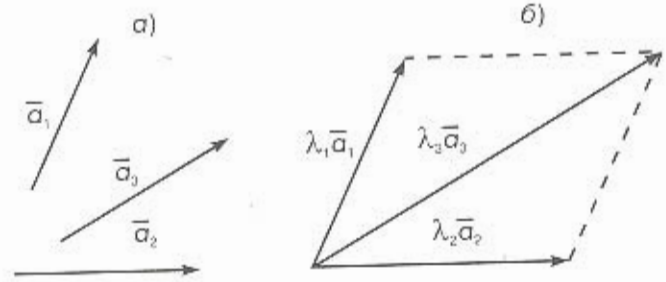

Рис. 13.1

Справедлива следующая теорема, которую мы приводим без доказательства.

Теорема 13.1. *В пространстве R^n любая система, содержащая m векторов, линейно зависима при $m > n$.*

Базис и ранг системы векторов

Рассмотрим систему векторов (13.11)

$$\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_k.$$

Максимально независимой подсистемой системы векторов (13.11) называется частичный набор векторов этой системы, удовлетворяющий двум условиям:

а) векторы этого набора линейно независимы;

б) любой вектор системы (13.11) линейно выражается через векторы этого набора.

Справедлива теорема, утверждающая, что все максимально независимые подсистемы данной системы векторов содержат одно и то же число векторов. Максимально независимая подсистема системы векторов (13.11) называется ее *базисом*; векторы, входящие в базис, называются *базисными векторами*. Будем называть *рангом* системы векторов число векторов ее базиса. Понятно, что если ранг системы векторов меньше числа k ее векторов, то она может иметь несколько базисов.

В силу теоремы 13.1 всякая система векторов пространства R^n содержащая более n векторов, является линейно зависимой.

Понятие базиса распространяется и на пространство R^n, которое является системой, содержащей всю бесконечную совокупность n-мерных векторов.

Определение 3. Система векторов называется *базисом пространства* R^n, если:

1) векторы этой системы линейно независимы;
2) всякий вектор из R^n линейно выражается через векторы данной системы.

Справедлива основная теорема о базисе пространства R^n которую мы приводим без доказательства.

Теорема 13.2. *Линейно независимая система векторов в* R^n *является базисом тогда и только тогда, когда число векторов этой системы равно* n.

13.3. Разложение вектора по базису

Представление вектора в произвольном базисе

Пусть система векторов

$$\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_m \tag{13.14}$$

является базисом, а вектор $\bar{b}$ — их линейная комбинация. Имеет место следующая теорема.

Теорема 13.3. *Разложение любого вектора в базисе, если оно существует, является единственным.*

Доказательство. Предположим, что вектор $\bar{b}$ может быть представлен в виде линейной комбинации векторов (13.14) двумя способами: $\bar{b} = \alpha_1\bar{a}_1 + \alpha_2\bar{a}_2 + \cdots + \alpha_m\bar{a}_m$ и $\bar{b} = \beta_1\bar{a}_1 + \beta_2\bar{a}_2 + \cdots + \beta_m\bar{a}_m$, где наборы чисел α_i и β_i, среди которых обязательно есть ненулевые значения, не совпадают. Вычитая одно равенство из другого, имеем

$$(\alpha_1 - \beta_1)\bar{a}_1 + (\alpha_2 - \beta_2)\bar{a}_2 + \cdots + (\alpha_m - \beta_m)\bar{a}_m = \bar{0}.$$

Мы получили, что линейная комбинация векторов системы (13.14), в которой коэффициенты не все равны нулю (в силу несовпадения α_i и β_i), равна нулю, т.е. данная система оказалась линейно зависимой, что противоречит условию теоремы. Полученное противоречие доказывает теорему.

Стало быть, в произвольном базисе пространства R^n

$$\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_n \tag{13.15}$$

любой вектор этого пространства обязательно представим в виде разложения по базисным векторам

$$\bar{b} = \alpha_1\bar{a}_1 + \alpha_2\bar{a}_2 + \cdots + \alpha_n\bar{a}_n, \tag{13.16}$$

причем это разложение является единственным для данного базиса. В наборе коэффициентов разложения

$$\alpha_1, \alpha_2, \cdots, \alpha_n \tag{13.17}$$

числа α_i, называются *координатами* вектора $\bar{b}$ в базисе (13.15), и, как следует из сказанного, этот набор единственный для любого вектора в данном базисе.

Задача нахождения коэффициентов разложения α_i в случае произвольного базиса (13.15) является, вообще говоря, непростой. Нужно приравнять соответствующие координаты линейной комбинации векторов слева и координаты вектора $\bar{b}$. Пусть

базисные векторы и вектор $\bar{b}$ заданы в следующей координатной форме:

$$\begin{aligned} \bar{a}_1 &= (a_{11}, a_{12}, \cdots, a_{1n}), \\ \bar{a}_2 &= (a_{21}, a_{22}, \cdots, a_{2n}), \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \bar{a}_n &= (a_{n1}, a_{n2}, \cdots, a_{nn}), \\ \bar{b} &= (b_1, b_2, \cdots, b_n). \end{aligned} \tag{13.18}$$

Выполнение процедуры, описанной выше, приводит к *системе n линейных уравнений* относительно n неизвестных координат α_i разложения вектора $\bar{b}$ в базисе (13.15):

$$\begin{cases} a_{11}\alpha_1 + a_{12}\alpha_2 + \cdots + a_{1n}\alpha_n = b_1, \\ a_{21}\alpha_1 + a_{22}\alpha_2 + \cdots + a_{2n}\alpha_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}\alpha_1 + a_{n2}\alpha_2 + \cdots + a_{nn}\alpha_n = b_n. \end{cases}$$

Такие системы уравнений и методы их решения представляют отдельные разделы линейной алгебры; они будут рассмотрены в следующих разделах.

Разложение вектора в ортогональном базисе

Рассмотрим базис пространства R^n, в котором каждый вектор ортогонален остальным векторам базиса:

$$\bar{e}_1, \bar{e}_2, \cdots, \bar{e}_n; \quad \bar{e}_i\bar{e}_j = 0, \quad i \neq j, \quad i, j = 1, 2, \cdots, n. \tag{13.19}$$

Ортогональные базисы известны и хорошо представимы на плоскости и в пространстве (рис. 13.2). Базисы такого вида удобны прежде всего тем, что координаты разложения произвольного вектора определяются весьма просто, без применения трудоемких вычислений.

Действительно, пусть требуется найти разложение произвольного вектора $\bar{b}$ в ортогональном базисе (13.19). Составим разложение этого вектора с неизвестными пока координатами разложения в данном базисе:

$$\bar{b} = \alpha_1\bar{e}_1 + \alpha_2\bar{e}_2 + \cdots + \alpha_n\bar{e}_n. \tag{13.20}$$

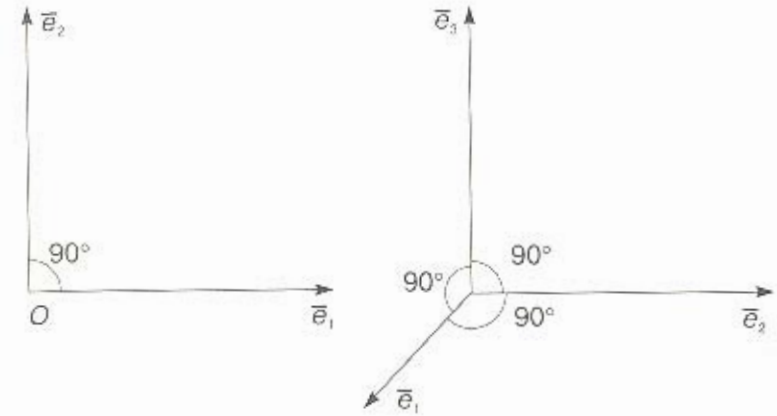

Рис. 13.2

Умножим обе части этого равенства, представляющие собой векторы, на вектор $\bar{e}_1$. В силу свойств 2° и 3° скалярного произведения векторов (см. параграф 13.1)

$$\bar{b}\bar{e}_1 = \alpha_1(\bar{e}_1\bar{e}_1) + \alpha_2(\bar{e}_2\bar{e}_1) + \cdots + \alpha_n(\bar{e}_n\bar{e}_1).$$

Однако в силу взаимной ортогональности векторов базиса (13.19) все скалярные произведения векторов базиса, за исключением первого, равны нулю, т.е. коэффициент α_1 определяется по формуле

$$\alpha_1 = \frac{\bar{b}\bar{e}_1}{\bar{e}_1\bar{e}_1} = \frac{\bar{b}\bar{e}_1}{|\bar{e}_1|^2}.$$

Умножая поочередно равенство (13.20) на другие базисные векторы, мы получаем простую формулу для вычисления коэффициентов разложения вектора $\bar{b}$:

$$\alpha_i = \frac{\bar{b}\bar{e}_i}{\bar{e}_i\bar{e}_i} = \frac{\bar{b}\bar{e}_i}{|\bar{e}_i|^2}, \qquad i = 1, 2, \cdots, n. \tag{13.21}$$

Нетрудно видеть, что соотношения (13.21) имеют смысл, поскольку $|\bar{e}_i|^2 \neq 0$.

Отметим особо частный случай ортогонального базиса, когда все векторы в (13.19) имеют единичную длину ($|\bar{e}_i| = 1$)

или *нормированы* по своей длине. В таком случае базис называют *ортонормированным* и координаты разложения (13.21) имеют наиболее простой вид:

$$a_i = \bar{b}\bar{c}_i, \qquad i = 1, 2, \cdots, n. \tag{13.22}$$

Глава 14
МАТРИЦЫ

14.1. Матрицы и операции над ними

Понятие матрицы

Определение 1. Прямоугольная таблица чисел вида

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{14.1}$$

называется *матрицей*. Здесь a_{ij} — действительные числа ($i = 1, 2, \cdots, m, j = 1, 2, \cdots, n$), называемые *элементами* матрицы, i и j — соответственно индексы строки и столбца. При этом произведение $m \times n$ числа строк на число столбцов называют *размером* матрицы $\mathbf{A}$. Часто матрицу (14.1) записывают в сокращенном виде

$$\mathbf{A} = \|a_{ij}\|, \qquad i = 1, 2, \cdots, m, \quad j = 1, 2, \cdots, n. \tag{14.2}$$

Матрица, все элементы которой равны нулю, называется *нулевой* матрицей. В том случае, когда $m = n$ (число строк равно числу столбцов)

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \tag{14.3}$$

матрица $\mathbf{A}$ называется *квадратной*.

Упорядоченная совокупность элементов $a_{11}, a_{22}, \cdots, a_{nn}$ называется *главной диагональю* квадратной матрицы. Квадратная матрица называется *диагональной*, если ее элементы удовлетворяют условию

$$a_{ij} = \begin{cases} a_{ij} \neq 0, & i = j, \\ a_{ij} = 0, & i \neq j, \end{cases}$$

т.е. ненулевыми могут быть только элементы главной диагонали, и матрица имеет вид

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}. \tag{14.4}$$

Единичной матрицей называется диагональная матрица, у которой все элементы главной диагонали равны единице:

$$\mathbf{E} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{14.5}$$

Определение 2. Две матрицы $\mathbf{A}$ и $\mathbf{B}$ называются *равными* ($\mathbf{A} = \mathbf{B}$), если они имеют одинаковые размеры и их соответствующие элементы равны.

Линейные операции над матрицами

Сумма матриц. *Суммой матриц* $\mathbf{A}$ и $\mathbf{B}$ одинакового размера называется матрица $\mathbf{C}$ того же размера, каждый элемент которой равен сумме соответствующих элементов матриц $\mathbf{A}$ и $\mathbf{B}$. Представим это в сокращенной записи. Пусть

$$\mathbf{A} = \|a_{ij}\|, \quad \mathbf{B} = \|b_{ij}\|; \quad i = 1, 2, \cdots, m, \quad j = 1, 2, \cdots, n.$$

Тогда сумма этих матриц $\mathbf{C} = \mathbf{A} + \mathbf{B}$ имеет вид

$$\mathbf{C} = \|c_{ij}\|, \quad c_{ij} = a_{ij} + b_{ij}; \quad i = 1, 2, \cdots, m, \quad j = 1, 2, \cdots, n.$$

27 Зак 6853

Пример 1. Пусть даны матрицы $\mathbf{A}$ и $\mathbf{B}$:

$$\mathbf{A} = \begin{pmatrix} -2 & 3 & 3 & 2 \\ 0 & -1 & 2 & 1 \\ 3 & 4 & -3 & 5 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 2 & 1 & -5 & 6 \\ 3 & 2 & 4 & -1 \\ 2 & -3 & 5 & -2 \end{pmatrix}.$$

Тогда их суммой, согласно определению, является матрица

$$\mathbf{C} = \begin{pmatrix} 0 & 4 & -2 & 8 \\ 3 & 1 & 6 & 0 \\ 5 & 1 & 2 & 3 \end{pmatrix}.$$

Умножение матрицы на действительное число. *Произведением матрицы* $\mathbf{A}$ *на действительное число* α называется матрица, каждый элемент которой получен умножением соответствующего элемента матрицы $\mathbf{A}$ на число α.

Пример 2. Пусть даны матрица $\mathbf{A}$ и число α:

$$\mathbf{A} = \begin{pmatrix} -1 & 2 & 3 & 5 \\ 8 & 4 & -2 & 7 \end{pmatrix}, \qquad \alpha = 3.$$

Тогда произведением матрицы $\mathbf{A}$ на число α является матрица

$$\mathbf{C} = \begin{pmatrix} -3 & 6 & 9 & 15 \\ 24 & 12 & -6 & 21 \end{pmatrix}.$$

Свойства операций суммирования матриц и произведения матрицы на число. Приведем свойства, непосредственно вытекающие из определения этих операций. Пусть $\mathbf{A}$, $\mathbf{B}$ и $\mathbf{C}$ — матрицы, имеющие одинаковый размер, а α и β — некоторые действительные числа. Тогда:

1°. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.

2°. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$.

3°. $\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$.

4°. $(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$.

5°. $(\alpha\beta)\mathbf{A} = (\alpha\mathbf{A})\beta$.

6°. $\mathbf{A} + \mathbf{O} = \mathbf{A}$, *где* $\mathbf{O}$ — *нулевая матрица*.

7°. $0\mathbf{A} = \mathbf{O}$.

Транспонирование матриц

Транспонированием матрицы называется замена строк матрицы на ее столбцы с сохранением их порядка (или, что то же самое, замена столбцов матрицы на ее строки).

Пусть дана исходная матрица $\mathbf{A}$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Тогда, согласно определению, транспонированная матрица $\mathbf{A}'$ имеет вид

$$\mathbf{A}' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}. \tag{14.6}$$

Сокращенная форма записи операции транспонирования матрицы:

$$\begin{gathered} \mathbf{A} = \|a_{ij}\|, \quad \mathbf{A}' = \|a_{ji}\|; \\ i = 1, 2, \cdots, m, \quad j = 1, 2, \cdots, n. \end{gathered} \tag{14.7}$$

Пример 3. Пусть даны матрицы $\mathbf{A}$ и $\mathbf{B}$

$$\mathbf{A} = \begin{pmatrix} 6 & 3 & 1 \\ 2 & 4 & -2 \\ 5 & 7 & 0 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 2 & 1 & 0 & 3 \\ 5 & 2 & 4 & 7 \end{pmatrix}.$$

Тогда соответствующие транспонированные матрицы имеют вид

$$\mathbf{A}' = \begin{pmatrix} 6 & 2 & 5 \\ 3 & 4 & 7 \\ 1 & -2 & 0 \end{pmatrix}, \qquad \mathbf{B}' = \begin{pmatrix} 2 & 5 \\ 1 & 2 \\ 0 & 4 \\ 3 & 7 \end{pmatrix}.$$

Нетрудно заметить две закономерности операции транспонирования матриц.

1. Дважды транспонированная матрица равна исходной матрице:

$$\mathbf{A}'' = \mathbf{A}.$$

2. При транспонировании квадратных матриц элементы, находящиеся на главной диагонали, не меняют своих позиций, т.е. главная диагональ квадратной матрицы не меняется при транспонировании.

Важную роль в алгебре и ее приложениях играют *симметрические* матрицы — квадратные матрицы, у которых элементы, симметричные относительно главной диагонали, равны, т.е. $a_{ij} = a_{ji}$. Транспонирование таких матриц не меняет их вида, так что равенство

$$\mathbf{A} = \mathbf{A}' \tag{14.8}$$

также можно полагать определением симметрической матрицы.

Умножение матриц

Умножение матриц — это специфическая операция, составляющая основу алгебры матриц. Строки и столбцы матриц можно рассматривать как векторы-строки и векторы-столбцы соответствующих размерностей, иными словами, любую матрицу можно интерпретировать как совокупность векторов-строк или векторов-столбцов.

Пусть даны две матрицы: $\mathbf{A}$ — размера $m \times n$ и $\mathbf{B}$ — размера $n \times k$. Будем рассматривать матрицу $\mathbf{A}$ как совокупность m векторов-строк $\bar{a}_i$ размерности n каждый, а матрицу $\mathbf{B}$ — как совокупность k векторов-столбцов $\bar{b}_j$, содержащих по n координат каждый:

$$\mathbf{A} = \begin{matrix} \bar{a}_1 \\ \bar{a}_2 \\ \cdots \\ \bar{a}_m \end{matrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{B} = \begin{matrix} \bar{b}_1 & \bar{b}_2 & \cdots & \bar{b}_k \end{matrix} \\ \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nk} \end{pmatrix}. \tag{14.9}$$

Векторы-строки матрицы $\mathbf{A}$ и векторы-столбцы матрицы $\mathbf{B}$ показаны в записи этих матриц (14.9). Длина строки матрицы $\mathbf{A}$ равна высоте столбца матрицы $\mathbf{B}$, и потому скалярное произведение этих векторов имеет смысл.

Определение 3. *Произведением матриц* $\mathbf{A}$ и $\mathbf{B}$ называется матрица $\mathbf{C}$, элементы которой c_{ij} равны скалярным произведениям векторов-строк $\bar{a}_i$ матрицы $\mathbf{A}$ на векторы-столбцы $\bar{b}_j$ матрицы $\mathbf{B}$:

$$\mathbf{C} = \mathbf{AB} = \|c_{ij}\|, \quad c_{ij} = \bar{a}_i \bar{b}_j = \sum_{s=1}^{n} a_{is} b_{sj}, \tag{14.10}$$
$$i = 1, 2, \cdots, m, \quad j = 1, 2, \cdots, k.$$

Произведение матриц $\mathbf{A}$ и $\mathbf{B}$ — матрица $\mathbf{C}$ — имеет размер $m \times k$, поскольку длина n векторов-строк и векторов-столбцов исчезает при суммировании скалярных произведений координат этих векторов, как показано в формулах (14.10). Таким образом, для вычисления элементов первой строки матрицы $\mathbf{C}$ необходимо последовательно получить скалярные произведения первой строки матрицы $\mathbf{A}$ на все столбцы матрицы $\mathbf{B}$; вторая строка матрицы $\mathbf{C}$ получается как скалярные произведения второго вектора-строки матрицы $\mathbf{A}$ на все векторы-столбцы матрицы $\mathbf{B}$, и так далее. Для удобства запоминания размера произведения матриц нужно перемножить отношения размеров матриц-сомножителей: $\frac{m}{n}\frac{n}{k} = \frac{m}{k}$, т.е. размер матрицы $\mathbf{C}$ равен $m \times k$.

В операции умножения матриц есть характерная особенность: произведение матриц $\mathbf{A}$ и $\mathbf{B}$ имеет смысл, если число столбцов в $\mathbf{A}$ равно числу строк в $\mathbf{B}$. Тогда, если $\mathbf{A}$ и $\mathbf{B}$ — прямоугольные матрицы, произведение матриц $\mathbf{B}$ и $\mathbf{A}$ может не иметь смысла, так как в скалярных произведениях, формирующих элементы соответствующей матрицы, должны участвовать векторы с одинаковым числом координат.

Если матрицы $\mathbf{A}$ и $\mathbf{B}$ квадратные, размера $n \times n$, имеют смысл как произведение матриц $\mathbf{AB}$, так и произведение матриц $\mathbf{BA}$, причем размер этих матриц такой же, как и у исходных сомножителей. При этом в общем случае перемножения матриц правило перестановочности не соблюдается, т. е. $\mathbf{AB} \neq \mathbf{BA}$.

Рассмотрим примеры на умножение матриц.

Пример 4.

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 2 \\ 0 & 2 & 3 \\ 1 & 4 & 1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & -2 \end{pmatrix}.$$

Поскольку число столбцов матрицы $\mathbf{A}$ равно числу строк матрицы $\mathbf{B}$, то произведение матриц $\mathbf{AB}$ имеет смысл. По формулам (14.10) получаем в произведении матрицу размера 3×2:

$$\mathbf{AB} = \begin{pmatrix} 0-1+4 & 1-1-4 \\ 0+2+6 & 0+2-6 \\ 0+4+2 & 1+4-2 \end{pmatrix} = \begin{pmatrix} 3 & -4 \\ 8 & -4 \\ 6 & 3 \end{pmatrix}.$$

Произведение $\mathbf{BA}$ не имеет смысла, так как число столбцов матрицы $\mathbf{B}$ не совпадает с числом строк матрицы $\mathbf{A}$.

Пример 5.

$$\mathbf{A} = \begin{pmatrix} 2 & 3 \\ 4 & -1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 5 & -2 \end{pmatrix}.$$

Здесь мы найдем произведения данных матриц $\mathbf{AB}$ и $\mathbf{BA}$:

$$\mathbf{AB} = \begin{pmatrix} 2 & 3 \\ 4 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 5 & -2 \end{pmatrix} = \begin{pmatrix} 2+15 & 0-6 \\ 4-5 & 0+2 \end{pmatrix} = \begin{pmatrix} 17 & -6 \\ -1 & 2 \end{pmatrix};$$

$$\mathbf{BA} = \begin{pmatrix} 1 & 0 \\ 5 & -2 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 4 & -1 \end{pmatrix} = \begin{pmatrix} 2+0 & 3+0 \\ 10-8 & 15+2 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 2 & 17 \end{pmatrix}.$$

Как видно из результата, матрица произведения зависит от порядка матриц в произведении. В обоих случаях произведения матриц имеют тот же размер, что и у исходных сомножителей: 2×2.

Пример 6.

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & -2 \\ 0 & 3 & 1 \end{pmatrix}, \qquad \mathbf{B} = \bar{b} = \begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix}.$$

В данном случае матрица $\mathbf{B}$ представляет собой вектор-столбец, т. е. матрицу, у которой три строки и один столбец. Вообще, векторы — это частные случаи матриц: вектор-строка длины n представляет собой матрицу с одной строкой

и n столбцами, а вектор-столбец высоты n — матрицу с n строками и одним столбцом. Размеры данных матриц соответственно 2×3 и 3×1, так что произведение этих матриц определено. Имеем

$$\mathbf{AB} = \mathbf{A}\bar{b} = \begin{pmatrix} 1 & 2 & -2 \\ 0 & 3 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix} = \begin{pmatrix} 2-2-10 \\ 0-3+5 \end{pmatrix} = \begin{pmatrix} -10 \\ 2 \end{pmatrix}.$$

В произведении получена матрица размера 2×1 или вектор-столбец высоты 2.

Пример 7. $\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 1 \\ 1 & 3 & 2 \end{pmatrix}$.

Найти матрицу $\mathbf{A}^3$.

Путем последовательного умножения матриц находим

$$\mathbf{A}^3 = \mathbf{A}^2\mathbf{A} = (\mathbf{AA})\mathbf{A} = \begin{pmatrix} 1+0+2 & 0+0+6 & 2+0+4 \\ 2+2+1 & 0+1+3 & 4+1+2 \\ 1+6+2 & 0+3+6 & 2+3+4 \end{pmatrix} \times$$

$$\times \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 1 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 6 & 6 \\ 5 & 4 & 7 \\ 9 & 9 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 1 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 21 & 24 & 24 \\ 20 & 25 & 28 \\ 36 & 36 & 45 \end{pmatrix}.$$

Свойства произведения матриц. Пусть $\mathbf{A}$, $\mathbf{B}$ и $\mathbf{C}$ — матрицы соответствующих размеров (чтобы произведения матриц были определены), а α — действительное число. Тогда имеют место следующие свойства произведения матриц:

1°. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

2°. $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$.

3°. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

4°. $\alpha(\mathbf{AB}) = (\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\alpha\mathbf{B})$.

Нетрудно убедиться, что в алгебре квадратных матриц единичная матрица $\mathbf{E}$ играет роль единицы, т.е. можно отметить

еще два свойства, связанных с умножением на эту матрицу слева и справа:

5°. $\mathbf{AE} = \mathbf{A}$.

6°. $\mathbf{EA} = \mathbf{A}$.

Иными словами, произведение любой матрицы на единичную матрицу, если оно имеет смысл, не меняет исходную матрицу.

Собственные значения и собственные векторы матрицы

Рассмотрим квадратные матрицы размера $n \times n$ или, что то же самое, матрицы порядка n.

При умножении матрицы порядка n на n-мерный вектор в произведении получается n-мерный вектор:

$$\mathbf{A}\bar{x} = \bar{b}.$$

Однако для любой матрицы существует набор особых векторов, таких, что произведение матрицы на вектор из такого набора равносильно умножению этого вектора на определенное число.

Определение 4. Число λ называется *собственным значением матрицы* $\mathbf{A}$ порядка n, если существует такой ненулевой вектор $\bar{x} \in R^n$, что выполняется равенство

$$\mathbf{A}\bar{x} = \lambda\bar{x}. \qquad (14.11)$$

При этом вектор $\bar{x}$ называется *собственным вектором матрицы* $\mathbf{A}$.

Иными словами, произведение матрицы на ее собственный вектор коллинеарно этому вектору, причем его длина меняется в $|\lambda|$ раз.

Уравнение (14.11) представлено в матричной форме. Группируя все слагаемые этого уравнения в левой части, его можно переписать в более удобном виде:

$$(\mathbf{A} - \lambda\mathbf{E})\bar{x} = \bar{0}, \qquad (14.12)$$

где $\mathbf{E}$ и $\bar{0}$ — соответственно единичная матрица и нулевой вектор.

Если a_{ij} — элементы матрицы $\mathbf{A}$, то *характеристическая* матрица $\mathbf{A} - \lambda\mathbf{E}$, согласно определениям умножения матрицы на число и суммы матриц, имеет вид

$$\mathbf{A} - \lambda\mathbf{E} = \begin{pmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{pmatrix}. \qquad (14.13)$$

Проблема отыскания собственных значений и собственных векторов матриц составляет основу специального раздела алгебры; в дальнейшем мы еще вернемся к этому вопросу. Здесь лишь отметим один важный результат алгебры матриц: для симметрических матриц (14.8) все n собственных значений являются действительными числами.

14.2. Обратная матрица

Ранг матрицы

Выше уже говорилось, что матрицы размера $m \times n$ можно рассматривать как системы, состоящие из m n-мерных векторов (или из n m-мерных векторов). Поскольку любая система векторов характеризуется рангом (см. параграф 13.2), то естественно встает вопрос о такой же характеристике и для матриц. Так как здесь имеют место две совокупности векторов — векторы-строки и векторы-столбцы, то у матрицы, вообще говоря, имеется два ранга: строчный и столбцовый. Ответ на вопрос об их равноправии дает следующая теорема.

Теорема 14.1. *Строчный и столбцовый ранги любой матрицы равны.*

Доказательство этой теоремы мы опускаем.

Стало быть, ранг любой матрицы размера $m \times n$ можно искать как ранг одной из двух систем векторов: либо m векторов-строк, либо n векторов-столбцов. Как следует из параграфа 13.2, для прямоугольной матрицы максимальный ранг $r = \min(m, n)$. Для квадратной матрицы размера $n \times n$ ее максимальный ранг не может превышать n, т. е. $r \leqslant n$.

Понятие обратной матрицы

Понятие обратной матрицы распространяется только на квадратные матрицы, поэтому здесь и далее мы будем иметь дело с матрицами порядка n.

Определение 1. Матрица порядка n называется *вырожденной*, если ее ранг $r < n$.

Определение 2. Матрица $\mathbf{A}^{-1}$ называется *обратной* по отношению к матрице $\mathbf{A}$, если их произведение равно единичной матрице:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{E}. \tag{14.14}$$

Способы нахождения обратной матрицы будут приведены в следующих главах. Несколько забегая вперед, отметим, что для вырожденной матрицы не существует обратной матрицы. Иными словами, *если для некоторой матрицы порядка n ее ранг $r < n$, то для нее не существует обратной матрицы.*

Глава 15
ОПРЕДЕЛИТЕЛИ

15.1. Операции над определителями и основные свойства

Понятие определителя

Любой квадратной матрице $\mathbf{A}$ порядка n ставится в соответствие по определенному закону некоторое число, называемое *определителем* или *детерминантом* n-го порядка этой матрицы. Начнем с определителей второго и третьего порядков.

Пусть дана матрица

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

тогда ее определитель второго порядка вводится по формуле

$$\Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{15.1}$$

Правило вычисления определителя второго порядка очевидно: из произведения элементов на главной диагонали вычитается произведение элементов на второй диагонали матрицы $\mathbf{A}$. Нетрудно видеть, что формула (15.1) представляет собой алгебраическую сумму двух попарных произведений элементов матрицы $\mathbf{A}$, стоящих в разных строках и разных столбцах.

В дальнейшем мы не будем приводить матрицу, для которой вычисляется определитель, так как в записи определителя содержатся все элементы соответствующей матрицы.

Определитель третьего порядка вычисляется по формуле

$$\Delta_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \tag{15.2}$$

$$= a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{12}a_{23}a_{31} -$$
$$- a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}.$$

Правило вычисления определителя третьего порядка таково. Это алгебраическая сумма шести тройных произведений элементов, стоящих в разных строках и разных столбцах; со знаком "плюс" берутся произведения, сомножители которых находятся на главной диагонали и в вершинах треугольников с основаниями, параллельными главной диагонали; со знаком "минус" — произведения, сомножители которых стоят на неглавной диагонали и в вершинах треугольников с основаниями, параллельными этой диагонали (рис. 15.1). Заметим, что каждое слагаемое в формуле (15.2) содержит по одному элементу из каждой строки и каждого столбца соответствующей матрицы.

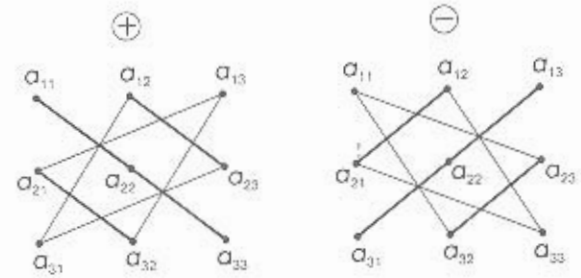

Рис. 15.1

В дальнейшем будем оперировать элементами, строками и столбцами определителей, не оговаривая специально, что это элементы, строки и столбцы соответствующих матриц. Порядок определителя полагается равным порядку соответствующей матрицы. Таким образом, определители выступают как самостоятельный объект изучения.

Итак, рассмотрим определитель n-го порядка

$$\Delta_n = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \qquad (15.3)$$

Теперь, с учетом подмеченных выше закономерностей, перейдем к определению для общего случая.

Определение 1. *Определителем* матрицы $\mathbf{A}$ n-го порядка называется алгебраическая сумма $n!$ произведений n-го порядка элементов этой матрицы, причем в каждое произведение входит по одному элементу из каждой строки и каждого столбца данной матрицы.

Основные свойства определителей

Из данного выше общего определения следуют основные свойства определителей.

1°. *Если некоторая строка (столбец) определителя состоит из нулей, то определитель равен нулю.*

Действительно, согласно общему определению, в каждое из $n!$ слагаемых обязательно войдет сомножителем элемент нулевой строки (нулевого столбца).

2°. *При перестановке двух строк (столбцов) определитель меняет знак.*

Это свойство легко проверяется на определителях второго и третьего порядков.

3°. *Определитель, содержащий две одинаковые строки (два одинаковых столбца), равен нулю.*

Действительно, поменяв местами эти строки, получаем $\Delta_n = -\Delta_n$, откуда и следует, что $\Delta = 0$.

4°. *Общий множитель любой строки (столбца) можно вынести за знак определителя.*

5°. *Если каждый элемент некоторой строки (столбца) определителя Δ_n представлен в виде суммы двух слагаемых, то этот определитель равен сумме двух определителей, в каждом из которых:*

а) все строки (столбцы), за исключением указанной строки (столбца), совпадают с аналогичными строками (столбцами) определителя Δ_n;

б) на месте указанной строки (столбца) первый определитель содержит первые слагаемые, а второй определитель — вторые слагаемые данной строки (столбца) определителя Δ_n.

Поясним это свойство на примере определителя третьего порядка:

$$\Delta_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a'_{21} + a''_{21} & a'_{22} + a''_{22} & a'_{23} + a''_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \Delta'_3 + \Delta''_3 =$$

$$= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a'_{21} & a'_{22} & a'_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a''_{21} & a''_{22} & a''_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

6°. *Определитель не изменится, если к элементам любой строки (столбца) прибавить соответствующие элементы другой строки (столбца), умноженные на любое число.*

Это свойство является следствием свойств 3°–5°.

7°. *При транспонировании матрицы определитель не меняется.*

Из перечисленных свойств следует, что определитель равен нулю, если по крайней мере одна из его строк (столбцов) является линейной комбинацией каких-либо других его строк (столбцов). Отсюда вытекает необходимое и достаточное условие равенства нулю определителя. *Определитель равен нулю тогда и только тогда, когда его строки (столбцы) линейно зависимы.*

Миноры и алгебраические дополнения

Рассмотрим определитель n-го порядка (15.3). Выделим в нем какой-либо элемент a_{ij} и вычеркнем i-ю строку и j-й столбец, на пересечении которых расположен этот элемент. Полученный определитель $(n-1)$-го порядка называется *минором* M_{ij} *элемента* a_{ij} определителя Δ_n.

Например, возьмем определитель четвертого порядка

$$\Delta_4 = \begin{vmatrix} 0 & -1 & 2 & 2 \\ 1 & 2 & 1 & 5 \\ 2 & 3 & 7 & 1 \\ 3 & 0 & 9 & 3 \end{vmatrix}.$$

Минор M_{32} элемента a_{32} получается вычеркиванием из данного определителя 3-й строки и 2-го столбца. Полученный определитель третьего порядка равен

$$\mathrm{M}_{32} = \begin{vmatrix} 0 & 2 & 4 \\ 1 & 1 & 5 \\ 3 & 9 & 3 \end{vmatrix} = 36 + 30 - 12 - 6 = 48.$$

Определение 2. *Алгебраическим дополнением* элемента a_{ij} определителя (15.3) называется число

$$\mathrm{A}_{ij} = (-1)^{i+j}\mathrm{M}_{ij}. \qquad (15.4)$$

Так, для приведенного выше примера алгебраическое дополнение

$$\mathrm{A}_{32} = (-1)^5 48 = -48.$$

Миноры и алгебраические дополнения играют важную роль в алгебре и ее приложениях. Одним из таких применений является основополагающая теорема о способе вычисления определителей.

Теорема 15.1. *Определитель равен сумме произведений элементов любой строки на их алгебраические дополнения:*

$$\Delta_n = a_{i1}\mathrm{A}_{i1} + a_{i2}\mathrm{A}_{i2} + \cdots + a_{in}\mathrm{A}_{in}, \quad 1 \leqslant i \leqslant n. \qquad (15.5)$$

Формула (15.5) называется *разложением* определителя по i-й строке. Доказательство этой теоремы мы опускаем. Аналогичное утверждение имеет место и для разложения определителя по любому столбцу.

Формула (15.5) сводит вычисление определителя n-го порядка к вычислению n определителей $(n-1)$-го порядка. Зная формулу (15.2) вычисления определителя третьего порядка, мы, например, можем найти определитель четвертого порядка путем разложения его на сумму алгебраических дополнений по формуле (15.5).

Пример 1. Вычислить определитель четвертого порядка

$$\Delta_4 = \begin{vmatrix} 1 & 2 & 4 & 7 \\ 0 & 3 & 0 & 2 \\ 2 & 4 & 3 & 2 \\ 6 & 3 & 1 & 1 \end{vmatrix}.$$

В принципе, разложить определитель можно по любой строке (столбцу), согласно формуле (15.5). Однако объем вычислений можно существенно уменьшить, если выбрать такую строку (столбец), в которой больше элементов, равных нулю. Наиболее подходящей в нашем случае является вторая строка. Разложение по ней определителя имеет вид

$$\Delta_4 = 0\mathrm{A}_{21} + 3\mathrm{A}_{22} + 0\mathrm{A}_{23} + 2\mathrm{A}_{24} = 3(-1)^4 \begin{vmatrix} 1 & 4 & 7 \\ 2 & 3 & 2 \\ 6 & 1 & 1 \end{vmatrix} +$$

$$+2(-1)^6 \begin{vmatrix} 1 & 2 & 4 \\ 2 & 4 & 3 \\ 6 & 3 & 1 \end{vmatrix} = 3(3 + 14 + 48 - 126 - 2 - 8) +$$

$$+2(4 + 24 + 36 - 48 - 9 - 4) = -207.$$

15.2. Применение определителей

Ранг матрицы и системы векторов

Пусть дана матрица, содержащая m строк и n столбцов

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \qquad (15.6)$$

Выделим в ней произвольным образом k строк и k столбцов. Элементы, которые находятся на пересечении выделенных строк и столбцов, образуют квадратную матрицу k-го порядка; определитель этой матрицы называется *минором k-го порядка матрицы* $\mathbf{A}$. Очевидно, что в общем случае таких миноров k-го порядка может быть несколько. При этом максимальный порядок миноров равен минимальному из чисел m и n, т. е.

$$\max k = \min(m, n). \tag{15.7}$$

Из всех возможных миноров матрицы $\mathbf{A}$ выделим те из них, которые отличны от нуля. В свою очередь, среди этих миноров можно найти по крайней мере один минор наибольшего порядка.

Определение 1. Наибольший порядок миноров, отличных от нуля, называется *рангом* матрицы (15.6).

Определение 2. Отличный от нуля минор матрицы, порядок которого равен рангу матрицы, называется *базисным минором* этой матрицы. Столбцы и строки матрицы, участвующие в образовании базисного минора, также называются *базисными*.

В курсе алгебры важную роль играет теорема о базисном миноре, которую мы приведем без доказательства.

Теорема 15.2. *Всякая строка (столбец) матрицы является линейной комбинацией ее базисных строк (столбцов).*

Заметим, что в общем случае у матрицы может быть несколько базисных миноров.

В параграфе 14.2 было дано определение ранга матрицы как наибольшего числа линейно независимых ее векторов-строк (столбцов). В курсе алгебры доказывается, что эти два определения эквивалентны. Приведенное в данном параграфе определение дает возможность вычислять ранг матрицы, а значит, и ранг системы векторов.

Пример 1. Найти ранг матрицы размера 4×6

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 & 5 & 4 & 2 \\ 2 & 2 & 2 & 2 & 2 & 2 \\ 5 & 5 & 5 & 5 & 5 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Нетрудно видеть, что максимальный порядок миноров этой матрицы, отличных от нуля, равен двум, поскольку миноры третьего порядка должны обязательно содержать элементы по крайней мере двух строк из строк матрицы $\mathbf{A}$ с номерами 2 по 4, но такие определители равны нулю либо по признаку пропорциональности двух строк, либо по признаку наличия в них нулевой строки. У этой матрицы существует три базисные строки (с первой по третью), и пять ее столбцов являются базисными (либо с первого по пятый, либо со второго по шестой); из них и формируются все базисные миноры второго порядка.

Рассмотрим квадратную матрицу порядка n, т.е. когда в матрице (15.6) $m = n$. Как было отмечено в параграфе 14.2, матрица порядка n является вырожденной и не имеет обратной матрицы, если ее ранг $r < n$. Максимальный порядок минора квадратной матрицы равен n; в этом случае минор равен определителю этой матрицы. Стало быть, *квадратная матрица является вырожденной, если ее определитель равен нулю.*

Вычисление обратной матрицы

Выведем формулу для вычисления обратной матрицы к заданной невырожденной матрице.

Пусть дана невырожденная матрица n-го порядка

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \tag{15.8}$$

так что ее определитель Δ не равен нулю. Рассмотрим матрицу, составленную из алгебраических дополнений к элементам матрицы $\mathbf{A}$:

$$\mathbf{A}^* = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{n1} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{A}_{1n} & \mathbf{A}_{2n} & \cdots & \mathbf{A}_{nn} \end{pmatrix}. \tag{15.9}$$

Следует обратить особое внимание, что элемент матрицы $\mathbf{A}^*$, стоящий на пересечении i-й строки и j-го столбца,

28 Зак. 6853

является алгебраическим дополнением к элементу a_{ij} матрицы $\mathbf{A}$. Матрица (15.9) называется *присоединенной*, или *взаимной, матрицей к матрице* (15.8).

Найдем произведения матриц $\mathbf{AA}^*$ и $\mathbf{A}^*\mathbf{A}$. Произведение матриц $\mathbf{AA}^*$ представляет собой, согласно определению (см. параграф 14.1), матрицу $\mathbf{C}$ с элементами

$$c_{ij} = a_{i1}\mathrm{A}_{j1} + a_{i2}\mathrm{A}_{j2} + \cdots + a_{in}\mathrm{A}_{jn}; \quad i, j = 1, 2, \cdots, n. \tag{15.10}$$

Согласно теореме 15.1, сумма произведений элементов i-й строки матрицы $\mathbf{A}$ на их алгебраические дополнения равна определителю этой матрицы, т. е. при $i = j$, на главной диагонали матрицы $\mathbf{C}$, мы получаем разложения одного и того же определителя Δ матрицы $\mathbf{A}$ по разным строкам, начиная с первой.

Что касается элементов c_{ij}, стоящих вне главной диагонали матрицы $\mathbf{C}$, то при $i \neq j$ мы имеем сумму произведений всех элементов некоторой строки определителя Δ на алгебраические дополнения соответствующих элементов другой строки этого же определителя. По теореме 15.1 такое разложение (15.10) равно определителю, который получен из определителя Δ заменой j-й строки на i-ю строку:

$$\Delta^{(j)} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{matrix} \\ \\ \leftarrow i\text{-я строка} \\ \\ \leftarrow j\text{-я строка}, \\ \\ \\ \end{matrix}$$

$$i \neq j; \quad i, j = 1, 2, \cdots, n.$$

Но по свойству 3^a (см. параграф 15.1) определитель, имеющий две одинаковые строки, равен нулю. Стало быть, матрица $\mathbf{C}$ является диагональной и имеет вид

$$\mathbf{C} = \mathbf{AA}^* = \begin{pmatrix} \Delta & 0 & 0 & \cdots & 0 \\ 0 & \Delta & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \Delta \end{pmatrix} =$$

$$= \Delta \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = \Delta \cdot \mathbf{E},$$

где $\mathbf{E}$ — единичная матрица.

Точно так же можно показать, что и произведение матриц $\mathbf{A}^*\mathbf{A}$ равно $\Delta \cdot \mathbf{E}$. Следовательно, мы получили, что

$$\mathbf{AA}^* = \mathbf{A}^*\mathbf{A} = \Delta \cdot \mathbf{E}.$$

Отсюда вытекает, что

$$\mathbf{A}\left(\frac{1}{\Delta}\mathbf{A}^*\right) = \left(\frac{1}{\Delta}\mathbf{A}^*\right)\mathbf{A} = \mathbf{E}.$$

Теперь, в силу определения обратной матрицы (см. параграф 14.2), окончательно имеем выражение для обратной матрицы:

$$\mathbf{A}^{-1} = \frac{1}{\Delta}\mathbf{A}^*, \tag{15.11}$$

где $\mathbf{A}^*$ определяется по формуле (15.9). Согласно определению умножения матрицы на число (см. параграф 14.1), получаем формулу для вычисления обратной матрицы к матрице (15.8):

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathrm{A}_{11}/\Delta & \mathrm{A}_{21}/\Delta & \cdots & \mathrm{A}_{n1}/\Delta \\ \mathrm{A}_{12}/\Delta & \mathrm{A}_{22}/\Delta & \cdots & \mathrm{A}_{n2}/\Delta \\ \cdots & \cdots & \cdots & \cdots \\ \mathrm{A}_{1n}/\Delta & \mathrm{A}_{2n}/\Delta & \cdots & \mathrm{A}_{nn}/\Delta \end{pmatrix}. \tag{15.12}$$

Из формул (15.11) и (15.12) видно, что для существования обратной матрицы необходимо, чтобы исходная матрица (15.8) имела ненулевой определитель.

Глава 16

СИСТЕМЫ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

Это один из основных разделов в алгебре. Нет такой отрасли науки и приложений, где в том или ином виде не использовались бы системы линейных алгебраических уравнений. При решении экономических задач системы линейных уравнений наиболее употребимы как в аппарате исследования, так и при рассмотрении частных проблем.

16.1. Основные понятия

Общий вид и свойства системы уравнений

Система m линейных уравнений с n неизвестными (переменными) $x_1, x_2, \cdots, x_n$ имеет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m. \end{cases} \tag{16.1}$$

Здесь a_{ij} и b_j — произвольные числа ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$), которые называются соответственно *коэффициентами при неизвестных* и *свободными членами* уравнений (16.1). Первый индекс у коэффициентов при неизвестных означает номер уравнения, второй индекс соответствует номеру неизвестного x_i.

Решением системы уравнений (16.1) называется набор n чисел $x_1 = \alpha_1, x_2 = \alpha_2, \ldots, x_n = \alpha_n$, при подстановке которых в эту систему каждое уравнение данной системы превращается в тождество.

Система уравнений (16.1) называется *совместной*, если она имеет хотя бы одно решение; если система не имеет решений, она называется *несовместной*. Совместная система уравнений имеет либо одно решение, и в таком случае она называется *определенной*, либо, если у нее больше одного решения, она называется *неопределенной*.

Системы уравнений вида (16.1) называются *эквивалентными*, если они имеют одно и то же множество решений. Элементарные преобразования исходной системы приводят к эквивалентной системе. К *элементарным преобразованиям* относятся:

1) вычеркивание уравнения $0x_1 + 0x_2 + \cdots + 0x_n = 0$ — нулевой строки;
2) перестановка уравнений или слагаемых $a_{ij}x_j$ в уравнениях;
3) прибавление к обеим частям одного уравнения соответственно обеих частей другого уравнения этой системы, умноженного на любое действительное число;
4) удаление уравнений, являющихся линейными комбинациями других уравнений системы.

Последнее свойство вытекает из третьего свойства: если какое-либо уравнение представляет собой линейную комбинацию других уравнений, то из него можно сформировать нулевую строку.

Матричная форма системы уравнений

Сведем коэффициенты при неизвестных в системе уравнений (16.1) в матрицу

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \tag{16.2}$$

Эта матрица состоит из m строк и n столбцов и называется *матрицей системы*. Введем в рассмотрение две матрицы-столбца: матрицу неизвестных $\mathbf{X}$ и матрицу свободных членов $\mathbf{B}$:

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}. \tag{16.3}$$

$\mathbf{X}$ и $\mathbf{B}$ представляют собой векторы-столбцы, однако в целях единого подхода в рамках матричной алгебры удобнее трактовать их именно как матрицы, состоящие соответственно из n и m строк и одного столбца.

Тогда систему линейных уравнений (16.1) можно записать в матричной форме, поскольку размер матрицы $\mathbf{A}$ равен $m \times n$, а размер $\mathbf{X}$ — $n \times 1$, значит, произведение этих матриц имеет смысл:

$$\mathbf{AX} = \mathbf{B}. \tag{16.4}$$

Произведение матриц $\mathbf{AX}$ является, как и $\mathbf{B}$, матрицей-столбцом размера $m \times 1$, состоящей из левых частей уравнений системы (16.1). Все уравнения этой системы вытекают из уравнения (16.4) в силу определения 2 равенства двух матриц (см. параграф 14.1).

Введем в рассмотрение еще одну матрицу; дополним матрицу системы $\mathbf{A}$ столбцом свободных членов и получим новую матрицу размера $m \times (n+1)$:

$$\mathbf{A_B} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}. \tag{16.5}$$

Матрица $\mathbf{A_B}$ называется *расширенной матрицей системы*. Эта матрица играет важную роль в вопросе о разрешимости системы уравнений.

Теорема 16.1 (теорема Кронекера-Капелли; критерий совместности системы). *Система линейных уравнений совместна тогда и только тогда, когда ранг матрицы системы равен рангу расширенной матрицы системы.*

Доказательство этой теоремы мы не приводим.

16.2. Методы решения систем линейных уравнений

Метод обратной матрицы и теорема Крамера

Рассмотрим частный случай системы (16.1), когда число уравнений равно числу неизвестных, т.е. $m = n$. Система уравнений имеет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n. \end{cases} \tag{16.6}$$

Составим квадратную матрицу $\mathbf{A}$ порядка n этой системы:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}. \tag{16.7}$$

В матричной форме система уравнений (16.6) имеет вид (16.4):

$$\mathbf{AX} = \mathbf{B}, \tag{16.8}$$

где матрицы $\mathbf{X}$ и $\mathbf{B}$ имеют размер $n \times 1$. Пусть матрица системы $\mathbf{A}$ является невырожденной, т.е. существует обратная матрица $\mathbf{A}^{-1}$. Умножив обе части этого уравнения слева на $\mathbf{A}^{-1}$, получаем решение системы (16.6) в матричной форме:

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}. \tag{16.9}$$

Вычисление обратной матрицы по заданной матрице $\mathbf{A}$ производится по формулам (15.4) и (15.12). В случае когда порядок n матриц $\mathbf{A}$ и $\mathbf{A}^{-1}$ достаточно велик, вычисление обратной матрицы может быть очень сложным.

Другой метод решения системы уравнений (16.6) основан на теореме Крамера. Составим определитель матрицы системы $\mathbf{A}$:

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}, \tag{16.10}$$

который называется также *определителем системы*. Заменим в этом определителе j-й столбец на столбец свободных членов

B, т.е. получим этой заменой другой определитель, который обозначим Δ_j:

$$\Delta_j = \begin{vmatrix} a_{11} & a_{12} & \cdots & \overset{(j)}{b_1} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & b_2 & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ a_{n1} & a_{n2} & \cdots & b_n & \cdots & a_{nn} \end{vmatrix}. \qquad (16.11)$$

Теорема 16.2 (ТЕОРЕМА КРАМЕРА). *Пусть Δ — определитель матрицы системы $\mathbf{A}$, а Δ_j — определитель, полученный из определителя Δ заменой j-го столбца столбцом свободных членов $\mathbf{B}$. Тогда, если $\Delta \neq 0$, система линейных уравнений (16.6) имеет единственное решение, определяемое по формулам*

$$x_j = \Delta_j/\Delta, \quad j = 1, 2, \ldots, n. \qquad (16.12)$$

Доказательство. Запишем в развернутом виде форму матричного решения системы (16.9). Из вида обратной матрицы (15.12), а также из определения умножения матрицы на число следует, что столбец неизвестных $\mathbf{X}$ выражается по формуле

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{21} & \cdots & \mathbf{A}_{n1} \\ \mathbf{A}_{12} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{A}_{1n} & \mathbf{A}_{2n} & \cdots & \mathbf{A}_{nn} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Выполнив умножение матриц в правой части этого равенства, мы имеем равенство двух матриц-столбцов, из которого, приравнивая соответствующие элементы, получаем систему равенств — выражений для неизвестных x_j:

$$x_j = \frac{1}{\Delta}(b_1\mathbf{A}_{1j} + b_2\mathbf{A}_{2j} + \cdots + b_n\mathbf{A}_{nj}), \qquad j = 1, 2, \ldots, n.$$

Но по теореме 15.1 (см. параграф 15.1) сумма в скобках правой части этого равенства представляет собой разложение определителя по j-му столбцу, в котором стоят элементы столбца свободных членов, а остальные столбцы этого определителя такие же, как и в определителе системы Δ. Иными словами, эта сумма и есть определитель Δ_j. Теорема доказана.

Формулы вычисления неизвестных (16.12) — решения системы (16.6) — носят название *формул Крамера.*

Пример 1. Найти решение системы уравнений

$$\begin{cases} x + 2y + z = 1, \\ 2x + 3y + 2z = 2, \\ x - y + 3z = 0. \end{cases}$$

Составим и вычислим определители системы Δ и Δ_j $(j = x, y, z)$:

$$\Delta = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & -1 & 3 \end{vmatrix} = -2, \quad \Delta_x = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 0 & -1 & 3 \end{vmatrix} = -3,$$

$$\Delta_y = \begin{vmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 0 & 3 \end{vmatrix} = 0, \quad \Delta_z = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & -1 & 0 \end{vmatrix} = 1.$$

Определитель системы отличен от нуля, стало быть, она имеет единственное решение, которое вычисляется по формулам (16.12):

$$x = \Delta_x/\Delta = 3/2, \quad y = \Delta_y/\Delta = 0, \quad z = \Delta_z/\Delta = -1/2.$$

Решение системы общего вида

Пусть задана система линейных уравнений общего вида (16.1), где $m \leqslant n$, т. е. число неизвестных не меньше числа уравнений. Представим общий порядок решения этой системы.

Необходимо определить совместность системы, т.е. определить сначала ранги матрицы системы $\mathbf{A}$ и расширенной матрицы $\mathbf{A_B}$. По теореме Кронекера–Капелли, если ранги этих матриц не совпадают, то система несовместна, и тогда нет смысла ее решать. Если же ранги матриц $\mathbf{A}$ и $\mathbf{A_B}$ равны, то система (16.1) совместна.

Определение. *Рангом совместной системы линейных алгебраических уравнений* называется ранг ее матрицы.

Пусть система (16.1) совместна и ранг ее равен k. Выделим в матрице системы (16.2) некоторый базисный минор; предположим, что именно первые r строк матриц $\mathbf{A}$ и $\mathbf{A_B}$ являются

базисными. Тогда по теореме о базисном миноре остальные строки матрицы являются линейными комбинациями остальных строк. В свою очередь, это означает, что в системе (16.1) первые r уравнений, соответствующие базисным строкам матрицы $\mathbf{A}$, являются базисными, а остальные — их линейными комбинациями. Тогда эти $(m-r)$ уравнений можно удалить из системы, причем в результате указанных элементарных преобразований мы получаем эквивалентную систему:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{r1}x_1 + a_{r2}x_2 + \cdots + a_{rn}x_n = b_r. \end{cases} \tag{16.13}$$

Система (16.13) характерна тем, что ее ранг равен числу уравнений в ней, причем $r \leqslant n$, т. е. ранг не превосходит число неизвестных. Поэтому возможны два случая: либо $r = n$, либо $r < n$.

В первом случае система (16.13) имеет квадратную невырожденную матрицу порядка r и, согласно теореме Крамера (теорема 16.2), существует единственное решение этой системы. Иными словами, если ранг системы равен числу неизвестных, то система имеет единственное решение, т. е. она является определенной.

Рассмотрим теперь случай, когда $r < n$. Перенесем в правые части уравнений (16.13) все слагаемые, содержащие неизвестные $x_{r+1}, x_{r+2}, \ldots, x_n$. Тогда система принимает вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1r}x_r = b_1 - a_{1,r+1}x_{r+1} - \cdots - a_{1n}x_n, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2r}x_r = b_2 - a_{2,r+1}x_{r+1} - \cdots - a_{2n}x_n, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{r1}x_1 + a_{r2}x_2 + \cdots + a_{rr}x_r = b_r - a_{r,r+1}x_{r+1} - \cdots - a_{rn}x_n. \end{cases} \tag{16.14}$$

Неизвестным $x_{r+1}, x_{r+2}, \ldots, x_n$ можно придавать любые значения, и потому они называются *свободными*. Неизвестные $x_1, x_2, \ldots, x_r$, соответствующие базисным столбцам, называются *базисными*. Из системы (16.14) легко найти выражения базисных неизвестных через свободные, согласно теореме

Крамера, рассматривая правые части этих уравнений как элементы столбца свободных членов, содержащие $x_{r+1}, x_{r+2}, \ldots, x_n$. Можно показать, что базисные неизвестные $x_1, x_2, \ldots, x_r$ линейно выражаются через свободные неизвестные. Поскольку свободные неизвестные могут принимать любые значения, то в случае, когда ранг совместной системы меньше числа неизвестных, эта система является неопределенной: она имеет бесчисленное множество решений.

Метод Гаусса

Следует заметить, что вычисления как по методу обратной матрицы, так и по методу Крамера являются очень трудоемкими. Оба они требуют порядка $n^2n!$ арифметических действий для нахождения решения системы линейных уравнений. При $n = 5$ это составит около 3000 действий, при $n = 10$ — около $3,6 \cdot 10^8$ действий. При решении серьезных задач приходится иметь дело с системами уравнений порядка $n = 100$ и более. При таких масштабах даже суперкомпьютерам потребуется долгое время для вычисления решения. Кроме того, погрешности компьютерного округления чисел приводят к большим ошибкам в численных расчетах решения систем уравнений большого порядка. Между тем существуют более экономичные методы решения систем линейных уравнений, основанные на предварительном преобразовании расширенной матрицы системы к специальному виду, в частности метод Гаусса. Рассмотрим его практическую реализацию.

Пусть для определенности в системе уравнений общего вида (16.1) $a_{11} \neq 0$ (если $a_{11} = 0$, то можно переставить на первое место ненулевое слагаемое или начать с другого уравнения). Умножим первое уравнение системы (16.1) на число a_{21}/a_{11} и вычтем его из второго уравнения этой системы. Затем умножим обе части первого уравнения на число a_{31}/a_{11} и вычтем его из третьего уравнения и так далее, т. е. процесс заключается в последовательном вычитании первого уравнения, умножаемого на числа $a_{i1}/a11$, из i-го уравнения $(i = 2, 3, \ldots, m)$. Таким образом, в результате элементарных преобразований мы получим эквивалентную систему, в которой, начиная со второго уравнения, отсутствуют слагаемые, содержащие неизвестную x_1:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1, \\ a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \cdots + a_{2n}^{(1)}x_n = b_2^{(1)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m2}^{(1)}x_2 + a_{m3}^{(1)}x_3 + \cdots + a_{mn}^{(1)}x_n = b_m^{(1)} \end{cases} \tag{16.15}$$

где верхний индекс в скобках означает новые коэффициенты, полученные после первого шага. Для уменьшения громоздкости записи удобнее оперировать расширенной матрицей системы, отделяя в ней вертикальной чертой столбец свободных членов. Итак, после первого шага, содержащего $(m-1)$ элементарных преобразований системы, мы переходим от расширенной матрицы (16.5) исходной системы к расширенной матрице

$$\mathbf{A}_{\mathrm{B}}^{(1)} = \left(\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2n}^{(1)} & b_2^{(1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & a_{m2}^{(1)} & a_{m3}^{(1)} & \cdots & a_{mn}^{(1)} & b_m^{(1)} \end{array}\right). \tag{16.16}$$

Второй шаг заключается в том, что теперь второе уравнение системы (16.13), или вторая строка матрицы (16.14), используется для аналогичных элементарных преобразований строк с третьей по m-ю: эта строка последовательно умножается на число $a_{i2}^{(1)}/a_{22}^{(1)}$ и вычитается из i-й строки $(i = 3, 4, \ldots, m)$. В результате этих $(m-2)$ элементарных преобразований получаем новую расширенную матрицу, соответствующую новой эквивалентной системе уравнений. Эта матрица имеет вид

$$\mathbf{A}_{\mathrm{B}}^{(2)} = \left(\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2n}^{(1)} & b_2^{(1)} \\ 0 & 0 & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} & b_3^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & a_{m3}^{(2)} & \cdots & a_{mn}^{(2)} & b_m^{(2)} \end{array}\right). \tag{16.17}$$

где верхний индекс означает новые коэффициенты. В случае если элемент $a_{22}^{(1)} = 0$, второе уравнение можно поменять местами с другим уравнением, у которого элемент $a_{i2} \neq 0$.

Прололжим этот процесс аналогичным образом (т.е. на третьем шаге преобразуются строки с 4-й по m-ю, на четвертом шаге — строки с 5-й по m-ю и т.д.) до тех пор, пока не дойдем до последней m-й строки. После $(r-1)$-го шага процесса последовательного исключения неизвестных мы получим следующую расширенную матрицу:

$$\mathbf{A}_{\mathrm{B}}^{(r-1)} = \left(\begin{array}{ccccccc|c} a_{11} & a_{12} & \cdots & a_{1r} & a_{1,r+1} & \cdots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & \cdots & a_{2r}^{(1)} & a_{2,r+1}^{(1)} & \cdots & a_{2n}^{(1)} & b_2^{(1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{rr}^{(r-1)} & a_{r,r+1}^{(r-1)} & \cdots & a_{rn}^{(r-1)} & b_r^{(r-1)} \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & b_{r+1}^{(r-1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & b_m^{(r-1)} \end{array}\right). \tag{16.18}$$

Последние $(m-r)$ строк этой матрицы соответствуют уравнениям эквивалентной системы уравнений

$$\begin{gathered} 0x_1 + 0x_2 + \cdots + 0x_n = b_i^{(r-1)}; \\ i = r+1, r+2, \ldots, m. \end{gathered} \tag{16.19}$$

Эти уравнения могут появиться, если соответствующие уравнения исходной системы (16.1) представляют собой линейные комбинации других уравнений этой системы. Здесь мы не исследовали заранее систему (16.1) на совместность; поэтому если эта система несовместна, то хотя бы одно из чисел $b_{r+1}^{(r-1)}, b_{r+2}^{(r-1)}, \ldots, b_m^{(r-1)}$ не равно нулю. Таким образом, метод Гаусса позволяет на определенном шаге установить возможную несовместность исходной системы линейных уравнений или выявить и удалить уравнения, являющиеся линейными комбинациями других уравнений системы (16.1), если она совместна.

Пусть система (16.1) совместна, тогда все правые части уравнений (16.19) равны нулю, и после удаления нулевых уравнений в эквивалентной системе и нулевых строк в расширенной матрице получаем матрицу специфического ступенчатого вида, ранг которой равен r. Все элементы этой матрицы, стоящие слева или ниже элементов a_{ii} равны нулю:

$$\mathrm{A}_{\mathrm{B}}^{(r-1)}=\left(\begin{array}{ccccccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1r} & \cdots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2r}^{(1)} & \cdots & a_{2n}^{(1)} & b_2^{(1)} \\ 0 & 0 & a_{33}^{(2)} & \cdots & a_{3r}^{(2)} & \cdots & a_{3n}^{(2)} & b_3^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a_{rr}^{(r-1)} & \cdots & a_{rn}^{(r-1)} & b_r^{(r-1)} \end{array}\right). \tag{16.20}$$

Эта расширенная матрица соответствует системе уравнений ранга r, которая имеет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1, \\ a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \cdots + a_{2n}^{(1)}x_n = b_2^{(1)}, \\ a_{33}^{(2)}x_3 + \cdots + a_{3n}^{(2)}x_n = b_3^{(2)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{rr}^{(r-1)}x_r + \cdots + a_{rn}^{(r-1)}x_n = b_r^{(r-1)}. \end{cases} \tag{16.21}$$

Полученная система уравнений (16.21) позволяет найти решение, осуществляя процесс снизу вверх, т.е. от последнего уравнения к первому. Переход от системы (16.1) к эквивалентной ей системе (16.21) называется *прямым ходом*, а нахождение неизвестных из системы (16.21) — *обратным ходом* метода Гаусса. Далее последовательность действий аналогична последовательности для решения системы линейных уравнений общего вида.

Если $r = n$, то система (16.21) имеет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1r}x_r = b_1, \\ a_{22}^{(1)}x_2 + \cdots + a_{2r}^{(1)}x_r = b_2^{(1)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{rr}^{(r-1)}x_r = b_r^{(r-1)}. \end{cases} \tag{16.22}$$

Поднимаясь снизу вверх, последовательно находим (обратный ход метода Гаусса):

- из последнего r-го уравнения неизвестное $x_r = \dfrac{b_r^{(r-1)}}{a_{rr}^{(r-1)}}$;
- из $(r-1)$-го уравнения неизвестное x_{r-1} путем подстановки в это уравнение уже найденного неизвестного x_r;
- из i-го уравнения неизвестное x_i при подстановке в него найденных величин $x_r, x_{r-1}, \ldots, x_{i+1}$;
- и так далее до первого уравнения, из которого при подстановке в него уже найденных величин $x_r, x_{r-1}, \ldots, x_2$ находим x_1.

Если ранг системы уравнений (16.21) $r < n$, объявляем неизвестные $x_{r+1}, x_{r+2}, \ldots, x_n$ свободными и формируем правые части уравнений (16.21), оставляя в левых частях слагаемые, содержащие базисные переменные $x_1, x_2, \ldots, x_r$:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1r}x_r = b_1 - a_{1,r+1}x_{r+1} - \cdots - a_{1n}x_n, \\ a_{22}^{(1)}x_2 + \cdots + a_{2r}^{(1)}x_r = b_2^{(1)} - a_{2,r+1}^{(1)}x_{r+1} - \cdots - a_{2n}^{(1)}x_n, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{rr}^{(r-1)}x_r = b_r^{(r-1)} - a_{r,r+1}^{(r-1)}x_{r+1} - \cdots - a_{rn}^{(r-1)}x_n. \end{cases} \tag{16.23}$$

Решение этой системы находится обратным ходом метода; теперь базисные неизвестные зависят от свободных неизвестных, которые могут принимать любые значения, а потому система (16.1) имеет бесчисленное множество решений.

Рассмотрим примеры решения систем линейных уравнений методом Гаусса.

Пример 2. Выпишем расширенную матрицу системы уравнений, приведенную в примере 1; справа в скобках укажем числа, на которые умножается соответствующая строка матрицы, с тем чтобы сложить ее с другими строками. Горизонтальными стрелками показаны переходы к расширенным матрицам эквивалентных систем. Первую строку расширенной матрицы исходной системы умножаем последовательно на (-2) и (-1) и прибавляем ее соответственно ко 2-й и 3-й строкам этой матрицы. После первого шага, состоящего в "обнулении" первого

столбца, согласно формуле (16.15), получаем (номера шагов показаны над стрелками перехода):

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 2 & 3 & 2 & 2 \\ 1 & -1 & 3 & 0 \end{array}\right) \begin{matrix}\\ (-2),(-1)\end{matrix} \xrightarrow{①} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & -3 & 2 & -1 \end{array}\right).$$

Второй шаг прямого хода метода Гаусса состоит в операциях с преобразованной расширенной матрицей: прибавляем вторую строку, умноженную на (-3), к 3-й строке

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & -3 & 2 & -1 \end{array}\right) (-3) \xrightarrow{②} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 2 & -1 \end{array}\right).$$

Последний вид расширенной матрицы является конечным этапом прямого хода метода [см. формулу (16.22)], после чего приступаем к обратному ходу, т.е. находим неизвестные, начиная с последнего. Полученная расширенная матрица соответствует системе уравнений

$$\begin{cases} x + 2y + z = 1, \\ -y = 0, \\ 2z = -1, \end{cases}$$

которая эквивалентна исходной системе. Отсюда последовательно находим: $z = -1/2$, $y = 0$, $x = 1 - 0 - (-1/2) = 3/2$.

Пример 3. Найти решение системы уравнений

$$\begin{cases} x_1 - 3x_2 + 2x_3 + 3x_4 = 1, \\ 2x_1 - x_2 + x_3 - 4x_4 = 0, \\ 2x_1 + x_2 - x_3 + 2x_4 = 2, \\ 3x_1 - 2x_2 + x_3 + 5x_4 = 3. \end{cases}$$

Составим расширенную матрицу этой системы, после чего выполним соответствующие шаги прямого хода метода Гаусса. Получаем:

$$\left(\begin{array}{cccc|c} 1 & -3 & 2 & 3 & 1 \\ 2 & -1 & 1 & -4 & 0 \\ 2 & 1 & -1 & 2 & 2 \\ 3 & -2 & 1 & 5 & 3 \end{array}\right) \begin{matrix}\\ (-2),(-2),(-3)\end{matrix} \xrightarrow{①}$$

$$\xrightarrow{①} \left(\begin{array}{cccc|c} 1 & -3 & 2 & 3 & 1 \\ 0 & 5 & -3 & -10 & -2 \\ 0 & 7 & -5 & -4 & 0 \\ 0 & 7 & -5 & -4 & 0 \end{array}\right) (7/5),(-7/5) \xrightarrow{②}$$

$$\xrightarrow{②} \left(\begin{array}{cccc|c} 1 & -3 & 2 & 3 & 1 \\ 0 & 5 & -3 & -10 & -2 \\ 0 & 0 & -\frac{4}{5} & 10 & \frac{14}{5} \\ 0 & 0 & -\frac{4}{5} & 10 & \frac{14}{5} \end{array}\right) (-1) \xrightarrow{③}$$

$$\xrightarrow{③} \left(\begin{array}{cccc|c} 1 & -3 & 2 & 3 & 1 \\ 0 & 5 & -3 & -10 & -2 \\ 0 & 0 & -\frac{4}{5} & 10 & \frac{14}{5} \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Последняя нулевая строка в расширенной матрице, полученной после третьего шага, появилась из-за того, что в исходной системе четвертое уравнение является суммой первого и третьего уравнений. Система совместная, и после удаления нулевой строки заключительный вид расширенной матрицы соответствует системе трех уравнений с четырьмя неизвестными (ранг системы меньше числа неизвестных). Полагая x_4 свободной переменной, получаем

$$\begin{cases} x_1 - 3x_2 + 2x_3 = 1 - 3x_4, \\ 5x_2 - 3x_3 = -2 + 10x_4, \\ -\frac{4}{5}x_3 = \frac{14}{5} - 10x_4. \end{cases}$$

Из этой системы получаем обратным ходом метода Гаусса

$$x_3 = -\frac{7}{2} + \frac{25}{2}x_4, \qquad x_2 = -\frac{5}{2} + \frac{19}{2}x_4, \qquad x_1 = \frac{1}{2} + \frac{1}{2}x_4.$$

Данная система уравнений имеет бесчисленное множество решений, поскольку x_4 может принимать любые значения.

Геометрическая интерпретация системы линейных уравнений

Как известно, уравнение с двумя переменными вида

$$Ax + By + C = 0$$

описывает на координатной плоскости Oxy прямую. Система двух уравнений такого вида означает, что ее решение, как точки на координатной плоскости, должны принадлежать одновременно двум прямым, соответствующим уравнениям этой системы. Отсюда возможны следующие варианты: а) прямые пересекаются, и тогда система имеет единственное решение; б) прямые параллельны, и система не имеет решения (несовместна); в) прямые совпадают, т.е. ранг системы равен единице, и система имеет бесчисленное множество решений.

Уравнение с тремя переменными вида

$$Ax + By + Cz - D = 0$$

описывает плоскость в трехмерном пространстве. Решение системы трех уравнений с тремя неизвестными — это точки пространства, которые должны принадлежать одновременно трем плоскостям, которые описываются уравнениями системы. В этом случае возможны следующие варианты:

а) три плоскости пересекаются в одной точке, и система имеет единственное решение;

б) три плоскости пересекаются по одной прямой — система имеет бесчисленное множество решений (все точки на этой прямой);

в) две плоскости совпадают, а третья пересекает их — бесчисленное множество решений (все точки прямой пересечения трех плоскостей), ранг системы равен двум;

г) все три плоскости совпадают — все точки общей плоскости являются решениями и ранг системы равен единице;

д) хотя бы одна из плоскостей параллельна какой-либо из двух других — система несовместна;

е) плоскости пересекаются попарно по параллельным прямым — система несовместна.

В последних двух случаях несовместность системы уравнений обусловлена тем, что нет таких точек трехмерного пространства, которые принадлежали бы одновременно всем трем плоскостям.

В случае системы уравнений с n неизвестными каждое уравнение вида

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - b_1 = 0, \qquad i = 1, 2, \cdots, m,$$

можно интерпретировать как гиперплоскость в координатном пространстве $^*A^n$. Решение системы (16.1) — это множество таких точек пространства A^n, которые принадлежат одновременно всем m гиперплоскостям, соответствующим уравнениям этой системы.

16.3. Однородные системы линейных уравнений

Определение. Система линейных уравнений называется *однородной*, если во всех ее уравнениях свободные члены равны нулю.

В общем случае однородная система (или система однородных уравнений) имеет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0. \end{cases} \tag{16.24}$$

Однородная система уравнений всегда совместна: действительно, набор значений неизвестных $x_i = 0$ $(i = 1, 2, \ldots, n)$ удовлетворяет всем уравнениям системы. Это решение однородной системы называется нулевым, или *тривиальным*.

Решение системы однородных уравнений

Вопрос о существовании ненулевого решения однородной системы линейных уравнений (16.24) разрешает следующая теорема.

Теорема 16.3. *Однородная система имеет ненулевое решение тогда и только тогда, когда ранг этой системы меньше числа ее неизвестных.*

Доказательство. а) Необходимость. Пусть система (16.24) имеет ненулевое решение, докажем, что тогда ее ранг $r < n$. Действительно, если $r = n$, то и $m = n$, и тогда по теореме Крамера (см. параграф 16.2) система с невырожденной квадратной матрицей имеет единственное решение, определяемое по формулам (16.12). Но так как каждый из определителей Δ_j содержит нулевой столбец свободных членов, это решение является нулевым. Следовательно, $r < n$.

б) Достаточность. Пусть $r < n$. Но тогда, поскольку ранг системы (16.24) меньше числа ее неизвестных, эта система имеет бесконечно много решений, а значит, и хотя бы одно ненулевое решение (см. параграф 16.2). Теорема доказана.

Из этой теоремы вытекают два важных следствия.

Следствие 1. *Если число уравнений однородной системы меньше числа ее неизвестных, то эта система имеет ненулевое решение.*

Следствие 2. *Если в однородной системе число уравнений равно числу неизвестных, то она имеет ненулевое решение тогда и только тогда, когда определитель матрицы системы равен нулю.*

Фундаментальная система решений

Решения однородной системы обладают следующими свойствами. Если вектор $\bar{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ является решением системы (16.24), то и для любого числа k вектор $k\bar{\alpha} = (k\alpha_1, k\alpha_2, \ldots, k\alpha_n)$ также будет решением этой системы. Если решением системы (16.24) является также и вектор $\bar{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_n)$, то и сумма $\bar{\alpha} + \bar{\gamma}$ будет решением этой системы. Отсюда следует, что *любая линейная комбинация решений однородной системы также является решением этой системы.*

Как мы знаем (см. параграф 13.2), всякая система n-мерных векторов, состоящая более чем из n векторов, является

линейно зависимой. Таким образом, из множества векторов-решений однородной системы (16.24) можно выбрать базис, т. е. любой вектор-решение данной системы будет линейной комбинацией векторов этого базиса. Любой такой базис называется *фундаментальной системой решений однородной системы линейных уравнений*. Справедлива следующая теорема, которую мы приводим без доказательства.

Теорема 16.4. *Если ранг r системы однородных уравнений* (16.24) *меньше числа неизвестных n, то всякая фундаментальная система решений системы* (16.24) *состоит из $n - r$ решений.*

Укажем теперь способ нахождения фундаментальной системы решений (ФСР). Пусть система однородных уравнений (16.24) имеет ранг $r < n$. Тогда, как следует из правила Крамера, базисные неизвестные этой системы $x_1, x_2, \ldots, x_r$ линейно выражаются через свободные переменные $x_{r+1}, x_{r+2}, \ldots, x_n$:

$$\begin{aligned} x_1 &= \beta_{11}x_{r+1} + \beta_{12}x_{r+2} + \cdots + \beta_{1,n-r}x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_r &= \beta_{r1}x_{r+1} + \beta_{r2}x_{r+2} + \cdots + \beta_{r,n-r}x_n. \end{aligned} \tag{16.25}$$

Выделим частные решения однородной системы (16.24) по следующему принципу. Для нахождения первого вектора-решения $\bar{x}_1$ положим значения свободных переменных $x_{r+1} = 1, x_{r+2} = x_{r+3} = \cdots = x_n = 0$. Затем находим второе решение $\bar{x}_2$: принимаем $x_{r+2} = 1$, а остальные $n - r - 1$ свободные переменные положим равными нулю. Иными словами, мы последовательно присваиваем каждой свободной переменной единичное значение, положив остальные нулями. Таким образом, ФСР в векторной форме, с учетом первых r базисных переменных (16.25), имеет вид

$$\begin{aligned} \bar{x}_1 &= (\beta_{11}, \beta_{12}, \ldots, \beta_{r1}, 1, 0, \ldots, 0), \\ \bar{x}_2 &= (\beta_{12}, \beta_{22}, \ldots, \beta_{r2}, 0, 1, 0, \ldots, 0), \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \bar{x}_{n-r} &= (\beta_{1,n-r}, \beta_{2,n-r}, \cdots, \beta_{r,n-r}, 0, \ldots, 0, 1). \end{aligned} \tag{16.26}$$

Фундаментальная система решений (16.26) является одним из фундаментальных наборов решений однородной системы (16.24).

Пример 1. Найти решение и ФСР системы однородных уравнений

$$\begin{cases} x_1 - x_2 + x_3 - x_4 + x_5 - 2x_6 = 0, \\ 2x_1 - 3x_2 - 2x_3 + x_4 - x_5 = 0, \\ -2x_1 + 3x_2 + 3x_3 + x_4 + x_5 - x_6 = 0. \end{cases}$$

Будем решать эту систему методом Гаусса. Поскольку число уравнений системы меньше числа неизвестных, положим x_1, x_2, x_3 базисными неизвестными, а x_4, x_5, x_6 — свободными переменными. Составим расширенную матрицу системы и выполним действия, предусмотренные прямым ходом метода:

$$\left(\begin{array}{ccc|c} 1 & -1 & 1 & x_4 - x_5 + 2x_6 \\ 2 & -3 & -2 & -x_4 + x_5 \\ -2 & 3 & 3 & -x_4 - x_5 + x_6 \end{array}\right) \begin{matrix}(-2),(2)\\ \\ \\ \end{matrix} \xrightarrow{①}$$

$$\xrightarrow{①} \left(\begin{array}{ccc|c} 1 & -1 & 1 & x_4 - x_5 + 2x_6 \\ 0 & -1 & -4 & -3x_4 + 3x_5 - 4x_6 \\ 0 & 1 & 5 & x_4 - 3x_5 + 5x_6 \end{array}\right) (1) \xrightarrow{②}$$

$$\xrightarrow{②} \left(\begin{array}{ccc|c} 1 & -1 & 1 & x_4 - x_5 + 2x_6 \\ 0 & -1 & -4 & -3x_4 + 3x_5 - 4x_6 \\ 0 & 0 & 1 & -2x_4 + x_6 \end{array}\right).$$

Преобразованная расширенная матрица соответствует системе уравнений, которая эквивалентна исходной однородной системе:

$$\begin{cases} x_1 - x_2 + x_3 = x_4 - x_5 + 2x_6, \\ -x_2 - 4x_3 = -3x_4 + 3x_5 - 4x_6, \\ x_3 = -2x_4 + x_6. \end{cases}$$

Обратный ход метода Гаусса дает значения базисных неизвестных, выраженные через свободные переменные

$$x_3 = -2x_4 + x_6, \quad x_2 = 11x_4 - 3x_5, \quad x_1 = 14x_4 - 4x_5 + x_6.$$

Поскольку ранг однородной системы равен трем, то ФСР для нее состоит из трех линейно независимых векторов. По формулам (16.26) при $n = 6$ и $r = 3$, беря последовательно для свободных переменных тройки чисел $(1, 0, 0)$, $(0, 1, 0)$ и $(0, 0, 1)$, получаем набор фундаментальных решений:

$$\bar{x}_1 = (14, 11, -2, 1, 0, 0),$$
$$\bar{x}_2 = (-4, -3, 0, 0, 1, 0),$$
$$\bar{x}_3 = (1, 0, 1, 0, 0, 1).$$

Характеристическое уравнение

В параграфе 14.1 было введено определение собственного значения и собственного вектора матрицы. Пусть $\bar{x}$ — собственный вектор квадратной матрицы $\mathbf{A}$ порядка n. Тогда имеет место матричное уравнение

$$\mathbf{A}\bar{x} = \lambda\bar{x},$$

или

$$(\mathbf{A} - \lambda\mathbf{E})\bar{x} = \bar{0}, \tag{16.27}$$

где λ — собственное значение матрицы $\mathbf{A}$, а $\mathbf{E}$ и $\bar{0}$ — соответственно единичная матрица и нулевой вектор-столбец. Уравнение (16.27) эквивалентно системе однородных уравнений

$$\begin{cases} (a_{11} - \lambda)x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0, \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \cdots + a_{2n}x_n = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - \lambda)x_n = 0. \end{cases} \tag{16.28}$$

В уравнениях (16.28) a_{ij} — элементы матрицы $\mathbf{A}$, x_j — координаты собственного вектора $\bar{x}$. Поскольку собственный вектор не является нулевым, то однородная система (16.28) должна иметь ненулевое решение, т.е. в силу следствия 2 определитель этой системы равен нулю:

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0. \tag{16.29}$$

Определитель системы однородных уравнений (16.28) называется *характеристическим многочленом*, а уравнение (16.29) — *характеристическим уравнением* матрицы **A**.

Уравнение (16.29) имеет степень n относительно неизвестной λ. Его корни являются собственными числами матрицы **A**. Определив набор этих чисел, для каждого из них можно найти соответствующий собственный вектор как решение однородной системы (16.28).

Пример 2. Найти собственные числа и собственные векторы матрицы

$$\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}.$$

Характеристическое уравнение для этой матрицы имеет вид

$$\begin{vmatrix} 3-\lambda & 2 \\ 1 & 4-\lambda \end{vmatrix} = 0,$$

откуда, раскрывая определитель, получаем

$$\lambda^2 - 7\lambda + 10 = 0.$$

Корни этого уравнения $\lambda_1 = 2, \lambda_2 = 5$. Для нахождения собственных векторов подставим найденные собственные значения в систему однородных уравнений (16.28) при $n = 2$, соответствующую заданной матрице **A**. Собственный вектор, соответствующий собственному значению $\lambda_1 = 2$, является решением системы

$$\begin{cases} x_1 + 2x_2 = 0, \\ x_1 + 2x_2 = 0. \end{cases}$$

По сути дела, это одно уравнение, поскольку определитель системы равен нулю. Полагая $x_2 = b$ свободной переменной, получаем первый собственный вектор $\bar{x}_1 = (-2b, b) = b(-2, 1)$. Подстановка второго собственного значения $\lambda_2 = 5$ приводит к системе уравнений

$$\begin{cases} -2x_1 + 2x_2 = 0, \\ x_1 - x_2 = 0, \end{cases}$$

которая через свободную переменную $x_2 = c$ определяет второй собственный вектор матрицы **A**: $\bar{x}_2 = (c, c) = c(1, 1)$.

Поскольку b и c — произвольные числа, то одному собственному значению может соответствовать несколько собственных векторов разной длины. Например, собственные векторы, соответствующие фундаментальным решениям однородных систем (в данном случае их будет по одному на каждое собственное значение), имеют вид: $\bar{x}_1 = (-2, 1)$, $\bar{x}_2 = (1, 1)$.

Глава 17
ПРИМЕНЕНИЕ ЭЛЕМЕНТОВ ЛИНЕЙНОЙ АЛГЕБРЫ В ЭКОНОМИКЕ

17.1. Использование алгебры матриц

Для решения многих экономических задач используются элементы алгебры матриц. Особенно при разработке и использовании баз данных: при работе с ними почти вся информация хранится и обрабатывается в матричной форме.

Матричные вычисления

Рассмотрим типичные задачи, использующие понятие вектора и его свойства.

Пример 1. Предприятие выпускает ежесуточно четыре вида изделий, основные производственно-экономические показатели которых приведены в табл. 17.1.

Таблица 17.1

Вид изделия	Количество изделий	Расход сырья, кг/изд.	Норма времени изготовления, ч/изд.	Цена изделия, ден. ед./изд.
1	20	5	10	30
2	50	2	5	15
3	30	7	15	45
4	40	4	8	20

Требуется определить следующие ежесуточные показатели: расход сырья S, затраты рабочего времени T и стоимость P выпускаемой продукции предприятия.

Решение. По данным табл. 17.1 составим четыре вектора, характеризующие весь производственный цикл:

- $\bar{q} = (20, 50, 30, 40)$ — вектор ассортимента;
- $\bar{s} = (5, 2, 7, 4)$ — вектор расхода сырья;
- $\bar{t} = (10, 5, 15, 8)$ — вектор затрат рабочего времени;
- $\bar{p} = (30, 15, 45, 20)$ — ценовой вектор.

Тогда искомые величины будут представлять собой соответствующие скалярные произведения вектора ассортимента $\bar{q}$ на три других вектора:

$$S = \bar{q}\bar{s} = 100 + 100 + 210 + 160 = 570 \text{ кг}, \quad T = \bar{q}\bar{t} = 1220 \text{ ч},$$
$$P = \bar{q}\bar{p} = 3500 \text{ ден. ед.}$$

Пример 2. Предприятие выпускает четыре вида изделий с использованием четырех видов сырья. Нормы расхода сырья даны как элементы матрицы A:

$$\mathbf{A} = \begin{array}{c} \text{Вид сырья} \\ \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \left(\begin{array}{cccc} 2 & 3 & 4 & 5 \\ 1 & 2 & 5 & 6 \\ 7 & 2 & 3 & 2 \\ 4 & 5 & 6 & 8 \end{array}\right) \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \text{ Вид изделия}$$

Требуется найти затраты сырья каждого вида при заданном плане выпуска каждого вида изделия: соответственно 60, 50, 35 и 40 ед.

Решение. Составим вектор-план выпуска продукции

$$\bar{q} = (60, 50, 35, 40).$$

Тогда решение задачи дается вектором затрат, координаты которого и являются величинами затрат сырья по каждому

его виду: этот вектор затрат вычисляется как произведение вектора $\bar{q}$ на матрицу A:

$$\bar{q}\mathbf{A} = (60, 50, 35, 40)\begin{pmatrix} 2 & 3 & 4 & 5 \\ 1 & 2 & 5 & 6 \\ 7 & 2 & 3 & 2 \\ 4 & 5 & 6 & 8 \end{pmatrix} =$$
$$= \begin{pmatrix} 120 + 50 + 245 + 160 \\ 180 + 100 + 70 + 200 \\ 240 + 250 + 105 + 240 \\ 300 + 300 + 70 + 320 \end{pmatrix} = \begin{pmatrix} 575 \\ 550 \\ 835 \\ 990 \end{pmatrix}.$$

Пример 3. Пусть затраты четырех видов сырья на выпуск четырех видов продукции характеризуются матрицей A, приведенной в предыдущей задаче. Требуется найти: а) общие затраты на сырье для каждого вида продукции и его перевозку; б) общие затраты на сырье и его транспортировку при условии заданного вектор-плана предыдущей задачи, если известны себестоимости каждого вида сырья и его доставки (соответственно 4, 6, 5, 8 и 2, 1, 3, 2 ден. ед.).

Решение. Составим матрицу себестоимости сырья и его доставки (соответственно 1-я и 2-я строки):

$$\mathbf{C} = \begin{pmatrix} 4 & 6 & 5 & 8 \\ 2 & 1 & 3 & 2 \end{pmatrix}.$$

Тогда ответ на первый вопрос задачи дается в виде произведения матрицы A на транспонированную матрицу $\mathbf{C}^{\mathrm{T}}$:

$$\mathbf{A}\mathbf{C}^{\mathrm{T}} = \begin{pmatrix} 2 & 3 & 4 & 5 \\ 1 & 2 & 5 & 6 \\ 7 & 2 & 3 & 2 \\ 4 & 5 & 6 & 8 \end{pmatrix}\begin{pmatrix} 4 & 2 \\ 6 & 1 \\ 5 & 3 \\ 8 & 2 \end{pmatrix} = \begin{pmatrix} 86 & 29 \\ 89 & 31 \\ 71 & 29 \\ 140 & 47 \end{pmatrix}.$$

Суммарные затраты на сырье и его доставку (в денежных единицах) при вектор-плане выпуска продукции $\bar{q} = (60, 50, 35, 40)$ определяются произведением вектора $\bar{q}$ на матрицу $\mathbf{A}\mathbf{C}^{\mathrm{T}}$:

$$\bar{q}\mathbf{A}\mathbf{C}^{\mathrm{T}} = (60, 50, 35, 40)\begin{pmatrix} 86 & 29 \\ 89 & 31 \\ 71 & 29 \\ 140 & 47 \end{pmatrix} = (17695, 6185).$$

Пример 4. В таблице 17.2 приведены данные о дневной производительности пяти предприятий, выпускающих четыре вида продукции с потреблением трех видов сырья, а также продолжительность работы каждого предприятия в году и цена каждого вида сырья.

Таблица 17.2

Вид изделия	Производительность предприятий, изд./день					Затраты сырья, ед. веса/изд.		
	1	2	3	4	5	1	2	3
1	4	5	3	6	7	2	3	4
2	0	2	4	3	0	3	5	6
3	8	15	0	4	6	4	4	5
4	3	10	7	5	4	5	8	6
	Количество рабочих дней в году					Цена видов сырья		
	200	150	170	120	140	40	50	60

Требуется определить:

1) годовую производительность каждого предприятия по каждому виду изделий;
2) годовую потребность каждого предприятия по каждому виду сырья;
3) годовую сумму кредитования каждого предприятия для закупки сырья, необходимого для выпуска изделий указанных видов и при определенном количестве рабочих дней.

Решение. Нужно составить матрицы, характеризующие весь интересующий нас экономический спектр производства, а затем при помощи соответствующих операций над ними получить решение данной задачи. Прежде всего приведем матрицу производительности предприятий по всем видам продукции:

$$\mathbf{A} = \begin{array}{c} \text{Производительность} \\ \text{предприятия} \\ \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \left(\begin{array}{ccccc} 4 & 5 & 3 & 6 & 7 \\ 0 & 2 & 4 & 3 & 0 \\ 8 & 15 & 0 & 4 & 6 \\ 3 & 10 & 7 & 5 & 4 \end{array}\right) \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \text{ Вид изделия.}$$

Каждый столбец этой матрицы соответствует дневной производительности отдельного предприятия по каждому виду продукции. Следовательно, годовая производительность j-го предприятия по каждому виду продукции получается умножением j-го столбца матрицы $\mathbf{A}$ на количество рабочих дней в году для этого предприятия ($i = 1, 2, 3, 4, 5$). Таким образом, годовая производительность каждого предприятия по каждому виду изделий описывается матрицей

$$\mathbf{A}_{\text{год}} = \begin{pmatrix} 800 & 750 & 510 & 720 & 980 \\ 0 & 300 & 680 & 360 & 0 \\ 1600 & 2250 & 0 & 480 & 840 \\ 600 & 1500 & 1190 & 600 & 560 \end{pmatrix}.$$

Матрица затрат сырья на единицу изделия (эти показатели по условию одинаковы для всех предприятий) имеет вид

$$\mathbf{B} = \begin{array}{c} \begin{array}{cccc} \text{Вид изделия} & & & \\ 1 & 2 & 3 & 4 \end{array} \\ \left(\begin{array}{cccc} 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 8 \\ 4 & 6 & 5 & 6 \end{array}\right) \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \text{ Вид сырья.}$$

Дневной расход по видам сырья на предприятиях описывается произведением матрицы $\mathbf{B}$ на матрицу $\mathbf{A}$:

$$\mathbf{BA} = \begin{pmatrix} 55 & 126 & 53 & 62 & 58 \\ 68 & 165 & 85 & 89 & 77 \\ 74 & 167 & 78 & 92 & 82 \end{pmatrix},$$

где i-я строка соответствует номеру вида сырья, а j-й столбец — номеру предприятия согласно табл. 17.2 ($i = 1, 2, 3$; $j = 1, 2, 3, 4, 5$). Нас интересует ответ на второй вопрос задачи — он получается, по аналогии с матрицей $\mathbf{A}_{\text{год}}$, умножением столбцов матрицы $\mathbf{BA}$ на соответствующее количество рабочих дней в году для предприятий. Таким образом, годовая потребность каждого предприятия в каждом виде сырья

$$\mathbf{BA}_{\text{год}} = \begin{pmatrix} 11000 & 18900 & 9010 & 7440 & 8120 \\ 13600 & 24750 & 14450 & 10680 & 10780 \\ 14800 & 25050 & 13260 & 11040 & 11480 \end{pmatrix}.$$

Введем вектор стоимости сырья

$$\bar{p} = (40, 50, 60).$$

Тогда стоимость общего годового запаса сырья для каждого предприятия получается умножением вектора $\bar{p}$ на матрицу $\mathbf{BA}_{\text{год}}$:

$$\bar{P} = \bar{p}\mathbf{BA}_{\text{год}} = \\ = (2008000, 3496500, 1878500, 1494000, 1552600).$$

Следовательно, суммы кредитования предприятий для закупки сырья определяются соответствующими компонентами вектора $\bar{P}$.

Пример 5. Отрасль состоит из n предприятий, выпускающих по одному виду продукции каждое: обозначим объем продукции i-го предприятия через x_i. Каждое из предприятий отрасли для обеспечения своего производства потребляет часть продукции, выпускаемой им самим и другими предприятиями. Например, в отрасли электротехнического оборудования часть продукции предприятий, выпускающих электродвигатели, силовые кабели, электрокары и т. д., употребляется практически всей отраслью. Пусть a_{ij} — доля продукции i-го предприятия, потребляемая j-м предприятием для обеспечения выпуска своей продукции объема x_j. Возникает естественный вопрос о величине y_i — количестве продукции i-го предприятия, предназначенной для реализации вне данной отрасли (объем конечного продукта). Эта величина легко может быть подсчитана по формуле

$$y_i = x_i - \sum_{j=1}^{n} a_{ij}x_j; \quad i = 1, 2, \ldots, n.$$

Введем в рассмотрение матрицу порядка n, описывающую внутреннее потребление отрасли

$$\mathbf{A} = ||a_{ij}||; \quad i, j = 1, 2, \ldots, n. \tag{17.1}$$

Тогда вектор конечного продукта является решением матричного уравнения

$$\bar{x} - \mathbf{A}\bar{x} = \bar{y},$$

с использованием единичной матрицы $\mathbf{E}$ получаем

$$(\mathbf{E} - \mathbf{A})\bar{x} = \bar{y}. \tag{17.2}$$

Рассмотрим конкретный пример при $n = 3$. Пусть вектор выпуска продукции отрасли и матрица внутреннего потребления имеют соответственно вид

$$\bar{x} = \begin{pmatrix} 300 \\ 200 \\ 400 \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} 0{,}3 & 0{,}1 & 0{,}2 \\ 0{,}2 & 0{,}3 & 0{,}1 \\ 0{,}2 & 0{,}2 & 0{,}6 \end{pmatrix}.$$

Используя формулу (17.2) и правило сложения матриц, получаем вектор объемов конечного продукта, предназначенного для реализации вне отрасли, состоящей из трех предприятий:

$$\bar{y} = (\mathbf{E} - \mathbf{A})\bar{x} = \begin{pmatrix} 0{,}7 & -0{,}1 & -0{,}2 \\ -0{,}2 & 0{,}7 & -0{,}1 \\ -0{,}2 & -0{,}2 & 0{,}4 \end{pmatrix} \begin{pmatrix} 300 \\ 200 \\ 400 \end{pmatrix} = \begin{pmatrix} 110 \\ 40 \\ 60 \end{pmatrix}.$$

Использование систем линейных уравнений

Рассмотрим задачи, приводящие к составлению и решению систем линейных алгебраических уравнений.

Пример 6. Предприятие выпускает три вида продукции, используя сырье трех видов. Необходимые характеристики производства указаны в табл. 17.3. Требуется определить объем выпуска продукции каждого вида при заданных запасах сырья. Задачи такого рода типичны при прогнозах и оценках функционирования предприятий, экспертных оценках проектов освоения месторождений полезных ископаемых, а также в планировании микроэкономики предприятий.

Таблица 17.3

Вид сырья	Расход сырья по видам продукции, вес. ед./изд.			Запас сырья, вес. ед.
	1	2	3	
1	6	4	5	2400
2	4	3	1	1450
3	5	2	3	1550

Решение. Обозначим неизвестные объемы выпуска продукции через x_1, x_2 и x_3.Тогда при условии полного расхода запасов каждого вида сырья можно записать балансовые соотношения, которые образуют систему трех уравнений с тремя неизвестными:

$$\begin{cases} 6x_1 + 4x_2 + 5x_3 = 2400, \\ 4x_1 + 3x_2 + x_3 = 1450, \\ 5x_1 + 2x_2 + 3x_3 = 1550. \end{cases}$$

Решая эту систему уравнений любым способом, находим, что при заданных запасах сырья объемы выпуска продукции составят по каждому виду соответственно (в условных единицах):

$$x_1 = 150, \quad x_2 = 250, \quad x_3 = 100.$$

Общая постановка задачи прогноза выпуска продукции. Пусть

$$\mathbf{C} = \|c_{ij}\|; \qquad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n, \tag{17.3}$$

— матрица затрат сырья m видов при выпуске продукции n видов. Тогда при известных объемах запаса каждого вида сырья, которые образуют соответствующий вектор

$$\bar{q} = (q_1, q_2, \ldots, q_m), \tag{17.4}$$

вектор-план $\bar{x} = (x_1, x_2, \ldots, x_n)$ выпуска продукции определяется из решения системы m уравнений с n неизвестными:

$$C\bar{x}^{\mathrm{T}} = \bar{q}^{\mathrm{T}}, \tag{17.5}$$

где индекс "т" означает транспонирование вектора-строки в вектор-столбец.

Пример 7. На предприятие с работниками четырех категорий привезли заработную плату в купюрах следующего достоинства: по 100 рублей — 1850 купюр, по 50 рублей — 230 купюр, по 10 рублей — 250 купюр, по 1 рублю — 740 купюр. Заработная плата работника 1-й категории составляет 962 руб.,

2-й категории — 713 руб., 3-й категории — 452 руб., 4-й категории — 261 руб. Определить, сколько сотрудников каждой категории работает на предприятии, если каждому сотруднику выдали заработную плату минимальным числом купюр.

Решение. Условие об оплате минимальным числом купюр является основным в определении количества купюр разного достоинства, выданных сотрудникам разных категорий. Исходя из величины заработной платы по категориям, однозначно определяем таблицу распределения купюр.

Пусть x_1, x_2, x_3, x_3 — количество работников категорий соответственно с первой по четвертую. Тогда по данным табл. 17.4 составляем уравнения "баланса", которые образуют систему четырех уравнений с четырьмя неизвестными:

$$\begin{cases} 9x_1 + 7x_2 + 4x_3 + 2x_4 = 1850, \\ x_1 + x_3 + x_4 = 230, \\ x_1 + x_2 + x_4 = 250, \\ 2x_1 + 3x_2 + 2x_3 + x_4 = 740. \end{cases}$$

Таблица 17.4

Достоинство купюры, руб.	Распределение купюр по категориям				Общее количество купюр
	1	2	3	4	
100	9	7	4	2	1850
50	1	–	1	1	230
10	1	1	–	1	250
1	2	3	2	1	740

Эту систему удобно решать методом Гаусса, для чего выпишем расширенную матрицу системы, предварительно переместив для удобства первое уравнение на последнее место. Прямой ход метода последовательно меняет вид матрицы (см. параграф 16.2):

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 230 \\ 1 & 1 & 0 & 1 & 250 \\ 2 & 3 & 2 & 1 & 740 \\ 9 & 7 & 4 & 2 & 1850 \end{array}\right) \begin{matrix} \\ (-1), (-2), (-9) \\ \\ \\ \end{matrix} \xrightarrow{\textcircled{1}}$$

$$\xrightarrow{①}\left(\begin{array}{cccc|c}1&0&1&1&230\\0&1&-1&0&20\\0&3&0&-1&280\\0&7&-5&-7&-220\end{array}\right)(-3),(-7)\xrightarrow{②}$$

$$\xrightarrow{②}\left(\begin{array}{cccc|c}1&0&1&1&230\\0&1&-1&0&20\\0&0&3&-1&220\\0&0&2&-7&-360\end{array}\right)\xrightarrow{③}$$

$$\xrightarrow{③}\left(\begin{array}{cccc|c}1&0&1&1&230\\0&1&-1&0&20\\0&0&2&-7&-360\\0&0&3&-1&220\end{array}\right)(-3/2)\xrightarrow{④}$$

$$\xrightarrow{④}\left(\begin{array}{cccc|c}1&0&1&1&230\\0&1&-1&0&20\\0&0&2&-7&-360\\0&0&0&19/2&760\end{array}\right).$$

Третий переход состоял в перемене мест 3-й и 4-й строк. Полученная в цепочке прямого хода расширенная матрица соответствует системе уравнений, эквивалентной исходной системе:

$$\begin{cases}x_1+x_3+x_4=230,\\x_2-x_3=20,\\2x_3-7x_4=-360,\\19/2x_4=760.\end{cases}$$

Обратным ходом метода получаем последовательно неизвестные:

$$x_4=80,\quad x_3=100,\quad x_2=120,\quad x_1=50.$$

17.2. Модель Леонтьева многоотраслевой экономики

Макроэкономика функционирования многоотраслевого хозяйства требует баланса между отдельными отраслями. Каждая отрасль, с одной стороны, является производителем, а с другой — потребителем продукции, выпускаемой другими отраслями. Возникает довольно непростая задача расчета связи

между отраслями через выпуск и потребление продукции разного вида. Впервые эта проблема была сформулирована в виде *математической модели* в трудах известного американского экономиста В. Леонтьева в 1936 г., который попытался проанализировать причины экономической депрессии США 1929–1932 гг. Эта модель основана на алгебре матриц и использует аппарат матричного анализа.

Балансовые соотношения

Для простоты будем полагать, что производственная сфера хозяйства представляет собой n отраслей, каждая из которых производит свой однородный продукт. Для обеспечения производства каждая отрасль нуждается в продукции других отраслей (производственное потребление). Обычно процесс производства рассматривается за некоторый период; в ряде случаев такой единицей служит год.

Введем следующие обозначения:

- x_i — общий объем продукции i-й отрасли (ее валовой выпуск);
- x_{ij} — объем продукции i-й отрасли, потребляемый j-й отраслью при производстве объема продукции x_j;
- y_i — объем продукции i-й отрасли, предназначенный для реализации (потребления) в непроизводственной сфере, или так называемый продукт конечного потребления. К нему относятся личное потребление граждан, удовлетворение общественных потребностей, содержание государственных институтов и т.д.

Балансовый принцип связи различных отраслей промышленности состоит в том, что валовой выпуск i-й отрасли должен быть равен сумме объемов потребления в производственной и непроизводственной сферах. В самой простой форме (гипотеза линейности, или простого сложения) балансовые соотношения имеют вид

$$x_i=x_{i1}+x_{i2}+\cdots+x_{in}+y_i,\qquad i=1,2,\ldots,n. \tag{17.6}$$

Уравнения (17.6) называются *соотношениями баланса*.

Поскольку продукция разных отраслей имеет разные измерения, в дальнейшем будем иметь в виду стоимостный баланс.

Линейная модель многоотраслевой экономики

В. Леонтьевым, на основании анализа экономики США в период перед второй мировой войной, был установлен важный факт: в течение длительного времени величины $a_{ij} = x_{ij}/x_j$ меняются очень незначительно и могут рассматриваться как постоянные числа. Это явление становится понятным в свете того, что технология производства остается на одном и том же уровне довольно длительное время, и, следовательно, объем потребления j-й отраслью продукции i-й отрасли при производстве своей продукции объема x_j есть технологическая константа.

В силу указанного факта можно сделать следующее допущение: для производства продукции j-й отрасли объема x_j нужно использовать продукцию i-й отрасли объема $a_{ij}x_i$, где a_{ij} — постоянное число. При таком допущении технология производства принимается *линейной*, а само это допущение называется *гипотезой линейности*. При этом числа a_{ij} называются *коэффициентами прямых затрат*. Согласно гипотезе линейности,

$$x_{ij} = a_{ij}x_j; \qquad i, j = 1, 2, \ldots, n. \tag{17.7}$$

Тогда уравнения (17.6) можно переписать в виде системы уравнений:

$$\begin{cases} x_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + y_1, \\ x_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + y_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n + y_n. \end{cases} \tag{17.8}$$

Введем в рассмотрение векторы-столбцы объемов произведенной продукции (вектор валового выпуска), объемов продукции конечного потребления (вектор конечного потребления) и матрицу коэффициентов прямых затрат:

$$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}. \tag{17.9}$$

Тогда система уравнений (17.8) в матричной форме имеет вид

$$\bar{x} = \mathbf{A}\bar{x} + \bar{y}. \tag{17.10}$$

Обычно это соотношение называют *уравнением линейного межотраслевого баланса*. Вместе с описанием матричного представления (17.9) это уравнение носит название *модели Леонтьева*.

Уравнение межотраслевого баланса можно использовать в двух целях. В первом, наиболее простом случае, когда известен вектор валового выпуска $\bar{x}$, требуется рассчитать вектор конечного потребления $\bar{y}$ — подобная задача была рассмотрена выше (см. параграф 17.1, пример 5).

Во втором случае уравнение межотраслевого баланса используется для целей планирования со следующей формулировкой задачи: для периода T (например, год) известен вектор конечного потребления $\bar{y}$ и требуется определить вектор $\bar{x}$ валового выпуска. Здесь необходимо решать систему линейных уравнений (17.10) с известной матрицей $\mathbf{A}$ и заданным вектором $\bar{y}$. В дальнейшем мы будем иметь дело именно с такой задачей.

Между тем система (17.10) имеет особенности, вытекающие из прикладного характера данной задачи: все элементы матрицы $\mathbf{A}$ и векторов $\bar{x}$ и $\bar{y}$ должны быть неотрицательными.

Продуктивные модели Леонтьева

Матрица $\mathbf{A}$, все элементы которой неотрицательны, называется *продуктивной*, если для любого вектора $\bar{y}$ с неотрицательными компонентами существует решение уравнения (17.10) — вектор $\bar{x}$, все элементы которого неотрицательны. В таком случае и модель Леонтьева называется продуктивной.

Для уравнения типа (17.10) разработана соответствующая математическая теория исследования решения и его особенностей. Укажем некоторые ее основные моменты. Приведем без доказательства важную теорему, позволяющую устанавливать продуктивность матрицы.

Теорема 17.1. *Если для матрицы $\mathbf{A}$ с неотрицательными элементами и некоторого вектора $\bar{y}$ с неотрицательными*

компонентами уравнение (17.10) *имеет решение* $\bar{x}$ *с неотрицательными компонентами, то матрица* $\mathbf{A}$ *продуктивна.*

Иными словами, достаточно установить наличие положительного решения системы (17.10) хотя бы для одного положительного вектора $\bar{y}$, чтобы матрица $\mathbf{A}$ была продуктивной. Перепишем систему (17.10) с использованием единичной матрицы $\mathbf{E}$ в виде

$$(\mathbf{E}-\mathbf{A})\bar{x}=\bar{y}. \tag{17.11}$$

Если существует обратная матрица $(\mathbf{E}-\mathbf{A})^{-1}$, то существует и единственное решение уравнения (17.11)

$$\bar{x}=(\mathbf{E}-\mathbf{A})^{-1}\bar{y}. \tag{17.12}$$

Матрица $(\mathbf{E}-\mathbf{A})^{-1}$ называется *матрицей полных затрат.*

Существует несколько критериев продуктивности матрицы $\mathbf{A}$. Приведем два из них.

Первый критерий продуктивности. Матрица $\mathbf{A}$ продуктивна тогда и только тогда, когда матрица $(\mathbf{E}-\mathbf{A})^{-1}$ существует и ее элементы неотрицательны.

Второй критерий продуктивности. Матрица $\mathbf{A}$ с неотрицательными элементами продуктивна, если сумма элементов по любому ее столбцу (строке) не превосходит единицы:

$$\sum_{i=1}^{n} a_{ij} \leqslant 1, \tag{17.13}$$

причем хотя бы для одного столбца (строки) эта сумма строго меньше единицы.

Рассмотрим применение модели Леонтьева на несложных примерах.

Пример 1. В таблице 17.5 приведены данные по балансу за некоторый период между пятью отраслями промышленности. Найти векторы конечного потребления и валового выпуска, а также матрицу коэффициентов прямых затрат и определить, является ли она продуктивной в соответствии с приведенными выше критериями.

Таблица 17.5

№ п/п	Отрасль	Потребление					Конечный продукт	Валовой выпуск
		1	2	3	4	5		
1	Станкостроение	15	12	24	23	16	10	100
2	Энергетика	10	3	35	15	7	30	100
3	Машиностроение	10	5	10	10	10	5	50
4	Автомобильная промышленность	10	5	10	5	5	15	50
5	Добыча и переработка углеводородов	7	15	15	10	3	50	100

Решение. Данные в таблице 17.5 приведены в соответствии с соотношениями (17.6): x_{ij} — первые пять столбцов, y_i — шестой столбец, x_i — последний столбец ($i, j=1,2,3,4,5$). Согласно формулам (17.7) и (17.9),

$$\bar{x}=\begin{pmatrix}100\\100\\50\\50\\100\end{pmatrix}, \qquad \bar{y}=\begin{pmatrix}10\\30\\5\\15\\50\end{pmatrix},$$

$$\mathbf{A}=\begin{pmatrix}0{,}15 & 0{,}12 & 0{,}48 & 0{,}46 & 0{,}16\\ 0{,}10 & 0{,}03 & 0{,}70 & 0{,}30 & 0{,}07\\ 0{,}10 & 0{,}05 & 0{,}20 & 0{,}20 & 0{,}10\\ 0{,}10 & 0{,}05 & 0{,}20 & 0{,}10 & 0{,}05\\ 0{,}07 & 0{,}15 & 0{,}30 & 0{,}20 & 0{,}03\end{pmatrix}.$$

Все элементы матрицы $\mathbf{A}$ положительны, однако нетрудно видеть, что их сумма в третьем и четвертом столбцах больше единицы. Следовательно, условия второго критерия продуктивности не соблюдены, и матрица $\mathbf{A}$ не является продуктивной. Экономическая причина этой непродуктивности заключается в том, что внутреннее потребление отраслей 3 и 4 слишком велико в соотношении с их валовым выпуском.

Пример 2. Таблица 17.6 содержит данные баланса трех отраслей промышленности за некоторый период. Требуется найти объем валового выпуска продукции, если конечное потребление по отраслям увеличить соответственно до 60, 70 и 30.

Таблица 17.6

№ п/п	Отрасль	Потребление			Конечный продукт	Валовой выпуск
		1	2	3		
1	Добыча и переработка углеводородов	5	35	20	40	100
2	Энергетика	10	10	20	60	100
3	Машиностроение	20	10	10	10	50

Решение. Выпишем векторы валового выпуска и конечного потребления и матрицу коэффициентов прямых затрат. Согласно формулам (17.7) и (17.9),

$$\bar{x} = \begin{pmatrix} 100 \\ 100 \\ 50 \end{pmatrix}, \quad \bar{y} = \begin{pmatrix} 40 \\ 60 \\ 10 \end{pmatrix},$$

$$A = \begin{pmatrix} 0{,}05 & 0{,}35 & 0{,}40 \\ 0{,}10 & 0{,}10 & 0{,}40 \\ 0{,}20 & 0{,}10 & 0{,}20 \end{pmatrix}.$$

Матрица $\mathbf{A}$ удовлетворяет второму критерию продуктивности. В случае заданного увеличения конечного потребления новый вектор конечного продукта будет иметь вид

$$\bar{y}_* = \begin{pmatrix} 60 \\ 70 \\ 30 \end{pmatrix}.$$

Требуется найти новый вектор валового выпуска $\bar{x}_*$, удовлетворяющий соотношениям баланса в предположении, что матрица $\mathbf{A}$ не изменяется. В таком случае компоненты x_1, x_2, x_3 неизвестного вектора $\bar{x}$ находятся из системы уравнений, которая, согласно (17.8), имеет в данном случае вид

$$\begin{cases} x_1 = 0{,}05x_1 + 0{,}35x_2 + 0{,}4x_3 + 60, \\ x_2 = \ 0{,}1x_1 + \ 0{,}1x_2 + 0{,}4x_3 + 70, \\ x_3 = \ 0{,}2x_1 + \ 0{,}1x_2 + 0{,}2x_3 + 30. \end{cases}$$

В матричной форме эта система выглядит следующим образом:

$$\bar{x}_* = \mathbf{A}\bar{x}_* + \bar{y}_*,$$

или

$$(\mathbf{E} - \mathbf{A})\bar{x}_* = \bar{y}_*,$$

где матрица $(\mathbf{E} - \mathbf{A})$ имеет вид

$$(\mathbf{E} - \mathbf{A}) = \begin{pmatrix} 0{,}95 & -0{,}35 & -0{,}40 \\ -0{,}10 & 0{,}90 & -0{,}40 \\ -0{,}20 & -0{,}10 & 0{,}80 \end{pmatrix}.$$

Отсюда рассчитывается новый вектор $\bar{x}_*$ как решение этого уравнения баланса:

$$\bar{x}_* = (\mathbf{E} - \mathbf{A})^{-1}\bar{y}_*.$$

Найдем обратную матрицу (матрицу полных затрат) $(\mathbf{E} - \mathbf{A})^{-1}$, с использованием формулы (15.12). Определитель матрицы $(\mathbf{E} - \mathbf{A})$

$$\det(\mathbf{E} - \mathbf{A}) = 0{,}514 \neq 0,$$

так что обратная матрица и решение указанной системы уравнений существуют. Вычисление обратной матрицы дает с точностью до третьего знака:

$$(\mathbf{E} - \mathbf{A})^{-1} = \frac{1}{0{,}514}\begin{pmatrix} 0{,}68 & 0{,}32 & 0{,}50 \\ 0{,}16 & 0{,}68 & 0{,}42 \\ 0{,}19 & 0{,}165 & 0{,}82 \end{pmatrix} =$$

$$= \begin{pmatrix} 1{,}323 & 0{,}623 & 0{,}973 \\ 0{,}311 & 1{,}323 & 0{,}817 \\ 0{,}370 & 0{,}321 & 1{,}595 \end{pmatrix}.$$

Заметим, что найденная обратная матрица удовлетворяет первому критерию продуктивности матрицы $\mathbf{A}$.

Теперь можно вычислить вектор валового выпуска $\bar{x}_*$:

$$\bar{x}_* = \begin{pmatrix} 1{,}323 & 0{,}623 & 0{,}973 \\ 0{,}311 & 1{,}323 & 0{,}817 \\ 0{,}370 & 0{,}321 & 1{,}595 \end{pmatrix}\begin{pmatrix} 60 \\ 70 \\ 30 \end{pmatrix} = \begin{pmatrix} 152{,}2 \\ 135{,}8 \\ 92{,}5 \end{pmatrix}.$$

Таким образом, для того чтобы обеспечить заданное увеличение компонент вектора конечного продукта, необходимо увеличить соответствующие валовые выпуски: добычу и переработку углеводородов на 52,2%, уровень энергетики — на 35,8% и выпуск машиностроения — на 85% по сравнению с исходными величинами, указанными в табл. 17.6.

17.3. Линейная модель торговли

Процесс взаимных закупок товаров анализируется с использованием понятий собственного числа и собственного вектора матрицы. Будем полагать, что бюджеты n стран, которые мы обозначим соответственно $x_1, x_2, \ldots, x_n$, расходуются на покупку товаров. Рассмотрим *линейную модель обмена*, или *модель международной торговли*.

Пусть a_{ij} — доля бюджета x_j, которую j-я страна тратит на закупку товаров у i-й страны. Введем матрицу коэффициентов a_{ij}:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}. \qquad (17.14)$$

Тогда, если весь бюджет расходуется только на закупки внутри страны и вне ее (это можно трактовать как торговый бюджет), справедливо равенство

$$\sum_{i=1}^{n} a_{ij} = 1; \quad j = 1, 2, \ldots, n. \qquad (17.15)$$

Матрица (17.14) со свойством (17.15), в силу которого сумма элементов ее любого столбца равна единице, называется структурной матрицей торговли. Для i-й страны общая выручка от внутренней и внешней торговли выражается формулой

$$P_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n. \qquad (17.16)$$

Условие сбалансированной (бездефицитной) торговли формулируется естественным образом: для каждой страны ее бюджет должен быть не больше выручки от торговли, т. е. $P_i \geqslant x_i$,

или

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \geqslant x_i; \quad i = 1, 2, \ldots, n. \qquad (17.17)$$

Докажем, что в условиях (17.17) не может быть знака неравенства. Действительно, сложим все эти неравенства при i от 1 до n. Группируя слагаемые с величинами бюджетов x_j, получаем

$$x_1(a_{11} + a_{21} + \cdots + a_{n1}) + x_2(a_{12} + a_{22} + \cdots + a_{n2}) + \cdots + \\ + x_n(a_{1n} + a_{2n} + \cdots a_{nn}) \geqslant x_1 + x_2 + \cdots + x_n.$$

Нетрудно заметить, что в скобках стоят суммы элементов матрицы $\mathbf{A}$ по ее столбцам, которые равны единице по условию (17.15). Стало быть, мы получили неравенство

$$x_1 + x_2 + \cdots + x_n \geqslant x_1 + x_2 + \cdots + x_n,$$

откуда следует, что возможен только знак равенства. Таким образом, условия (17.17) принимают вид равенств:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = x_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = x_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = x_n. \end{cases} \qquad (17.18)$$

Введем вектор бюджетов $\bar{x}$, каждая компонента которого характеризует бюджет соответствующей страны: тогда систему уравнений (17.18) можно записать в матричной форме:

$$\mathbf{A}\bar{x} = \bar{x}. \qquad (17.19)$$

Это уравнение означает, что собственный вектор структурной матрицы $\mathbf{A}$, отвечающий ее собственному значению $\lambda = 1$, состоит из бюджетов стран бездефицитной международной торговли.

Перепишем уравнение (17.19) в виде, позволяющем определить $\bar{x}$:

$$(\mathbf{A} - \mathbf{E})\bar{x} = \bar{0}. \qquad (17.20)$$

Пример 1. Структурная матрица торговли четырех стран имеет вид

$$\mathbf{A} = \begin{pmatrix} 0{,}2 & 0{,}3 & 0{,}2 & 0{,}2 \\ 0{,}4 & 0{,}3 & 0{,}1 & 0{,}2 \\ 0{,}3 & 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}1 & 0{,}1 & 0{,}2 & 0{,}4 \end{pmatrix}.$$

Найти бюджеты этих стран, уловлетворяющие сбалансированной бездефицитной торговле при условии, что сумма бюджетов задана:

$$x_1 + x_2 + x_3 + x_4 = 6270.$$

Решение. Необходимо найти собственный вектор x, отвечающий собственному значению $\lambda = 1$ заданной структурной матрицы $\mathbf{A}$, т. е. решить уравнение (17.20), которое в нашем случае имеет вид

$$\begin{pmatrix} -0{,}8 & 0{,}3 & 0{,}2 & 0{,}2 \\ 0{,}4 & -0{,}7 & 0{,}1 & 0{,}2 \\ 0{,}3 & 0{,}3 & -0{,}5 & 0{,}2 \\ 0{,}1 & 0{,}1 & 0{,}2 & -0{,}6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Поскольку ранг этой системы равен трем, то одна из неизвестных является свободной переменной, остальные выражаются через нее. Решая систему методом Гаусса, находим компоненты собственного вектора $\bar{x}$:

$$x_1 = \frac{140}{121}c, \quad x_2 = \frac{146}{121}c, \quad x_3 = \frac{20}{11}c, \quad x_4 = c.$$

Подставив найденные значения в заданную сумму бюджетов, определим величину c:

$$c = 1210.$$

Откуда окончательно получаем искомые величины бюджетов стран при бездефицитной торговле:

$$x_1 = 1400, \quad x_2 = 1460, \quad x_3 = 2200, \quad x_4 = 1210.$$

Часть 3. ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ И РАЗНОСТНЫЕ УРАВНЕНИЯ

Дифференциальные уравнения занимают особое место в математике и имеют многочисленные приложения в большом спектре наук. Исследования природных процессов и изучение закономерностей общественных процессов приводят к построению математических моделей, в основе которых лежат дифференциальные уравнения. В дифференциальных уравнениях неизвестная функция содержится вместе со своими производными. Основной задачей теории дифференциальных уравнений является изучение функций, представляющих собой решения этих уравнений. В этой части излагаются элементы теории *обыкновенных* дифференциальных уравнений, когда неизвестные функции зависят от одной переменной. Теория дифференциальных уравнений, когда неизвестные функции зависят от нескольких переменных (*уравнения в частных производных*), более сложна.

Глава 18
ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ ПЕРВОГО ПОРЯДКА

18.1. Основные понятия

Определение дифференциального уравнения первого порядка

Определение 1. Уравнение вида

$$F(x, y, y') = 0, \tag{18.1}$$

где x — независимая переменнная, y и y' — соответственно неизвестная функция и ее производная, называется *дифференциальным уравнением первого порядка.*

Примеры дифференциальных уравнений первого порядка:

$$(y')^3y^2 + 2xy = 0, \quad y' = x^2 \sin y, \quad \sqrt{x + y'} = 3xy^2.$$

В случае когда из этого уравнения можно выразить y', оно имеет вид

$$y' = f(x, y). \tag{18.2}$$

Уравнение (18.2) называется уравнением первого порядка, *разрешенным относительно производной*. В дальнейшем будем рассматривать уравнения первого порядка именно такого вида. Примеры уравнений, разрешенных относительно производной:

$$y' = x^2 + y^2, \quad y' = xy^3 + \operatorname{tg}(xy), \quad y' = 2x + y, \quad y' = \frac{\ln x}{xy}.$$

Примеры уравнений, которые можно разрешить относительно производной неизвестной функции y':

1) $(y')^2 = x^2 + y^2$, откуда получаем два уравнения первого порядка: $y' = \pm\sqrt{x^2 + y^2}$;

2) $\operatorname{tg} y' - xy^2 = 0$, откуда $y' = \operatorname{arctg}(xy^2)$;

3) $\sqrt{y'} = x + \cos y$, откуда $y' = (x + \cos y)^2$.

Существование решения дифференциального уравнения

Определение 2. Решением дифференциального уравнения первого порядка называется функция $y = \varphi(x)$, определенная на некотором интервале (a, b), которая при подстановке в уравнение обращает его в тождество.

Например, функция $y = x^2$ тождественно обращает в нуль левую часть уравнения $xy' - 2x^2 = 0$ и потому представляет собой решение этого уравнения.

В теории дифференциальных уравнений основной задачей является вопрос о существовании и единственности решения. Ответ на него дает теорема Коши, которую мы приводим без доказательства.

Теорема 18.1 (ТЕОРЕМА КОШИ). *Пусть дано дифференциальное уравнение* (18.2). *Если функция* $f(x, y)$ *и ее частная производная* $f'_y(x, y)$ *непрерывны в некоторой области* D *плоскости* Oxy, *то в некоторой окрестности любой внутренней точки* (x_0, y_0) *этой области существует единственное решение уравнения* (18.2), *удовлетворяющее условию* $y = y_0$ *при* $x = x_0$.

График решения дифференциального уравнения называется *интегральной кривой*. В области D содержится бесконечно много интегральных кривых. Теорема Коши гарантирует, что при соблюдении определенных условий через каждую внутреннюю точку области D проходит только одна интегральная кривая. Условия, которые задают значение функции y_0 в фиксированной точке x_0, называют *начальными условиями* (*условиями Коши*) и записывают в такой форме:

$$y\Big|_{x_0} = y_0. \tag{18.3}$$

Задача нахождения решения уравнения (18.2), удовлетворяющих условию (18.3), называется *задачей Коши* — из множества интегральных кривых выделяется та, которая проходит через заданную точку (x_0, y_0) области D.

В ряде случаев, когда условия теоремы Коши не выполнены, через некоторые точки плоскости Oxy либо не проходит ни одна интегральная кривая, либо проходит более одной интегральной кривой; эти точки называются *особыми точками* данного дифференциального уравнения.

Определение 3. *Общим решением* уравнения (18.2) называется функция $y = \varphi(x, C)$, удовлетворяющая этому уравнению при произвольном значении постоянной C.

Определение 4. *Частным решением* уравнения (18.2) в области D называется функция $y = \varphi(x, C_0)$, полученная при определенном значении постоянной $C = C_0$.

Общее решение $y = \varphi(x, C)$ описывает семейство интегральных кривых на плоскости Oxy. Условия Коши (18.3) фиксируют произвольную постоянную C и позволяют выбрать из семейства интегральных кривых уравнения (18.2) одну интегральную кривую $y = \varphi(x, C_0)$, проходящую через заданную точку (x_0, y_0).

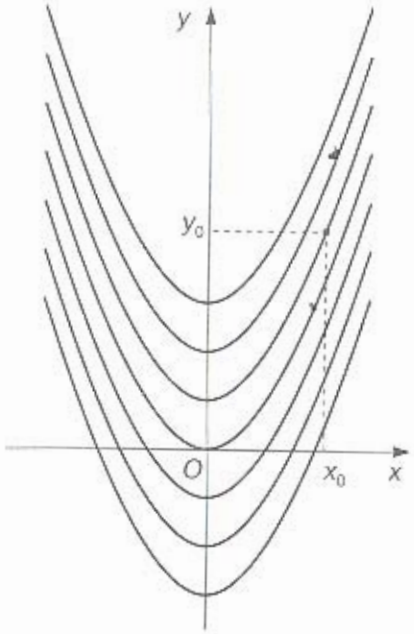

Рис. 18.1

Например, рассмотрим уравнение $y' = 2x$. Правая часть этого уравнения удовлетворяет условиям теоремы Коши во всех точках плоскости Oxy (функции $f(x,y) = 2x$ и $f'_y(x,y) \equiv 0$ определены и непрерывны на всей плоскости Oxy). Общим решением уравнения является функция $y = x^2 + C$, где C — произвольная постоянная, описывающая семейство парабол (рис. 18.1). Для отыскания частного решения зададим произвольные начальные условия (18.3) и подставим их в формулу общего решения; получаем, что $C = y_0 - x_0^2$, откуда находим частное решение $y = x^2 + y_0 - x_0^2$. Это частное решение выделяет из семейства парабол одну, проходящую через точку (x_0, y_0).

Геометрический смысл уравнения первого порядка

Рассмотрим уравнение $y' = f(x,y)$. Пусть $y = \varphi(x)$ — его решение, график которого представляет собой непрерывную интегральную кривую, причем в каждой точке ее существует касательная. Из дифференциального уравнения следует,

что угловой коэффициент касательной к интегральной кривой в каждой ее точке равен правой части этого уравнения. Следовательно, уравнение первого порядка задает угловой коэффициент y' касательной к интегральной кривой как функцию двух переменных. Если каждой точке (x,y) сопоставить отрезок, направленный под углом наклона $\alpha = \operatorname{arctg}(f(x,y))$ к оси Ox, то мы получим *поле направлений* данного уравнения. В этом и заключается геометрический смысл дифференциального уравнения первого порядка.

Поле направлений позволяет проанализировать решение дифференциального уравнения и даже приближенно построить интегральные кривые.

Пример 1. Построить поле направлений уравнения $y'=x^2-y$. Нетрудно видеть, что правая часть этого уравнения удовлетворяет условиям теоремы Коши единственности и существования решения при любых x и y, т. е. интегральные кривые заполняют всю плоскость Oxy. Найдем линии, на которых наклон направлений одинаков, так называемые *изоклины*. На рис. 18.2 изоклины показаны черточками. Так, если $y' = 0$, то $x^2 - y = 0$, т. е. на параболе $y = x^2$ касательные к интегральным кривым горизонтальны. При $y' = 1$ имеем $x^2 - y = 1$, т. е. касательные к интегральным кривым направлены под углом 45° к оси Ox на параболе $y = x^2 - 1$. Наконец, на параболе $y = x^2 + 1$ угол наклона касательных равен 135°. По полю направлений можно приближенно восстановить ход интегральных кривых (сплошные линии на рис. 18.2).

18.2. Виды уравнений и методы решения

Уравнения с разделяющимися переменными

Определение 1. Дифференциальное уравнение вида

$$y' = f_1(x) f_2(y), \tag{18.4}$$

где $f_1(x)$ и $f_2(y)$ — непрерывные функции, называется *уравнением с разделяющимися переменными*.

Подчеркнем, что правая часть уравнения представляет собой произведение, в котором один сомножитель зависит только

31 Зак. 6853

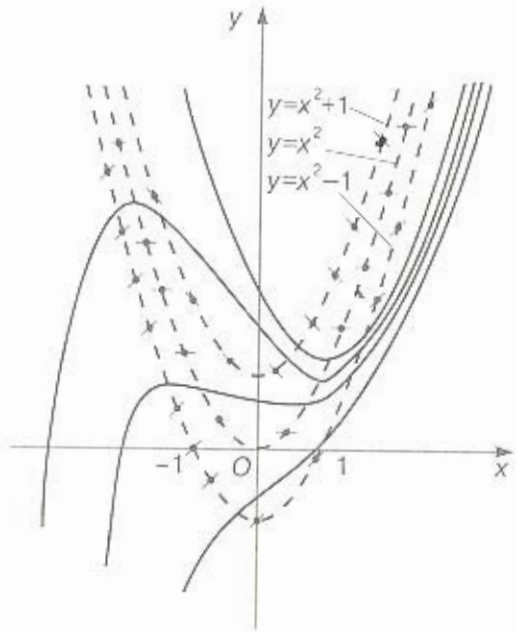

Рис. 18.2

от x, а другой — только от y. Метод решения такого вида уравнений носит название *разделения переменных*. Запишем производную y' в ее эквивалентной форме как отношение дифференциала функции к дифференциалу независимой переменной $\frac{dy}{dx}$; умножим на dx и поделим на $f_2(y)$ обе части уравнения (18.4), полагая, что $f_2(y) \neq 0$; получаем

$$\frac{dy}{f_2(y)} = f_1(x)dx. \tag{18.5}$$

В этом уравнении переменная y входит только в левую часть, а переменная x — только в правую, т.е. переменные разделены.

Пусть $y = \varphi(x)$ является решением уравнения (18.4), тогда при подстановке этого решения в уравнение (18.5) получаем тождество: два дифференциала равны друг другу, только в правой части дифференциал выражен через независимую переменную x, а в левой части — через функцию y. Поскольку

дифференциалы равны, то их неопределенные интегралы различаются на постоянную величину, т.е., интегрируя слева по переменной y, а справа по переменной x, получаем

$$\int \frac{dy}{f_2(y)} = \int f_1(x)\,dx + C, \tag{18.6}$$

где C — произвольная постоянная.

Рассмотрим примеры решения уравнений методом разделения переменных.

Пример 1. Найти частное решение уравнения $xy' - y = 0$ по начальным условиям: $y_0 = 4$ при $x_0 = -2$.

Решение. Разделим переменные, для чего перенесем y в правую часть, поделим обе части полученного уравнения на xy и умножим их на dx; получим

$$\frac{dy}{y} = \frac{dx}{x}.$$

Интегрируя обе части этого уравнения (правую по x, а левую — по y), имеем

$$\ln|y| = \ln|x| + \ln|C|,$$

где C — произвольная постоянная. При потенцировании получаем

$$|y| = |Cx|,$$

что эквивалентно уравнению $y = \pm Cx$, или $y = C_1 x$. Полученная функция представляет семейство интегральных кривых. Для выделения частного решения при указанных начальных условиях подставим в эту формулу $x = -4$ и $y = 2$, откуда получим значение для C_1: $C_1 = -1/2$. Окончательно частное решение имеет вид

$$y = -\frac{1}{2}x.$$

Пример 2. Найти частное решение уравнения $y' = x\dfrac{\sqrt{y^2+1}}{y}$, проходящее через точку $(0, 1)$.

Решение. Разделяя переменные, получаем уравнение в дифференциалах:

$$\frac{y\,dy}{\sqrt{y^2+1}} = x\,dx.$$

Интегрируя, имеем

$$\int \frac{y\,dy}{\sqrt{y^2+1}} = \int x\,dx + C,$$

где C — произвольная постоянная величина. После интегрирования (интеграл в левой части берется путем замены переменной) имеем уравнение семейства интегральных кривых

$$\sqrt{y^2+1} = \frac{1}{2}x^2 + C.$$

Выделение частного решения, проходящего через точку $(0,1)$, приводит к определению произвольной постоянной: $C = \sqrt{2}$, т.е. эта кривая описывается уравнением (с учетом выбора знака)

$$y = \sqrt{(x^2/2+\sqrt{2})^2 - 1}.$$

Неполные уравнения

Определение 2. Дифференциальное уравнение первого порядка (18.2) называется *неполным*, если функция f явно зависит только от одной переменной: либо от x, либо от y.

Различают два случая такой зависимости.

1. Пусть функция f зависит только от x. Переписав это уравнение в виде

$$dy = f(x)\,dx,$$

нетрудно убедиться, что его решением является функция

$$y = \int f(x)\,dx + C.$$

2. Пусть функция f зависит только от y, т. е. уравнение (18.2) имеет вид

$$y' = f(y). \tag{18.7}$$

Дифференциальное уравнение такого вида называется *автономным*. Такие уравнения часто употребляются при математическом моделировании и исследованиях природных и физических процессов, когда, например, независимая переменная x играет роль времени, не входящего в соотношения, описывающие законы природы. В этом случае особый интерес вызывают так называемые *точки равновесия*, или стационарные точки — нули функции $f(y)$, где производная $y' = 0$.

Решение уравнения (18.7) методом разделения переменных приводит к функциональному уравнению для определения неизвестной функции $y = \varphi(x)$ (или $x = \psi(y)$):

$$\int \frac{dy}{f(y)} = x + C. \tag{18.8}$$

В общей теории дифференциальных уравнений развита теория качественного анализа, основанная на исследовании характера стационарных точек.

Линейные дифференциальные уравнения первого порядка

Определение 3. Уравнение вида

$$y' + p(x)y = q(x), \tag{18.9}$$

где $p(x)$ и $q(x)$ — непрерывные функции, называется *линейным дифференциальным уравнением первого порядка*.

Неизвестная функция и ее производная входят в указанное уравнение в первой степени *линейно*, что и объясняет название уравнения.

Если $q(x) \equiv 0$, уравнение (18.9) называется *линейным однородным* уравнением; если же функция $q(x)$ не равна тождественно нулю, то уравнение (18.9) называется *линейным неоднородным* уравнением.

Пусть линейное однородное уравнение

$$y' + p(x)y = 0, \qquad (18.10)$$

соответствует неоднородному уравнению (18.9). Мы рассмотрим так называемый метод *вариации постоянной* — метод решения неоднородного уравнения (18.9), основанный на предварительном решении однородного уравнения (18.10). Уравнение (18.10) можно решить методом разделения переменных:

$$\frac{dy}{y} = -p(x), \text{ откуда } \ln|y| = -\int p(x)\,dx + \ln|C_1|.$$

Потенцируя, получаем общее решение уравнения (18.9):

$$y = Ce^{-\int p(x)\,dx}, \qquad (18.11)$$

где $C = \pm C_1$.

Общее решение уравнения (18.9) ищем в виде (18.11), полагая C новой неизвестной функцией от аргумента x:

$$y = C(x)e^{-\int p(x)\,dx}.$$

Подставим это решение в уравнение (18.9) с тем, чтобы найти функцию $C(x)$:

$$C'(x)e^{-\int p(x)\,dx} - C(x)p(x)e^{-\int p(x)\,dx} + \\ + p(x)C(x)e^{-\int p(x)\,dx} = q(x),$$

откуда после приведения подобных получаем уравнение для $C'(x)$:

$$C'(x) = q(x)e^{\int p(x)\,dx}. \qquad (18.12)$$

Интегрирование уравнения (18.12) даст выражение для $C(x)$:

$$C(x) = \int q(x)e^{\int p(x)\,dx}\,dx + C_1,$$

подстановка которого в формулу общего решения (18.11) приводит к окончательному виду решения неоднородного уравнения (18.9):

$$y(x) = e^{-\int p(x)\,dx}\left[C_1 + \int q(x)e^{\int p(x)\,dx}\,dx\right], \qquad (18.13)$$

где C_1 — произвольная постоянная.

Следует отметить, что некоторые нелинейные уравнения приводятся к линейным уравнениям соответствующими заменами неизвестной функции $y(x)$. К таковым относится *уравнение Бернулли*:

$$y' + p(x)y = q(x)y^n, \qquad (18.14)$$

где p и q — непрерывные функции, а n — некоторое постоянное число. При $n = 0$ имеем линейное неоднородное уравнение, а при $n = 1$ — линейное однородное уравнение

$$y' + (p - q)y = 0.$$

Пусть $n \neq 0$, $n \neq 1$. Введем новую функцию

$$z = y^{1-n}, \qquad (18.15)$$

тогда

$$z' = (1 - n)y^{-n}y'.$$

Поделим обе части уравнения (18.14) на y^n

$$y^{-n}y' + py^{1-n} = q.$$

Умножая обе части этого уравнения на $(1 - n)$, с учетом выражений для новой функции z и ее производной, получаем линейное дифференциальное неоднородное уравнение относительно неизвестной функции $z(x)$:

$$z' + (1 - n)pz = (1 - n)q. \qquad (18.16)$$

В этом уравнении, метод решения которого нам известен, функция $z(x)$ связана с искомой функцией $y(x)$ соотношением (18.15).

Рассмотрим примеры решения неоднородных уравнений первого порядка.

Пример 3. $y' + x^2 y = x^2$.

Это линейное неоднородное уравнение первого порядка. Сначала решим соответствующее ему однородное уравнение

$$y' + x^2 y = 0.$$

Разделяя переменные, получаем

$$\frac{dy}{y} = -x^2\,dx.$$

Интегрирование левой части этого уравнения по y и правой части по x дает решение однородного уравнения

$$\ln y = -\frac{x^3}{3} + \ln|C|, \text{ откуда } y = Ce^{-x^3/3}.$$

Полагая C функцией от x и подставляя найденное решение в исходное неоднородное уравнение, получаем после приведения подобных дифференциальное уравнение для $C(x)$:

$$C'(x) = x^2 e^{x^3/3}.$$

После интегрирования этого уравнения и подстановки в уже найденное решение однородного уравнения получим искомое решение исходного уравнения:

$$y(x) = C_1 e^{-x^3/3} + 1.$$

Пример 4. $xy' + y = e^x$.

Опять начинаем с однородного уравнения

$$xy' + y = 0.$$

Проводя разделение переменных, получаем после интегрирования его решение

$$y = \frac{C}{x}.$$

Полагая, что $C = C(x)$, получаем после подстановки в неоднородное уравнение:

$$C'(x) = e^x, \text{ откуда } C(x) = e^x + C_1.$$

Стало быть, общее решение исходного уравнения имеет вид

$$y(x) = \frac{1}{x}(e^x + C_1).$$

Пример 5. $y' + xy = xy^3$.

Данное нелинейное уравнение представляет собой уравнение Бернулли при $n = 3$. Заменой искомой функции $z = y^{-2}$, согласно (18.15), мы получим линейное неоднородное уравнение относительно $z(x)$:

$$\frac{dz}{dx} - 2xz = -2x.$$

По формуле (18.13) получаем общее решение этого уравнения

$$z(x) = C_1 e^{x^2} + 1.$$

Теперь, выполняя обратную замену $y = \pm 1/\sqrt{z}$, получаем решение исходного нелинейного уравнения:

$$y = \pm(C_1 e^{x^2} + 1)^{-1/2}.$$

Глава 19
ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ ВЫСШИХ ПОРЯДКОВ

19.1. Дифференциальные уравнения второго порядка

Основные понятия

Определение 1. *Дифференциальным уравнением второго порядка* называется уравнение вида

$$F(x, y, y', y'') = 0, \tag{19.1}$$

где x — независимая переменная, y — искомая функция, y' и y'' — соответственно ее первая и вторая производные.

Примеры дифференциальных уравнений второго порядка:

$$y'' + yy' - xy^3 - \cos x = 0; \quad y^2y'' + xy' + x^2 \cos y = 0.$$

Будем рассматривать уравнения, которые можно записать в виде, разрешенном относительно второй производной:

$$y'' = f(x, y, y') \tag{19.2}$$

Как и в случае уравнения первого порядка, решением уравнения (19.1) называется функция $y = \varphi(x)$, определенная на некотором интервале (a, b), которая обращает это уравнение в тождество. График решения называется *интегральной кривой*. Имеет место теорема существования и единственности решения уравнения второго порядка.

Теорема 19.1 (ТЕОРЕМА КОШИ). *Пусть функция $f(x, y, y')$ и ее частные производные f'_y и $f'_{y'}$ непрерывны в некоторой области D пространства переменных (x, y, y'). Тогда для любой внутренней точки $M_0(x_0, y_0, y'_0)$ этой области существует единственное решение уравнения* (19.2), *удовлетворяющее условиям:*

$$x = x_0; \quad y = y_0; \quad y' = y'_0. \tag{19.3}$$

Геометрический смысл этой теоремы (ее доказательство мы не приводим) заключается в том, что через заданную точку (x_0, y_0) на координатной плоскости Oxy проходит единственная интегральная кривая с заданным угловым коэффициентом y'_0 касательной (рис. 19.1).

Условия (19.3) называются *начальными условиями*, а задачу отыскания решения уравнения (19.2) по заданным начальным условиям называют *задачей Коши*.

Общим решением уравнения (19.2) в некоторой области D называется функция $y = \varphi(x, C_1, C_2)$, если она является решением этого уравнения при любых постоянных величинах C_1 и C_2, которые могут быть определены единственным образом при заданных начальных условиях (19.3). *Частным решением* уравнения (19.2) называется общее решение этого уравнения при фиксированных значениях постоянных C_1 и C_2: $y = \varphi(x, C_1^0, C_2^0)$.

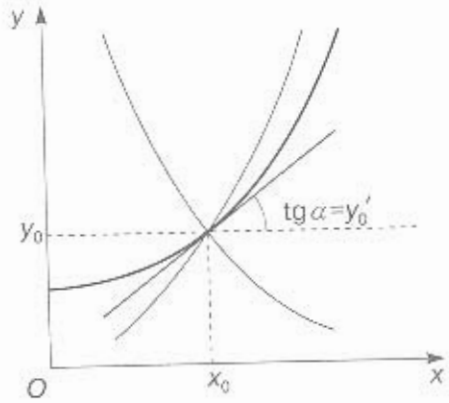

Рис. 19.1

Рассмотрим для пояснения уравнение $y'' = 0$. Его общее решение получается при двукратном интегрировании уравнения

$$y = C_1x + C_2,$$

где C_1 и C_2 — произвольные постоянные. Это решение представляет собой семейство прямых в произвольных направлениях, причем через каждую точку плоскости Oxy проходит бесконечное число таких прямых. Поэтому для выделения частного решения, проходящего через заданную точку (x_0, y_0), следует задать еще и угловой коэффициент прямой, совпадающей в данном случае со своей касательной. Например, найдем частное решение, удовлетворяющее начальным условиям

$$y\Big|_{x=1} = 2, \quad y'\Big|_{x=1} = 1,$$

т.е. нужно найти прямую, проходящую через точку $(1, 2)$, с угловым коэффициентом, равным единице. Подстановка начальных условий в общее решение уравнения приводит к системе двух линейных уравнений относительно постоянных C_1 и C_2:

$$\begin{cases} C_1 + C_2 = 2, \\ C_1 = 1, \end{cases}$$

откуда $C_1 = 1$, $C_2 = 1$. Таким образом, искомое частное решение это прямая $y = x + 1$.

Уравнения, допускающие понижение порядка

Существует три вида уравнения $y'' = f(x, y, y')$, которые путем замены переменной (искомой функции) сводятся к уравнениям первого порядка.

1. Уравнение вида

$$y'' = f(x). \tag{19.4}$$

Введем новую функцию $z(x)$ путем замены $z(x) = y'$. Тогда исходное уравнение второго порядка преобразуется в неполное уравнение первого порядка: $z' = f(x)$, решением которого является функция $z(x) = \int f(x)\,dx + C_1$. Поскольку $z(x) = y'$, то повторным интегрированием находим общее решение уравнения (19.4):

$$y(x) = \int \left[\int f(x)\,dx\right] dx + C_1 x + C_2, \tag{19.5}$$

где C_1 и C_2 — произвольные постоянные.

2. Уравнение вида

$$y'' = f(x, y'), \tag{19.6}$$

т.е. уравнение не содержит в явном виде y. Как и в предыдущем случае, положим $z(x) = y'$. Тогда получаем уравнение первого порядка общего вида $z' = f(x, z)$. Найдя общее решение этого уравнения $z = \varphi(x, C_1)$, повторным интегрированием получим искомое общее решение уравнения (19.6):

$$y(x) = \int \varphi(x, C_1)\,dx + C_2, \tag{19.7}$$

где C_1 и C_2 — произвольные постоянные.

3. Уравнение вида

$$y'' = f(y, y'), \tag{19.8}$$

т.е. уравнение не содержит независимой переменной x. Введем новую функцию, зависящую от y, полагая $z(y) = y'$. Тогда, поскольку по правилу дифференцирования сложной функции

$$y'' = \frac{d}{dx}[z(y)] = \frac{dz}{dy}y' = \frac{dz}{dy}z,$$

уравнение (19.8) преобразуется в дифференциальное уравнение первого порядка относительно функции $z(y)$:

$$z\frac{dz}{dy} = f(y, z).$$

Пусть общее решение этого уравнения $z = \varphi(y, C_1)$. Тогда обратной заменой получаем неполное уравнение первого порядка относительно функции $y(x)$

$$\frac{dy}{dx} = \varphi(y, C_1),$$

из которого методом разделения переменных получаем функциональное соотношение для определения общего решения уравнения (19.8):

$$\int \frac{dy}{\varphi(y, C_1)} = x + C_2, \tag{19.9}$$

где C_1 и C_2 — произвольные постоянные.

Рассмотрим два примера решения дифференциальных уравнений второго порядка.

Пример 1. Найти решение уравнения $y'' + x(y')^2 = 0$.

Это уравнение вида (19.6), поскольку оно не содержит в явном виде y. Заменой $z(x) = y'$ приведем его к уравнению первого порядка $\dfrac{dz}{dx} = -xz^2$, откуда $z = \dfrac{2}{x^2 + C_1}$, или $y' = \dfrac{2}{x^2 + C_1}$. Интегрируя это уравнение, получаем общее решение исходного уравнения:

$$y(x) = \int \frac{2dx}{x^2 + C_1} + C_2,$$

где C_1 и C_2 — произвольные постоянные. В зависимости от выбора знака C_1 интеграл в правой части этого равенства (обозначим его через I) может иметь разные выражения

$$\begin{cases} \left[\ln\left|(x - \sqrt{-C_1})/(x + \sqrt{-C_1})\right|\right]/\sqrt{-C_1}, & C_1 < 0, \\ -2/x, & C_1 = 0, \\ 2/\sqrt{C_1}\,\operatorname{arctg}(x/\sqrt{C_1}), & C_1 > 0. \end{cases}$$

Пример 2. Найти решение уравнения $y'' - (y')^2 = 0$.

Это уравнение вида (19.8), т.е. оно не содержит явно независимой переменной x. Заменой $z(y) = y'$ сведем его к уравнению первого порядка

$$z\frac{dz}{dy} - z^2 = 0.$$

Первое решение этого уравнения: $z = 0$, или $y = C$, где C — постоянная величина. Сокращая после этого обе части уравнения на z, получаем $\dfrac{dz}{dy} - z = 0$. Решение этого уравнения методом разделения переменных y и z дает $z = C_1 e^y$. Наконец, обратная замена приводит к уравнению первого порядка

$$\frac{dy}{dx} = C_1 e^y.$$

Разделение переменных x и y приводит к общему решению исходного уравнения: $e^{-y}\,dy = C_1\,dx$, откуда $e^{-y} = C_1 x + C_2$, или окончательно

$$y(x) = -\ln(C_1 x + C_2),$$

где C_1 и C_2 — произвольные постоянные. Нетрудно видеть, что это решение включает в себя и решение $y = C$, полученное выше (при $C_1 = 0$, $C_2 \neq 0$).

Линейные дифференциальные уравнения второго порядка

Определение 2. *Линейным дифференциальным уравнением второго порядка* называется уравнение вида

$$y'' + p(x)y' + q(x)y = f(x), \tag{19.10}$$

где y — искомая функция, а $p(x)$, $q(x)$ и $f(x)$ — функции, непрерывные на некотором интервале (a, b).

Если $f(x) \equiv 0$, уравнение (19.10) называется *линейным однородным уравнением*, в противном случае оно называется *линейным неоднородным уравнением*. Если разрешить уравнение (19.10) относительно второй производной, легко увидеть,

что оно является частным случаем уравнения (19.2) и удовлетворяет условиям теоремы Коши. Поэтому для любых начальных условий (19.3) при $x_0 \in (a, b)$ это уравнение имеет единственное решение задачи Коши.

Линейные однородные уравнения второго порядка

Рассмотрим свойства решений линейных однородных дифференциальных уравнений второго порядка

$$y'' + p(x)y' + q(x)y = 0. \tag{19.11}$$

Теорема 19.2. *Пусть функции $y_1(x)$ и $y_2(x)$ — решения уравнения (19.11). Тогда функция $y = C_1 y_1(x) + C_2 y_2(x)$ также является решением этого уравнения при любых постоянных C_1 и C_2.*

Доказательство. Подставив указанную в условии теоремы функцию в уравнение (19.11) и собрав члены при C_1 и C_2, получим тождества, равные нулю, поскольку по условию теоремы $y_1(x)$ и $y_2(x)$ — решения этого уравнения. Теорема доказана.

В параграфе 11.2 было введено понятие зависимости функций. Напомним, что функции $y_1(x)$ и $y_2(x)$ называются *линейно зависимыми* на (a, b), если можно найти такие числа λ_1 и λ_2, не равные нулю одновременно, что для любого $x \in (a, b)$ выполняется равенство

$$\lambda_1 y_1(x) + \lambda_2 y_2(x) = 0. \tag{19.12}$$

В этом случае функции $y_1(x)$ и $y_2(x)$ очевидным образом пропорциональны, например $y_1/y_2 = -\lambda_2/\lambda_1 = k$ при $\lambda_1 \neq 0$ и $y_2(x) \neq 0$. Обратное также верно: если две функции пропорциональны на (a, b), то они линейно зависимы на этом интервале.

Если нельзя найти двух одновременно ненулевых чисел λ_1 и λ_2, чтобы выполнялось равенство (19.12), то функции $y_1(x)$ и $y_2(x)$ называются *линейно независимыми* на (a, b). В таком случае эти функции уже не будут пропорциональными.

Введем *определитель Вронского*[1], который для случая двух функций имеет вид

$$W(x) = \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} = y_1 y_2' - y_2 y_1'. \tag{19.13}$$

[1] Юзеф Вронский (1776–1853) — польский математик.

Теорема 19.3. *Если функции $y_1(x)$ и $y_2(x)$ линейно зависимы на интервале (a,b), то определитель Вронского, составленный из них, равен нулю на этом интервале; если же функции линейно независимы на (a,b), то определитель Вронского отличен от нуля на (a,b).*

Доказательство. Пусть y_1 и y_2 линейно зависимы на интервале (a,b). Тогда, как уже отмечалось, эти функции пропорциональны на (a,b), т. е. $y_1 = ky_2$, а значит, и $y_1' = ky_2'$. Отсюда в определителе $W(x)$ будут пропорциональные столбцы, что и означает его равенство нулю на интервале (a,b).

Доказательство второй части теоремы проводится от противного. Пусть y_1 и y_2 линейно независимы на (a,b); предположим, что определитель $W(x)$ равен нулю на этом интервале. Тогда его столбцы необходимо пропорциональны, т. е. пропорциональны и эти функции, из чего следует их линейная зависимость на (a,b), что противоречит условию теоремы. Теорема доказана.

Оказывается, при решении уравнения второго порядка важно найти два линейно независимых решения, о чем говорится в теореме 19.4.

Теорема 19.4. *Пусть решения уравнения* (19.11) $y_1(x)$ *и* $y_2(x)$ *линейно независимы на* (a,b). *Тогда функция*

$$y(x) = C_1y_1(x) + C_2y_2(x), \qquad (19.14)$$

где C_1 *и* C_2 *— произвольные постоянные, является общим решением однородного уравнения* (19.11).

Доказательство. В силу теоремы 19.2 функция (19.14) является решением уравнения (19.11); нужно показать, что она представляет собой общее решение, т. е. что из этой функции можно выделить частное решение, удовлетворяющее любым начальным условиям (19.3).

Возьмем любые числа $x_0 \in (a,b)$, y_0 и y_0' и составим из них начальные условия для решения (19.14):

$$y\big|_{x=x_0} = y_0, \qquad y'\big|_{x=x_0} = y_0'. \qquad (19.15)$$

Подставляя в левые части этих условий функцию (19.14), получим систему двух линейных уравнений относительно неизвестных чисел C_1 и C_2:

$$\begin{cases} C_1y_1(x_0) + C_2y_2(x_0) = y_0, \\ C_1y_1'(x_0) + C_2y_2'(x_0) = y_0'. \end{cases}$$

Определителем этой системы является определитель Вронского $W(x_0)$ и в силу линейной независимости y_1 и y_2 он не равен нулю, т.е. система имеет единственное решение $C_1 = C_1^0$, $C_2 = C_2^0$ при любых правых частях y_0 и y_0'. Подставляя эти значения в решение (19.14), получим частное решение, удовлетворяющее произвольно выбранным начальным условиям (19.15). Таким образом, решение (19.14) является общим для уравнения (19.11), что и требовалось доказать.

Рассмотрим примеры: установить, будут ли указанные функции общими решениями уравнений (C_1 и C_2 — произвольные постоянные).

Пример 3. $y = C_1\sin x + C_2\cos x$, $y'' + y = 0$.

По теореме 19.4 указанное решение будет общим для данного уравнения, если функции $\sin x$ и $\cos x$ являются решениями этого уравнения и линейно независимы. Поскольку $(\sin x)'' = -\sin x$ и $(\cos x)'' = -\cos x$, то обе функции являются решениями данного уравнения. Для установления их линейной независимости вычислим соответствующий определитель Вронского по формуле (19.13):

$$W(x) = \begin{vmatrix} \sin x & \cos x \\ \cos x & -\sin x \end{vmatrix} = -\sin^2 x - \cos^2 x = -1 \neq 0,$$

т. е. функция $C_1\sin x + C_2\cos x$ является общим решением уравнения $y'' + y = 0$.

Пример 4. $y = C_1e^{2x} + C_2e^{3x}$, $y'' - 5y' + 6y = 0$.

Как и в предыдущем примере, сначала проверяем подстановкой в данное уравнение, являются ли функции e^{2x} и e^{3x} его решениями:

$$4e^{2x} - 10e^{2x} + 6e^{2x} \equiv 0, \quad 9e^{3x} - 15e^{3x} + 6e^{3x} \equiv 0.$$

Затем вычисляем определитель $W(x)$, составленный из этих функций:

$$W(x) = \begin{vmatrix} e^{2x} & e^{3x} \\ 2e^{2x} & 3e^{3x} \end{vmatrix} = 3e^{5x} - 2e^{5x} = e^{5x} \neq 0.$$

32 Зак. 6853

Указанная функция является общим решением данного уравнения.

Пример 5. $y = C_1 x^{3/2} + C_2$, $2xy'' = y'$, $0 < x < 1$.

Предлагаем читателю убедиться, что функции $y_1 = x^{3/2}$ и $y_2 = 1$ удовлетворяют данному уравнению и являются линейно независимыми, иными словами, что указанная функция является общим решением этого уравнения второго порядка.

Линейные неоднородные уравнения второго порядка

Рассмотрим основные свойства решения неоднородного дифференциального уравнения

$$y'' + p(x)y' + q(x)y = f(x). \tag{19.16}$$

Теорема 19.5. *Общее решение уравнения* (19.16) *состоит из суммы его частного решения и общего решения соответствующего однородного уравнения* (19.11).

Доказательство. Пусть $\tilde{y}(x) = C_1 y_1(x) + C_2 y_2(x)$ — общее решение соответствующего однородного уравнения, где C_1 и C_2 — произвольные постоянные, а $Y(x)$ — частное решение неоднородного уравнения (19.16). Сначала покажем, что функция

$$y(x) = Y(x) + \tilde{y}(x) \tag{19.17}$$

— это решение уравнения (19.16). Действительно, подставляя в это уравнение производные функций $Y(x)$ и $y(x)$ и группируя слагаемые, составляющие соответственно однородное и неоднородное уравнения, без особого труда получим требуемое в силу свойств этих функций как решений указанных уравнений.

Теперь нужно показать, что функция (19.17) является общим решением неоднородного уравнения (19.16). Рассмотрим разность $y(x) - Y(x)$, где $y(x)$ — любое решение (19.17) уравнения (19.16). Эта разность является решением однородного уравнения (19.11), что несложно проверить прямой подстановкой в него и группированием соответствующих слагаемых:

$$[y - Y]'' + p(x)[y - Y]' + q(x)[y - Y] =$$

$$= [y'' + p(x)y' + q(x)y] - [Y'' + p(x)Y' + q(x)Y] =$$
$$= f(x) - f(x) \equiv 0.$$

Стало быть, эту разность можно записать в виде частного решения однородного уравнения (19.11):

$$y - Y(x) = C_1^0 y_1(x) + C_2^0 y_2(x),$$

где C_1^0 и C_2^0 — определенные значения постоянных C_1 и C_2 (в формуле общего решения однородного уравнения). Следовательно, любое решение $y(x)$ уравнения (19.16) можно получить по формуле (19.17) путем соответствующего подбора произвольных постоянных C_1 и C_2, а это и означает, что функция (19.17) является общим решением неоднородного уравнения (19.16). Теорема доказана.

Эта теорема указывает способ нахождения общего решения неоднородного уравнения (19.16): нужно найти общее решение соответствующего однородного уравнения (19.11) и какое-либо частное решение уравнения (19.16). Задача нахождения частного решения неоднородного уравнения в общем случае довольно сложна; однако существует метод *вариации постоянных*, основанный на использовании известного решения однородного уравнения.

Пусть $\tilde{y} = C_1 y_1(x) + C_2 y_2(x)$ — общее решение однородного уравнения (19.11). Будем полагать, что частное решение неоднородного уравнения имеет аналогичный вид, но произвольные постоянные C_1 и C_2 являются функциями переменной x:

$$Y = C_1(x)y_1(x) + C_2(x)y_2(x). \tag{19.18}$$

Дифференцируя это равенство, получаем

$$Y' = C_1'(x)y_1(x) + C_1(x)y_1'(x) + C_2'(x)y_2(x) + C_2(x)y_2'(x).$$

Положим функции $C_1(x)$ и $C_2(x)$ такими, чтобы выполнялось равенство

$$C_1'(x)y_1(x) + C_2'(x)y_2(x) = 0, \tag{19.19}$$

тогда

$$Y' = C_1(x)y_1'(x) + C_2(x)y_2'(x).$$

Дифференцируя это равенство, найдем выражение для Y'':

$$Y'' = C_1'(x)y_1'(x) + C_1(x)y_1''(x) + C_2'(x)y_2'(x) + C_2(x)y_2''(x).$$

Подставляя выражения для Y, Y' и Y'' в уравнение (19.16) и группируя слагаемые при $C_1(x)$ и $C_2(x)$, получаем

$$\begin{gathered} C_1(x)[y_1''(x) + p(x)y_1'(x) + q(x)y_1(x)] + \\ + C_2(x)[y_2''(x) + p(x)y_2'(x) + q(x)y_2(x)] + \\ + C_1'(x)y_1'(x) + C_2'(x)y_2'(x) = f(x). \end{gathered}$$

Так как $y_1(x)$ и $y_2(x)$ — решения однородного уравнения, то выражения в квадратных скобках равны нулю. Равенство упрощается:

$$C_1'(x)y_1'(x) + C_2'(x)y_2'(x) = f(x).$$

Объединяя его с равенством (19.19), получаем систему линейных уравнений относительно $C_1'(x)$ и $C_2'(x)$:

$$\begin{cases} C_1'(x)y_1(x) + C_2'(x)y_2(x) = 0, \\ C_1'(x)y_1'(x) + C_2'(x)y_2'(x) = f(x). \end{cases} \tag{19.20}$$

Определитель этой системы представляет собой определитель Вронского (19.13); в силу теоремы 19.3 он не равен нулю, поскольку составлен из линейно независимых функций $y_1(x)$ и $y_2(x)$. Значит, система (19.20) имеет единственное решение. Пусть

$$C_1'(x) = \varphi_1(x), \qquad C_2'(x) = \varphi_2(x)$$

— решение системы уравнений (19.20). Интегрируя эти выражения, получим $C_1(x)$ и $C_2(x)$; подставив их в (19.18), найдем частное решение неоднородного уравнения $Y(x)$.

Пример 6. Найти общее решение неоднородного уравнения

$$y'' - 5y' + 6y = x.$$

Решение. Как мы знаем из примера 4, общее решение соответствующего однородного уравнения $y'' - 5y' + 6y = 0$ имеет

вид $\tilde{y} = C_1e^{2x} + C_2e^{3x}$, где C_1 и C_2 — произвольные постоянные. Следовательно, частное решение неоднородного уравнения, согласно методу вариации постоянных, будем искать в виде

$$Y = C_1(x)e^{2x} + C_2(x)e^{3x},$$

где производные функций $C_1(x)$ и $C_2(x)$ определяются из системы вида (19.20):

$$\begin{cases} C_1'(x)e^{2x} + C_2'(x)e^{3x} = 0, \\ 2C_1'(x)e^{2x} + 3C_2'(x)e^{3x} = x. \end{cases}$$

Умножая первое уравнение этой системы на 2 и вычитая его из второго, получаем $C_2'(x) = xe^{-3x}$. Тогда из первого уравнения имеем $C_1'(x) = -C_2'(x)e^x = -xe^{-2x}$. Интегрируя эти выражения, получаем

$$C_1(x) = -\int xe^{-2x}\,dx = \frac{1}{2}e^{-2x}\left(x + \frac{1}{2}\right),$$

$$C_2(x) = \int xe^{-3x}\,dx = -\frac{1}{3}e^{-3x}\left(x + \frac{1}{3}\right).$$

Здесь при интегрировании мы опустили произвольные константы неопределенного интеграла, так как нам нужно найти любое частное решение неоднородного уравнения. Согласно (19.18), оно имеет вид

$$Y = \frac{1}{2}\left(x + \frac{1}{2}\right) - \frac{1}{3}\left(x + \frac{1}{3}\right) = \frac{1}{6}\left(x + \frac{5}{6}\right).$$

Искомое общее решение заданного неоднородного дифференциального уравнения, в силу теоремы 19.5, выражается формулой

$$y = \frac{1}{6}\left(x + \frac{5}{6}\right) + C_1e^{2x} + C_2e^{3x},$$

где C_1 и C_2 — произвольные постоянные.

Линейные дифференциальные уравнения второго порядка с постоянными коэффициентами

Рассмотрим важный и весьма распространенный случай, когда в уравнении вида (19.16) функции $p(x)$ и $q(x)$ — постоянные величины. Уравнения такого вида называются *линейными уравнениями с постоянными коэффициентами*. Как и в общем случае линейных дифференциальных уравнений второго порядка, основой в построении решения являются однородные уравнения.

1. Рассмотрим линейное однородное уравнение

$$y'' + py' + qy = 0, \tag{19.21}$$

где p и q — вещественные числа. Будем искать решение этого уравнения в виде $y = e^{kx}$, где k — некоторое число. Подставляем эту функцию в уравнение (19.21):

$$k^2e^{kx} + pke^{kx} + qe^{kx} = 0.$$

Сокращая обе части этого равенства на e^{kx}, получаем квадратное уравнение:

$$k^2 + pk + q = 0. \tag{19.22}$$

Стало быть, если число k является корнем уравнения (19.22), то функция $y = e^{kx}$ есть решение однородного уравнения (19.21). Уравнение (19.22) называется *характеристическим уравнением* для уравнения (19.21).

Вид решения уравнения (19.21) существенно зависит от того, какие корни имеет характеристическое уравнение (19.22). Обозначим эти корни через k_1 и k_2. Справедлива следующая теорема.

Теорема 19.6. *Если корни характеристического уравнения вещественные и $k_1 \neq k_2$ общее решение однородного уравнения* (19.21) *имеет вид*

$$y = C_1e^{k_1x} + C_2e^{k_2x}. \tag{19.23}$$

Если корни уравнения (19.22) *вещественные и равные* ($k_1 = k_2 = k$), *общее решение уравнения* (19.21) *имеет вид*

$$y = C_1e^{kx} + C_2xe^{kx}. \tag{19.24}$$

Если корни характеристического уравнения комплексные ($k_1 = a + bi$, $k_2 = a - bi$, *где* $i = \sqrt{-1}$, a *и* b — *вещественные числа*), *общее решение уравнения* (19.21) *имеет вид*

$$y = e^{ax}(C_1\cos bx + C_2\sin bx), \tag{19.25}$$

где $a = -p/2$, $b = \sqrt{q - p^2/4}$.

Во всех трех случаях C_1 *и* C_2 — *произвольные постоянные.*

Доказательство. Достаточно убедиться в том, что каждое из слагаемых в правых частях равенств (19.23)–(19.25) является решением уравнения (19.21), а затем по определителю Вронского удостовериться в том, что каждая пара функций в этих равенствах линейно независима.

Заметим, что в последнем случае корни характеристического многочлена с действительными коэффициентами представляют собой комплексно-сопряженные числа в алгебраической форме.

Рассмотрим примеры нахождения общих решений однородных уравнений.

Пример 7. $y'' - 5y' + 4y = 0$.

Характеристическое уравнение данного дифференциального уравнения имеет вид

$$k^2 - 5k + 4 = 0.$$

Его корни вещественные и различны: $k_1 = 1$, $k_2 = 4$. Следовательно, общее решение данного уравнения имеет вид

$$y = C_1e^x + C_2e^{4x}.$$

Пример 8. $y'' - 6y' + 9 = 0$.

Составим характеристическое уравнение:

$$k^2 - 6k + 9, \text{ или } (k-3)^2 = 0.$$

Оно имеет кратный корень $k = 3$; следовательно, общее решение данного однородного уравнения имеет вид

$$y = e^{3x}(C_1 + C_2x).$$

Пример 9. $y'' - 2y' + 2y = 0$.

Соответствующее характеристическое уравнение

$$k^2 - 2k + 2 = 0$$

имеет комплексно-сопряженные корни: $k_1 = 1 + i$, $k_2 = 1 - i$, где $i = \sqrt{-1}$ — *мнимая единица*. Следовательно, общее решение данного уравнения таково:

$$y = e^x(C_1 \sin x + C_2 \cos x).$$

2. Что касается решения неоднородных дифференциальных уравнений с постоянными коэффициентами, то оно полностью основывается на методе вариации постоянных. Однако в ряде случаев бывает проще "угадать" или подобрать частное решение неоднородного уравнения по виду его правой части.

Рассмотрим линейное неоднородное уравнение второго порядка

$$y'' + py' + qy = f(x).$$

Как известно, общее решение этого уравнения представляет собой сумму общего решения однородного уравнения и частного решения неоднородного уравнения. Укажем некоторые случаи нахождения частного решения неоднородного уравнения методом неопределенных коэффициентов, не требующим интегрирования. Эти случаи типизируются по виду правой части уравнения $f(x)$.

1) Правая часть имеет вид многочлена степени n:

$$f(x) = P_n(x).$$

Тогда частное решение ищется в виде:

$$Y = Q_n(x)x^s,$$

где $Q_n(x)$ — многочлен степени n с неопределенными коэффициентами, а s — кратность корня характеристического уравнения (19.22), равного нулю. Неопределенные коэффициенты многочлена $Q_n(x)$ находятся из системы линейных алгебраических уравнений, получаемой приравниванием коэффициентов при одинаковых степенях x в левой и правой частях уравнения (после подстановки в левую часть выражения для Y).

2) Правая часть имеет вид

$$f(x) = e^{\alpha x} P_n(x),$$

где $P_n(x)$ — многочлен степени n. В этом случае частное решение ищется в виде:

$$Y = Q_n(x)x^s e^{\alpha x},$$

где $Q_n(x)$ — многочлен степени n с неопределенными коэффициентами, s — кратность корня характеристического уравнения (19.22), равного α (при $\alpha = 0$ имеем случай 1).

3) Правая часть имеет вид

$$f(x) = a \sin \beta x + b \cos \beta x.$$

В этом случае частное решение ищется в виде:

$$Y = (A \sin \beta x + B \cos \beta x)x^s,$$

где A и B — неизвестные коэффициенты, s — кратность корня характеристического уравнения (19.22), равного $i\beta$.

4) Правая часть задана в виде

$$f(x) = e^{\alpha x}[P_n(x) \sin \beta x + Q_m(x) \cos \beta x],$$

где $P_n(x)$ и $Q_m(x)$ — соответственно многочлены степени n и m. Частное решение ищется в виде:

$$Y = x^s e^{\alpha x}[Q_k^{(1)}(x) \sin \beta x + Q_k^{(2)}(x) \cos \beta x],$$

где $Q_k^{(1)}(x)$ и $Q_k^{(2)}(x)$ — многочлены степени $k = \max(n, m)$, а s — кратность корня характеристического уравнения (19.22), равного $\alpha + i\beta$.

В заключение приведем теорему, которую можно применять при решении линейных неоднородных уравнений, правая часть которых представляет собой сумму нескольких слагаемых разных типов. *Если Y_1 — частное решение уравнения*

$$y'' + py' + qy = f_1(x),$$

а Y_2 — частное решение уравнения

$$y'' + py' + qy = f_2(x),$$

то сумма $Y_1 + Y_2$ является решением уравнения

$$y'' + py' + qy = f_1(x) + f_2(x).$$

Доказательство этой теоремы очевидно.

Краевая задача для дифференциального уравнения второго порядка

Как уже было сказано, в силу основной теоремы существования и единственности решения для уравнения второго порядка

$$y'' = f(x, y, y') \tag{19.26}$$

определена задача Коши: в точке $x = x_0$ заданы значения неизвестной функции и ее производной:

$$x = x_0, \qquad y = y_0, \qquad y' = y'_0. \tag{19.27}$$

Если выполняются условия теоремы 19.1, то задача Коши (19.26), (19.27) однозначно определяет частное решение.

Однако существует и другой тип задач для дифференциальных уравнений второго порядка с использованием задания неизвестной функции в двух разных точках. Иными словами, при решении уравнения (19.26) на интервале (a, b) рассмотрим *граничные условия* наиболее простого вида на концах интервала

$$x = x_1, \quad y = y_1; \quad x = x_2, \quad y = y_2. \tag{19.28}$$

В этом случае уравнение (19.26) совместно с условиями (19.28) называется *первой краевой задачей для уравнения второго порядка*. Поскольку второе условие в (19.28) равносильно заданию второго условия в (19.27), то указанная краевая задача может иметь единственное решение, т.е. определять единственным образом частное решение дифференциального уравнения (19.26), проходящее через точки (x_1, y_1), (x_2, y_2). Так, для линейного дифференциального уравнения второго порядка первая краевая задача имеет решение, если определитель системы линейных алгебраических уравнений относительно произвольных постоянных C_1 и C_2

$$\begin{cases} C_1\tilde{y}_1(x_1) + C_2\tilde{y}_2(x_1) + Y(x_1) = y_1, \\ C_1\tilde{y}_1(x_2) + C_2\tilde{y}_2(x_2) + Y(x_2) = y_2, \end{cases}$$

реализующей краевые условия (19.28), отличен от нуля. Здесь, в соответствии с (19.18), $Y(x)$ — частное решение неоднородного уравнения, $\tilde{y}_1(x)$ и $\tilde{y}_2(x)$ — линейно независимые решения соответствующего однородного уравнения. В таком случае краевая задача с условиями (19.28) однозначно определяет частное решение дифференциального уравнения (19.16).

Пример 10. Найти частное решение уравнения

$$y'' - 5y' + 6y = x,$$

удовлетворяющее краевым условиям

$$x = 0, \quad y = 1; \quad x = \ln 2, \quad y = 2.$$

Общее решение этого уравнения было найдено в примере 6. Для выделения частного решения подставим это решение в краевые условия. Получаем систему линейных уравнений относительно произвольных постоянных C_1 и C_2:

$$\begin{cases} C_1 + C_2 = \dfrac{31}{36}, \\ 4C_1 + 8C_2 = \dfrac{67 - 6\ln 2}{36}. \end{cases}$$

Нетрудно видеть, что определитель этой системы не равен нулю, т. е. данная краевая задача имеет решение. Вычитая из второго уравнения первое, умноженное на 4, получаем C_2, а затем из первого уравнения получим C_1:

$$C_2 = -(57 + 6\ln 2)/144, \quad C_1 = (181 + 6\ln 2)/144.$$

Отсюда решение данной краевой задачи как частное решение дифференциального уравнения, проходящее через точки $(0, 1)$ и $(\ln 2, 2)$, имеет вид

$$y = \frac{1}{6}\left(x + \frac{5}{6}\right) + \frac{181 + 6\ln 2}{144}e^{2x} - \frac{57 - 6\ln 2}{144}e^{3x}.$$

19.2. Дифференциальные уравнения высших порядков

Дифференциальное уравнение высшего (n-го) порядка имеет вид

$$F(x, y, y', \ldots, y^{(n)}) = 0, \tag{19.29}$$

где x — независимая переменная; $y, y', \ldots, y^{(n)}$ — соответственно искомая функция и ее производные вплоть до n-го порядка. В случае когда уравнение (19.29) разрешимо относительно старшей производной, оно имеет вид

$$y^{(n)} = f\left(x, y, y', \ldots, y^{(n-1)}\right), \tag{19.30}$$

где f — некоторая функция от $n+1$ переменной.

Решением дифференциального уравнения (19.29) или (19.30) называется такая функция $y = y(x)$, которая при подстановке в это уравнение обращает его в тождество. График решения дифференциального уравнения называется *интегральной кривой*.

Далее мы рассмотрим условия существования и единственности решения для обыкновенного дифференциального уравнения n-го порядка вида (19.30).

Система дифференциальных уравнений первого порядка

Для удобства вместо одного дифференциального уравнения n-го порядка вводят в рассмотрение систему, состоящую из n дифференциальных уравнений первого порядка. Наряду с искомой функцией y вводят еще $n-1$ вспомогательных искомых функций $y_1, y_2, \ldots, y_{n-1}$:

$$\frac{dy}{dx} = y_1, \frac{dy_1}{dx} = y_2, \ldots, \frac{dy_{n-2}}{dx} = y_{n-1}. \tag{19.31}$$

Из этих соотношений следует, что

$$y_k = \frac{d^k y}{dx^k} = y^{(k)}, \qquad k = 1, 2, \ldots, n-1.$$

Поэтому $y^{(n)} = dy_{n-1}/dx$, и тогда уравнение (19.30) принимает вид

$$\frac{dy_{n-1}}{dx} = f(x, y, y_1, \ldots, y_{n-1}). \tag{19.32}$$

Уравнения (19.31) и (19.32) представляют систему n дифференциальных уравнений с n искомыми функциями $y, y_1, \ldots,$

y_{n-1}. Система такого вида, где в левых частях стоят производные от искомых функций, а правые части суть функции от независимой переменной и искомых функций, называется *системой n дифференциальных уравнений первого порядка нормальной формы*.

В целях обобщения вводят в рассмотрение вместо системы специального вида, представленной выше, систему n дифференциальных уравнений первого порядка общего вида относительно n искомых функций $y, y_1, \ldots, y_n$:

$$\begin{cases} \dfrac{dy_1}{dx} = f_1(x, y_1, y_2, \ldots, y_n), \\ \dfrac{dy_2}{dx} = f_2(x, y_1, y_2, \ldots, y_n), \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \dfrac{dy_n}{dx} = f_n(x, y_1, y_2, \ldots, y_n). \end{cases} \tag{19.33}$$

Для этой системы справедлива следующая теорема, которая приведена без доказательства.

Теорема 19.7 (ТЕОРЕМА КОШИ). *Пусть для системы дифференциальных уравнений первого порядка* (19.33) *выполняются следующие условия:*

а) *функции f_i $(i = 1, 2, \ldots, n)$ непрерывны по всем аргументам в замкнутой области D:*

$$x_0 - a \leqslant x \leqslant x_0 + a, \quad y_i^{(0)} - b_i \leqslant y_i \leqslant y_i^{(0)} + b_i \quad (i = 1, 2, \ldots, n);$$

б) *частные производные $\partial f_i/\partial y_k$ непрерывны в области D. Тогда существует одна и только одна система решений уравнений* (19.33):

$$y_1 = y_1(x), \ y_2 = y_2(x), \ \ldots, \ y_n = y_n(x), \tag{19.34}$$

определенная в некоторой окрестности точки x_0 и удовлетворяющая при $x = x_0$ заданным начальным условиям:

$$y_1(x_0) = y_1^{(0)}, \ y_2(x_0) = y_2^{(0)}, \ \ldots, \ y_n(x_0) = y_n^{(0)}. \tag{19.35}$$

Теорема Коши утверждает существование частного решения системы обыкновенных дифференциальных уравнений (19.33), удовлетворяющего данным начальным условиям. Геометрически это означает, что существует одна интегральная кривая, проходящая через точку $(x_0, y_1^{(0)}, y_2^{(0)}, \ldots, y_n^{(0)})$.

Введем вектор искомых функций $\bar{y} = (y_1, y_2, \ldots, y_n)$ и вектор правых частей уравнений (19.33) $\bar{f} = (f_1, f_2, \ldots, f_n)$. Тогда векторная форма записи системы уравнений первого порядка имеет вид

$$\frac{d\bar{y}}{dx} = \bar{f}(x, \bar{y}). \tag{19.36}$$

Возвращаясь к случаю одного дифференциального уравнения n-го порядка, на основе этой теоремы Коши сформулируем соответствующую теорему существования и единственности решения, о чем говорилось в начале этого параграфа:

Теорема 19.8. *Уравнение* (19.30), *правая часть которого непрерывна по всем аргументам и дифференцируема по ним в некоторой замкнутой области* $D = \{x_0 - a \leqslant x \leqslant x_0 + a, y_0^{(i)} - b_i \leqslant y^{(i)} \leqslant y_0^{(i)} + b_i; i = 0, 1, 2, \ldots, n-1\}$, *имеет единственное решение, удовлетворяющее начальным условиям при* $x = x_0$:

$$x = x_0; \quad y = y_0,\ y' = y_0',\ \ldots,\ y^{(n-1)} = y_0^{(n-1)}, \tag{19.37}$$

где $x_0, y_0, y_0', \ldots, y_0^{(n-1)}$ — *заданные числа.*

Таким образом, теорема Коши определяет *частное решение* уравнения n-го порядка; *общее решение* этого уравнения содержит n произвольных постоянных:

$$y = \varphi(x, C_1, C_2, \ldots, C_n). \tag{19.38}$$

Задача Коши и краевая задача для дифференциального уравнения n-го порядка

Условия (19.37) выделяют частное решение уравнения (19.30), т.е. по заданным значениям искомой функции y_0 и ее производных $y_0', \ldots, y_0^{(n-1)}$ в точке x_0 из множества интегральных кривых, проходящих через эту точку, выделяется только

одна. Таким образом, по заданным условиям (19.37) находятся значения постоянных в равенстве (19.38), которые и определяют вид частного решения уравнения n-го порядка:

$$Y(x) = \varphi(x, C_1^0, C_2^0, \ldots, C_n^0). \tag{19.39}$$

Задача нахождения частного решения дифференциального уравнения n-го порядка, когда все n условий заданы в одной точке $x = x_0$ называется *задачей Коши* для этого уравнения.

Бывают задачи, в которых требуется определить частное решение, когда часть условий (r условий) задана в одной точке, а остальные $(n - r)$ — в другой точке, например:

$$\begin{aligned} x = a, \quad & y = y_1,\ y' = y_1',\ \ldots,\ y^{(r-1)} = y_1^{(r-1)}; \\ x = b, \quad & y^{(r)} = y_2^{(r)},\ \ldots,\ y^{(n-1)} = y_2^{(n-1)}, \end{aligned} \tag{19.40}$$

где $y_1, y_1', \ldots, y_1^{(r-1)}, y_2^{(r)}, \ldots, y_2^{(n-1)}$ — заданные числа. Задача такого рода называется *краевой задачей для дифференциального уравнения n-го порядка*. Бывают случаи, когда вместо значений функции и ее производных в точках $x = a$ и $x = b$ заданы значения их линейных комбинаций; тогда такие условия называются *смешанными*, а задача определения частного решения дифференциального уравнения n-го порядка ($n \geqslant 2$) называется *смешанной краевой задачей*.

Вопрос о существовании и единственности решения краевых задач для уравнений n-го порядка требует своего решения в каждом конкретном случае дифференциального уравнения и типа краевых условий.

Линейные дифференциальные уравнения n-го порядка

Уравнение вида

$$y^{(n)} + a_1(x)y^{(n-1)} + \cdots + a_n(x)y = f(x), \tag{19.41}$$

где $a_1(x), \ldots a_n(x), f(x)$ — известные функции, непрерывные на некотором интервале (a, b), называется *линейным дифференциальным уравнением n-го порядка*. Если $f(x) \equiv 0$, такое уравнение называется *однородным*:

$$y^{(n)} + a_1(x)y^{(n-1)} + \cdots + a_n(x)y = 0. \tag{19.42}$$

Если же $f(x)$ не обращается тождественно в нуль на интервале определения решения $y(x)$, это уравнение называется *неоднородным*.

По аналогии с линейным уравнением второго порядка для однородного уравнения n-го порядка можно выделить совокупность линейно независимых частных решений

$$y_1(x), y_2(x), \ldots, y_n(x), \tag{19.43}$$

которая называется *фундаментальной системой решений*. Определитель Вронского n-го порядка, составленный из решений фундаментальной системы,

$$W_n(x) = \begin{vmatrix} y_1 & y_2 & \cdots & y_n \\ y_1' & y_2' & \cdots & y_n' \\ \cdots & \cdots & \cdots & \cdots \\ y_1^{(n-1)} & y_2^{(n-1)} & \cdots & y_n^{(n-1)} \end{vmatrix}$$

не обращается в нуль.

Основные свойства линейных однородных уравнений n-го порядка таковы.

1°. *Если функции* (19.43) *являются решениями, то их линейная комбинация* $C_1y_1(x) + C_2y_2(x) + \cdots + C_ny_n(x)$, *где* $C_1, C_2, \ldots, C_n$ — *произвольные постоянные, также будет решением этого уравнения.*

2°. *Если* (19.43) *являются фундаментальной системой решений, то любое общее решение уравнения выражается в виде линейной комбинации*

$$y(x) = C_1y_1(x) + C_2y_2(x) + \cdots + C_ny_n(x), \tag{19.44}$$

где $C_1, C_2, \ldots, C_n$ — *произвольные постоянные, не равные нулю одновременно.*

В теории дифференциальных уравнений по аналогии с уравнениями второго порядка устанавливается, что общее решение неоднородного уравнения n-го порядка (19.41) состоит из любого частного решения этого уравнения и общего решения соответствующего однородного уравнения (19.42):

$$y(x) = Y(x) + C_1y_1(x) + C_2y_2(x) + \cdots + C_ny_n(x). \tag{19.45}$$

Линейные дифференциальные уравнения n-го порядка с постоянными коэффициентами

Рассмотрим частный случай уравнения (19.41). Уравнение вида

$$y^{(n)} + a_1y^{(n-1)} + \cdots + a_ny = f(x), \tag{19.46}$$

где $a_1, a_2, \ldots a_n$ — постоянные величины, называется *дифференциальным уравнением с постоянными коэффициентами*. Такие уравнения широко используются в математике и в различных приложениях.

Рассмотрим соответствующее однородное уравнение

$$y^{(n)} + a_1y^{(n-1)} + \cdots + a_ny = 0. \tag{19.47}$$

Пусть функция $y = e^{kx}$, где k — постоянное число, является решением уравнения (19.47); подставляя ее в это уравнение, получаем

$$k^ne^{kx} + a_1k^{n-1}e^{kx} + \cdots + a_ne^{kx} =$$
$$= e^{kx}(k^n + a_1k^{n-1} + \cdots + a_n) = 0.$$

Сокращая обе части на e^{kx}, получаем *характеристическое уравнение* однородного дифференциального уравнения n-го порядка:

$$k^n + a_1k^{n-1} + \cdots + a_{n-1}k + a_n = 0. \tag{19.48}$$

Это алгебраическое уравнение n-го порядка относительно k, его корни соответствуют линейно независимым решениям вида e^{kx} однородного уравнения (19.47).

Разные случаи корней характеристического уравнения (19.48) соответствуют разным видам линейно независимых решений фундаментальной системы. По аналогии с теоремой 19.6 (см. параграф 19.1) можно указать четыре основных случая.

1. Уравнение (19.48) имеет n разных вещественных корней $k_{1,2}, \ldots, k_n$; тогда общее решение уравнения (19.47) имеет вид

$$y(x) = C_1e^{k_1x} + C_2e^{k_2x} + \cdots + C_ne^{k_nx}, \tag{19.49}$$

где $C_1, C_2, \ldots, C_n$ — произвольные постоянные.

33 Зак. 6853

2. Уравнение (19.48) имеет кратные вещественные корни. Тогда каждому кратному корню в фундаментальной системе (19.43), а значит, и в общем решении (19.44) соответствует набор решений вида

$$e^{k_s x}, xe^{k_s x}, x^2 e^{k_s x}, \ldots, x^{r-1} e^{k_s x}, \tag{19.50}$$

где r — кратность корня k_s.

3. Простым комплексно-сопряженным корням вида $k = a \pm bi$, где a и b — действительные числа, соответствуют решения вида

$$e^{ax}(C_1 \sin bx + C_2 \cos bx), \tag{19.51}$$

входящие в общее решение однородного уравнения (19.47), где C_1 и C_2 — произвольные постоянные.

4. Наконец, комплексно-сопряженным корням кратности r соответствуют решения вида

$$\begin{aligned} &e^{ax}\cos bx,\ xe^{ax}\cos bx,\ \ldots,\ x^{r-1}e^{ax}\cos bx; \\ &e^{ax}\sin bx,\ xe^{ax}\sin bx,\ \ldots,\ x^{r-1}e^{ax}\sin bx. \end{aligned} \tag{19.52}$$

Общее решение неоднородного дифференциального линейного уравнения n-го порядка с постоянными коэффициентами может быть найдено, согласно формуле (19.45), как сумма любого частного решения этого уравнения и общего решения однородного уравнения с постоянными коэффициентами.

Рассмотрим примеры нахождения общих решений дифференциальных уравнений с постоянными коэффициентами.

Пример 1. $y''' - 6y'' + 11y' - 6y = 0$.

Характеристическое уравнение для этого случая имеет вид

$$k^3 - 6k^2 + 11k - 6 = 0.$$

Нетрудно видеть, что корни этого уравнения $k_1 = 1$, $k_2 = 2$, $k_3 = 3$. Следовательно, общее решение данного дифференциального уравнения дается формулой

$$y(x) = C_1 e^x + C_2 e^{2x} + C_3 e^{3x},$$

где C_1, C_2, C_3 — произвольные постоянные.

Пример 2. $y^{(4)} - 2y''' + 5y'' - 8y' + 4y = 0$.

Составим соответствующее характеристическое уравнение:

$$k^4 - 2k^3 + 5k^2 - 8k + 4 = 0.$$

Решая его, находим: $k_1 = k_2 = 1$, $k_{3,4} = \pm 2i$. Следовательно, согласно приведенным выше пунктам 2 и 3, общее решение данного уравнения четвертого порядка выражается формулой

$$y(x) = e^x(C_1 + xC_2) + C_3 \cos 2x + C_4 \sin 2x,$$

где C_1, C_2, C_3, C_4 — произвольные постоянные.

Пример 3. $y''' - 2y' - 4y = 3$.

Это неоднородное уравнение. Нетрудно видеть, что одно из его частных решений $Y = -3/4$. Теперь составим характеристическое уравнение:

$$k^3 - 2k - 4 = 0.$$

Решая, находим его корни: $k_1 = 2, k_{2,3} = -1 \pm i$. Отсюда общее решение исходного неоднородного уравнения имеет вид

$$y(x) = C_1 e^{2x} + e^{-x}(C_2 \cos x + C_3 \sin x) - 3/4,$$

где C_1, C_2, C_3 — произвольные постоянные.

Глава 20
ПРИМЕНЕНИЕ АППАРАТА ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЭКОНОМИКЕ

В этой главе мы рассмотрим некоторые примеры применения теории дифференциальных уравнений в непрерывных моделях экономики, где независимой переменной является время t. Такие модели достаточно эффективны при исследовании эволюции экономических систем на длительных интервалах времени; они являются предметом исследования *экономической динамики*.

20.1. Дифференциальные уравнения первого порядка

Модель естественного роста выпуска

Будем полагать, что некоторая продукция продается по фиксированной цене P. Обозначим через $Q(t)$ количество продукции, реализованной на момент времени t; тогда на этот момент времени получен доход, равный $PQ(t)$. Пусть часть указанного дохода расходуется на инвестиции в производство реализуемой продукции, т.е.

$$I(t) = mPQ(t), \tag{20.1}$$

где m — норма инвестиции, постоянное число, причем $0 < m < 1$.

Если исходить из предположения о ненасыщаемости рынка (или о полной реализации производимой продукции), то в результате расширения производства будет получен прирост дохода, часть которого опять будет использована для расширения выпуска продукции. Это приведет к росту скорости выпуска (акселерации), причем скорость выпуска пропорциональна увеличению инвестиций, т. е.

$$Q' = lI, \tag{20.2}$$

где $1/l$ — норма акселерации. Подставив в (20.2) формулу (20.1), получим

$$Q' = kQ, \quad k = lmP. \tag{20.3}$$

Дифференциальное уравнение (20.3) представляет собой уравнение первого порядка с разделяющимися переменными. Общее решение этого уравнения имеет вид

$$Q = Ce^{kt},$$

где C — произвольная постоянная. Пусть в начальный момент времени $t = t_0$ зафиксирован (задан) объем выпуска продукции Q_0:

$$Q_0 = Ce^{kt_0}.$$

Тогда из этого условия можно выразить постоянную C:

$$C = Q_0e^{-kt_0}.$$

Отсюда получаем частное решение уравнения (20.3) — решение задачи Коши для этого уравнения:

$$Q = Q_0e^{k(t-t_0)}. \tag{20.4}$$

Заметим, что математические модели обладают свойством общности. Так, из результатов биологических опытов следует, что процесс размножения бактерий также описывается уравнением (20.3). Процесс радиоактивного распада подчиняется закономерности, установленной формулой (20.4), и т.д.

Рост выпуска в условиях конкуренции

В этой модели мы не будем предполагать, что рынок не насыщается. Пусть $P = P(Q)$ — убывающая функция, т.е. с увеличением объема продукции на рынке цена на нее падает: $dP/dQ < 0$. Теперь из формул (20.1)–(20.3) мы получаем нелинейное дифференциальное уравнение первого порядка относительно Q с разделяющимися переменными:

$$Q' = \alpha P(Q)Q, \qquad \alpha = lm. \tag{20.5}$$

Поскольку все сомножители в правой части этого уравнения положительны, то $Q' > 0$, т.е. функция $Q(t)$ возрастающая.

Характер возрастания функции определяется ее второй производной. Из уравнения (20.5) получаем

$$Q'' = \alpha\left[Q'P(Q) + Q\frac{dP}{dQ}Q'\right] = \alpha Q'\left(P + \frac{dP}{dQ}Q\right).$$

Это равенство можно преобразовать, введя эластичность спроса $E(P) = \dfrac{dQ}{dP}\dfrac{P}{Q}$, откуда $Q'' = \alpha Q'P\left(1 + \dfrac{dP}{dQ}\dfrac{Q}{P}\right)$, или, так как $\dfrac{dQ}{dP} < 0$ и, значит, $E < 0$, окончательно получаем

$$Q'' = \alpha Q'P(1 - 1/|E|). \tag{20.6}$$

Из уравнения (20.6) следует, что при эластичном спросе, т.е. когда $|E| > 1$, $Q'' > 0$ и график функции $Q(t)$ имеет направление выпуклости вниз, что означает прогрессирующий рост; при неэластичном спросе $|E| < 1$, $Q'' < 0$ — направление выпуклости функции $Q(t)$ вверх, что означает замедленный рост (насыщение).

Для простоты примем зависимость $P(Q)$ в виде линейной функции

$$P(Q) = a - bQ, \qquad a > 0, \qquad b > 0, \tag{20.7}$$

(рис. 20.1). Тогда уравнение (20.5) имеет вид

$$Q' = \alpha(a - bQ)Q, \tag{20.8}$$

откуда

$$Q'' = \alpha Q'(a - 2bQ). \tag{20.9}$$

Из соотношений (20.8) и (20.9) получаем: $Q' = 0$ при $Q = 0$ и при $Q = a/b$, $Q'' > 0$ при $Q < a/(2b)$ и $Q'' < 0$ при $Q > a/(2b)$; $Q = a/(2b)$ — точка перегиба графика функции $Q = Q(t)$. Приведенный на рис 20.2 график этой функции [одной из интегральных кривых дифференциального уравнения (20.8)] носит название *логистической кривой*.

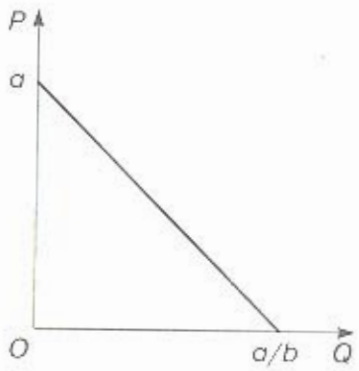

Рис. 20.1

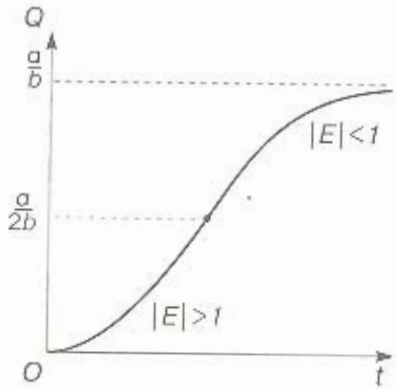

Рис. 20.2

Аналогичные кривые характеризуют и другие процессы, например размножение бактерий в ограниченной среде обитания, динамику эпидемий внутри ограниченной общности биологических организмов и др.

Динамическая модель Кейнса

Рассмотрим простейшую балансовую модель, включающую в себя основные компоненты динамики расходной и доходной частей экономики. Пусть $Y(t)$, $E(t)$, $S(t)$, $I(t)$ — соответственно национальный доход, государственные расходы, потребление и инвестиции. Все эти величины рассматриваются как функции времени t. Тогда справедливы следующие соотношения:

$$\begin{cases} Y(t) = S(t) + I(t) + E(t), \\ S(t) = a(t)Y(t) + b(t), \\ I(t) = k(t)Y'(t), \end{cases} \tag{20.10}$$

где $a(t)$ — коэффициент склонности к потреблению ($0<a(t)<1$); $b(t)$ — автономное (конечное) потребление; $k(t)$ — норма акселерации. Все функции, входящие в уравнения (20.10), положительны.

Поясним смысл уравнений (20.10). Сумма всех расходов должна быть равной национальному доходу — этот баланс отражен в первом уравнении. Общее потребление состоит из внутреннего потребления некоторой части национального дохода в народном хозяйстве плюс конечное потребление — эти составляющие показаны во втором уравнении. Наконец, размер инвестиций не может быть произвольным: он определяется произведением нормы акселерации, величина которой характеризуется уровнем технологии и инфраструктуры данного государства, на предельный национальный доход.

Будем полагать, что функции $a(t)$, $b(t)$, $k(t)$ и $E(t)$ заданы — они являются характеристиками функционирования и эволюции данного государства. Требуется найти динамику национального дохода или Y как функцию времени t.

Подставим выражения для $S(t)$ из второго уравнения и $I(t)$ из третьего уравнения в первое уравнение. После приведения подобных получаем дифференциальное неоднородное линейное уравнение первого порядка для функции $Y(t)$:

$$Y' = \frac{1 - a(t)}{k(t)}Y - \frac{b(t) + E(t)}{k(t)}. \tag{20.11}$$

Как известно из параграфа 18.2, существует достаточно сложная формула общего решения этого уравнения. Мы проанализируем более простой случай, полагая основные параметры задачи a, b, k и E постоянными числами. Тогда уравнение (20.11) упрощается до случая линейного дифференциального уравнения первого порядка с постоянными коэффициентами:

$$Y' = \frac{1-a}{k}Y - \frac{b+E}{k}. \tag{20.12}$$

Как известно, общее решение неоднородного уравнения есть сумма какого-либо его частного решения и общего решения соответствующего однородного уравнения. В качестве частного решения уравнения (20.12) возьмем так называемое *равновесное* решение, когда $Y' = 0$, т.е.

$$Y_p = \frac{b+E}{1-a}. \tag{20.13}$$

Нетрудно видеть, что эта величина положительна. Общее решение однородного уравнения дается формулой $y = C\exp\left(\frac{1-a}{k}t\right)$, так что общее решение уравнения (20.12) имеет вид

$$Y(t) = \frac{b+E}{1-a} + Ce^{\frac{1-a}{k}t}. \tag{20.14}$$

Интегральные кривые уравнения (20.12) показаны на рис. 20.3. Если в начальный момент времени $Y_0 < Y_p$, то $C = Y_0 - Y_p < 0$ и кривые уходят вниз от равновесного решения (20.13), т.е. национальный доход со временем падает при заданных параметрах задачи a, b, k и E, так как показатель экспоненты в (20.14) положителен. Если же $Y_0 > Y_p$, то $C > 0$ и национальный доход растет — интегральные кривые уходят вверх от равновесной прямой $Y = Y_p$. Для автономного дифференциального уравнения (20.12) стационарная точка (20.13) является точкой неустойчивого равновесия.

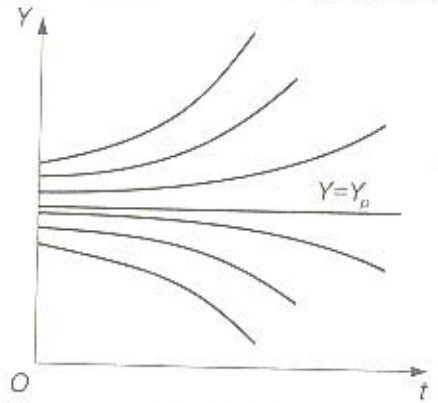

Рис. 20.3

Неоклассическая модель роста

Пусть $Y = F(K, L)$ — национальный доход, где F — однородная производственная функция первого порядка ($F(tK, tL) = tF(K, L)$), K — объем капиталовложений (производственных фондов), L — объем затрат труда. Введем в рассмотрение величину фондовооруженности $k = K/L$, тогда производительность труда выражается формулой

$$f(k) = \frac{F(K,L)}{L} = F(k, l). \tag{20.15}$$

Целью рассматриваемой задачи является описание динамики фондовооруженности или представление ее как функции от времени t. Поскольку любая модель базируется на определенных предпосылках, нам нужно сделать некоторые предположения и ввести ряд определяющих параметров. В данном случае будем полагать, что выполняются следующие предположения:

1) имеет место естественный прирост во времени трудовых ресурсов

$$L' = \alpha L; \tag{20.16}$$

2) инвестиции расходуются на увеличение производственных фондов и на амортизацию, т. е.

$$I = K' + \beta K, \tag{20.17}$$

где β — норма амортизации.

Тогда, если l — норма инвестиций, $I = lY = K' + \beta K$, или

$$K' = lF(K, L) - \beta K. \tag{20.18}$$

Из определения фондовооруженности k вытекает, что

$$\ln k = \ln K - \ln L.$$

Дифференцируя это равенство по t, имеем

$$\frac{k'}{k} = \frac{K'}{K} - \frac{L'}{L}.$$

Подставив в полученное равенство выражения (20.16) и (20.18), получаем уравнение относительно неизвестной функции k:

$$k' = lf(k) - (\alpha + \beta)k, \tag{20.19}$$

где $f(k)$ определена по формуле (20.15).

Полученное соотношение (20.19) представляет собой нелинейное дифференциальное уравнение первого порядка с разделяющимися переменными, которое является автономным. Выделим стационарное решение этого уравнения; из условия $k' = 0$ следует, что

$$lf(k) - (\alpha + \beta)k = 0, \tag{20.20}$$

т. е. $k_{st} = const$ — постоянная величина, являющаяся корнем этого нелинейного алгебраического уравнения.

Рассмотрим конкретную задачу: для производственной функции $F(K, L) = \sqrt{KL}$ найти интегральные кривые уравнения (20.19) и стационарное решение.

Из (20.15) следует, что $f(k) = \sqrt{k}$, и тогда уравнение (20.19) имеет вид

$$\frac{dk}{dt} = l\sqrt{k} - (\alpha + \beta)k. \tag{20.21}$$

Стационарное решение этого автономного уравнения следует из равенства

$$l\sqrt{k} - (\alpha + \beta)k = 0,$$

откуда получаем ненулевое частное решение уравнения (20.19): $k_{st} = l^2/(\alpha + \beta)^2$.

Дифференциальное уравнение (20.21) решаем методом разделения переменных:

$$\frac{dk}{\sqrt{k}\left[l - (\alpha + \beta)\sqrt{k}\right]} = dt.$$

Интегрируя это уравнение с заменой переменной $\sqrt{k} = z$, получаем его общее решение в окончательном виде

$$k(t) = \left[\frac{l}{\alpha + \beta} + C\exp\left(-\frac{\alpha + \beta}{2}t\right)\right]^2. \tag{20.22}$$

Семейство интегральных кривых сходится сверху и снизу к стационарному решению (рис. 20.4): т. е. $k \to k_{st}$ при $t \to \infty$.

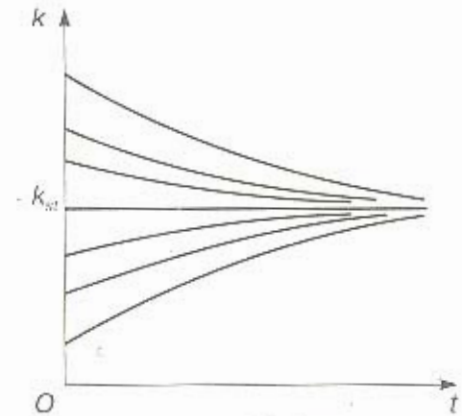

Рис. 20.4

Следовательно, при неизменных входных параметрах задачи l, α и β функция фондовооруженности устойчиво стремится к стационарному значению независимо от начальных условий. В таком случае говорят, что $k = k_{st}$ является точкой устойчивого равновесия.

20.2. Дифференциальные уравнения второго порядка

Рассмотрим модель рынка с прогнозируемыми ценами. В простых моделях рынка обычно полагают, что спрос и предложение зависят только от текущей цены на товар. Однако в

реальных ситуациях они зависят еще и от тенденции ценообразования, и от темпов изменения цены. В моделях с непрерывными и дифференцируемыми по времени t функциями эти характеристики описываются соответственно первой и второй производными функции цены $P(t)$.

Рассмотрим конкретный пример. Пусть функции спроса D предложения S имеют следующие зависимости от цены P и ее производных:

$$\begin{aligned} D(t) &= 3P'' - P' - 2P + 18, \\ S(t) &= 4P'' + P' + 3P + 3. \end{aligned} \tag{20.23}$$

Принятые в (20.23) зависимости вполне реалистичны: поясним слагаемые с производными функции цены.

1. *Спрос* "подогревается" темпом изменения цены: если темп роста растет ($P'' > 0$), рынок увеличивает интерес к товару, и наоборот. Быстрый рост цены отпугивает покупателя, поэтому слагаемое с первой производной функции цены входит со знаком "минус".

2. *Предложение* в еще большей мере усиливается темпом изменения цены, поэтому коэффициент при P'' в функции $S(t)$ больше, чем в $D(t)$. Рост цены также увеличивает предложение — потому слагаемое, содержащее P', входит в выражение для $S(t)$ со знаком "плюс".

Требуется установить зависимость цены от времени. Поскольку равновесное состояние рынка характеризуется равенством $D = S$, приравняем правые части уравнений (20.23). После приведения подобных получаем

$$P'' + 2P' + 5P = 15. \tag{20.24}$$

Соотношение (20.24) представляет линейное неоднородное дифференциальное уравнение второго порядка относительно функции $P(t)$. Как было установлено в параграфе 19.1, общее решение такого уравнения состоит из суммы какого-либо его частного решения и общего решения соответствующего однородного уравнения

$$P'' + 2P' + 5P = 0. \tag{20.25}$$

Характеристическое уравнение имеет вид

$$k^2 + 2k + 5 = 0.$$

Его корни — комплексно-сопряженные числа: $k_{1,2} = -1 \pm 2i$, и, следовательно, общее решение уравнения (20.25) дается формулой

$$\tilde{P}(t) = e^{-t}(C_1 \cos 2t + C_2 \sin 2t),$$

где C_1 и C_2 — произвольные постоянные. В качестве частного решения неоднородного уравнения (20.24) возьмем решение $P = P_{st}$ — постоянную величину как *установившуюся* цену. Подстановка в уравнение (20.24) дает значение P_{st}:

$$P_{st} = 3.$$

Таким образом, общее решение уравнения (20.24) имеет вид

$$P(t) = 3 + e^{-t}(C_1 \cos 2t + C_2 \sin 2t). \tag{20.26}$$

Нетрудно видеть, что $P(t) \to P_{st} = 3$ при $t \to \infty$, т.е. все интегральные кривые имеют горизонтальную асимптоту $P = 3$ и колеблются около нее. Это означает, что все цены стремятся к установлению с колебаниями около установившейся цены P_{st}, причем амплитуда этих колебаний затухает со временем.

Приведем частные решения этой задачи в двух вариантах: задача Коши и смешанная задача.

Задача Коши. Пусть в начальный момент времени известны цена и тенденция ее изменения:

$$t = 0; \qquad P = 4, \qquad P' = 1. \tag{20.27}$$

Подставляя первое условие в формулу (20.26), получаем $P(0) = C_1 + 3 = 4$, откуда $C_1 = 1$, т.е.

$$P(t) = 3 + e^{-t}(\cos 2t + C_2 \sin 2t). \tag{20.28}$$

Дифференцируя, имеем отсюда

$$P'(t) = e^{-t}[(2C_2 - 1)\cos 2t - (C_2 + 2)\sin 2t].$$

Теперь реализуем второе условие задачи Коши: $P'(0) = 2C_2 - 1 = 1$, откуда $C_2 = 1$. Окончательно получаем, что решение задачи Коши имеет вид

$$P(t) = 3 + e^{-t}(\cos 2t + \sin 2t),$$

или в более удобной форме:

$$P(t) = 3 + \sqrt{2}e^{-t}\cos(2t - \pi/4). \qquad (20.29)$$

На рис. 20.5 изображена интегральная кривая, соответствующая решению данной задачи Коши (кривая 1).

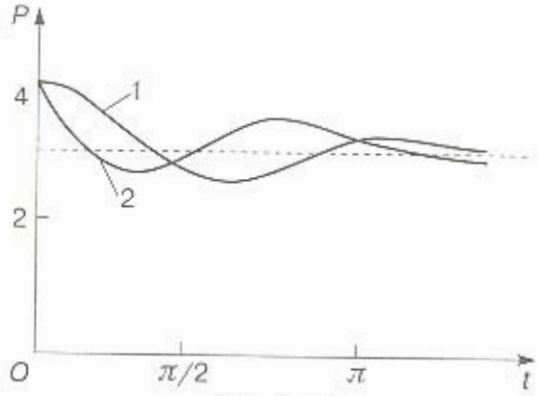

Рис. 20.5

Смешанная задача. Пусть в начальный момент времени известны цена и спрос:

$$t = 0; \qquad P = 4, \qquad D = 16. \qquad (20.30)$$

Поскольку первое начальное условие такое же, как и в предыдущем случае, то имеем и здесь решение (20.28). Тогда производные функции $P(t)$ выражаются формулами

$$P'(t) = e^{-t}[(2C_2 - 1)\cos 2t - (C_2 + 2)\sin 2t],$$
$$P''(t) = -e^{-t}[(4C_2 + 3)\cos 2t + (3C_2 - 5)\sin 2t].$$

Отсюда $P'(0) = 2C_2 - 1$ и $P''(0) = -4C_2 - 3$. Подставляя эти равенства во второе условие задачи, т.е. $D(0) = 16$, имеем

с учетом вида $D(t)$ из первой формулы (20.23): $C_2 = -1$. Итак, решение данной задачи имеет вид

$$P(t) = 3 + e^{-t}(\cos 2t - \sin 2t),$$

или в более удобной форме:

$$P(t) = 3 - \sqrt{2}e^{-t}\sin(2t - \pi/4). \qquad (20.31)$$

На рис. 20.5 решению данной смешанной задачи соответствует интегральная кривая 2.

Глава 21
ЭЛЕМЕНТЫ ТЕОРИИ ЛИНЕЙНЫХ РАЗНОСТНЫХ УРАВНЕНИЙ

21.1. Основные понятия

Сетки и сеточные функции

Определение 1. *Сеткой на отрезке* $[a, b]$ называется любое конечное множество точек этого отрезка.

Точки сетки обычно называют ее *узлами*. Мы уже имели дело с сетками и их узлами, когда занимались приближенным вычислением определенных интегралов по формулам трапеций и Симпсона (см. параграф 7.8). Будем обозначать через ω_N сетку, удовлетворяющую условиям

$$a = x_0 < x_1 < x_2 < \cdots < x_N = b. \qquad (21.1)$$

Это означает, что сетка ω_N состоит из $N + 1$ узла $x_i \in \omega_N$.

Если узлы сетки делят отрезок $[a, b]$ на неравные отрезки, то такие сетки называются *неравномерными*. Множество точек на $[a, b]$

$$\omega_N = \{x_i = a + ih, i = 0, 1, 2, \ldots, N\}, \qquad (21.2)$$

называется *равномерной сеткой с шагом* $h = (b - a)/N$.

Определение 2. Функция, определенная в точках сетки, называется *сеточной функцией*.

Соответствующие значения сеточной функции в узлах сетки обозначаются посредством символов y_i или f_i. В случае если сеточная функция определена на равномерной сетке, ее значения удобно обозначать через $y(i)$, где i — номер узла сетки, $i = 0, 1, 2, \ldots, N$.

Таким образом, сеточная функция является конечным множеством чисел, когда каждому узлу сетки ω_N ставится по определенному закону в соответствие лишь одно число из этого множества. Сетки и сеточные функции представляют собой *конечные дискретные множества точек*. В случае равномерной сетки удобно ввести зависимость сеточной функции от номера узла, т.е. рассматривать сеточную функцию как функцию от целочисленного аргумента.

Пример 1. Составить равномерную сетку и определенную на ней сеточную функцию при $N = 4$, если функция $y = x^2$ определена на отрезке $[0, 1]$.

Решение. Шаг сетки $h = 0,25$, стало быть, сетка $\omega_N = \{0;\ 0{,}25;\ 0{,}5;\ 0{,}75;\ 1\}$; сеточная функция также состоит из пяти чисел — $\{0;\ 0{,}0625;\ 0{,}25;\ 0{,}5625;\ 1\}$.

Сетки и сеточные функции широко используются в самых различных приложениях; наиболее распространенными примерами таких приложений являются разнообразные численные методы решений задач. В настоящее время разработана теория разностных уравнений, получившая свое самостоятельное развитие и имеющая многочисленные приложения. В этой главе приведены основные элементы указанной теории.

Линейные обыкновенные разностные уравнения

Определение 3. Уравнение вида

$$a_n(i)y(i+n) + a_{n-1}(i)y(i+n-1) + \cdots + a_0(i)y(i) = f(i), \quad (21.3)$$

где $a_j(i)$ и $f(i)$ — известные функции целочисленного аргумента; $j = 0, 1, \ldots, n$; $y(i + j)$ — неизвестная функция целочисленного аргумента из некоторого множества целых чисел I, причем $a_n(i)$ и $a_0(i)$ отличны от нуля для любого i, называется *линейным разностным неоднородным уравнением n-го порядка с переменными коэффициентами.*

Из записи разностного уравнения видно, что его порядок определяется максимальным количеством узлов сетки, входящих в неизвестную функцию $y(i + n)$, начиная с номера $i + 1$. Например, равенство

$$(3i^2 + 1)y(i + 2) - 5iy(i + 1) + 2y(i) = e^i$$

является линейным разностным уравнением второго порядка с переменными коэффициентами.

Существует аналогия между обыкновенными дифференциальными уравнениями и разностными уравнениями. Действительно, заменив зависимость от целочисленного аргумента (текущего индекса) i на зависимость от x, а $y(i + j)$ на $y^{(j)}$ $(j = 0, 1, \ldots, n)$, мы получим вместо (21.3) обыкновенное линейное дифференциальное неоднородное уравнение n-го порядка вида (19.41)

$$a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_0(x)y = f(x). \quad (21.4)$$

По аналогии с дифференциальными уравнениями разностные уравнения разделяются на неоднородные вида (21.3) и однородные вида

$$a_n(i)y(i + n) + a_{n-1}(i)y(i + n - 1) + \cdots + a_0(i)y(i) = 0. \quad (21.5)$$

Свойства решений линейных разностных уравнений

Для решений линейных разностных уравнений характерны те же свойства, что и для линейных дифференциальных уравнений. Перечислим их без вывода.

Определение 4. *Решением разностного уравнения* (21.3) называется такая функция $y = y(x)$, последовательность значений $y(i)$ которой удовлетворяет данному уравнению для некоторой области I значений целочисленного аргумента i.

Разностные уравнения могут иметь, как и их дифференциальные аналоги, бесконечное множество решений, которое описывается *общим решением* вида

$$y = y(x, C_1, C_1, \ldots, C_n), \quad (21.6)$$

где $C_1, C_2, \ldots, C_n$ — произвольные постоянные; их количество равно порядку разностного уравнения (21.3). *Частные решения* уравнения (21.5) могут быть выделены из общего решения (21.6) заданием значений функции $y(i)$ в n произвольных, но *расположенных подряд* точках $i_0, i_0+1, \ldots, i_0+n-1$ множества I:

$$y(i_0), y(i_0+1), \ldots, y(i_0+n-1). \tag{21.7}$$

Эти условия называются *начальными условиями* для разностного уравнения n-го порядка; они определяют произвольные постоянные в формуле общего решения (21.6).

Пример 2. Найти общее решение уравнения

$$y(i+1) - e^i y(i) = 0.$$

Это линейное разностное уравнение первого порядка с переменными коэффициентами. Перенеся второе слагаемое в правую часть, последовательно получаем

$$y(i+1) = e^i y(i) = e^i e^{i-1} y(i-1) = e^i e^{i-1} e^{i-2} y(i-2) = \cdots =$$
$$= y(1) \exp\left(\sum_{k=1}^{i} k\right) = y(1) \exp[i(i+1)/2].$$

Поскольку величина $y(1)$ не определена, получаем формулу для общего решения данного уравнения:

$$y(i) = C \exp[i(i-1)/2],$$

где C — произвольная постоянная.

Линейные однородные уравнения. Введем понятие линейно независимых решений однородного уравнения (21.5). Сеточные функции $y_1(i), y_2(i), \ldots, y_n(i)$ называются *линейно независимыми решениями уравнения* (21.5), если:

1) они принимают конечные значения и удовлетворяют уравнению (21.5);

2) равенство

$$C_1 y_1(i) + C_2 y_2(i) + \cdots + C_n y_n(i) = 0 \tag{21.8}$$

при любых постоянных $C_1, C_2, \ldots, C_n$, не равных нулю одновременно, не выполняется ни для каких значений $i \in I$.

Укажем теперь свойства решений линейного однородного разностного уравнения.

1°. Если $y_1(i), y_2(i), \ldots, y_m(i)$ — *решения уравнения* (21.5), *то функция*

$$y(i) = C_1 y_1(i) + C_2 y_2(i) + \cdots + C_m y_m(i), \tag{21.9}$$

где $C_1, C_2, \ldots, C_m$ — произвольные постоянные, также является решением этого уравнения.

2°. *Если $y_1(i), y_2(i), \ldots, y_n(i)$ — линейно независимые решения однородного уравнения* (21.5) *(фундаментальная система решений этого уравнения), то общее решение данного уравнения имеет вид*

$$y(i) = C_1 y_1(i) + C_2 y_2(i) + \cdots + C_n y_n(i). \tag{21.10}$$

По аналогии с определителем Вронского вводится *определитель Казоратти*, составленный из значений решений однородного разностного уравнения n-го порядка:

$$K[y_1(i), \ldots, y_n(i)] =$$
$$= \begin{vmatrix} y_1(i) & y_2(i) & \cdots & y_n(i) \\ y_1(i+1) & y_2(i+1) & \cdots & y_n(i+1) \\ \cdots & \cdots & \cdots & \cdots \\ y_1(i+n-1) & y_2(i+n-1) & \cdots & y_n(i+n-1) \end{vmatrix}.$$

Для того чтобы решения однородного разностного уравнения были линейно независимыми, необходимо и достаточно, чтобы этот определитель не обращался в нуль ни при одном значении $i \in I$.

Неоднородные разностные уравнения. Справедлива следующая теорема.

Теорема 21.1. *Общее решение уравнения* (21.3) *является суммой его частного решения Y и общего решения* (21.10) *соответствующего однородного уравнения* (21.8) :

$$y(i) = Y(i) + C_1 y_1(i) + C_2 y_2(i) + \cdots + C_n y_n(i). \tag{21.11}$$

Для нахождения общего решения неоднородного линейного разностного уравнения (21.3) разработан метод *вариации постоянных*, во многом аналогичный методу для дифференциальных уравнений. Однако это достаточно громоздкий метод и во многих случаях проще подобрать какое-либо частное решение $Y(i)$ неоднородного уравнения или "угадать" его по виду правой части $f(i)$.

Решение линейных уравнений с постоянными коэффициентами

Определение 5. Уравнение вида

$$a_n y(i+n) + a_{n-1} y(i+n-1) + \cdots + a_0 y(i) = f(i), \qquad (21.12)$$

где $a_n, a_{n-1}, \ldots, a_0$ — действительные числа, называется *разностным уравнением n-го порядка с постоянными коэффициентами относительно неизвестной функции y*, определенной на некотором множестве целых чисел I.

Соответствующее однородное уравнение имеет вид

$$a_n y(i+n) + a_{n-1} y(i+n-1) + \cdots + a_0 y(i) = 0. \qquad (21.13)$$

Будем искать частные решения этого уравнения в виде $u(i) = q^i$, где q — число, подлежащее определению. Подставляя $u(i)$ в уравнение (21.13) вместо $y(i)$, получаем

$$a_n q^{i+n} + a_{n-1} q^{i-n-1} + \cdots + a_0 q^i = 0.$$

Так как мы ищем ненулевое решение, то, сокращая обе части этого уравнения на q^i, получим уравнение для q:

$$a_n q^n + a_{n-1} q^{n-1} + \cdots + a_0 = 0. \qquad (21.14)$$

Это уравнение называется *характеристическим уравнением* для однородного уравнения (21.13). Корни характеристического уравнения соответствуют линейно независимым решениям уравнения (21.13).

Рассмотрим отдельно каждый возможный случай корней характеристического уравнения.

1. Пусть уравнение (21.14) имеет n различных простых корней $q_1, q_2, \cdots, q_n$. Тогда функции

$$y_1(i) = q_1^i,\ y_2(i) = q_2^i,\ \ldots,\ y_n(i) = q_n^i \qquad (21.15)$$

являются линейно независимыми решениями уравнения (21.13). Иными словами, общее решение уравнения (21.12), согласно свойству 2° и формуле (21.10), имеет вид

$$y(i) = C_1 q_1^i + C_2 q_2^i + \cdots + C_n q_n^i, \qquad (21.16)$$

где $C_1, C_2, \ldots, C_n$ — произвольные постоянные.

2. Пусть среди корней характеристического уравнения (21.14) имеются комплексно-сопряженные корни вида $\alpha \pm \beta i^*$ ($i^* = \sqrt{-1}$). Тогда этим корням соответствуют линейно независимые решения уравнения (21.13) вида

$$\rho^i \cos i\varphi \ \text{ и } \ \rho^i \sin i\varphi, \quad \rho = \sqrt{\alpha^2 + \beta^2}, \quad \varphi = \operatorname{arctg} \frac{\beta}{\alpha}, \qquad (21.17)$$

где ρ и φ — соответственно модуль и аргумент комплексного корня.

3. Если характеристическое уравнение (21.14) имеет действительный корень q_m кратности r, то ему соответствуют r линейно независимых решений уравнения (21.13) вида

$$q_m^i, iq_m^i, \ldots, i^{r-1} q_m^i. \qquad (21.18)$$

4. Наконец, если характеристическое уравнение (21.14) имеет комплексно-сопряженный корень (см. п. 2), кратность которого равна r, то ему соответствуют $2r$ линейно независимых решений уравнения (21.13) вида

$$\rho^i \cos i\varphi,\ \rho^i \sin i\varphi;\ \rho^i i \cos i\varphi,\ \rho^i i \sin i\varphi;\ \ldots;\ \rho^i i^{r-1} \cos i\varphi,\ \rho^i i^{r-1} \sin i\varphi. \qquad (21.19)$$

Рассмотрим несколько примеров на решение линейных разностных уравнений с постоянными коэффициентами.

Пример 3. Найти общее решение разностного уравнения

$$y(i+2) - 7y(i+1) + 10y(i) = 0.$$

Это однородное уравнение второго порядка; его характеристическое уравнение

$$q^2 - 7q + 10 = 0$$

имеет простые корни $q_1 = 2$ и $q_2 = 5$. Согласно п. 1, общее решение данного уравнения имеет вид

$$y(i) = C_1 2^i + C_2 5^i.$$

Пример 4. Найти общее решение неоднородного уравнения

$$y(i+2) - 3y(i+1) - 4y(i) = 6 \cdot 5^i.$$

Соответствующее однородное уравнение имеет постоянные коэффициенты; у его характеристического уравнения

$$q^2 - 3q - 4 = 0$$

два простых корня: $q_1 = -1$ и $q_2 = 4$. Следовательно, общее решение однородного уравнения имеет вид

$$\tilde{y}(i) = C_1(-1)^i + C_2 \cdot 4^i,$$

где C_1 и C_2 — произвольные постоянные. Далее одно из частных решений заданного неоднородного уравнения будем искать, сообразуясь с видом правой части: $Y(i) = C \cdot 5^i$, где C — постоянная, подлежащая определению. Подставив указанное частное решение в исходное разностное уравнение, получаем: $C \cdot 5^{i+2} - 3C \cdot 5^{i+1} - 4C \cdot 5^i = 6 \cdot 5^i$ или $25C \cdot 5^i - 15C \cdot 5^i - 4C \cdot 5^i = 6 \cdot 5^i$, откуда $C = 1$. Таким образом, общее решение заданного уравнения как сумма его частного решения и общего решения однородного уравнения имеет вид

$$y(i) = C_1(-1)^i + C_2 \cdot 4^i + 5^i,$$

где C_1 и C_2 — произвольные постоянные.

Пример 5. Найти общее решение неоднородного уравнения

$$y(i+4) - 3y(i+3) - 6y(i+2) + 28y(i+1) - 24y(i) = 8.$$

Это неоднородное уравнение четвертого порядка с постоянными коэффициентами. Найдем его частное решение в виде постоянной величины: $Y = C$; подстановка в уравнение дает $C = -2$. Характеристическое уравнение четвертого порядка

$$q^4 - 3q^3 - 6q^2 + 28q - 24 = (q-2)^3(q+3) = 0$$

имеет корень $q_1 = 2$ кратности 3 и простой корень $q_2 = -3$. Следовательно, в силу п. 3 общее решение данного уравнения имеет вид

$$y(i) = 2^i(C_1 + C_2 i + C_3 i^2) + C_4(-3)^i - 2,$$

где C_1, C_2, C_3, C_4 — произвольные постоянные.

Пример 6. Найти решение задачи Коши для уравнения четвертого порядка

$$y(j+4) + 4y(j+2) + 4y(j) = 0, \quad j \geqslant 4,$$

с заданными начальными условиями

$$y(0) = 1, \quad y(1) = 0, \quad y(2) = -6, \quad y(3) = 4\sqrt{2}.$$

Решение. Характеристическое уравнение

$$q^4 + 4q^2 + 4 = (q^2+2)^2 = 0$$

имеет два комплексно-сопряженных корня кратности 2 каждый:

$$q_{1,2} = \sqrt{2}\left(\cos\frac{\pi}{2} \pm i\sin\frac{\pi}{2}\right), \quad i = \sqrt{-1}.$$

Тогда, в соответствии с п. 4, общее решение данного однородного уравнения имеет вид

$$y(j) = 2^{j/2}\left[(C_1 + C_2 j)\cos\frac{j\pi}{2} + (C_3 + C_4 j)\sin\frac{j\pi}{2}\right],$$

где C_1, C_2, C_3, C_4 — произвольные постоянные.

Теперь найдем частное решение — решение задачи Коши, для чего подставим найденное общее решение в заданные начальные условия. Получаем систему линейных алгебраических уравнений относительно неизвестных величин C_1, C_2, C_3, C_4:

$$\begin{aligned} y(0)=1, &\quad \text{откуда} \quad C_1=1,\\ y(1)=0, &\quad \text{откуда} \quad \sqrt{2}(C_3+C_4)=0,\\ y(2)=-6, &\quad \text{откуда} \quad -2(C_1+2C_2)=-6,\\ y(3)=4\sqrt{2}, &\quad \text{откуда} \quad -2\sqrt{2}(C_3+3C_4)=4\sqrt{2}. \end{aligned}$$

Решая эту систему, находим: $C_2=1$, $C_3=1$, $C_4=-1$. Окончательно решение задачи Коши имеет вид

$$y(j)=2^{j/2}\left[(1+j)\cos\frac{j\pi}{2}+(1-j)\sin\frac{j\pi}{2}\right].$$

Пример 7. Найти решение краевой задачи для однородного уравнения третьего порядка

$$y(i+3)-3y(i+2)+4y(i)=0, \quad 0\leqslant i\leqslant 2N-3,$$

при следующих краевых условиях

$$y(0)=2, \quad y(1)=0, \quad y(2N)=0.$$

Решение. Характеристическое уравнение

$$q^3-3q^2+4=0$$

имеет корень $q_1=2$ кратности 2 и простой корень $q_2=-1$. Следовательно, общее решение данного уравнения имеет вид

$$y(i)=2^i(C_1+C_2i)+C_3(-1)^i,$$

где C_1, C_2, C_3 — произвольные постоянные. Подставляя это решение в заданные граничные условия, получаем систему трех линейных уравнений относительно указанных постоянных:

$$\begin{cases} C_1+C_3=2,\\ 2(C_1+C_2)-C_3=0,\\ (C_1+2NC_2)2^{2N}+C_3=0. \end{cases}$$

Решая эту систему, мы приходим к частному решению данного разностного уравнения, определяемому краевыми условиями задачи:

$$y(i)=\frac{1}{1+(3N-1)2^{2N}}\left\{\left[2(1+2^{2N}N)-(2^{2N}+2)i\right]2^i+\right.$$
$$\left.+(2N-1)2^{2N+1}(-1)^i\right\}.$$

Системы линейных разностных уравнений первого порядка

По аналогии с системами дифференциальных уравнений (см. параграф 19.2) рассматриваются также и системы разностных уравнений. Приведем систему n линейных разностных уравнений первого порядка с постоянными коэффициентами a_{jk}, которая имеет вид

$$\begin{cases} y_1(i+1)=a_{11}y_1(i)+a_{12}y_2(i)+\cdots+a_{1n}y_n(i)+f_1(i),\\ y_2(i+1)=a_{21}y_1(i)+a_{22}y_2(i)+\cdots+a_{2n}y_n(i)+f_2(i),\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\ y_n(i+1)=a_{n1}y_1(i)+a_{n2}y_2(i)+\cdots+a_{nn}y_n(i)+f_n(i). \end{cases} \tag{21.20}$$

Здесь $y_1, y_2, \dots, y_n$ — неизвестные функции, $f_1, f_2, \dots, f_n$ — известные функции целочисленного аргумента i.

К системе разностных уравнений первого порядка можно перейти от линейного разностного уравнения n-го порядка путем введения новых функций соответствующими заменами:

$$y(i+1)=y_1(i), \quad y_1(i+1)=y_2(i), \dots,$$
$$y_{n-1}(i+1)=\left\{-\left[a_{n-1}y_{n-2}(i)+\cdots+a_0y(i)\right]+f(i)\right\}/a_n.$$

Введем векторы-столбцы неизвестных функций и правых частей, а также матрицу коэффициентов:

$$\bar{y}(i)=\begin{pmatrix} y_1(i)\\ y_2(i)\\ \dots\\ y_n(i) \end{pmatrix}, \quad \bar{f}(i)=\begin{pmatrix} f_1(i)\\ f_2(i)\\ \dots\\ f_n(i) \end{pmatrix}, \quad \mathrm{A}=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}. \tag{21.21}$$

Тогда систему разностных уравнений (21.20) можно записать в матричной форме:

$$\bar{y}(i+1) = \mathbf{A}\bar{y}(i) + \bar{f}(i). \tag{21.22}$$

Пусть известен начальный вектор $\bar{y}(0)$. Тогда последовательно получаем из (21.22):

$$\bar{y}(1) = \mathbf{A}\bar{y}(0) + \bar{f}(0),$$
$$\bar{y}(2) = \mathbf{A}\bar{y}(1) + \bar{f}(1) = \mathbf{A}^2\bar{y}(0) + \mathbf{A}\bar{f}(0) + \bar{f}(1)$$

и так далее. Методом математической индукции можно доказать следующую формулу:

$$\bar{y}(i) = \mathbf{A}^i\bar{y}(0) + \sum_{k=0}^{i-1} \mathbf{A}^{i-k-1}\bar{f}(k). \tag{21.23}$$

Здесь матрица $\mathbf{A}^i$ называется *матрицей изменения состояния* для системы (21.20).

21.2. Применение разностных уравнений в экономике

Модель рынка с запаздыванием сбыта

В реальности часто складывается такая рыночная ситуация, что цикл производства продукции отстает от цикла ее реализации. Это характерно, например, для сельского хозяйства. Да и в промышленном производстве предложение формируется на основе цены в предшествующий период. Таким образом, функция предложения S сдвинута по времени относительно цены P, т.е. будем полагать, что $S(t) = S(P(t-1))$, в то время как функция спроса D одномоментно отвечает цене: $D(t) = D(P(t))$. Для простоты рассмотрим линейные зависимости спроса и предложения от цены:

$$D(t) = a - bP(t), \qquad S(t) = -c + dP(t-1); \qquad a, b, c, d > 0. \tag{21.24}$$

Условие равновесия предполагает равенство предлагаемого и востребованного объемов товара: $D(t) = S(t)$, откуда с учетом (21.24) имеем

$$bP(t) + dP(t-1) = a + c.$$

Поделив обе части этого равенства на b и переходя для удобства на шаг вперед по времени, получаем линейное неоднородное разностное уравнение первого порядка относительно цены P с постоянными коэффициентами:

$$P(t+1) + \frac{d}{b}P(t) = \frac{a+c}{b}. \tag{21.25}$$

Характеристическое уравнение имеет единственный корень, равный $-d/p$. Частное же решение уравнения (21.25) удобно искать в виде постоянной величины; после подстановки в это уравнение оно легко определяется:

$$P_* = \frac{a+c}{b+d}. \tag{21.26}$$

Величина P_* является равновесной ценой. Общее решение уравнения (21.25) определяется формулой

$$P(t) = C\left(-\frac{d}{b}\right)^t + P_*, \tag{21.27}$$

где C — произвольная постоянная. Пусть в начальный момент времени известна цена $P(0) = P_0$ (задача Коши), тогда подстановкой в равенство (21.27) находим $C = P_0 - P_*$, так что в окончательном виде получаем

$$P(t) = (P_0 - P_*)\left(-\frac{d}{b}\right)^t + P_*. \tag{21.28}$$

Проанализируем полученное решение. В зависимости от входных параметров задачи b и d формул (21.24) динамика цены во времени может быть разной. Здесь возможны три варианта:

а) $d/b > 1$ — текущая цена расходится с равновесной ценой с увеличением размаха колебаний вокруг нее;

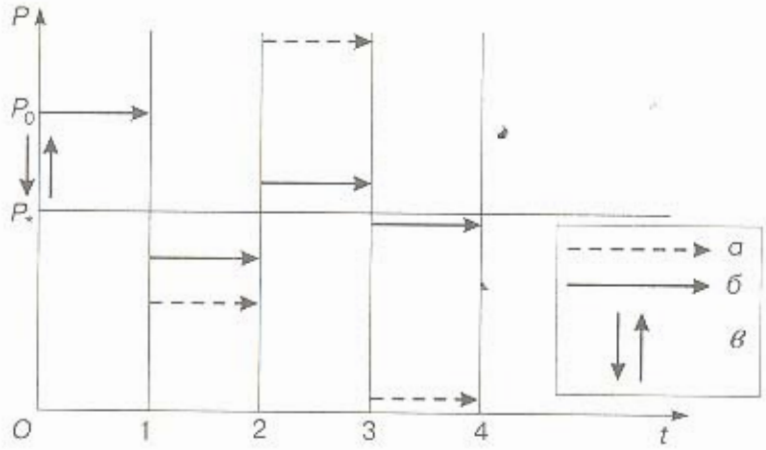

Рис. 21.1

б) $d/b < 1$ — текущая цена стремится к равновесной цене с колебаниями около нее;

в) $d/b = 1$ — две точки равновесия: в зависимости от четности t имеет место колебание от одной точки к другой.

На рис. 21.1 эти случаи показаны соответствующими сериями дискретных отрезков.

Рыночная модель с запасами

В этой модели предполагаются запасы товара как разность между предложением S и спросом D. Примем следующие допущения.

1. Спрос $D(t)$ и предложение $S(t)$ представляют собой линейные функции от текущей цены $P(t)$:

$$D(t) = a - bP(t), \quad S(t) = -c + dP(t); \qquad a, b, c, d > 0. \tag{21.29}$$

2. Цена, устанавливаемая на рынке, зависит от объема запаса продукции на предшествующий период, причем разница в ценах во времени пропорциональна отрицательной величине запаса с некоторым коэффициентом k (при наличии запаса цена на товар в последующий период падает):

$$P(t+1) = P(t) - k[S(t) - D(t)]. \tag{21.30}$$

Подстановка соотношений (21.29) в (21.30) приводит к линейному разностному уравнению первого порядка с постоянными коэффициентами относительно цены $P(t)$:

$$P(t+1) - [1 - k(b+d)]P(t) = k(a+c). \tag{21.31}$$

Пусть P_0 — значение цены в начальный момент времени $t = 0$, тогда решение этого уравнения имеет вид

$$P(t) = (P_0 - P_*)\,[1 - k(b+d)]^t + P_*, \tag{21.32}$$

где $P_* = (a+c)/(b+d)$ — равновесная цена, или стационарное решение уравнения (21.31).

Сходимость $P(t)$ во времени к значению P_* существенно зависит от величины и знака основания степени в (21.32) $\alpha = 1 - k(b+d)$. Рассмотрим все возможные случаи сочетания этих параметров задачи:

а) $0 < \alpha < 1$, откуда $k < 1/(b+d)$ — монотонная сходимость $P(t)$ к равновесному значению P_*;

б) $\alpha = 0$, откуда $k = 1/(b+d)$, т.е. $P(t) \equiv P_*$;

в) $-1 < \alpha < 0$, откуда $1/(b+d) < k < 2/(b+d)$ — сходимость цены $P(t)$ к равновесному значению P_* с колебаниями около него;

г) $\alpha = -1$, т.е. $k = 2/(b+d)$ — две точки равновесия: P_* и P_0, на каждом шаге по времени цена "перескакивает" с одного значения на другое;

д) $\alpha < -1$, т.е. $k > 2/(b+d)$ — цена $P(t)$ расходится с увеличением амплитуды колебаний.

Все указанные случаи приведены на рис. 21.2 в виде соответствующих серий дискретных отрезков.

Динамическая модель Леонтьева

В параграфе 17.2 была рассмотрена продуктивная модель межотраслевого баланса безотносительно ко времени, т.е. все ее компоненты полагались осредненными за некоторый временной интервал и "одномоментными". В реальности продукт,

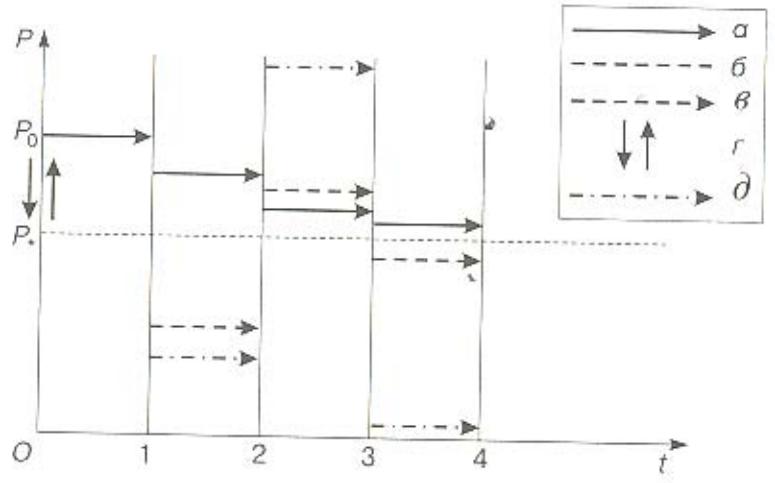

Рис. 21.2

предназначенный для внутреннего и конечного потребления в период t, определяется не текущим выпуском, а выпуском в последующий период $t+1$. Эта задержка производства обусловлена многими факторами, в частности инерцией планирования и перенастройки, мобилизацией внутренних ресурсов и изменением транспорта сырья и т.д.

С учетом этого система уравнений баланса, в предположении о постоянстве доли внутреннего потребления каждой отраслью, будет иметь следующий вид

$$\begin{cases} x_1(t+1) = a_{11}x_1(t) + \cdots + a_{1n}x_n(t) + y_1(t), \\ x_2(t+1) = a_{21}x_1(t) + \cdots + a_{2n}x_n(t) + y_2(t), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n(t+1) = a_{n1}x_1(t) + \cdots + a_{nn}x_n(t) + y_n(t). \end{cases} \tag{21.33}$$

Соотношения (21.33) составляют систему линейных разностных уравнений первого порядка с постоянными коэффициентами a_{ij}.

Как и прежде, введем вектор валового выпуска $\bar{x}(t)$, матрицу прямых затрат $\mathbf{A} = \|a_{ij}\|$ и вектор конечного потребления $\bar{y}(t)$. Тогда систему (21.33) можно переписать в матричной форме:

$$\bar{x}(t+1) = \mathbf{A}\bar{x}(t) + \bar{y}(t). \tag{21.34}$$

Теперь уже задача формулируется следующим образом: при заданном векторе конечного потребления $\bar{y}(t)$ и матрице прямых затрат $\mathbf{A}$ определить динамику (изменение во времени) вектора валового выпуска $\bar{x}(t)$.

Одна из основных задач прогноза с использованием динамической модели Леонтьева такова: известны динамика вектора конечного потребления $\bar{y}(t)$ и вектор валового выпуска $\bar{x}_0$ в начальный момент времени $t=0$; требуется рассчитать вектор валового выпуска на момент времени $t \geqslant 1$. Эту задачу можно решить при помощи формулы (21.23):

$$\bar{x}(t) = \mathbf{A}^t\bar{x}(0) + \sum_{k=0}^{t-1} \mathbf{A}^{t-k-1}\bar{y}(k). \tag{21.35}$$

Пример 1. Обратимся к данным табл. 17.6 (см. параграф 17.2). Пусть известны продуктивная матрица затрат $\mathbf{A}$, а также заданные в момент времени $t=0$ векторы валового выпуска $\bar{x}(0)$ и $\bar{y}(0)$, указанные в этой таблице:

$$\bar{x}(0) = \begin{pmatrix} 100 \\ 100 \\ 50 \end{pmatrix}, \quad \bar{y}(0) = \begin{pmatrix} 40 \\ 60 \\ 10 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0{,}05 & 0{,}35 & 0{,}40 \\ 0{,}10 & 0{,}10 & 0{,}40 \\ 0{,}20 & 0{,}10 & 0{,}20 \end{pmatrix}.$$

Требуется рассчитать вектор валового выпуска на момент времени $t=2$, если все компоненты вектора конечного потребления $\bar{y}$ увеличиваются на 30% за каждый период.

Решение. Вектор конечного потребления, согласно условию задачи, имеет вид

$$\bar{y}(t) = \bar{y}(0) \cdot (1,3)^t.$$

Применяя формулу (21.35), получаем

$$\bar{x}(2) = \mathbf{A}^2\bar{x}(0) + \mathbf{A}\bar{y}(0) + \bar{y}(1).$$

Теперь нужно вычислить матрицу $\mathbf{A}^2$ изменения состояния и вектор $\bar{y}(1)$ и подставить их в это уравнение. Выполнив указанные действия, получим решение поставленной задачи:

$$\bar{x}(2) = \begin{pmatrix} 0{,}1175 & 0{,}0925 & 0{,}24 \\ 0{,}095 & 0{,}085 & 0{,}16 \\ 0{,}06 & 0{,}1 & 0{,}16 \end{pmatrix} \begin{pmatrix} 100 \\ 100 \\ 50 \end{pmatrix} +$$

$$+\begin{pmatrix}0{,}05 & 0{,}35 & 0{,}4\\ 0{,}1 & 0{,}1 & 0{,}4\\ 0{,}2 & 0{,}1 & 0{,}2\end{pmatrix}\begin{pmatrix}40\\60\\10\end{pmatrix}+\begin{pmatrix}52\\78\\13\end{pmatrix}=$$

$$=\begin{pmatrix}33\\26\\24\end{pmatrix}+\begin{pmatrix}27\\14\\16\end{pmatrix}+\begin{pmatrix}52\\78\\13\end{pmatrix}=\begin{pmatrix}112\\118\\53\end{pmatrix}.$$

Таким образом, при указанном темпе роста продукта конечного потребления необходимо через два временных цикла увеличить компоненты валового выпуска соответственно на 12, 18 и 6% по сравнению с исходными величинами на начальный момент времени.

Часть 4. ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Глава 22
ОСНОВНЫЕ ПОЛОЖЕНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

События, происходящие в окружающем нас мире, можно разделить на три вида: достоверные, невозможные и случайные. *Достоверным* относительно комплекса условий S называется событие, которое обязательно произойдет при осуществлении этого комплекса условий. Например, если гладкий желоб с лежащим внутри него тяжелым шариком наклонить, то шарик обязательно покатится по желобу в сторону уклона. *Невозможным* называется событие, которое заведомо не произойдет при осуществлении комплекса условий S. Например, из герметически изолированного сосуда вода не может вылиться. *Случайным* относительно комплекса условий S называется событие, которое при осуществлении указанного комплекса условий может либо произойти, либо не произойти. Например, если вы уронили фарфоровую чашку на пол, то она может как разбиться, так и остаться неповрежденной.

Теория вероятностей имеет дело со случайными событиями. Однако она не может предсказать, произойдет единичное событие или нет. Теория вероятностей изучает *вероятностные закономерности массовых однородных случайных событий*. Ее методы получили широкое распространение в различных областях естествознания и в прикладных проблемах техники. Теория вероятностей легла в основу теории массового

обслуживания и теории надежности. В последние годы аппарат теории вероятностей активно используется в экономике.

22.1. Основные понятия теории вероятностей

Некоторые формулы комбинаторики

Пусть задано конечное множество элементов некоторой природы. Из них можно составлять определенные комбинации, количества которых изучает комбинаторика. Некоторые ее формулы используются в теории вероятности; приведем их.

Комбинации, состоящие из одной и той же совокупности n различных элементов и отличающиеся только порядком их расположения, называются *перестановками*. Число всех возможных перестановок определяется произведением чисел от единицы до n:

$$P_n = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n = n!$$

Пример 1. Сколько четырехзначных чисел можно составить из цифр 1, 2, 3 и 4 с использованием всех указанных цифр в каждом числе ?

Решение. Искомое число равно $P_4 = 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$.

Комбинации по m элементов, составленные из n различных элементов ($m \leqslant n$), отличающиеся друг от друга либо элементами, либо их порядком, называются *размещениями*. Число всевозможных размещений

$$A_n^m = n(n-1)(n-2)\ldots(n-m+1).$$

Пример 2. Сколько трехзначных чисел можно составить из семи различных цифр при отсутствии среди них нуля?

Решение. Искомое количество цифр

$$A_7^3 = 7 \cdot 6 \cdot 5 = 210.$$

Комбинации, содержащие по m элементов каждая, составленные из n различных элементов ($m \leqslant n$) и различающиеся

хотя бы одним элементом, называются *сочетаниями*. Число сочетаний дается формулой

$$C_n^m = \frac{n!}{m!\,(n-m)!}.$$

Можно показать, что справедливы формулы

$$C_n^m = C_n^{n-m}, \quad A_n^m = P_m C_n^m. \tag{22.1}$$

В частности, первую из формул удобно использовать в расчетах, когда $m > n/2$.

Напомним формулу бинома Ньютона, в которой участвуют коэффициенты (22.1):

$$(p+q)^n = C_n^n p^n q^0 + C_n^{n-1} p^{n-1} q^1 + \ldots + \\ + C_n^1 p^1 q^{n-1} + C_n^0 p^0 q^n. \tag{22.2}$$

Пример 3. Сколькими способами можно выбрать а) по три карты, б) по 32 карты из колоды, содержащей 36 игральных карт?

Решение. Искомое число способов:

$$\text{а) } C_{36}^2 = \frac{36!}{2!\,34!} = \frac{35\,36}{2} = 630,$$

$$\text{б) } C_{36}^{32} = C_{36}^4 = \frac{36!}{4!\,32!} = 58905.$$

Виды случайных событий

Выше было введено определение случайного события. Обычно в теории вероятностей вместо "совокупности условий" употребляют термин "испытание", и тогда событие трактуется как результат испытания. Например, стрельба по мишени: выстрел — это испытание, попадание в мишень — это событие. Другой пример: подбрасывание монеты вверх — это испытание, выпадение орла (или решки) — это событие.

Определение 1. События называют *несовместными*, если в одном и том же испытании появление одного из них исключает

появление других. Например, выпадение орла при подбрасывании монеты исключает появление в этом же испытании решки и наоборот.

Определение 2. Несколько событий образуют *полную группу*, если в результате испытания появление хотя бы одного из них является достоверным событием. Например, при произведении выстрела по мишени (испытание) обязательно будет либо попадание, либо промах; эти два события образуют полную группу.

Следствие. *Если события, образующие полную группу, попарно несовместны, то в результате испытания появится одно и только одно из этих событий.*

Этот частный случай будет использован далее.

Классическое определение вероятности

Назовем каждый из возможных результатов испытания *элементарным событием*, или *исходом*. Те элементарные исходы, которые интересуют нас, называются *благоприятными* событиями.

Определение 3. Отношение числа благоприятствующих событию A элементарных исходов к общему числу равновозможных несовместных элементарных исходов, образующих полную группу, называется *вероятностью события* A.

Вероятность события A обозначается $P(A)$. Понятие вероятности является одним из основных в теории вероятностей. Данное выше определение является классическим. Из него вытекают некоторые свойства.

Свойство 1. Вероятность достоверного события равна единице.

Свойство 2. Вероятность невозможного события равна нулю.

Свойство 3. Вероятность случайного события есть положительное число:

$$0 < P(A) < 1.$$

Следовательно, вероятность любого события удовлетворяет неравенству

$$0 \leqslant P(A) \leqslant 1.$$

Отметим, что современные курсы теории вероятностей основаны на теоретико-множественном подходе, в котором элементарные события являются точками пространства элементарных событий Ω; при этом событие A отождествляется с подмножеством элементарных исходов, благоприятствующих этому событию, $A \subset \Omega$.

Приведем примеры непосредственного вычисления вероятностей.

Пример 4. В коробке лежит 10 шаров: 6 белых и 4 черных. Найти вероятность того, что из пяти взятых наугад шаров будет 4 белых.

Решение. Найдем число благоприятных исходов: число способов, которыми можно взять 4 белых шара из 6 имеющихся, равно $C_6^4 = C_6^2 = \dfrac{6!}{2!\,4!} = 15$. Общее число исходов определяется числом сочетаний из 10 по 5: $C_{10}^5 = 252$. Согласно определению 3 искомая вероятность $P = 15/252 \approx 0{,}06$.

Пример 5. Какова вероятность того, что при заполнении карточки спортивной лотереи "6 из 36" будет угадано 4 номера?

Решение. Общее число исходов равно $C_{36}^6 = 1947792$. Число благоприятных исходов равно $C_6^4 = 15$. Отсюда искомая вероятность равна $7{,}7 \cdot 10^{-6}$.

Пример 6. В ящике находится 10 стандартных и 5 нестандартных деталей. Какова вероятность, что среди наугад взятых 6 деталей будет 4 стандартных и 2 нестандартных?

Решение. Общее число исходов равно C_{15}^6. Число благоприятных исходов определяется произведением $C_{10}^4 C_5^2$, где первый сомножитель соответствует числу вариантов изъятия из ящика 4-х стандартных деталей из 10, а второй — числу вариантов изъятия из ящика 2-х нестандартных деталей из пяти. Отсюда следует, что искомая вероятность равна

$$P = C_{10}^4 C_5^2 / C_{15}^6 = 0{,}42.$$

22.2. Теорема сложения вероятностей

Несовместные события

Определение 1. *Суммой* двух событий A и B называют событие $C = A + B$, которое состоит в появлении либо события A, либо события B, либо событий A и B одновременно.

Это определение напоминает сумму множеств (см. гл. 1) и используется в теоретико-множественном подходе теории вероятностей. Примеры суммы событий: произведены два выстрела, и события A и B — попадания при первом и втором выстрелах соответственно; тогда $A + B$ — попадание либо при первом выстреле, либо при втором, либо в обоих выстрелах. Если события A и B несовместные, то их сумма — это событие, состоящее в появлении какого-либо из этих событий.

Аналогично определяется сумма *нескольких* событий, состоящая в появлении хотя бы одного из этих событий.

Теорема 22.1. *Вероятность появления какого-либо из нескольких попарно несовместных событий равна сумме их вероятностей:*

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n). \quad (22.3)$$

Следствие. *Вероятность появления какого-либо из двух несовместных событий равна сумме вероятностей этих событий:*

$$P(A + B) = P(A) + P(B).$$

Пример 1. Стрелок стреляет по мишени, разделенной на 4 концентрические зоны. Вероятности попадания в эти области соответственно равны 0,4, 0,3, 0,2 и 0,1. Найти вероятность попадания либо в первую, либо во вторую зоны.

Решение. Пусть событие A — попадание в первую зону мишени, а событие B — попадание во вторую зону мишени. Эти события несовместны, поэтому применимы теорема 22.1 и формула (22.3) сложения вероятностей. Искомая вероятность равна

$$P(A + B) = P(A) + P(B) = 0{,}4 + 0{,}3 = 0{,}7.$$

Полная группа событий

Теорема 22.2. *Сумма вероятностей событий, образующих полную группу, равна единице:*

$$P(A_1) + P(A_2) + \ldots + P(A_n) = 1.$$

Пример 2. На складе готовой продукции находятся изделия, среди которых 5% нестандартных. Найти вероятность того, что при выдаче изделия со склада оно будет стандартным.

Решение. Вероятность получения нестандартного изделия равна 0,05; события выдачи стандартного и нестандартного изделия образуют полную группу. Следовательно, сумма их вероятностей равна единице, и тогда искомая вероятность равна 0,95.

Противоположные события

Определение 2. Два единственно возможных события, образующих полную группу, называются *противоположными*.

Если событие обозначено через A, то противоположное ему событие обозначается через $\bar{A}$. Из теоремы 22.2 следует, что

$$P(A) + P(\bar{A}) = 1. \quad (22.4)$$

Например, если при стрельбе по мишени попадание — это событие A, то событие $\bar{A}$ — это промах; сумма их вероятностей равна единице — при выстреле обязательно будет либо попадание, либо промах. То же самое и при подбрасывании монеты: обязательно выпадет либо орел, либо решка.

Пример 3. В магазине имеется 10 телевизоров, из которых 2 неисправных. Найти вероятность того, что среди наугад взятых 3-х телевизоров будет хотя бы один неисправный.

Решение. События "среди взятых телевизоров нет ни одного неисправного" и "есть хотя бы один неисправный" — противоположные. Первое из них обозначим через A, а второе —

через $\bar{A}$. Общее число способов, которыми можно взять 3 изделия из десяти, равно C_{10}^3. Число исправных телевизоров равно 8, число способов выборки из них трех изделий равно C_8^3, так что вероятность $P(A) = C_8^3/C_{10}^3$. Искомая вероятность определяется из формулы (22.4):

$$P(\bar{A}) = 1 - P(A) = 1 - C_8^3/C_{10}^3 = 1 - 0{,}47 = 0{,}53.$$

22.3. Теорема умножения вероятностей

Произведение событий и условная вероятность

Определение 1. *Произведением* двух событий A и B называется событие AB, означающее совместное появление этих событий (см. гл. 1.1, произведение множеств).

Например, если событие A — шар, событие B — белый цвет, то их произведение AB — белый шар. Аналогично определяется произведение *нескольких* событий, как совместное появление их всех.

Если при вычислении вероятности события никаких других ограничений кроме необходимого комплекса условий S не налагается, то такая вероятность называется *безусловной*. Если же налагаются другие дополнительные условия, содержащие случайные события, то вероятность такого события называется *условной*.

Определение 2. Вероятность события B в предположении о наличии события A называют *условной вероятностью* $P_A(B)$.

Пример 1. В ящике лежит 11 деталей, 3 из них нестандартные. Из ящика дважды берут по одной детали, не возвращая их обратно. Найти вероятность того, что во второй раз из ящика будет извлечена стандартная деталь — событие B, если в первый раз была извлечена нестандартная деталь — событие A.

Решение. После первого извлечения в ящике из 10 деталей осталось 8 стандартных, и, следовательно, искомая вероятность

$$P_A(B) = 0{,}8.$$

Пусть теперь известны вероятность $P(A)$ события A и условная вероятность $P_A(B)$ события B. Тогда справедлива следующая теорема.

Теорема 22.3. *Вероятность произведения двух событий определяется формулой*

$$P(AB) = P(A)P_A(B). \tag{22.5}$$

Пример 2. В условиях примера 1 найти вероятности того, что в первый раз извлечена нестандартная деталь, а во второй раз — стандартная, и наоборот.

Решение. Итак, событие A — это извлечение из ящика нестандартной детали, а событие B — стандартной. Тогда возможны два случая. 1) Вероятность $P(A) = 3/11$, а условная вероятность $P_A(B) = 0{,}8$. Искомая вероятность произведения этих событий (их совместного появления в указанном порядке) равна, согласно теореме 22.3,

$$P(AB) = P(A)P_A(B) = (3/11)0{,}8 \approx 0{,}22.$$

2) Вероятность $P(B) = 8/11$, а условная вероятность $P_B(A) = 0{,}3$. Мы видим, что и в этом случае вероятность произведения событий $P(BA) = P(B)P_B(A) \approx 0{,}22$.

В этом примере мы проверили известное в теории равенство

$$P(A)P_A(B) = P(B)P_B(A).$$

Теорема 22.3 допускает обобщение на случай произведения любого числа событий A_1, A_2, A_3, ..., A_n:

$$P(A_1A_2A_3\dots A_n) = P(A_1)P_{A_1}(A_2)P_{A_1A_2}(A_3)\dots \\ \dots P_{A_1A_2\dots A_{n-1}}(A_n), \tag{22.6}$$

т.е. вероятность совместного появления n событий равна произведению n вероятностей, где $P_{A_1A_2\dots A_{k-1}}(A_k)$ — условные вероятности событий A_k в предположении, что события $A_1A_2\dots A_{k-1}$ уже произошли ($k = 1, 2, \dots, n$).

Пример 3. В урне находится 4 белых шара, 5 красных и 3 синих. Наудачу извлекают по одному шару, не возвращая его

обратно. Найти вероятность того, что в первый раз появится белый шар (событие A), во второй раз — красный (событие B), в третий — синий (событие C).

Решение. Вероятность появления белого шара в первом извлечении $P(A) = 1/3$; условная вероятность появления красного шара во втором извлечении при условии появления в первый раз белого шара $P_A(B) = 5/11$; условная вероятность появления синего шара в третьем извлечении при условиях появления в предыдущих извлечениях белого и красного шаров $P_{AB}(C) = 0{,}3$. Искомая вероятность определяется по формуле (22.6) при $n = 3$:

$$P(ABC) = P(A)P_A(B)P_{AB}(C) = (1/3)(5/11)0{,}3 \approx 0{,}045.$$

Независимые события

Определение 3. Событие B называется *независимым* от события A, если условная вероятность события B равна его безусловной вероятности (появление события A не влияет на вероятность события B):

$$P_A(B) = P(B).$$

Отсюда следует, что и событие A также независимо от события B:

$$P_B(A) = P(A).$$

Для независимых событий теорема умножения вероятностей 22.3 в общей форме, которая следует из (22.6), имеет вид

$$P(A_1A_2A_3\ldots A_n)=P(A_1)P(A_2)P(A_3)\ldots P(A_n),\ n>1. \quad (22.7)$$

Равенство (22.7) принимается за определение независимых событий. При этом если события независимы, то независимы также и соответствующие им противоположные события.

Пример 4. Найти вероятность поражения цели при совместной стрельбе тремя орудиями, если вероятности поражения цели орудиями соответственно равны 0,9, 0,8 и 0,7 (события A, B и C).

Решение. Поскольку события A, B и C являются независимыми, то искомая вероятность вычисляется, согласно формуле (22.7), при $n = 3$:

$$P(ABC) = P(A)P(B)P(C) = 0{,}9 \cdot 0{,}8 \cdot 0{,}7 = 0{,}504.$$

Когда в результате испытания может иметь место n независимых событий с известными вероятностями их появления, особый интерес представляет случай нахождения вероятности наступления хотя бы одного из них (например, в случае трех событий найти вероятность наступления либо одного, либо двух, либо трех событий). Обозначим это событие через A. Справедлива следующая теорема.

Теорема 22.4. *Вероятность появления хотя бы одного из независимых событий A_1, A_2, ..., A_n определяется формулой*

$$P(A) = 1 - q_1q_2\ldots q_n, \quad (22.8)$$

где $q_i = 1 - p_i$ — вероятности соответствующих противоположных событий $\bar{A}_i$ $(i = 1, 2, \ldots, n)$.

В частном случае, когда все события A_i имеют одинаковую вероятность p, из формулы (22.8) следует, что

$$P(A) = 1 - q^n, \quad q = 1 - p. \quad (22.9)$$

Пример 5. В условиях примера 4 найти вероятность поражения цели (хотя бы одного попадания) при залповой стрельбе орудий.

Решение. Вероятности противоположных событий (промахов) соответственно равны $q_1 = 0{,}1$, $q_2 = 0{,}2$, $q_3 = 0{,}3$. Искомая вероятность находится по формуле (22.8) при $n = 3$:

$$P(A) = 1 - q_1q_2q_3 = 1 - 0{,}006 = 0{,}994.$$

Из этого примера наглядно видно преимущество совместного воздействия случайных событий с целью достижения общего результата.

Пример 6. На перевозку груза направлены 4 автомобиля. Вероятность нахождения каждой из машин в исправном состоянии равна 0,8. Найти вероятность того, что в работе участвует хотя бы один из выделенных для этого автомобилей.

Решение. Вероятность противоположного события (машина неисправна) равна $q = 1 - 0{,}8 = 0{,}2$. По формуле (22.9) находим искомую вероятность при $n = 4$:

$$P(A) = 1 - q^4 = 1 - 0{,}2^4 = 0{,}9984.$$

Пример 7. Вероятность обслуживания клиента одним операционистом в банке равна 0,6. Какое минимальное число операционистов должно работать в банке, чтобы вероятность обслуживания клиента была не менее 0,95?

Решение. Вероятность противоположного события (отказ в обслуживании клиента операционистом) равна 0,4. Пусть n — количество операционистов, удовлетворяющее условию задачи, т.е.

$$P(A) = 1 - q^n = 1 - 0{,}4^n \geqslant 0{,}95.$$

Решая это неравенство, получаем

$$0{,}4^n \leqslant 0{,}05.$$

Логарифмирование обеих частей этого неравенства дает

$$n \geqslant \lg 0{,}05 / \lg 0{,}4 = 3{,}27.$$

Поскольку n должно быть целым числом, окончательно получаем, что в банке должны работать не менее 4 операционистов.

22.4. Обобщения теорем сложения и умножения

Появление только одного из независимых событий

Рассмотрим примеры совместного применения теорем сложения и умножения. Пусть два независимых события A_1 и A_2 имеют вероятности появления соответственно p_1 и p_2. Найдем вероятность появления только одного из этих событий. Для этого введем новые события: B_1 и B_2. Событие B_1 состоит в том, что событие A_1 наступило, а событие A_2 не наступило; иными словами, $B_1 = A_1\bar{A}_2$. Аналогичным образом определяется и событие $B_2 = \bar{A}_1 A_2$ (совместное ненаступление события A_1 и наступление события A_2). Поскольку события A_1 и A_2

независимы, то независимы также и противоположные события $\bar{A}_1$ и $\bar{A}_2$; тогда события B_1 и B_2 являются несовместными. Вероятность наступления только одного из событий A_1 и A_2 находится как сумма вероятностей несовместных событий B_1 и B_2:

$$\begin{aligned} P(B_1 + B_2) &= P(B_1) + P(B_2) = P(A_1\bar{A}_2) + P(\bar{A}_1 A_2) = \\ &= P(A_1)P(\bar{A}_2) + P(\bar{A}_1)P(A_2) = p_1 q_2 + p_2 q_1, \end{aligned} \tag{22.10}$$

где $q_1 = 1 - p_1$, $q_2 = 1 - p_2$.

Аналогичным образом можно убедиться в справедливости формулы вероятности наступления только одного из трех независимых событий A_1, A_2, A_3 с вероятностями наступления соответственно p_1, p_2 и p_3:

$$P(B_1 + B_2 + B_3) = p_1 q_2 q_3 + p_2 q_1 q_3 + p_3 q_1 q_2, \tag{22.11}$$

где B_i — произведения наступившего события A_i и двух других ненаступивших событий ($i = 1, 2, 3$). Для случая n независимых событий формула вероятности наступления только одного из них имеет аналогичный вид — сумма n слагаемых, каждый член которой представляет собой произведение вероятности наступления одного из событий на вероятности $(n-1)$ других противоположных событий.

Пример 1. Предприниматель решил вложить свои средства поровну в два контракта, каждый из которых принесет ему прибыль в размере 100%. Вероятность того, что любой из контрактов не "лопнет", равна 0,8. Какова вероятность того, что по истечении контрактов предприниматель по меньшей мере ничего не потеряет?

Решение. Предприниматель по крайней мере ничего не потеряет, если либо не "лопнет" один из контрактов (другой возместит ему потери), либо будут выполнены оба контракта. Пусть события A_1 и A_2 — это выполнение соответствующих контрактов (вероятность $p = 0{,}8$); эти события являются независимыми. Противоположные им события $\bar{A}_1$ и $\bar{A}_2$ — невыполнение контрактов (вероятность $q = 0{,}2$). Тогда собы-

тия $B_1 = A_1\bar{A}_2$, $B_2 = \bar{A}_1 A_2$ и $A_1 A_2$ являются несовместными (последнее событие — это выполнение обоих контрактов). Искомая вероятность определяется с учетом формул (22.10) и (22.7):

$$P(B_1 + B_2 + A_1A_2) = p_1q_2 + p_2q_1 + p_1p_2 = 2pq + p^2 = 0{,}96.$$

Теорема сложения вероятностей совместных событий

Определение 1. События A и B называют совместными, если в одном и том же испытании появление одного из них не исключает появления другого.

Для таких событий справедлива следующая теорема.

Теорема 22.5. *Вероятность суммы совместных событий равна сумме их вероятностей без вероятности их произведения:*

$$P(A + B) = P(A) + P(B) - P(AB). \qquad (22.12)$$

Из формулы (22.12) получается ряд следующих частных случаев.

1. Для независимых событий с учетом формулы (22.7)

$$P(A + B) = P(A) + P(B) - P(A)P(B). \qquad (22.13)$$

2. Для зависимых событий с учетом формулы (22.5)

$$P(A + B) = P(A) + P(B) - P(A)P_A(B).$$

3. Для несовместных событий $P(AB) = 0$, и в этом случае имеем подтверждение теоремы 22.1 и формулы (22.3):

$$P(A + B) = P(A) + P(B).$$

Пример 2. Вероятности поражения цели первым и вторым стрелками равны соответственно 0,8 и 0,9. Найти вероятность поражения цели при залпе.

Решение. Поскольку вероятности поражения цели стрелками (события A и B соответственно) не зависят от результатов стрельбы каждого из напарников, то эти события независимы. Искомая вероятность рассчитывается по формуле (22.13):

$$P(A + B) = P(A) + P(B) - P(A)P(B) = \\ = 0{,}8 + 0{,}9 - 0{,}72 = 0{,}98.$$

Аналогичный результат можно было бы получить и с применением формулы (22.8). Пусть событие A — поражение цели, $\bar{A}$ и $\bar{B}$ — события, соответствующие промахам стрелков, тогда

$$P(A) = 1 - P(\bar{A})P(\bar{B}) = 1 - 0{,}2 \cdot 0{,}1 = 0{,}98.$$

Формула полной вероятности

Пусть события $B_1, B_2, \ldots, B_n$ несовместны и образуют полную группу, т.е., согласно теореме 22.2, выполняется равенство

$$P(B_1) + P(B_2) + \ldots + P(B_n) = 1.$$

Пусть также событие A может наступить при условии появления одного из событий B_i, причем известны как вероятности $P(B_i)$, так и условные вероятности $P_{B_i}(A)$ $(i = 1, 2, \ldots, n)$. В таком случае формула для вероятности события A определяется следующей теоремой.

Теорема 22.6. *Вероятность события A, появление которого возможно лишь при наступлении одного из несовместных событий B_i, образующих полную группу $(i = 1, 2, \ldots, n)$, равна сумме попарных произведений каждого из этих событий на соответствующую условную вероятность появления события A:*

$$P(A) = P(B_1)P_{B_1}(A) + \\ + P(B_2)P_{B_2}(A) + \ldots + P(B_n)P_{B_n}(A). \qquad (22.14)$$

Пример 3. В двух урнах находятся белые и красные шары: в первой — 4 белых и 5 красных, во второй — 7 белых и 3 красных. Из второй урны наудачу взяли шар и переложили его

в первую урну. Найти вероятность того, что наудачу взятый после этого из первой урны шар будет белым.

Решение. Перекладывание из второй урны в первую белого шара (событие B_1) и красного шара (событие B_2) образует полную группу независимых событий. Их вероятности соответственно $P(B_1) = 0{,}7$ и $P(B_2) = 0{,}3$. Условные вероятности извлечения из первой урны белого шара (событие A) при добавлении туда белого или красного шара из второй урны соответственно равны $P_{B_1}(A) = 0{,}5$ и $P_{B_2}(A) = 0{,}4$. Искомая вероятность находится по формуле (22.14) при $n = 2$:

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) =$$

$$= 0{,}7 \cdot 0{,}5 + 0{,}3 \cdot 0{,}4 = 0{,}47.$$

Пример 4. В двух ящиках находятся детали: в первом — 10 штук и из них 3 нестандартные, а во втором — 20 штук и из них 8 нестандартных. Из каждого ящика наудачу вынуто по одной детали, а потом из этих двух деталей наудачу взята одна. Найти вероятность того, что эта деталь окажется стандартной.

Решение. При первой выборке двух деталей возможны четыре случая, которые образуют полную группу независимых событий. События B_{ss}, B_{sn}, B_{ns}, B_{nn} соответствуют случаям изъятия: из первого и второго ящиков по стандартной детали, из первого ящика — стандартной и из второго — нестандартной деталей, из первого ящика — нестандартной и из второго — стандартной, из первого и второго ящиков по нестандартной детали. В свою очередь события B_{ik} $(i, k = s, n)$ представляют собой произведения независимых событий — изъятия из каждого ящика по детали, и потому их вероятности равны соответствующим произведениям вероятностей этих изъятий: $P(B_{ss}) = 0{,}7 \cdot 0{,}6 = 0{,}42$; $P(B_{sn}) = 0{,}7 \cdot 0{,}4 = 0{,}28$; $P(B_{ns}) = 0{,}3 \cdot 0{,}6 = 0{,}18$; $P(B_{nn}) = 0{,}3 \cdot 0{,}4 = 0{,}12$. Условные вероятности выборки из двух деталей стандартной, согласно перечисленным выше возможным случаям, равны:

$$P_{B_{ss}}(A) = 1, \quad P_{B_{sn}}(A) = 0{,}5,$$

$$P_{B_{ns}}(A) = 0{,}5, \quad P_{B_{nn}}(A) = 0.$$

Теперь, согласно теореме 22.6 и формуле (22.14), получаем искомую вероятность события A:

$$P(A) = P(B_{ss})P_{B_{ss}}(A) + P(B_{sn})P_{B_{sn}}(A) +$$
$$+P(B_{ns})P_{B_{ns}}(A) + P(B_{nn})P_{B_{nn}}(A) =$$
$$= 0{,}42 \cdot 1 + 0{,}28 \cdot 0{,}5 + 0{,}18 \cdot 0{,}5 + 0{,}12 \cdot 0 = 0{,}65.$$

Формулы Байеса

Пусть события B_1, B_2, ..., B_n несовместны и образуют полную группу, а событие A может наступить при условии появления одного из них. События B_i называют *гипотезами*, так как заранее неизвестно, какое из них наступит. Пусть произведено испытание и в результате появилось событие A. Тогда оказывается возможным определить условные вероятности гипотез B_i по следующим формулам:

$$P_A(B_i) = \frac{P(B_i)P_{B_i}(A)}{P(B_1)P_{B_1}(A)+P(B_2)P_{B_2}(A)+\ldots+P(B_n)P_{B_n}(A)} =$$
$$= P(B_i)P_{B_i}(A)/\sum_{k=1}^{n} P(B_k)P_{B_k}(A) = P(B_i)P_{B_i}(A)/P(A), \tag{22.15}$$
$$i = 1, 2, \ldots, n.$$

Формулы (22.15) называются *формулами Байеса*, по имени их автора. Они позволяют оценить вероятность гипотезы B_i во всех испытаниях, где наступает событие A. Иными словами, зная вероятность $P(B_i)$ до проведения испытания, мы можем переоценить ее после проведения испытания, в результате которого появилось событие A.

Пример 5. Вероятность изготовления изделия с браком равна 0,08. После изготовления все изделия подвергаются проверке, в результате которой изделия без брака признаются годными с вероятностью 0,95, а изделия с браком — с вероятностью 0,06. Найти долю изделий, выпущенных после проверки, а также вероятность того, что выпущенное после проверки изделие окажется без брака.

36 Зак. 6853

Решение. Независимые события (гипотезы), образующие полную группу, — это B_1 (изделие без брака) и B_2 (изделие с браком). Пусть событие A заключается в том, что при проверке изделие признается годным. Ответ на первый вопрос задачи дает формула (22.14):

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) = \\ = 0{,}92 \cdot 0{,}95 + 0{,}08 \cdot 0{,}06 = 0{,}8788.$$

Следовательно, после проверки признаются годными около 88% всех изготовленных изделий.

Ответ на второй вопрос задачи дает формула Байеса (22.15) при $n = 2$ и $i = 1$:

$$P_A(B_1) = \frac{P(B_1)P_{B_1}(A)}{P(A)} = \frac{0{,}92 \cdot 0{,}95}{0{,}8788} = 0{,}995.$$

Иными словами, среди изделий, прошедших проверку, содержится 99,5% изделий без брака.

Пример 6. В среднем из каждых 100 клиентов отделения банка 60 обслуживаются первым операционистом и 40 — вторым операционистом. Вероятность того, что клиент будет обслужен без помощи заведующего отделением, только самим операционистом, составляет 0,9 и 0,75 соответственно для первого и второго служащих банка. Найти вероятность полного обслуживания клиента первым операционистом.

Решение. Вероятность того, что клиент попадет к первому операционисту (событие B_1), составляет 0,6, ко второму — 0,4 (событие B_2). Искомая вероятность полного обслуживания клиента первым операционистом (событие A) определяется по формулам (22.14) и (22.15):

$$P_A(B_1) = \frac{P(B_1)P_{B_1}(A)}{P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A)} = \\ = \frac{0{,}6 \cdot 0{,}9}{0{,}6 \cdot 0{,}9 + 0{,}4 \cdot 0{,}75} = 0{,}64.$$

Иными словами, 64% клиентов, попавших на обслуживание к первому операционисту, будут обслужены им полностью.

22.5. Схема независимых испытаний

Формула Бернулли

Определение 1. Если при проведении нескольких испытаний вероятность события A в каждом испытании не зависит от исходов других событий, то эти испытания называются *независимыми относительно события* A.

Будем рассматривать только такие независимые испытания, в которых событие A имеет одинаковую вероятность. Пусть производится n независимых испытаний, в каждом из которых событие A может появиться с вероятностью p. Тогда вероятность противоположного события — ненаступления события A — также постоянна в каждом испытании и равна $q = 1 - p$. В теории вероятностей представляет особый интерес случай, когда в n испытаниях событие A осуществится k раз и не осуществится $n - k$ раз.

Вероятность этого сложного события, состоящего из n испытаний, определяется *формулой Бернулли*

$$P_n(k) = C_n^k p^k q^{n-k}, \text{ или } P_n(k) = \frac{n!}{k!\,(n-k)!} p^k q^{n-k}. \qquad (22.16)$$

Пример 1. Монету бросают 6 раз. Найти вероятности того, что герб выпадет: 1) 2 раза, 2) не менее двух раз.

Решение. Вероятности выпадения любой из двух сторон монеты одинаковы, т.е. $p = q = 0{,}5$. 1) В этом случае $n = 6$, $k = 2$. Отсюда согласно формуле (22.16) получаем

$$P_6(2) = C_6^2(0{,}5)^2(0{,}5)^4 = 15(0{,}5)^6 = 0{,}23.$$

2) $P_6(2 \leqslant k \leqslant 6) = 1 - P_6(0) - P_6(1) =$

$= 1 - C_6^0(0{,}5)^6 - C_6^1(0{,}5)^6 = 1 - (0{,}5)^6(1 + 6) = 0{,}89.$

Пример 2. Вероятность покупки бракованного комплекта посуды равна 0,1. Найти вероятность того, что из 7 купленных комплектов 5 будет без брака.

Решение. Вероятность покупки комплекта без брака $p = 0{,}9$, $q = 0{,}1$ — это дано по условию задачи. Тогда искомая вероятность находится по формуле (22.16):

$$P_7^5 = C_7^5 p^5 q^2 = C_7^2(0{,}9)^5(0{,}1)^2 = 0{,}12.$$

Пример 3. Контрольный тест состоит из 4 вопросов. На каждый вопрос предлагается 4 варианта ответа, среди которых только один правильный. Найти вероятность правильного ответа на два, три и четыре вопроса теста для неподготовленного человека (выбор ответа наудачу).

Решение. Искомые значения вероятности находятся по формуле Бернулли (22.16) с учетом того, что вероятность события A (правильный ответ) в каждом испытании (выбор ответа на вопрос теста) равна 0,25, а $q = 0{,}75$. Отсюда получаем:

$$P_4(2) = C_4^2(0{,}25)^2(0{,}75)^2 = 0{,}21;$$

$$P_4(3) = C_4^3(0{,}25)^3(0{,}75)^1 = C_4^1(0{,}25)^3(0{,}75) = 0{,}047;$$

$$P_4(4) = C_4^4(0{,}25)^4(0{,}75)^0 = C_4^0(0{,}25)^4 = (0{,}25)^4 = 0{,}004.$$

Локальная теорема Лапласа

Использование формулы Бернулли (22.16) при больших значениях n и k представляется затруднительным ввиду увеличения объема вычислений и операций с большими числами. В этом случае применима формула, устанавливаемая следующей *локальной теоремой Лапласа.*

Теорема 22.7. *Пусть вероятность p появления события A в каждом испытании постоянна, причем $0 < p < 1$. Тогда вероятность $P_n(k)$ того, что событие A появится в n испытаниях ровно k раз, приближенно равна значению функции $\varphi(x)$:*

$$y = \frac{1}{\sqrt{npq}}\varphi(x) = \frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \qquad (22.17)$$

где $x = (k - np)/\sqrt{npq}$.

Точность формулы (22.17) возрастает с увеличением n. Имеются таблицы с вычисленными значениями функции $\varphi(x)$ (см. Приложение), по которым можно с достаточно высокой степенью точности найти практически любое значение этой функции, а значит, и вычислить нужную вероятность. Поскольку функция $\varphi(x)$ четная,то в таблицах даются ее значения только для положительных значений x; иными словами, знак аргумента не играет роли. Формула (22.17) носит название *асимптотической* формулы.

Пример 4. Вероятность выпуска бракованного изделия равна 0,3. Найти вероятность того, что среди 100 выпущенных изделий будет ровно 60 изделий без брака.

Решение. Вероятность появления события A в одном испытании (изделие без брака) $p = 0{,}7$, тогда $q = 0{,}3$; в нашем случае $n = 100$, $k = 60$. Последовательно вычисляем:

$$\sqrt{npq} = \sqrt{100 \cdot 0{,}7 \cdot 0{,}3} = 4{,}58,$$
$$|x| = |k - np|/\sqrt{npq} = 10/4{,}58 = 2{,}18.$$

Теперь для найденного аргумента x находим по табл. 1 (см. Приложение) соответствующее значение $\varphi(x)$; оно равно 0,0371. Подстановка этого числа в формулу (22.17) дает приближенное значение искомой вероятности:

$$P_{100}(60) = 0{,}0371/4{,}58 = 0{,}008.$$

Интегральная теорема Лапласа

Опять предположим, что в каждом из произведенных n испытаний событие A появляется с одинаковой вероятностью p. В прикладных вопросах теории вероятностей наиболее употребимы определения вероятности события A в n испытаниях, когда k изменяется в заданном интервале значений: $l < k < m$. Соответствующую вероятность обозначают $P_n(l, m)$. Формула для приближенного вычисления этой вероятности устанавливается следующей *интегральной теоремой Лапласа.*

Теорема 22.8. *Пусть вероятность p наступления события A в каждом испытании постоянна, причем $0 < p < 1$. Тогда вероятность того, что событие A появится в n испытаниях от l до m раз, приближенно равна определенному интегралу:*

$$P_n(l, m) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-z^2/2}dz, \qquad (22.18)$$

где $a = (l - np)/\sqrt{npq}$, $b = (m - np)/\sqrt{npq}$.

Формула (22.18), как и (22.17), применима в случае больших значений n и k. При вычислениях по этой формуле пользуются специальными таблицами для интеграла

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-z^2/2}\, dz,$$

поскольку соответствующий неопределенный интеграл не выражается через элементарные функции. Функцию $\Phi(x)$ часто называют *интегралом ошибок*, соответствующая таблица ее значений приведена в Приложении. Эта функция является нечетной, поэтому в таблицах обычно приводят значения $\Phi(x)$ для положительных значений верхнего предела интегрирования x. Более удобно использовать формулу (22.18) в виде формулы Ньютона–Лейбница:

$$\begin{gathered} P_n(l,m) = \Phi(b) - \Phi(a), \\ a = (l - np)/\sqrt{npq}, \quad b = (m - np)/\sqrt{npq}. \end{gathered} \tag{22.19}$$

Пример 5. Вероятность выпуска бракованных деталей равна 0,3. Найти вероятность того, что среди 100 выпущенных деталей будет не менее 75 стандартных.

Решение. По условию задачи $p = 0{,}7$, $q = 0{,}3$, $n = 100$. Условие "не менее" означает, что число стандартных деталей k заключено в пределах от $l = 75$ до $m = 100$. Согласно формуле (22.19) производим предварительные вычисления:

$$\begin{gathered} a = (l - np)/\sqrt{npq} = 5/\sqrt{21} = 1{,}09, \\ b = (m - np)/\sqrt{npq} = 6{,}55. \end{gathered}$$

Далее по табл. 2 Приложения находим соответствующие значения интегральной функции $\Phi(x)$, подставляем их в формулу (22.19) и получаем значение искомой вероятности:

$$P_n(l,m) = \Phi(b) - \Phi(a) = 0{,}5 - 0{,}36 = 0{,}24.$$

Пример 6. В страховой компании 10 тыс. клиентов, застраховавших свою недвижимость. Страховой взнос составляет 2000 р., вероятность несчастного случая $p = 0{,}005$, страховая выплата клиенту при несчастном случае составляет 200 тыс. р. Определить размер прибыли страховой компании с вероятностью P: 1) 0,9, 2) 0,995.

Решение. Прибыль компании зависит от числа страховых выплат k при несчастных случаях. Будем полагать, что величина ее равна разности между суммами страховых взносов и страховых выплат:

$$R = (20 - 0{,}2k) \text{ млн р.}$$

Теперь задача состоит в нахождении такого числа N, чтобы вероятность несчастного случая $P_{10\,000}(k > N)$ была не больше заданной величины $1 - P$, или, что то же самое, чтобы выполнялось условие

$$P_{10\,000}(N < k < 10\,000) \leqslant 1 - P.$$

Тогда с вероятностью P прибыль компании составит $(20 - 0{,}2N)$ млн р. Предварительные вычисления значений аргумента функции $\Phi(x)$ при $n = 10\,000$, $l = N$ и $m = 10\,000$ по формулам (22.19) дают:

$$a = (N - np)/\sqrt{npq} = \frac{N - 50}{7{,}05}, \quad b = 9550/7{,}05 = 1411{,}34.$$

Из табл. 2 находим, что $\Phi(x) = 0{,}5$ при $|x| > 5$. Подставляя в указанное выше неравенство, получаем

$$0{,}5 - \Phi\left(\frac{N - 50}{7{,}05}\right) \leqslant 1 - P.$$

1. В этом случае имеем неравенство

$$0{,}5 - \Phi\left(\frac{N - 50}{7{,}05}\right) \leqslant 0{,}1, \text{ или } \Phi\left(\frac{N - 50}{7{,}05}\right) \geqslant 0{,}4.$$

По табл. 2 находим, что при значении функции $\Phi = 0{,}4$ аргумент x равен 1,28; поскольку функция $\Phi(x)$ является монотонно возрастающей, то неравенство между значениями $\Phi(x)$ переходит в неравенство такого же смысла и для соответствующих аргументов:

$$\frac{N - 50}{7{,}05} \geqslant 1{,}28.$$

Отсюда получаем, что $N \geqslant 50 + 9{,}02$, или $N \geqslant 60$. В этом случае с вероятностью 0,9 страховой компании гарантирована прибыль

$$R = 20 - 0{,}2 \cdot 60 = 8 \text{ млн р.}$$

2. Проводя для этого случая аналогичные вычисления, получим

$$\Phi\left(\frac{N-50}{7{,}05}\right) \geqslant 0{,}495.$$

Из табл. 2 находим, что при $\Phi(x) = 0{,}495$ аргумент $x = 2{,}57$, т.е.

$$\frac{N-50}{7{,}05} \geqslant 2{,}57.$$

Из последнего неравенства получаем $N \geqslant 69$, и в этом случае с вероятностью 0,995 компании гарантирована прибыль

$$R = 20 - 0{,}2 \cdot 69 = 6{,}2 \text{ млн р.}$$

Из решенной задачи хорошо видно, что увеличение риска страхования может привести к возрастанию прибыли компании. Это есть реализация известного принципа в предпринимательской деятельности: менее рискованные, но более надежные финансовые операции не приносят сверхприбылей.

Отклонение относительной частоты от постоянной вероятности

Определение 2. Отношение числа испытаний, в которых событие A появилось, к общему числу фактически проведенных испытаний называют *относительной частотой события*

$$W(A) = m/n,$$

где m — число появлений события A, n — общее число испытаний.

Разница между вероятностью и относительной частотой состоит в том, что первая вычисляется до опыта, а вторая — после него. Одной из важных характеристик независимых испытаний с постоянной вероятностью появления события A в каждом испытании $(0 < p < 1)$ является отклонение относительной частоты от вероятности p. Вероятность того, что это отклонение не превышает по модулю заданного числа ε:

$$|m/n - p| \leqslant \varepsilon,$$

определяется формулой

$$P(|m/n - p| \leqslant \varepsilon) \cong 2\Phi(\varepsilon\sqrt{n/(pq)}). \qquad (22.20)$$

Пример 7. Вероятность получения нестандартной детали $p = 0{,}1$. Найти вероятность того, что среди случайно взятых 200 деталей относительная частота появления нестандартной детали отклонится от вероятности p по абсолютной величине не более чем на 0,03.

Решение. В данном случае $n = 200$, $q = 0{,}9$, $\varepsilon = 0{,}03$. По формуле (22.20) имеем

$$P(|m/n - p| \leqslant \varepsilon) \cong 2\Phi(\varepsilon\sqrt{n/(pq)}) =$$
$$= 2\Phi(0{,}03 \cdot \sqrt{2222}) = 2\Phi(1{,}41) \approx 0{,}84.$$

Смысл полученного результата состоит в том, что при достаточно большом числе проб, каждая из которых содержит 200 случайно выбранных деталей, в 84% случаев отклонение относительной частоты от постоянной вероятности $p = 0{,}1$ по абсолютной величине не превысит 0,03.

Глава 23
СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

23.1. Случайные величины и законы их распределения

Виды случайных величин

В главе 22 рассматривались события, состоящие в появлении того или иного числа. Например, среди трех изъятых деталей может оказаться до трех стандартных.

Определение 1. Величину называют *случайной*, если в результате испытания она примет лишь одно возможное значение, заранее не известное и зависящее от случайных причин.

Каждой случайной величине соответствует множество чисел — это множество значений, которые она может принимать. Например, число мальчиков среди 100 новорожденных — это случайная величина, которая может принимать значения от 0

до 100. Далее будем обозначать случайные величины прописными буквами, а их возможные значения — строчными буквами; например, случайная величина X имеет два возможных значения x_1 и x_2. Другой пример: случайная величина Y принимает возможные значения, принадлежащие интервалу (a, b). Различают два вида случайных величин.

Определение 2. Случайная величина, принимающая отдельные возможные значения с определенными вероятностями, называется *дискретной случайной величиной*.

Определение 3. *Непрерывной* называется случайная величина, которая может принимать все значения из некоторого промежутка.

Как следует из определения 2, для задания дискретной случайной величины нужно задать не только перечень ее возможных значений, но и их вероятности. Иными словами, каждому возможному значению случайной величины соответствует определенное значение вероятности появления этой величины.

Дискретные случайные величины

Определение 4. Соответствие между отдельными возможными значениями и их вероятностями называется *законом распределения дискретной случайной величины*.

Как и в случае функциональной зависимости, этот закон можно задать таблицей, аналитически (формулой) и графически. В случае табличного задания закона распределения дискретной случайной величины соответствующая таблица состоит из двух строк — первая указывает возможные значения, а вторая — их вероятности:

$$\begin{array}{cccccc} X & x_1 & x_2 & \dots & x_n \\ P & p_1 & p_2 & \dots & p_n. \end{array} \qquad (23.1)$$

Поскольку в одном испытании случайная величина принимает только одно возможное значение, то события $X{=}x_1$, $X{=}x_2$, ..., $X{=}x_n$ образуют полную группу, т.е. сумма их вероятностей равна единице:

$$p_1 + p_2 + \dots + p_n = 1.$$

Если множество возможных значений X дискретной случайной величины бесконечно, то соответствующий ряд вероятностей сходится и его сумма равна единице:

$$p_1 + p_2 + \dots + p_n + \dots = 1.$$

Пример 1. В денежной лотерее на 100 билетов разыгрывается один выигрыш в 20 р., два выигрыша по 10 р. и 10 выигрышей по 1 р. Найти закон распределения случайной величины X возможного выигрыша на один билет.

Решение. Возможные значения X: $x_1{=}20$, $x_2{=}10$, $x_3{=}1$, $x_4{=}0$. Соответственно их вероятности равны: $p_1{=}0{,}01$, $p_2{=}0{,}02$, $p_3{=}0{,}1$, $p_4{=}1-(p_1+p_2+p_3){=}1-0{,}13=0{,}87$. Таким образом, искомый закон распределения имеет вид

X	20	10	1	0
P	0,01	0,02	0,1	0,87.

Пример 2. Партия из 8 изделий содержит 5 стандартных. Наудачу отбираются 3 изделия. Составить таблицу закона распределения числа стандартных изделий среди отобранных.

Решение. Случайная величина X — число стандартных деталей среди отобранных — может принимать 4 возможных значения: 0, 1, 2 и 3. Вероятность нахождения k стандартных изделий среди трех отобранных определяется формулой

$$P(X=k) = C_5^k C_3^{3-k} / C_8^3, \quad k = 0, 1, 2, 3.$$

Варьируя значения k от 0 до 3, получаем искомое распределение:

X	0	1	2	3
P	0,0179	0,2679	0,5357	0,1785.

Пример 3. Вероятностный прогноз для величины X — процентного изменения стоимости акций по отношению к их текущему курсу в течение шести месяцев — дан в виде закона распределения:

X	5	10	15	20	25	30
P	0,1	0,1	0,2	0,3	0,2	0,1.

Найти вероятность того, что покупка акций будет более выгодна, чем помещение денег на банковский депозит под 36% годовых.

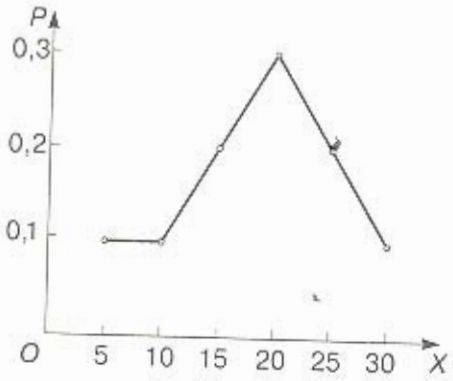

Рис. 23.1

Решение. Прирост суммы на банковском депозите при условии 3% в месяц составит через 6 месяцев $[(1{,}03)^6-1]100\% = 19{,}4\%$. Вероятность того, что покупка акций выгоднее банковского депозита, определяется суммой вероятностей, соответствующих более высокому росту курса акций:

$$P(X > 19{,}4) = p_4 + p_5 + p_6 = 0{,}3 + 0{,}2 + 0{,}1 = 0{,}6.$$

Закон распределения дискретной случайной величины можно изобразить графически, соединив в прямоугольной системе координат XOP точки (x_i, p_i) отрезками прямых. Так, на рис. 23.1 показан закон распределения из примера 3. Такая фигура называется *многоугольником распределения*.

Биномиальное распределение

Пусть производится n независимых испытаний и в каждом из них событие A может либо появиться, либо не появиться. Пусть также вероятность p появления события A в каждом испытании постоянна (см. раздел 22.5). В качестве дискретной случайной величины X рассмотрим число появления события A в этих n испытаниях. Очевидно, что $x_1{=}0$, $x_2{=}1$, $x_3{=}2$, ..., $x_{n+1}{=}n$. Вероятности этих возможных значений k даются формулой Бернулли (см. формулу (22.16)):

$$P_n(k) = C_n^k p^k q^{n-k}, \qquad (23.2)$$

где $q = 1-p$ — вероятность противоположного события (непоявление события A в одном испытании). Формула (23.2) представляет собой аналитическую форму закона распределения случайной величины (числа появления события A в n независимых испытаниях), которое называется *биномиальным*. Этот закон получил свое название потому, что правая часть в (23.2) представляет собой общий член разложения бинома Ньютона (22.2). Таким образом, табличная форма биномиального закона с учетом формулы (23.2) имеет вид

$$\begin{array}{cccccccc} X & n & n-1 & \ldots & k & \ldots & 0 \\ P & p^n & np^{n-1}q & \ldots & C_n^k p^k q^{n-k} & \ldots & q^n. \end{array} \qquad (23.3)$$

Можно показать, что сумма всех вероятностей второй строки этой таблицы равна единице, т.е.

$$\sum_{k=0}^{n} C_n^k p^k q^{n-k} = p^n + np^{n-1}q + \ldots + C_n^k p^k q^{n-k} + \ldots + q^n = 1.$$

Пример 4. Банк выдает 5 кредитов. Вероятность невозврата кредита равна 0,2 для каждого из заемщиков. Составить таблицу закона распределения количества заемщиков, не вернувших кредит по окончании срока кредитования.

Решение. Примем за A событие невозврата кредита. Так как заемщики действуют независимо, то выдачу 5 кредитов можно считать за 5 независимых событий. Вероятность невозврата k кредитов из 5 описывается биномиальным распределением (23.2), где $p = 0{,}2$, $q = 0{,}8$, k принимает значения от нуля до 5. Искомая таблица закона распределения составляется, согласно (23.3), при $n = 5$:

$$\begin{array}{ccccc} X & 5 & & 4 & & 3 \\ P & (0{,}2)^5 & & 5(0{,}2)^4 0{,}8 & & 10(0{,}2)^3(0{,}8)^2 \end{array}$$

$$\begin{array}{ccccc} X & 2 & & 1 & & 0 \\ P & 10(0{,}2)^2(0{,}8)^3 & & 5(0{,}2)(0{,}8)^4 & & (0{,}8)^5, \end{array}$$

или окончательно:

$$\begin{array}{cccccccc} X & 5 & & 4 & & 3 & 2 & 1 & 0 \\ P & 0{,}00032 & & 0{,}0064 & & 0{,}0512 & 0{,}2048 & 0{,}4096 & 0{,}32768. \end{array}$$

Распределение Пуассона

Пусть в каждом из n производимых испытаний вероятность появления события A равна p. Как мы знаем, для определения вероятности k появлений события A используется формула Бернулли (23.2); при больших n пользуются асимптотической формулой Лапласа (22.17). Однако эта формула плохо подходит для случая, когда p мало. Для случая малых значений p и больших значений n используется *асимптотическая формула Пуассона*. Эта формула выведена при важном допущении, что произведение np является постоянной величиной, т.е. $np = \lambda$. Тогда вероятность того, что событие A наступит ровно k раз, дается формулой, которая представляет собой *закон распределения Пуассона вероятностей массовых и редких событий*:

$$P_n(k) = \lambda^k e^{-\lambda}/k!. \qquad (23.4)$$

Пример 5. На базу отправлено 10 000 изделий. Вероятность того, что изделие в пути получит повреждение, равна 0,0003. Найти вероятность того, что на базу прибудут 4 поврежденных изделия.

Решение. По условию задачи $n = 10\,000$, $p = 0{,}0003$, $k = 4$. Находим λ, а затем по формуле (23.4) и искомую вероятность:

$$\lambda = np = 10\,000 \cdot 0{,}0003 = 3,$$

$$P_{10\,000}(4) = 3^4 e^{-3}/4! = 0{,}168.$$

23.2. Числовые характеристики дискретных случайных величин

Установленный закон распределения полностью характеризует случайную величину. Однако часто используются *числовые характеристики случайной величины*, которые дают некоторое осредненное описание случайной величины, получаемое на базе закона ее распределения.

Математическое ожидание дискретной случайной величины

Пусть случайная величина X может принимать значения $x_1, x_2, \ldots, x_n$ с вероятностями соответственно $p_1, p_2, \ldots, p_n$.

Определение 1. *Математическим ожиданием дискретной случайной величины* называется сумма произведений всех ее возможных значений на соответствующие вероятности:

$$M(X) = x_1p_1 + x_2p_2 + \cdots + x_np_n = \sum_{i=1}^{n} x_ip_i. \qquad (23.5)$$

Из этого определения следует, что математическое ожидание есть некоторая постоянная (неслучайная) величина. Вероятностный же смысл математического ожидания состоит в том, что оно приближенно равно (в особенности для большого числа испытаний) *среднему арифметическому значений случайной величины*. Это хорошо видно в случае, когда вероятности всех возможных значений дискретной случайной величины равны: $p_i = p = 1/n$; из формулы (23.5) получаем

$$M(X) = (x_1 + x_2 + \cdots + x_n)/n. \qquad (23.6)$$

Пример 1. Найти математическое ожидание количества очков, выпадающих при бросании игральной кости.

Решение. Выпадение каждой грани кубика от одного очка до шести имеет одинаковую вероятность $p = 1/6$. Следовательно, по формуле (23.6) получаем искомое математическое ожидание:

$$M(X) = (1 + 2 + 3 + 4 + 5 + 6)/6 = 3{,}5.$$

Пример 2. Найти математическое ожидание числа невозврата кредитов по данным примера 4 п. 23.1.

Решение. Воспользуемся итоговой таблицей распределения дискретной случайной величины, полученной в этом примере, и формулой (23.6); находим

$$M(X) = 5 \cdot 0{,}00032 + 4 \cdot 0{,}0064 + 3 \cdot 0{,}0512 +$$
$$+2 \cdot 0{,}2048 + 1 \cdot 0{,}4096 + 0 \cdot 0{,}32768 = 1.$$

Свойства математического ожидания

Математическое ожидание обладает рядом свойств, которые указаны ниже.

Свойство 1. Математическое ожидание постоянной величины C равно этой постоянной:

$$M(C) = C.$$

Свойство 2. Постоянный множитель можно выносить за знак математического ожидания:

$$M(CX) = CM(X).$$

Свойство 3. Математическое ожидание суммы случайных величин равно сумме их математических ожиданий:

$$M(X_1 + X_2 + \cdots + X_n) = M(X_1) + M(X_2) + \cdots + M(X_n).$$

Свойство 4. Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий:

$$M(X_1 X_2 \ldots X_n) = M(X_1)M(X_2)\ldots M(X_n).$$

Пример 3. Пусть ежедневные расходы на обслуживание и рекламу автомобилей в некотором автосалоне составляют в среднем 100 тыс. р., а число продаж X автомашин в течение дня подчиняется следующему закону распределения:

X	0	1	2	3	4	5	6	7	8	9
P	0,25	0,2	0,1	0,1	0,1	0,1	0,05	0,05	0,025	0,025.

Найти математическое ожидание ежедневной прибыли при цене на машину 150 тыс. р.

Решение. Ежедневная прибыль подсчитывается по формуле

$$\text{П} = (150X - 100) \text{ тыс. р.}$$

Искомая характеристика $M(\text{П})$ находится с использованием указанных выше свойств математического ожидания (в тыс. р.):

$$M(\text{П}) = M(150X - 100) = 150M(X) - 100 =$$
$$= 150 \cdot 2{,}675 - 100 = 301{,}25.$$

Если в n независимых испытаниях вероятность появления в каждом из них события A постоянна, то ответ на вопрос о среднем числе появления события A дает следующая теорема.

Теорема 23.1. *Математическое ожидание $M(X)$ числа появлений события A в n независимых испытаниях равно произведению числа испытаний на вероятность появления события в каждом испытании:*

$$M(X) = np. \tag{23.7}$$

Пример 4. Найти математическое ожидание числа выигрышных лотерейных билетов, если вероятность выигрыша по одному билету равна 0,015, причем куплено 200 билетов.

Решение. Поскольку приобретение каждого билета является независимым испытанием относительно появления события A — выпадения выигрыша, то здесь применимы теорема 23.1 и формула (23.7). В нашем случае $n = 200$, $p = 0{,}015$, откуда мы получаем

$$M(200) = 200 \cdot 0{,}015 = 3.$$

Дисперсия дискретной случайной величины

Как уже говорилось выше, математическое ожидание является средней характеристикой случайной величины. Однако оно не характеризует случайную величину достаточно полно, и по этой причине рассматриваются и другие числовые характеристики. Пусть X — случайная величина, а $M(X)$ — ее математическое ожидание.

Определение 2. Разность между случайной величиной и ее математическим ожиданием называется *отклонением*.

Пусть закон распределения случайной величины X дается формулой (23.1), тогда отклонение $X - M(X)$ имеет следующий закон распределения:

37 Зак. 6853

$X-M(X)$	$x_1-M(X)$	$x_2-M(X)$	$\ldots$	$x_n-M(X)$
p	p_1	p_2	$\ldots$	p_n.

(23.8)

Отклонение имеет важное свойство, которое устанавливается непосредственно из свойств математического ожидания:

$$M[X - M(X)] = 0, \tag{23.9}$$

т.е. *математическое ожидание отклонения равно нулю.*

Пример 5. По данным примера 3 найти закон распределения отклонения числа проданных за день автомашин.

Решение. Как было подсчитано в примере 3, $M(X) = 2{,}675$. Тогда, согласно (23.8), искомый закон определяется следующей таблицей:

$X - M(X)$	$-2{,}675$	$-1{,}675$	$-0{,}675$	$0{,}325$	$1{,}325$
p	$0{,}25$	$0{,}2$	$0{,}1$	$0{,}1$	$0{,}1$
$X - M(X)$	$2{,}325$	$3{,}325$	$4{,}325$	$5{,}325$	$6{,}325$
p	$0{,}1$	$0{,}05$	$0{,}05$	$0{,}025$	$0{,}025$.

На практике важной характеристикой является рассеяние возможных значений случайной величины вокруг ее среднего значения. Среднее значение отклонения, согласно (23.9), равно нулю, так как суммируются отрицательные и положительные отклонения (см. пример 5), поэтому целесообразно ввести в рассмотрение абсолютные значения отклонений или их квадраты.

Определение 3. Математическое ожидание квадрата отклонения называется *дисперсией*, или *рассеянием*:

$$D(X) = M[X - M(X)]^2. \tag{23.10}$$

Пусть случайная величина задана законом распределения (23.1), тогда квадрат отклонения этой случайной величины имеет следующий закон распределения:

$[X-M(X)]^2$	$[x_1-M(X)]^2$	$[x_2-M(X)]^2$	$\ldots$	$[x_n-M(X)]^2$
p	p_1	p_2	$\ldots$	p_n.

Отсюда, согласно формуле (23.10), получаем формулу дисперсии в развернутом виде:

$$D(X) = [x_1 - M(X)]^2 + [x_2 - M(X)]^2 + \ldots + [x_n - M(X)]^2.$$

При вычислении дисперсии часто бывает удобно воспользоваться формулой, которая непосредственно выводится из формулы (23.10):

$$D(X) = M(X^2) - [M(X)]^2. \tag{23.11}$$

Пример 6. Найти дисперсию ежедневной продажи числа автомашин по данным примера 3.

Решение. Закон распределения случайной величины X^2 имеет вид

X^2	0	1	4	9	16	25	36	49	64	81
P	0,25	0,2	0,1	0,1	0,1	0,1	0,05	0,05	0,025	0,025.

Математическое ожидание $M(X^2)$ подсчитывается из этой таблицы:

$$M(X^2) = 0 \cdot 0{,}25 + 1 \cdot 0{,}2 + 4 \cdot 0{,}1 + 9 \cdot 0{,}1 + 16 \cdot 0{,}1 + 25 \cdot 0{,}1 +$$
$$+36 \cdot 0{,}05 + 49 \cdot 0{,}05 + 64 \cdot 0{,}025 + 81 \cdot 0{,}025 = 13{,}475.$$

Математическое ожидание $M(X) = 2{,}675$. Следовательно, согласно формуле (23.11), получаем искомую величину дисперсии:

$$D(X) = M(X^2) - [M(X)]^2 = 13{,}475 - 7{,}156 = 6{,}319.$$

Свойства дисперсии

Приведем здесь основные свойства дисперсии.

Свойство 1. Дисперсия постоянной величины C равна нулю:

$$D(C) = 0.$$

Свойство 2. Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:

$$D(CX) = C^2 D(X). \qquad (23.12)$$

Свойство 3. Дисперсия суммы независимых случайных величин равна сумме их дисперсий:

$$D(X_1+X_2+\ldots+X_n)=D(X_1)+D(X_2)+\ldots+D(X_n). \qquad (23.13)$$

Перечисленные свойства дисперсии используются при вычислениях, когда мы имеем дело с несколькими случайными величинами. Из свойств 1 и 3 следует важный вывод: $D(X+C) = D(X)$, где C — постоянная величина. Кроме того, справедлива следующая теорема.

Теорема 23.2. *Дисперсия числа появления события A в n независимых испытаниях с вероятностью появления p в каждом из них этого события вычисляется по формуле*

$$D(X) = np(1-p) = npq. \qquad (23.14)$$

Приведем здесь еще два важных результата: для случайной величины, распределенной по закону Пуассона (23.4), математическое ожидание и дисперсия равны параметру данного распределения.

Пример 7. Найти дисперсию числа выигрышных лотерейных билетов по данным примера 4.

Решение. Имеем 200 независимых испытаний с вероятностью появления выигрышного билета $p = 0{,}015$. Стало быть, $q = 1 - 0{,}015 = 0{,}985$, откуда и получаем искомую дисперсию:

$$D(X) = npq = 200 \cdot 0{,}015 \cdot 0{,}985 = 2{,}955.$$

Пример 8. Банк выдал ссуды n разным заемщикам в размере S р. каждому под ставку ссудного процента r. Найти математическое ожидание и дисперсию прибыли банка, а также условие на ставку ссудного процента, если вероятность возврата ссуды заемщиком равна p.

Решение. Поскольку заемщики между собой не связаны, то можно полагать, что мы имеем n независимых испытаний.

Вероятность утери ссуды для банка в каждом испытании равна $q = 1-p$. Пусть X — число заемщиков, возвративших ссуду с ссудным процентом, тогда прибыль банка определяется формулой

$$\Pi = (1 + r/100)SX - nS,$$

где X является случайной величиной с биномиальным законом распределения. Тогда, согласно теореме 23.1, математическое ожидание прибыли определяется с использованием формулы (23.7):

$$M(\Pi) = (1+r/100)SM(X) - nS =$$
$$= (1 + r/100)Snp - Sn = Sn(rp/100-q).$$

Поскольку выдача ссуды имеет смысл лишь при положительном математическом ожидании прибыли (положительная средняя величина прибыли), то из условия $M(\Pi) > 0$ вытекает условие на ставку ссудного процента:

$$r > 100q/p, \text{ или } r > 100(1-p)/p.$$

Дисперсия прибыли банка находится согласно теореме 23.2, с использованием формулы (23.14) и свойств 1–3:

$$D(\Pi) = D((1 + r/100)SX - nS) = (1 + r/100)^2 S^2 npq.$$

Среднее квадратическое отклонение

Одной из основных оценок рассеяния возможных значений случайной величины служит *среднее квадратическое отклонение*.

Определение 4. Средним квадратическим отклонением случайной величины X (стандартом) называется квадратный корень из ее дисперсии:

$$\sigma(X) = \sqrt{D(X)}. \qquad (23.15)$$

Согласно этому определению, из свойства 3 и формулы (23.13) следует, что в случае суммы взаимно независимых случайных величин справедлива формула

$$\sigma(X_1 + X_2 + \ldots + X_n) = \sqrt{\sigma^2(X_1) + \sigma^2(X_2) + \ldots + \sigma^2(X_n)}.$$

Пример 9. Найти дисперсию и среднее квадратическое отклонение случайной величины X, заданной следующим распределением:

X	-5	2	3	4	6
p	0,1	0,2	0,3	0,3	0,1.

Решение. Имеем $M(X) = 2{,}6$. Составим таблицу распределения случайной величины X^2:

X^2	25	4	9	16	36
p	0,1	0,2	0,3	0,3	0,1.

Отсюда получаем, что $M(X^2) = 14{,}4$. По формулам (23.11) и (23.15) окончательно получаем искомые значения $D(X)$ и $\sigma(X)$:

$$D(X) = M(X^2) - [M(X)]^2 = 7{,}64,$$
$$\sigma(X) = \sqrt{D(X)} = 2{,}76.$$

Пример 10. Законы распределения независимых случайных величин X и Y приведены соответственно в таблицах:

X	-2	0	1	3	4
p	0,1	0,2	0,3	0,3	0,1;

Y	2	4	6	8
p	0,2	0,4	0,3	0,1.

Найти дисперсию и среднее квадратическое отклонение случайной величины $Z = 2X + 3Y$.

Решение. Согласно свойствам 2 и 3 дисперсии (формулы (23.12) и (23.13)), имеем

$$D(Z) = D(2X + 3Y) = 4D(X) + 9D(Y).$$

Для вычисления дисперсий $D(X)$ и $D(Y)$ составляем соответствующие таблицы — законы распределения случайных величин X^2 и Y^2:

X^2	4	0	1	9	16
p	0,1	0,2	0,3	0,3	0,1;

Y^2	4	16	36	64
p	0,2	0,4	0,3	0,1.

Отсюда получаем

$$D(X) = M(X^2) - [M(X)]^2 = 5 - 1{,}4^2 = 3{,}04;$$
$$D(Y) = M(Y^2) - [M(Y)]^2 = 24{,}4 - 4{,}6^2 = 3{,}24.$$

Искомые дисперсия и среднее квадратичное отклонение случайной величины Z равны:

$$D(Z) = D(2X + 3Y) = 4D(X) + 9D(Y) =$$
$$= 4 \cdot 3{,}04 + 9 \cdot 3{,}24 = 41{,}32;$$
$$\sigma(Z) = \sqrt{D(Z)} = 6{,}43.$$

Пример 11. В условиях примера 8 найти математическое ожидание и среднее квадратическое отклонение прибыли при $n = 1000$, $p = 0{,}8$, $S = 100$ тыс. р. и $r = 30\%$.

Решение. Ставка ссудного процента удовлетворяет условию, чтобы математическое ожидание прибыли было положительным: $30 > 100(1 - 0{,}8)/0{,}8$. Математическое ожидание прибыли:

$$M(\Pi) = Sn(rp/100 - q) =$$
$$= 100 \cdot 1000(30 \cdot 0{,}8/100 - 0{,}2) = 4 \text{ млн. р.}$$

Среднее квадратическое отклонение прибыли:

$$\sigma(X) = \sqrt{D(X)} = (1 + r/100)S\sqrt{npq} =$$
$$= 1{,}3 \cdot 100\sqrt{1000 \cdot 0{,}8 \cdot 0{,}2} = 1644{,}38 \text{ тыс. р.}$$

Начальные и центральные моменты

Определение 5. *Начальным моментом порядка k* случайной величины X называется математическое ожидание величины X^k:

$$\nu_k = M(X^k). \tag{23.16}$$

В частности,

$$\nu_1 = M(X), \quad \nu_2 = M(X^2),$$

и тогда формула (23.11) для вычисления дисперсии принимает вид

$$D(X) = \nu_2 - \nu_1^2. \tag{23.17}$$

Определение 6. *Центральным моментом порядка* k случайной величины X называется математическое ожидание k-й степени отклонения:

$$\mu_k = M[(X - M(X))^k]. \tag{23.18}$$

В частности, согласно формуле (23.9), $\mu_1 = 0$, а дисперсия случайной величины X является центральным моментом второго порядка:

$$D(X) = \mu_2 = \nu_2 - \nu_1^2. \tag{23.19}$$

Соотношения, связывающие начальные и центральные моменты, также могут быть легко получены. Приведем их здесь для моментов третьего и четвертого порядков (они наряду с моментами первого и второго порядков широко применяются в статистике):

$$\begin{gathered}\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3, \\ \mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4.\end{gathered} \tag{23.20}$$

Моменты более высоких порядков применяются крайне редко.

Моменты, рассмотренные в этом разделе, называют *теоретическими*. В отличие от них моменты, вычисляемые по данным наблюдений в математической статистике, называют *эмпирическими*.

23.3. Система двух случайных величин

Двумерная случайная величина

До сих пор мы рассматривали дискретные случайные величины, которые называют одномерными: их возможные значения определялись одним числом. Кроме одномерных величин рассматривают также величины, возможные значения которых определяются несколькими числами. Двумерную случайную величину обозначают через (X, Y); каждая из величин

X и Y называется компонентой (составляющей). Обе величины X и Y, рассматриваемые одновременно, образуют систему двух случайных величин. Например, при штамповке стальных пластинок их длина и ширина прелставляют собой двумерную случайную величину.

Определение 1. Законом распределения двумерной случайной величины (X, Y) называют множество возможных пар чисел (x_i, y_j) и их вероятностей $p(x_i, y_j)$. Двумерную случайную величину можно трактовать как случайную точку $A(X, Y)$ на координатной плоскости.

Закон распределения двумерной случайной величины обычно задается в виде таблицы, в строках которой указаны возможные значения x_i случайной величины X, а в столбцах — возможные значения y_j случайной величины Y, на пересечениях строк и столбцов указаны соответствующие вероятности p_{ij}. Пусть случайная величина X может принимать n значений, а случайная величина Y — m значений. Тогда закон распределения двумерной случайной величины (X, Y) имеет вид

$Y/X \to$ $\downarrow$	x_1	x_2	$\dots$	x_i	$\dots$	x_n
y_1	p_{11}	p_{12}	$\dots$	p_{1i}	$\dots$	p_{1n}
y_2	p_{21}	p_{22}	$\dots$	p_{2i}	$\dots$	p_{2n}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_j	p_{j1}	p_{j2}	$\dots$	p_{ji}	$\dots$	p_{jn}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_m	p_{m1}	p_{m2}	$\dots$	p_{mi}	$\dots$	p_{mn}.

(23.21)

Из этой таблицы можно найти законы распределения каждой из случайных компонент. Например, вероятность того, что случайная величина X примет значение x_k, равна, согласно теореме сложения вероятностей независимых событий,

$$P(x_k) = p_{1k} + p_{2k} + \dots + p_{mk}, \quad k = 1, 2, \dots, n. \tag{23.22}$$

Иными словами, для нахождения вероятности $P(x_k)$ нужно просуммировать все m вероятностей по k-му столбцу таблицы (23.21). Аналогично получается вероятность того, что случайная величина Y примет возможное значение y_r: $P(y_r)$ получается суммированием всех n вероятностей r-й строки таблицы (23.21) ($r = 1, 2, \dots, m$). Отсюда следует, что

сумма всех вероятностей в законе распределения (23.21) равна единице:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} = 1. \qquad (23.23)$$

Пример 1. Задано распределение двумерной случайной величины:

$Y/X \to$	1	2	3
$\downarrow$			
1	0,1	0,15	0,12
2	0,2	0,22	0,21.

Найти распределения X, Y и $X+Y$.

Решение. В нашем случае возможные значения случайной величины X: $x_1 = 1$, $x_2 = 2$, $x_3 = 3$. Тогда, согласно формуле (23.22), имеем $P(x_1) = 0{,}1+0{,}2 = 0{,}3$, $P(x_2) = 0{,}15+0{,}22 = 0{,}37$, $P(x_3) = 0{,}12 + 0{,}21 = 0{,}33$. Отсюда получаем закон распределения X:

X	1	2	3
p	0,3	0,37	0,33.

Аналогично получаем и для распределения Y: $y_1 = 1$, $y_2 = 2$; $P(y_1) = 0{,}1+0{,}15+0{,}12 = 0{,}37$, $P(y_2) = 0{,}2+0{,}22+0{,}21 = 0{,}63$;

Y	1	2
p	0,37	0,63.

Теперь найдем распределение $X+Y$. Возможные значения этой случайной величины: 2, 3, 4 и 5. Соответствующие вероятности $P(2) = 0{,}1$, $P(3) = 0{,}15+0{,}2 = 0{,}35$, $P(4) = 0{,}12+0{,}22 = 0{,}34$, $P(5) = 0{,}21$. Отсюда находим искомое распределение:

$X+Y$	2	3	4	5
p	0,1	0,35	0,34	0,21.

В случае системы двух случайных величин используются кроме математических ожиданий и дисперсий еще и другие числовые характеристики, описывающие их взаимосвязь.

Корреляционный момент

Определение 2. *Корреляционным моментом* случайных величин X и Y (или *ковариацией*) называется математическое ожидание произведений их отклонений:

$$\mu_{xy} = \operatorname{Cov}(X, Y) = M\{[X - M(X)][Y - M(Y)]\}.$$

Корреляционный момент служит для описания связи между случайными величинами X и Y. Из свойств математического ожидания легко убедиться в том, что μ_{xy} можно записать в следующем виде:

$$\mu_{xy} = M(XY) - M(X)M(Y).$$

Для непосредственного вычисления корреляционного момента (ковариации) используется формула (см. распределение (23.21))

$$\mu_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_{ij} - M(X)M(Y). \qquad (23.24)$$

Теорема 23.3. *Корреляционный момент двух независимых случайных величин X и Y равен нулю.*

Если корреляционный момент μ_{xy} не равен нулю, то, стало быть, величины X и Y являются зависимыми.

Коэффициент корреляции

Из определения корреляционного момента следует, что его размерность равна произведению размерностей величин X и Y; например, если X и Y измерены в сантиметрах, то μ_{xy} имеет размерность см^2.

Это обстоятельство затрудняет сравнение корреляционных моментов различных систем случайных величин. Для устранения этого недостатка вводят безразмерную числовую характеристику — *коэффициент корреляции*, величина которого не зависит от выбора системы измерения случайных величин.

Определение 3. Коэффициентом корреляции случайных величин X и Y называется отношение их корреляционного момента к произведению средних квадратических отклонений этих величин:

$$r_{xy} = \mu_{xy}/(\sigma_x\sigma_y) = \mu_{xy}/\sqrt{D(X)D(Y)}. \qquad (23.25)$$

Из определения и свойств математического ожидания и дисперсии следует важный вывод, что абсолютная величина коэффициента корреляции не превосходит единицы:

$$|r_{xy}| \leqslant 1 \text{ или } -1 \leqslant r_{xy} \leqslant 1. \qquad (23.26)$$

Определение 4. Две случайные величины X и Y называются *коррелированными*, если их корреляционный момент (коэффициент корреляции) отличен от нуля; если же их корреляционный момент равен нулю, то X и Y называются *некоррелированными*.

Таким образом, две коррелированные случайные величины (т.е. при $r_{xy} \neq 0$) являются также и зависимыми. Обратное утверждение неверно, т.е. две зависимые величины могут быть как коррелированными, так и некоррелированными.

Пример 2. Найти корреляционный момент и коэффициент корреляции двух случайных величин X и Y, распределения которых заданы в предыдущем примере 1.

Решение. Воспользуемся формулами (23.24), (23.26), а также формулой вычисления центрального момента второго порядка (23.19); последовательно вычисляем: $M(X)$=2,03, $M(Y)$=1,63, $D(X) = 0{,}629$, $D(Y) = 0{,}233$,

$$\mu_{xy} = \sum_{i=1}^{3}\sum_{j=1}^{2} x_i y_j p_{ij} - M(X)M(Y) =$$

$$= 1\cdot 1\cdot 0{,}1 + 1\cdot 2\cdot 0{,}2 + 2\cdot 1\cdot 0{,}15 + 2\cdot 2\cdot 0{,}22 +$$

$$+3\cdot 1\cdot 0{,}12 + 3\cdot 2\cdot 0{,}21 - 2{,}03\cdot 1{,}63 = -0{,}0089,$$

$$r_{xy} = -0{,}0089/\sqrt{0{,}629\cdot 0{,}233} =$$

$$= -0{,}023.$$

В данном случае коэффициент корреляции близок к нулю; это означает, что случайные величины X и Y *слабокоррелированы*.

Линейная регрессия

Пусть (X, Y) — двумерная случайная величина, где X и Y — зависимые случайные величины. Оказывается возможным приближенное представление величины Y в виде линейной функции величины X:

$$Y \approx g(x) = aX + b, \qquad (23.27)$$

где a и b — параметры, подлежащие определению. Обычно эти величины определяются с помощью метода наименьших квадратов (см. п. 8.5).

Определение 5. Функция (23.27) называется *наилучшим приближением* в смысле метода наименьших квадратов, если математическое ожидание $M[Y - g(X)]^2$ принимает наименьшее возможное значение. Функцию $g(x)$ называют *среднеквадратической регрессией* Y на X.

Теорема 23.4. *Линейная средняя квадратическая регрессия Y на X имеет вид*

$$g(X) = m_y + r_{xy}\frac{\sigma_y}{\sigma_x}(X - m_x), \qquad (23.28)$$

где r_{xy} определяется формулой (23.25), $m_y = M(Y)$ *и* $m_x = M(X)$ — *математические ожидания соответственно случайных величин Y и X.*

Коэффициент $b = r_{xy}\sigma_y/\sigma_x$ называют *коэффициентом регрессии* Y на X, а прямую

$$y - m_y = r_{xy}\frac{\sigma_y}{\sigma_x}(x - m_x),$$

реализующую линейную зависимость (23.28) случайной величины Y от случайной величины X, называют *прямой среднеквадратической регрессии* X на Y. Поскольку зависимость (23.28) является приближенной, то существует погрешность этого приближения, называемая *остаточной дисперсией*:

$$\varepsilon^2 = \sigma_y^2(1 - r_{xy}^2). \qquad (23.29)$$

Аналогичную форму записи имеет прямая среднеквадратическая регрессия X на Y:

$$x - m_x = r_{xy}\frac{\sigma_x}{\sigma_y}(y - m_y).$$

Пример 3. Найти линейную среднюю квадратическую регрессию и остаточную дисперсию случайной величины Y на случайную величину X по данным примеров 1 и 2.

Решение. Для двумерной случайной величины (X, Y), приведенной в примере 1, все необходимые числовые характеристики указаны в решении примера 2: $m_x = 2{,}03$, $m_y = 1{,}63$, $r_{xy} = -0{,}023$, $\sigma_x = \sqrt{D(X)} = 0{,}793$, $\sigma_y = \sqrt{D(Y)} = 0{,}483$. Из уравнения (23.28) получаем искомое соотношение:

$$g(X) = 1{,}63 - 0{,}014(X - 2{,}03).$$

Остаточная дисперсия рассчитывается по формуле (23.29):

$$\varepsilon^2 = \sigma_y^2(1 - r_{xy}^2) = 0{,}483^2(1 - 0{,}023^2) = 0{,}233.$$

Для оценки среднеквадратичной погрешности линейной регрессии обычно используют величину ε, в нашем случае она составляет

$$\varepsilon = 0{,}483.$$

23.4. Непрерывные случайные величины

Функция распределения и ее свойства

Пусть X — непрерывная случайная величина (см. определение 3 п. 23.1), значения которой сплошь заполняют интервал (a, b). Теперь уже нельзя составить перечень всех возможных значений X, как это было сделано в случае дискретной случайной величины. Тем не менее существует способ задания любых видов случайных величин. Пусть x — действительное число. Обозначим вероятность события того, что X примет значение, меньшее x, через $F(x)$.

Определение 1. *Функцией распределения случайной величины* X называется функция $F(x)$, определяющая вероятность того, что X примет значение, меньшее x:

$$F(x) = P(X < x). \qquad (23.30)$$

Геомстрический смысл приведенного определения: $F(x)$ — это вероятность того, что случайная величина X примет значение, изображаемое точкой на числовой оси левее точки x. По виду функции $F(x)$ определяется и вид случайной величины. Уточним понятие непрерывной случайной величины.

Определение 2. Случайная величина называется *непрерывной*, если ее функция распределения есть непрерывная кусочно-дифференцируемая функция с непрерывной производной.

Таким образом, дискретную случайную величину можно считать кусочно-непрерывной.

Функция распределения обладает рядом фундаментальных свойств, указанных ниже.

Свойство 1. Область значений функции распределения лежит на отрезке $[0, 1]$:

$$0 \leqslant F(x) \leqslant 1.$$

Свойство 2. Функция распределения является неубывающей, т.е.

$$F(x_2) \geqslant F(x_1) \text{ при } x_2 > x_1.$$

Свойство 3. Если возможные значения случайной величины находятся на интервале (a, b), то $F(x) = 0$ при $x \leqslant a$ и $F(x) = 1$ при $x \geqslant b$.

Из указанных свойств вытекают важные следствия.

1. Вероятность того, что случайная величина X принимает значения, заключенные внутри интервала (α, β), равна разности значений функции распределения на концах этого интервала:

$$P(\alpha \leqslant X < \beta) = F(\beta) - F(\alpha). \qquad (23.31)$$

2. Вероятность того, что непрерывная случайная величина X примет одно определенное значение, равна нулю.

3. Если возможные значения непрерывной случайной величины X расположены на всей числовой оси, то

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1.$$

График функции распределения непрерывной случайной величины показан на рис. 23.2.

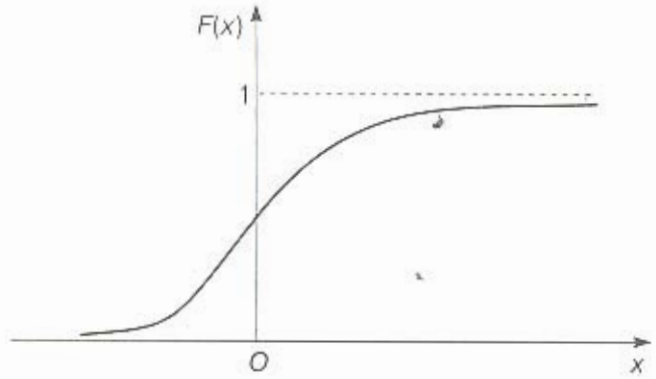

Рис. 23.2

Пример 1. Найти функцию распределения процентного изменения стоимости акций по данным примера 3 п. 23.1 и построить ее график.

Решение. Перепишем таблицу распределения дискретной случайной величины в порядке возрастания ее возможных значений:

X	0	5	10	15	20	25	30
P	0	0,1	0,1	0,2	0,3	0,2	0,1.

Если $x \leqslant 5$, то $F(x) = 0$. Если $5 < x \leqslant 10$, то $F(x) = 0{,}1$. На интервале $10 < x \leqslant 15$ применяем теорему сложения вероятностей, так как события $X < 10$ и $10 < X \leqslant 15$ несовместны: $F(x) = 0{,}1 + 0{,}1 = 0{,}2$. Аналогично определяются значения $F(x)$ на других интервалах: при $15 < x \leqslant 20$ $F(x) = 0{,}4$; при $20 < x \leqslant 25$ $F(x) = 0{,}7$; при $25 < x \leqslant 30$ $F(x) = 0{,}9$; при $x > 30$ имеем достоверное событие (все случаи изменения стоимости акций исчерпаны), т.е. $F(x) = 1$. Таким образом, искомая функция распределения имеет следующую аналитическую форму записи:

$$F(x) = \begin{cases} 0, & x \leqslant 5; \\ 0{,}1, & 5 < x \leqslant 10; \\ 0{,}2, & 10 < x \leqslant 15; \\ 0{,}4, & 15 < x \leqslant 20; \\ 0{,}7, & 20 < x \leqslant 25; \\ 0{,}9, & 25 < x \leqslant 30; \\ 1, & x > 30. \end{cases}$$

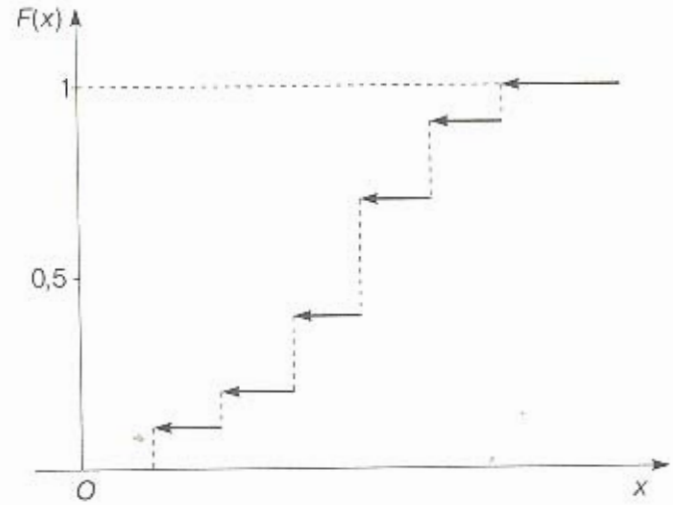

Рис. 23.3

График этой функции распределения показан на рис. 23.3.

Плотность распределения вероятностей и ее свойства

Определение 3. Производная от функции распределения непрерывной случайной величины X называется плотностью распределения вероятностей X:

$$f(x) = F'(x). \tag{23.32}$$

Из этого определения следует, что функция распределения является первообразной для плотности распределения или неопределенным интегралом от нее. Плотность распределения — это "скорость" изменения вероятности $P(X < x)$. Из свойства 2 функции распределения следует справедливость следующей фундаментальной теоремы.

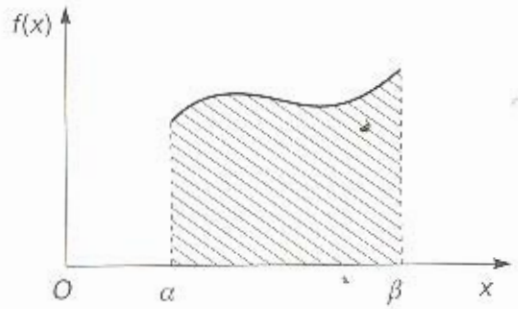

Рис. 23.4

Теорема 23.5. *Вероятность того, что непрерывная случайная величина X примет значение на интервале (α, β), определяется по формуле*

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)\,dx. \tag{23.33}$$

Вспоминая геометрический смысл определенного интеграла (см. п. 7.5), можно сказать, что вероятность того, что непрерывная случайная величина X примет значение, принадлежащее интервалу (α, β), равна площади криволинейной трапеции, ограниченной сверху кривой плотности распределения $f(x)$, снизу — осью Ox, а с краев — вертикальными прямыми $x = \alpha$ и $x = \beta$ (рис. 23.4).

Связь между функцией распределения и плотностью распределения вероятностей устанавливается, согласно (23.32), формулой

$$F(x) = P(X < x) = \int_{-\infty}^{x} f(z)\,dz. \tag{23.34}$$

Пример 2. Случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0, & x \leqslant 0; \\ \sin x, & 0 < x \leqslant \pi/2; \\ 1, & x > \pi/2. \end{cases}$$

Найти плотность распределения X.

Решение. Функция $F(x)$ является кусочно-дифференцируемой. Согласно формуле (23.32), дифференцируя $F(x)$ по интервалам ее задания, получаем

$$f(x) = F'(x) = \begin{cases} 0, & x \leqslant 0; \\ \cos x, & 0 < x \leqslant \pi/2; \\ 0, & x > \pi/2. \end{cases}$$

Пример 3. Непрерывная случайная величина X задана плотностью распределения на всей числовой оси:

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

Найти вероятность того, что X примет значение на интервале $(-1, 1)$.

Решение. Согласно формуле (23.33), искомая вероятность равна

$$P(-1 < X < 1) = \frac{1}{\pi}\int_{-1}^{1} \frac{1}{1+x^2}\,dx =$$

$$= \frac{1}{\pi}\operatorname{arctg} x\Big|_{-1}^{1} = \frac{1}{\pi}\left(\frac{\pi}{4} + \frac{\pi}{4}\right) = \frac{1}{2}.$$

Плотность распределения обладает рядом свойств, основные из них указаны ниже.

Свойство 1. Плотность распределения является неотрицательной функцией:

$$f(x) \geqslant 0.$$

Это следует из характера функции распределения: она является неубывающей, и, значит, ее производная неотрицательна.

Свойство 2. Несобственный интеграл от плотности распределения в пределах интегрирования по всей числовой оси равен единице:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

Это равенство означает достоверность события, что случайная величина X примет значение, принадлежащее интервалу $(-\infty, \infty)$, т.е. вероятность этого события $P(-\infty < X < \infty)$=1.

Так, если все возможные значения случайной величины X лежат внутри интервала (a, b), то

$$\int_a^b f(x)\,dx = 1. \qquad (23.35)$$

Числовые характеристики непрерывных случайных величин

Определения числовых характеристик дискретных случайных величин распространяются и на непрерывные величины. Разница состоит в том, что вместо сумм в формулах (23.5) и (23.10) берутся их интегральные аналоги.

Определение 4. *Математическим ожиданием непрерывной случайной величины* X, возможные значения которой находятся на отрезке $[a, b]$, называется определенный интеграл:

$$M(X) = \int_a^b xf(x)\,dx. \qquad (23.36)$$

В том случае, когда возможные значения случайной величины X заполняют всю ось Ox, пределы интегрирования a и b бесконечны: $a = -\infty$, $b = \infty$. Возможны также случаи, когда один из пределов интегрирования бесконечен (возможные значения X лежат на полупрямой).

Определение 5. *Дисперсией непрерывной случайной величины* X называется математическое ожидание квадрата ее отклонения:

$$D(X) = \int_a^b [x - M(X)]^2 f(x)\,dx. \qquad (23.37)$$

Все сказанное выше о случаях бесконечных пределов интегрирования остается справедливым и для дисперсии.

Среднее квадратическое отклонение непрерывной случайной величины определяется, как и прежде, по формуле (23.15):

$$\sigma(X) = \sqrt{D(X)}.$$

Для вычисления дисперсии употребляется более удобная формула, которая выводится из (23.37):

$$D(X) = \int_a^b x^2 f(x)dx - [M(X)]^2. \qquad (23.38)$$

Пример 4. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X, заданной плотностью распределения на отрезке $[0, 1]$:

$$f(x) = 1, \quad x \in [0, 1].$$

Решение. Согласно формулам (23.36), (23.38) и (23.15) последовательно вычисляем искомые величины:

$$M(X) = \int_0^1 xf(x)dx = \int_0^1 x\,dx = \frac{1}{2}x^2\Big|_0^1 = \frac{1}{2},$$

$$D(X) = \int_0^1 x^2 f(x)dx - [M(X)]^2 = \int_0^1 x^2\,dx - (1/2)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12},$$

$$\sigma(X) = \sqrt{D(X)} = 1/2\sqrt{3} \approx 0{,}289.$$

Пример 5. Найти основные числовые характеристики непрерывной случайной величины X, заданной функцией распределения на положительной полуоси Ox:

$$F(x) = 1 - e^{-x}, \quad x \in (0, \infty).$$

Решение. Найдем сначала плотность распределения:

$$f(x) = F'(x) = e^{-x}, \quad x \in (0, \infty).$$

Затем, как и в предыдущем примере, вычисляем соответствующие интегралы; при их вычислении применяем правило интегрирования по частям для определенного интеграла. В итоге

получаем искомые величины:

$$M(X)=\int\limits_0^\infty x\,e^{-x}dx=-\int\limits_0^\infty x\,d(e^{-x})=$$

$$=-xe^{-x}\Big|_0^\infty+\int\limits_0^\infty e^{-x}dx=-e^{-x}\Big|_0^\infty=1,$$

$$D(X)=\int\limits_0^\infty x^2e^{-x}dx-[M(X)]^2=-\int\limits_0^\infty x^2d(e^{-x})-1=$$

$$=-x^2e^{-x}\Big|_0^\infty+2\int\limits_0^\infty x\,e^{-x}dx-1=2-1=1,$$

$$\sigma(X)=\sqrt{D(X)}=1.$$

23.5. Основные распределения непрерывных случайных величин

Равномерное распределение

Определение 1. Распределение вероятностей называется *равномерным*, если на интервале возможных значений случайной величины плотность распределения является постоянной.

Пусть на интервале (a,b) плотность распределения является постоянной величиной: $f(x)=C$. Определим значение C из условия (23.35):

$$\int\limits_a^b f(x)dx=\int\limits_a^b C\,dx=C(b-a)=1,$$

откуда получаем, что $f(x)=C=1/(b-a)$. Значит, искомая плотность равномерного распределения дается формулой

$$f(x)=\begin{cases}0, & x\leqslant a;\\ 1/(b-a), & a<x\leqslant b;\\ 0, & x>b.\end{cases}\qquad(23.39)$$

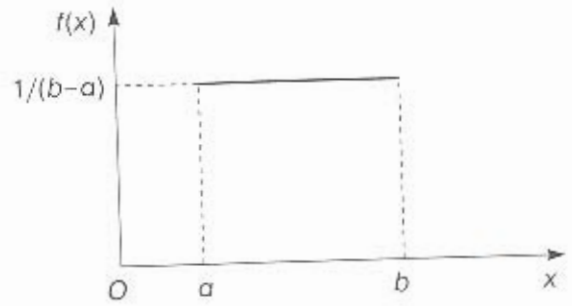

Рис. 23.5

График плотности равномерного распределения указан на рис. 23.5.

Пример 1. Найти среднеквадратическое отклонение случайной величины X, распределенной равномерно на интервале (1,5).

Решение. Согласно формуле (23.39), плотность распределения указанной случайной величины является ненулевой и равна 0,25 на интервале (1,5). По формулам (23.36) и (23.38) последовательно вычисляем:

$$M(Y)=\int\limits_a^b x\,f(x)dx=\int\limits_1^5 x\,0{,}25dx=0{,}125\,x^2\Big|_1^5=0{,}125(25-1)=3,$$

$$D(X)=\int\limits_a^b x^2f(x)dx-[M(X)]^2=\int\limits_1^5 x^2 0{,}25dx-3^2=$$

$$=0{,}25\,\frac{x^3}{3}\Big|_1^5-9=1{,}33,\qquad \sigma(X)=\sqrt{1{,}33}=1{,}15.$$

Пример 2. Радиус круга измерен приближенно на интервале (a,b). Полагая, что радиус является случайной величиной X, распределенной равномерно в этом интервале, найти математическое ожидание и дисперсию площади круга.

Решение. Площадь круга также является случайной величиной, вычисляемой по формуле $Y=\pi X^2$; она имеет то же равномерное распределение, что и случайная величина X. По

формулам (23.36) и (23.38) получаем

$$M(Y) = \int_a^b y\,f(x)\,dx = \frac{\pi}{b-a}\int_a^b x^2 dx = \frac{\pi}{b-a}\int_a^b x^2 dx =$$

$$= \frac{\pi}{b-a}\,\frac{b^3-a^3}{3} = \frac{\pi}{3}(a^2+ab+b^2),$$

$$D(Y) = \int_a^b y^2 f(x)\,dx - [M(Y)]^2 =$$

$$= \frac{\pi^2}{b-a}\int_a^b x^4 dx - \frac{\pi^2}{9}(a^2+ab+b^2)^2 = \frac{\pi^2}{b-a}\,\frac{b^5-a^5}{5} -$$

$$-\frac{\pi^2}{9}(a^2+ab+b^2)^2 = \frac{\pi^2}{45}(b-a)^2(4b^2+7ab+4a^2).$$

Нормальное распределение

Определение 2. *Общим нормальным* распределением вероятностей непрерывной случайной величины X называется распределение с плотностью

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\alpha)^2/2\sigma^2}. \qquad (23.40)$$

Нормальное распределение задается двумя параметрами: a и σ. Согласно определениям математического ожидания и дисперсии (формулы (23.36) и (23.38)), после выполнения соответствующих интегрирований можно вывести, что для нормального распределения справедливы формулы

$$M(X) = a, \quad D(X) = \sigma^2, \quad \sigma(X) = \sigma.$$

Определение 3. Нормальное распределение с параметрами $a = 0$ и $\sigma = 1$ называется *нормированным*; его плотность равна

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}. \qquad (23.41)$$

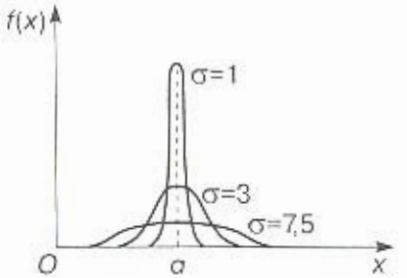

Рис. 23.6

Рассмотрим функцию нормального распределения как первообразную плотности распределения вероятностей. Для случая нормированного нормального распределения (23.41) она, согласно формуле (23.34), имеет вид

$$F(x) = \int_{-\infty}^{x} f(z)\,dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2}dz. \qquad (23.42)$$

Поскольку функция (23.41) является четной, то неопределенный интеграл от нее является нечетной функцией, и потому вместо функции распределения (23.42) используется функция Лапласа (см. п. 22.5)

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-z^2/2}dz. \qquad (23.43)$$

Функции (23.41) и (23.43) табулированы (см. Приложение).

График плотности нормального распределения (23.40) для разных значений σ показан на рис. 23.6.

Определение 4. *Модой* $M_o(X)$ называется возможное значение случайной величины X, при котором плотность распределения имеет максимум.

Определение 5. *Медианой* $M_e(X)$ называется такое возможное значение случайной величины X, что вертикальная прямая $x = M_e(X)$ делит пополам площадь, ограниченную кривой плотности распределения.

Нетрудно видеть, что график плотности нормального распределения симметричен относительно прямой $x = a$, и потому

и мода, и медиана в данном случае совпадают с математическим ожиданием:

$$M_o(X) = M_e(X) = M(X) = a.$$

Пусть случайная величина X задана плотностью нормального распределения (23.40), тогда вероятность того, что X примет значение на интервале (α, β), согласно формуле (23.33), равна

$$P(\alpha < X < \beta) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-(x-a)^2/2\sigma^2} dx.$$

Преобразование этой формулы путем введения новой переменной интегрирования $z = (x - a)/\sigma$ приводит к удобной вычислительной формуле:

$$P(\alpha < X < \beta) = \Phi\left((\beta - a)/\sigma\right) - \Phi\left((\alpha - a)/\sigma\right), \qquad (23.44)$$

где Φ — функция Лапласа, определенная по формуле (23.43).

Пример 3. Случайная величина распределена по нормальному закону с математическим ожиданием и средним квадратическим отклонением, соответственно равными 10 и 5. Найти вероятность того, что X примет значение на интервале $(20, 30)$.

Решение. Воспользуемся формулой (23.44). По условию $a = 10, \sigma = 5$, $\alpha = 20$ и $\beta = 30$. Следовательно,

$$P(10<X<30)=\Phi((30-10)/5)-\Phi((20-10)/5)=\Phi(4)-\Phi(2).$$

По табл. 2 Приложения находим соответствующие значения функции Лапласа и окончательно получаем

$$P(10 < X < 30) = 0{,}5 - 0{,}4772 = 0{,}0228.$$

Пример 4. Магазин производит продажу мужских костюмов. По данным статистики, распределение по размерам является нормальным с математическим ожиданием и средним квадратическим отклонением, соответственно равными 48 и 2. Определить процент спроса на 50-й размер при условии разброса значений этой величины в интервале (49,51).

Решение. По условию задачи $a = 48$, $\sigma = 2$, $\alpha = 49$, $\beta = 51$. Используя формулу (23.44), получаем, что вероятность спроса на 50-й размер в заданном интервале равна

$$P(49 < X < 50) = \Phi((51 - 48)/2) - \Phi((49 - 48)/2) =$$
$$= \Phi(1{,}5) - \Phi(0{,}5) = 0{,}4332 - 0{,}1915 = 0{,}2417.$$

Следовательно, спрос на 50-й размер костюмов составит около 24%, и магазину нужно предусмотреть это в общем объеме закупки.

Асимметрия и эксцесс

В прикладных задачах, например в математической статистике, при теоретическом изучении *эмпирических* распределений, отличающихся от нормального распределения, возникает необходимость количественных оценок этих различий. Для этой цели введены специальные безразмерные характеристики.

Определение 6. *Асимметрией* теоретического распределения называется отношение центрального момента третьего порядка к кубу среднего квадратического отклонения:

$$A_s = \mu_3/\sigma^3. \qquad (23.45)$$

Определение 7. *Эксцессом* теоретического распределения называется величина, определяемая равенством

$$E_k = \mu_4/\sigma^4 - 3, \qquad (23.46)$$

где μ_4 — центральный момент четвертого порядка.

Для нормального распределения $A_s = E_k = 0$. При отклонении от нормального распределения асимметрия положительна, если "длинная" и более пологая часть кривой распределения расположена справа от точки на оси абсцисс, соответствующей моде; если эта часть кривой расположена слева от моды, то асимметрия отрицательна (рис. 23.7,*а*,*б*).

Эксцесс характеризует "крутизну" подъема кривой распределения по сравнению с нормальной кривой: если эксцесс положителен, то кривая имеет более высокую и острую вершину; в случае отрицательного эксцесса сравниваемая кривая имеет более низкую и пологую вершину (рис. 23.7,*в*).

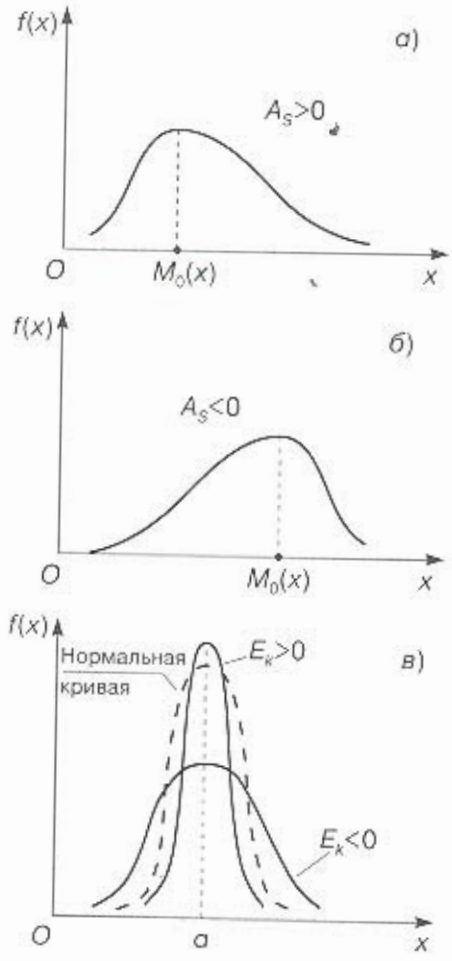

Рис. 23.7

Следует иметь в виду, что при использовании указанных характеристик сравнения опорными являются предположения об одинаковых величинах математического ожидания и дисперсии для нормального и теоретического распределений.

Пример 5. Пусть дискретная случайная величина X задана законом следующего распределения:

X	1	2	3	5	8
p	0,1	0,2	0,4	0,2	0,1.

Найти асимметрию и эксцесс теоретического распределения.

Решение. Найдем сначала математическое ожидание случайной величины:

$$M(X) = \nu_1 = 1 \cdot 0{,}1 + 2 \cdot 0{,}2 + 3 \cdot 0{,}4 + \\ +5 \cdot 0{,}2 + 8 \cdot 0{,}1 = 3{,}5.$$

Затем вычисляем начальные и центральные моменты 2, 3 и 4-го порядков и среднее квадратическое отклонение (см. формулы (23.27)–(23.31)):

$$\nu_2 = M(X^2) = 1 \cdot 0{,}1 + 4 \cdot 0{,}2 + 9 \cdot 0{,}4 + 25 \cdot 0{,}2 + 64 \cdot 0{,}1 = 15{,}9;$$

$$\nu_3 = M(X^3) = 1 \cdot 0{,}1 + 8 \cdot 0{,}2 + 27 \cdot 0{,}4 + 125 \cdot 0{,}2 + 512 \cdot 0{,}1 = 88{,}7;$$

$$\nu_4 = M(X^4) = 1 \cdot 0{,}1 + 16 \cdot 0{,}2 + 81 \cdot 0{,}4 + 625 \cdot 0{,}2 + 4096 \cdot 0{,}1 = 570{,}3;$$

$$D(X) = \mu_2 = \nu_2 - \nu_1^2 = 15{,}9 - 12{,}25 = 3{,}65;$$

$$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 = 88{,}7 - 3 \cdot 15{,}9 \cdot 3{,}5 + 2 \cdot 3{,}5^3 = 7{,}5;$$

$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4 = 570{,}3 - 4 \cdot 88{,}7 \cdot 3{,}5 + \\ +6 \cdot 15{,}9 \cdot 3{,}5^2 - 3 \cdot 3{,}5^4 = 570{,}3 - 1241{,}8 + 1168{,}65 - 450{,}19 = 46{,}96;$$

$$\sigma(X) = \sqrt{D(X)} = 1{,}91.$$

Теперь по формулам (23.45) и (23.46) находим искомые величины:

$$A_s = \mu_3/\sigma^3 = 7{,}5/1{,}91^3 = 1{,}076;$$

$$E_k = \mu_4/\sigma^4 - 3 = 40{,}96/1{,}91^4 - 3 = 0{,}53.$$

В данном случае "длинная" часть кривой распределения расположена справа от моды, причем сама кривая является несколько более островершинной, чем нормальная кривая с теми же величинами математического ожидания и дисперсии.

23.6. Некоторые элементы математической статистики

Задачи математической статистики

Первой задачей математической статистики является указание методов сбора и группировки статистических сведений, которые получены в результате экспериментов или наблюде-

ний. Вторая задача — это разработка методов анализа статистических данных: оценки неизвестных вероятности события, а также функции и параметров распределения; оценка зависимости случайной величины от других случайных величин; проверка статистических гипотез о виде и величинах параметров неизвестного распределения. Рассмотрим некоторые из этих вопросов.

Выборки

На практике сплошное исследование (каждого объекта из интересующей нас совокупности) проводят крайне редко. К тому же если эта совокупность содержит большое число объектов или исследование объекта требует нарушения его функционального стандарта, то сплошное исследование нереально. В таких случаях из всей совокупности случайно отбирают ограниченное число объектов и подвергают их исследованию.

Введем основные понятия, связанные с выборками. *Генеральной совокупностью* называется совокупность объектов, из которых производится выборка. *Выборочной* совокупностью (*выборкой*) называется совокупность случайно отобранных объектов из генеральной совокупности. Число объектов в совокупности называется ее *объемом*.

Пример 1. Пусть из 2000 изделий отобрано для обследования 100 изделий. Тогда объем генеральной совокупности $N = 2000$, а объем выборки $n = 100$.

Выборку можно осуществлять двумя способами. Если после исследования объект из выборки возвращается в генеральную совокупность, то такая выборка называется *повторной*; если объект не возвращается в генеральную совокупность, то выборка называется *бесповторной*.

Выборка называется *репрезентативной* (представительной), если по ее данным можно достаточно уверенно судить об интересующем признаке генеральной совокупности.

Способы отбора

Различают два способа отбора: без расчленения генеральной совокупности на части и с расчленением. К первому относятся *простые случайные* отборы (либо повторный, либо бесповторный), когда объекты извлекают по одному из всей генеральной совокупности; такой отбор можно производить с использованием таблицы случайных чисел.

Второй способ отбора включает в себя следующие разновидности, соответствующие способам расчленения генеральной совокупности. Отбор, при котором объекты отбираются из каждой "типической" части генеральной совокупности, называется *типическим*. Например, отбор деталей из продукции каждого станка, а не из их общего количества является типическим. Если генеральную совокупность делят на число групп, равное объему выборки, с последующим отбором из каждой группы по одному объекту, то такой отбор называется *механическим*. *Серийным* называется отбор, при котором объекты отбираются не по одному, а сериями; этот способ используется, когда исследуемый признак имеет незначительные колебания в различных сериях.

На практике часто употребляется комбинирование указанных выше способов отбора. Например, генеральную совокупность разбивают на серии одинакового объема, затем случайным образом отбирают несколько серий и в завершение случайным извлечением отдельных объектов составляют выборку. Конкретная комбинация способов отбора объектов из генеральной совокупности определяется требованием репрезентативности выборки.

Статистическое распределение выборки

Пусть из генеральной совокупности извлечена выборка объема n, в которой значение x_1 некоторого исследуемого признака X наблюдалось n_1 раз, значение x_2 — n_2 раз, ..., значение x_k — n_k раз. Значения x_i называются *вариантами*, а их последовательность, записанная в возрастающем порядке,— *вариационным рядом*. Числа n_i называются *частотами*, а их отношения к объему выборки

$$W_i = n_i/n \tag{23.47}$$

— *относительными частотами*. При этом $\sum n_i = n$. *Модой* M_o называется варианта, имеющая наибольшую частоту. *Медианой* m_e называется варианта, которая делит вариационный ряд на две части с одинаковым числом вариант в каждой. Если

число вариант нечетно, т.е. $k = 2l + 1$, то $m_e = x_{l+1}$; если же число вариант четно ($k = 2l$), то $m_e = (x_l + x_{l+1})/2$. *Размахом варьирования* называется разность между максимальной и минимальной вариантами или длина интервала, которому принадлежат все варианты выборки:

$$R = x_{\max} - x_{\min}. \tag{23.48}$$

Перечень вариант и соответствующих им частот называется *статистическим распределением выборки*. Здесь имеется аналогия с законом распределения случайной величины: в теории вероятностей — это соответствие между возможными значениями случайной величины и их вероятностями, а в математической статистике — это соответствие между наблюдаемыми вариантами и их частотами (относительными частотами). Нетрудно видеть, что сумма относительных частот равна единице: $\sum W_i = 1$.

Пример 2. Выборка задана в виде распределения частот:

x_i	4	7	8	12	17
n_i	2	4	5	6	3.

Найти распределение относительных частот и основные характеристики вариационного ряда.

Решение. Найдем объем выборки: $n = 2+4+5+6+3 = 20$. Относительные частоты соответственно равны $W_1 = 2/20 = 0{,}1$; $W_2 = 4/20 = 0{,}2$; $W_3 = 5/20 = 0{,}25$; $W_4 = 6/20 = 0{,}3$; $W_5 = 3/20 = 0{,}15$. Контроль: $0{,}1 + 0{,}2 + 0{,}25 + 0{,}3 + 0{,}15 = 1$. Искомое распределение относительных частот имеет вид

x_i	4	7	8	12	17
W_i	0,1	0,2	0,25	0,3	0,15.

Мода этого вариационного ряда равна 12. Число вариант в данном случае нечетно: $k = 2 \cdot 2 + 1$, поэтому медиана $m_e = x_3 = 8$. Размах варьирования, согласно формуле (23.48), $R = 17 - 4 = 13$.

Эмпирическая функция распределения

Пусть n_x — число наблюдений, при которых значение признака X меньше x. При объеме выборки, равном n, относительная частота события $X < x$ равна n_x/n.

Определение 8. Функция

$$F^*(x) = n/n, \tag{23.49}$$

определяющая для каждого значения x относительную частоту события $X < x$, называется *эмпирической функцией распределения*, или функцией распределения выборки.

В отличие от эмпирической функции распределения $F^*(x)$ выборки функция распределения $F(x)$ генеральной совокупности называется *теоретической функцией распределения*. Различие между ними состоит в том, что функция $F(x)$ определяет вероятность события $X < x$, а $F^*(x)$ — относительную частоту этого события. Из теоретических результатов общей теории вероятностей (закон больших чисел) следует, что при больших n вероятность отличия этих функций друг от друга близка к единице:

$$\lim_{n\to\infty} P[|F(x) - F^*(x)| < \varepsilon] = 1, \quad \varepsilon > 0. \tag{23.50}$$

Нетрудно видеть, что $F^*(x)$ обладает всеми свойствами $F(x)$, что вытекает из ее определения (23.49):

1) значения $F^*(x)$ принадлежат отрезку $[0, 1]$;

2) $F^*(x)$ является неубывающей функцией;

3) если x_1 — наименьшая варианта, то $F^*(x) = 0$ при $x \leqslant x_1$; если x_k — максимальная варианта, то $F^*(x) = 1$ при $x > x_k$.

Сама же функция $F^*(x)$ служит для оценки теоретической функции распределения $F(x)$ генеральной совокупности.

Пример 3. Построить эмпирическую функцию по заданному распределению выборки:

x_i	2	4	6
n_i	10	15	25.

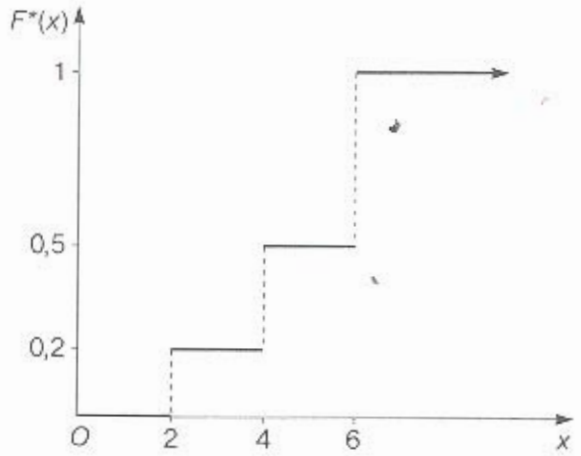

Рис. 23.8

РЕШЕНИЕ. Находим объем выборки: $n = 10 + 15 + 25 = 50$. Наименьшая варианта равна 2, поэтому $F^*(x) = 0$ при $x \leqslant 2$. Значение $X < 4$ (или $x_1 = 2$) наблюдалось 10 раз, значит, $F^*(x) = 10/50 = 0{,}2$ при $2 < x < 4$. Значения $X < 6$ (а именно $x_1 = 2$ и $x_2 = 4$) наблюдались $10 + 15 = 25$ раз, значит, при $4 < x < 6$ функция $F^*(x) = 25/50 = 0{,}5$. Поскольку $x = 6$ — максимальная варианта, то $F^*(x) = 1$ при $x > 6$. Напишем формулу искомой эмпирической функции:

$$F^*(x) = \begin{cases} 0, & x \leqslant 2; \\ 0{,}2, & 2 < x \leqslant 4; \\ 0{,}5, & 4 < x \leqslant 6; \\ 1, & x > 6. \end{cases}$$

График этой функции показан на рис. 23.8.

Полигон и гистограмма

Каждую пару значений (x_i, n_i) из распределения выборки можно трактовать как точку на координатной плоскости. Точно так же можно рассматривать и пары значений (x_i, W_i) относительного распределения выборки. Ломаная, отрезки которой соединяют точки (x_i, n_i), называется *полигоном частот*. Ломаная, соединяющая на координатной плоскости точки (x_i, W_i), называется *полигоном относительных частот*. На

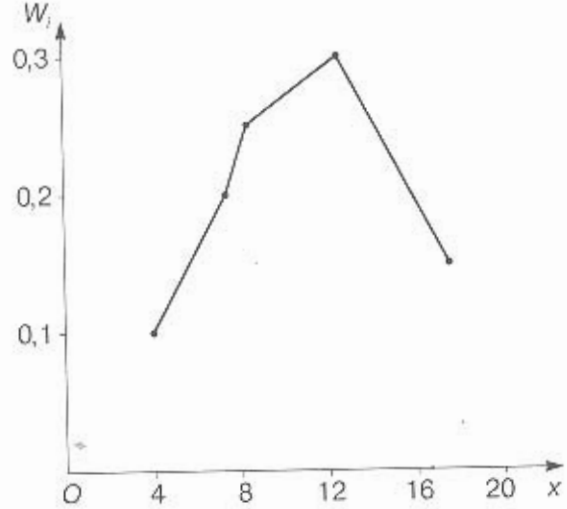

Рис. 23.9

рис. 23.9 показан полигон относительных частот для распределения, приведенного в примере 2.

Для случая непрерывного признака X удобно разбить интервал $(x_{\min}, x_{\max})$ его наблюдаемых значений на несколько частичных интервалов длиной h каждый и найти для каждого из этих интервалов сумму частот n_j, попавших в него. Ступенчатая фигура, состоящая из прямоугольников с основаниями длиной h и высотами n_j/h (плотность частоты), называется *гистограммой* частот. Геометрический смысл гистограммы: нетрудно видеть, что площадь ее равна сумме всех частот или объему выборки. На рис. 23.10 изображена гистограмма объема $n = 100$.

Аналогичным образом определяется и *гистограмма относительных частот*; в этом случае высоты прямоугольников, составляющих ступенчатую фигуру, определяются отношениями сумм относительных частот, попадающих в интервал $(x_{\min} + (j-1)h, x_{\min} + jh)$, к длине интервала h, т.е. величинами W_j/h. Нетрудно видеть, что площадь гистограммы относительных частот равна единице (сумме относительных частот выборки).

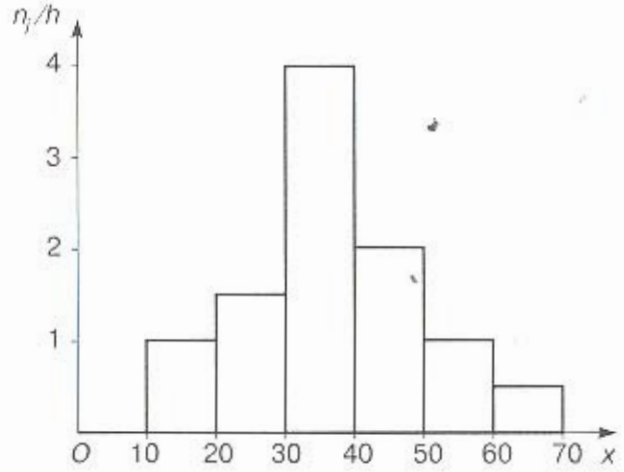

Рис. 23.10

Статистические оценки параметров распределения

Значения количественного признака x_1, x_2, ..., x_k в выборке можно рассматривать как независимые случайные величины. В таком случае нахождение статистической оценки неизвестного параметра теоретического распределения означает отыскание функции от наблюдаемых случайных величин, которая и даст нам приближенное значение искомого параметра. Укажем виды статистических оценок.

Несмещенной называется статистическая оценка ϑ_*, математическое ожидание которой равно оцениваемому параметру ϑ при любой выборке:

$$M(\vartheta_*) = \vartheta. \tag{23.51}$$

Смещенной называется оценка, при которой условие (23.51) не выполнено. *Эффективной* называется оценка, которая имеет минимальную дисперсию при заданном объеме выборки n. *Состоятельной* называется статистическая оценка типа (23.50), которая при $n > \infty$ стремится по вероятности к оцениваемому параметру.

Теперь укажем виды числовых характеристик оценок. Прежде всего, это средние. *Генеральная средняя* для изучаемого количественного признака X по генеральной совокупности

$$\bar{x}_\text{г} = (x_1 + x_2 + \ldots + x_N)/N$$

и *выборочная средняя*

$$\bar{x}_\text{в} = (x_1 + x_2 + \ldots + x_n)/n.$$

Если значения признака x_1, x_2, ..., x_k в выборке имеют соответственно частоты n_1, n_2, ..., n_k, то последнюю формулу можно переписать в виде

$$\bar{x}_\text{в} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i. \tag{23.52}$$

Можно показать, что выборочная средняя (23.52) является несмещенной оценкой; это аналог математического ожидания случайной величины. *Общая средняя* — это средняя арифметическая выборочных.

Введем в рассмотрение величины, характеризующие отклонение значений количественного признака X от своего среднего значения. Это *генеральная дисперсия*:

$$D_\text{г} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}_\text{г})^2$$

и *выборочная дисперсия*:

$$D_\text{в} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_\text{в})^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x}_\text{в})^2. \tag{23.53}$$

Можно показать, что для вычисления этих характеристик справедливы более удобные формулы, аналогичные дисперсии случайной величины; так, формула (23.53) принимает вид

$$D_\text{в} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - (\bar{x}_\text{в})^2. \tag{23.54}$$

Генеральное среднее квадратическое отклонение определяется как

$$\sigma_\text{г} = \sqrt{D_\text{г}}.$$

Аналогично вводится и *выборочное среднее квадратическое отклонение*

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}}. \qquad (23.55)$$

Пример 4. Выборка задана таблицей распределения

x_i	1	2	3	5
n_i	15	20	10	5.

Найти выборочные характеристики: среднюю, дисперсию и среднее квадратическое отклонение.

Решение. По формуле (23.52) сначала находим $\bar{x}_{\text{в}}$:

$$\bar{x}_{\text{в}} = \frac{15 \cdot 1 + 20 \cdot 2 + 10 \cdot 3 + 5 \cdot 5}{15 + 20 + 10 + 5} = \frac{110}{50} = 2{,}2.$$

Затем по формулам (23.54) и (23.55) находим две другие искомые величины:

$$D_{\text{в}} = (15 \cdot 1 + 20 \cdot 4 + 10 \cdot 9 + 5 \cdot 25)/50 - 2{,}2^2 = 6{,}2 - 4{,}84 = 1{,}36,$$

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}} = \sqrt{1{,}36} \approx 1{,}166.$$

Виды дисперсий

Часто значения количественного признака X совокупности разбиваются на определенное число групп. Каждую группу можно рассматривать как самостоятельную выборку, и для каждой группы можно определить групповую среднюю и дисперсию. Пусть r — число групп. *Групповой дисперсией* называется дисперсия значений признака в группе относительно групповой средней:

$$D_j = \frac{1}{N_j} \sum_{i=1}^{N_j} n_i (x_i - \bar{x}_j)^2, \; j = 1, 2, \ldots, r, \qquad (23.56)$$

где n_i — частота значения x_i в группе, j — номер группы, $\bar{x}_j$ — групповая средняя j-й группы, $N_j = \sum n_i$ — объем j-й группы.

Зная дисперсию каждой группы, можно определить их среднюю арифметическую. *Внутригрупповой дисперсией* называется средняя арифметическая дисперсий, где каждое слагаемое входит с весом объема группы:

$$D_{\text{внгр}} = \frac{1}{n} \sum_{j=1}^{r} N_j D_j. \qquad (23.57)$$

В свою очередь, зная для всех групп средние $\bar{x}_j$ и общую среднюю $\bar{x}$, введем еще одно понятие. *Межгрупповой дисперсией* называется дисперсия групповых средних относительно общей средней:

$$D_{\text{мгр}} = \frac{1}{n} \sum_{j=1}^{r} N_j (\bar{x}_j - \bar{x})^2, \qquad (23.58)$$

где $n = \sum\limits_{i=1}^{r} N_j$ — объем всей совокупности.

Для общей дисперсии всей совокупности справедлива следующая теорема, которая приводится здесь без доказательства.

Теорема 23.6. *Если совокупность состоит из нескольких групп, то общая дисперсия равна сумме внутригрупповой и межгрупповой дисперсий:*

$$D_{\text{общ}} = D_{\text{мгр}} + D_{\text{внгр}}, \qquad (23.59)$$

где слагаемые в правой части определяются соответственно формулами (23.57) *и* (23.58).

Поясним сказанное в этом пункте на примере.

Пример 5. Совокупность состоит из двух следующих групп:

Первая группа				Вторая группа		
x_i	2	4	5	x_i	4	9
n_i	1	7	2	n_i	3	2

Найти групповые, внутригрупповую, межгрупповую и общую дисперсии.

Решение. Объемы групп соответственно равны N_1=10 и N_2=5. Общий объем совокупности: $n = 10 + 5 = 15$. Найдем групповые средние:

$$\bar{x}_1 = (1 \cdot 2 + 7 \cdot 4 + 2 \cdot 5)/10 = 4; \; \bar{x}_2 = (3 \cdot 4 + 2 \cdot 9)/5 = 6.$$

Теперь находим групповые дисперсии по формуле (23.56):

$$D_1 = (1\cdot(2-4)^2 + 7\cdot(4-4)^2 + 2\cdot(5-4)^2)/10 = 0{,}6;$$

$$D_2 = (3\cdot(4-6)^2 + 2\cdot(9-6)^2)/5 = 6.$$

Внутригрупповая дисперсия, согласно формуле (23.57), равна:

$$D_{\text{внгр}} = (10\cdot 0{,}6 + 5\cdot 6)/15 = 2{,}4.$$

Теперь найдем межгрупповую дисперсию по формуле (23.58), для чего сначала определим общую среднюю:

$$\bar{x} = (1\cdot 2 + 7\cdot 4 + 2\cdot 5 + 3\cdot 4 + 2\cdot 9)/15 = 70/15 = 14/3;$$

$$D_{\text{мгр}} = (10(4-14/3)^2 + 5(6-14/3)^2)/15 = 8/9.$$

Наконец, общая дисперсия, согласно формуле (23.59), равна:

$$D_{\text{общ}} = 2{,}4 + 8/9 \approx 3{,}29.$$

Эмпирические моменты

Для вычисления сводных характеристик выборок используют эмпирические моменты, аналогичные соответствующим теоретическим моментам. *Обычным эмпирическим моментом порядка s* называется среднее значение s-х степеней разностей $x_i - C$, где x_i — наблюдаемая варианта, C — произвольная постоянная (ложный нуль — либо мода, либо любая варианта, расположенная примерно в середине вариационного ряда):

$$M'_s = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - C)^s. \qquad (23.60)$$

При $C = 0$ имеем *начальные эмпирические моменты порядка s*; в частности, в случае $s = 1$

$$M_1 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \bar{x}_{\text{в}}.$$

Центральным эмпирическим моментом порядка s называется обычный момент (23.60) при $C = \bar{x}_{\text{в}}$:

$$m_s = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - x_{\text{в}})^s. \qquad (23.61)$$

В частности,

$$m_2 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - x_{\text{в}})^2 = D_{\text{в}}.$$

Иными словами, выборочная дисперсия равна центральному эмпирическому моменту второго порядка. Центральные моменты выражаются через обычные по формулам, полностью аналогичным (23.19) и (23.20).

Асимметрия и эксцесс эмпирического распределения

Нормальное распределение является одним из самых распространенных в применениях математической статистики. Для оценки отклонения эмпирического распределения от нормального используют характеристики, аналогичные для теоретического распределения (см. предыдущий раздел 23.6).

Асимметрия эмпирического распределения определяется следующим равенством:

$$a_s = m_3/\sigma_{\text{в}}^3. \qquad (23.62)$$

Эксцесс эмпирического распределения определяется следующим равенством:

$$e_k = m_4/\sigma_{\text{в}}^4 - 3. \qquad (23.63)$$

В формулы (23.62) и (23.63) входят центральные эмпирические моменты, определяемые формулами (23.61), а также выборочное среднее квадратическое отклонение (23.55).

Пример 6. Найти асимметрию и эксцесс эмпирического распределения:

варианта	1	2	3	4	5	6	10
частота	5	10	15	35	16	15	4.

Решение. Найдем сначала $\bar{x}_{\text{в}}$ и $\sigma_{\text{в}}$ с использованием формул (23.52)–(23.55):

$$\bar{x}_{\text{в}} = \frac{5\cdot 1+10\cdot 2+15\cdot 3+35\cdot 4+16\cdot 5+15\cdot 6+4\cdot 10}{5+10+15+35+16+15+4} = 4{,}2;$$

$$D_{\text{в}}=\frac{5\cdot1+10\cdot4+15\cdot9+35\cdot16+16\cdot25+15\cdot36+4\cdot100}{100}-4{,}2^2=3{,}16;$$
$$\sigma_{\text{в}}=1{,}787.$$

Далее, используя формулы (23.61), определяем центральные эмпирические моменты третьего и четвертого порядков:

$$m_3=(5(-3{,}2)^3+10(-2{,}2)^3+15(-1{,}2)^3+35(-0{,}2)^3+16\cdot0{,}8^3+$$
$$+15\cdot1{,}8^3+4\cdot5{,}8^3)/100=579{,}6/100=5{,}796;$$

$$m_4=(5(-3{,}2)^4+10(-2{,}2)^4+15(-1{,}2)^4+35(-0{,}2)^4+16\cdot0{,}8^4+$$
$$+15\cdot1{,}8^4+4\cdot5{,}8^4)/100=5480{,}32/100=54{,}8032.$$

Затем по формулам (23.62) и (23.63) находим искомые величины:

$$a_s=5{,}796/1{,}787^3=1{,}016;$$
$$e_k=54{,}8032/1{,}787^4=5{,}374.$$

В заключение отметим, что все оценки, приведенные выше, определяются одним числом, т.е. являются *точечными*. При малых объемах выборки точечная оценка может приводить к большим ошибкам и значительно отличаться от оцениваемого параметра.

Часть 5. ОСНОВНЫЕ АСПЕКТЫ ЭКОНОМИКО-МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Глава 24
ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МОДЕЛИ

Решение современных экономических проблем и анализ экономической ситуации невозможен без использования в той или иной мере математических моделей. Сама по себе формулировка основного принципа экономической деятельности — максимум эффективности при минимуме затрат ресурсов — подразумевает активное участие математического моделирования в обработке обширных баз данных, в оперативной оценке ситуации, в прогнозировании деятельности фирмы.

Математическую модель можно определить как внутренне непротиворечивую замкнутую систему математических соотношений (объект конечной сложности), предназначенную для воспроизведения определенного качества (или группы определенных качеств) изучаемого реального явления или процесса. Математические модели представляют собой основу компьютерного моделирования и обработки информации. Они развивают наши представления о закономерностях экономических процессов и способствуют формированию образа мышления и анализа на новом, более высоком уровне.

В последнее время, для обозначения специфичности класса математических моделей, используемых в экономике, используют термин "экономико-математическое моделирование". И это не случайно, поскольку экономическая теория давно уже использует элементы математики в своих выводах. Более того, настоятельность решения актуальных экономических проблем часто инициирует и развитие математического аппарата

(например, появление класса продуктивных матриц в линейной алгебре обусловлено исследованием моделей межотраслевого баланса; математическое программирование в своей основе имеет сугубо экономичекий аспект оптимального планирования распределения ограниченных ресурсов).

Простейшие математические модели на уровне таблиц и формул использовались еще Ф. Кене в 1758 г. ("Экономические таблицы"), А. Смитом (классическая макроэкономическая модель), Д. Рикардо (модель международной торговли). В XIX веке усилиями Л. Вальраса, О. Курно, В. Парето, Ф. Эджворта и др. были сформированы основы математического подхода к исследованию рыночной экономики. В XX веке математическое моделирование экономических процессов получило бурное развитие; достаточно упомянуть Нобелевских лауреатов в области экономики — В. Леонтьева, Д. Хикса, Р. Солоу, П. Самуэльсона. Значительный вклад в развитие экономико-математических моделей внесли отечественные исследователи: Е. Е. Слуцкий, В. С. Новожилов, Л. В. Канторович.

24.1. Предназначение модели

Развитие микро- и макроэкономики и прикладных экономических дисциплин связано со все возрастающим уровнем их формализации и наличием высокой степени абстракции. Основа для этого была заложена прогрессом в области прикладной математики в XX веке: развитием теории игр, математической статистики, математического программирования. Использование математики в экономических приложениях, сформировавшее область экономико-математического моделирования, позволяет указать следующие основные аспекты:

— выделить и формально описать наиболее существенные связи экономических объектов;

— из полученных модельных соотношений путем обработки базы данных исходной информации получать дедуктивным методом выводы, адекватные исследуемому объекту в пределах функциональной надежности модели;

— получить новые знания об объекте и зависимостях входящих в него формализованных параметрах;

— компактно формулировать основные положения и выводы экономической теории;

— разрабатывать стратегии управления экономическими объектами и поведения фирмы в условиях рынка.

Укажем логическую цепочку достаточно общих принципов построения математических моделей.

1. Формулировка предмета и цели исследования реального объекта. Таким объектом выступает некоторая совокупность каких-либо качеств исследуемого явления или процесса.

2. Выделение в экономическом объекте наиболее важных структурных и функциональных элементов и их характеристик.

3. Формализация определяющих элементов экономического объекта и их взаимосвязей.

4. Определение вида исходной информации (входные параметры модели) и выходной информации (расчетные параметры модели).

5. Постановка задачи — создание основы математической модели — получение замкнутой и внутренне непротиворечивой совокупности математических соотношений, предназначенных для описания исследуемого экономического объекта через расчетные переменные. В информационном аспекте модель является оператором отражения информационного поля реального объекта в конечную совокупность расчетных информационных признаков. Выбор этого оператора зависит от автора модели.

6. Определение функциональной надежности модели — установление области ее адекватности исследуемому объекту.

Следует особо отметить отличие математической модели от математической задачи. Поясним это на примере. Пусть X — область исходной информации модели, Y — область прогнозной информации модели, а F — оператор пересчета множества X в множество Y. Тогда формула

$$Y = F(X) \tag{24.1}$$

определяет модель переработки (отражения) множества X в множество Y. Если говорить о математической задаче, то допустимая область X определяется только законностью выполнения математических операций, входящих в оператор F. Для математической модели еще требуются и ограничения естественного характера как на множество X, так и на множество Y: последнее ограничение означает установление области адекватности. Например, удлинение ΔL металлического стержня

при его нагреве на температуру ΔT подсчитывается по формуле

$$\Delta L = \alpha \Delta T, \qquad (24.2)$$

где α — коэффициент относительного удлинения материала стержня (физическая константа). С точки зрения математической задачи уравнение (24.2) представляет собой линейную функцию с неограниченными областями определения и изменения. Однако с точки зрения математической модели реального физического процесса, во-первых, нет таких материалов, которые могли бы выдержать нагрев, скажем, на 10 тысяч градусов, т.е. формула (24.1) приемлема лишь в определенном диапазоне нагрева для данного материала; во-вторых, вне этой зоны процесс удлинения является уже нелинейным и нагрев приводит к совершенно другим физическим процессам в металлическом стержне.

Поскольку модель создается для получения определенной прогнозной информации, то в указанную выше логическую цепочку 1-6 следует добавить еще несколько этапов, имеющих отношение к процессу моделирования.

7. Подбор оптимального метода решения математической задачи, составляющей основу модели (в том числе и выбор вычислительной схемы решения задачи).

8. Выполнение прогнозного этапа моделирования — "проигрывание" на модели различных сценариев (сочетаний исходных параметров модели) как проведение многовариантных расчетов с целью создания базы расчетной информации, как количественного образа исследуемого объекта.

Поясним несколько подробнее суть двух последних этапов. Погрешность математического моделирования, как меру отклонения модели от реального объекта, можно упрощенно представить в виде суммы:

$$\sigma = \sigma_m + \sigma_c + \sigma_i, \qquad (24.3)$$

где σ_m — погрешность собственно модели, σ_c — погрешность вычислительной схемы, σ_i — погрешность в исходной информации.

Погрешность σ_m обусловлена упрощениями, которые имеют место при переходе от реального объекта бесконечной сложности к математической модели — объекту конечной сложности. В этом плане модель — это как бы фотоснимок, она всегда "беднее" оригинала. Говоря языком теории множеств, если M и R — соответственно множества информационных признаков модели и реального объекта, то модель всегда представляет собой отображение множества R в множество M. Отображение "в" (или гомоморфизм) несет в себе потерю информативности в отличие от отображения "на" (или изоморфизм), которое является взаимно однозначным.

Погрешность σ_c обусловлена вычислительной реализацией математической модели: упрощения при переходе от исходных уравнений модели к разностным, особенности вычислительных схем, конечная разрядная сетка компьютера и др. Например, при решении систем линейных алгебраических уравнений высокого порядка (задачи математического программирования, задачи межотраслевого баланса, численное решение систем дифференциальных и разностных уравнений) накопление вычислительной погрешности только за счет конечного представления чисел в разрядной сетке компьютера может существенно искажать расчетные результаты. Именно поэтому для многих моделей этап 7 в упомянутой выше схеме является существенным и важным.

Погрешность σ_i характеризует детальность задания исходной информации. Как правило, эта величина обусловлена рядом причин: трудности при проведении эксперимента на объекте исследования, несовершенство самого экспериментального исследования, статистическая неопределенность результатов измерений. Зачастую в силу объективных причин определение входных параметров модели имеет погрешность в десятки процентов и более, а иногда часть параметров неизвестна вообще (например, в случае распределенных параметров — когда коэффициенты уравнений представляют собой функциональные зависимости). В таком случае модель является некорректно сформулированной, с информационными "дырами". Как правило, при удачно подобранной математической модели и хорошей схеме ее численной реализации основной вклад в погрешность моделирования, согласно формуле (24.3), вносит неточность входной информации. Это объективная реальность, и игнорирование этого факта невозможно.

Для устранения информационных "дыр" в моделях часто возникает необходимость подбора недостающей исходной информации. Вообще говоря, эта задача является некорректной с точки зрения математической классификации и требует применения сложного аппарата решения обратных задач. В практике моделирования обычно применяется более простая схема, когда путем многовариантных расчетов подбирается недостающая входная информация так, чтобы она наилучшим образом соответствовала некоторому известному объему выходной информации, а также некоторым априорным и экспертным оценкам. Процесс заполнения информационных "дыр" является частью общего процесса моделирования (этап 8). В свете этого представляется нецелесообразным выбор чрезмерно усложненной математической модели и численной схемы ее реализации повышенной точности: все равно ошибка моделирования будет определяться погрешностью исходной информации. Более того, требование оперативности моделирования при многовариантности расчетов делает предпочтительным выбор более простых модельных и вычислительных схем, тем более что даже незначительное повышение точности расчетов (уменьшение σ_m и σ_c) приводит к значительному увеличению сложности и громоздкости модели.

В свете сказанного можно сформулировать основные требования, которым должны удовлетворять математические модели.

А) Модель не должна быть чрезмерно сложной, так как это приводит к неоправданно большим затратам ресурсов при ее реализации. Следует соотносить сложность и детальность модели с уровнем достоверности исходной информации.

Б) Не следует строить модель всеобъемлющего прогноза реального объекта. Это приводит к чрезвычайно громоздким, необозримым и плохо анализируемым математическим моделям, которые к тому же могут оказаться еще и плохо обусловленными (неустойчивыми). Если возникает необходимость в прогнозе ряда разнородных качеств реального процесса, то целесообразно построить совокупность или иерархию соподчиненных относительно простых математических моделей.

В) Сложность модели должна соответствовать степени разработанности математического аппарата, а не превосходить

ее: в противном случае математическая модель будет неразрешимой.

Вообще говоря, нельзя сформулировать единые жесткие правила создания математических моделей и в этом плане можно согласиться с Е. С. Вентцель, что разработка моделей — это искусство [2]. Более того, у разных исследователей модели одного и того же процесса могут существенно отличаться, и потому целесообразна конкуренция или "спор" моделей как способ их селекции.

24.2. Классификация экономико-математических моделей

В экономико-математическом моделировании установлена следующая классификация по основным типам моделей. Модели разделяются на классы по ряду признаков, относящихся к особенностям моделируемых объектов, целям моделирования и используемого инструментария.

Макроэкономические модели описывают экономику как единое целое со связями между агрегированными материальными и финансовыми показателями (ВВП, потребление, инвестиции, занятость, денежная масса, государственный долг, инфляция и др.).

Микроэкономические модели описывают взаимодействие структурных и функциональных составляющих экономики либо их поведение в отдельности в рыночной среде.

Теоретические модели являются аппаратом изучения общих свойств экономики и ее составляющих на основе дедукции выводов из формальных предпосылок.

Прикладные модели представляют собой аппарат оценок параметров конкретных экономических объектов, выработки рекомендаций для принятия экономических решений и разработки стратегий поведения фирм на рынке.

Равновесные модели описывают такие состояния экономики, когда результирующая всех воздействий на нее равна нулю. Как правило, равновесные модели являются дескриптивными, описательными.

Оптимизационные модели используются в теории рыночной экономики на микроуровне (оптимизация деятельности потребителя, производителя или фирмы). На макроуровне ре-

40 Зак. 6853

зультат выбора экономическими субъектами рационального поведения может приводить к состоянию относительного равновесия.

Статические модели описывают состояние экономических объектов в определенный момент или осредненно за некоторый период времени. При этом все параметры статических моделей полагаются фиксированными величинами, независящими от времени.

Динамические модели включают зависимость и взаимосвязи переменных модели во времени. Они используют обычно аппарат дифференциальных и разностных уравнений и вариационного исчисления, где независимой переменной является время.

Детерминированные модели предполагают в своей основе только жесткие функциональные связи между переменными модели.

Стохастические модели допускают наличие случайных связей между переменными модели. Эти модели используют аппарат теории вероятностей и математической статистики.

Модели с элементами неопределенности используются для моделирования ситуаций, когда для определяющих факторов невозможно собрать статистические данные, и их значения не определены. В этих моделях используются аппараты теории игр и имитационного моделирования.

Глава 25
АППАРАТ ПРОИЗВОДСТВЕННЫХ ФУНКЦИЙ

25.1. Производственные функции

Понятие производственной функции имеет основополагающее значение в экономической теории. Под *производственной функцией* понимается такая функция, независимая переменная которой имеет смысл объема используемого ресурса (фактора), а зависимая переменная — объема выпускаемой продукции. В формуле

$$y = f(x) \tag{25.1}$$

используемый ресурс $x \geqslant 0$, объем выпускаемой продукции $y \geqslant 0$. Производственная функция вида (25.1) называется *однофакторной* или одноресурсной. Знак функции f рассматривается как характеристика *производственной системы*, преобразующей ресурс x в выпуск y.

Многофакторная производственная функция характеризуется функцией нескольких переменных

$$y = f(\overline{x}) = f(x_1, x_2, \ldots, x_n), \tag{25.2}$$

где независимые неотрицательные переменные x_1, x_2, ..., x_n имеют смысл объемов n используемых ресурсов, а значение функции $f(x_1, x_2, \ldots, x_n)$ имеет смысл величины объема выпуска. Областью определения многофакторной производственной функции по экономическому смыслу является множество n-мерных векторов с неотрицательными компонентами, а зависимая переменная $y \geqslant 0$ — скалярная величина.

Производственные функции (ПФ) реализуют схему "затраты — выпуск" и имеют различные области использования в экономике. На уровне микроэкономики в качестве производственной системы может выступать отдельная фирма, отрасль, межотраслевой комплекс, а сама ПФ используется для описания взаимосвязи между величиной ресурса, затрачиваемого за некоторый период времени, и итоговым выпуском продукции за этот же временной промежуток (например, год). В этом случае ПФ строятся и используются в основном для решения задач анализа и планирования, реже — для прогноза. На макроэкономическом уровне в роли производственной системы выступает регион и страна в целом; здесь ПФ строятся и используются для решения всех трех видов указанных задач.

Выбор единиц измерения затрачиваемых ресурсов и выпускаемой продукции различен для разных уровней: на микроэкономическом уровне затраты и выпуск могут измеряться и в натуральных, и в стоимостных единицах (показателях); на макроэкономическом уровне затраты и выпуск измеряются в основном в стоимостных показателях.

25.2. Экономический смысл производственной функции

Следует отметить, что в микро- и макроэкономике зависимая переменная y в производственной функции имеет разный смысл. В микроэкономической теории полагается, что y — это

максимально возможный объем выпуска продукции при объеме затраченного ресурса x. В макроэкономике полагается, что при другом распределении ресурса x между структурными единицами экономики выпуск может быть иным, т.е. под производственной функцией понимается статистически устойчивая связь между затратой ресурса и выпуском.

Вообще говоря, для производственной функции более корректной является развернутая запись

$$y = f(\overline{x}, \overline{a}), \tag{25.3}$$

где $\overline{x} = (x_1, x_2, \ldots, x_n)$ — вектор ресурсов (независимых переменных), $\overline{a}$ — вектор параметров ПФ.

Пример 1. Рассмотрим производственную функцию вида $y = ax^b$. Здесь x — величина затрачиваемого ресурса, $f(x)$ — объем выпускаемой продукции, a и b — параметры ПФ (положительные числа).

Производственная функция вида (25.1) или (25.2) называется *статической*, если сама функция и ее параметры не зависят от времени t. Производственная функция называется *динамической*, если выполнено хотя бы одно из следующих условий:

1) время t входит в функцию f в качестве независимой переменной (временно́й ресурс);

2) параметры производственной функции зависят от времени t.

В частности, в динамической ПФ можно учесть научно-технический прогресс путем введения в функциональную зависимость (25.1) множителя e^{pt}, где параметр $p > 0$ характеризует темп прироста выпуска вследствие НТП:

$$y(t) = e^{pt} f(x_1(t), x_2(t), \ldots, x_n(t)). \tag{25.4}$$

25.3. Основные виды производственных функций

Приведем два основных, наиболее употребимых вида производственной функции.

1. Линейная ПФ. Она имеет вид:

$$y = a_0 + a_1 x_1 + a_2 x_2, \quad a_i > 0 \quad (i = 0, 1, 2) \tag{25.5}$$

— двухфакторная линейная производственная функция. Многофакторная ПФ имеет вид:

$$y = a_0 + a_1 x_1 + \ldots + a_n x_n.$$

Линейная производственная функция принадлежит к классу *аддитивных* функций.

2. Мультипликативная функция. Двухфакторная функция имеет вид:

$$y = a_0 x_1^{a_1} x_2^{a_2}. \tag{25.6}$$

Переход от мультипликативной функции к аддитивной производится с помощью логарифмирования:

$$\ln y = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2.$$

Обозначая $\ln y = w$, $\ln a_0 = b$, $\ln x_1 = u$, $\ln x_2 = v$, получаем аддитивную ПФ

$$w = b + a_1 u + a_2 v.$$

Из мультипликативных производственных функций в микро- и макроэкономике наиболее часто используется двухфакторная производственная функция Кобба–Дугласа (так она названа по именам американских ученых, предложившим ее в 1929 г. на основе обработки большого числа статистических данных). В приложениях этой функции одним из ресурсов является объем используемого капитала K, другим — затраты труда L, т.е. функция Кобба–Дугласа имеет вид:

$$Y = a_0 K^{\alpha_1} L^{\alpha_2}. \tag{25.7}$$

Функция Кобба–Дугласа широко используется в микро- и макроэкономических приложениях благодаря своей структурной простоте. Согласно статистической обработке экономических данных, проводившейся различными авторами, наблюдаются следующие закономерности:

$$\alpha_1 < 1, \ \alpha_2 < 1, \ \alpha_1 < \alpha_2, \ \alpha_1 + \alpha_2 \approx 1. \tag{25.8}$$

25.4. Формальные свойства производственных функций

Приведем свойства производственной функции на примере двухфакторной функции $f(x_1, x_2)$.

1. $f(0, x_2) = f(x_1, 0) = 0$.

Свойство 1 означает, что при отсутствии хотя бы одного ресурса нет выпуска продукции.

2. При $x_1' > x_1''$ $\quad f(x_1', x_2) > f(x_1'', x_2)$;
аналогично: при $x_2' > x_2''$ $\quad f(x_1, x_2') > f(x_1, x_2'')$.

Свойство 2 означает, что с увеличением объема использования любого ресурса объем выпуска растет.

3. При $\overline{x} > \overline{0}$ $\quad \partial f/\partial x_i > 0$, $i = 1, 2$.

Это свойство является следствием свойства 2: первые частные производные производственной функции положительны — это означает, что с ростом потребления одного из ресурсов при неизменном объеме другого ресурса объем выпуска возрастает. Здесь и далее условная запись $\overline{x} > \overline{0}$ означает, что обе компоненты вектора $\overline{x} = (x_1, x_2)$ строго положительны.

4. При $\overline{x} > \overline{0}$ $\quad \partial^2 f/\partial x_i^2 \leqslant 0$, $i = 1, 2$.

Свойство 4 (вторая производная производственной функции по любой координате неположительна) означает, что с ростом объема затрат одного из ресурсов при неизменном объеме использования другого ресурса величина прироста выпуска на каждую дополнительную единицу i-го ресурса не увеличивается (это свойство известно как закон убывающей эффективности).

5. При $\overline{x} > \overline{0}$ $\quad \partial^2 f/(\partial x_1 \partial x_2) \geqslant 0$.

Это свойство означает, что с ростом затрат одного из ресурсов предельная эффективность другого ресурса возрастает.

6. $f(tx_1, tx_2) = t^p f(x_1, x_2)$.

Свойство 6 означает, что производственная функция является однородной функцией степени p. Иными словами, при переходе от вектора затрат ресурсов $\overline{x}$ к вектору $t\overline{x}$ объем выпуска изменяется в t^p раз. При $p > 1$ имеем рост выпуска в t^p раз с ростом масштаба производства в t раз; при $p < 1$ имеем снижение выпуска в t^p раз с ростом масштаба производства в t раз. При $p = 1$ имеем постоянную эффективность производства независимо от роста его масштаба.

25.5. Средние и предельные значения производственной функции

Рассмотрим многофакторную производственную функцию (25.2)

$$y = f(\overline{x}) = f(x_1, x_2, \ldots, x_n).$$

1. *Средней производительностью* i-го ресурса, или средним выпуском по i-му ресурсу, называется величина

$$A_i = f(\overline{x})/x_i, \quad i = 1, 2, \ldots, n. \tag{25.9}$$

2. *Предельной (маржинальной) производительностью* i-го ресурса, или предельным выпуском по i-му ресурсу, называется первая частная производная

$$M_i = \partial f/\partial x_i, \quad i = 1, 2, \ldots, n. \tag{25.10}$$

В приращениях функции и аргумента частную производную можно приближенно представить в виде

$$M_i \approx \Delta_i f(\overline{x})/\Delta x_i, \quad i = 1, 2, \ldots, n.$$

Следовательно, предельная производительность i-го фактора производства приближенно показывает, на сколько единиц увеличится объем выпуска при росте объема затрат i-го ресурса на одну единицу.

3. Отношение предельной производительности i-го ресурса к его средней производительности

$$E_i = M_i/A_i = (x_i/f(\overline{x}))(\partial f/\partial x_i), \quad i = 1, 2, \ldots, n \tag{25.11}$$

называется *эластичностью выпуска* по i-му ресурсу (частной эластичностью выпуска). В экономической теории для эластичности часто используется разностный аналог формулы (25.11):

$$E_i = (\Delta f/f 100\%)/(\Delta x_i/x_i 100\%).$$

Из этой формулы следует, что эластичность выпуска по i-му ресурсу равна относительному изменению объема выпуска при изменении затрат этого ресурса на один процент.

Сумма всех эластичностей

$$E_x = E_1 + E_2 + \ldots + E_n$$

называется *эластичностью производства*.

Средние и маржинальные показатели, а также эластичность производственной функции являются одними из основных характеристик, используемых в экономике.

Пример 2. Найдем средние и маржинальные значения, а также эластичности для двухфакторных мультипликативной и аддитивной функции.

Решение. а) Для мультипликативной ПФ типа Кобба–Дугласа (25.6) имеем:

$$A_1 = f(\overline{x})/x_1 = a_0 x_1^{a_1-1} x_2^{a_2}, \quad A_2 = f(\overline{x})/x_2 = a_0 x_1^{a_1} x_2^{a_2-1};$$
$$M_1 = \partial f/\partial x_1 = a_1 A_1, \quad M_2 = \partial f/\partial x_2 = a_2 A_2;$$
$$E_1 = M_1/A_1 = a_1, \quad E_2 = M_2/A_2 = a_2.$$

Из полученных выражений с учетом соотношений (25.8) следует, что

$$M_i < A_i, \quad i = 1, 2,$$

т.е. предельная производительность i-го ресурса меньше средней производительности этого ресурса. Кроме того, мультипликативные функции имеют постоянные частные эластичности.

б) Для аддитивной функции вида (25.5) имеем:

$$A_1 = f(\overline{x})/x_1 = a_0/x_1 + a_1 + a_2 x_2/x_1,$$
$$A_2 = f(\overline{x})/x_2 = a_0/x_2 + a_2 + a_1 x_1/x_2;$$
$$M_1 = \partial f/\partial x_1 = a_1, \quad M_2 = \partial f/\partial x_2 = a_2;$$
$$E_1 = M_1/A_1 = a_1/(a_0/x_1 + a_1 + a_2 x_2/x_1),$$
$$E_2 = M_2/A_2 = a_2/(a_0/x_2 + a_2 + a_1 x_1/x_2).$$

Из полученных соотношений следует, что и для аддитивной производственной функции предельная производительность i-го ресурса меньше средней производительности этого ресурса.

Наряду со средними и маржинальными показателями в экономике используется еще один важный показатель. Величина

$$R_{ij} = -dx_j/dx_i \quad (i, j = 1, 2) \tag{25.12}$$

называется *предельной нормой замены* i-го ресурса j-м ресурсом при постоянном объеме выпуска y (или предельной нормой замены ресурсов) для двухфакторной производственной функции. Из вида разностного аналога этой формулы $R_{ij} \approx -\Delta x_j/\Delta x_i$ следует, что предельная норма замены ресурсов показывает, на сколько единиц увеличатся (уменьшатся) затраты j-го ресурса при уменьшении (увеличении) затрат i-го ресурса. Для многофакторной производственной функции этот показатель определяется аналогично.

Выведем выражение для R_{ij}. Поскольку $f(\overline{x}) = \text{const}$, то $df = 0$, т.е. первый дифференциал равен нулю, откуда следует:

$$\frac{\partial f(\overline{x})}{\partial x_1} dx_1 + \frac{\partial f(\overline{x})}{\partial x_2} dx_2 = 0.$$

Поделив это уравнение на $\dfrac{\partial f(\overline{x})}{\partial x_j} dx_i$ $(i, j = 1, 2)$, получаем выражение для предельной нормы замены i-го ресурса j-м ресурсом:

$$R_{ij} = -\frac{dx_j}{dx_i} = \left(\frac{\partial f(\overline{x})}{\partial x_i}\right) / \left(\frac{\partial f(\overline{x})}{\partial x_j}\right), \tag{25.13}$$
$$(i \neq j, \ i, j = 1, 2).$$

В силу свойства 3 производственной функции этот показатель всегда положителен: $R_{ij} > 0$.

Для двухфакторной производственной функции справедливо равенство

$$R_{ij} = (E_1 x_2)/(E_2 x_1), \tag{25.14}$$

что непосредственно проверяется прямой подстановкой выражений для эластичности (25.11) в соотношение (25.13).

Пример 3. Найдем выражения для предельных норм замены в случаях: а) производственной функции Кобба–Дугласа (25.7) и б) аддитивной двухфакторной ПФ (25.5).

Решение. а) Подставляя в формулу (25.14) выражения эластичности для мультипликативной функции (см. пример 2а), получаем:

$$R_{12} = (\alpha_1 L)/(\alpha_2 K), \quad R_{21} = (\alpha_2 K)/(\alpha_1 L).$$

б) Аналогичная подстановка в случае аддитивной ПФ (пример 2б) приводит к еще более простым выражениям:

$$R_{12} = a_1/a_2, \quad R_{21} = a_2/a_1.$$

25.6. Убывающая эффективность производства

Рассмотрим производственную функцию Кобба–Дугласа. Поделив обе части уравнения (25.7) на переменную L, получаем с учетом равенства $\alpha_1 + \alpha_2 = 1$ однофакторную производственную функцию:

$$y = a_0 k^{\alpha_1}, \tag{25.15}$$

где $y = Y/L$ — производительность труда, $k = K/L$ — фондовооруженность. График функции (25.15) показан на рис. 25.1. Поскольку $\alpha_1 < 1$, то вторая производная

$$d^2y/dk^2 = a_0\alpha_1(\alpha_1 - 1)k^{\alpha_1 - 1} < 0,$$

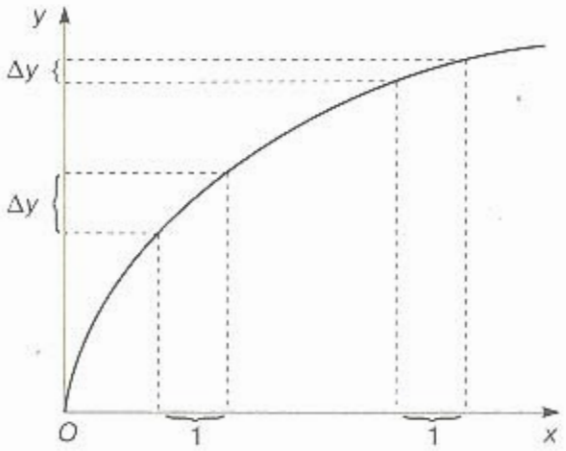

Рис. 25.1

т.е. кривая направлена выпуклостью вверх и становится все более пологой с ростом k. Это означает, что каждая дополнительная единица ресурса фондовооруженности дает все меньший прирост производительности труда. Это же свойство прослеживается и в зависимости объема выпуска от затрат ресурсов в двухфакторной производственной функции (свойство 4) — замедление темпов роста объема выпуска с увеличением затрат по каждому ресурсу (сечения трехмерного графика этой функции плоскостями $x_1 = \text{const}$ и $x_2 = \text{const}$ приводят к графикам, аналогичным графику функции (25.15)). Указанное обстоятельство в сочетании со свойствами 2 и 3 (рост величины выпуска) отражает фундаментальное положение экономической теории, которое называется *законом убывающей эффективности производства*. Образно говоря, это можно назвать законом "все меньшей отдачи" с ростом производства.

Согласно закону убывающей эффективности производства целесообразно определить, начиная с какого объема затрат ресурсов дальнейшее их увеличение становится нерентабельным; для этого достаточно определить критическое значение темпа роста объема выпуска.

Аппарат производственных функций активно используется в моделях поведения производителей, основанных на максимизации прибыли, а также в моделях фирмы. В этом плане производственная функция описывает технологию фирмы. Рассмотрим это на следующем примере.

Пример 4. Выпуск однопродуктовой фирмы определяется ПФ Кобба–Дугласа $Y = 3\,K^{1/3}L^{2/3}$. Найти распределение фондов K и затрат труда L, при котором выпуск будет максимальным, если на аренду фондов и оплату труда выделено 150 денежных единиц (д.е.), стоимость аренды фондов $w_K = 10$ д.е. на единицу фондов, ставка заработной платы $w_L = 5$ д.е./чел.

Решение. Пусть K и L — искомые распределение фондов и затраты труда, тогда по условию задачи имеем $Kw_K + Lw_L = 150$. Из этого уравнения связи выразим одно искомое неизвестное через другое, например L через K: $L = 150/w_L - Kw_K/w_L = 30 - 2K$. Далее подставляем полученное выражение в уравнение производственной функции, откуда получаем функцию одной переменной — аргумента K: $Y = 3\,K^{1/3}\times$

$\times(30-2K)^{2/3}$. Теперь находим точку максимума этой функции, приравнивая нулю производную:

$$Y' = (30-2K)^{2/3}/K^{2/3} - 4K^{1/3}/(30-2K)^{1/3} = 0.$$

Отсюда последовательно получаем, что $K = 5$ е.ф., $L = 20$ чел.

Глава 26
МОДЕЛЬ ПОТРЕБИТЕЛЬСКОГО ВЫБОРА

Будем полагать, что имеется n различных товаров. Тогда некоторый набор товаров будем обозначать через n-мерный вектор

$$\overline{x} = (x_1, x_2, \ldots, x_n), \tag{26.1}$$

где x_i — количество i-го товара. Естественные ограничения на набор товаров — это $x_i \geqslant 0$ ($i = 1,2,...,n$); в условной записи $\overline{x} \geqslant \overline{0}$, где $\overline{0}$ — нулевой вектор. Множество всех наборов товаров X называется пространством, так как в нем можно сложить любые два набора и умножить любой набор на любое неотрицательное число (последнее справедливо для случая безграничной делимости и умножения товаров).

Выбор индивида-потребителя характеризуется отношением предпочтения: считается, что потребитель может сказать о любых двух наборах, какой более желателен, или он не видит между ними разницы. Отношение предпочтения транзитивно: если набор $\overline{x}$ предпочтительнее набора $\overline{y}$, а набор $\overline{y}$ предпочтительнее набора $\overline{z}$, то набор $\overline{x}$ предпочтительнее набора $\overline{z}$. Будем полагать, что поведение потребителя полностью описывается *аксиомой индивида-потребителя*: каждый индивид-потребитель принимает решения о потреблении, покупках и т.п., исходя только из своей системы предпочтений.

26.1. Функция полезности

На множестве потребительских наборов X определена функция

$$u(\overline{x}) = u(x_1, x_2, \ldots, x_n), \tag{26.2}$$

значение которой на потребительском наборе $\overline{x}$ равно потребительской оценке индивида для этого набора. Функция $u(\overline{x})$ называется *функцией полезности потребителя* или функцией потребительского предпочтения. Согласно этому определению каждый потребитель имеет свою функцию полезности. Однако все множество потребителей можно разделить на определенные классы потребителей (по возрасту, имущественному положению и т.п.) и каждому классу приписать некоторую, может быть, осредненную функцию полезности.

Таким образом, функция $u(\overline{x})$ является потребительской оценкой или уровнем удовлетворения потребностей индивида при приобретении данного набора $\overline{x}$. Если набор $\overline{x}$ предпочтительнее набора $\overline{y}$ для данного индивида, то $u(\overline{x}) > u(\overline{y})$.

Укажем свойства функции полезности в предположении о ее дифференцируемости.

$$1.\ \partial u/\partial x_i = u'_i > 0, \quad i = 1,2,\ldots,n. \tag{26.3}$$

Первые частные производные функции полезности называются *предельными полезностями продуктов*. Из этого свойства вытекает, что если $x'_k > x''_k$ при неизменных количествах других продуктов, то $u(x_1, \ldots, x'_k, x_{k+1}, \ldots, x_n) > u(x_1, \ldots, x''_k, x_{k+1}, \ldots, x_n)$, т.е. возрастание потребления одного продукта при неизменном потреблении других продуктов приводит к росту потребительской оценки. Заметим, что вектор $\partial u/\partial \overline{x} = (\partial u/\partial x_1, \partial u/\partial x_2, ..., \partial u/\partial x_n)$ является градиентом функции $u(\overline{x})$; он показывает направление наибольшего роста функции. Для функции полезности $u(\overline{x})$ ее градиент представляет собой вектор предельных полезностей продуктов.

$$2.\ \partial^2 u/\partial x_i^2 = u''_{ii} < 0, \quad i = 1,2,\ldots,n, \tag{26.4}$$

т.е. предельная полезность любого товара уменьшается с ростом его потребления. Это утверждение называется *законом Госсена* (немецкий экономист XIX века, впервые на практике установивший эту закономерность).

$$3.\ \partial^2 u/(\partial x_i \partial x_j) = u''_{ij} > 0, \quad i \neq j\ (i,j = 1,2,\ldots,n). \tag{26.5}$$

Это свойство означает, что предельная полезность каждого продукта увеличивается с ростом количества другого продукта.

Нетрудно видеть, что указанные свойства функции полезности аналогичны свойствам 3–5 производственной функции (см. раздел 25.4). Формально это следует из трактования функции полезности как производственной функции, ресурсами которой являются товары, а выпуском — уровень удовлетворения потребностей индивида.

Приведем несколько видов функций полезности, удовлетворяющих принятым допущениям.

а) Неоклассическая:

$$u(x_1, x_2) = x_1^{\alpha} x_2^{\beta}, \quad \alpha, \beta > 0, \quad \alpha + \beta < 1. \tag{26.6}$$

б) Квадратическая:

$$u(\overline{x}) = \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} x_i x_j, \tag{26.7}$$

где матрица $\|b_{ij}\|$ отрицательно определена и $a_j + \sum\limits_{j=1}^{n} b_{ij} x_i > 0$ для $j = 1, 2, \ldots, n$.

в) Логарифмическая функция:

$$u(\overline{x}) = \sum_{i=1}^{n} a_i \log_d (x_i - b_i), \tag{26.8}$$

где $a_i > 0$, $x_i > b_i \geqslant 0$ для $j = 1, 2, \ldots, n$.

26.2. Линии безразличия

В прикладных задачах и моделях потребительского выбора часто используется частный случай набора из двух товаров, т.е. когда функция полезности зависит от двух переменных. Здесь вводится понятие *линии безразличия* — это линия, соединяющая потребительские наборы, имеющие один и тот же уровень удовлетворения потребностей индивида. По сути дела линии безразличия представляют собой линии уровня функции $u = f(x_1, x_2)$ или сечения поверхности $f(x_1, x_2)$ плоскостями, параллельными плоскости $x_1 O x_2$; уравнения линий безразличия:

$$u(x_1, x_2) = \text{const}. \tag{26.9}$$

Множество линий безразличия называется картой линий безразличия (карта линий уровня — это снесенные на плоскость $x_1 O x_2$ линии уровня, соответствующие различным значениям постоянной величины в правой части (26.9)). Карта линий безразличия изображается на плоскости $x_1 O x_2$ в виде семейства кривых. Понятие линий безразличия и оперирование с ними встречаются во многих экономических приложениях, имеющих цель оптимизации экономических процессов.

В случае многомерной функции полезности (26.2), определенной в пространстве многомерных наборов (26.1), аналогичным образом можно ввести в рассмотрение *поверхность безразличия* — поверхность уровня функции (26.2) или гиперповерхность размерности $n - 1$.

Укажем основные свойства линий безразличия.

1. Линии безразличия, соответствующие разным уровням удовлетворения потребностей, не касаются и не пересекаются. Это следует из вида их определения (26.9).

2. Линии безразличия убывают. Рассмотрим уравнение этой линии в виде

$$x_2 = \varphi(x_1), \tag{26.10}$$

которое можно получить из (26.9). Дифференциал функции вдоль линии уровня равен нулю, т.е. $du(x_1, x_2) = u_1' dx_1 + u_2' dx_2 = 0$, откуда следует, что первая производная функции (26.10)

$$dx_2/dx_1 = -u_1'/u_2' \tag{26.11}$$

отрицательна в силу свойства 1 функций полезности (формулы (26.3)).

3. Линии безразличия выпуклы вниз. Действительно, вторая производная функции (26.10), согласно правилу дифференцирования частного, вычисляется по формуле:

$$d(dx_2/dx_1)/dx_1 = d^2 x_2/dx_1^2 = -(u_{11}'' u_2' - u_{21}'' u_1')/(u_2')^2.$$

Нетрудно видеть, что согласно свойствам 2 и 3 функций полезности выражение в правой части этого равенства отрицательно, т.е. график функции (26.10) является выпуклым вниз.

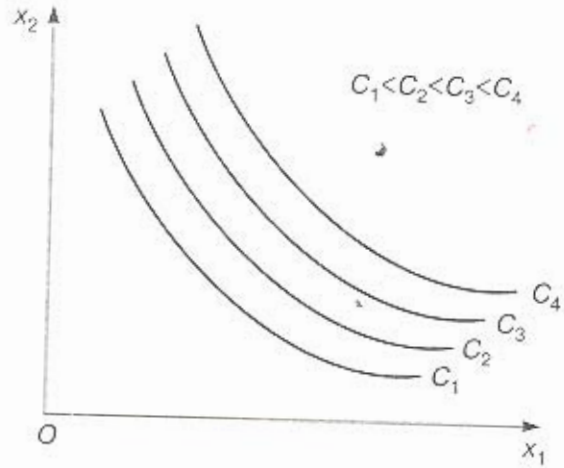

Рис. 26.1

Характерный вид линий безразличия функции полезности показан на рис. 26.1.

Из формулы (26.11) следует важное приближенное равенство

$$-\Delta x_2/\Delta x_1 \approx u_1'/u_2'.$$

Это соотношение показывает, на сколько индивид должен увеличить (уменьшить) потребление второго продукта при уменьшении (увеличении) потребления первого продукта на одну единицу без изменения уровня удовлетворения своих потребностей. Отношение $-\Delta x_2/\Delta x_1$ называется *нормой замены* первого продукта вторым, а величину dx_2/dx_1 — *предельной нормой замены* первого продукта вторым.

Пример 1. Если предельная полезность первого товара равна 6, а второго товара — 2, то при уменьшении потребления первого товара на единицу нужно увеличить потребление второго товара на 3 единицы при том же уровне удовлетворения потребностей.

26.3. Бюджетное множество

Пусть $\overline{p} = (p_1, p_2, \ldots, p_n)$ — вектор цен на набор из n продуктов $\overline{x}$. Зафиксируем некоторую денежную сумму I — доход

индивида, который он готов потратить на приобретение набора продуктов $\overline{x} \in X$ (на самом деле это расход индивида на набор продуктов). Множество наборов товаров стоимостью не более I при данных ценах $\overline{p}$ называется *бюджетным множеством* $B \subset X$. При этом множество наборов стоимостью ровно I называется границей G множества B. Таким образом, множество B ограничено границей G, а также естественными ограничениями $\overline{x} \geqslant \overline{0}$.

Поскольку граница G определяется соотношением $\overline{p}\,\overline{x} = I$, то бюджетное множество B описывается системой следующих неравенств:

$$\overline{x} \geqslant \overline{0}, \quad \overline{p}\,\overline{x} \leqslant I, \tag{26.12}$$

или в развернутой форме:

$$x_i \geqslant 0\ (i = 1, 2, \ldots, n), \quad p_1x_1 + p_2x_2 + \ldots + p_nx_n \leqslant I. \tag{26.13}$$

Для случая набора из двух товаров бюджетное множество представляет собой треугольник в системе координат x_1Ox_2, ограниченный координатными осями и прямой $p_1x_1 + p_2x_2 = I$ (рис. 26.2).

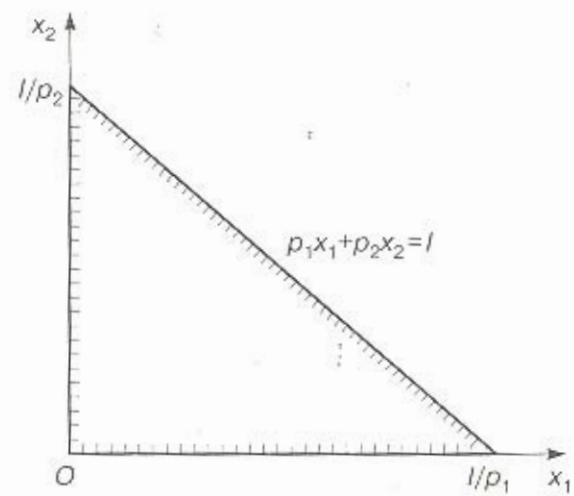

Рис. 26.2

26.4. Задача потребительского выбора

В теории потребления полагается, что потребитель всегда стремится максимизировать свою полезность и единственным ограничением для него является ограниченность дохода I, который он может потратить на покупку набора товаров. В общем виде задача потребительского выбора (задача рационального поведения потребителя на рынке) формулируется следующим образом: найти потребительский набор

$$\overline{x}^* = (x_1^*, x_2^*, \ldots, x_n^*), \tag{26.14}$$

который максимизирует его функцию полезности при заданном бюджетном ограничении. Формальный вид задачи потребительского выбора — ищется точка максимума (26.14) функции полезности на бюджетном множестве

$$\begin{gathered} u(\overline{x}) = u(x_1, x_2, \ldots, x_n) \to \max; \\ x_i \geqslant 0 \ \ (i = 1, 2, \ldots, n), \quad p_1 x_1 + p_2 x_2 + \ldots + p_n x_n \leqslant I. \end{gathered} \tag{26.15}$$

Рассмотрим задачу потребительского выбора для случая набора из двух товаров: найти набор $\overline{x}^* = (x_1^*, x_2^*)$ такой что

$$u = u(x_1, x_2) \to \max;\ x_1 \geqslant 0,\ x_2 \geqslant 0,\ p_1 x_1 + p_2 x_2 \leqslant I. \tag{26.16}$$

Поиск оптимального набора $\overline{x}^*$ можно интерпретировать графически как последовательный переход на линии безразличия более высокого уровня полезности (рис. 26.3) — вправо-вверх — до тех пор, пока эти линии имеют общие точки с бюджетным множеством. Из геометрии бюджетного множества B (рис. 26.2) и формы линий безразличия (рис. 26.1) следует, что искомая точка (x_1^*, x_2^*) лежит на границе G, т.е. на прямой $p_1 x_1 + p_2 x_2 = I$. Таким образом, задача потребительского выбора сводится к задаче на условный экстремум — найти точку (x_1^*, x_2^*) максимума функции $u = u(x_1, x_2)$ при условии, что эта точка находится на прямой $p_1 x_1 + p_2 x_2 = I$. В условной записи:

$$u = u(x_1, x_2) \to \max; \quad p_1 x_1 + p_2 x_2 = I. \tag{26.17}$$

В общем случае бюджетного множества n-мерных наборов B также справедливо, что точка $\overline{x}^*$ лежит на границе G (на гиперплоскости), определяемой условием $\overline{p}\,\overline{x} = I$. Поэтому задача потребительского выбора формулируется аналогичным

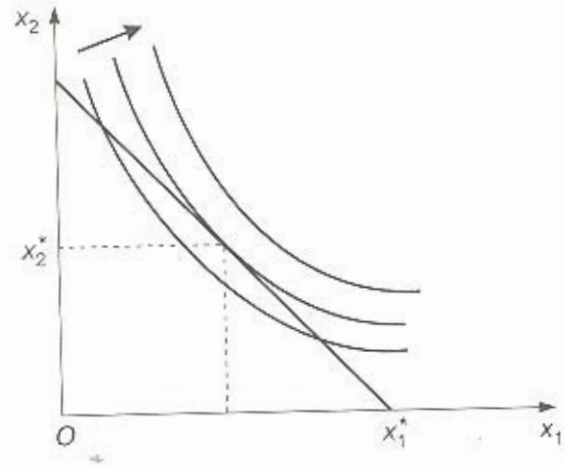

Рис. 26.3

образом в виде задачи на условный экстремум: при заданной функции $u(\overline{x})$ и известных векторе $\overline{p}$ и величине I найти точку (26.14) такую, что

$$\begin{gathered} u(\overline{x}) = u(x_1, x_2, \ldots, x_n) \to \max; \\ \overline{p}\,\overline{x} = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n = I. \end{gathered} \tag{26.18}$$

Последнее уравнение в (26.18) называется уравнением связи.

26.5. Решение задачи потребительского выбора

Как известно, для решения задачи (26.17) используется метод Лагранжа. Выпишем функцию Лагранжа

$$L(x_1, x_2, \lambda) = u(x_1, x_2) - \lambda(p_1 x_1 + p_2 x_2 - I), \tag{26.19}$$

где λ — неопределенный множитель Лагранжа. Экономический смысл этого множителя: если цены и доход I меняются в одно и то же число раз λ, то функция полезности, а значит, и решение задачи потребительского выбора не изменится. Далее ищется точка максимума функции L: все три частные производные этой функции приравниваются нулю, т.е. получаем систему трех уравнений:

$$\begin{cases} \partial L/\partial x_1 = u_1' - \lambda p_1 = 0, \\ \partial L/\partial x_2 = u_2' - \lambda p_2 = 0, \\ \partial L/\partial \lambda = p_1 x_1 + p_2 x_2 - I = 0. \end{cases}$$

Исключив из этих уравнений неизвестную λ, получаем систему двух уравнений с двумя неизвестными x_1 и x_2:

$$\begin{cases} u'_1/u'_2 = p_1/p_2, \\ p_1x_1 + p_2x_2 = I. \end{cases} \tag{26.20}$$

Решение (x_1^*, x_2^*) системы уравнений (26.20) является решением задачи потребительского выбора.

В многомерном случае (26.18) также используется метод Лагранжа. Функция Лагранжа имеет вид:

$$L(\overline{x}, \lambda) = u(\overline{x}) - \lambda(\overline{p}\,\overline{x} - I). \tag{26.21}$$

Далее, согласно условиям максимума, составляется система из $(n+1)$ уравнений относительно $(n+1)$ неизвестных $\overline{x}$ и λ:

$$\begin{gathered} \partial L/\partial x_i = u'_i - \lambda p_i = 0, \quad i = 1, 2, \ldots, n; \\ \partial L/\partial \lambda = \overline{p}\,\overline{x} - I = 0. \end{gathered} \tag{26.22}$$

Исключая из этих уравнений неизвестную λ, получаем систему n уравнений относительно n неизвестных — координат точки условного максимума:

$$u'_i/u'_j = p_i/p_j, \quad i \neq j; \quad \overline{p}\,\overline{x} - I = 0. \tag{26.23}$$

Решение $\overline{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ этой системы уравнений — точка условного экстремума — является решением общей задачи потребительского выбора.

Если функция полезности строго вогнута вниз и удовлетворяет условиям дифференцируемости (26.3)–(26.5), то при естественных ограничениях на векторы цен и набора товаров (все компоненты должны быть неотрицательными) задача определения оптимального набора товаров, как задача на условный экстремум, имеет единственное решение $\overline{x}^*$ при любом положительном доходе I.

Точка $\overline{x}^*$ называется *точкой локального рыночного равновесия*. Первое соотношение (26.23) показывает, что отношение предельных полезностей продуктов в точке локального рыночного равновесия, или *предельная норма замены* i-го продукта j-м продуктом $u'_i(\overline{x}^*)/u'_j(\overline{x}^*)$ равна отношению рыночных цен на эти продукты. Этот результат играет важную роль в экономической теории.

26.6. Функции спроса

Решение задачи потребительского выбора $\overline{x}^*$ называется *точкой спроса*. Эта точка спроса зависит от цен $\overline{p}$ и дохода I. Иными словами, точка спроса является функцией цен и дохода, т.е. *функцией спроса*. В свою очередь, функция спроса представляет собой вектор-функцию $n+1$ аргумента (в общем случае n цен $p_1, p_2, \ldots, p_n$ и дохода I), состоящую из n компонент:

$$\overline{x}^* = \overline{x}^*(p_1, p_2, \ldots, p_n, I).$$

Таким образом, функция спроса — это набор n функций, каждая из которых зависит от $n+1$ аргумента:

$$\begin{aligned} x_1^* &= x_1^*(p_1, p_2, \ldots, p_n, I), \\ x_2^* &= x_2^*(p_1, p_2, \ldots, p_n, I), \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n^* &= x_n^*(p_1, p_2, \ldots, p_n, I). \end{aligned} \tag{26.24}$$

Функции (26.24) называются функциями *спроса соответствующих товаров*.

Пример 2. Для набора из двух товаров на рынке, известных ценах на них p_1 и p_2 и дохоле I найти функции спроса, если функция полезности имеет вид $u(x_1, x_2) = x_1^{0,5} x_2^{0,5}$.

Решение. Дифференцируя данную функцию полезности, получаем $u'_1 = 0{,}5x_2^{0,5}/x_1^{0,5}$, $u'_2 = 0{,}5x_1^{0,5}/x_2^{0,5}$. Далее, подстановка в формулы (26.20) приводит к системе уравнений:

$$\begin{cases} x_2/x_1 = p_1/p_2, \\ p_1x_1 + p_2x_2 = I. \end{cases}$$

Из первого уравнения следует, что затраты денежных средств на оба вида товаров должны быть одинаковыми, т.е. $p_1x_1 = p_2x_2$. Следовательно, из второго уравнения получаем, что функции спроса имеют вид:

$$x_1^* = I/2p_1, \quad x_2^* = I/2p_2.$$

Иными словами, в данном случае расход на каждый товар составляет половину дохода потребителя, а количество приобретенного товара равно затраченной на него сумме, поделенной на цену товара.

26.7. Модель Р. Стоуна

Модель потребительского спроса с функцией полезности мультипликативного вида (неоклассическая)

$$u(\overline{x}) = \prod_{i=1}^{n}(x_i - a_i)^{\alpha_i}; \quad x_i > a_i,\ a_i > 0,\ i = 1, 2, \ldots, n \quad (26.25)$$

носит имя автора этой функции. Здесь a_i — минимально необходимое количество i-го товара, которое приобретается в любом случае. Набор $\overline{a} = (a_1, a_2, \ldots, a_n)$ можно рассматривать как минимальную корзину потребления. Для приобретения минимального набора $\overline{a}$ необходимо, чтобы доход был больше стоимости этого набора, т.е.

$$I > \sum_{i=1}^{n} p_i a_i. \quad (26.26)$$

Показатели степеней $\alpha_i > 0$ в (26.25) характеризуют относительную "ценность" соответствующих товаров для потребителя.

Добавив к целевой функции (26.25) бюджетные ограничения, получим задачу потребительского выбора (26.15), называемую моделью Р.Стоуна. Приравняв нулю частные производные функции Лагранжа L по всем переменным, получим систему $n + 1$ уравнений:

$$\frac{\alpha_i u(\overline{x})}{x_i - a_i} - \lambda p_i = 0, \quad i = 1, 2, \ldots, n; \quad \sum_{i=1}^{n} p_i x_i - I = 0.$$

Из соотношений первого вида получаем выражения для неизвестных x_i:

$$x_i = a_i + \frac{\alpha_i u(\overline{x})}{\lambda p_i}, \quad i = 1, 2, \ldots, n. \quad (26.27)$$

Умножим каждое из уравнений (26.27) на λp_i и просуммируем их по i; получаем:

$$u(\overline{x}) \sum_{i=1}^{n} \alpha_i + \lambda \left(\sum_{i=1}^{n} p_i x_i - \sum_{i=1}^{n} p_i a_i \right) = 0.$$

Отсюда отношение $u(\overline{x})/\lambda$ с учетом условия связи имеет вид:

$$\frac{u(\overline{x})}{\lambda} = -\left(I - \sum_{i=1}^{n} p_i a_i \right) / \sum_{i=1}^{n} \alpha_i.$$

Подставляя это выражение в (26.27), получаем в окончательном виде формулы для функций спроса:

$$x_i^* = a_i + \frac{\alpha_i}{p_i} \left(I - \sum_{i=1}^{n} p_i a_i \right) / \sum_{i=1}^{n} \alpha_i, \quad i = 1, 2, \ldots, n. \quad (26.28)$$

Полученные функции спроса легко интерпретировать по последовательности действий.

1. Вначале приобретается набор $\overline{a}$ необходимых количеств товаров.

2. Затем рассчитывается оставшаяся сумма денег.

3. Эта сумма денег распределяется пропорционально весам "ценности" α_i товаров.

4. Поделив распределенные на приобретение каждого товара суммы денег на соответствующие цены этих товаров, получаем приобретаемое сверх минимума a_i количество i-го товара.

Укажем частный случай модели Стоуна.

$$a_i = 0, \quad \alpha_i = \alpha, \quad i = 1, 2, \ldots, n. \quad (26.29)$$

В этом случае из (26.28) получаем:

$$x_i^* = a_i + \frac{I}{n p_i}, \quad i = 1, 2, \ldots, n, \quad (26.30)$$

т.е. доход I делится на n равных частей и количество приобретаемого i-го товара обратно пропорционально его цене. Здесь спрос растет при росте дохода с единичной эластичностью и уменьшается с ростом цены с эластичностью, равной минус единице.

26.8. Уравнение Слуцкого

В теории потребительского выбора и в исследовании функций спроса основополагающую роль играет уравнение, полученное российским математиком Е.Е. Слуцким в 1915 г. Приведем здесь вывод уравнения Слуцкого в общем случае набора из n товаров[1].

[1] Приводимый здесь вывод был выполнен доц. Г.А. Шапошниковой.

Рассмотрим функции спроса (26.24). При неизменной величине функции полезности в точке локального рыночного равновесия

$$u(x_1^*, x_2^*, \ldots, x_n^*) = \text{const} \tag{26.31}$$

(на гиперповерхности уровня этой функции) приращение функции x_i^*, обусловленное изменением только величины цены p_j, равно:

$$\Delta x_i^* = \left(\frac{\partial x_i^*}{\partial p_j}\right)_{I=\text{const}} \Delta p_j + \frac{\partial x_i^*}{\partial I}\Delta I, \quad j = 1, 2, \ldots, n.$$

Поделив обе части на Δp_j и переходя к пределу при $\Delta p_j \to 0$, получаем:

$$\left(\frac{\partial x_i^*}{\partial p_j}\right)_{u=\text{const}} = \left(\frac{\partial x_i^*}{\partial p_j}\right)_{I=\text{const}} + \frac{\partial x_i^*}{\partial I}\frac{\partial I}{\partial p_j}; \quad j = 1, 2, \ldots, n. \tag{26.32}$$

Поскольку $I = p_1 x_1^* + p_2 x_2^* + \ldots + p_n x_n^*$, то при изменении цены p_j на величину Δp_j приращение дохода ΔI составит:

$$\Delta I = p_1 \Delta x_1^* + p_2 \Delta x_2^* + \ldots + p_n \Delta x_n^* + \Delta p_j x_j^*, \quad j = 1, 2, \ldots, n. \tag{26.33}$$

Как следствие, из (26.33) получаем, что

$$\frac{\partial I}{\partial p_j} = p_1 \frac{\partial x_1^*}{\partial p_j} + p_2 \frac{\partial x_2^*}{\partial p_j} + \ldots + p_n \frac{\partial x_n^*}{\partial p_j} + x_j^*; \quad j = 1, 2, \ldots, n. \tag{26.34}$$

Так как дифференциал от функции (26.31) равен нулю, имеем соотношение:

$$\frac{\partial u}{\partial x_1^*}\Delta x_1^* + \frac{\partial u}{\partial x_2^*}\Delta x_2^* + \ldots + \frac{\partial u}{\partial x_n^*}\Delta x_n^* = 0. \tag{26.35}$$

В точке локального рыночного равновесия, в силу (26.23), имеем соотношения пропорциональности первых производных функции полезности и соответствующих цен:

$$u_i' = \lambda p_i, \quad i = 1, 2, \ldots, n.$$

После подстановки этих соотношений в (26.35) и сокращения обеих частей на сомножитель λ, получаем:

$$p_1 \Delta x_1^* + p_2 \Delta x_2^* + \ldots + p_n \Delta x_n^* = \bar{p}\Delta\bar{x}^* = 0. \tag{26.36}$$

После деления соотношения (26.36) на Δp_j и предельного перехода при $\Delta p_j \to 0$ получаем, что

$$p_1 \frac{\partial x_1^*}{\partial p_j} + p_2 \frac{\partial x_2^*}{\partial p_j} + \ldots + p_n \frac{\partial x_n^*}{\partial p_j} = 0, \quad j = 1, 2, \ldots, n.$$

Подстановка этих уравнений в (26.34) приводит к формулам:

$$\frac{\partial I}{\partial p_j} = x_j^*; \quad j = 1, 2, \ldots, n,$$

т.е. в точке рыночного равновесия относительное изменение дохода, вызванное изменением цены на товар, равно количеству этого товара. Подставляя эти формулы в формулы (26.32), окончательно получаем:

$$\left(\frac{\partial x_i^*}{\partial p_j}\right)_{u=\text{const}} = \left(\frac{\partial x_i^*}{\partial p_j}\right)_{I=\text{const}} + \frac{\partial x_i^*}{\partial I} x_j^*; \quad j = 1, 2, \ldots, n. \tag{26.37}$$

Поскольку первое слагаемое в правой части представляет собой действие эффекта изменения цены при неизменном доходе на компенсационную добавку величины продукта, то это слагаемое называется компенсационным. Само же уравнение Слуцкого получается при переносе второго слагаемого в другую часть и используется в следующей форме:

$$\frac{\partial x_i^*}{\partial p_j} = \left(\frac{\partial x_i^*}{\partial p_j}\right)_{comp} - \frac{\partial x_i^*}{\partial I} x_j^*; \quad j = 1, 2, \ldots, n. \tag{26.38}$$

Поясним экономический смысл уравнения Слуцкого (26.38). Левая часть уравнения (26.38) представляет собой "отклик" точки спроса при неизменном доходе на изменение цены j-го товара. Второе слагаемое в правой части описывает действие эффекта дохода: это произведение, в котором одним из сомножителей является "отклик" точки спроса на изменение дохода

I, а другим — величина спроса на j-й товар. Первое слагаемое в правой части (26.38) интерпретируется следующим образом: при изменении цены j-го товара на Δp_j при неизменных остальных ценах и доходе изменяются точка спроса и максимальная полезность; изменим доход так, чтобы значение максимальной полезности осталось неизменным — это и является компенсацией.

К уравнению Слуцкого присоединяется ряд соотношений, которые получаются в процессе его вывода. Перечислим их, поскольку они носят четкий экономический характер. Часть их мы приводим ниже без вывода. Итак, в точке локального рыночного равновесия (26.24) справедливы следующие соотношения.

1. Уравнение (26.36):

$$p_1\Delta x_1^* + p_2\Delta x_2^* + \ldots + p_n\Delta x_n^* = 0.$$

Это уравнение связывает прежние цены с приращениями товаров, что придает сложный характер изменению спроса: спрос на одни блага растет, на другие он падает.

2. Из уравнения (26.33) с учетом (26.36) получаем:

$$dI = x_1^* dp_1 + x_2^* dp_2 + \ldots + x_n^* dp_n. \qquad (26.39)$$

Это соотношение означает, что при увеличении цен компенсация дохода должна быть положительной, а при уменьшении цен — отрицательной, что достаточно очевидно.

3. $$\left(\frac{\partial x_i^*}{\partial p_i}\right)_{comp} < 0. \qquad (26.40)$$

Это означает, что при повышении цены спрос на товар падает даже при компенсации дохода.

4. $$\left(\frac{\partial x_1^*}{\partial I}\right) p_1 + \left(\frac{\partial x_2^*}{\partial I}\right) p_2 + \ldots + \left(\frac{\partial x_n^*}{\partial I}\right) p_n = 1. \qquad (26.41)$$

Это уравнение показывает, что часть слагаемых суммируется со знаком плюс, а часть — со знаком минус. Товар называется "ценным", если при повышении дохода спрос на него растет, в противном случае товар является малоценным. Из (26.41) следует, что не все слагаемые отрицательны, т.е. существуют ценные товары. Например, высококачественные

продукты и предметы роскоши являются ценными, маргарин — малоценным. При повышении дохода употребление маргарина снижается, в то время как потребление высококачественного масла возрастает.

5. $$\left(\frac{\partial x_1^*}{\partial p_i}\right)_{comp} p_1 + \left(\frac{\partial x_2^*}{\partial p_i}\right)_{comp} p_2 + \ldots + \left(\frac{\partial x_n^*}{\partial p_i}\right)_{comp} p_n = 0. \qquad (26.42)$$

Так как $p_i > 0$, то при условии (26.40) следует, что найдется какой-то j-й товар такой, что $(\partial x_j^*/\partial p_i)_{comp} > 0$ — иными словами, при повышении цены на i-й товар потребление j-го товара растет при компенсационном изменении дохода I. В таком случае j-й товар называется заменяющим i-й товар. Например, для большинства потребителей рыба может заменить мясо при повышении цены на него, соответственно картофель — хлеб, а растительное масло — сливочное масло.

Глава 27

ЭЛЕМЕНТЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ В ЭКОНОМИКЕ

Теория систем массового обслуживания (СМО) начала развиваться в начале XX столетия. Иохансен в 1907 г. сформулировал основные предпосылки новой теории. В 1909 г. шведский математик Эрланг применил теорию вероятностей к исследованию зависимости обслуживания телефонных вызовов от числа поступающих на телефонную станцию вызовов. В нашей стране известный математик А. Я. Хинчин систематизировал основные положения теории СМО в монографии "Теория очередей". Именно такое название теории СМО используется за рубежом.

В последние годы применение теории систем массового обслуживания в экономике приобрело особую актуальность в связи с использованием ряда ее аспектов в финансово-экономической сфере (банки различных типов, страховые организации, налоговые инспекции, аудиторские службы). Теория СМО широко применяется также в сфере обслуживания (различные системы связи, погрузочно-разгрузочные комплексы, АЗС, ма-

газины, билетные кассы, ремонтные предприятия, больницы и т.д.) и в современных высоких технологиях (компьютерные сети, базы данных, поточные линии, военные системы ПВО и ПРО).

27.1. Структура и классификация систем массового обслуживания

Системы массового обслуживания представляют собой системы специфического вида. *Системой* называется целостное множество взаимосвязанных элементов, которое нельзя разделить на независимые подмножества. Основой СМО является определенное число обслуживающих устройств — *каналы обслуживания*. Роль каналов в реальности могут выполнять приборы, операторы, продавцы, линии связи и пр.

Предназначение СМО состоит в обслуживании *потока заявок* (*требований*), представляющих последовательность событий, поступающих нерегулярно и в заранее неизвестные и случайные моменты времени. Само *обслуживание заявок* также имеет непостоянный характер, происходит в случайные промежутки времени и зависит от многих и даже неизвестных причин. Случайный характер потока заявок и времени их обслуживания обусловливает неравномерность загрузки СМО: на входе могут накапливаться необслуженные заявки (перегрузка СМО) либо заявок нет или их меньше, чем свободных каналов (недогрузка СМО). Структура систем массового обслуживания показана схематически на рис. 27.1. В СМО поступает поток заявок; часть из них принимается на обслуживание в каналы, часть ждет в очереди на обслуживание, часть покидает систему необслуженными.

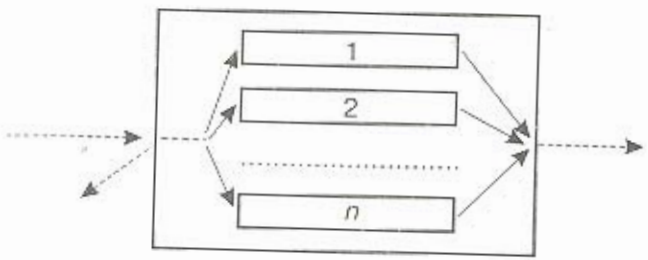

Рис. 27.1

Основными элементами СМО являются:

1) входной поток заявок,

2) очередь,

3) каналы обслуживания,

4) выходной поток заявок (обслуженные заявки).

Эффективность функционирования СМО определяется ее *пропускной способностью* — относительным числом обслуженных заявок.

По числу каналов n все СМО разделяются на одноканальные ($n = 1$) и многоканальные ($n > 1$). Многоканальные СМО могут быть как однородными (по каналам), так и разнородными (по продолжительности обслуживания заявок).

По дисциплине обслуживания различаются три класса СМО.

1. СМО с *отказами* (нулевое ожидание или явные потери). "Отказная" заявка вновь поступает в систему, чтобы ее обслужили (например, вызов абонента через АТС).

2. СМО с *ожиданием* (неограниченное ожидание или очередь). При занятости системы заявка поступает в очередь и в конце концов будет выполнена (торговля, сфера бытового и медицинского обслуживания).

3. СМО *смешанного типа* (ограниченное ожидание). Имеется ограничение на длину очереди (сервис по обслуживанию автомобилей). Ограничение на время пребывания заявки в СМО (ПВО, особые условия обслуживания в банке) также может рассматриваться.

В дальнейшем мы будем рассматривать *открытые* (поток заявок неограничен), *упорядоченные* (заявки обслуживаются в порядке их поступления) и *однофазные* (однородные каналы выполняют одну и ту же операцию) СМО.

Целью теории систем массового обслуживания является выработка рекомендаций по рациональному построению СМО и рациональной организации их работы и регулированию потока заявок. Отсюда вытекают задачи, связанные с теорией массового обслуживания: установление зависимостей работы СМО от ее организации, характера потока заявок, числа каналов и их производительности, правил работы СМО.

27.2. Основные показатели эффективности работы СМО

Эффективность работы систем массового обслуживания характеризуют показатели, которые можно разбить на три группы.

1. Группа показателей эффективности использования СМО:

— абсолютная пропускная способность — среднее число заявок, обслуживаемых в единицу времени (A);

— относительная пропускная способность — отношение абсолютной пропускной способности к среднему числу заявок, поступивших в систему за единицу времени (Q);

— средняя продолжительность периода занятости СМО ($T_в$);

— коэффициент использования СМО — средняя доля времени, в течение которого система занята обслуживанием заявок.

2. Показатели качества обслуживания заявок:

— среднее время ожидания заявки в очереди ($\overline{T}_{\text{line}}$);

— среднее время пребывания заявки в СМО ($\overline{T}_{\text{sys}}$);

— вероятность отказа заявки в обслуживании без ожидания;

— вероятность немедленного приема заявки;

— закон распределения времени ожидания заявки в очереди в СМО;

— среднее число заявок в очереди ($\overline{N}_{\text{line}}$);

— среднее число заявок, находящихся в СМО ($\overline{N}_{\text{sys}}$).

3. Показатели эффективности функционирования пары "СМО — потребитель" (вся совокупность заявок или их источник, например, средний доход в единицу времени от СМО). Эта группа полезна, когда доход от СМО и затраты на ее обслуживание измеряются в одних и тех же единицах, и отражает специфику работы СМО.

27.3. Случайный процесс в СМО

Процессы поступления и обслуживания заявок в СМО являются случайными, что обусловлено случайным характером потока заявок и длительности их обслуживания. Как известно, случайный процесс описывается случайной функцией, когда каждому значению аргумента (например, времени) ставится в соответствие случайная величина — одно не известное заранее какое именно число из данного числового множества. Случайные процессы упрощают исследование СМО.

В основе СМО будем предполагать *марковский случайный процесс* (процесс без последействия или без памяти), когда вероятность состояния СМО в будущем зависит только от ее состояния в настоящем и не зависит от прошлого (название по имени известного российского математика А. А. Маркова). Условие марковского случайного процесса: необходимо, чтобы все потоки событий, при которых система переходит из одного состояния в другое (потоки заявок, потоки обслуживания и т.д.), были пуассоновскими. По характеру потоков событий СМО разделяются на марковские и немарковские; в дальнейшем мы будем иметь дело с системами первого типа. Преимущества и полезность такого подхода состоят в том, что для выработки обоснованных рекомендаций нужно знать не точные характеристики СМО, а лишь их приближенные значения.

Пуассоновский поток событий обладает следующими свойствами:

— отсутствия последействия (для любых двух непересекающихся промежутков времени число событий, наступающих за один из них, не зависит от числа событий, наступающих за другой промежуток времени);

— ординарности (вероятность наступления за малый промежуток времени более одного события пренебрежимо мала по сравнению с вероятностью наступления одного события за этот промежуток времени);

— стационарности (вероятность наступления того или иного числа событий за какой-либо промежуток времени зависит только от длины этого промежутка и не зависит от его начала).

В простейшем пуассоновском потоке П событий случайная величина — промежуток времени между двумя соседними событиями — распределена по показательному закону:

$$f(t) = \lambda e^{-\lambda t}, \quad (t \geqslant 0) \tag{27.1}$$

где λ — интенсивность потока П.

Рассмотрим ряд примеров: что принять за СМО, каналы обслуживания и их число, поток заявок на входе, поток обслуживания на выходе?

Пример 1. Вызов абонента, имеющего только один телефонный номер, через АТС. Здесь поток заявок является случайным, если абонент занят, очередная поступающая заявка получает отказ в обслуживании; АТС — это одноканальная СМО (канал обслуживания — линия связи с телефонным номером абонента) с отказами.

Пример 2. Инструментальная кладовая с тремя кладовщиками, выдающими рабочим по их требованию одинаковые наборы инструментов во время работы; если все кладовщики заняты, очередному рабочему инструмент не выдается. Здесь кладовую следует принять за СМО (поток заявок и поток обслуженных заявок являются случайными и зависят только от настоящего состояния), число каналов обслуживания — три; СМО с отказами.

Пример 3. Работа телефонной справочной Центрального железнодорожного агентства. Многоканальная СМО, число каналов — количество дежурных операторов, СМО с ограниченной очередью (ограничение на длину очереди — память накопителя вызовов).

Пример 4. Железнодорожная станция принимает на 5 путей пассажирские поезда и электрички, которые прибывают по расписанию каждые 15 минут на каждый из них и отбывают после обслуживания также по расписанию через 12 мин. В данном случае теория систем массового обслуживания неприменима, поскольку входной и выходной потоки (приход и уход обслуженных поездов организован по расписанию) не являются случайными.

27.4. Одноканальная СМО с отказами

Пусть СМО включает только один канал обслуживания и на ее вход подается пуассоновский поток заявок с интенсивностью λ, т.е. непрерывная случайная величина T — время между двумя соседними заявками — распределена по закону Пуассона (27.1); другая случайная величина T_s — время обслуживания каналом одной заявки — также распределена по закону Пуассона с параметром μ:

$$f_1(t) = \mu e^{-\mu t}. \tag{27.2}$$

Параметры λ и μ называются соответственно *интенсивностью потока заявок* и *интенсивностью потока обслуживания*. Например, среднее значение $\overline{T}_s$ равно математическому ожиданию $M(T_s)$, откуда и следует формула и название μ:

$$M(T_s) = \int_0^T t f(t) dt = \int_0^T t\mu e^{-\mu t} dt \approx 1/\mu. \tag{27.3}$$

Состояния СМО характеризуются простаиванием или занятостью ее канала, т.е. система может находиться в одном из двух состояний: s_0 — канал свободен (простаивает) или s_1 — канал занят. Из состояния s_0 в состояние s_1 систему переводит поток входящих заявок, а из состояния s_1 в состояние s_0 — поток обслуживаний. *Плотности вероятностей перехода* из состояния s_0 в состояние s_1 и обратно равны соответственно λ и μ. Граф состояний СМО показан на рис. 27.2.

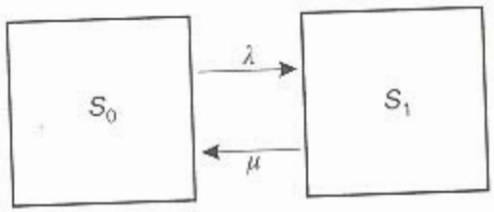

Рис. 27.2

Пусть $p_0(t)$ и $p_1(t)$ — вероятности событий, состоящих в том, что в момент времени t СМО находится в состояниях s_0 и s_1 соответственно (вероятности состояний). Очевидно, что справедливо нормировочное условие:

$$p_0(t) + p_1(t) = 1. \tag{27.4}$$

В силу того что случайный процесс, протекающий в СМО, является марковским, вероятности $p_0(t)$ и $p_1(t)$ удовлетворяют системе уравнений Колмогорова [9]:

$$\begin{cases} \dfrac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t), \\ \dfrac{dp_1(t)}{dt} = -\mu p_1(t) + \lambda p_0(t). \end{cases} \tag{27.5}$$

Подстановка нормировочного условия (27.4) в эту систему приводит к одному обыкновенному дифференциальному уравнению относительно $p_0(t)$

$$\frac{dp_0(t)}{dt} + (\lambda + \mu)p_0(t) = \mu, \quad t \geqslant 0. \tag{27.6}$$

При условии, что в начальный момент времени $t = 0$ канал свободен:

$$p_0(0) = 1; \quad p_1(t) = 0, \tag{27.7}$$

получаем решение уравнения (27.6):

$$p_0(t) = [\mu + \lambda \exp(-(\lambda + \mu)t)]/(\lambda + \mu). \tag{27.8}$$

Затем из нормировочного условия (27.4) получаем формулу для $p_1(t)$:

$$p_1(t) = \lambda[1 - \exp(-(\lambda + \mu)t)]/(\lambda + \mu). \tag{27.9}$$

Из формул (27.8) и (27.9) видно, что с ростом времени t вторые слагаемые в квадратных скобках стремятся к нулю, т.е. $p_0(t)$ и $p_1(t)$ заметно отличны от постоянных величин лишь в начальный интервал времени после начала работы СМО. Поскольку нас интересуют осредненные характеристики работы СМО (установившийся режим работы СМО), то естественно в дальнейшем рассматривать предельные значения вероятностей состояния СМО, т.е. при $t \to \infty$; тогда из (27.8) и (27.9) получаем предельным переходом:

$$\begin{aligned} &\lim_{t\to\infty} p_0(t) = p_0 = \mu/(\lambda + \mu), \\ &\lim_{t\to\infty} p_1(t) = p_1 = \lambda/(\lambda + \mu). \end{aligned} \tag{27.10}$$

Заметим, что при $\mu < \lambda$ вероятность отказа выше 0,5 и превышает вероятность обслуживания.

Теперь обратимся к установлению основных характеристик СМО (см. п. 27.2). Поскольку вероятность обслуживания поступивших заявок равна p_0, а относительная пропускная способность Q равна отношению среднего числа обслуженных заявок к среднему числу поступивших заявок за единицу времени, то $Q = p_0$, т.е. для одноканальной СМО с отказами:

$$Q = \mu/(\lambda + \mu). \tag{27.11}$$

Абсолютная пропускная способность СМО — это среднее число заявок, обслуживаемое СМО в единицу времени, или интенсивность выходящего потока обслуженных заявок — это часть интенсивности входящего потока заявок:

$$A = \lambda\mu/(\lambda + \mu). \tag{27.12}$$

Из второй группы показателей (качества работы СМО) укажем вероятность отказа в обслуживании заявки, когда канал занят; это вероятность p_1:

$$p_r = p_1 = \lambda/(\lambda + \mu). \tag{27.13}$$

Среднее время обслуживания заявки, как говорилось выше, есть величина, обратная μ:

$$\overline{T}_s = 1/\mu. \tag{27.14}$$

Аналогично, среднее время простоя канала равно:

$$\overline{T}_{st} = 1/\lambda. \tag{27.15}$$

Наконец, среднее время пребывания заявки в системе рассчитывается по формуле:

$$\overline{T}_{\text{sys}} = p_0\overline{T}_s = 1/(\lambda + \mu) = \overline{T}_s\overline{T}_{st}/(\overline{T}_s + \overline{T}_{st}). \tag{27.16}$$

Обратимся теперь к решению задач.

Пример 5. Телефонная АТС имеет одну линию, на которую в среднем приходит 0,8 вызова в минуту. Среднее время разговора 1,5 мин. Вызов, пришедший во время разговора, не обслуживается. Считая потоки вызовов пуассоновскими, найти абсолютную и относительную пропускную способности станции и вероятность отказа абоненту.

Решение. Телефонную станцию рассматриваем как одноканальную СМО с отказами. $\overline{T}_s = 1/\mu = 1{,}5$ мин; интенсивности поступающего и обслуженного потоков заявок равны соответственно $\lambda = 0{,}8$, $\mu = 0{,}67$. Тогда по формулам (27.11) и (27.12) имеем: $Q = 0{,}455$, $p_r = 0{,}545$, $A = \lambda Q = 0{,}364$ выз./мин. Заметим здесь, что абсолютная пропускная способность СМО оказалась почти вдвое меньше интенсивности μ потока обслуживания; это обусловлено случайным характером потока заявок.

Пример 6. На телефонную линию приходит простейший поток вызовов с интенсивностью $\lambda = 0{,}9$ выз./мин; производительность линии 0,7 выз./мин. Вызов, пришедший на линию во время ее занятости, не обслуживается. Найти абсолютную пропускную способность линии, среднее время обслуживания одного вызова, вероятность отказа в обслуживании, а также среднее время пребывания заявки в системе.

Решение. Используя формулы (27.12), (27.13), (27.14) и (27.16) последовательно получаем: $A = 0{,}394$ выз/мин; $p_{\text{т}} = 0{,}5625$; $\overline{T}_s = 1{,}43$ мин; $\overline{T}_{sys} = 0{,}625$ мин.

27.5. Многоканальная СМО с отказами

Рассмотрим работу многоканальной СМО ($n > 1$) с отказами. На ее вход поступает простейший поток заявок с интенсивностью λ, а поток обслуживаний каждым каналом является также простейшим с интенсивностью $\mu = 1/\overline{T}_s$. Занумеруем состояния системы по числу занятых каналов (по числу заявок, находящихся в системе, — каждый канал в любой момент времени либо свободен, либо обслуживает только одну заявку). Таким образом СМО может находиться только в одном из $n+1$ состояний: от состояния s_0 (все каналы свободны) до состояния s_n — все каналы заняты. Заявка, поступившая в систему, когда заняты все n каналов, получает отказ в обслуживании и покидает СМО.

Граф состояний многоканальной СМО с отказами представлен на рис. 27.3.

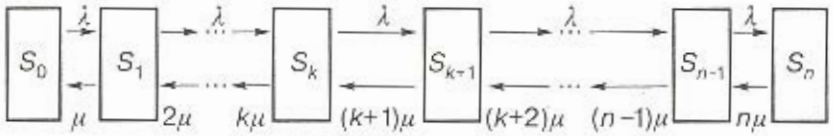

Рис. 27.3

Входной поток заявок с интенсивностью λ переводит систему из любого состояния s_k ($k = 1, \ldots, n-1$) в соседнее состояние справа s_{k+1}, причем плотность вероятности этого перехода одинакова и равна λ. В силу ординарности входного потока заявок СМО может переходить слева направо только в соседние состояния без перескока через какое-либо состояние.

Переход СМО в направлении справа налево из состояния s_k, когда k каналов заняты ($k = 1, \ldots n$), в состояние s_{k-1}, когда освобождается один из занятых каналов, происходит под воздействием суммарного потока обслуживаний с интенсивностью $k\mu$. Следовательно, плотность вероятности такого перехода равна $k\mu$. В этом случае система также не может перескакивать через состояние, а только последовательно переходит из одного состояния в другое.

Воспользуемся правилом составления дифференциальных уравнений Колмогорова, по которому изменение вероятности какого-либо состояния равно сумме входящих и выходящих из этого состояния (квадратика на графе) потоков, умноженных на их интенсивности. При этом выходящий поток считается со знаком минус, а входящий — со знаком плюс (аналог правила Кирхгофа из электрофизики). Тогда для вероятностей состояний $p_0(t)$, $p_1(t)$, ..., $p_n(t)$ получаем систему из $n+1$ дифференциальных уравнений первого порядка, содержащую $n+1$ неизвестных функций:

$$\begin{cases} p_0'(t) = -\lambda p_0(t) + \mu p_1(t), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ p_k'(t) = \lambda p_{k-1}(t) - (\lambda + k\mu) p_k(t) + \\ \qquad +(k+1)\mu p_{k+1}(t), \quad k = 1, \ldots, n-1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ p_n'(t) = \lambda p_{n-1}(t) - n\mu p_n(t). \end{cases} \tag{27.17}$$

Вероятности состояний в любой момент времени t удовлетворяют нормировочному условию:

$$p_0(t) + p_1(t) + \ldots + p_n(t) = 1, \tag{27.18}$$

т.е. в любой момент времени t система находится в одном из $n+1$ состояний, указанных на графе рис. 27.3.

Поскольку нас интересует предельный режим работы СМО в установившемся состоянии, то мы будем рассматривать предельные вероятности состояний системы, т.е. когда

$$p_k = \lim_{t \to \infty} p_k(t), \quad p_k' = 0, \quad k = 0, 1, \ldots, n. \tag{27.19}$$

Подстановка этих соотношений в уравнения (27.17) приводит к системе $n+1$ линейных алгебраических уравнений относительно $n+1$ неизвестных — предельных вероятностей p_0, p_1, ..., p_n:

$$\begin{cases} -\lambda p_0 + \mu p_1 = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \lambda p_{k-1} - (\lambda + k\mu)p_k + \\ \quad +(k+1)\mu p_{k+1} = 0, \quad k = 2, \dots, n-1, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \lambda p_{n-1} - n\mu p_n = 0. \end{cases} \tag{27.20}$$

Поскольку эта система является вырожденной (существует ненулевое решение однородной системы уравнений), то к этим уравнениям следует добавить и уравнение нормировки в стационарной форме:

$$p_0 + p_1 + \dots + p_n = 1. \tag{27.21}$$

Решение системы уравнений (27.20), (27.21) выражаются по следующим формулам:

$$p_0 = 1 \Big/ \left(\sum_{i=0}^{n} \rho^i / i! \right), \quad p_k = p_0 \rho^k / k!, \quad k = 1, \dots, n. \tag{27.22}$$

В формулах (27.22)

$$\rho = \lambda / \mu \tag{27.23}$$

— *приведенная интенсивность* входящего потока (или показатель нагрузки СМО, или трафик), измеряемая в эрлангах. Формулы (27.22) называются *формулами Эрланга*. Они выражают предельные вероятности состояний СМО через число каналов n и показатель нагрузки ρ. Следует иметь в виду, что $0! = 1$ (первая формула (27.22)).

Теперь обратимся к основным характеристикам многоканальной СМО с отказами.

Отказ в обслуживании заявки, по определению СМО с отказами, наступает, когда все каналы заняты и система находится в состоянии s_0; следовательно, вероятность отказа равна вероятности p_n:

$$p_r = p_n = p_0 \rho^n / n!. \tag{27.24}$$

Поскольку событие обслуживания заявки и событие отказа в ее обслуживании являются противоположными, то вероятность обслуживания заявки (вероятность того, что свободен хотя бы один канал) равна:

$$p_s = 1 - p_r = 1 - p_n. \tag{27.25}$$

Относительная пропускная способность СМО определяется вероятностью ее обслуживания (см. раздел 27.4):

$$Q = p_s = 1 - p_r = 1 - p_n. \tag{27.26}$$

Абсолютная пропускная способность СМО (она же интенсивность потока обслуженных заявок):

$$A = \lambda Q = \lambda(1 - p_n). \tag{27.27}$$

Для многоканальной СМО одной из важных характеристик ее эффективности является *среднее число занятых каналов* $\overline{K}$ — отношение абсолютной пропускной способности к интенсивности канала обслуживания; в случае СМО с отказами эта величина совпадает со средним числом заявок, находящихся в системе $\overline{N}_{\text{sys}}$:

$$\overline{K} = \overline{N}_{\text{sys}} = A/\mu = \rho(1 - p_n) = \lambda(1 - p_n)/\mu. \tag{27.28}$$

Наконец, приведем формулу для среднего времени пребывания заявки в СМО; эта характеристика относится ко всем заявкам — как к обслуженным, так и получившим отказ:

$$\overline{T}_{\text{sys}} = \overline{N}_{\text{sys}}/\lambda. \tag{27.29}$$

Рассмотрим задачи на применение многоканальных СМО с отказами.

Пример 7. В отделении банка на обслуживании клиентов работают 3 оператора. Среднее время обслуживания одного клиента оператором 12 минут. В среднем за час в банк обращаются 15 клиентов. Если все операторы заняты, клиенты не обслуживаются банком. Найти основные средние характеристики работы банка, а также вероятность того, что не менее двух каналов простаивают.

Решение. При условии того, что потоки заявок являются простейшими, банк можно принять как СМО с отказами, число каналов $n = 1$. Приняв за единицу времени час, получаем, что интенсивность входного потока заявок $\lambda = 15$ кл./час, интенсивность одного канала обслуживания $\mu = 5$ кл./час. Тогда $\rho = \lambda/\mu = 3$. Используя формулы (27.22), (27.24)–(27.29), последовательно получаем:

$$p_0 = 1/(1 + 3 + 3^2/2 + 3^3/6) = 0{,}077;$$
$$p_r = p_3 = p_0 3^3/6 = 0{,}346;$$
$$Q = p_s = 1 - p_r = 1 - p_n = 0{,}654.$$

Из этих расчетов следует, что из каждых 100 клиентов, обратившихся в банк, в среднем будут обслужены около 65 клиентов. При этом абсолютная пропускная способность СМО составит $A = \lambda Q = 9{,}81$ кл./час (это меньше суммарной интенсивности потока обслуженных заявок $n\mu = 15$ кл./час из-за случайности потоков заявок); среднее число занятых каналов или заявок в системе $\overline{K} = \overline{N}_{\text{sys}} = A/\mu = 1{,}962$; среднее время пребывания заявки в СМО $\overline{T}_{\text{sys}} = \overline{N}_{\text{sys}}/\lambda = 0{,}131$ час $= 7{,}85$ мин. Наконец, вероятность того, что не менее двух каналов простаивают, равна сумме вероятностей $p_0 + p_1 = 0{,}077(1+3) = 0{,}308$.

Пример 8. В условиях предыдущей задачи рассчитать, как изменятся основные характеристики работы СМО, если квалификация операторов такова, что позволяет каждому из них тратить на обслуживание одного клиента в среднем 10 минут.

Решение. В этом случае $\mu = 6$ кл./час, т.е. производительность каждого оператора повышается на 20%. Тогда при прежней интенсивности входного потока заявок получаем, что $\rho = 2{,}5$. Теперь имеем:

$$p_0 = 1/(1 + 2{,}5 + 2{,}5^2/2 + 2{,}5^3/6) = 0{,}108;$$
$$p_r = p_3 = p_0 2{,}5^3/6 = 0{,}28; \quad Q = p_s = 1 - p_r = 0{,}72;$$
$$A = \lambda Q = 10{,}8 \text{ кл./час}; \quad \overline{K} = \overline{N}_{\text{sys}} = A/\mu = 1{,}8;$$
$$\overline{T}_{\text{sys}} = \overline{N}_{\text{sys}}/\lambda = 0{,}12 \text{ часа } = 7{,}2 \text{ мин}.$$

Из этих расчетов можно сделать следующие выводы: в результате повышения производительности труда каждого из операторов на 20% вероятность обслуживания возросла на 10% (из

каждых 100 клиентов, обратившихся в банк, в среднем будут обслужены 72), пропускная способность СМО возросла на 7%, среднее число занятых каналов (среднее число заявок в системе) и среднее время пребывания заявки в системе уменьшились на 8%.

Пример 9. В условиях примера 7 определить минимально целую среднюю производительность операторов (интенсивность потока обслуживания одним каналом μ), чтобы вероятность обслуживания клиентов банка была бы не ниже 0,9.

Решение. Из условия задачи $1 - p_n \geqslant 0{,}9$ следует, что $p_n \leqslant 0{,}1$. Путем подбора определяем: при $\mu = 11$ $p_3 = 0{,}114$; при $\mu = 12$ $p_3 = 0{,}097$, т.е. $p_s = 0{,}903$, что и приводит к ответу: $\mu \geqslant 12$ кл/час.

27.6. Многоканальная СМО с ожиданием и ограничением на длину очереди

Рассмотрим многоканальную СМО ($n \geqslant 1$) с ожиданием, т.е. заявка, поступившая в СМО в момент времени, когда все каналы заняты, в отличие от СМО с отказами, не покидает систему необслуженной, а становится в очередь и ожидает обслуживания. Следует отметить, что большинство обсуживающих фирм и учреждений устроены как раз по такому принципу. Пусть максимальное число мест в очереди равно $m \geqslant 1$, т.е. в очереди могут ожидать своего обслуживания не более m заявок. Поэтому заявка, пришедшая на вход в СМО в момент, когда в очереди уже находятся m заявок, получает отказ и покидает систему. Иными словами, "заполнение" СМО заявками из входного потока идет в два этапа: сначала происходит загрузка каналов обслуживания, затем заполняется очередь. Нумерация состояний системы в этом случае имеет следующий вид: от состояния s_0 (в СМО нет заявок и все каналы свободны) до состояния s_n (в СМО n заявок и все каналы заняты) очереди нет; от состояния s_{n+1} (в СМО $n+1$ заявка, все каналы заняты и одна заявка находится в очереди) до состояния s_{n+m} (все каналы заняты и все m мест в очереди заняты заявками) происходит заполнение очереди.

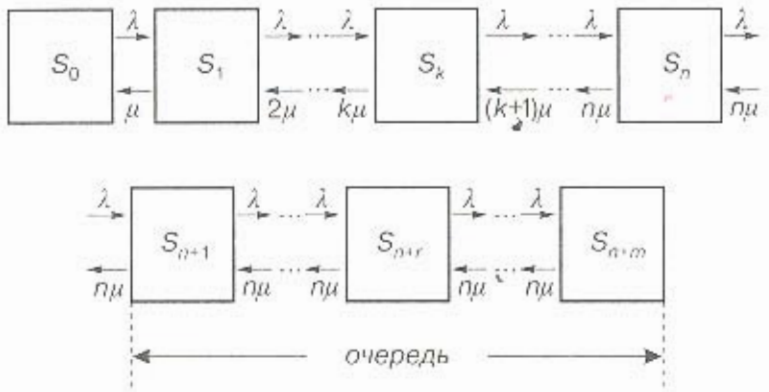

Рис. 27.4

Граф состояний СМО показан на рис. 27.4. Переход системы из состояния s_k в состояние s_{k+1} слева направо ($k = 0, 1, ..., n+m-1$) происходит под воздействием одного и того же входного потока заявок интенсивности λ; следовательно, плотности вероятности перехода из состояния в состояние слева направо одинаковы и равны λ.

Переход системы из состояния в состояние справа налево происходит с разными плотностями вероятностей внутри двух циклов состояний, отмеченных выше. Если заявка продолжает оставаться в очереди (состояние s_k, $n+1 \leqslant k \leqslant n+m$), т.е. все каналы заняты, то эти переходы имеют плотность вероятности, равную $n\mu$ (перемещение системы из состояния в состояние обусловлено общей работой n каналов). Если система находится в состоянии, когда занято k каналов ($1 \leqslant k \leqslant n$), то переход ее в левое состояние обусловлен потоком, представляющим собой сумму k потоков обслуживаний (общей работой k каналов); в таком случае плотность вероятности перехода равна $k\mu$.

Следует отметить, что, как и в предыдущих случаях, система не может "перескакивать" через промежуточное состояние, а переходит из состояния в состояние последовательно: либо слева направо, либо справа налево по графу состояний.

Предельные вероятности $p_0, p_1, \ldots, p_n, p_{n+1}, \ldots, p_{n+m}$ соответствующих состояний СМО удовлетворяют системе линейных однородных алгебраических уравнений, которая получается из системы дифференциальных уравнений Колмогорова

путем, аналогичным описанному в п. 27.3. Эта система имеет вид:

$$\begin{cases} -\lambda p_0 + \mu p_1 = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \lambda p_{k-1} - (\lambda + k\mu)p_k + \\ \quad +(k+1)\mu p_{k+1} = 0, \quad k = 2, \ldots, n-1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \lambda p_{n-1} - (\lambda + n\mu)p_n + n\mu p_{n+1} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \lambda p_{n+k-1} - (\lambda + n\mu)p_{n+k} + \\ \quad +n\mu p_{n+k+1} = 0, \quad k = 1, \ldots, m-1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \lambda p_{n+m-1} - n\mu p_{n+m} = 0. \end{cases} \tag{27.30}$$

К этой системе уравнений необходимо добавить нормировочное условие

$$\sum_{k=0}^{n+m} p_k = 1. \tag{27.31}$$

Введем величину $\psi = \rho/n = \lambda/(n\mu)$ — *показатель нагрузки на один канал*. Решение системы уравнений (27.30), (27.31) выражается, как и в предыдущем случае СМО с отказами, через вероятность простоя системы (или вероятность того, что все каналы свободны) p_0:

$$p_0 = \left(\sum_{k=0}^{n} \frac{n^k}{k!}\psi^k + \frac{n^n}{n!} \sum_{k=n+1}^{n+m} \psi^k \right)^{-1}.$$

Второе слагаемое в правой части этого равенства является суммой m членов геометрической прогрессии с первым членом ψ^{n+1} и знаменателем ψ, т.е. формула для p_0 упрощается:

$$p_0 = \begin{cases} \left(\sum\limits_{k=0}^{n} \dfrac{n^k}{k!}\psi^k + \dfrac{n^n}{n!}\dfrac{\psi^{n+1}(1-\psi^m)}{1-\psi} \right)^{-1}, & \psi \neq 1; \\ \left(\sum\limits_{k=0}^{n} \dfrac{n^k}{k!}\psi^k + \dfrac{n^n}{n!}m \right)^{-1}, & \psi = 1. \end{cases} \tag{27.32}$$

Остальные предельные вероятности состояний имеют вид, аналогичный формулам (27.22):

$$p_k = \begin{cases} (n^k/k!)\psi^k p_0, & k = 1, 2, \ldots, n; \\ (n^n/n!)\psi^k p_0, & k = n+1, \ldots, n+m. \end{cases} \qquad (27.33)$$

Вероятность отказа (заявка, поступившая в момент, когда заняты все n каналов и все m мест в очереди) есть вероятность того, что СМО находится в состояний s_{n+1}, откуда получаем:

$$p_r = p_{n+m} = (n^n/n!)\psi^{n+m} p_0. \qquad (27.34)$$

Так как события отказа заявки и приема ее в СМО являются противоположными, то вероятность приема заявки в СМО равна вероятности p_{sys} и абсолютной пропускной способности системы:

$$p_{\text{sys}} = Q = 1 - p_r = 1 - (n^n/n!)\psi^{k+m} p_0. \qquad (27.35)$$

Отсюда абсолютная пропускная способность:

$$A = \lambda Q = \lambda[1 - (n^n/n!)\psi^{k-m} p_0]. \qquad (27.36)$$

Так как каждый занятый канал обслуживает в среднем μ заявок в единицу времени, то среднее число заявок, находящихся под обслуживанием (или среднее число занятых каналов), подсчитывается по формуле:

$$\overline{N}_s = \overline{K} = A/\mu = \rho Q = \rho[1 - (n^n/n!)\psi^{k+m} p_0]. \qquad (27.37)$$

Для вычисления среднего числа заявок $\overline{N}_{\text{line}}$, находящихся в очереди, рассмотрим дискретную случайную величину N_{line} — число заявок в очереди. Закон распределения N_{line} имеет вид:

N_{line}	0	1	2	...	m
P	p_s	p_{n-1}	p_{n+2}	...	p_{n+m}

Здесь $p_s = \sum_{k=1}^{n} p_k$, поскольку событие, состоящее в том, что в очереди нет ни одной заявки, является объединением событий, состоящих в том, что СМО находится в одном из состояний s_0, $s_1, \ldots, s_n$. Так как $\overline{N}_{\text{line}}$ является математическим ожиданием случайной величины N_{line}, то отсюда получаем и преобразуем:

$$\overline{N}_{\text{line}} = M[N_{\text{line}}] = \sum_{k=1}^{m} k p_{k+n} = (n^n/n!)p_0 \sum_{k=1}^{m} k\psi^{k+n} =$$

$$= (n^n/n!)p_0\psi^{n+1} \sum_{k=1}^{m} k\psi^{k-1}.$$

Сумма в последнем равенстве имеет выражение в виде компактной формулы как при $\psi \neq 1$ (формула суммирования), так и при $\psi = 1$ (сумма отрезка натурального ряда чисел). Соответственно мы получаем окончательное выражение для среднего числа заявок в очереди:

$$\overline{N}_{\text{line}} = \begin{cases} \dfrac{n^n}{n!}\psi^{n-1}\dfrac{1 - \psi^m(m+1-m\psi)}{(1-\psi)^2} p_0, & \psi \neq 1; \\ \dfrac{n^n}{n!}\dfrac{m(m+1)}{2} p_0, & \psi = 1. \end{cases} \qquad (27.38)$$

Среднее число заявок, находящихся в системе, равно сумме средних чисел обслуживаемых заявок и заявок, находящихся в очереди:

$$\overline{N}_{\text{sys}} = \overline{N}_{\text{line}} + \overline{N}_s. \qquad (27.39)$$

Для средних величин времени обслуживания заявки, времени ожидания заявки в очереди и времени пребывания заявки в системе имеем соответственно следующие формулы (формулы Литтла):

$$\overline{T}_s = \overline{N}_s/\lambda, \qquad (27.40)$$

$$\overline{T}_{\text{line}} = \overline{N}_{\text{line}}/\lambda, \qquad (27.41)$$

$$\overline{T}_{\text{sys}} = \overline{T}_s + \overline{T}_{\text{line}} = (\overline{N}_s + \overline{N}_{\text{line}})/\lambda = \overline{N}_{\text{sys}}/\lambda. \qquad (27.42)$$

Формулы (27.32)–(27.42) позволяют рассчитать все характеристики работы многоканальной СМО с ожиданием и ограничением на длину очереди. Рассмотрим несколько примеров.

Пример 10. В пункте валютного обмена работают два оператора, каждый из которых обслуживает клиента в среднем за 2,5 мин. По условиям безопасности в помещении пункта может

находиться одновременно не более 5 человек, включая обслуживаемых клиентов. Если помещение заполнено, то очередной клиент не становится в очередь, а уходит. В среднем клиенты приходят каждые 2 мин. Найти основные характеристики работы обменного пункта.

Решение. Математической моделью данного обменного пункта является двухканальная СМО ($n = 2$) с ожиданием и ограничением на длину очереди ($m = 3$). Предполагается, что потоки прибывающих клиентов и обслуживания — простейшие. Интенсивность входящего потока $\lambda = 0{,}5$ кл./мин, интенсивность потока обслуживания $\mu = 0{,}4$ кл./мин; отсюда показатель нагрузки СМО $\rho = \lambda/\mu = 1{,}25$ эрланга, а показатель нагрузки на один канал $\psi = \rho/n = 0{,}625$ эрланга. Вероятность простоя системы (оба канала свободны) определяется по формуле (27.32) при $n = 2$, $m = 3$ и $\psi \neq 1$:

$$p_0 = [1 + (2/1!) \cdot 0{,}625 + (2^2/2!) \cdot 0{,}625^2 + \\ + (2^2/2!) \cdot (0{,}625^3(1 - 0{,}625^3)/(1 - 0{,}625)]^{-1} = 0{,}249.$$

Вероятность отказа в соответствии с формулой (27.34) равна

$$p_r = p_5 = (2^2/2!) \cdot 0{,}625^5 \cdot 0{,}249 = 0{,}047,$$

т.е. из каждых 100 клиентов, обратившихся в обменный пункт, в среднем обслуживаются 95 клиентов.

Относительная пропускная способность (формула (27.35)):

$$Q = 1 - 0{,}047 = 0{,}953.$$

Абсолютная пропускная способность, согласно формуле (27.36):

$$A = \lambda Q = 0{,}5 \cdot 0{,}953 = 0{,}477.$$

Далее находим по формулам (27.37)–(27.42) остальные характеристики работы СМО:

— среднее число клиентов, находящихся под обслуживанием (среднее число занятых операторов)

$$\overline{N}_s = \overline{K} = A/\mu = \rho/Q = 1{,}25 \cdot 0{,}953 = 1{,}191;$$

— среднее число клиентов, ожидающих в очереди

$$\overline{N}_{\text{line}} = (2^2/2!) \cdot 0{,}625^3 \cdot [(1 - 4 \cdot 0{,}625^3 + \\ - 3 \cdot 0{,}625^4)/(1 - 0{,}625)^2] \cdot 0{,}249 = 0{,}416;$$

— среднее число клиентов в системе

$$\overline{N}_{\text{sys}} = \overline{N}_{\text{line}} + \overline{N}_s = 1{,}191 + 0{,}416 = 1{,}607;$$

— среднее время обслуживания клиента

$$\overline{T}_s = \overline{N}_s/\lambda = 1{,}191/0{,}5 = 2{,}382 \text{ мин};$$

— среднее время пребывания клиента в очереди

$$\overline{T}_{\text{line}} = \overline{N}_{\text{line}}/\lambda = 0{,}416/0{,}5 = 0{,}832 \text{ мин};$$

— среднее время пребывания клиента в пункте валютного обмена

$$\overline{T}_{\text{sys}} = \overline{T}_{\text{s}} + \overline{T}_{\text{line}} = 2{,}382 + 0{,}832 = 3{,}21 \text{ мин}.$$

Пример 11. В условиях примера 10 определить, как повлияет на вероятность отказа увеличение числа операторов до трех человек.

Решение. В данном случае число каналов $n = 3$, число мест в очереди $m = 2$; вероятность простоя системы, согласно формуле (27.32)

$$p_0 = [1 + (2/1!) \cdot 0{,}625 + (2^2/2!) \cdot 0{,}625^2 + (2^3/3!) \cdot 0{,}625^3 + \\ + (2^3/3!) \cdot 0{,}625^4(1 - 0{,}625^2)/(1 - 0{,}625)]^{-1} = 0{,}271.$$

Тогда вероятность отказа равна

$$p_r = p_5 = (2^3/3!) \cdot 0{,}625^5 \cdot 0{,}271 = 0{,}034.$$

По сравнению с предыдущим случаем вероятность отказа изменилась весьма незначительно: теперь из каждых 100 клиентов, обратившихся в пункт, в среднем будут обслуживаться около 97 человек, т.е. характеристика показателя числа обслуживаемых клиентов изменилась в сторону улучшения всего на 2%. По-видимому, в данном случае целесообразнее будет не увеличивать число операторов, а уменьшить время обслуживания (увеличить производительность μ обслуживания каждым оператором).

Формулы (27.32)–(27.42) легко распространяются на **частный случай одноканальной СМО с ожиданием и ограничением на длину очереди**. Полагая $n = 1$ и $\psi = \rho$, получаем вероятность простоя системы:

$$p_0 = \begin{cases} \left(\sum_{k=0}^{m+1} \rho^k\right)^{-1} = \dfrac{1-\rho^{m+2}}{1-\rho}, & \rho \neq 1 \\ \left(\sum_{k=0}^{m+1}\right)^{-1} = \dfrac{1}{m+2}, & \rho = 1. \end{cases} \tag{27.43}$$

Вероятности отказа в обслуживании и принятия заявки в систему определяются формулами:

$$p_r = p_{m+1} = \rho^{m+1} p_0; \quad p_{\text{sys}} = 1 - \rho^{m+1} p_0. \tag{27.44}$$

Аналогичным образом выводятся остальные характеристики работы одноканальной СМО: в формулах (27.36)–(27.42) нужно только положить $n = 1$ и $\psi = \rho$.

27.7. Многоканальная СМО с ожиданием и неограниченной очередью

Рассмотрим СМО, состоящую из n каналов ($n \geqslant 1$), с ожиданием и без ограничения на длину очереди и время ожидания. Такие приложения часто встречаются в практике, когда на длину очереди заявок, ожидающих обслуживания, не накладывается никаких ограничений (например, кассы вокзалов и метрополитена, магазины и торговые точки и пр.). Заявка, поступившая в систему в момент времени, когда все n каналов заняты, становится в очередь и ожидает своего обслуживания; любая поступившая заявка будет обслужена. По-прежнему поток заявок и потоки обслуживания каждым каналом являются простейшими с интенсивностями соответственно λ и μ. Данная СМО может находиться в одном из бесконечного множества состояний:

— s_k ($k = 0, 1, \ldots, n$) — k каналов заняты и очереди нет;

— s_m ($m = 0, 1, \ldots$) все n каналов заняты и в очереди находятся m заявок.

Граф состояний такой СМО аналогичен случаю n-канальной СМО с ожиданием и ограничением на длину очереди с той

лишь разницей, что правого крайнего состояния нет и граф является бесконечным. Предельный (стационарный) режим марковского случайного процесса, протекающего в СМО, будет существовать при условии, что интенсивность обслуживания всех n каналов будет выше интенсивности λ входящего потока, т.е.

$$\lambda < n\mu, \text{ или } \psi < 1. \tag{27.45}$$

Если это условие не выполняется, то очередь заявок в системе будет неограниченно расти с течением времени (переполнение системы).

Для получения расчетных формул для основных характеристик СМО нужно перейти к пределу при $m \to \infty$ в формулах (27.32)–(27.38). Вычисляя соответствующие пределы, получаем сначала формулы для вероятностей состояний:

$$p_0 = \left(\sum_{k=0}^{n} \frac{n^k}{k!}\psi^k + \frac{n^n}{n!}\frac{\psi^{n+1}}{1-\psi}\right)^{-1}; \tag{27.46}$$

$$p_k = \begin{cases} (n^k/k!)\psi^k p_0, & k = 1, 2, \ldots, n; \\ (n^n/n!)\psi^k p_0, & k = n+1, n+2, \ldots. \end{cases} \tag{27.47}$$

Можно непосредственно проверить, что найденные вероятности удовлетворяют нормировочному условию $p_0 + p_1 + \ldots = 1$.

Вычисление вероятности отказа приема в заявке, как и следовало ожидать, при предельном переходе во второй строке (27.47) дает:

$$p_r = 0. \tag{27.48}$$

Тогда вероятность того, что пришедшая заявка будет принята в систему,

$$p_{\text{sys}} = 1, \tag{27.49}$$

что соответствует исходному предположению о свойстве СМО — любая поступившая заявка будет обслужена.

Существенно упрощаются другие формулы. Приведем их. Относительная пропускная способность системы, согласно (27.35):

$$Q = 1. \tag{27.50}$$

43 Зак. 6853

Из (27.36) получаем, что абсолютная пропускная способность СМО определяется только интенсивностью входящего потока:

$$A = \lambda. \qquad (27.51)$$

Аналогично получаем из (27.37), что среднее число занятых каналов (заявок, находящихся под обслуживанием), равно показателю нагрузки СМО:

$$\overline{N}_s = \overline{K} = A/\mu = \rho. \qquad (27.52)$$

Наконец, из (27.38) предельным переходом при $m \to \infty$ получаем формулу для среднего числа заявок, находящихся в очереди:

$$\overline{N}_{\text{line}} = \frac{n^n}{n!} \cdot \frac{\psi^{n+1}}{(1-\psi)^2} p_0. \qquad (27.53)$$

Далее из формул (27.39)–(27.42) получаются остальные характеристики СМО и ее работы.

Частный случай одноканальной СМО с ожиданием и неограниченной очередью получается из формул (27.46), (27.47) и (27.53) при $n = 1$ и $\psi = \rho$:

$$p_k = \rho^k(1-\rho), \quad k = 0, 1, 2, \ldots \qquad (27.54)$$

$$\overline{N}_{\text{line}} = \rho^2/(1-\rho). \qquad (27.55)$$

Пример 12. В кассе метрополитена, продающей карточки на проезд, работают два окна. В среднем один кассир тратит на обслуживание одного пассажира 0,5 мин. В среднем к кассе подходит 3 чел./мин. Найти основные характеристики работы кассы.

Решение. Касса метрополитена моделируется двухканальной СМО с ожиданием и без ограничения на длину очереди. Параметры системы: $n = 2$, $\lambda = 3$ чел./мин, $\mu = 2$ чел./мин. Нагрузка системы $\rho = \lambda/\mu = 1{,}5$, показатель нагрузки на один канал $\psi = \rho/n = 0{,}75$ — т.е. предельный режим работы системы существует, и возможно осуществить расчет основных характеристик эффективности СМО в данном режиме.

Вероятность простоя системы (оба кассира свободны) равна

$$p_0 = 1/[1 + 2 \cdot 0{,}75 + (2^2/2!) \cdot 0{,}75^2 + \\ + (2^2/2!) \cdot 0{,}75^3/(1 - 0{,}75)] = 0{,}143.$$

Среднее число занятых кассиров (среднее число пассажиров под обслуживанием) равно $\overline{N}_s = \overline{K} = A/\mu = \rho = 1{,}5$. Среднее число пассажиров в очереди определяется по формуле (27.53):

$$\overline{N}_{\text{line}} = (2^2/2!) \cdot 0{,}75^3/(1 - 0{,}75)^2 \cdot 0{,}143 = 1{,}931.$$

По формуле (27.39) подсчитываем среднее число пассажиров у касс:

$$\overline{N}_{\text{sys}} = \overline{N}_{\text{line}} + \overline{N}_{\text{s}} = 1{,}931 + 1{,}5 = 3{,}431.$$

По формулам Литтла определяем среднее время пассажира в очереди и среднее время, которое тратит пассажир на приобретение карточки на проезд:

$$\overline{T}_{\text{line}} = \overline{N}_{\text{line}}/\lambda = 1{,}931/3 = 0{,}644 \text{ мин};$$
$$\overline{T}_{\text{sys}} = \overline{N}/\lambda = 3{,}431/3 = 1{,}144 \text{ мин}.$$

Глава 28
ЭЛЕМЕНТЫ УПРАВЛЕНИЯ РИСКАМИ В ЭКОНОМИКЕ

Любая сфера человеческой деятельности, в особенности экономика и бизнес, связана с принятием решений в условиях неполноты информации, которая, в свою очередь, может быть обусловлена разнообразными причинами — как объективными, так и субъективными. Особенно распространенными при формировании инвестиционных портфелей являются ситуации, когда выбор решения осуществляется в условиях рисков: существует неопределенность в виде множества частных исходов результата принятия решения, причем вероятности появления этих исходов либо определяемы тем или иным способом, либо неизвестны или не имеют смысла.

В этом плане проблема соотношения между риском и прибылью является одной из ключевых в экономической деятельности и, в частности, в управлении финансами. Слово "риск" в буквальном переводе означает принятие решения, результат которого заранее неизвестен и может быть небезопасен. Под *риском* принято понимать угрозу потери действующим финансовым лицом части своих ресурсов или появления дополнительных расходов в результате осуществления определенной финансовой политики. В этом разделе мы будем рассматривать *инвестиционный* риск, обусловленный возможным обесцениванием инвестиционно-финансового портфеля, состоящего из собственных и приобретенных ценных бумаг.

28.1. Оптимизация портфелей банка

Математическая постановка задачи оптимального планирования системы портфелей банка заключается в следующем. Требуется найти неизвестные векторы активов и пассивов банка — соответственно

$$\overline{A} = (A_1, A_2, \ldots, A_n) \text{ и } \overline{L} = (L_1, L_2, \ldots, L_n), \qquad (28.1)$$

максимизирующие линейную форму прибыли системы портфелей:

$$Prf(\overline{A}, \overline{L}) = \sum_{k=1}^{n} A_k D_k - \sum_{i=1}^{m} L_i E_i \rightarrow \max, \qquad (28.2)$$

где A_k — сумма инвестиций в отдельный тип активов в портфеле, D_k — доходность отдельного типа активов, k — индекс отдельного типа активов, n — число типов активов в портфеле; L_i — сумма привлечения по отдельному типу обязательств в портфеле пассивов, E_i — процентные расходы по привлечению отдельного типа ресурсов, i — индекс отдельного типа пассивов, m — число типов пассивов в портфеле.

Для формулировки задачи математического программирования необходимо к целевой функции (28.2) добавить условия (ограничения) на компоненты векторов (28.1), которые формируют *допустимое множество*. Эти ограничения разделяются на две группы. К первой из них относятся экономические

нормативы ЦБ. Эта группа содержит следующие обязательные показатели: норматив достаточности капитала; нормативы ликвидности — текущей, мгновенной и долгосрочной; минимально допустимое значение ликвидных активов и суммарных активов кредитной организации; максимальный размер привлеченных денежных вкладов населения; норматив использования собственных средств кредитных организаций для приобретения долей (акций) других юридических лиц; норматив риска собственных вексельных обязательств. Ко второй группе относятся собственные и технологические ограничения банка. Хотя целевая функция (28.2) является линейной, ограничения экономических нормативов ЦБ нелинейны из-за операций деления показателей. Следует отметить, что в число нормативов ЦБ также входят и показатели рисков, связанных с кредитованием, однако в формальные ограничения допустимого множества они не включаются.

В рамках сформулированной задачи можно выделить управление активами банка, суть которого выражается в следующем. Нужно максимизировать целевую функцию

$$M_A = p \cdot f_M(r_M, r_A, r_B, \ldots, \overline{W}) \rightarrow \max \text{ дохода}, \qquad (28.3)$$

где M_A — стоимость активов банка, p — абсолютный уровень цен, r_M — ставка доходности денег, r_A — ставка доходности актива A, r_B — ставка доходности актива B, $\overline{W}$ — норма доходности от величины состояния банка, выраженного в труднореализуемых активах (недвижимость, драгоценные металлы, антиквариат и т.п.). Активы A и B — это ликвидные ценные бумаги (облигации, акции, векселя и т.д.).

Структура портфеля активов должна строиться таким образом, чтобы получить максимально возможную прибыль в конкретной ситуации на финансовом рынке. Эта конкретная ситуация определяет допустимое множество (на котором ищется максимум целевой функции дохода) в виде ограничений на комбинации переменных, входящих в форму (28.3): балансовое, ликвидное и кредитное ограничения. В принципе, управление активами банка можно свести к традиционной задаче линейного программирования в случае линейности целевой функции (28.3) и форм ограничений. Однако при формировании портфеля активов встает еще одна задача, имеющая противоположную цель, — минимизация риска, т.е. на допустимом

множестве решений ищется такое, чтобы помимо цели (28.3) целевая функция удовлетворяла бы также условию:

$$M_A = p \cdot f_M(r_M, r_A, r_B, \ldots, \overline{W}) \to \min \text{ риска.} \qquad (28.4)$$

Риск фактически заложен уже в доходностях D_i и ставках доходностей r_A и r_B, входящих соответственно в целевые функции (28.2) и (28.3). Учет фактора риска в рассмотренных моделях, как правило, задается либо ограничением на определенные группы (виды) активов (пассивов), либо ссылками на экономические нормативы банковской деятельности, которые якобы уже учитывают риски. Далее необходимо ввести понятие меры риска.

28.2. Портфельный анализ

Начало *портфельного анализа* было положено в работах Нобелевского лауреата в области экономики, профессора Чикагского университета Г. Марковица в 1952 г. Им впервые была предложена модель формирования оптимального портфеля и были даны методы построения таких портфелей.

Поскольку ценные бумаги различных видов отличаются по доходности и степени надежности, инвесторы вкладывают средства в приобретение ценных бумаг нескольких видов, стремясь достичь наилучшего соотношения риск — доходность. Принимая решение о приобретении набора ценных бумаг, инвестор должен иметь в виду, что доходность портфеля в предстоящий период владения неизвестна. Однако можно оценить предполагаемую доходность различных ценных бумаг, основываясь на некоторых предположениях. Уровень доходности является случайной величиной, и основными ее характеристиками являются ожидаемое или среднее значение и стандартное отклонение. Именно последнюю характеристику предлагается использовать как меру риска. Инвестор максимизирует ожидаемую доходность и минимизирует риск, т.е. неопределенность. Подход Г. Марковица к принятию решения даст возможность адекватно учесть обе цели, противоречащие друг другу.

Метод, применяемый для выбора наиболее желательного портфеля, использует кривые безразличия. Иными словами, для инвестора существует функция полезности, зависящая от

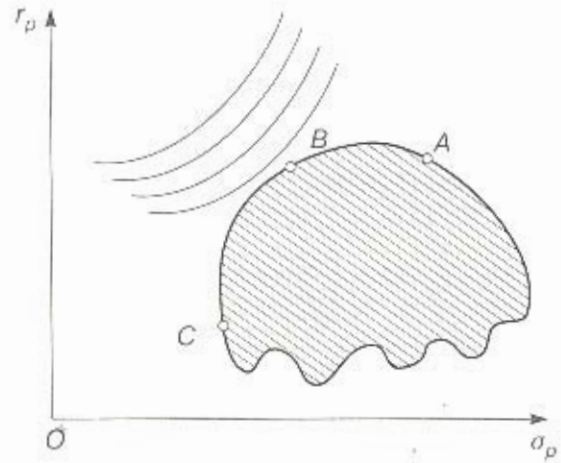

Рис. 28.1

двух аргументов — ожидаемой доходности портфеля r_p и стандартного отклонения (меры риска) σ_p :

$$u = u(r_p, \sigma_p). \qquad (28.5)$$

Все портфели, лежащие на одной линии безразличия или линии уровня функции (28.5)

$$u(r_p, \sigma_p) = I, \qquad (28.6)$$

являются равноценными для инвестора. Линии безразличия отражают отношение инвестора к риску и доходности портфеля и представляют собой кривые в координатах $\sigma_p - r_p$ (рис. 28.1). Инвестор считает любой портфель, лежащий на линии безразличия выше и левее, более привлекательным, чем портфели, лежащие на линии безразличия, которая ниже и правее. Ожидаемая доходность портфеля, состоящего из n ценных бумаг, равна

$$r_p = \sum_{i=1}^{n} x_i r_i, \qquad (28.7)$$

где x_i — доля начальной стоимости портфеля, инвестированная в i-й вид ценных бумаг, r_i — ожидаемая доходность i-го

вида ценных бумаг, n — количество видов ценных бумаг в портфеле.

Стандартное отклонение портфеля $\sigma(r_p)$ вычисляется следующим образом. Дисперсия доходности портфеля — это дисперсия суммы случайных величин; как известно из теории математической статистики, она равна ковариации:

$$D(r_p) = \mathrm{Cov}\left(\sum_{i=1}^{n} x_i r_i, \sum_{j=1}^{n} x_j r_j\right) = \\ = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \mathrm{Cov}\,(r_i, r_j). \tag{28.8}$$

Здесь $\mathrm{Cov}\,(r_i, r_j)$ — ковариация ожидаемых доходностей ценных бумаг i и j, вычисляемая по формуле

$$\mathrm{Cov}\,(r_i, r_j) = \mu_{ij}\sqrt{D(r_i)D(r_j)} = \\ = \mu_{ij}\sigma_i\sigma_j, \quad i, j = 1, 2, \ldots, n, \tag{28.9}$$

где μ_{ij} — коэффициент корреляции между доходностями i-й и j-й ценных бумаг, D и σ — соответственно дисперсия и стандартное (среднее квадратическое отклонение) доходностей ценных бумаг. Как известно,

$$-1 \leqslant \mu_{ij} \leqslant 1.$$

Формула для стандартного отклонения портфеля имеет вид:

$$\sigma(r_p) = \sqrt{D(r_p)}. \tag{28.10}$$

Пример 1. Найти ожидаемую доходность и стандартное отклонение доходности портфеля, состоящего из 30% акций компании A и 70% акций компании B, если их доходности некоррелированы и равны соответственно 20% и 10%, а стандартные отклонения — 10% и 5%.

Решение. По формуле (28.7) получаем: $r_p = 0{,}3 \cdot 20\% + 0{,}7 \cdot 10\% = 13\%$. Поскольку доходности бумаг некоррелированы, то $\mu_{ij} = 0$ при $i \neq j$ и тогда $\sigma(r_p) = \sqrt{0{,}09 \cdot 100 + 0{,}49 \cdot 25} = 4{,}6\%$.

Приведенный пример показывает, что портфель ценных бумаг обладает меньшим риском, чем некоторые отдельные составляющие его бумаги. Это свойство портфеля называется *диверсификацией*: увеличение количества видов ценных бумаг при одновременном сокращении их долей в общей ожидаемой доходности уменьшает риск портфеля. Проиллюстрируем свойство диверсификации еще на одном примере.

Пример 2. Найти ожидаемую доходность и ее стандартное отклонение для портфеля, состоящего из 10 видов ценных бумаг с некоррелированными доходностями. Доли ценных бумаг x_i, их доходности r_i и стандартные отклонения σ_i приведены в таблице.

Параметры	Номера ценных бумаг i									
	1	2	3	4	5	6	7	8	9	10
x_i, %	10	10	10	10	10	10	20	10	5	5
r_i, %	15	15	18	12	25	20	10	28	35	40
σ_i, %	8	8	10	7	12	10	5	15	20	25

Решение.

$$r_p = 0{,}1 \cdot 15 + 0{,}1 \cdot 15 + 0{,}1 \cdot 18 + 0{,}1 \cdot 12 + 0{,}1 \cdot 25 + \\ + 0{,}1 \cdot 20 + 0{,}2 \cdot 10 + 0{,}1 \cdot 28 + 0{,}05 \cdot 35 + 0{,}05 \cdot 40 = 19{,}05\%.$$

Так как случайные величины доходностей бумаг являются независимыми, то дисперсия доходности портфеля равна

$$D(r_p) = 0{,}01 \cdot 64 + 0{,}01 \cdot 64 + 0{,}01 \cdot 100 + 0{,}01 \cdot 49 + \\ + 0{,}01 \cdot 144 + 0{,}01 \cdot 100 + 0{,}04 \cdot 25 + 0{,}01 \cdot 225 + \\ + 0{,}0025 \cdot 400 + 0{,}0025 \cdot 625 = 11{,}02.$$

Тогда $\sigma(r_p) = \sqrt{D(r_p)} = 3{,}32\%$. Видим, что стандартное отклонение доходности портфеля оказалось ниже минимального значения для ценной бумаги с номером 7, а "пиковые" значения стандартных отклонений ценных бумаг с номерами 9 и 10 попросту "растворились" в общей величине $\sigma(r_p)$.

На рис. 28.1 показано *достижимое множество*, представляющее собой все портфели, которые можно сформировать из

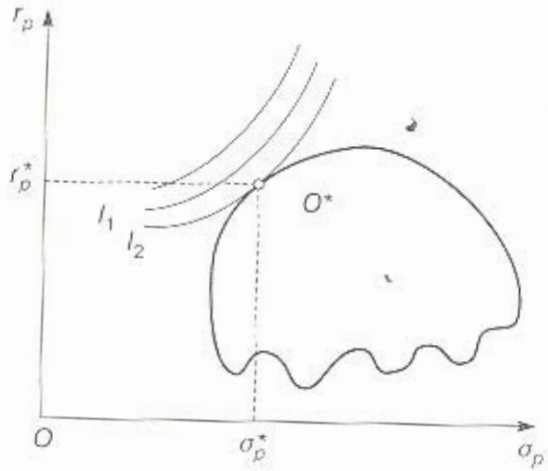

Рис. 28.2

n видов ценных бумаг. Множество портфелей, обеспечивающих минимальный риск при меняющемся уровне ожидаемой доходности, находится на левой части границы достижимого множества, расположенной между точками A и C.

Справедлива теорема об *эффективном множестве портфелей*: инвестор выбирает свой оптимальный портфель из такого множества портфелей, каждый из которых: а) максимизирует ожидаемую доходность для некоторого уровня риска, б) минимизирует риск для некоторого уровня ожидаемой доходности. Согласно этой теореме, инвестора удовлетворяют только портфели, находящиеся на верхней и левой границе достижимого множества, т.е. эффективное множество портфелей представляет собой участок границы AB. На этом множестве инвестор будет выбирать самый оптимальный.

Для выбора оптимального портфеля инвестор должен совместить свои линии безразличия с эффективным множеством (рис. 28.2). Оптимальный портфель будет соответствовать точке, в которой кривая безразличия касается эффективного множества — портфель O^* на кривой безразличия I_2.

Понятно, что оптимальный портфель O^* существенно зависит от формы линий безразличия (28.6), которые, в свою очередь, зависят от функции полезности (28.5), являющейся характеристикой стратегии инвестора. Если инвестор осторо-

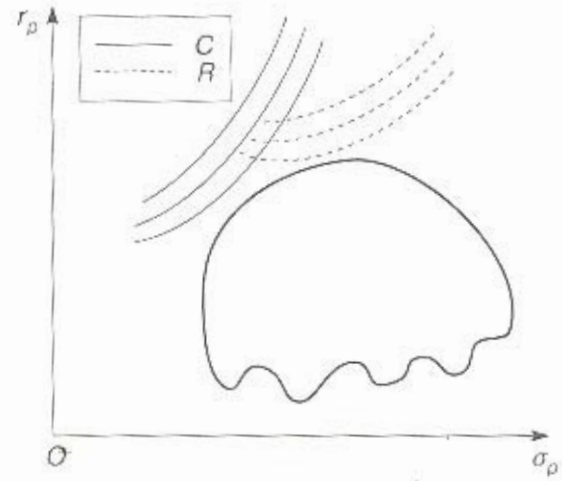

Рис. 28.3

жен и тяготеет к уменьшению риска за счет снижения ожидаемой доходности портфеля, то линии безразличия являются менее выпуклыми вниз (семейство линий C на рис. 28.3). Повышение риска инвестора для достижения более высокого уровня ожидаемой доходности выражается в том, что выпуклость вниз линий безразличия увеличивается (семейство линий R). Эта трактовка соответствует понятию функции полезности Неймана–Моргенштерна, используемой для оценки действий инвестора.

Известный американский ученый-экономист Д. Тобин обобщил задачу минимизации рыночного риска портфелей, показав, что оптимальная структура портфеля ценных бумаг не зависит от склонности (несклонности) инвестора к риску.

28.3. Методика VAR

Классический способ измерения величины риска как стандартного отклонения ожидаемой доходности инвестиционного портфеля (от рыночной стоимости) имеет три существенных недостатка.

1. Завышение оценок риска, поскольку на величину дисперсии (28.8) одинаково влияют как отклонения от средних до-

ходностей ценных бумаг со знаком минус, характеризующие потери (отрицательное изменение стоимости), так и со знаком плюс, характеризующие прибыль (положительное изменение стоимости). Иными словами, формально наименьшая степень риска соответствует ситуации, когда любые отклонения от средних доходностей ценных бумаг, в том числе и в положительную сторону, будут по возможности минимальными.

2. В случае распределения вероятностей, существенно отличного от нормального, особенно для асимметричных распределений, происходит искажение (как увеличение, так и уменьшение) величины риска.

3. Оценка риска в виде стандартного отклонения затрудняет процесс управления риском, так как не показывает величину возможных денежных потерь.

Между тем для инвесторов риск означает прежде всего потери или неблагоприятный результат, связанный с утерей части инвестированных средств или возникновением необходимости дополнительных расходов. Именно поэтому с середины 90-х годов XX столетия Базельским комитетом банковского надзора в качестве альтернативы была предложена методика VAR (Value-at-Risk) для оценки рисков потенциальных убытков при неблагоприятной конъюнктуре рынка.

В условиях диверсификации инвестиционных портфелей ожидаемую доходность (28.7) можно рассматривать, согласно центральной предельной теореме, как случайную величину с нормальным законом распределения (большое число слагаемых случайных величин в сумме). Как и прежде, стандартное отклонение $\sigma(r_p)$ является здесь основной характеристикой. Но в классическом случае, рассмотренном выше, это мера риска, а в новой методике величина $\sigma(r_p)$ используется для определения параметра VAR, который и является мерой риска. Параметр VAR характеризует величину потерь, при которой потери в стоимости инвестиционного портфеля за определенный период времени T_0 не превысят этой величины (потери понимаются как *отрицательное* изменение стоимости). Заметим, что в Базельской методике оценивается *вероятность отрицательного изменения стоимости (потерь) выше определенного уровня потерь*; этот уровень потерь и представляет собой определяемый параметр VAR.

Под термином VAR понимают как методику (совокупность отдельных методов) оценки риска, так и количественную оценку рыночного риска в виде единственного параметра. VAR — это выраженная в данных денежных единицах (базовой валюте) оценка максимальных, ожидаемых в течение данного периода времени с данной вероятностью потерь данного портфеля под воздействием рыночных факторов риска.

ЛИТЕРАТУРА

1. В. Е. Гмурман. *Теория вероятностей и математическая статистика.* М.: Высшая школа, 1998.

2. Е. С. Вентцель. *Исследование операций. Задачи, принципы, методология.* М.: Наука, 1988.

3. О. О. Замков, А. В. Толстопятенко, Ю. Н. Черемных. *Математические методы в экономике.* 2-е изд. М.: Дело и Сервис, 1999.

4. М. Интрилигатор. *Математические методы оптимизации и экономическая теория.* М.: Прогресс, 1975.

5. В. А. Ильин, Э. Г. Позняк. *Основы математического анализа.* Ч. 1 М.: Наука, 1971.

6. В. А. Ильин, Э. Г. Позняк. *Основы математического анализа.* Ч. 2 М.: Наука, 1973.

7. М. С. Красс, Б. П. Чупрынов. *Основы математики и ее приложения в экономическом образовании.* 2-е изд. М.: Дело, 2001.

8. В. Б. Кутуков. *Основы финансовой и страховой математики.* М.: Дело, 1998.

9. Л. Г. Лабскер, Л. О. Бабешко. *Теория массового обслуживания в экономической сфере.* М.: ЮНИТИ, 1998.

10. М. Грегори Мэнкью. *Макроэкономика.* М.: Изд-во МГУ, 1994.

11. В. В. Степанов. *Курс дифференциальных уравнений.* М.: Изд-во физ.-мат. литературы, 1958.

12. В. В. Федосеев. *Экономико-математические модели и методы в маркетинге.* М.: Финстатпром, 1996.

13. И. Ф. Цисарь, В. П. Чистов, А. И. Лукьянов. *Оптимизация финансовых портфелей банков, страховых компаний, пенсионных фондов.* М.: Дело, 1998.

14. Е. М. Четыркин. *Финансовая математика.* М.: Дело, 2000.

15. В. С. Шипачев. *Высшая математика.* М.: Высшая школа, 1985.

16. В. С. Шипачев. *Задачи по высшей математике.* М.: Высшая школа, 1996.

ПРИЛОЖЕНИЕ

Таблица 1

Значения функции $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1093	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034

Окончание табл. 1

x	0	1	2	3	4	5	6	7	8	9
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

44 Зак. 6853

Таблица 2

Значения функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-z^2/2} dz$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,39	0,1517	0,78	0,2823	1,17	0,3790	1,56	0,4406
0,01	0,0040	0,40	0,1554	0,79	0,2852	1,18	0,3810	1,57	0,4418
0,02	0,0080	0,41	0,1591	0,80	0,2881	1,19	0,3830	1,58	0,4429
0,03	0,0120	0,42	0,1628	0,81	0,2910	1,20	0,3849	1,59	0,4441
0,04	0,0160	0,43	0,1664	0,82	0,2939	1,21	0,3869	1,60	0,4452
0,05	0,0199	0,44	0,1700	0,83	0,2967	1,22	0,3883	1,61	0,4463
0,06	0,0239	0,45	0,1736	0,84	0,2995	1,23	0,3907	1,62	0,4474
0,07	0,0279	0,46	0,1772	0,85	0,3023	1,24	0,3925	1,63	0,4484
0,08	0,0319	0,47	0,1808	0,86	0,3051	1,25	0,3944	1,64	0,4495
0,09	0,0359	0,48	0,1844	0,87	0,3078	1,26	0,3962	1,65	0,4505
0,10	0,0398	0,49	0,1879	0,88	0,3106	1,27	0,3980	1,66	0,4515
0,11	0,0438	0,50	0,1915	0,89	0,3133	1,28	0,3997	1,67	0,4525
0,12	0,0478	0,51	0,1950	0,90	0,3159	1,29	0,4015	1,68	0,4535
0,13	0,0517	0,52	0,1985	0,91	0,3186	1,30	0,4032	1,69	0,4545
0,14	0,0557	0,53	0,2019	0,92	0,3212	1,31	0,4049	1,70	0,4554
0,15	0,0596	0,54	0,2054	0,93	0,3238	1,32	0,4066	1,71	0,4564
0,16	0,0636	0,55	0,2088	0,94	0,3264	1,33	0,4082	1,72	0,4573
0,17	0,0675	0,56	0,2123	0,95	0,3289	1,34	0,4099	1,73	0,4582
0,18	0,0714	0,57	0,2157	0,96	0,3315	1,35	0,4115	1,74	0,4591
0,19	0,0753	0,58	0,2190	0,97	0,3340	1,36	0,4131	1,75	0,4599
0,20	0,0793	0,59	0,2224	0,98	0,3365	1,37	0,4147	1,76	0,4608
0,21	0,0832	0,60	0,2257	0,99	0,3389	1,38	0,4162	1,77	0,4616
0,22	0,0871	0,61	0,2291	1,00	0,3413	1,39	0,4177	1,78	0,4625
0,23	0,0910	0,62	0,2324	1,01	0,3438	1,40	0,4192	1,79	0,4633
0,24	0,0948	0,63	0,2357	1,02	0,3461	1,41	0,4207	1,80	0,4641
0,25	0,0987	0,64	0,2389	1,03	0,3485	1,42	0,4222	1,81	0,4649
0,26	0,1026	0,65	0,2422	1,04	0,3508	1,43	0,4236	1,82	0,4656
0,27	0,1064	0,66	0,2454	1,05	0,3531	1,44	0,4251	1,83	0,4664
0,28	0,1103	0,67	0,2486	1,06	0,3554	1,45	0,4265	1,84	0,4671
0,29	0,1141	0,68	0,2516	1,07	0,3577	1,46	0,4279	1,85	0,4678
0,30	0,1179	0,69	0,2549	1,08	0,3599	1,47	0,4292	1,86	0,4686
0,31	0,1217	0,70	0,2580	1,09	0,3621	1,48	0,4306	1,87	0,4693
0,32	0,1255	0,71	0,2611	1,10	0,3643	1,49	0,4319	1,88	0,4699
0,33	0,1293	0,72	0,2642	1,11	0,3665	1,50	0,4332	1,89	0,4706
0,34	0,1331	0,73	0,2673	1,12	0,3686	1,51	0,4345	1,90	0,4713
0,35	0,1368	0,74	0,2704	1,13	0,3708	1,52	0,4357	1,91	0,4719
0,36	0,1406	0,75	0,2734	1,14	0,3729	1,53	0,4370	1,92	0,4726
0,37	0,1443	0,76	0,2764	1,15	0,3749	1,54	0,4382	1,93	0,4732
0,38	0,1480	0,77	0,2794	1,16	0,3770	1,55	0,4394	1,94	0,4738

Окончание табл. 2

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,95	0,4744	2,16	0,4846	2,42	0,4922	2,68	0,4963	2,94	0,4984
1,96	0,4750	2,18	0,4854	2,44	0,4927	2,70	0,4965	2,96	0,4985
1,97	0,4756	2,20	0,4861	2,46	0,4931	2,72	0,4967	2,98	0,4986
1,98	0,4761	2,22	0,4868	2,48	0,4934	2,74	0,4969	3,00	0,49865
1,99	0,4767	2,24	0,4875	2,50	0,4938	2,76	0,4971	3,20	0,49931
2,00	0,4772	2,26	0,4881	2,52	0,4941	2,78	0,4973	3,40	0,49966
2,02	0,4783	2,28	0,4887	2,54	0,4945	2,80	0,4974	3,60	0,499841
2,04	0,4793	2,30	0,4893	2,56	0,4948	2,82	0,4976	3,80	0,499928
2,06	0,4803	2,32	0,4898	2,58	0,4951	2,84	0,4977	4,00	0,499968
2,08	0,4812	2,34	0,4904	2,60	0,4953	2,86	0,4979	4,50	0,499997
2,10	0,4821	2,36	0,4909	2,62	0,4956	2,88	0,4980	5,00	0,499997
2,12	0,4830	2,38	0,4913	2,64	0,4959	2,90	0,4981		
2,14	0,4838	2,40	0,4918	2,66	0,4961	2,92	0,4982		

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Абсолютная величина 22
Абсолютно сходящийся ряд 260
Аксиоматический метод 19
Алгебраическая форма комплексного числа 503
Алгебраический многочлен 54, 177
Аналитический способ задания функции 47
Аппликата 325
Аргумент 46
– комплексного числа 533
Асимптота 149
– гиперболы 97
Асимметрия 603, 617
Алгебраическое дополнение 430

Базис 412
Базисные неизвестные 442
Балансовый анализ 467
Бесконечно большая последовательность 26
– – функция 70
– малая последовательность 26
– – функция 69
Бесконечный предел 32
– промежуток 21
Билинейная форма 348
Бином Ньютона 41, 120, 343, 547
Большая полуось эллипса 95

Варианта 607
Вариационный ряд 607
Вектор 301, 406
– нулевой 407
– столбец 406
– строка 406
– n-мерный 406
Векторная функция 354
Величина случайная
– двумерная 584
– дискретная 570
– непрерывная 570, 590
Величины случайные
– зависимые 588
– коррелированные 588
– независимые 587
– некоррелированные 588
Вероятность 548
– безусловная 552
– плотность распределения 593
– попадания в интервал непрерывной случайной величины 591
Вертикальная асимптота 149
Верхний предел интегрирования 189
Верхняя грань множества 21
– сумма Дарбу 191
Вещественная ось 20
Вещественные числа 17
Внутренняя точка множества 285
Возрастающая последовательность 39
– функция 89
Вторая производная 118, 338
Второй дифференциал 122, 342
– замечательный предел 67
Входящий поток заявок 653
Выборка 606
– бесповторная 606
– повторная 606
– репрезентативная 606
Выпуклая функция 357–358
Выпуклое множество 357
Выпуклость вверх 144
– вниз 144
Выходящий поток заявок 653

Гармонический анализ 273
Геометрический смысл
– – дифференциала 105, 324
– – дифференциального уравнения первого порядка 480–481
– – определенного интеграла 212
– – производной 99
– – теорем Ферма, Ролля, Лагранжа 125–128
Геометрическое изображение функции двух переменных 287–288
Гессиан 350
Гипербола 52
Гиперболический параболоид 315
Гиперплоскость 290
Гистограмма 611
Горизонтальная асимптота 149
Градиент 335, 336
Граница области 353
Граничная точка области 285, 321
Граф состояний 637, 660, 666
График функции 49
Графический способ задания функции 49

Двойной интеграл:
– – вычисление путем замены переменной 398
– – геометрические приложения 402–405
– – геометрический смысл 391
– – основные свойства 393–394
– – сведение к повторному 394
Двуполостный гиперболоид 313
Детерминант 426
Диагональ матрицы 417
– – главная 417
– определителя 427
Диаметр области 391
Дисконтирование 231
Дисперсия:
– внутригрупповая 615
– выборочная 613
– генеральная 613
– групповая 614
– дискретной случайной величины 577
– межгрупповая 615
– непрерывной случайной величины 596

Дифференциал:
– инвариантность формы 115, 331
– высшего порядка 121, 342
– применение в приближенных вычислениях 105
– свойства 106, 332
– функции нескольких переменных 206–207
Дифференциальное исчисление 98
Дифференциальное уравнение 477
– – второго порядка 489
– – – – допускающее понижение порядка 492–493
– – – – линейное 494
– – – – – неоднородное 494, 498
– – – – – общее решение 498
– – – – – однородное 495
– – – – – с постоянными коэффициентами 502
– – – – – – однородное 502
– – – – разрешенное относительно второй производной 490
– – первого порядка 477
– – – – автономное 485
– – – – линейное 485
– – – – – общее решение 487
– – – – неполное 484–485
– – – – разрешенное относительно производной 478
– – – – с разделяющимися переменными 481
– – *n*-го порядка 507
– – – – разрешенное относительно старшей производной 508
– – – – линейное 511
– – – – – общее решение 512
– – – – – с постоянными коэффициентами 513
– – – – сведение к системе уравнений первого порядка 508–509
Дифференцирование:
– обратной функции 110
основные правила 106
– простейших элементарных функций 107–110
– сложной функции 113, 327
– таблица основных производных 117
– функции, заданной параметрически 123
Дифференцируемая функция 101, 321
Длина вектора 302, 408
– дуги плоской кривой 218–221
– частичного отрезка 187
Дробно-рациональная функция 54, 177

Евклидово пространство 282

Зависимая переменная 46
Зависимость функций 379
Задача Коши 479, 490, 510, 535
Замкнутое множество 285
Знакопеременный ряд 259

Интеграл с переменным верхним пределом 204
Интегральная кривая 479, 490, 508
– сумма 188
– – геометрический смысл 188
Интегральное исчисление 162
Интегральный признак сходимости ряда 267
Интегрирование 163
– некоторых иррациональных функций 180
– – трансцендентных функций 186
– непосредственное 167
– подстановкой 167–170, 207–208
– по частям 170–173, 209–210
– рациональных функций 173
– тригонометрических функций 185–186
Интегрируемая функция 189
Интегрируемость функции:
– необходимое условие 190
– – – и достаточное условия 194–195
Интервал 20
– сходимости степенного ряда 265
Интенсивность
– потока заявок 657
– потока обслуживания 657
Исследование функций 140
Иррациональные числа 17

Канал обслуживания 653
Каноническое уравнение:
– – гиперболы 96
– – гиперболического параболоида 315
– – двуполостного гиперболоида 313
– – конуса второго порядка 317
– – однополостного гиперболоида 312
– – параболы 97
– – эллипса 95
– – эллипсоида 311
– – эллиптического параболоида 314
Кардиоида 218, 221
Касательная 144
– плоскость 323
Квадратичная форма: 348
– – знакопеременная 349
– – отрицательно определенная 349
– – положительно определенная 349
Квадратная матрица 416
Коллинеарные векторы 302
Компланарные векторы 411
Комплексное число 503, 514, 533
Комплексно-сопряженные числа 503, 514, 533
Компоненты вектора 406
Координаты:
– точки 20, 283
– вектора 406, 413
Координатная плоскость 282
Координатное пространство 283
Коэффициенты:
– корреляции 587
– регрессии 589
– степенного ряда 262, 268
– ряда Фурье 275–277, 278
Краевая задача 506, 511
Кратный корень 502, 514, 533
Криволинейная трапеция 212
Критерий продуктивности матрицы 470
Кусочно-непрерывная функция 81

Линейная комбинация:
– – векторов 409

– – решений дифференциального уравнения 496
– – разностного уравнения 531
– функция 54
Линия:
– безразличия 638, 679
– второго порядка 95
– уровня 299
Логарифмическая производная 115
Локальный максимум 140, 346, 350
– минимум 140, 346, 350
– экстремум 140, 346

Максимальное значение функции 152, 353
Математическая модель 13, 619
Математическое ожидание
– дискретной случайной величины 575
– непрерывной случайной величины 596
Матрица: 416
– вырожденная 426, 433
– Гессе 350
– единичная 417
– квадратичной формы 349
– квадратная 416
– невырожденная 433
– нулевая 416
– обратная 426
– – вычисление 433–435
– присоединенная 433
– расширенная 438
– симметрическая 349, 420
– системы уравнений 437
– транспонированная 420
– Якоби 355
Метод вариации постоянных 499–500
– Гаусса 443–447
– наименьших квадратов 367
– неопределенных коэффициентов 178
– – множителей Лагранжа 385, 643
Методика VAR 684
Минимальное значение функции 353
Минор 430
– главный 349
Мнимая единица 503
Многочлен 54
Множество 15
– бюджетное 640–641
– действительных чисел 17
– допустимое 676
– достижимое 681
– значений функции 46
– замкнутое 285
– открытое 285
– пустое 16
Мода 601, 607
Модель 620
– динамическая Кейнса 519
– Леонтьева 466–469
– – динамическая 541–542
– обмена (международной торговли) 474
– потребительского выбора 636
– Стоуна 646
Модуль вектора 302, 408
– числа 22
Момент корреляционный 587
– начальный теоретический 583
– центральный теоретический 584
– эмпирический 616

Монотонная последовательность 39
– функция 89

Наклонная асимптота 149
Направленный отрезок 301
Направляющий вектор прямой 308
Натуральный логарифм 68
Начальные условия 479, 490, 511, 530
Невозрастающая последовательность 39
– функция 89
Независимая переменная 46
Нелинейное программирование 677
Неограниченная последовательность 26
Неопределенный интеграл (см. также *Интегрирование*) 163
– – свойства 164
– – таблица 165
– – "неберущиеся" интегралы 166
Непрерывность вещественных чисел 19
– функции 74
– – в интервале 78
– – в точке 75
– – слева, справа 74
– – нескольких переменных 299
– – на отрезке 78
– обратной функции 91
Неравенство 18
– Коши–Буняковского 409
Несобственный интеграл 237
– – второго рода 242
– – первого рода 238
– – геометрический смысл 238–239
– – сходящийся 238, 242
– – расходящийся 238, 242
Неубывающая последовательность 39
– функция 89
Неявная функция 371
– вектор-функция 373, 375
Нижняя грань множества 21
Нормаль к поверхности 325
Нормальный вектор плоскости 306

Область значений функции 46, 287
– определения функции 46, 49, 287
Обратная функция 89, 377–378
Обратные тригонометрические функции 112
Общее уравнение прямой 92
– решение дифференциального уравнения 479, 490, 510
– – разностного уравнения 531
Общий член последовательности 25
Объединение множеств 16
Объем:
– выборки 606
– криволинейного цилиндра 402
– тела вращения 222–223
Ограниченная последовательность 26, 287
– функция 84
Ограниченное множество 21
Однополостный гиперболоид 312

Однородная функция 330
Однородное уравнение
– – дифференциальное 486, 495, 511
– – разностное 529
Окрестность точки 31, 284
Окружность, уравнение 47, 95
Операции:
– над векторами 303, 407
– – – свойства 303–304, 407–408
– над вещественными числами 18
– – – – свойства 18
– над матрицами 417–418, 420–421
– – – свойства 418–419, 423
Определенный интеграл 188
– – геометрические приложения 211–228
– – основные свойства 198–200
– – оценки 200–204
– – условия интегрируемости 196–198
– – экономические приложения 229–232
Определитель 426
– Вронского 495, 512
– Казоратти 531
– свойства 428–429
– системы линейных алгебраических уравнений 439
Оптимизация банковского портфеля 676
Ортогональная система функций 273
Ортогональные векторы 409
Ортонормированный базис 415–416
Отрезок 20
Отсутствие последействия 655
Оценка
– несмещенная 612
– смещенная 612
– состоятельная 612
– точечная 618
– эффективная 612

Парабола 97
Параболоид 313, 314
Параллельный перенос графика функции 52
Параметр 123
Параметрическое задание функции 123
Паутинная модель рынка 56
Первообразная 162
Первый замечательный предел 65
Переменная интегрирования 163, 189
Период 272
Периодическая функция 272
Периодическое продолжение функции 276
Плотность распределения вероятностей 592
Площадь:
– криволинейного сектора 216–217
– криволинейной трапеции 212
– плоской фигуры 211–213
– поверхности 404
– – вращения 225–227
Поверхность уровня 336
Погрешность 235, 236, 369
Подынтегральная функция 163, 189
Показательная функция 111
Поле направлений 481
Полигон 610
Полная группа событий 548
Полное приращение функции 321
Полуось:
– гиперболоида двуполостного 313
– – однополостного 312
– эллипса 96
– эллипсоида 311
Полярные координаты 215–216
Портфельный анализ 678
Порядок матрицы 424
Правило Лопиталя 130
Предел:
– последовательности 30, 286
– функции 58
– – левый, правый 61
– – нескольких переменных 291
– – по Гейне 60
– – по Коши 60
Предельная норма замены ресурсов 633
– полезность продуктов 637
Предельная себестоимость 155
Предельные издержки 158
Предельный доход 158
Признак сходимости:
– – Даламбера 254
– – Дирихле–Абеля 247
– – Лейбница 259
– – несобственных интегралов 245–247
– сравнения 245, 252–253
Приращение аргумента 75
– функции 75
Произведение вектора на число 303, 407
– вещественных чисел 18
– матриц 420
– матрицы на число 418
– последовательностей 26
– последовательности на число 26
Производная (см. также *Дифференцирование*) 98
– вектор-функции 355
– высшего порядка 118, 338
– левая 101
– неявной функции 374–375
– обратной функции 110
– по направлению 233–334
– правая 101
– применение в экономике 155
– сложной функции 113
– физический смысл 100
Производственная функция 290, 320, 326, 364, 626
– предельное значение 631
– среднее значение 631
– формальные свойства 630
Промежуток 21
Промежуточная переменная 114
Простейшие элементарные функции 54

Равенство векторов 302, 407
– матриц 417
Равновесная цена 56
Равномерная непрерывность 87
Радиус сходимости степенного ряда 265
Разложение вектора по базису 413
– функций в ряд Маклорена 269
– – – – Фурье 275–276, 277–278
– – по формуле Тейлора 133, 344

– – – – Маклорена 136
– рациональной функции на простые дроби 177–178
Размер матрицы 416
Размерность пространства 283
Разностные уравнения 528
Ранг:
– матрицы 425, 432
– системы векторов 412, 425
Раскрытие неопределенностей 129
Распределение:
– биномиальное 572
– Пуассона 574
– нормальное 600–601
– равномерное 598–599
Расстояние между точками 283
– от точки до плоскости 307
– – – до прямой 94, 310
Рациональная функция 54, 180
Рациональные числа 17
Регрессия линейная 589
– средняя квадратическая 589
Рекуррентная формула 173, 176
Решение дифференциального уравнения 487, 498, 512
– разностного уравнения 529
– системы линейных алгебраических уравнений 436
Ряд:
– Маклорена 269
– Фурье 272, 275
– – для функций с периодом $2l$ 281
– для четных и нечетных функций 277

Связное множество 286
Связь между бесконечно большими и бесконечно малыми последовательностями 28
– – – – – – – функциями 70
Свойства решений линейных уравнений:
– – – – дифференциальных 494, 498, 512
– – – – разностных 531
Сегмент 20
Система:
– дифференциальных уравнений первого порядка 509
– – – в матричной форме 510
– линейных алгебраических уравнений 436
– – – в матричной форме 437–438
– – – геометрический смысл 450–451
Система массового обслуживания:
– – – многоканальная 653
– – – одноканальная 653
– – – предельный режим работы 658, 661
– – – с неограниченной очередью 653, 672
– – – с ожиданием и ограниченной очередью 653, 665
– – – с отказами 653, 660
Система разностных уравнений первого порядка 537
– – – – – – в матричной форме 538
Скалярное произведение векторов 305, 408
– – – свойства 305, 408
Сложная функция 88, 299
Собственное значение матрицы 424

Собственный вектор матрицы 424
События
– независимые 554
– несовместные 547
– противоположные 551
– совместные 558
Спираль Архимеда 217
Способы задания функций 46
Сравнение бесконечно малых функций 71, 294
Среднее квадратическое отклонение 581
– – – выборочное 614
– – – генеральное 613
Средняя 613
– выборочная 613
– генеральная 613
– общая 613
Стационарность потока 655
Степенная функция 54
Степенной ряд 262
Строго монотонная последовательность 39
Строгое неравенство 19
Сумма векторов 303, 407
– матриц 417
– последовательностей 26
– ряда 250
Суперпозиция функций 54
Сходимость по вероятности 568
Сходящаяся последовательность 30
Сходящийся ряд 250

Табличный интеграл 166
Теорема:
– Больцано–Коши вторая 83
– – – первая 82
– Вейерштрасса вторая 85
– – первая 84
– Дирихле–Абеля 261
– Кантора 88
– Коши 128
– – о существовании и единственности решения дифференциального уравнения 479, 490, 509
– Лагранжа 127
– Лопиталя 129
– о вложенных отрезках 45
– о замене переменной в двойном интеграле 398
– – – – в определенном интеграле 207–208
– о непрерывности обратной функции 91
– о смешанных производных функции нескольких переменных 340–341
– о среднем 202–203
– о существовании точных граней ограниченного множества 22
– о сходимости монотонной ограниченной последовательности 40
– – – несобственных интегралов 245, 247
– – об устойчивости знака непрерывной функции 81
– сложения вероятностей 550
– – – несовместных событий 550
– – – совместных событий 558
– Ролля 125
– умножения вероятностей 553
– Ферма 124
Теоремы о бесконечно малых последовательностях 28–30

– – – функциях 69
– – дифференцируемых функциях 106, 322–323
– – направлении выпуклости и точках перегиба 144–145
– – непрерывных функциях 81–86
– – первообразных 162–163
– – пределах функций 64
Точные грани множества 21
– – функции 84
Тригонометрические функции 77

Убывание функций, достаточное условие 140
Убывающая последовательность 39
Угловой коэффициент 91
Угол между векторами 305, 408
– – двумя прямыми 93, 309
– – плоскостями 307
– – прямой и плоскостью 310
Уравнение касательной 144
– линии на плоскости 91
– общее линии второго порядка 98, 305
– плоскости общее 306
– поверхности 305
– Слуцкого 64
Уравнения прямых 92, 308
Условный экстремум 343, 642

Фокус:
– гиперболы 96
– параболы 97
– эллипса 95
Формула:
– интегрирования по частям в неопределенном интеграле 170
– конечных приращений 128
– Коши 129
– Лагранжа 128
– Лейбница 120
– Маклорена 136
– Ньютона–Лейбница 206
– среднего значения 202–203
– Симпсона 235–236
– Тейлора 135, 344
– трапеций 233–235
Формулы:
– Байеса 561
– Крамера 440
Фундаментальная система решений:
– – – линейного дифференциального уравнения 496
– – – – разностного уравнения 531
– – – системы линейных алгебраических уравнений 452–453
Функция:
– Дирихле 49
– иррациональная 54
– Кобба–Дугласа 290, 320, 326
– нескольких переменных 287
– полезности 636–637, 678
– прибыли 158, 359, 361, 364
– производственная 626
– – аддитивная 629
– – мультипликативная 629
– распределения 590
– – теоретическая 609
– – эмпирическая 609
– спроса 645
– четная, нечетная 151
– явная 371

Характеристическое уравнение
– – линейного дифференциального уравнения с постоянными коэффициентами 502, 513
– – линейного разностного уравнения с постоянными коэффициентами 532
– – для собственных чисел матрицы 424, 455

Целая часть числа 48
Целевая функция 676
Ценовая дискриминация 361
Циклоида 214, 221

Частичная сумма ряда 250
Частная производная 318–319
Частное приращение функции 318
– решение дифференциального уравнения 479, 490, 504
– – разностного уравнения 531
Частные производные высших порядков 338
Частота 568
– относительная 568, 607
Число *e* 41, 68
Числовая ось (прямая) 21
– последовательность 25
Числовой ряд 249
Член последовательности 25

Эквивалентные бесконечно малые 71
– системы уравнений 437
Экономическая динамика 515
Эксцентриситет гиперболы 96
– эллипса 96
Эксцесс 603, 617

Учебник

Максим Семенович КРАСС

МАТЕМАТИКА ДЛЯ ЭКОНОМИЧЕСКИХ СПЕЦИАЛЬНОСТЕЙ

Главный редактор *Ю.В. Луизо*
Зав. редакцией *Г.Г. Кобякова*
Художник *Н.В. Пьяных*
Компьютерная подготовка
оригинал-макета *Д.С. Теляковский*
Технический редактор *Л.А. Зотова*
Корректоры *Е.Ю. Иванова, Л.И. Трифонова*

Гигиеническое заключение № 77.99.2.953.П.16308.12.00 от 01.12.2000 г.

Подписано в печать 20.11.2002. Формат 60 × 90 1/16. Бумага офсетная.
Гарнитура Ньютон. Печать офсетная. Усл. печ. л. 44,0.
Тираж 5000 экз. Заказ № 7021 . Изд. № 380.

Издательство "Дело"
119571, Москва, пр-т Вернадского, 82

Коммерческий отдел тел.: 433-2510, 433-2502
E-mail: delo@ane.ru
Internet: http://www.delo.ane.ru

ФГУИПП «Янтарный сказ»
236000, Калининград, ул. К. Маркса, 18

ISBN 5-7749-0264-1

9 785774 902644 >