КОНСПЕКТ ЛЕКЦИЙ

Анализ-2

Лектор В. И. Богачев (НМУ, весна 2022)

Оглавление

Глава 1

Дифференцирование в многомерных пространствах

Эта глава посвящена дифференцированию в многомерных (в том числе и бесконечномерных) пространствах. В ней предполагается знакомство с производной функций на прямой и с понятиями линейного пространства и линейного оператора, в том числе с координатным пространством $\mathbb{R}^n$ и линейными операторами в нем.

1.1. Линейные пространства и линейные операторы

Далее рассматриваются вещественные линейные пространства. Одно из простейших и важнейших — $\mathbb{R}^n$ векторов $x = (x_1, \ldots, x_n)$. На $\mathbb{R}^n$ есть стандартное скалярное произведение

$$(x, y) = \sum_{i=1}^{n} x_i y_i,$$

которое порождает стандартную норму

$$|x| = \sqrt{(x, x)},$$

а норма порождает расстояние

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}.$$

На абстрактном линейном пространстве X нормой $\|\cdot\|$ называется неотрицательная функция $x \mapsto \|x\|$, равная нулю только в нуле и удовлетворяющая таким условиям:

$$1)\ \|\lambda x\| = |\lambda|\,\|x\|, \quad 2)\ \|x + y\| \leqslant \|x\| + \|y\|.$$

Последнее называется неравенством треугольника. Оно позволяет задать метрику
$$d(x,y)=\|x-y\|.$$
Некоторые (но отнюдь не все) нормы порождаются скалярными произведениями $(\cdot,\cdot)$ по формуле
$$\|x\|=\sqrt{(x,x)}.$$
Критерием того, что норма получена из какого-то скалярного произведения является тождество параллелограмма
$$\|x+y\|^2+\|x-y\|^2=2\|x\|^2+2\|y\|^2.$$
Например, на плоскости норма
$$\|x\|_1=|x_1|+|x_2|$$
не удовлетворяет этому тождеству, поэтому эта норма не задается скалярным произведением. Такова же и норма
$$\|x\|_\infty=\max(|x_1|,|x_2|).$$
На ненулевом линейном пространстве бесконечно много различных норм. Говорят, что норма p оценивается через норму q, если $p(x)\leqslant Cq(x)$ для некоторого числа $C>0$, не зависящего от x. Если также и q оценивается через p, то нормы p и q называются эквивалентными. На всяком бесконечномерном линейном пространстве найдутся неэквивалентные нормы. Иное положение в конечномерных пространствах.

1.1.1. Теорема. *На* $\mathbb{R}^n$ *все нормы эквивалентны.*

Доказательство. Проверим, что всякая норма p эквивалентна стандартной норме. Для конкретных норм это легко увидеть явно: например, для нормы $\|x\|_1=|x_1|+\cdots+|x_n|$ имеют место неравенства
$$\|x\|_1^2\leqslant n\sum_{i=1}^n |x_i|^2\leqslant n\|x\|_1^2.$$
Правда, в одну сторону так можно оценить и неизвестную норму p: записав
$$x=\sum_{i=1}^n x_ie_i$$

с помощью стандартного базиса из векторов e_i (имеющих 1 на месте i и 0 на остальных местах), ввиду неравенства треугольника и неравенства Коши получаем оценку

$$p(x) = p(x_1 e_1 + \cdots + x_n e_n) \leqslant |x_1| p(e_1) + \cdots + |x_n| p(e_n) \leqslant \Big(\sum_{i=1}^n p(e_i)^2\Big)^{1/2} |x|.$$

Откуда взять оценку в другую сторону? Заметим, что установленная оценка означает, что функция p с точки зрения обычной метрики на $\mathbb{R}^n$ удовлетворяет условию Липшица с постоянной $L = \big(\sum_{i=1}^n p(e_i)^2\big)^{1/2}$:

$$|p(x) - p(y)| \leqslant L|x - y|.$$

В самом деле,

$$|p(x) - p(y)| \leqslant p(x - y),$$

ибо по неравенству треугольника

$$p(x) - p(y) \leqslant p(x - y), \quad p(y) - p(x) \leqslant p(x - y).$$

Для применения неравенства треугольника надо воспользоваться равенствами $x - y + y = x$ и $p(x - y) = p(y - x)$. Липшицева функция непрерывна, поэтому функция p имеет минимум m на компактной единичной сфере по обычной норме. Этот минимум положителен, так как $p > 0$ вне нуля. Итак, $p(x) \leqslant m|x|$ при $|x| = 1$, но из-за однородности обеих функций это неравенство остается в силу для всех x. $\square$

Обычно используется стандартная норма на $\mathbb{R}^n$, хотя иногда более удобными оказываются упомянутые выше нормы $\|x\|_1$ и $\|x\|_\infty$. Но вот введение «стандартной» нормы на пространстве $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ линейных операторов из $\mathbb{R}^n$ в $\mathbb{R}^k$ уже не такое банальное дело. Это пространство тоже конечномерно (размерности nk), поэтому по доказанной теореме все нормы на нем эквивалентны. Какую же конкретную взять? Один разумный кандидат — норма Гильберта – Шмидта, которая оператор A трактует как матрицу $(a_{ij})_{i\leqslant k, j\leqslant n}$ и сопоставляет ей число

$$\|A\|_2^2 = \sum_{i,j} |a_{ij}|^2.$$

Иначе говоря, это квадрат нормы вектора из $\mathbb{R}^{nk}$, полученного вытягиванием матрицы в один столбец. Можно записать это в несколько более инвариантном виде:

$$\|A\|_2 = \mathrm{trace}(A^*A),$$

где A^* — сопряженный оператор, т. е.

$$(Ax, y) = (x, A^*y).$$

Матрица A^* получается транспонированием матрицы A. Эта норма получается из скалярного произведения на пространстве операторов, заданного формулой

$$(A, B)_2 = \mathrm{trace}(B^*A).$$

В стандартном базисе $\{e_i\}$ в $\mathbb{R}^n$ оно имеет вид

$$(A, B)_2 = \sum_{i=1}^{n}(Ae_i, Be_i).$$

Однако у нормы Гильберта – Шмидта есть серьезный конкурент — операторная норма.

1.1.2. Определение. *Пусть A — линейный оператор из нормированного пространства X в нормированное пространство Y. Операторной нормой A называется величина*

$$\|A\| = \sup_{\|x\|\leqslant 1}\|Ax\|,$$

если она конечна.

Почему нужна оговорка, что норма конечна? Оказывается, на всяком бесконечномерном нормированном пространстве есть линейная функция с бесконечной нормой. В самом деле, по условию найдется линейно независимая последовательность векторов v_n единичной длины. Как известно из линейной алгебры (кому неизвестно, следует сделать известным), последовательность $\{v_n\}$ можно дополнить до алгебраического базиса $\{v_\alpha\}$ всего пространства (т. е. линейно независимого семейства, конечные линейные комбинации которого дают все векторы). Теперь положим $l(v_n) = n$ при всех $n \in \mathbb{N}$ и $l(v_\alpha) = 0$ для остальных векторов базиса, что позволяет по линейности продолжить l на все пространство. Ясно, что $\|l\| = \infty$.

Разумеется, на $\mathbb{R}^n$ такого не бывает, ибо всякая линейная функция непрерывна (будучи линейной комбинацией координатных функций) и потому ограничена на компактном единичном шаре. Значит, на $\mathbb{R}^n$ всякий линейный оператор имеет конечную норму.

Норма оператора есть наименьшее число C, с которым он удовлетворяет оценке

$$\|Ax\| \leqslant C\|x\|,$$

так как для $C = \|A\|$ это верно при $\|x\| = 1$ по определению, а тогда ввиду однородности верно при всех x.

Заметим, что множество $\mathcal{L}(X,Y)$ линейных операторов с конечной нормой само является нормированным пространством с операторной нормой. В самом деле:

$$\|\lambda A\| = |\lambda|\,\|A\|$$

для всякого скаляра λ и

$$\|A+B\| \leqslant \|A\| + \|B\|,$$

ибо

$$\|(A+B)x\| \leqslant \|Ax\| + \|Bx\| \leqslant \|A\| + \|B\| \quad \text{при } \|x\| \leqslant 1.$$

Вот еще одно полезное тривиальное наблюдение: оператор $A \in \mathcal{L}(X,Y)$ является липшицевым отображением с постоянной $\|A\|$:

$$\|Ax - Ay\| = \|A(x-y)\| \leqslant \|A\|\,\|x-y\|.$$

В частности, оператор с конечной нормой непрерывен. С другой стороны, верно и обратное: всякий непрерывный линейный оператор A имеет конечную норму. Действительно, из непрерывности хотя бы в нуле следует, что найдется такое $\delta > 0$, что $\|Ax\| \leqslant 1$ при $\|x\| \leqslant \delta$. Значит, $\|Ax\| \leqslant \delta^{-1}$ при $\|x\| \leqslant 1$, т. е. $\|A\| \leqslant \delta^{-1}$.

Таким образом, $\mathcal{L}(X,Y)$ есть множество всех непрерывных линейных операторов из X в Y. Эти операторы называют также ограниченными, но следует иметь в виду, что такая терминология отличается от принятой для общих отображений, когда ограниченными называют отображения, переводящие все пространство в ограниченное множество. Здесь же ограниченным оказывается образ шара, но отнюдь не всего пространства X (образ всего пространства при линейном отображение может быть ограниченным лишь для тождественно нулевого оператора).

Выше мы видели, что норма Гильберта–Шмидта явно вычисляется по матрице оператора. Однако операторная норма редко может быть найдена явно даже для оператора на $\mathbb{R}^n$ со стандартной нормой. Посмотрим, как в принципе ее можно найти. Очевидно равенство

$$\|A\|^2 = \sup_{|x|=1} |Ax|^2 = \sup_{|x|=1} (Ax, Ax) = (A^*Ax, x).$$

Оператор A^*A в отличие от исходного оператора симметричен. Как известно из линейной алгебры (кому неизвестно, надо сделать известным), существует ортонормированный базис $\varphi_1, \ldots, \varphi_n$ в $\mathbb{R}^n$ из собственных векторов A^*A:

$$A^*A\varphi_i = k_i\varphi_i.$$

В этом базисе задача сводится к нахождению максимума функции

$$\sum_{i=1}^{n} k_i x_i^2$$

на единичной сфере. Заметим, что $k_i \geqslant 0$, ибо

$$k_i = (A^* A\varphi_i, \varphi_i) = |A\varphi_i|^2 \geqslant 0.$$

Ясно, что максимальное значение есть наибольшее из чисел k_i, т. е. собственных чисел оператора A^*A. Такое число безусловно есть, но вот явно его найти обычно нельзя. Итак, $\|A\| = \max_i \sqrt{k_i}$.

Квадрат нормы Гильберта – Шмидта оператора A равен сумме чисел k_i. В большой размерности сумма может быть много больше максимального из чисел (скажем, операторная норма единичного оператора равна 1, а норма Гильберта – Шмидта равна $\sqrt{n}$). Эта одна из причин, почему в приложениях операторная норма нередко оказывается предпочтительнее. Конечно, еще одно ее преимущество в том, что для нее не нужно скалярное произведение.

Далее при обсуждении дифференцируемости будет видно, зачем нужны непрерывные линейные операторы и функционалы.

Хотя основной целью наших обсуждений являются дифференцируемые отображения в $\mathbb{R}^n$, полезно посмотреть на них с несколько более общей точки зрения, для чего и привлекаются нормированные пространства. Приведем характерные примеры нормированных пространств.

1.1.3. Пример. (i) Пространство l^2 состоит из бесконечных последовательностей $x = (x_n)$ вещественных чисел со сходящимся рядом $\sum_n x_n^2$. На нем задано скалярное произведение

$$(x, y) = \sum_n x_n y_n.$$

Отметим, что ряд сходится абсолютно, ибо $|x_n y_n| \leqslant x_n^2 + y_n^2$. Это скалярное произведение порождает норму

$$\|x\| = \Bigl(\sum_n x_n^2\Bigr)^{1/2}$$

и соответствующее расстояние.

Все обычные $\mathbb{R}^n$ изометрично вложены в l^2 как линейные подпространства в l^2 из векторов вида $(x_1, \ldots, x_n, 0, 0, \ldots)$. Объединение l_0^2 таких подпространств само является линейным подпространством в l^2, всюду плотным в l^2. Пространство l^2 универсально для пространств со скалярным произведением в следующем смысле: всякое пространство E со скалярным произведением, содержащее всюду плотное счетное множество, можно с помощью линейной изометрии отождествить с некоторым линейным подпространством в l^2.

(ii) Пространство $C[a,b]$ непрерывных функций на $[a,b]$ можно наделить естественной нормой

$$\|f\| = \sup_{t\in[a,b]} |f(t)|.$$

Конечно, вместо $[a,b]$ можно взять иной компакт. В этом пространстве всюду плотно счетное множество многочленов с рациональными коэффициентами, а также множество кусочно линейных функций, у которых рациональны коэффициенты наклона на участках линейности, а также концы промежутков, где меняются эти коэффициенты. Пространство $C[0,1]$ универсально в следующем смысле: всякое нормированное пространство, обладающее счетным всюду плотным множеством, линейной изометрично некоторому линейному подпространству в $C[0,1]$.

(iii) На пространстве $C[a,b]$ есть и другие разумные нормы, например:

$$\|f\|_1 = \int_a^b |f(t)|\,dt,$$

$$\|f\|_2 = \left(\int_a^b |f(t)|^2\,dt\right)^{1/2},$$

причем последняя получена из скалярного произведения

$$(f,g)_2 = \int_a^b f(t)g(t)\,dt.$$

Эти две нормы не эквивалентны между собой, а также не эквиваленты sup-норме. Однако односторонние оценки очевидны:

$$\|f\|_1 \leqslant (b-a)^{1/2}\|f\|_2 \leqslant (b-a)\|f\|.$$

Первое есть следствие неравенства Коши – Буняковского $|(f,1)_2| \leqslant \|f\|_2\|1\|_2$, которое для общего скалярного произведения имеет вид

$$(u,v)^2 \leqslant (u,u)(v,v).$$

Упомянутая естественность sup-нормы заключается в том, что она превращает $C[a,b]$ в полное метрическое пространство (где всякая фундаментальная последовательность имеет предел).

1.1.4. Определение. *Полное нормированное пространство называется банаховым. Полное евклидово пространство называется гильбертовым.*

Названия связаны с именами выдающихся математиков Стефана Банаха и Давида Гильберта.

Например, l^2 — гильбертово пространство (докажите!), $C[a,b]$ с sup-нормой банахово (докажите!), но $C[0,1]$ с нормами $\|f\|_1$ и $\|f\|_2$ не полны. Например, в них не имеет предела фундаментальная последовательность функций f_n такого вида: $f_n(t)=0$ при $t\leqslant 1/2-1/n$, $f_n(t)=1$ при $t\geqslant 1/2$, $f_n(t)=n(t-1/2+1/n)$.

Подпространство l^2_0 конечных последовательностей в l^2 тоже не является полным: векторы $v_n=(1,\ldots,2^{-n},0,0,\ldots)$ не имеют в нем предела.

В качестве упражнения полезно доказать, что если пространства X и Y банаховы, то пространство непрерывных линейных операторов $\mathcal{L}(X,Y)$ с операторной нормой тоже банахово.

Поскольку замкнутый шар в $\mathbb{R}^n$ со стандартной нормой компактен, то в силу доказанной выше эквивалентности норм на $\mathbb{R}^n$ он компактен по любой норме. Поэтому замкнутые шары в конечномерных нормированных пространствах компактны.

Это характеризует конечномерные пространства: во всяком бесконечномерном нормированном пространстве X можно найти последовательность векторов x_n с $\|x_\|=1$, для которых $\|x_n-x_k\|\leqslant 1$ при $n\neq k$. Из такой последовательности нельзя выбрать сходящуюся. Векторы x_n строятся по индукции. Пусть $x_1,\ldots,x_n$ уже построены. Есть вектор $v\in X$ с $\|v\|=1$, не лежащий в их линейной оболочке L. Функция $f(x)=\|v-x\|$ на L непрерывна (даже липшицева). Мы уже знаем, что замкнутый шар U радиуса 2 в L компактен. Поэтому на нем функция f достигает минимума в некоторой точке w. Поскольку $\|v\|=1$, то $\|v-w\|\leqslant 1$. На самом деле точка w дает минимум и на всем L, ибо $\|u-v\|>1$ при $\|u\|>2$. Положим $x_{n+1}=(v-w)/\|v-w\|$. Тогда $\|x_{n+1}\|=1$, причем $\|x_{n+1}-x\|\geqslant 1$ для всех $x\in L$. В самом деле, имеем

$$\Bigl\|\frac{v-w}{\|v-w\|}-x\Bigr\|=\frac{1}{\|v-w\|}\Bigl\|v-w-\|v-w\|x\Bigr\|\geqslant\frac{\|v-w\|}{\|v-w\|}=1,$$

так как $\bigl\|v-w-\|v-w\|x\bigr\|\geqslant\|v-w\|$ из-за включения $w+\|v-w\|x\in L$.

1.2. Теорема Хана – Банаха

Поскольку на $\mathbb{R}^n$ линейные функции относятся к наиболее явно задаваемым, то в случае абстрактного бесконечномерного нормированного пространства возникает вопрос о ненулевых непрерывных линейных функциях на нем. Сразу следует отметить, что непрерывность линейной функции, заданной с помощью алгебраического базиса, не удастся проверить. На самом деле и сам способ задания через базис нехорош тем, что в бесконечномерных пространствах базисы не задаются явно (их существование выводится из аксиомы выбора). Однако есть иной способ построения непрерывных линейных функционалов.

1.2.1. Теорема. (Теорема Хана – Банаха) *Пусть X_0 — линейное подпространство нормированного пространства X и $l_0\colon X_0 \to \mathbb{R}$ — линейная функция с конечной нормой на X_0. Тогда существует ее линейное продолжение на X с той же нормой.*

Доказательство. Можно считать, что $\|l_0\| = 1$. Основная проблема — продолжить l_0 с сохранением нормы на какое-либо большее подпространство, хотя бы полученное добавлением еще одного единичного вектора h. Этот этап нетривиален и в двумерном случае. Нам надо так ухитриться задать $c = l(h)$, что для всех $x_0 \in X_0$ и всех $t \in \mathbb{R}$ будет верно неравенство

$$l(x_0 + th) \leqslant \|x_0 + th\|,$$

тогда оно автоматически будет верно и с модулем, ибо $-l(x + th) = l(-x - th)$ и $\| - x - th\| = \|x + th\|$. Нужное неравенство запишем в виде

$$l_0(x_0) + ct \leqslant \|x_0 + th\|.$$

При $t > 0$ это есть

$$c \leqslant \|x_0/t + h\| - l_0(x_0/t),$$

а при $t = -s < 0$

$$c \geqslant -\|x_0/s - h\| + l_0(x_0/s).$$

Удивительным образом оказывается, что всегда

$$-\|x_0/s - h\| + l_0(x_0/s) \leqslant \|x_0/t + h\| - l_0(x_0/t),$$

причем между этими числами есть число, не зависящее от $x_0 \in X_0$, $t, s \geqslant 0$. Для этого положим

$$m = \sup\{-\|u - h\| + l_0(u)\colon u \in X_0\},$$
$$M = \inf\{\|v + h\| - l_0(v)\colon v \in X_0\}.$$

Покажем, что это конечные величины, причем $m \leqslant M$. Тогда в качестве c можно взять любое число из $[m, M]$. Так как $l_0(u) \leqslant \|u\|$ и $-\|u-h\| \leqslant -\|u\| + \|h\|$ (так как $\|u\| = \|u-h+h\| \leqslant \|u+h\| + \|h\|$), то $m \leqslant \|h\|$. Далее,

$$-\|u-h\| + l_0(u) \leqslant \|v+h\| - l_0(v),$$

ибо это можно записать как

$$l_0(v) + l_0(u) \leqslant \|v+h\| + \|u-h\|,$$

что верно, ибо левая часть есть $l_0(v+h)+l_0(u-h) \leqslant \|v+h\|+\|u-h\|$. Итак, всегда можно продолжить l_0 на дополнительный вектор с сохранением нормы. По индукции можно продолжить на подпространство, порожденное счетным набором векторов. В общем случае нужно применить лемму Цорна. Она говорит следующее: если в частично упорядоченном множестве $(\mathcal{M}, \leqslant)$ для всякого набора $\mathcal{P}$ попарно сравнимых элементов есть элемент $m \in \mathcal{M}$ с $p \leqslant m$ при всех $p \in \mathcal{P}$, то в $\mathcal{M}$ есть элемент μ, который максимален в таком смысле: нет большего. В нашей ситуации в качестве $\mathcal{M}$ берется множество всех пар (L, l), где L — линейное подпространство в X, $X_0 \subset L$, l — линейная функция на L с нормой $\|l_0\|$, продолжающая l_0. Частичный порядок простой: $(L, l) \leqslant (L', l')$, если $L \subset L'$ и l' продолжает l. Если набор состоит из попарно сравнимых элементов (L_t, l_t), то $L = \bigcap_t L_t$ есть линейное пространство, на котором корректно задано общее продолжение: $l(x) = l_t(x)$, если $x \in L_t$. Максимальный элемент из леммы Цорна дает искомую пару (X, l), ибо в противном случае по доказанному мы бы еще продолжили вопреки максимальности. $\square$

Теорема Хана – Банаха дает весьма неочевидное следствие уже при продолжении с одномерного подпространства X_0, порожденного вектором v единичной нормы. Можно положить $l_0(tv) = t$, что даст линейную функцию на X_0 с $\|l_0\| = 1$. Продолжение по теореме Хана – Банаха дает линейную функцию на X с $\|l\| = 1$ и $l(v) = 1$. Более общим образом, можно продолжать линейные функции с конечномерных подпространств с сохранением нормы.

1.3. Производная

Для функций f на прямой есть два основных равносильных определения производной в точке x_0:

это конечный предел

$$f'(x_0) = \lim_{h \to 0, h \neq 0} \frac{f(x_0+h) - f(x_0)}{h},$$

либо такое число $\lambda(x_0)$, что

$$f(x_0+h)-f(x_0)=\lambda(x_0)h+r(h), \quad r(h)=o(h),$$

т. е. $r(h)/h \to 0$ при $h \to 0$, $h \neq 0$. Во втором случае полагают $f'(x_0):=\lambda(x_0)$.

Здесь мы обсудим многомерные (и бесконечномерные) версии этих двух определений. Пусть сначала f — вещественная функция на $\mathbb{R}^n$. На первый взгляд, первое определение не кажется многообещающим, если его пытаться делить на вектор. Не поможет делить на его норму: ведь не так на прямой делается. Поэтому обратимся сначала ко второй опции. Здесь напрашивается такая аналогия:

производной функции f точке x_0 будет называть такую линейную функцию $\lambda(x_0)$, что

$$f(x_0+h)-f(x_0)=\lambda(x_0)(h)+r(h), \quad r(h)=o(h),$$

т. е. $r(h)/|h| \to 0$ при $h \to 0$, $h \neq 0$. Эту линейную функцию в случае $\mathbb{R}^1$ записываем в виде умножения на число, вот это число и есть классическая производная (коэффициент линейной функции).

На $\mathbb{R}^n$ при $n>1$ линейная функция l не задается одним числом. Ее можно задать с помощью вектора v в виде

$$l(h)=(v,h)=\sum_{i=1}^{n} v_i h_i.$$

Представляющий производную вектор называют градиентом функции f в точке x_0 и обозначают символами

$$\operatorname{gard} f(x_0) \text{ или } \nabla f(x_0)$$

т. е.

$$f(x_0+h)-f(x_0)=(\nabla f(x_0),h)+0(h).$$

После этого нетрудно догадаться и до следующего общего определения.

1.3.1. Определение. *Пусть X и Y — нормированные пространства. Отображение $F\colon X \to Y$ называется дифференцируемым в точке x_0 по Фреше, если существует такой непрерывный линейный оператор $A\colon X \to Y$, что*

$$F(x_0+h)-F(x)=Ah+r(h), \quad r(h)=o(h),$$

т. е. $\|r(h)\|/\|h\| \to 0$ при $h \to 0$, $h \neq 0$. При этом полагаем

$$F'(x_0):=DF(x_0)=A.$$

Как легко заметить, для придания смысла такому определению достаточно, чтобы отображение F было определено в некоторой окрестности точки x_0, а не на всем пространстве.

Из определения следует непрерывность дифференцируемого отображения: если $h \to 0$, то $F(x_0 + h) \to F(x_0)$, ибо $r(h) \to 0$ и $Ah \to 0$ из-за непрерывности.

Теперь возобновим попытки что-то сделать в духе первого определения производной на прямой. Для упрощения будем иметь дело с вещественной функцией f на $\mathbb{R}^n$. Пусть $e_1, \ldots, e_n$ — стандартный базис в $\mathbb{R}^n$.

1.3.2. Определение. *Частной производной функции f в точке x_0 по переменной x_i назовем конечный предел*

$$\partial_{x_i} f(x_0) := \frac{\partial f}{\partial x_i}(x_0) = \lim_{t\to 0, t\neq 0} \frac{f(x_0 + te_i) - f(x_0)}{t}$$

в случае его существования.

Таким образом, при вычислении частной производной по x_i берется приращение только вдоль e_i. Понятно, что дифференцируемость влечет существование частных производных, поскольку для $h = te_i$ получаем

$$f(x_0 + te_i) - f(x_0) = f'(x_0)(te_i) + o(t) = tf'(x_0)(e_i) + o(t),$$

а также

$$\partial_{x_i} f(x_0) = f'(x_0)(e_i).$$

Ясно, что дифференцируемость не следует из существования частных производных: взяв $f(x) = 0$ на координатных прямых и $f(x) = 1$ вне их, получаем $\partial_{x_i} f(0) = 0$, но дифференцируемости нет просто из-за очевидной разрывности f в нуле.

Этот простой пример подсказывает более интересный вопрос: пусть существуют не только частные производные по направлениям базисных векторов, но и частные производные вдоль всех векторов:

$$\partial_h f(x_0) = \lim_{t\to 0, t\neq 0} \frac{f(x_0 + te_i) - f(x_0)}{t};$$

равносильно ли это более сильное условие дифференцируемости? Ответ опять отрицательный: возьмем функцию f на плоскости, равную нулю всюду, кроме точек на параболе $y = x^2$ при $x > 0$. Эта функция тоже разрывна в нуле, значит, не дифференцируема в нуле, но $\partial_h f(0) = 0$ при всех h, ибо при фиксированном h при достаточно малых t имеем $f(th) = 0$, поскольку th не лежит на указанной кривой.

Рассмотренное в этом примере явление приводит к такому общему определению.

1.3.3. Определение. *Отображение $F\colon X \to Y$ между нормированными пространствами называется дифференцируемым в точке x_0 по Гато, если найдется такой линейный оператор $A\colon X \to Y$, что для всякого вектора $h \in X$ верно соотношение*

$$F(x_0 + th) - F(x_0) = tAh + o(t).$$

Оператор A называется производной Гато в точке x_0 и обозначается символами $F'(x_0)$ или $DF(x_0)$.

В предыдущем примере производная Гато равна нулю, но она не является производной Фреше и даже не обеспечивает непрерывность в точке.

Для сокращения терминологии под дифференцируемостью понимается именно дифференцируемость по Фреше, а прочие виды дополнительно оговариваются.

Фреше (Fréchet) и Гато (Gateaux) — французские математики XX века, причем правильное написание фамилии последнего отличается от французского слова gâteaux (пирожное).

1.4. Свойства дифференцируемых отображений

С помощью производной Гато можно получить такое обобщение теоремы о среднем.

1.4.1. Теорема. *Пусть отображение F из открытого шара U в нормированном пространстве X в нормированное пространство Y дифференцируемо по Гато во всех точках, причем $\|F'(x)\| \leqslant M$. Тогда*

$$\|F(x) - F(y)\| \leqslant M\|x - y\| \quad \forall\, x, y \in U.$$

Доказательство. Пусть сначала $Y = \mathbb{R}$. Зафиксируем векторы $x, y \in U$. При всех $t \in [0,1]$ имеем $ty + (1-t)x \in U$ в силу неравенства треугольника. Функция

$$g(t) = F(ty + (1-t)x)$$

дифференцируема на $[0,1]$. В самом деле, пусть $z = ty + (1-t)x$. Тогда

$$\begin{aligned} g(t+r) - g(t) &= F(ty + ry + (1-t)x - rx) - F(ty + (1-t)x) \\ &= F(z + r(y-x)) - F(z) = rF'(z)(y-x) + o(r), \end{aligned}$$

т. е. $g'(t) = F'(z)(y-x)$, значит, $|g'(t)| \leqslant \|F'(z)\|\,\|y-x\| \leqslant M\|y-x\|$. В силу одномерной теоремы о среднем $|g(1) - g(0)| \leqslant M\|y - x\|$, что и требовалось доказать.

Перейдем к общему случаю. Опять зафиксируем $x, y \in X$. По следствию теоремы Хана–Банаха найдется такой непрерывный линейный функционал l на Y, что $\|l\| = 1$ и

$$l(F(y) - F(x)) = \|F(x) - F(y)\|.$$

Рассмотрим вещественную функцию $f(u) = l(F(u))$. Эта функция дифференцируема в U по Гато:

$$\begin{aligned} f(u+th) - f(u) &= l(F(u+th) - F(u)) = l(tF'(u)(h) + r(t)) \\ &= tl(F'(u)(h)) + l(r(t)), \end{aligned}$$

где $r(t) = o(t)$. Тогда $l(r(t)) = o(t)$, ибо $|l(r(t))| \leqslant \|l\|\,\|r(t)\|$. При этом $f'(u)(h) = l(F'(u)(h))$, поэтому при $\|h\| \leqslant 1$ имеем $\|F'(u)(h)\| \leqslant \|F'(u)\| \leqslant M$ и $|l(F'(u)(h))| \leqslant \|F'(u)(h)\| \leqslant M$, откуда $\|f'(u)\| \leqslant M$. По доказанному

$$|f(y) - f(x)| \leqslant M\|y - x\|,$$

а в силу выбора l имеем $|f(y) - f(x)| = \|F(y) - F(x)\|$. □

В доказательстве мы фактически дважды вручную проверили дифференцируемость композиции. Рассмотрим этот важный вопрос отдельно.

1.4.2. Теорема. *Пусть X, Y, Z — нормированные пространства, $F\colon X \to Y$ дифференцируемо по Фреше в x_0, $G\colon Y \to Z$ дифференцируемо по Фреше в $y_0 = F(x_0)$. Тогда $G\circ F\colon X \to Z$ дифференцируемо по Фреше в x_0 и*

$$(G\circ F)'(x_0) = G'(F(x_0)) \circ F'(x_0)).$$

Доказательство. По условию

$$F(x_0 + h) = F(x_0) + F'(x_0)h + o(h),$$

поэтому

$$\begin{aligned} &G(F(x_0 + h)) = G(F(x_0) + F'(x_0)h + o(h)) \\ &\quad = G(F(x_0)) + G'(F(x_0))(F'(x_0)h + o(h)) + r(F'(x_0)h + o(h)), \end{aligned}$$

где $r(u) = o(u)$, т. е. $r(u) = \delta(u)\|u\|$, где $\delta(u) \to 0$ при $u \to 0$. Так как

$$\|G'(F(x_0))(o(h))\| \leqslant \|G'(F(x_0))\|\,\|o(h)\|,$$

то это есть $o(h)$. Далее, по определению $o(h) = \varepsilon(h)\|h\|$, где $\varepsilon(h) \to 0$ при $h \to 0$. Значит, $\|F'(x_0)h + o(h)\| \leqslant C\|h\|$, где C — постоянная.

Поэтому

$$\|r(F'(x_0)h + o(h))\| = \|\delta(F'(x_0)h + o(h))\| \, \|F'(x_0)h + o(h)\| \\ \leqslant C\|h\|\|\delta(F'(x_0)h + o(h))\|.$$

Это есть $o(h)$, ибо $\|\delta(F'(x_0)h + o(h))\| \to 0$ при $h \to 0$ ввиду оценки $\|F'(x_0)h + o(h)\| \leqslant C\|h\|$ и равенства $\lim\limits_{u\to 0} \delta(u) = 0$. □

Однако выше мы дифференцировали композиции дифференцируемых по Гато, а не по Фреше отображений. Как дела с ними? Легко видеть, что предыдущая теорема для них неверна. Например, отображение $g\colon t \mapsto (t, t^2)$ из $\mathbb{R}^1$ в $\mathbb{R}^2$ очевидным образом дифференцируемо на прямой. Рассмотрим композицию $f(g(t))$, где функция f на плоскости взята из рассмотренного выше примера: $f(t, t^2) = 0$ при всех $t \neq 0$, $f(x, y) = 0$ в остальных точках. Мы видели, что в нуле эта функция имеет нулевую производную Гато, но разрывна. Для композиции получаем $f(g(t)) = 0$ при $t \neq 0$, $f(g(0)) = 0$, т. е. композиция разрывна в нуле, поэтому недифференцируема.

Однако в доказательстве вроде бы все было ОК с дифференцируемостью композиции (неужели нет???). В чем дело? Оказывается, там неприятности были исключены из-за специального вида композиций.

1.4.3. Предложение. *Пусть X, Y, Z — нормированные пространства, $F\colon X \to Y$ дифференцируемо по Гато в x_0, $T\colon Y \to Z$ и $S\colon Z \to X$ — непрерывные линейные операторы, $x_0 = Sz_0$. Тогда $T\circ F$ дифференцируемо по Гато в x_0 и $(T\circ F)'(x_0) = T\circ F'(x_0)$, а также $F\circ S$ дифференцируемо по Гато в z_0 и $(F\circ S)'(z_0) = F'(x_0)\circ S$.*

Доказательство. Пусть $h \in X$. Тогда

$$T(F(x_0 + th)) = T(F(x_0) + tF'(x_0)h + r(t)) \\ = T(F(x_0)) + tTF'(x_0)h + T(r(t)), \quad r(t) = o(t).$$

Так как $\|T(r(t))\| \leqslant \|T\| \, \|r(t)\|$, то $T(r(t)) = 0(t)$.

Во втором случае пусть $h \in Z$. Тогда

$$F(S(z_0 + th)) = F(Sz_0 + tSh) = F(Sz_0) + tF'(Sz_0)Sh + r(tSh),$$

где $r(tSh) = o(t)$, что означает дифференцируемость по Гато. □

Легко видеть также, что сумма дифференцируемых по Фреше или Гато отображений из X в Y дифференцируема в таком же смысле и

$$(F + G)' = F' + G'.$$

1.4.4. Пример. (i) Если F — непрерывный линейный оператор, то из определения ясно, что

$$F'(x_0) = F$$

для всех x_0.

(ii) Координатные функции дифференцируемы на $\mathbb{R}^n$ и имеют место равенство $\partial_{x_i} x_j = \delta ij$, т. е. $\nabla x_i = e_i$.

(iii) Квадрат стандартной нормы $|x|^2 = x_i 6 + \cdots + x_n^2$ дифференцируем на $\mathbb{R}^n$, поэтому норма дифференцируема вне нуля (корень дифференцируем вне нуля),

$$\nabla |x| = \frac{x}{|x|}, \ x \neq 0.$$

(iv) Норма $|x|_\infty = \max_i |x_i|$ дифференцируема не во всех точках единичной сферы. Для упрощения обозначений рассмотрим случай плоскости и точку $(1,1)$. Если предположить, что в этой точке есть производная Фреше, то частные производные равны нулю, ибо на сторонах квадрата норма постоянна. Тогда должна быть нулевой и производной вдоль диагонали, что не так. Рассмотренная норма имеет мало точек недифференцируемости, но можно построить пример нормы на плоскости с бесконечным числом точек недифференцируемости. С другой стороны, можно показать, что всякая норма на $\mathbb{R}^n$ имеет бесконечно много точек дифференцируемости.

Заметим, что дифференцируемость отображения $F = (F_1, \ldots, F_k)$ из $\mathbb{R}^n$ в $\mathbb{R}^k$ по Фреше или Гато равносильна дифференцируемости в соответствующем смысле его компонент F_i.

Полезно следующее достаточное (но не необходимое) условие дифференцируемости в терминах частных производных.

1.4.5. Теорема. *Пусть функция f на открытом множестве U в $\mathbb{R}^n$ имеет частные производные $\partial_{x_i} f$, которые в некоторой точке x_0 непрерывны. Тогда f дифференцируема в точке x_0.*

Доказательство. Пусть $c_i = \partial_{x_i} f(x_0)$. Перейдя к новой функции $f(x) - c_1 x_1 - \cdots - c_n x_n$, можно считать, что $\partial_{x_i} f(x_0) = 0$. Кроме того, можно считать, что $x_0 = 0$ и $f(x_0) = 0$. Покажем, что f имеет нулевую производную в нуле. Пусть $\varepsilon > 0$. В силу непрерывности $\partial_{x_i} f$ в нуле найдется такое $\delta > 0$, что $|\partial_{x_i} f(x)| < \varepsilon/n^2$ при $|x| < \delta$. Покажем, что $|f(x)| \leqslant \varepsilon |x|$ для таких x. Пусть $z = (x_1, \ldots, x_{n-1}, 0)$. По теореме о среднем для функций на отрезке имеем

$$|f(x) - f(z)| = |\partial_{x_n} f(x + \xi e_n) x_n| \leqslant \varepsilon |x_n|/n^2.$$

Далее сравним аналогично значения в z и $y = (x_1, \ldots, x_{n-2}, 0, 0)$, что даст оценку $|f(z) - f(y)| \leqslant \varepsilon |x_{n-1}|/n^2$. Повторяя это и учитывая, что аналогично $|f(x_1, 0, \ldots, 0)| \leqslant \varepsilon |x_1|/n^2$, получаем окончательную оценку $|f(x)| \leqslant \varepsilon(|x_1| + \cdots + |x_n|)/n \leqslant \varepsilon |x|$, откуда $f'(0) = 0$. $\square$

Теперь мы вернемся к вопросу о разностных отношениях. Можно ли описать дифференцируемость через них аналогично одномерному случаю? Конечно, если есть дифференцируемость по Гато, то

$$\frac{f(x + t_n h) - f(x)}{t_n} \to f'(x)h \quad \text{при } t_n \to 0.$$

Мы видели выше, что вроде бы обратное неверно. Однако давайте чуть модифицируем обратное.

1.4.6. Предложение. *Пусть X — нормированное пространство, $f\colon X \to \mathbb{R}$, $x_0 \in X$, l — такой непрерывный линейный функционал на X, что*

$$\frac{f(x_0 + t_n h) - f(x_0)}{t_n} \to l(h) \quad \textit{при } t_n \to 0 \textit{ для всякого } h \in X.$$

Тогда l — производная Гато в x_0. Она является производной Фреше в точности тогда, когда указанная сходимость имеет место равномерно по h из единичного шара.

Доказательство. Зафиксируем h. Тогда

$$f(x_0 + th) - f(x_0) = tl(h) + r(t).$$

Так можно записать для какой угодно функции, но нам нужно показать, что $r(t) = o(t)$. Если это неверно, то найдется последовательность $t_n \to 0$, $t_n \neq 0$, для которой $r(t_n)/t_n \not\to 0$. Однако такое противоречит предполагаемому соотношению.

Предположим теперь, что это соотношению выполнено равномерно по h с $\|h\| \leqslant 1$. Зададим r равенством $f(x_0+h)-f(x_0) = l(h)+r(h)$. Покажем, что $r(h) = o(h)$. Если это неверно, то найдется последовательность ненулевых векторов h_n с $\|h_n\| \to 0$ и $r(h_n)/\|h_n\| \not\to 0$. Взяв подпоследовательность, можно считать, что $\|r(h_n)\| \geqslant c\|h_n\|$ при некотором $c > 0$. Пусть $v_n = h_n/\|h_n\|$, $t_n = \|h_n\|$. Тогда $\|v_n\| = 1$, $h_n = t_n v_n$, $t_n \to 0$. По предположению

$$\sup_{\|v\| \leqslant 1} t_n^{-1}(f(x_0 + t_n v) - f(x_0)) - l(v) \to 0.$$

Значит, при всех достаточно больших n имеем

$$|t_n^{-1}(f(x_0 + t_n v_n) - f(x_0)) - l(v_n)| < c/2.$$

Левая часть равна $|r(t_n v_n)/t_n$. Итак, $|r(h_n)| = |r(t_n v_n)| < ct_n/2$ вопреки тому, что $\|r(h_n)\| \geqslant c\|h_n\|$.

Наконец, пусть есть дифференцируемость по Фреше в x_0. Пусть $t_n \to 0$. Если указанное соотношение не выполнено равномерно по h с $\|h\| \leqslant 1$, то найдутся последовательность векторов h_n с $\|h_n\| \leqslant 1$ и число $c > 0$, для которых $|t_n^{-1}(f(x_0 + t_n h_n) - f(x_0)) - l(h_n)| \geqslant c$. Однако определение дифференцируемости говорит, что

$$|f(x_0 + h) - f(x_0) - l(h)| \leqslant c\|h\|/2$$

для всех h с достаточно малой нормой. Значит, при достаточно больших n получаем

$$|f(x_0 + t_n h_n) - f(x_0) - l(t_n h_n)| \leqslant c\|t_n h_n\|/2,$$

откуда $|t_n^{-1}(f(x_0+t_n h_n) - f(x_0)) - l(h_n)| \leqslant c/2$ вопреки нашему предположению. $\square$

Отметим, что для отображений общих нормированных пространств есть естественные аналоги теоремы 1.4.5. Во-первых, в качестве задачи предлагается доказать, что если отображение $F\colon X \to Y$ нормированных пространств дифференцируемо по Гато в точках некоторого шара, причем отображение $x \mapsto F'(x)$ из X в пространство $\mathcal{L}(X, Y)$ с операторной нормой непрерывно в точке x_0, то $F'(x_0)$ — производная Фреше. Это не вполне то, что было в теореме 1.4.5, где не требовалась дифференцируемость Гато, однако из сформулированной задачи и цитированной теоремы следует, что достаточным условием дифференцируемости по Фреше в x_0 непрерывного отображения F из банахова пространства X является существование производных по всем направлениям $\partial_h F(x)$ и их непрерывность в x_0. В самом деле, можно считать, что $x_0 = 0$. Тогда дифференцируемы в нуле сужения F на конечномерные подпространства. Поэтому возникает линейный оператор $Ah = \lim\limits_{n\to\infty} nF(h/n)$. Допредельные отображения $h \mapsto nF(h/n)$ непрерывны ввиду непрерывности F, поэтому по следствию теоремы Бэра оператор A имеет точку непрерывности. В силу линейности он непрерывен, т. е. получаем дифференцируемость по Гато.

Рассмотрим бесконечномерный пример.

1.4.7. Пример. Пусть Ψ — функция на прямой с непрерывной производной. На пространстве $C[0, 1]$ с его $\sup$-нормой функция

$$F(f) = \int_0^1 \Psi(f(s))\, ds$$

дифференцируема по Фреше, причем

$$F'(f)(h) = \int_0^1 \Psi'(f(s))h(s)\,ds.$$

Действительно,

$$\frac{F(f+th)-F(f)}{t} = \\ = \int_0^1 \frac{\Psi(f(s)+th(s))-\Psi(f(s))}{t}\,ds \to \int_0^1 \Psi'(f(s))h(s)\,ds$$

при $t \to 0$, ибо по теореме о среднем отношение под интегралом равно $\Psi'(\xi_{s,t})h(s)$, где $\xi_{s,t}$ — некоторая точка между $f(s)$ и $f(s)+th(s)$. В силу непрерывности Ψ' при $t \to 0$ величины $\Psi'(\xi_{s,t})h(s)$ стремятся к $\Psi'(f(s)$ равномерно по s. Поэтому предел интегралов равен указанному выражению. Это показывает дифференцируемость по Гато, но производная Гато непрерывна по f (также ввиду непрерывности Ψ'), поэтому в силу упомянутой выше задачи есть и дифференцируемость по Фреше.

Однако ситуация меняется, если на $C[0,1]$ ввести интегральную норму

$$\|f\|_1 = \int_0^1 |f(s)|\,ds.$$

Здесь для $\Psi(u) = \sin u$ уже нет дифференцируемости по Фреше (дифференцируемость по Гато легко проверяется). В самом деле, если бы в точке $f_0 = 0$ была дифференцируемость по Фреше, то согласно сказанному выше для всякой стремящейся к нулю последовательности чисел t_n и всякой последовательности элементов h_n с $\|h_n\|_1 \leqslant 2$ мы бы имели $F(t_nh_n)/t_n - F'(0)(h_n) \to 0$, где $F'(0)(h_n)$ есть интеграл от $h_n(s)$ по $[0,1]$. Иначе говоря, интеграл от $\sin(t_nh_n(s))/t_n - h_n(s)$ стремился бы к нулю. Но это не так: возьмем $t_n = n^{-1}$ и такие функции $h_n \in C[0,1]$, что $h_n(t) = n$ при $t \in [0,1/n]$, а модуль интеграла от $n\sin(n^{-1}h_n(s)) - h_n(s)$ по $[1/n,1]$ меньше $1/n$. Этого легко добиться, положив $h_n(s) = 0$ при $s \geqslant 1/n + \delta_n$ для достаточно малого δ_n. Тогда $\sin(t_nh_n(s))/t_n - h_n(s) = n(\sin 1 - 1)$ при $s \in [0,1/n]$, что после интегрирования дает ненулевую константу $\sin 1 - 1$, которую не может компенсировать стремящийся к нулю интеграл по $[1/n,1]$.

1.5. Обратные и неявные функции

Во многих вопросах современного анализа используется следующее утверждение, называемое теоремой о сжимающем отображении или принципом сжимающих отображений. Сжимающим отображением (или сжатием) метрического пространства (X, d) называют такое отображение $f\colon X \to X$, что при некотором $L < 1$ верна оценка

$$d(f(x), f(y)) \leqslant Ld(x, y) \quad \forall\, x, y \in X.$$

Для дифференцируемого отображения в $\mathbb{R}^n$ достаточным условием является оценка $\|Df(x)\| \leqslant L < 1$.

1.5.1. Теорема. *Всякое сжатие f непустого полного метрического пространства X имеет единственную неподвижную точку $\widehat{x}$, т. е. $f(\widehat{x}) = \widehat{x}$. При этом $d(\widehat{x}, x_n) \leqslant L^n(1-L)^{-1}d(x_1, x_0)$ для всякой точки $x_0 \in X$, где $x_{n+1} := f(x_n)$.*

Доказательство. Покажем, что последовательность $\{x_n\}$ фундаментальна. Для этого заметим, что

$$d(x_{k+1}, x_k) \leqslant Ld(x_k, x_{k-1}) \leqslant \cdots \leqslant L^k d(x_1, x_0).$$

Поэтому $d(x_{n+m}, x_n)$ оценивается через

$$\begin{aligned} d(x_{n+m}, x_{n+m-1}) + d(x_{n+m-1}, x_{n+m-2}) + \cdots + d(x_{n+1}, x_n) \leqslant \\ \leqslant L^{n+m-1}d(x_1, x_0) + L^{n+m-2}d(x_1, x_0) + \cdots + L^n d(x_1, x_0), \end{aligned}$$

что дает $d(x_{n+m}, x_n) \leqslant L^n(1-L)^{-1}d(x_1, x_0)$. Из этой оценки и условия $L < 1$ следуют фундаментальность $\{x_n\}$ и существование предела $\widehat{x} = \lim\limits_{n\to\infty} x_n$. Очевидно, что

$$f(\widehat{x}) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = \widehat{x}$$

ввиду непрерывности f. Единственность неподвижной точки ясна из того, что $d(\widehat{x}, y) = d\bigl(f(\widehat{x}), f(y)\bigr) \leqslant Ld(\widehat{x}, y)$ для другой неподвижной точки y. Очевидна и оценка скорости сходимости. $\square$

Нам понадобится следующее простое наблюдение, вытекающее из доказательства.

1.5.2. Предложение. *Предположим, что X — полное метрическое пространство и отображения f и g из X в X липшицевы с постоянной $L < 1$. Тогда для неподвижных точек x_f и x_g этих отображений* *справедливо неравенство*

$$d(x_f, x_g) \leqslant (1-L)^{-1} \sup_{x\in X} d\bigl(f(x), g(x)\bigr).$$

Доказательство. Положим $y_n = g^n(x_f)$. Из доказательства теоремы выше ясно, что $x_g = \lim\limits_{n\to\infty} y_n$. Поэтому $d(x_f, x_g) = \lim\limits_{n\to\infty} d(x_f, y_n)$. Как и в доказательстве теоремы, для каждого n мы имеем

$$\begin{aligned} d(x_f, y_n) &\leqslant d\bigl(x_f, g(x_f)\bigr)(1 + L + \cdots + L^n) \leqslant \\ &\leqslant (1-L)^{-1} d\bigl(x_f, g(x_f)\bigr) = (1-L)^{-1} d\bigl(f(x_f), g(x_f)\bigr), \end{aligned}$$

откуда вытекает доказываемое утверждение. □

Перейдем к теореме об обратной функции — основному результату этого параграфа.

1.5.3. Теорема. *Пусть отображение $F\colon \mathbb{R}^n \to \mathbb{R}^n$ имеет непрерывную производную в окрестности точки a, а оператор $DF(a)$ обратим. Тогда F отображает взаимно однозначно некоторую окрестность точки a на некоторый открытый шар с центром в $F(a)$, на котором обратное отображение $G = F^{-1}$ непрерывно дифференцируемо и*

$$DG(F(a)) = DF(a)^{-1}.$$

Доказательство. Можно считать, что $a = 0$ и $F(a) = 0$, перейдя к $F(x) - F(a)$. Кроме того, можно считать, что $DF(a)$ есть единичный оператор, перейдя к композиции $DF(a)^{-1}F(x)$, ибо по правилу дифференцирования композиции производная нового отображения есть единичный оператор. В силу непрерывности производной найдется такой замкнутый шар U радиуса r с центром в нуле, что $\|DF(x) - I\| < 1/2$ при всех $x \in U$. Мы хотим добиться однозначной разрешимости уравнения

$$F(x) = y$$

для всех y из окрестности нуля. Это уравнение можно переписать в виде

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

Положим

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

По теореме о среднем $\|Df(x)\| = \|DF(x) - I\| < 1/2$ при всех $x \in U$. Поэтому $|f(x) - f(0)| \leqslant 2^{-1}|x|$ для всех x из U, что дает оценку $|f(x)| \leqslant |y| + 2^{-1}|x| \leqslant r$ при $|x| \leqslant r, |y| \leqslant r/2$. Таким образом, при

$|y| \leqslant r/2$ отображение f является сжимающим в замкнутом шаре радиуса r и переводит этот шар в него же. Поэтому уравнение $F(x) = y$ имеет единственное решение в U при каждом y из шара радиуса $r/2$. Это решение мы и возьмем в качестве $G(y)$. Из предыдущего предложения легко усмотреть, что G непрерывно. Более того, $|G(y)| \leqslant 2|y|$.

Покажем, что G имеет единичную производную в нуле (это и даст нужную формулу в общем случае). Надо доказать, что

$$G(y) = y + r(y), \quad r(y) = o(|y|).$$

Задав $r(y)$ равенством $r(y) = G(y) - y$, имеем $y = F(y + r(y))$, откуда

$$y = F(G(y)) = F(G(y)) - F(0) = F'(0)(G(y)) + o(G(y)) = y + r(y) + o(G(y)).$$

Значит, $r(y) + o(G(y)) = 0$. Так как $|G(y)| \leqslant 2|y|$, то $o(G(y))$ есть $o(|y|)$. Итак, $r(y) = o(|y|)$. Прообраз открытого шара радиуса $r/2$ при F в пересечении с U дает искомую окрестность. Непрерывность производной G следует из формулы $DG(y) = DF(G(y))$ и непрерывности DF и G. $\square$

Отметим, что непрерывное обратное отображение может существовать и при вырождении производной F в данной точке (например, для $F(x) = x^3$ в окрестности нуля), но для дифференцируемости обратного невырожденность $DF(a)$ необходима, ибо по правилу дифференцирования композиции произведение операторов $DG(F)$ и DF должно быть единичным оператором.

Следующий важный результат — теорема о неявной функции.

1.5.4. Теорема. *Пусть* $F\colon \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^k$ *имеет непрерывную производную в окрестности точки* (a, b), $F(a, b) = 0$, *причем производная отображения* $y \mapsto F(a, y)$ *из* $\mathbb{R}^k$ *в* $\mathbb{R}^k$ *в точке* b *обратима. Тогда существует такое непрерывно дифференцируемое отображение* f *из окрестности* b *со значениями в* $\mathbb{R}^k$, *что*

$$F(x, f(x)) \equiv 0, \quad f(a) = b.$$

В малой окрестности a *такое отображение единственно, а в малой окрестности* (a, b) *все решения уравнения* $F(x, y) = 0$ *задаются формулой* $y = f(x)$.

Доказательство. Рассмотрим вспомогательное отображение, заданное формулой

$$\Psi\colon \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n \times \mathbb{R}^k, \quad \Psi(x, y) = (x, F(x, y)).$$

Можно считать, что $a = 0$, $b = 0$. Отображение Ψ имеет непрерывную производную, $\Psi(0,0) = (0,0)$, причем производная Ψ в нуле обратима, так как
$$D\Psi(0,0)(h_1,h_2) = (h_1, DF(0,0)(h_2)),$$
где $DF(0,0)(h_2)) \neq 0$ при $h_2 \neq 0$ по условию теоремы. По теореме об обратной функции отображение $\Phi = \Psi^{-1}$ определено и непрерывно дифференцируемо в окрестности нуля. Это обратное отображение имеет вид
$$\Phi\colon (x,z) \mapsto (x, \varphi(x,z)).$$
Положим $f(x) := \varphi(x,0)$. Это отображение непрерывно дифференцируемо, ибо таково Φ. По построению $\Psi(x,\varphi(x,0)) = (x,0)$, откуда $F(x,\varphi(x,0)) = 0$. Единственность f с указанными свойствами легко усмотреть из локальной обратимости Ψ. $\square$

Смысл теоремы состоит в том, что мы хотим выразить y через x из соотношения $F(x,y) = 0$, про которое изначально дано, что $(0,0)$ ему удовлетворяет. Однако заранее даже не сказано, что есть другие решения. Их может и не быть: $x^2 + y^2 = 0$, или для них может оказаться невозможным выразить y через x: $x(1+y^2) = 0$. В чем разница между этими функциями и $x - y = 0$, для которой можно y выразить через x? Теорема говорит, что нужно условие $\partial F/\partial y(0,0) \neq 0$, нарушенное в указанных случаях. Если же условие выполнено, то не только есть другие решения, но есть целая кривая решений вида $(x, f(x))$ с дифференцируемой функцией f.

Следует иметь в виду, что утверждение о единственности не означает, то нет других отображений g, для которых $F(x,g(x)) = 0$ в окрестности a. Например, для уравнения $x^2 + y^2 = 1$ точки $(1,0)$ для какой угодно функции ε со значениями -1 и 1 мы имеем
$$x^2 + \Bigl(\varepsilon(x)\sqrt{1-x^2}\Bigr)^2 = 1.$$
Исключить такие решения нельзя и требованием $y(1) = 0$, но требование вхождения $(x, f(x))$ в малую окрестность $(1,0)$ дает нужный эффект.

Для функции $F(x,y)$ двух переменных условие теоремы требует неравенство $\partial_y F(x,y) \neq 0$ в точке (a,b), т. е. невырожденной должна быть производная по той переменной, которую мы хотим выразить из уравнения через остальные. В многомерном случае, конечно, невырожденность производной по y означает не просто отличие от нуля матрицы из элементов $\partial_{y_j} F_i$, где $F(x,y) = (F_1(x,y), \ldots, F_k(x,y))$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^k$, размера $k \times k$, но отличие от нуля ее определителя.

1.6. Производные высших порядков

Если отображение F из открытого множества U в $\mathbb{R}^n$ со значениями в $\mathbb{R}^k$ дифференцируемо во всех точках U, то возникает новое отображение DF из U в пространство $L(\mathbb{R}^n, \mathbb{R}^k)$ линейных операторов из $\mathbb{R}^n$ в $\mathbb{R}^k$. Это пространство можно отождествить с пространством матриц размера $n \times k$, а тем самым с $\mathbb{R}^{nk}$. Как уже говорилось, на пространстве операторов есть как операторная норма $\|A\| = \sup_{|h|=1} |Ah|$, так и норма Гильберта – Шмидта, но обе нормы оценивают друг друга с постоянными, поэтому для дальнейшего безразлично, какой из них пользоваться. В случае бесконечномерных пространств используется операторная норма. Если отображение DF дифференцируемо в точке x, то его производная называется второй производной отображения F в x и обозначается символами $F''(x)$ или $D^2F(x)$. Она принимает значения в пространстве $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^k))$ линейных отображений из $\mathbb{R}^n$ в $L(\mathbb{R}^n, \mathbb{R}^k)$. Это пространство можно отождествить с пространством билинейных отображений из $\mathbb{R}^n$ в $\mathbb{R}^k$, т. е. отображений $\Psi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^k$ таких, что линейны отображения $u \mapsto \Psi(u, v)$ и $v \mapsto \Psi(u, v)$. В самом деле, для $\Lambda \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^k))$ полагаем

$$\Psi(u, v) = \Lambda(u)(v),$$

ибо применение Λ к $u \in \mathbb{R}^n$ дает оператор, который можно применить к v и получить вектор из $\mathbb{R}^k$. Наоборот, если дано билинейное отображение Ψ, то по нему строится оператор Λ так: $\Lambda(u)(v) = \Psi(u, v)$.

Для числовой функции F (т. е. $k = 1$) пространство $L(\mathbb{R}^n, \mathbb{R}^1)$ линейных функций можно отождествить с самим $\mathbb{R}^n$, поэтому можно считать, что вторая производная принимает значения в пространстве матриц $n \times n$. Матрица второй производной в точке x имеет матричные элементы $\partial_{x_i}\partial_{x_j}F(x)$, состоящие из вторых частных производных по переменным x_i. Эти смешанные частные производные обозначают также символами $F_{x_i x_j}$ и

$$\frac{\partial^2 F(x)}{\partial x_i \partial x_j}.$$

1.6.1. Пример. (i) Вторая производная линейного отображения нулевая, так как первая производная не зависит от точки. Кроме того, отображение с нулевой второй производной имеет вид $F(x) = Ax + b$, где A линейно и b постоянно. Это следует из того, что по доказанному ранее для первых производных первая производная $DF(x)$ есть постоянный линейный оператор A. Значит, $F(x) - Ax$ имеет нулевую производную и потому постоянно.

(ii) Вторая производная квадратичной формы $f(x) = (Qx, x)$, заданной симметричным оператором Q, равна $2Q$, ибо первая производная есть $Df(x)(h) = 2(Qx, h)$, т. е. $\nabla f(x) = 2Qx$. Если же оператор Q не симметричен, то можно перейти к симметричному оператору $(Q+Q^*)/2$, задающему ту же самую квадратичную форму, и получить $D^2 f(x) = (Q + Q^*)/2$.

(iii) Если дважды дифференцируемая функция f имеет не зависящую от точки вторую производную $D^2 f(x) = Q$, то

$$f(x) = 2^{-1}(Qx, x) + (x, v) + c,$$

где v — постоянный вектор и c — число. В самом деле, в силу (i) разность $f(x) - 2^{-1}(Qx, x)$ есть линейная функция плюс постоянная.

Очень важным свойством второй производной является то, что симметрична, т. е.

$$D^2F(x)(u)(v) = D^2F(x)(v)(u),$$

что можно записать также в виде

$$\partial_{x_i}\partial_{x_j}F(x) = \partial_{x_j}\partial_{x_i}F(x). \tag{1.6.1}$$

Доказательство можно прочитать в главе X учебника Зорича. Мы приведем доказательство следующего более специального утверждения, связанного лишь с частными производными.

1.6.2. Теорема. *Пусть отображение $F\colon U \to \mathbb{R}^k$ для некоторой пары индексов $i \neq j$ имеет в U частные производные $\partial_{x_i}\partial_{x_j}F$ и $\partial_{x_j}\partial_{x_i}F$, которые непрерывны в некоторой точке x. Тогда в этой точке верно равенство* (1.6.1).

Доказательство. Достаточно рассмотреть случай $k = 1$. Можно считать, что речь идет о производных в нуле. Кроме того, утверждение сводится к случаю функции $F(x, y)$ двух переменных. Пусть $h > 0$. Выражение

$$F(h, h) - F(h, 0) - F(0, h) + F(0, 0)$$

представим в виде $G(h) - G(0)$, где $G(x) = F(x, h) - F(x, 0)$, и по теореме о среднем получим, что оно равно

$$hG'(\xi_h) = h\frac{\partial F(\xi_h, h)}{\partial x} - h\frac{\partial F(\xi_h, 0)}{\partial x}.$$

где $\xi_h \in [0,h]$. Еще раз применяя теорему о среднем, находим точку $\eta_h \in [0,h]$, для которой последняя разность равна

$$h^2 \frac{\partial^2 F(\xi_h, \eta_h)}{\partial y \partial x}.$$

С другой стороны, указанное выше выражение есть $\Psi(h) - \Psi(0)$, где $\Psi(y) = F(h,y) - F(0,y)$, что дает точку $\zeta_h \in [0,h]$, для которой оно равно

$$h\Psi'(\zeta_h) = h\frac{\partial F(h,\zeta_h)}{\partial y} - h\frac{\partial F(0,\zeta_h)}{\partial y},$$

т. е.

$$h^2 \frac{\partial^2 F(\tau_h, \zeta_h)}{\partial x \partial y}$$

при некотором $\tau_h \in [0,h]$. Значит,

$$\frac{\partial^2 F(\xi_h, \eta_h)}{\partial y \partial x} = \frac{\partial^2 F(\tau_h, \zeta_h)}{\partial x \partial y}.$$

При $h \to 0$ имеем $\xi_h \to 0$, $\eta_h \to 0$, $\zeta_h \to 0$, $\tau_h \to 0$, поэтому в силу непрерывности смешанных частных производных в нуле получаем нужное равенство. $\square$

Приведем иное обоснование при немного более сильном условии. Это обоснование предполагает минимальное знание интеграла Римана.

1.6.3. Теорема. *Пусть непрерывное отображение* $F\colon U \to \mathbb{R}^k$ *для некоторой пары индексов* $i \neq j$ *имеет в* U *непрерывные частные производные* $\partial_{x_i} F$, $\partial_{x_j} F$, $\partial_{x_i}\partial_{x_j} F$, $\partial_{x_j}\partial_{x_i} F$. *Тогда верно* (1.6.1).

Доказательство. Как и выше, достаточно рассмотреть случай $k = 1$, причем можно считать, что речь идет о производных в нуле для функции двух переменных. Рассмотрим интеграл функции $\partial_{x_1}\partial_{x_2} F$ по маленькому квадрату с вершинами в $(0,0)$, $(h,0)$, (h,h), $(0,h)$, где $h > 0$. Вычислим интеграл этот интеграл по теореме Фубини. При фиксированном x_2 интеграл по x_1 из $[0,h]$ по формуле Ньютона – Лейбница равен $\partial_{x_2} F(h, x_2) - \partial_{x_2} F(0, x_2)$. Интеграл этой разности по x_2 из $[0,h]$ равен $F(h,h) - F(h,0) - F(0,h) + F(0,0)$. Это же значение мы получаем для интеграла от $\partial_{x_2}\partial_{x_1} F$. При этом эти интегралы заключены между минимумами и максимумами соответствующих смешанных частных производных, умноженными на h^2. Следовательно, отношения указанных равных интегралов к h^2 при $h \to 0$ стремятся к $\partial_{x_1}\partial_{x_2} F(0,0)$ и $\partial_{x_2}\partial_{x_1} F(0,0)$ соответственно, что доказывает совпадение этих величин. $\square$

Правда, надо иметь в виду, что без дополнительных условий это утверждение может быть неверно (хотя в случае существования полной производной, а не частных, ничего больше не требуется). Например, рассмотрим функцию

$$f(x,y) = xy\frac{x^2 - y^2}{x^2 + y^2},$$

доопределенную нулем в нуле. Тогда $\partial_x f(0,y) = -y$ при $y \neq 0$, $\partial_x f(0,0) = 0$, откуда $\partial_y \partial_x f(0,0) = -1$. Аналогично $\partial_y f(x,0) = x$ при $x \neq 0$, $\partial_y f(0,0) = 0$, откуда $\partial_x \partial_y f(0,0) = 1$.

Построение следующих производных осуществляется по индукции, причем индукцию можно вести двумя способами. А именно можно производную $D^p F(x)$ порядка $p > 2$ определить как производную первого порядка производной $D^{p-1}F$ порядка $p-1$, но можно ввести ее также как производную порядка $p-1$ первой производной. В качестве упражнения полезно проверить, что оба способа равносильны.

Аналогично второй производной можно отождествить производную $D^p F(x)$ порядка p с некоторым p-линейным отображением из произведения p экземпляров пространства $\mathbb{R}^n$ в $\mathbb{R}^k$. Значение этого отображения на векторах $h_1, \ldots, h_p$ задается формулой

$$D^p F(x)(h_1, \ldots, h_p) = \partial_{h_p} \cdots \partial_{h_1} F(x),$$

где

$$\partial_h F(x) = \lim_{t \to 0} \frac{F(x + th) - F(x)}{t}.$$

Из (1.6.1) следует симметричность производной порядка p, т. е. инвариантность $D^p F(x)(h_1, \ldots, h_p)$ относительно перестановок векторов h_i.

Приведем теперь многомерную формулу Тейлора.

1.6.4. Теорема. *Пусть U — открытое множество в $\mathbb{R}^n$ и отображение $F\colon U \to \mathbb{R}^k$ имеет в U производные до порядка $p-1$.*

(i) *Если в некоторой точке $x_0 \in U$ существует производная $D^p F(x_0)$, то при $\|h\| \to 0$ имеем*

$$\|F(x_0 + h) - F(x_0) - DF(x_0)h - \cdots - \frac{1}{p!} D^p F(x_0)(h, \ldots, h)\| = o(\|h\|).$$

(ii) Для числовой функции F, имеющей в U непрерывную производную порядка p

$$F(x+h) = F(x) + DF(x)h + \cdots + \frac{1}{(p-1)!}D^{p-1}F(x)(h,\ldots,h) +$$
$$+ \frac{1}{p!}\int_0^1 (1-t)^{p-1} D^p F(x+th)(h,\ldots,h)\,dt. \qquad (1.6.2)$$

Доказательство (i) аналогично одномерному случаю. Утверждение (ii) вытекает из одномерного случая. Для этого надо рассмотреть функцию одного переменного

$$f(t) = F(x+th).$$

Из условия следует, что эта функция имеет p непрерывных производных на $[0,1]$, причем

$$f^{(p)}(0) = D^p F(x)(h,\ldots,h).$$

Например,

$$f'(0) = \lim_{t\to 0}\frac{F(x+th)-F(x)}{t} = DF(x)(h).$$

Остается вспомнить формулу Тейлора для f.

Функция f на $\mathbb{R}^n$ называется бесконечно дифференцируемой, если она имеет во всех точках производные всех порядков. Для этого достаточно (конечно, и необходимо) существование и непрерывность всех частных производных (всех порядков). Однако от непрерывности частных производных отказаться нельзя.

1.6.5. Пример. Для $z \in \mathbb{C}$ положим $f(z) = \exp(-z^{-4})$ при $z \neq 0$, $f(0) = 0$. Тогда f как функция на $\mathbb{R}^2$ имеет все частные производные $\partial_x^n \partial_y^k f$ на всей плоскости, но разрывна в нуле.

Простым примером бесконечно дифференцируемой функции многих переменных служит бесконечно дифференцируемая функция одного переменного. Вот менее банальный пример:

$$f(x) = g((x,x)),$$

где g — гладкая функция на прямой. По правилу дифференцирования композиции имеем

$$\nabla f(x) = 2g'((x,x))x,$$
$$D^2 f(x)(h_1,h_2) = 4g''((x,x))(x,h_1)(x,h_2) + 2g'((x,x))(h_1,h_2).$$

Если g имеет носитель в $[-R,R]$, то f имеет носитель в шаре радиуса $R^{1/2}$.

Глава 2

Мера Лебега

2.1. Основные задачи теории меры

Большинство теорий интеграла в том или ином виде имеют дело с теорией меры. Во многих задачах совершенно разных областей бывает полезно задавать меру множеств. Задача определения длин, площадей и объемов восходит к глубокой древности. Великие ученые античности дали блестящие образцы решения этой задачи в различных частных случаях, решения, которые относятся к величайшим достижениям науки. Однако формализация самой постановки этой задачи появилась не так уж давно, во второй половине XIX века, причем для классических мер. Как можно определить длину множества E в $[0, 1]$? Для элементарного множества, т. е. конечного объединения дизъюнктных промежутков, естественно назвать длиной сумму длин этих промежутков (считая, что уже договорились, что это означает).

Как быть с неэлементарными множествами? Конструкция Пеано – Жордана близка античному подходу и состоит в том, что рассматриваются вписанные в E элементарные множества и берется точная верхняя грань их длин (это дает внутреннюю меру Жордана $\mathrm{mes}_i(E)$), а также содержащие E элементарные множества и точная нижняя грань их мер (это дает внешнюю меру Жордана $\mathrm{mes}_e(E)$).

Если $\mathrm{mes}_i(E) = \mathrm{mes}_e(E)$, то множество называется измеримым по Жордану, а полученное общее значение объявляется его мерой Жордана. Можно проверить, что это равносильно интегрируемости по Риману индикатора I_E множества E. Например, в множество Q рациональных чисел в $[0, 1]$ можно вписать лишь конечные элементарные множества, поэтому его внутренняя мера Жордана равна нулю. Всякое же элементарное множество, содержащее все рациональные числа, может быть лишь всем отрезком без конечного числа иррациональных точек. Значит, внешняя мера Жордана множества Q равна 1 (при этом оно очевидным образом имеет лебеговскую меру нуль!).

Таким образом, класс измеримых по Жордану множеств имеет недостатки, похожие на недостатки интеграла Римана. В конце XIX века предпринимались попытки усовершенствовать определение меры, но лишь в самом начале XX века ощутимый прогресс был достигнут молодым французским математиком Анри Лебегом.

2.1.1. Определение. *Внешняя мера Лебега множества E в отрезке определяется формулой*

$$\lambda^*(E) = \inf\Bigl\{\sum_{i=1}^{\infty} \lambda(U_i), \quad E \subset \bigcup_{i=1}^{\infty} U_i\Bigr\}, \tag{2.1.1}$$

где $\inf$ *берется по всем счетным или конечным покрытиям интервалами U_i, $\lambda(U_i)$ — обычная длина. Аналогично определяется внешняя мера $\lambda^*(E) = \lambda_d^*(E)$ множества E в кубе в $\mathbb{R}^d$, в качестве U_i берутся открытые кубы с ребрами, параллельными координатным осям.*

Ясно, что $0 \leqslant \lambda^*(E) \leqslant \lambda(I)$. Для всякого отрезка J получаем $\lambda^*(J) = \lambda(J)$, ибо сумма длин покрывающих интервалов не может стать меньше длины отрезка J (счетное покрытие имеет конечное подпокрытие, поэтому здесь все сводится к конечным покрытиям).

Внешнюю меру определили на всех множествах, но в большинстве приложений полезно иметь дело со счетно-аддитивными мерами, причем на областях определения, допускающих счетные теоретико-множественные операции. Еще античные математики вычисляли объемы сложных тел методом, который в современных терминах можно назвать счетным разбиением на элементарные множества.

Таким образом, естественное требование к области определения меры состоит в том, чтобы она была замкнута относительно счетных объединений, пересечений и взятия дополнения. Это то, что в следующем параграфе будет названо сигма-алгеброй. От самой же меры хотят счетную аддитивность, т.е. чтобы значение меры на счетном объединении дизъюнктных частей из области определения было равно сумме ряда мер этих частей. Конечная аддитивность — такое равенство для конечных объединений.

Вот с мерой Жордана проблема как раз в незамкнутости области определения относительно счетных объединений. Конечно, классическая проблема сложна дополнительным условием: на неких множествах мера уже предписана, поэтому речь идет не о переопределении, а о продолжении. Если забыть об уже измеренных множествах, то никаких проблем нет: можно нулем все сделать, а если желать получить вероятностные меры (неотрицательные со значением 1 на всем пространстве), то можно взять меру Дирака δ_a в какой-либо точке a, т.е.

$\delta_a(A) = 1$, если $a \in A$, $\delta_a(A) = 0$, если $a \notin A$, либо похожим образом устроенную меру со счетным носителем, сложив меры Дирака δ_{a_n} в точках a_n с весами $c_n > 0$ из сходящегося ряда. Такие меры находят много применений, но не согласуются с предписанными значениями классических объемов. Наконец, в классической ситуации продолжения предписанных объемов еще один нюанс возникает: на исходных элементарных множествах площадь и объем инвариантны относительно некоторых преобразований типа сдвигов или поворотов, но надо ли и можно ли требовать это же от продолжений?

Какие же ответы удается получить на эти вопросы?

• Длину, площадь, объем можно продолжить до счетно-аддитивных мер на сигма-алгебрах, содержащих исходные элементарные множества. При этом продолжения на наименьшие такие сигма-алгебры единственны и даются формулой Лебега (2.1.1).

• Нам не дано знать, задает ли формула Лебега счетно-аддитивную функцию на всех множествах. Если принять аксиому выбора, то ответ отрицательный (но может быть положительным при иных аксиомах).

• Нам также не дано знать, можно ли длину продолжить счетно-аддитивно на все множества каким-либо иным способом, отличным от формулы Лебега. Однако и здесь при неких дополнительных теоретико-множественных аксиомах (но не аксиоме выбора) ответ отрицательный. Похожа ситуация с инвариантными продолжениями.

• Если оставить в стороне меры с предписанными значениями на элементарных множествах и обратиться к общим ненулевым счетно-аддитивным мерам, то можно эффективно строить вероятностные меры указанного выше дираковского типа на всех множествах, однако без дополнительных теоретико-множественных предположений нельзя узнать, есть ли счетно-аддитивные меры на классе всех множеств, обращающиеся в нуль на всех точках (их нет при неких аксиомах).

• Не следует пытаться строить меры на всех множествах, а надо удовольствоваться наименьшими сигма-алгебрами, содержащими элементарные множества, т.е. борелевскими сигма-алгебрами. Борелевские множества не имеют явного описания, да и вряд ли можно понять, что это такое, но к ним можно привыкнуть, как мы привыкаем к телефонам и компьютерам, не вполне зная их устройство. Немного привыкнув к ним, можно успешно обращаться с общими борелевскими мерами. Они не связаны с классической мерой Лебега, но удивительным образом получаются из нее измеримыми преобразованиями.

Далее эти тезисы наполняются некоторыми техническими деталями.

2.2. Алгебры и σ-алгебры

2.2.1. Определение. *Класс подмножеств $\mathcal{A}$ множества X называется алгеброй, если он содержит все X и допускает взятие дополнения, конечного объединения и пересечения.*

Класс подмножеств $\mathcal{A}$ множества X называется σ-алгеброй (сигма-алгеброй), если он содержит все X и допускает взятие дополнения, счетного объединения и счетного пересечения.

Условие с пересечением включено для симметрии и наглядности, но технически оно вытекает из первых двух, ибо

$$\bigcap_{n=1}^{\infty} A_n = X \setminus \bigcup_{n=1}^{\infty} (X \setminus A_n).$$

Ясно, что σ-алгебра есть алгебра, замкнутая относительно счетных объединений.

2.2.2. Пример. (i) Наименьшая σ-алгебра в X состоит из пустого множества $\varnothing$ и всего X, наибольшая — класс 2^X всех подмножеств в X.

(ii) Класс всяческих конечных объединений промежутков вида (a, b), $[a, b]$, $[a, b)$, $(a, b]$ в фиксированном отрезке есть алгебра, но не σ-алгебра.

(iii) Класс всех конечных подмножеств X и их дополнений есть алгебра, но не σ-алгебра для бесконечного X.

(iv) Класс всех не более чем счетных подмножеств X и их дополнений есть σ-алгебра.

2.2.3. Определение. *Для всякого класса $\mathcal{F}$ подмножеств X есть наименьшая σ-алгебра, содержащая $\mathcal{F}$. Она обозначается символом $\sigma(\mathcal{F})$ и задается формулой*

$$\sigma(\mathcal{F}) = \textit{пересечение всех } \sigma\textit{-алгебр, содержащих.}$$

Эта σ-алгебра называется порожденной классом $\mathcal{F}$.

Борелевской σ-алгеброй $\mathcal{B}(X)$ топологического пространства X называется σ-алгебра, порожденная всеми открытыми множествами. Множества из нее называют борелевскими.

Конечно, следует проверить, что указанная формула в самом деле дает σ-алгебру.

Эта формула лишь кажется явной, но ничего конструктивного в ней нет. Как правило, с ее помощью не удается проверить принадлежность к $\sigma(X)$. Скажем, практическая проверка проверка борелевости множества заключается в том, что удается указать счетную последовательность шагов получения этого множества счетными операциями.

Например, замкнутые множества борелевы как дополнения открытых. Счетные объединения замкнутых (скажем, множество рациональных чисел) борелевы, хотя могут не быть открытыми или замкнутыми.

Возникает искушение строить борелевские множества индуктивно: начать с класса $\mathcal{B}_0$ открытых множеств, класс $\mathcal{B}_{n+1}$ получать из $\mathcal{B}_n$ присоединением дополнений, счетных объединений и счетных пересечений множеств из $\mathcal{B}_n$. Можно показать (это довольно непросто: нелегко предъявить пример даже множества из класса $\mathcal{B}_3$, не входящего в $\mathcal{B}_2$), что все классы $\mathcal{B}_n$ различны и не являются σ-алгебрами, причем даже их объединение не σ-алгебра!

Таким образом, нет явного описания борелевских множеств. Правда, предыдущую попытку можно довести до успеха, если использовать трансфинитную индукцию и счетные ординалы, но ничего явного это тоже не даст. Хотя сейчас уже привыкли к этому обстоятельству, но сто лет назад оно считалось большой проблемой. В попытке ее решить студент Московского университета Михаил Суслин ввел так называемую операцию Суслина (А-операцию). Она задается так. Предположим, что для каждого упорядоченного конечного набора натуральных чисел $n_1, \ldots, n_k$ задан интервал $I_{n_1,\ldots,n_k}$. Тогда говорят, что задана таблица Суслина $\{I_{n_1,\ldots,n_k}\}$. Результат исполнения операции Суслина над этой таблицей есть множество

$$S\{I_{n_1,\ldots,n_k}\} = \bigcup_{(n_i)} \bigcap_{k=1}^{\infty} I_{n_1,\ldots,n_k},$$

где объединение берется по всем бесконечным последовательностям натуральных чисел (n_i), т.е. объединение континуально. Множества этого вида называются суслинскими. Оказывается, что все борелевские множества кодируются такими таблицами (это было основной целью Суслина — дать конструктивное задание борелевских множеств). На первый взгляд, это не очень удивительно из-за континуальности объединения, ведь почему бы не всем вообще множествам иметь такие кодировки? Однако нетрудно сообразить, что таблиц континуум, поэтому заведомо суслинских множеств тоже континуум, т.е. гораздо меньше, чем всех множеств. Суслин установил (кстати, в дипломной работе), что бывают неборелевские суслинские множества. Это выглядело как отрицательный результат, но на самом деле привело к чрезвычайно важному объекту современного анализа.

Борелевские σ-алгебры — самые типичные области определения счетно-аддитивных мер.

Банальное наблюдение: борелевская σ-алгебра порождается не только открытыми, но и замкнутыми множествами. На прямой она порождается интервалами с рациональными концами. Вместо интервалов можно брать отрезки или полуинтервалы (или лучи с рациональными концами), но недостаточно брать только точки (докажите!).

В $\mathbb{R}^n$ борелевская σ-алгебра порождается открытыми кубами с ребрами, параллельными осям. Можно брать и замкнутые кубы или шары (докажите!), но недостаточно брать только шары с центром в нуле.

В сепарабельном метрическом пространстве (все ли помнят после карантина, что это такое?) борелевская σ-алгебра тоже порождается открытыми шарами, скажем, рационального радиуса с центрами из точек счетного всюду плотного множества (докажите!). В несепарабельном метрическом пространстве это может быть неверно. Скажем, если на прямой взять дискретную метрику (все ненулевые расстояния равны 1), то шары — отдельные точки и все пространство. Порождаемая ими σ-алгебра состоит из счетных множеств и их дополнений, а борелевская из всех множеств, ибо точки этого странного пространства оказываются открытыми множествами, значит, таковы и их произвольные объединения.

2.3. Аддитивные и счетно-аддитивные функции множества

Перейдем к мерам.

2.3.1. Определение. *Числовая функция μ на σ-алгебре $\mathcal{A}$ называется счетно-аддитивной, если*

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n)$$

для всех наборов попарно непересекающихся множеств $A_n \in \mathcal{A}$.

Более слабое свойство, когда равенство верно для конечных объединений попарно непересекающихся множеств, называют конечной аддитивностью.

Разумеется, конечную аддитивность достаточно проверять для пар множеств. На σ-алгебрах существуют конечно-аддитивные функции множества, не являющиеся счетно-аддитивными, но конструктивных примеров нет (используется аксиома выбора). Легко привести конструктивный пример на алгебре.

2.3.2. Пример. *На алгебре конечных множеств в* $\mathbb{N}$ *и их дополнений введем функцию* m *так:* $m(A) = 0$, $m(\mathbb{N}\backslash A) = 1$ *для конечных* A. *Тогда* m *аддитивна, но не счетно-аддитивна.*

2.3.3. Замечание. Если функция μ аддитивна на $\mathcal{A}$, то ее счетная аддитивность равносильна тому, что для всякой последовательности множеств $B_n \in \mathcal{A}$ с $B_n \subset B_{n+1}$ (т.е. возрастающей) и объединением B из $\mathcal{A}$ мы имеем соотношение $\mu(B_n) \to \mu(B)$. Иначе говоря:

$$B_n \uparrow B \Rightarrow \mu(B_n) \to \mu(B).$$

Для этого для дизъюнктных A_n в качестве B_n берем $A_1 \cup \cdots \cup A_n$.

Это же равносильно тому, что для всякой последовательности множеств $C_n \in \mathcal{A}$ с $C_{n+1} \subset C_n$ и $\cap_n C_n = \varnothing$, т.е. убывающей к пустому множеству, мы имеем $\mu(C_n) \to 0$. Это называется «непрерывностью в нуле». Иначе говоря:

$$C_n \downarrow \varnothing \Rightarrow \mu(C_n) \to 0.$$

Действительно, в этом случае множества $B_n = C_1 \backslash C_n$ возрастают к C_1.

Явных примеров счетно-аддитивных мер на сигма-алгебрах мало. Вот два основных (в математике много — когда больше двух).

2.3.4. Пример. (i) Пусть $p_n \geqslant 0$, $\sum_n p_n = 1$. На всяком множестве $A \subset \mathbb{N}$ положим

$$\mu(A) = \sum_{n \in A} p_n.$$

Докажите, что получена счетно-аддтивная мера. Аналогично строятся меры на любых счетных множествах.

(ii) На σ-алгебре $\mathcal{A}$ всех подмножеств несчетного множества X, которые либо счетны, либо дополнительны к счетным, положим

$$\mu(A) = 0, \quad \mu(X \backslash A) = 1 \quad \text{для счетных } A.$$

Тогда мера μ счетно-аддитивна. В самом деле, если множества $A_n \in \mathcal{A}$ попарно непересекаются, то среди них не может быть более одного множества со счетным дополнением. Если таково A_1, то $\mu(A_1) = 1$, $\mu(A_n) = 0$ для остальных, причем дополнение к $\bigcup_n A_n$ тоже счетно, так что $\mu(\bigcup_n A_n) = 1$. Если же все A_n счетны, то таково и их объединение.

Поскольку бывают аддитивные не счетно-аддитивные функции, возникает вопрос о проверке счетной аддитивности не по определению (или предыдущему замечанию). Важнейший практически способ таков.

2.3.5. Определение. *Класс $\mathcal{K}$ подмножеств абстрактного пространства X называется компактным, если для всякой последовательности множеств $K_n \in \mathcal{K}$ с $\bigcap_{n=1}^{\infty} K_n = \varnothing$ есть такое N, что $\bigcap_{n=1}^{N} K_n = \varnothing$.*

Основной пример — произвольный набор компактов в данном пространстве (проверьте!). Для теории меры это нужно вот зачем.

2.3.6. Теорема. *Пусть μ — неотрицательная аддитивная функция множества на некоторой алгебре множеств $\mathcal{A}$ в пространстве X, причем есть компактный класс $\mathcal{K} \subset \mathcal{A}$, приближающий меру в следующем смысле: для всякого $A \in \mathcal{A}$ и всякого $\varepsilon > 0$ найдется такое $K_\varepsilon \in \mathcal{K}$, что $K_\varepsilon \subset A$ и $\mu(A \backslash K_\varepsilon) < \varepsilon$. Тогда μ счетно-аддитивна на $\mathcal{A}$.*

Доказательство. Проверим непрерывность меры в нуле. Пусть множества $A_n \in \mathcal{A}$ убывают и их пересечение пусто. Покажем, что $\mu(A_n) \to 0$. Пусть $\varepsilon > 0$. Возьмем такие $K_n \in \mathcal{K}$, что $K_n \subset A_n$ и $\mu(A_n \backslash K_n) < \varepsilon 2^{-n}$. Ясно, что $\bigcap_{n=1}^{\infty} K_n \subset \bigcap_{n=1}^{\infty} A_n = \varnothing$. По условию найдется такое m, что $\bigcap_{n=1}^{m} K_n = \varnothing$. Таким образом,

$$A_m = A_m \backslash \bigcap_{n=1}^{m} K_n = \bigcup_{n=1}^{m} (A_m \backslash K_n) \subset \bigcup_{n=1}^{m} (A_n \backslash K_n),$$

откуда

$$\mu(A_m) \leqslant \sum_{n=1}^{m} \mu(A_n \backslash K_n) \leqslant \sum_{n=1}^{m} \varepsilon 2^{-n} \leqslant \varepsilon.$$

Итак, $\mu(A_n) \to 0$, что влечет счетную аддитивность μ. $\square$

С помощью этого утверждения мы проверим счетную аддитивность меры Лебега на алгебре, порожденной промежутками или кубами. Однако это еще не полностью решает проблему: ведь нужны меры на сигма-алгебрах. Этому посвящен следующий параграф, но сейчас отметим, что данное выше условие счетной аддитивности весьма универсально, так что весьма непросто найти счетно-аддитивную меру, не обладающую приближающим компактным классом (хотя такие есть). А именно верна следующая теорема (доказательство — задача).

2.3.7. Теорема. *Пусть μ — неотрицательная счетно-аддитивная мера на борелевской σ-алгебре $\mathcal{B}(\mathbb{R}^n)$ пространства $\mathbb{R}^n$. Тогда для всякого борелевского множества $B \subset \mathbb{R}^n$ и всякого $\varepsilon > 0$ найдутся такие открытое множество U_ε и компакт K_ε, что $K_\varepsilon \subset B \subset U_\varepsilon$ и $\mu(U_\varepsilon \backslash K_\varepsilon) < \varepsilon$.*

2.4. Внешняя мера и продолжения мер

Поскольку одной из первичных целей является построение счетно-аддитивных мер на сигма-алгебрах, а совсем явно они могут быть построены в немногих случаях, возникает план строить меры на алгебрах и продолжать их. Эффективное и выполняющееся во всех приложениях условие счетной аддитивности аддитивной функции множества рассмотрено выше: компактные приближающие классы. Но вот есть хорошая мера на алгебре, скажем, лебеговская, можно ли ее продолжить на сигма-алгебру? Вообще говоря, даже если счетно-аддитивная мера уже задана на сигма-алгебре, то не всегда ее можно продолжить на данную более широкую сигма-алгебру.

2.4.1. Пример. В борелевской σ-алгебре прямой возьмем под-σ-алгебру множеств первой категории и их дополнений, а также меру μ из примера 2.3.4(ii), равную 0 на множествах первой категории и 1 на их дополнениях. Эту меру нельзя продолжить до счетно-аддитивной меры на борелевской σ-алгебре. Действительно, если ν — такое продолжение, то $\nu(\mathbb{Q}) = 0$. По теореме 2.3.7 множество иррациональных чисел должно содержать компакт K с $\nu(K) > 0$. Однако такой компакт нигде не плотен, поэтому есть множество первой категории, значение на котором ν совпадает с нулевым значением исходной меры μ.

Тем не менее замечательным образом оказывается, что всякая счетно-аддитивная мера μ на алгебре $\mathcal{A}$ продолжается до счетно-аддитивной меры на минимальной сигма-алгебре, содержащей данную алгебру (предыдущий пример этому не противоречит). Более того, продолжение осуществляется явной формулой (впрочем, степень явности этой формулы не стоит переоценивать), которая задает так называемую *внешнюю меру* на ВСЕХ множествах:

$$\mu^*(E) = \inf\Bigl\{\sum_{j=1}^{\infty}\mu(A_j)\colon A_j \in \mathcal{A},\, E \subset \bigcup_{j=1}^{\infty} A_j\Bigr\}.$$

Это в точности аналог внешней меры Лебега для класической длины.

Будет ли внешняя мера хотя бы аддитивна на всех множествах? Ответ зависит от теоретико-множественных установок. Если принять аксиому выбора, то ответ отрицательный. Аксиома выбора разрешает образовать множество, содержащее ровно по одному элементу из данного набора непустых дизъюнктных множеств. Точнее говоря, имеется шкала аксиом выбора для разных мощностей наборов множеств. Самая слабая — счетные наборы счетных множеств. Такая счетная форма аксиомы выбора необходима даже для того, чтобы менять порядок суммирования в двойном ряде.

2.4.2. Пример. (Пример Витали) Точки x и y прямой объявим эквивалентыми, если $x-y$ рационально. Знающие, что такое отношение эквивалентности, должны быстро проверить, что такое отношение получено. Аксиома выбора позволяет взять множество V (множество Витали), содержащее ровно по одному представителю из каждого класса эквивалентности (всякий класс есть некий представитель и все отличающиеся на него рациональными сдвигами). Можно взять представителей из $[0,1]$. С таким множеством внешняя мера Лебега не может быть аддитивна.

В самом деле, если $\lambda^*(V) = 0$, то $\lambda^*(V + r) = 0$ для всякого рационального r. Легко проверить (ниже это сделано в общем случае), что тогда объединение счетного числа множеств $V + r$ тоже имеет нулевую внешнюю меру, что невозможно, ибо это вся прямая. Остается случай $\lambda^*(V) = \alpha > 0$. Здесь тоже противоречие: ясно, что $\lambda^*(V+r) = \lambda^*(V)$ при всех r, но при разных $p, q \in \mathbb{Q}$ множества $V+p$ и $V+q$ не пересекаются, ибо если $v_1+p = v_2+q$, то v_1 и v_2 эквивалентны, а в V лишь по одному представителю. Значит, при рациональных $r \in [0,1]$ все $V+r$ дизъюнктны и лежат в $[0,2]$, причем имеют равные внешние меры. Если бы внешняя мера была аддитивна, то на $[0,2]$ была бы бесконечна.

Есть два пути бороться с этой проблемой: 1) запретить аксиому выбора, 2) уменьшить область определения внешней меры. Лебег, хотя и не поверил в множество Витали, пошел по второму пути и ввел такое определение.

2.4.3. Определение. *Множество E называется измеримым относительно внешней меры μ^*, порожденной счетно-аддитивной мерой μ на алгебре $\mathcal{A}$, если*

$$\mu^*(E) + \mu^*(X \backslash E) = \mu(X).$$

Класс измеримых множеств обозначим через $\mathcal{A}_\mu$.

Таким образом, условие измеримости — минимум того, что надо сделать для аддитивности μ^* хотя бы на паре множеств E и $X \backslash E$.

Классикам всегда везет: оказалось, что введенный им класс — σ-алгебра, содержащая $\mathcal{A}$, причем на ней внешняя мера счетно-аддитивна.

Сейчас это докажем, но нужны две технические леммы, чтобы разгрузить доказательство основной теоремы.

2.4.4. Лемма. *Имеем* (i) $\mu^*(\varnothing) = 0$,

$$\text{(ii)} \quad \mu^*(A) \leqslant \mu^*(B) \quad \textit{при } A \subset B,$$

$$\text{(iii)} \quad \mu^*\Bigl(\bigcup_{n=1}^{\infty} E_n\Bigr) \leqslant \sum_{n=1}^{\infty} \mu^*(E_n) \quad \textit{для всех } E_n \subset X.$$

Доказательство. Первые два свойства очевидны. Для доказательства (iii) возьмем $\varepsilon > 0$ и для каждого n покроем E_n множествами $A_{nj} \in \mathcal{A}$ так, что

$$\sum_{j=1}^{\infty} \mu(A_{nj}) \leqslant \mu^*(E_n) + \varepsilon 2^{-n}.$$

Это можно, если помнить, что такое inf. Тогда $E = \bigcup_n E_n$ покрыто всеми A_{nj} и

$$\mu^*(E) \leqslant \sum_{n,j} \mu^*(A_{n,j}) \leqslant \sum_{n=1}^{\infty} (\mu^*(E_n) + \varepsilon 2^{-n}) = \sum_{n=1}^{\infty} \mu^*(E_n) + \varepsilon,$$

что дает нужную оценку при $\varepsilon \to 0$. □

2.4.5. Следствие. *Если $E \in \mathcal{A}$, то $\mu^*(E) = \mu(E)$ и $E \in \mathcal{A}_\mu$.*

Доказательство. Ясно, что $\mu^*(E) \leqslant \mu(E)$, ибо E само себя покрывает и входит в $\mathcal{A}$. Пусть $E \subset \bigcup_n A_n$, $A_n \in \mathcal{A}$. Заменив A_n на $E \cap A_n \in \mathcal{A}$, можно считать, что $A_n \subset E$. Если бы A_n были дизъюнктны, то получили бы равенство $\mu(E)$ сумме ряда из мер A_n. Но они могут пересекаться, поэтому в общем случае оценка, которая следует из того, что $\mu\bigl(\bigcup_{n=1}^N A_n\bigr) \to \mu(E)$ (свойство счетной аддитивности), причем $\mu\bigl(\bigcup_{n=1}^N A_n\bigr) \leqslant \sum_{n=1}^N \mu(A_n)$ из-за конечной аддитивности и неотрицательности (очевидно из дробления $A_1, \ldots, A_N$ на дизъюнктные куски). Итак, $\mu(E) \leqslant \sum_n \mu(A_n)$, так что меньше $\mu(E)$ сделать не получится, т.е. $\mu^*(E) = \mu(E)$. Это же верно и для дополнения, откуда $E \in \mathcal{A}_\mu$. □

Можно заметить, что в самой лемме не использовалась счетная аддитивность μ, а в следствии она была важна.

2.4.6. Лемма. *Имеют место соотношения*

$$\mu^*(A) = \lim_{n\to\infty} \mu(A_n), \quad \textit{если } A_n \in \mathcal{A},\ \ A_n \subset A_{n+1},\ \ A = \bigcup_n A_n,$$

$$\mu^*(A \cup B) + \mu^*(A \cap B) \leqslant \mu^*(A) + \mu^*(B) \quad \textit{для всех } A, B.$$

Доказательство. Ясно, что $\mu(A_n) \leqslant \mu^*(A)$. Пусть $\varepsilon > 0$. Найдем такие возрастающие номера n_k, что $\mu(A_{n_k}) > \lim\limits_{n\to\infty} \mu(A_n) - \varepsilon 2^{-k}$. Тогда $\mu(A_{n_{k+1}} \backslash A_{n_k}) < \varepsilon 2^{-k}$, причем множества A_{n_1}, $A_{n_2} \backslash A_{n_1}$ и т.д. покрывают A. Поэтому

$$\mu^*(A) \leqslant \mu(A_{n_1}) + \mu(A_{n_2} \backslash A_{n_1}) + \cdots \leqslant \mu(A_{n_1}) + \varepsilon \leqslant \lim_{n\to\infty} \mu(A_n) + \varepsilon,$$

что ввиду произвольности ε дает нужное равенство.

Для доказательства оставшейся оценки возьмем $\varepsilon > 0$ и найдем множества U и V вида $U = \bigcup_n U_n$, $V = \bigcup_n V_n$, где $U_n, V_n \in \mathcal{A}$, $U_n \subset U_{n+1}$, $V_n \subset V_{n+1}$, для которых $\mu^*(A) \geqslant \mu^*(U) - \varepsilon$, $\mu^*(B) \geqslant \mu^*(V) - \varepsilon$. Это можно сделать по определению внешней меры: покрыв U множествами $I_j \in \mathcal{A}$ с $\mu^*(U) \geqslant \sum_j \mu(I_j) - \varepsilon$ и взяв $U_n = \bigcup_{j=1}^n I_j$ и аналогично для V. Тогда $A \cup B \subset U \cup V$, $A \cap B \subset U \cap V$, значит,

$$\mu^*(A \cup B) + \mu^*(A \cap B) \leqslant \mu^*(U \cup V) + \mu^*(U \cap V).$$

По доказанному,

$$\mu^*(U \cup V) = \lim_{n\to\infty} \mu(U_n \cup V_n), \quad \mu^*(U \cap V) = \lim_{n\to\infty} \mu(U_n \cap V_n).$$

В силу аддитивности меры на $\mathcal{A}$ имеем

$$\mu(U_n \cup V_n) + \mu(U_n \cap V_n) \leqslant \mu(U_n) + \mu(V_n).$$

Так как $\mu(U_n) \to \mu^*(U)$, $\mu(V_n) \to \mu^*(V)$, получаем

$$\mu^*(U \cup V) + \mu^*(U \cap V) \leqslant \mu^*(U) + \mu^*(V) \leqslant \mu^*(A) + \varepsilon + \mu^*(B) + \varepsilon.$$

Вспоминая, что лева часть не меньше $\mu^*(A \cup B) + \mu^*(A \cap B)$, при $\varepsilon \to 0$ получаем нужную оценку. □

2.4.7. Теорема. *Класс $\mathcal{A}_\mu$ является σ-алгеброй, содержит $\mathcal{A}$, поэтому и $\sigma(\mathcal{A})$, внешняя мера μ^* счетно-аддитивна на $\mathcal{A}_\mu$ и продолжает μ.*

Доказательство. Уже знаем, что $\mathcal{A} \subset \mathcal{A}_\mu$ и μ^* продолжает μ. Если бы знать, что $\mathcal{A}_\mu$ — алгебра, а μ^* аддитивна, то все быстро проверяется: если $A_n \in \mathcal{A}_\mu$ дизъюнктны, то для $A = \bigcup_n A_n$ при всех N имеем

$$\sum_{n=1}^N \mu^*(A_n) = \mu^*\Bigl(\bigcup_{n=1}^N A_n\Bigr) \leqslant \mu^*(A) \leqslant \sum_n \mu^*(A_n),$$

откуда $\mu^*(A) = \sum_n \mu^*(A_n)$. Кроме того, $X\backslash A \subset X\backslash \bigcup_{n=1}^N A_n$ для всякого N, поэтому

$$\mu^*(A)+\mu^*(X\backslash A) \leqslant \sum_n \mu^*(A_n)+\mu(X)-\sum_{n=1}^{N} \mu^*(A_n) = \mu(X)+\sum_{N+1}^{\infty} \mu^*(A_n),$$

что при $N \to \infty$ дает оценку $\mu^*(A) + \mu^*(X\backslash A) \leqslant \mu(X)$. По лемме есть и противоположная оценка. Это означает измеримость A, т.е. замкнутость относительно и счетных объединений.

Проверим, что $\mathcal{A}_\mu$ есть алгебра. Дополнения можно брать по определению. Пусть $A, B \in \mathcal{A}_\mu$. Надо показать, что

$$\mu^*(A \cup B) + \mu^*(X\backslash(A \cup B)) \leqslant \mu(X),$$

так как обратное верно по лемме. Мы имеем равенство

$$\mu^*(A) + \mu^*(X\backslash A) + \mu^*(B) + \mu^*(X\backslash B) = 2\mu(X),$$

а также неравенства из последней леммы

$$\mu^*(A \cup B) + \mu^*(A \cap B) \leqslant \mu^*(A) + \mu^*(B),$$

$$\mu^*(X\backslash(A \cup B)) + \mu^*(X\backslash(A \cap B)) \leqslant \mu^*(X\backslash A) + \mu^*(X\backslash B),$$

где во второй оценке использованы тождества

$$X\backslash(A \cup B) = (X\backslash A) \cap (X\backslash B),\ X\backslash(A \cap B) = (X\backslash A) \cup (X\backslash B).$$

Сумма левых частей последних двух неравенств равна $2\mu(X)$, что возможно только в том случае, когда эти неравенства обращаются в равенства, ибо $\mu^*(A \cup B) + \mu^*(X\backslash(A \cup B)) \geqslant \mu(X)$ и аналогично для пересечения. Итак, $\mathcal{A}_\mu$ алгебра. Остается проверить, что $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$ для дизъюнктных $A, B \in \mathcal{A}_\mu$. Это видно из второго из предыдущих равенств (в которые превратились неравенства), ибо для измеримых A, B оно принимает вид $\mu^*(X\backslash(A \cup B)) + \mu(X) = 2\mu(X) - \mu^*(A) - \mu^*(B)$, где слева с учетом измеримости $A \cup B$ имеем $2\mu(X) - \mu^*(A \cup B)$. $\square$

СОГЛАШЕНИЕ: *далее счетно-аддитивная мера μ^* на σ-алгебре $\mathcal{A}_\mu$ обозначается прежним символом μ, как исходная мера. Значок μ^* сохраняется для значений на потенциально неизмеримых множествах.*

Отметим, что попутно мера μ продолжена на порожденную сигма-алгебру $\sigma(\mathcal{A})$. Отчего бы сразу не использовать явную формулу для продолжения только на $\sigma(\mathcal{A})$ и зачем эта морока с $\mathcal{A}_\mu$? Попробуйте так сделать и расскажите, если получится.

2.4.8. Следствие. *Описанное выше продолжение — единственно возможно неотрицательное счетно-аддитивное продолжение меры* μ *на порожденную сигма-алгебру* $\sigma(\mathcal{A})$, *а также единственное такое продолжение на* $\mathcal{A}_\mu$.

Доказательство. Пусть ν — неотрицательное счетно-аддитивное продолжение меры μ на $\sigma(\mathcal{A})$. Пусть $A \in \sigma(\mathcal{A})$ и $\varepsilon > 0$. Возьмем такие $A_n \in \mathcal{A}$, что $A \subset \bigcup_n A_n$ и $\mu^*(A) \geqslant \sum_n \mu(A_n) - \varepsilon$. Пусть $B = \bigcup_n A_n$. Тогда

$$\nu(A) \leqslant \nu(B) \leqslant \sum_n \nu(A_n) = \sum_n \mu(A_n) \leqslant \mu^*(A) + \varepsilon.$$

Значит, $\nu(A) \leqslant \mu^*(A)$. Точно так же $\nu(X\backslash A) \leqslant \mu^*(X\backslash A)$. Однако $\mu^*(A) + \mu^*(X\backslash A) = \nu(A) + \nu(X\backslash A) = \mu(X)$, поэтому $\nu(A) \leqslant \mu^*(A)$. Случай $\mathcal{A}_\mu$ аналогичен. □

2.4.9. Замечание. Короткий комментарий про первую возможность борьбы с неизмеримыми множествами: запрет аксиомы выбора. Как уже сказано, нехорошо совсем ее запретить, надо хотя бы счетную форму оставить, иначе даже рядов не будет (впрочем, кому-то это не покажется ужасным). Кроме того, что-то запретив, надо хоть что-то разрешить взамен. У аксиомы выбора есть конкурент: аксиома детерминированности (D). Важно при этом, что ее можно ввести при счетной аксиоме выбора. Аксиома (D) такова. Два студента матфака долго играют в следующую игру (не преферанс): дано непустое множество A на прямой, точки которой представлены в двоичном разложении, можно считать, что речь идет о множествах в $[0, 1]$. Игроки по очереди пишут 0 или 1, зная результаты предыдущих ходов. Если в конце пары полученная бинарная последовательность дает число из A, то выиграл первый, а если не из A, то вторая. Говорят, что для A один из играющих имеет выигрывающую стратегию, если можно может делать ходы, приводящие к победе. Например, если A состоит из одной точки $a = (a_i)$, то второй играющий выигрывает, просто заменив a_2. Ясно, что так же можно действовать для конечных множеств. Первый же может выиграть аналогично, если A есть дополнение одной точки. Множество A называется детерминированным, если один из играющих имеет выигрывающую стратегию. Аксиома (D): на матфаке все множества детерминированы. Можно доказать, что тогда на матфаке все множества измеримы по Лебегу. Впрочем, возможен иной взгляд на вещи: можно считать, что признаются за множества лишь детерминированные множества. Тут невозможно спорить, ведь неизвестно, что такое множество.

2.5. Пополнение мер и классическая мера Лебега

Выше уже говорилось, что специальная σ-алгебра $\mathcal{A}_\mu$, возникшая из алгебры $\mathcal{A}$, может быть шире наименьшей σ-алгебры $\sigma(\mathcal{A})$, содержащей $\mathcal{A}$. Скажем, для меры Дирака в нуле на σ-алгебре в $[0,1]$, состоящей из нуля, пустого множества и всего отрезка, σ-алгебра $\mathcal{A}_\mu$ есть класс всех множеств в отрезке. Но сколь велика разница в смысле меры? Ответ очень простой.

2.5.1. Предложение. *Пусть μ — счетно-аддитивная мера на σ-алгебре $\mathcal{A}$. Тогда $\mathcal{A}_\mu$ есть класс множеств вида $A \cup Z$, где $A \in \mathcal{A}$ и $\mu^*(Z) = 0$, т.е. к $\mathcal{A}$ надо добавить всевозможные подмножества множеств меры нуль из $\mathcal{A}$.*

Доказательство. Ясно, что указанные множества лежат в $\mathcal{A}_\mu$. Пусть теперь $E \in \mathcal{A}_\mu$. Для всякого k найдется множество $A_k \in \mathcal{A}$ с $E \subset A_k$ и $\mu(A_k) \leqslant \mu^*(E) + 1/k$. Это ясно из определения внешней меры. Тогда $A = \bigcap_k A_k \in \mathcal{A}$, $E \subset A$, $\mu(E) = \mu^*(E) = \mu(A)$. Аналогично для $X \backslash E$ находим $B \in \mathcal{A}$ с $X \backslash E \subset B$ и $\mu(X \backslash E) = \mu(B)$. Для $D = X \backslash B$ получаем $D \in \mathcal{A}$, $D \subset E \subset A$,

$$\mu(D) = \mu(X) - \mu(B) = \mu(E) = \mu(A).$$

Итак, $Z = E \backslash D \subset A \backslash D \in \mathcal{A}$, $E = D \cup Z$, $\mu(A \backslash D) = 0$. $\square$

Введенное выше множество $A \in \mathcal{A}$ для E существует для всякого множества E, даже неизмеримого. Оно называется измеримой оболочкой E. Такое множество не одно, но все они имеют равную меру $\mu^*(E)$, причем минимальны в том смысле, что между E и A нет множеств из $\mathcal{A}$ положительной меры.

Мера μ на $\mathcal{A}_\mu$ полна в том смысле, что в $\mathcal{A}_\mu$ входят и все подмножества множеств меры нуль, чего могло не быть на исходной σ-алгебре. Поэтому $\mathcal{A}_\mu$ называют также лебеговским пополнением $\mathcal{A}$ относительно меры μ.

Из доказанного следует, что если применять теорему о продолжении к $\mathcal{A} = \mathcal{A}_\mu$, то расширения не произойдет, т.е. $(\mathcal{A}_\mu)_\mu = \mathcal{A}_\mu$. Однако это отнюдь не означает, что нельзя продолжить μ на еще более широкую σ-алгебру. Например, если меру Лебегу рассмотреть на σ-алгебре в $[0,1]$ из множеств меры нуль и их дополнений, то сразу получится полная мера, но ее можно продолжить на всю σ-алгебру измеримых множеств, хотя и не с помощью внешней меры.

В качестве задачи полезно доказать, что если дано множество E не из σ-алгебры, на которой задана мера μ, то эту меру можно продолжить на σ-алгебру, порожденную $\mathcal{A}$ и дополнительным множеством E.

Классическая мера Лебега на отрезке прямой или на кубе в $\mathbb{R}^n$ задается как результат применения предыдущих построений к обычному объему на алгебре $\mathcal{A}$ конечных объединений произведений промежутков $P = J_1 \times \cdots \times J_n$ в $[0,1]^n$, где J_i — промежутки из $[0,1]$. Функция $\lambda_n(P) := \lambda(J_1) \times \cdot \times \lambda(J_n)$, заданная на таких произведениях и продолженная по аддитивности на их дизъюнктные конечные объединения, счетно-аддитивна, ибо обладает приближающим компактным классом из компактных множеств. Теорема о продолжении дает счетно-аддитивную меру λ_n на σ-алгебре $\mathcal{L}_n([0,1]^n)$ измеримых по Лебегу множеств в кубе. Эта σ-алгебра включает борелевскую σ-алгебру куба. Описанное продолжение называется мерой Лебега на кубе.

Из сказанного выше явствует, что для всякого измеримого по Лебегу множества A верно равенство

$$\mu(A) = \sup\{\mu(K)\colon K \text{ — компакт в } A.\}$$

Если сначала как-то задать меру Лебега на компактах, то в принципе так можно определять ее дальше.

Аналогично вводятся измеримые по Лебегу множества и мера Лебега на всяком кубе $[-N,N]^n$. Теперь можно ввести σ-алгебру $\mathcal{L}_n$ всех измеримых по Лебегу множеств в $\mathbb{R}^n$, состоящую из множеств, пересечения которых с кубами измеримы.

Чтобы ввести еще и меру Лебега на $\mathcal{L}_n$ заметим, что точно так же, как для числовых функций, можно ввести понятие счетно-аддитивной меры со значениями в $[0,+\infty]$, полагая, что сумма с бесконечностью есть бесконечность. Такие меры будет называть именно мерами со значениями в $[0,+\infty]$, а термин «мера» сохраним для конечных мер.

Счетно-аддитивная в этом смысле мера Лебега λ_n на $\mathcal{L}_n$ задается формулой

$$\lambda_n(A) = \sum_{j=1}^{\infty} \lambda_n(A \cap K_j),$$

где K_j — целочисленные сдвиги единичного куба $[0,1]^n$. Эти сдвиги могут пересекаться, но пересечения имеют нулевую меру. В качестве упражнения надо убедиться, что и правда получена счетно-аддитивная мера на σ-алгебре.

Из определения очевидно, что мера Лебега инвариантна при сдвигах: сдвиг измеримого множества A на вектор h измерим и

$$\lambda_n(A+h) = \lambda_n(A).$$

Этим свойством мера Лебега задается однозначно с точностью до множителя.

2.5.2. Лемма. *Если борелевская мера μ на $I = [0,1]^n$ такова, что $\mu(B) = \mu(B+h)$ для всех таких $B \in \mathcal{B}(I)$ и h, что $B \subset I$, $B+h \subset I$, то $\mu = c\lambda_n$ при некотором c.*

Доказательство. Из счетной аддитивности и инвариантности ясно, что грани кубов имеют нулевую μ-меру (куб содержит бесконечное число сдвигов грани). Поэтому значение $c = \mu(I)$ однозначно определяет значения меры μ на кубах вида $2^{-kn}I + h$ из I, полученных как произведения промежутков длины 2^{-k} после деления $[0,1]$ на 2^k промежутков равной длины (промежутки могут быть замкнутыми, открытыми, полуоткрытыми, поэтому кубы могут быть без части граней, ребер или вершин). Тем самым мера μ совпадает с мерой $c\lambda_n$ на алгебре конечных объединений кубов указанного вида. Тогда равенство остается в силе и на порожденной этой алгеброй σ-алгебре, а это борелевская σ-алгебра. □

2.5.3. Следствие. *Мера Лебега — единственная с точностью до множителя счетно-аддитивная мера со значениями в $[0,+\infty]$ на борелевской σ-алгебре в $\mathbb{R}^n$, инвариантная при сдвигах и конечная на $[0,1]^n$.*

Из этого вытекает, что единственность имеет место и на $\mathcal{L}_n$.

В качестве задачи (предлагавшейся в свое время на питерской городской олимпиаде для школьников) предлагается доказать такой интересный факт.

2.5.4. Теорема. *Для всякого $\varepsilon > 0$ найдется такое конечное число дизъюнктных шаров в $[0,1]$, что их суммарный объем больше $1-\varepsilon$.*

2.5.5. Следствие. *Всякое непустое открытое множество в $\mathbb{R}^n$ с точностью до множества меры нуль есть объединение последовательности дизъюнктных шаров.*

Шары в этом утверждении можно брать открытые или замкнутые. В качестве задачи докажите, что $\mathbb{R}^n$ нельзя представить в виде точного объединения замкнутых шаров с ненулевыми радиусами и попарно не пересекающимися внутренностями (такая задача предлагалась жюри на одной из последних студенческих олимпиад СССР, где уровень участников, среди которых был Г. Перельман, позволял надеяться на получение решения).

Из приведенного факта вытекает, что в определении внешней меры Лебега вместо покрытий кубами можно использовать покрытия шарами. Действительно, каждый куб из покрытия можно покрыть конечным набором шаров с суммой объемов, сколь угодно близкой к объему куба.

Отметим следующий простой и полезный факт.

2.5.6. Лемма. *Пусть $A \subset \mathbb{R}^n$ измеримо по Лебегу. Тогда для всякого липшицевого с постоянной L отображения F множество $F(A)$ измеримо и*

$$\lambda_n(F(A)) \leqslant L^n \lambda_n(A).$$

ДОКАЗАТЕЛЬСТВО. Достаточно рассмотреть ограниченные множества. В силу сказанного выше множество A есть объединение последовательности компактов и множества меры нуль. Образ компакта при F компактен, поэтому достаточно показать, что F переводит множества меры нуль в множества меры нуль. Пусть $\lambda_n(E) = 0$ и $\varepsilon > 0$. Покроем E кубами B_j с ребрами r_j и $\sum_j r_j^n \leqslant \varepsilon$. Тогда $F(B_j)$ лежит в шаре радиуса не более $Ln^{1/2}r_j$, т.е. в кубе с ребром $2Ln^{1/2}r_j$. Это дает $\lambda_n(F(E)) \leqslant 2^n L^n n^{n/2}\varepsilon$, откуда $\lambda_n(F(E)) = 0$. Итак, $F(A)$ измеримо. Чтобы доказать оценку для мер, надо применить упомянутую выше возможно вычислять внешние меры по покрытиям шарами, ибо для шара эта оценка очевидна. $\square$

2.5.7. Следствие. *Пусть $A \subset \mathbb{R}^n$ измеримо по Лебегу. Тогда множества αA и $U(A)$ для всяких числа α и ортогонального оператора U измеримы, причем*

$$\lambda_n(\alpha A) = |\alpha|^n \lambda_n(A), \quad \lambda_n(U(A)) = \lambda_n(A).$$

Более того, для всякого линейного оператора T имеем

$$\lambda_n(T(A)) = |\det T| \lambda_n(A).$$

ДОКАЗАТЕЛЬСТВО. Первые два утверждения следуют из результата о единственности меры Лебега, если заметить, что левые части равенств задают инвариантные борелевские меры (для этого нужна вытекающая из первого утверждения леммы измеримость αA и $U(A)$, оценка из лемма здесь не нужна). Последнее утверждение аналогично, но в нем для нахождения множителя пропорциональности еще полезно заметить, что невырожденный оператор есть композиция ортогональных операторов и растяжений по осям. $\square$

Глава 3

Интеграл Лебега

3.1. Определение интеграла Лебега

Пусть дана мера μ на пространстве X с σ-алгеброй $\mathcal{A}$, в качестве которой может быть взята и $\mathcal{A}_\mu$. Задание меры на множествах позволяет естественным образом ввести интеграл на функциях с конечным числом значений, принимаемых на множествах из $\mathcal{A}$, т.е. функциях вида

$$f(x) = c_1 I_{A_1}(x) + \cdots + c_n I_{A_n}(x),$$

где c_i — числа, $A_i \in \mathcal{A}$, I_A — индикатор множества A, т.е.

$$I_A(x) = 1 \text{ при } x \in A, \quad I_A(x) = 0 \text{ при } x \notin A.$$

Такие функции называют простыми. Конечно, они не такие уж простые, так как множества A_i могут быть сложными. Не следует путать эти функции со ступенчатыми на отрезки — те еще проще.

Для простой функции f ее интеграл Лебега задается формулой

$$\int_X f\, d\mu := \int_X f(x)\, \mu(dx) := \sum_{i=1}^n c_i \mu(A_i).$$

Функцию f можно разными способами записать в указанном виде, но легко проверить с помощью аддитивности меры, что указанная сумма от этого не зависит. Поэтому можно использовать представление, в котором все c_i различны (т.е. это все значения, взятые один раз), а множества A_i дизъюнктны (это не предполагается заранее).

Простые функции — функции с конечным числом значений с измеримыми прообразами значений. Уже на этом этапе интеграл Лебега приписывает интеграл неинтегрируемым по Риману функциям, например функции Дирихле (равной 1 на множестве рациональных чисел и 0 на иррациональных). Такая функция проста с точки зрения Лебега, ее интеграл равен нулю. Из определения очевидна оценка

$$\int_X f\,d\mu \leqslant \int_X g\,d\mu \quad \text{при } f \leqslant g,$$

а также оценка

$$\left|\int_X f\,d\mu\right| \leqslant \sup_x |f(x)|\,\mu(X).$$

Естественность такого введения интеграла при заданной мере состоит в том, что желательно иметь интеграл, линейный по f и дающий меру множества на ее индикаторе. Тем самым никак иначе на простых не задать. Но как продолжить интебграл на более сложные функции?

Естественный шаг такой. Из предыдущей оценки видно, что если простые функции сходятся равномерно, то их интегралы имеют конечный предел.

3.1.1. Определение. *Пусть функция f является равномерным пределом последовательности простых функций f_n. Положим*

$$\int_X f\,d\mu := \lim_{n\to\infty} \int_X f_n\,d\mu$$

Важно, что интеграл не изменится, если вместо f_n взять другие простые функции g_n, тоже равномерно сходящиеся к f. В самом деле, последовательность $f_1, g_1, f_2, g_2, \ldots$ равномерно сходится, поэтому сходятся и интегралы.

Из этого определения ясно, что интеграл линеен:

$$\int_X (\alpha f + \beta g)\,d\mu = \alpha \int_X f\,d\mu + \beta \int_X g\,d\mu,$$

ибо это верно для простых функций, а если f есть равномерный предел функций f_n, а g есть равномерный предел функций g_n, то функции $\alpha f_n + \beta g_n$ равномерно сходятся к $\alpha f + \beta g$.

Однако возникает вопрос: для каких же ограниченных функций мы сумели определить интеграл? Кроме того, остается вопрос и про неограниченные функции.

Ключевую роль играет концепция измеримой функции.

3.1.2. Определение. *Функция f на пространстве X с σ-алгеброй $\mathcal{A}$ называется измеримой относительно $\mathcal{A}$ или $\mathcal{A}$-измеримой, если*

$$\{x\colon f(x) < c\} \in \mathcal{A} \quad \textit{для всех чисел } c.$$

Здесь очень важен квантор «для всех». Иногда вместо этого говорят «для любого» c, после чего многие сдающие экзамен думают, что достаточно взять какое-то c, какое им любо. Но нет, надо именно для всякого!

Обратим внимание, что в определении измеримости никаких мер нет, только их потенциальные области определения!

В следующем параграфе мы увидим, что равномерные пределы простых функций — в точности все ограниченные измеримые функции. Таким образом, пока у нас определен интеграл Лебега для всякой ограниченной измеримой функции.

Ниже мы проверим, что для измеримой функции f ее положительная и отрицательная части

$$f^+ = \max(f, 0), \quad f^- = \max(-f, 0)$$

тоже измеримы.

Кроме того, для измеримой функции f и всякого числа N измерима срезка $[f]_N$ по уровню N, заданная формулой $[f]_N = \min(f, N)$, т.е. $[f]_N(x) = f(x)$ при $f(x) \leqslant N$, $[f]_N(x) = N$ при $f(x) \geqslant N$.

С помощью срезок интеграл Лебега можно распространить на неограниченные функции так.

3.1.3. Определение. *Неотрицательная $\mathcal{A}$-измеримой функция f считается интегрируемой, если интегралы от ее срезок $[f]_N$ равномерно ограничены по $N \in \mathbb{N}$; в этом случае эти интегралы возрастают и имеют конечный предел, который и называется интегралом функции f и обозначается через $\int_X f\,d\mu$.*

Знакопеременная функция $f = f^+ - f^-$ называется интегрируемой, если таковы f^+ и f^-, а ее интеграл есть разность их интегралов:

$$\int_X f\,d\mu = \int_X f^+\,d\mu - \int_X f^-\,d\mu.$$

Однако можно единообразно ввести интеграл Лебега для $\mathcal{A}$-измеримых функций, не выделяя случай ограниченных функций. Ниже мы проверим, что для ограниченных функций это определение согласуется с предыдущим и равносильно предыдущему определению.

3.1.4. Определение. *Для $\mathcal{A}$-измеримой функции $f \geqslant 0$ положим*

$$\int_X f\,d\mu := \sup_{\varphi} \int_X \varphi\,d\mu,$$

где $\sup$ *берется по всем простым функциям $\varphi \leqslant f$. Функция f называется интегрируемой, если этот* $\sup$ *конечен.*

Знакопеременная функция $f = f^+ - f^-$ называется интегрируемой, если таковы f^+ и f^-, а ее интеграл есть разность их интегралов:

$$\int_X f\,d\mu = \int_X f^+\,d\mu - \int_X f^-\,d\mu.$$

Сразу обратим внимание на то, что в этом определении нет никаких «если что-то от чего-то не зависит»: здесь сразу никто ни от кого не зависит, интеграл неотрицательной функции всегда есть, либо бесконечный, либо конечный (как сумма неотрицательного ряда), в последнем случае функция объявляется интегрируемой. Для знакопеременной функции отдельно рассматриваются две функции, тоже нет никаких «условностей».

Наконец, сделаем еще один технический шаг, в котором в измеримости все же появляется мера.

Будем говорить, что некоторое свойство точек выполнено почти всюду относительно меры μ или μ-п.в., если множество точек без этого свойства входит в $\mathcal{A}_\mu$ и имеет меру нуль.

Например, можно сказать, что функция почти всюду определена или почти всюду неотрицательна.

Подмножество $E \subset X$ можно наделить следовой σ-алгеброй $\mathcal{A}|_E$, состоящей из пересечений $A \cap E$, где $A \in \mathcal{A}$. Тем самым E превращается тоже в измеримое пространство.

3.1.5. Определение. *Будем называть μ-измеримой такие функции f, что f определена вне некоторого множества Z меры нуль и измерима в указанном выше смысле на $X\backslash Z$ относительно $\mathcal{A}_\mu|_{X\backslash Z}$.*

Смысл этого шага — разрешить функции не быть определенной на множестве меры нуль, т.е. рассматривать п.в. определенные функции. В главном определении никаких мер не было, функция должна была быть всюду задана. Теперь это слегка ослабили. При этом не требуется как-то принудительно доопределять не всюду заданную функцию или переопределять, если у нее где-то было бесконечное значение — такие значения не считаем значениями, они возможно лишь на множестве меры нуль. Тем самым всюду равная бесконечности функция у нас не считается функцией.

3.1.6. Определение. *Причислим к интегрируемым такие μ-измеримые функции, которые которые интегрируемы в указанном смысле на своей области определения.*

В качестве задаче следует проверить, что μ-измеримая функция станет $\mathcal{A}_\mu$-измеримой после доопределения ее произвольными числовыми значениями (скажем, нулем) на множестве меры нуль, где она не была определена. Как легко проверить (это следует из нижеследующего) значения интеграла и сам факт его существования от этого не зависят.

Теперь проверим, что если функция f есть равномерный предел простых функций f_n, то общее определение 3.1.4 дает прежнее значение интеграла, равное пределу интегралов от f_n. Достаточно проверить это для $f \geqslant 0$. Общее определение дает не меньшее значение, чем этот предел. С другой стороны для всякого $\varepsilon > 0$ и всякой простой функции $\varphi \leqslant f$ при достаточно большом n имеем $\varphi \leqslant f_n + \varepsilon$, ибо $f_n \geqslant f - \varepsilon$ при достаточно больших n. Тогда для интегралов простых функций получаем

$$\int_X \varphi \, d\mu \leqslant \int_X (f_n + \varepsilon) \, d\mu.$$

Значит, супремум по простым $\varphi \leqslant f$ даст не больше, чем предел интегралов от f_n плюс $\varepsilon\mu(X)$. В силу произвольности ε супремум равен этому пределу.

Из этого следует и равносильность определений 3.1.3 и 3.1.4. Понятно, что проверять это надо для неотрицательных функций f. Если интегралы срезок $[f]_N$ имеют конечный предел I и $\varphi \leqslant f$ — простая функция, то при $N \geqslant \max \varphi$ имеем $\varphi \leqslant [f]_N$, поэтому интеграл от φ не больше I. Значит, и суперемум в определении 3.1.4 не больше I. С другой стороны, как показано выше, для каждого N интеграл от $[f]_N$ есть супремум интегралов от простых функций $\varphi \leqslant [f]_N$, а этот супремум оценивается через супремум в определении 3.1.4, ибо $[f]_N \leqslant f$. Поэтому I не больше последнего супремума.

3.2. Свойства измеримых функций

Этот параграф имеет технический характер — проверяются основные свойства измеримых функций. Даже если считать, что все функции измеримы по Лебегу, то такая проверка необходима, поскольку часто приходится иметь дело с борелевскими функциями или функциями, измеримыми относительно иных сигма-алгебр.

3.2.1. Определение. *Функция f на топологическом пространстве X называется борелевской или измеримой по Борелю, если она измерима относительно борелевской σ-алгебры $\mathcal{B}(X)$, т.е. наименьшей σ-алгебры, содержащей все открытые множества.*

Например, непрерывная функция f измерима по Борелю, ибо множества $\{x\colon f(x) < c\}$ открыты.

Нас будут интересовать борелевские множества на прямой и в $\mathbb{R}^n$.

3.2.2. Лемма. *Функция f измерима относительно σ-алгебры $\mathcal{A}$ в точности тогда, когда $f^{-1}(B) \in \mathcal{A}$ для всякого борелевского множества $B \subset \mathbb{R}$.*

Доказательство. Класс множеств

$$\mathcal{E} = \{E \in \mathcal{B}(\mathbb{R})\colon f^{-1}(E) \in \mathcal{A}\}$$

является σ-алгеброй для какой угодно функции. Проверка — обязательное упражнение. Если f измерима, то $\mathcal{E}$ содержит лучи, а тогда и наименьшую сигма-алгебру, содержащую лучи, а это борелевская сигма-алгебра. Значит, $\mathcal{E} = \mathcal{B}(\mathbb{R})$. Обратное очевидно. $\square$

3.2.3. Теорема. (i) *Для всякой $\mathcal{A}$-измеримой функции f композиция $g \circ f$ тоже $\mathcal{A}$-измерима для всякой борелевской функции g.*

(ii) *Линейные комбинации и произведения $\mathcal{A}$-измеримых функций $\mathcal{A}$-измеримы.*

(iii) *Если $f(x) = \lim\limits_{n\to\infty} f_n(x)$, где f_n $\mathcal{A}$-измеримы, то f тоже $\mathcal{A}$-измерима.*

Доказательство. (i) Имеем $(g \circ f)^{-1}(B) = f^{-1}(g^{-1}(B)) \in \mathcal{A}$ для всякого борелевского B, ибо $g^{-1}(B)$ тоже борелевское, поэтому применима лемма.

(ii) Достаточно проверить, что сумма $\mathcal{A}$-измеримых f и g измерима. Легко проверить (проверьте!), что

$$\{x\colon f(x) + g(x) < c\} = \bigcup_{r\in\mathbb{Q}} \{x\colon f(x) < c - r,\ g(x) < r\}.$$

В правой части стоит счетное объединение пересечений множеств из $\mathcal{A}$. Теперь измеримость fg следует из равенства

$$fg = ((f+g)^2 - f^2 - g^2)/2,$$

где квадраты измеримы в силу (i) с учетом борелевости непрерывной функции $t \mapsto t^2$.

(iii) Аналогично предыдущему (снова нужна проверка!) имеем

$$\{x\colon f(x) < c\} = \bigcup_{k\geqslant 1} \bigcup_{N\geqslant 1} \bigcap_{n\geqslant N} \{x\colon f_n(x) < c - 1/k\},$$

где множество в скобках входит в $\mathcal{A}$. $\square$

Из доказанного вытекает, что и частное f/g измеримых функций измеримо, если $g \neq 0$. Для этого надо проверить лишь измеримость $1/g$, что легко сделать по определению, но можно и сослаться на (i), заметив, что функция $\varphi(t) = 1/t$, $\varphi(0) = 0$ борелева.

Вот еще одно описание измеримых функций.

3.2.4. Следствие. *Функция f является $\mathcal{A}$-измеримой в точности тогда, когда она есть поточечный предел простых функций.*

При этом для ограниченной функции эти простые можно взять равномерно сходящимися, а для неотрицательной поточечно возрастающими к ней.

Доказательство. Приближения строятся банально и явно (что очень важно для интеграла Лебега: делим прямую (или отрезок, в который попадают значения) на промежутки вида $[k2^{-n}, k2^{-n}+2^{-n})$, т.е. $[0, 2^{-n})$ и т.д., и задаем f_n так:

$$f_n(x) = k2^{-n} \quad \text{при} \quad f(x) \in [k2^{-n}, k2^{-n} + 2^{-n}).$$

Множества $f^{-1}([k2^{-n}, k2^{-n} + 2^{-n}))$ входят в $\mathcal{A}$, на них f_n постоянна, поэтому $\mathcal{A}$-измерима (а почему?). Наконец, $|f(x) - f_n(x)| \leqslant 2^{-n}$ для всех x. Для ограниченной функции сразу получены простые функции, а в общем случае возьмем простые g_n так: $g_n(x) = f_n(x)$ при $|f_n(x)| \leqslant 2^n$, $g_n(x) = 0$ при $|f_n(x)| > 2^n$. Проверьте, что получилось то, что нужно. □

В качестве несложных задач предлагается доказать следующее утверждение.

3.2.5. Предложение. *Пусть f_n — $\mathcal{A}$-измеримые функции. Если в каждой точке x последовательность $\{f_n(x)\}$ ограничена, то $\mathcal{A}$-измеримы функции*

$$\inf_n f_n(x), \quad \sup_n f_n(x).$$

Кроме того, множество всех точек, где есть конечный предел чисел $f_n(x)$, входит в $\mathcal{A}$.

В теории интеграла очень важны две следующие именные теоремы.

3.2.6. Теорема. (Теорема Д.Ф. Егорова) *Пусть μ — конечная мера на $\mathcal{A}$ и $f(x) = \lim f_n(x)$ для каждого x, причем f_n $\mathcal{A}$-измеримы. Тогда для всякого $\varepsilon > 0$ найдется такое множество $X_\varepsilon \in \mathcal{A}$, что $\mu(X \backslash X_\varepsilon) < \varepsilon$ и на X_ε сходимость равномерна.*

Доказательство. Перейдя к $f - f_n$, можно считать, что $f = 0$. Тогда для каждого x и каждого $k \in \mathbb{N}$ найдется такой номер N, что $|f_n(x)| \leqslant 2^{-k}$ при всех $n \geqslant N$. Множества

$$X_{k,N} = \{x\colon |f_n(x)| \leqslant 2^{-k} \ \forall\, n \geqslant N\}$$

входят в $\mathcal{A}$ и возрастают к X. Поэтому $\mu(X_{k,N}) \to \mu(X)$, т.е. найдется N_k с $\mu(X\backslash X_{k,N_k}) < \varepsilon 2^{-k}$. Множество

$$X_\varepsilon = X\backslash\Big(X\backslash\bigcup_k X_{k,N_k}\Big) = \bigcap_k X_{k,N_k}$$

входит в $\mathcal{A}$, $\mu(X\backslash X_\varepsilon) < \varepsilon$ и сходимость на нем равномерна. В самом деле, для заданного $\delta > 0$ найдем такое k, что $2^{-k} < \delta$. Так как $X_\varepsilon \subset X_{k,N_k}$, то при $n \geqslant N_k$ имеем $|f_n(x)| \leqslant 2^{-k}$. $\square$

Ясно, что из доказанного сразу следует такая модификация: то же самое верно для μ-измеримых функций, которые сходятся μ-почти всюду.

Закончим этот параграф следующей важной теоремой. Те, кто еще помним, что такое полное сепарабельное метрическое пространство, могут считать, что оно фигурирует вместо $\mathbb{R}$.

3.2.7. Теорема. (Теорема Н.Н. Лузина) *Пусть μ — конечная борелевская мера на прямой. Функция f является μ-измеримой в точности тогда, когда для всякого $\varepsilon > 0$ найдется такой компакт K_ε с $\mu(\mathbb{R}\backslash K_\varepsilon) < \varepsilon$, что сужение f на K_ε непрерывно. При этом найдется непрерывная функция g на $\mathbb{R}$, для которой $\mu(x\colon f(x) \neq g(x)) < \varepsilon$.*

Доказательство. Ясно, что достаточно доказать это для ограниченных функций (взяв гомеоморфизм между прямой и интервалом). Кроме того, достаточно рассматривать f и μ на отрезке $[a,b]$ (сразу взяв отрезок с мерой, близкой к мере прямой). Пусть f $\mathcal{A}$-измерима. Доопределим f нулем на множестве меры нуль, где она могла быть не определена. Как показано выше, есть равномерно сходящаяся к f последовательность простых функций f_n. Каждая функция f_n постоянна на неких дизъюнктных измеримых множествах $A_{n,k} \subset [a,b]$, $k \leqslant N_n$. Впишем в $A_{n,k}$ компакт $K_{n,k}$ так, что μ-мера разности $\bigcup_k A_{n,k} = [a,b]$ и $S_n = \bigcup_k K_{n,k}$ стане меньше $\varepsilon 2^{-n}$. На компакте S_n функция f_n непрерывна. Значит, на их пересечении S непрерывны сразу все функции f_n, что дает и непрерывность их равномерного предела f. При этом $\mu([a,b]\backslash S) \leqslant \sum_n \varepsilon 2^{-n} = \varepsilon$. Наконец, можно продолжить f с S на $\mathbb{R}$

до непрерывной функции (это обычно откуда-то знают, то ли из анализа, то ли из топологии, а кто не знает, должен сделать в качестве упражнения).

Теперь надо еще доказать обратное. Почему f измерима, если есть такие компакты? В этом случае прямая исчерпывается последовательностью таких компактов K_n с точностью до множества меры нуль. Тогда множество $\{x\colon f(x) < c\}$ лишь множеством меры нуль может отличаться от счетного объединения множеств $\{x \in K_n\colon f(x) < c\}$, которые из-за непрерывности f на K_n представляют собой пересечения K_n с открытыми множествами (открытые подмножества пространства K_n есть его пересечения с открытыми на прямой), т.е. это борелевские множества. $\square$

3.3. Свойства интеграла Лебега

Интеграл Лебега был определен сначала для простых функций, а затем расширен двумя равносильными способами, в первом из которых следующий переход делался на ограниченные функции, а во втором на неотрицательные. Из определения интеграла Лебега следует, что две почти всюду равные функции либо одновременно интегрируемы, либо нет, причем их интегралы совпадают в случае интегрируемости. Это достаточно проверить для неотрицательных функций, когда нужное равенство следует из того, что если φ и ψ — простые функции и $\varphi \leqslant \psi$ почти всюду, то интеграл от φ не больше интеграла от ψ.

Из определения сразу следует, что если f и g интегрируемы и $f(x) \leqslant g(x)$, то

$$\int_X f\,d\mu \leqslant \int_X g\,d\mu.$$

Более того, если $0 \leqslant f(x) \leqslant g(x)$, где f и g измеримы и g интегрируема, то определение влечет и интегрируемость f. Из неравенства выше вытекает уже отмечавшаяся оценка для ограниченных функций:

$$\left|\int_X f\,d\mu\right| \leqslant \sup_x |f(x)|\,\mu(X).$$

Кроме того, из этого же вытекает одно из важнейших неравенств теории интеграла:

НЕРАВЕНСТВО ЧЕБЫШЁВА: если f интегрируема, то

$$\mu(x\colon\ |f(x)| \geqslant R) \leqslant \frac{1}{R}\int_X |f|\,d\mu, \quad R > 0.$$

В самом деле, $R\mu(x\colon\ |f(x)| \geqslant R)$ есть интеграл от простой функции g, равной R на множестве $\{x\colon\ |f(x)| \geqslant R\}$ и 0 вне его. При этом $g \leqslant |f|$.

Еще отметим, что если $f \geqslant 0$ имеет нулевой интеграл, то $f(x) = 0$ почти всюду. В самом деле, $\mu(\{f \geqslant 1/n\}) = 0$, ибо f не меньше простой функции $n^{-1}I_{f\geqslant 1/n}$ с интегралом $n^{-1}\mu(\{f \geqslant 1/n\}) > 0$. При этом $\{f > 0\} = \bigcup_n \{f \geqslant 1/n\}$. Конечно, сказанное следует и из неравенства Чебышёва.

Однако при нашем определении интеграла неочевидна его линейность по функции (ниже приведены равносильные определения, при которых линейность более очевидна, а также другие, при которых еще менее очевидна). Линейность будет получена из следующей леммы, которая полезна тем, что при вычислении интеграла позволяет перебирать не все простые функции меньше данной неотрицательной, а взять какую-либо возрастающую к ней последовательность.

3.3.1. Лемма. *Пусть $f \geqslant 0$ интегрируема и f_n — простые, причем $f_n(x)$ возрастает к $f(x)$ почти всюду. Тогда интеграл от f есть предел интегралов от f_n.*

Доказательство. Переопределив на множестве меры нуль, считаем, что сходимость есть всюду. Интегралы от f_n возрастают к некоторому числу I не больше интеграла $I(f)$ от f. Поэтому достаточно проверить, что для всякой простой функции $\varphi \leqslant f$ интеграл не больше I. Функции $\varphi_n = \min(f_n, \varphi)$ являются простыми и возрастают к φ, ибо для тех x, где $\varphi(x) < f(x)$, при всех достаточно больших n имеем $\varphi_n(x) = \varphi(x)$. Покажем, что интегралы от φ_n возрастают к интегралу от φ. Можно считать, что $\mu(X) = 1$ и $0 \leqslant \varphi_n \leqslant \varphi \leqslant 1$. Пусть $\varepsilon > 0$. По теореме Егорова найдется такое измеримое множество X_ε с $\mu(X\backslash X_\varepsilon) < \varepsilon/4$ и $|\varphi(x) - \varphi_n(x)| \leqslant \varepsilon/4$ для всех $x \in X_\varepsilon$ и некоторого n. Для простых функций интеграл линеен, поэтому интеграл по X есть сумма интегралов по X_ε и $X\backslash X_\varepsilon$. Интегралы от φ_n и φ по $X\backslash X_\varepsilon$ не больше $\varepsilon/4$, разность интегралов по X_ε не больше $\varepsilon/4$. В итоге разность интегралов не больше ε. $\square$

3.3.2. Следствие. *Если f и g интегрируемы и неотрицательны, то $f + g$ интегрируема и*

$$\int_X (f+g)\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu.$$

Доказательство. Выше было показано, что f и g есть пределы возрастающих последовательностей простых функций f_n и g_n соответственно, тогда $f+g$ есть предел f_n+g_n, причем такие же соотношения верны и для интегралов. □

Теперь можно доказать линейность интеграла в полном объеме.

3.3.3. Предложение. *Если f и g интегрируемы, то $\alpha f+\beta g$ интегрируема для всех $\alpha, \beta \in \mathbb{R}$ и*

$$\int_X (\alpha f+\beta g)\, d\mu = \alpha \int_X f\, d\mu + \beta \int_X g\, d\mu.$$

Доказательство. Если $g=0$, то утверждение сводится к $f \geqslant 0$ и доказывается так же, как следствие выше. Поэтому можно считать, что $\alpha=\beta=1$. Так как $f=f^+-f^-$, $g=g^+-g^-$, то $f+g=f^++g^+-(f^-+g^-)$. Для f^++g^+ и f^-+g^- все уже доказано, поэтому остается разобрать случай, когда $f \geqslant 0$, $g \leqslant 0$, т.е. случай разности неотрицательных функций. Эта функция интегрируема, ибо $|f-g| \leqslant f+g$. На множестве $Y := \{x\colon f(x) \geqslant g(x)\}$ имеем $(f-g)^+ = f-g$ и $(f-g)^-=0$, на $X\backslash Y$ имеем $(f-g)^- = g-f$ и $(f-g)^+=0$. Значит,

$$\int_X (f-g)^+\, d\mu = \int_X (f-g) I_Y\, d\mu = \int_X f I_Y\, d\mu - \int_X g I_Y\, d\mu,$$

ибо на Y имеем $f = f-g+g$ и $f-g \geqslant 0, g \geqslant 0$. Аналогично

$$\int_X (f-g)^-\, d\mu = \int_X (g-f) I_{X\backslash Y}\, d\mu = \int_X g I_{X\backslash Y}\, d\mu - \int_X f I_{X\backslash Y}\, d\mu.$$

По доказанному

$$\int_X f\, d\mu = \int_X f I_Y\, d\mu + \int_X f I_{X\backslash Y}\, d\mu,$$

$$\int_X g\, d\mu = \int_X g I_Y\, d\mu + \int_X g I_{X\backslash Y}\, d\mu.$$

Из этих соотношений вытекает нужное равенство. □

Для всякого измеримого множества A и всякой интегрируемой функции f положим

$$\int_A f\, d\mu = \int_X f I_A\, d\mu.$$

Из доказанного следует, что это то же самое, что интеграл от f по самостоятельному пространству A с ограниченной на него мерой.

Закончим этот раздел важной теоремой.

3.3.4. Теорема. (Абсолютная непрерывность интеграла Лебега) *Пусть функция f интегрируема по мере μ. Тогда для всякого $\varepsilon > 0$ найдется такое $\delta > 0$, что*

$$\int_A |f|\,d\mu \leqslant \varepsilon \quad \textit{при } \mu(A) \leqslant \delta.$$

Доказательство. Найдем такую простую функцию φ, что $0 \leqslant \varphi \leqslant |f|$ почти всюду и

$$\int_X \varphi\,d\mu \geqslant \int_X |f|\,d\mu - \varepsilon/2.$$

Возьмем $\delta = \varepsilon/(2M+2)$, $M = \max_x \varphi(x)$. При $\mu(A) \leqslant \delta$ имеем

$$\int_A |f|\,d\mu = \int_A \varphi\,d\mu + \int_A (|f| - \varphi)\,d\mu \leqslant M\mu(A) + \int_X (|f| - \varphi)\,d\mu \leqslant \varepsilon,$$

ибо $M\mu(A) \leqslant M\delta \leqslant \varepsilon/2$. □

3.4. Предельный переход в интеграле Лебега

Пусть интегрируемые функции f_n сходятся почти всюду к функции f. Когда f интегрируема и когда ее интеграл равен пределу интегралов от f_n?

Ясно, что без дополнительных условий этой неверно: всякая измеримая функция есть предел простых, но даже если она интегрируема, то ее интеграл может отличаться от предела интегралов f_n и в том случае, когда он есть, а ведь его может и не быть. Скажем, простые функции $f_n = nI_{(0,1/n]}$ на $[0,1]$ сходятся к нулю и имеют интегралы 1 по мере Лебега. Чтобы исправить дело, нужные какие-то оценки. Следующие три базовые теоремы различаются этими оценками.

ВО ВСЕХ ТРЕХ ЕСТЬ ОБЩЕЕ УСЛОВИЕ:

$$f_n(x) \to f(x) \quad \text{почти всюду, } f_n \text{ интегрируемы.}$$

3.4.1. Теорема. (Теорема Лебега о мажорируемой сходимости) *Пусть есть интегрируемая Φ (общая мажоранта), для которой для каждого n*

$$|f_n(x)| \leqslant \Phi(x) \quad \textit{п.в.}.$$

Тогда f интегрируема, причем

$$\int_X f\,d\mu = \lim_{n\to\infty} \int_X f_n\,d\mu, \quad \lim_{n\to\infty} \int_X |f - f_n|\,d\mu = 0.$$

3.4.2. Теорема. (Теорема Беппо Леви о монотонной сходимости) *Пусть $f_n(x) \leqslant f_{n+1}(x)$ п.в. для каждого n. Если*

$$\sup_n \int_X f_n \, d\mu < \infty,$$

то верно заключение предыдущей теоремы.

3.4.3. Теорема. (Теорема Фату) *Пусть $f_n(x) \geqslant 0$ п.в. для каждого n. Если*

$$\sup_n \int_X f_n \, d\mu < \infty,$$

то f интегрируема и

$$\int_X f \, d\mu \leqslant \liminf_{n\to\infty} \int_X f_n \, d\mu,$$

но равенства может не быть.

Заметим, что в условии нет разницы верно ли нечто почти всюду для каждого отдельного n или сразу для всех n, ибо счетное объединение множеств меры нуль тоже имеет меру нуль.

В качестве упражнения полезно вывести каждую из этих теорем из каждой другой. Мы докажем теорему Лебега, а из нее вывод остальных тривиален. Впрочем, и сама теорема Лебега легко доказывается.

Доказательство теоремы Лебега. Модуль разности интегралов не больше интеграла от разности модулей, поэтому достаточно проверить второй утверждение. Оно сводится к случаю $f = 0$ переходом к $f_n - f$ (для них мажоранта 2Φ). Кроме того, можно считать, что $\mu(X) = 1$. Пусть $\varepsilon > 0$. Аболютная непрерывность интеграла дает такое $\delta > 0$, что

$$\int_A \Phi \, d\mu < \varepsilon \quad \text{при } \mu(A) < \delta.$$

Теорема Егорова дает множество X_δ с $\mu(X \backslash X_\delta) < \delta$, на котором сходимость равномерна. Значит, есть такое N, что $|f_n(x)|\varepsilon$ при всех $x \in X_\delta$ и $n \geqslant N$. При таких n интеграл от $|f_n|$ не больше

$$\int_{X_\delta} |f_n| \, d\mu + \int_{X \backslash X_\delta} |f_n| \, d\mu \leqslant \varepsilon + \int_{X \backslash X_\delta} \Phi \, d\mu \leqslant 2\varepsilon,$$

что показывает нужную сходимость. □

Доказательство теоремы Беппо Леви. Можно считать, что $f_n \geqslant 0$, взяв $f_n - f_1$. Тогда $f_n \leqslant f$. Покажем, что f интегрируема, тогда она будет мажорантой из теоремы Лебега. Интегралы от f_n не больше некоторого M. Проверим, что интеграл от f не больше M. Пусть φ — простая, $0 \leqslant \varphi \leqslant f$. Тогда функции $\min(f_n, \varphi)$ возрастают к φ, т.е. φ оказалась их интегрируемой мажорантной, значит, по теореме Лебега их интегралы стремятся к интегралу от φ. Эти интегралы не больше интегралов от f_n, поэтому они не больше M, а тогда интеграл φ не больше M. По определению интеграл от f не больше M. $\square$

Доказательство теоремы Фату. Функции $g_n = \inf_{k \geqslant n} f_n$ измеримы и возрастают почти всюду к f. При этом $g_n \leqslant f_n$, значит, интеграл от g_n не больше интеграла от f_n. По теореме о монотонной сходимости f интегрируема, причем ее интеграл равен пределу интегралов от g_n. Последний не больше $\liminf$ интегралов от f_n. $\square$

3.4.4. Замечание. Докажите в качестве задачи небольшое усиление теоремы Беппо Леви: пусть интегрируемые f_n почти всюду возрастают (возможно, к бесконечности), а их интегралы ограничены общей константой. Тогда их предел почти всюду конечен (что и предполагалось в основной теореме).

Для всякой меры μ (возможно, бесконечной) на $\mathcal{A}$ и всякой μ-интегрируемой функции f возникает новая мера

$$\nu(A) := (f \cdot \mu)(A) := \int_A f\, d\mu.$$

Счетная аддитивность следует из теоремы Лебега: если множества $A_n \in \mathcal{A}$ дизъюнктны и A — их объединение, то $I_A f = \sum_{n=1}^{\infty} I_{A_n} f$, причем $\sum_{n=1}^{N} I_{A_n}|f| \leqslant |f|$, поэтому данный ряд можно проинтегрировать почленно.

Функция f называется плотность (или плотностью Радона – Никодима) меры $\nu = f \cdot \mu$ относительно меры μ и обозначается символом $d\nu/d\mu$.

Как по мере ν узнать, задается ли она плотностью относительно меры μ? Мы рассмотрим случай неотрицательных мер.

3.4.5. Определение. *Мера $\nu \geqslant 0$ на $\mathcal{A}$ называется абсолютно непрерывной относительно меры $\mu \geqslant 0$ на $\mathcal{A}$, если $\nu(A) = 0$ для всякого множества A с $\mu(A) = 0$.*

Ясно, что мера с плотностью относительно μ абсолютно непрерывна относительно μ. Следующая теорема Радона – Никодима утверждает, что верно и обратное.

3.4.6. Теорема. *Если мера ν абсолютно непрерывна относительно μ, то она задается плотностью относительно μ.*

Эта теорема имеет большое значение не только в описанной абстрактной ситуацией (нужной, скажем, для построения условных математических ожиданий в теории вероятностей), но и для обычной меры Лебега на $\mathbb{R}^n$. В этом случае имеется следующий способ явного нахождения плотности Радона – Никодима.

3.4.7. Теорема. *Пусть f — интегрируемая функция на $\mathbb{R}^n$ с мерой Лебега. Тогда для почти всякого x верны равенства*

$$f(x) = \lim_{r\to 0} |B_r(x)|^{-1} \int_{B_r(x)} f(y)\,dy = \lim_{r\to 0} r^{-n} \int_{K_r(x)} f(y)\,dy,$$

где $B_r(x)$ — шар с центром в x радиуса r, $|B_r(x)|$ — его объем, $K_r(x)$ — куб с центром в x и ребрами длины r, параллельными координатным осям.

Таким образом, если ν — мера с плотностью f относительно меры Лебега, то эта плотность вычисляется по формулам

$$f(x) = \lim_{r\to 0} \frac{\nu(B_r(x))}{|B_r(x)|} = \lim_{r\to 0} \frac{\nu(K_r(x))}{r^n}.$$

Следует отметить, что доказательство этой теоремы сложнее, чем доказательство теоремы Радона – Никодима.

3.5. Критерии интегрируемости по Лебегу и связь с интегралом Римана

Удобные критерии интегрируемости по Лебегу формулируются в терминах рядов и интегралов Римана.

Сначала рассмотрим пример.

3.5.1. Пример. Пусть функция f принимает значения c_n на дизъюнктных измеримых множествах A_n. Тогда ее интегрируемость равносильна условию

$$\sum_{n=1}^{\infty} |c_n|\mu(A_n) < \infty.$$

В случае интегрируемости

$$\int_X f\,d\mu = \sum_{n=1}^{\infty} c_n\mu(A_n).$$

Доказательство. По определению все сводится к $c_n \geqslant 0$. Простые функции g_n, равные c_i на A_i при $i \leqslant n$ и 0 на остальных, возрастают к f, их интегралы — частичные суммы ряда, так что остается применить результаты предыдущего параграфа. □

3.5.2. Теорема. *Измеримая функция f интегрируема относительно конечной меры μ в точности тогда, когда*

$$\sum_{n=1}^{\infty} n\mu(x\colon n \leqslant |f(x)| < n+1) < \infty.$$

Это равносильно также условию

$$\sum_{n=1}^{\infty} \mu(x\colon |f(x)| \geqslant n) < \infty$$

или же интегральному условию

$$\int_0^{\infty} \mu(x\colon |f(x)| > t)\, dt < \infty.$$

Доказательство. Пусть $g(x) = n$ при $n \leqslant |f(x)| < n+1$, $n \geqslant 0$. Функция g измерима и $g \leqslant f \leqslant g+1$. Значит, интегрируемость g равносильна интегрируемости f. В силу примера первый ряд есть интеграл от g. Его сходимость равносильна сходимости второго ряда, ибо

$$\mu(|f| \geqslant n) = \mu(n \leqslant |f| < n+1) + \mu(n+1 \leqslant |f| < n+2) + \cdots .$$

Функция $t \mapsto \mu(x\colon |f(x)| > t)$ убывает, ее несобственный интеграл конечен в точности тогда, когда сходится первый ряд. □

Из теории римановского интеграла известно, что функция f на отрезке интегрируема по Риману в точности тогда, когда она ограничена и почти всюду непрерывна.

Однако и без этого факта легко сравнить интегралы Римана и Лебега.

3.5.3. Теорема. (i) *Если f на $[a, b]$ интегрируема по Риману, то она интегрируема и по Лебегу и оба интеграла равны.*

(ii) *Это же верно, если f имеет абсолютный несобственный интеграл Римана. Однако если f имеет несобственный интеграл Римана, но не абсолютный, то она не интегрируема по Лебегу.*

Доказательство. (i) Пусть $[a,b]=[0,1]$. Пусть $n \geqslant 1$. Поделим отрезок точками $k/2^n$ и положим $f_n(x)=\inf_{[k2^{-n},(k+1)2^{-n})} f(x)$, $F_n(x)=\sup_{[k2^{-n},(k+1)2^{-n})} f(x)$. Тогда $f_n \leqslant f \leqslant F_n$, $f_n \leqslant f_{n+1}$, $F_{n+1} \leqslant F_n$. Функции f_n и F_n ступенчатые, их интегралы Римана равны интегралам Лебегам и стремятся к интегралу Римана от f. Функции $g(x)=\lim\limits_{n\to\infty} f_n$ и $G=\lim\limits_{n\to\infty} F_n$ измеримы, причем $g \leqslant f \leqslant F$. Кроме того, интеграл Лебега от g есть предел интегралов (Лебега и Римана) от f_n, а интеграл от G есть предел интегралов от F_n, т.е. g и G имеют равные интегралы. Тогда $g=G$ п.в., откуда $f=g=G$ п.в. и верно равенство для интегралов.

(ii) Если $f \geqslant 0$ имеет несобственный интеграл Римана нуль — единственная особая точка, то функции $f_n = fI_{[1/n,1]}$ возрастаю к f, их интегралы Лебега равны римановским и стремятся к несобственному интегралу от f. По теореме Беппо Леви получаем доказываемое. Случай нескольких особых точек аналогичен. Так как $|f|$ должна быть интегрируемой по Лебегу вместе с f, то неабсолютно несобственно интегрируемая функция не будет интегрируемой по Лебегу (иначе римановские интегралы от $|f|I_{[1/n,1]}$ равномерно ограничены). □

3.6. Произведение мер

Если даны два пространства с мерами $(X,\mathcal{A},\mu)$ и $(Y,\mathcal{B},\nu)$, то в произведении $X\times Y$ возникает класс множеств вида $A\times B$, где $A\in\mathcal{A}, B\in\mathcal{B}$, называемых измеримыми прямоугольниками, которым естественно приписать меру $\mu(A)\nu(B)$, что дает основания рассчитывать, что на $X\times Y$ имеется мера $\mu\otimes\nu$ с

$$\mu\otimes\nu(A\times B)=\mu(A)\nu(B).$$

Однако по нашей доктрине мера $\mu\otimes\nu$ должна быть определена на σ-алгебре, а класс прямоугольников даже не алгебра. Проблема будет решена, если мы сможем разумно продолжить $\mu\otimes\nu$ на сигма-алгебру, порожденную прямоугольниками. Эта сигма-алгебра обозначается символом $\mathcal{A}\otimes\mathcal{B}$. Например, в качестве упражнения полезно проверить, что $\mathcal{B}(\mathbb{R})\otimes\mathcal{B}(\mathbb{R})=\mathcal{B}(\mathbb{R}^2)$. Оказывается, так можно сделать.

Для всякого множества $E\in\mathcal{A}\otimes\mathcal{B}$ положим

$$E_x=\{y\colon (x,y)\in E\},\quad E^y=\{x\colon (x,y)\in E\}.$$

3.6.1. Определение. *Произведение мер $\mu\otimes\nu$ зададим формулой*

$$\mu\otimes\nu(E)=\int_Y \mu(E^y)\,\nu(dy),\quad E\in\mathcal{A}\otimes\mathcal{B}.$$

Разумеется, надо еще проверить, что формула имеет смысл: для всех сечений $E^y \in \mathcal{A}$, так что можно взять $\mu(E^y)$, а полученное будет $\mathcal{B}$-измеримой функцией, что даст возможность проинтегрировать ее по мере ν. Прежде чем это проверять, заметим, что для $E = A \times B$ мы имеем $E^y = A$ при $y \in B$, $E^y = \varnothing$ при $y \notin B$, так что интеграл равен желаемому $\mu(A)\nu(B)$.

Кроме того, как функция от E интеграл счетно-аддитивен, ибо если $E = \cup_n E_n$, где E_n дизъюнктны, то $E^y = \cup_n E_n^y$ при каждом y, причем E_n^y тоже дизъюнктны. Значит,

$$\mu(E^y) = \sum_{n=1}^{\infty} \mu(E_n^y).$$

По теореме Беппо Леви (или по теореме Лебега, здесь обе годятся) интеграл от левой части равен сумме ряд из интегралов от $\mu(E_n^y)$, т.е. $\mu \otimes \nu(E) = \sum_n \mu \otimes \nu(E_n)$.

Теперь займемся проверкой существования интеграла.

3.6.2. Лемма. *Для всякого $E \in \mathcal{A} \otimes \mathcal{B}$ имеем $E^y \in \mathcal{A}$ при всех y.*

Доказательство. Для прямоугольников это уже проверили. Пусть $\mathcal{E}$ — класс всех таких $E \in \mathcal{A} \otimes \mathcal{B}$, что утверждение верно. Этот класс есть сигма-алгебра, ибо операции над множествами дают такие же операции над сечениями (например: $(\cup_n E_n)^y = \cup_n E_n^y$). Значит, $\mathcal{A} \otimes \mathcal{B}$ входит в $\mathcal{E}$, откуда $\mathcal{A} \otimes \mathcal{B} = \mathcal{E}$. $\square$

Со вторым свойством так дешево не разделаться. Для прямоугольников оно тоже тривиально, но вот что делать дальше? Применение меры к сечениям усложняет дело. Выручить может замечательная теорема Серпинского о монотонных классах.

Семейство $\mathcal{E}$ подмножеств множества X называется *монотонным классом*, если $\bigcup_{n=1}^{\infty} E_n \in \mathcal{E}$ для каждой возрастающей последовательности множеств $E_n \in \mathcal{E}$ (т. е. $E_n \subset E_{n+1}$ при всех n) и $\bigcap_{n=1}^{\infty} E_n \in \mathcal{E}$ для каждой убывающей последовательности множеств $E_n \in \mathcal{E}$ (т. е. $E_{n+1} \subset E_n$ при всех n).

Для всякого класса $\mathcal{E}$ подмножеств X существует минимальный монотонный класс, содержащий $\mathcal{E}$ и называемый *монотонным классом, порожденным* $\mathcal{E}$. Таким минимальным классом является пересечение всех монотонных классов, содержащих $\mathcal{E}$.

3.6.3. Теорема. *Пусть $\mathcal{A}$ — алгебра множеств. Тогда σ-алгебра, порожденная $\mathcal{A}$, совпадает с монотонным классом, порожденным $\mathcal{A}$.*

Следовательно, если алгебра $\mathcal{A}$ входит в некоторый монотонный класс $\mathcal{M}$, то $\sigma(\mathcal{A}) \subset \mathcal{M}$.

Доказательство. Пусть $\mathcal{M}(\mathcal{A})$ — монотонный класс, порожденный $\mathcal{A}$. Тогда $\mathcal{M}(\mathcal{A}) \subset \sigma(\mathcal{A})$, ибо $\sigma(\mathcal{A})$ — монотонный класс. Докажем обратное включение. Для этого покажем, что $\mathcal{M}(\mathcal{A})$ есть σ-алгебра. По определению монотонного класса достаточно установить, что $\mathcal{M}(\mathcal{A})$ — алгебра. Докажем сначала, что класс $\mathcal{M}(\mathcal{A})$ замкнут относительно взятия дополнения. Пусть

$$\mathcal{M}_0 := \{B\colon\ B, X\backslash B \in \mathcal{M}(\mathcal{A})\}.$$

Класс $\mathcal{M}_0$ является монотонным, что очевидно из монотонности класса $\mathcal{M}(\mathcal{A})$ и равенств

$$X\backslash \bigcap_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} (X\backslash B_n), \quad X\backslash \bigcup_{n=1}^{\infty} B_n = \bigcap_{n=1}^{\infty} (X\backslash B_n).$$

Поскольку $\mathcal{A} \subset \mathcal{M}_0 \subset \mathcal{M}(\mathcal{A})$, то получаем равенство $\mathcal{M}_0 = \mathcal{M}(\mathcal{A})$.

Проверим теперь замкнутость $\mathcal{M}(\mathcal{A})$ относительно конечных пересечений. Пусть $A \in \mathcal{M}(\mathcal{A})$. Введем класс множеств

$$\mathcal{M}_A := \bigl\{B \in \mathcal{M}(\mathcal{A})\colon\ A \cap B \in \mathcal{M}(\mathcal{A})\bigr\}.$$

Если $B_n \in \mathcal{M}_A$ — возрастающие множества, то получаем

$$A \cap \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} (A \cap B_n) \in \mathcal{M}(\mathcal{A}).$$

Аналогично рассматривается случай, когда B_n убывают. Поэтому $\mathcal{M}_A$ — монотонный класс. Если $A \in \mathcal{A}$, то имеем $\mathcal{A} \subset \mathcal{M}_A \subset \mathcal{M}(\mathcal{A})$, откуда получаем $\mathcal{M}_A = \mathcal{M}(\mathcal{A})$. Пусть теперь $A \in \mathcal{A}$ и $B \in \mathcal{M}(\mathcal{A})$. Тогда по доказанному $A \cap B \in \mathcal{M}(\mathcal{A})$, т. е. $A \in \mathcal{M}_B$. Таким образом, имеем включения $\mathcal{A} \subset \mathcal{M}_B \subset \mathcal{M}(\mathcal{A})$. Следовательно, $\mathcal{M}_B = \mathcal{M}(\mathcal{A})$ при всех $B \in \mathcal{M}(\mathcal{A})$, что означает замкнутость $\mathcal{M}(\mathcal{A})$ относительно взятия пересечения двух множеств. Из доказанного следует, что $\mathcal{M}(\mathcal{A})$ — алгебра, что и требовалось. $\square$

3.6.4. Следствие. *Функция $y \mapsto \mu(E^y)$ является $\mathcal{B}$-измеримой для всякого $E \in \mathcal{A} \otimes \mathcal{B}$.*

Доказательство. Класс всех множеств $E \in \mathcal{A} \otimes \mathcal{B}$ с желаемым свойством является монотонным: если E_n возрастают к E, то сечения E_n^y возрастают к E^y, поэтому $\mu(E_n^y) \to \mu(E^y)$ при каждом y. Остается вспомнить, что предел измеримых функций измерим. Аналогично с убывающими E_n. $\square$

То же рассуждение показывает, что существует и интеграл

$$\int_X \nu(E_x)\,\mu(dx)$$

и задает меру на $\mathcal{A}\otimes\mathcal{B}$. Так как эта мера равна $\mu\otimes\nu$ на прямоугольниках, то обе меры равны на конечных дизъюнктных объединениях прямоугольников. Легко проверить (проверьте!), что класс таких объединений есть алгебра. Тогда меры равны и на порожденной сигма-алгебре, значит, имеем тождество

$$\mu\otimes\nu(E) = \int_Y \mu(E^y)\,\nu(dy) = \int_X \nu(E_x)\,\mu(dx).$$

Из этого тождества следует, что если $\mu\otimes\nu(E) = 0$, то для ν-п.в. y имеем $\mu(E^y) = 0$ и для μ-п.в. x имеем $\nu(E_x) = 0$.

3.6.5. Следствие. (Теорема Фубини для $\mathcal{A}\otimes\mathcal{B}$-измеримых ограниченных функций) *Пусть $\mathcal{A}\otimes\mathcal{B}$-измеримая функция f ограничена. Тогда*

$$\begin{aligned}\int_{X\times Y} f\,d(\mu\otimes\nu) &= \int_Y\int_X f(x,y)\,\mu(dx)\,\nu(dy)\\ &= \int_X\int_Y f(x,y)\,\nu(dy)\,\mu(dx),\end{aligned}$$

где оба повторных интеграла существуют.

Доказательство. Для индикаторов множеств это уже доказано, причем внутренние интегралы измеримы по оставшейся переменной. Значит, это же верно для простых функций. Если простые f_n равномерно сходятся к f, то внутренние интегралы тоже сходятся равномерно по оставшейся переменной. Поэтому получаются ограниченные измеримые (по y и x соответственно) функции. В пределе остается верным и двойное равенство. $\square$

Это следствие еще не охватывает всех даже ограниченных $\mu\otimes\nu$-измеримых функций, даже не все индикаторы множеств из пополнения $\mathcal{A}\otimes\mathcal{B}$. Например, для меры Лебега на квадрате $[0,1]\times[0,1]$ всякое подмножество отрезка вещественной оси имеет меру нуль, но множество Витали не войдет в произведение сигма-алгебр измеримых по Лебегу множеств на отрезках, ибо для множеств из произведения сигма-алгебр выше было показано, что все их сечения входят в сигма-алгебру на сомножителе. Таким образом, даже произведение пополненных сигма-алгебр не обязано быть полным относительно произведения мер, т.е. $(\mathcal{A}\otimes\mathcal{B})_{\mu\otimes\nu}$ может быть шире $\mathcal{A}\otimes\mathcal{B}$. Однако нетрудно распространить доказанное и на функции, измеримые относительно пополненной сигма-алгебры (напомним, что у нас измеримыми относительно меры считаются функции, измеримые относительно пополненной сигма-алгебры).

3.6.6. Теорема. (Теорема Фубини) *Пусть функция f интегрируема относительно меры $\mu\otimes\nu$. Тогда*

$$\begin{aligned}\int_{X\times Y} f\,d(\mu\otimes\nu) &= \int_Y\int_X f(x,y)\,\mu(dx)\,\nu(dy)\\ &= \int_X\int_Y f(x,y)\,\nu(dy)\,\mu(dx),\end{aligned}$$

где в правой части при почти каждом y первый внутренний интеграл по x существует и задает ν-интегрируемую функцию, и аналогично со вторым внутренним интегралом.

Доказательство. Достаточно рассмотреть $f\geqslant 0$. Сначала предположим, что эта функция $\mathcal{A}\otimes\mathcal{B}$-измерима. Для ограниченных f уже все сделано. Стандартные срезки $f_n=\min(f,n)$ возрастают к f. Поэтому их интегралы по $\mu\otimes\nu$ возрастают к интегралу от f. Рассмотрим первый повторный интеграл по μ и ν. При фиксированном y интегралы

$$F_n(y)=\int_X f_n(x,y)\,\mu(dx)$$

возрастают к пределу $F(y)$, равному интегралу от $f(x,y)$ по x, если он конечен, и к бесконечности в противном случае. Такой противный случай в самом деле может представиться для некоторых y, однако множество всех таких y имеет ν-меру нуль в силу теоремы Беппо Леви (со сделанным после нее замечанием), так как интегралы от F_n равномерно ограничены. Из этого же следует, что F интегрируема и ее интеграл по y есть предел интегралов от F_n. Если при каком-то y предел $F(y)$ конечен, то получаем интегрируемость $f(x,y)$ по x при этом y и равенство интеграла числу $F(y)$. Это дает первое равенство в теореме. Второе аналогично.

В общем случае интегрируема функция f всюду равна некоторой $\mathcal{A}\otimes\mathcal{B}$-измеримой функции, для которой все доказано. Поэтому достаточно рассмотреть случай $f=0$ почти всюду. По-прежнему считаем, что $f \geqslant 0$. Пусть $f \leqslant N$. Так как f отлична от нуля на множестве Z меры нуль, то найдется $E \in \mathcal{A} \otimes \mathcal{B}$ с $Z \subset E$. Тогда $Z^y \subset E^y$ при каждом y, значит, достаточно получить равенство $\mu(E^y) = 0$ для ν-п.в. y. Так оно и есть: ведь интеграл от $\mu(E^y)$ по y есть $\mu \otimes \nu(E) = 0$. Таким образом, при каждом N функция $\min(f(x,y), N)$ при почти каждом y имеет нулевой интеграл по x. Поэтому при почти всех y так будет сразу для всех N, что при $N \to \infty$ дает тоже самое и для $f(x,y)$. $\square$

Важно, что без предположения априорной интегрируемости на произведении $X \times Y$ из существования повторных интегралов не следует их равенство, причем даже если они вдруг равны, то f не обязана быть интегрируемой. Наконец, может быть и так, что один повторный интеграл существует, а другой нет. Все эти примеры надо обязательно построить.

Однако есть один специальный случай, когда хватает повторных интегралов.

Задача. Доказать, такую теорему Тонелли:

если неотрицательная функция f измерима относительно $\mu \otimes \nu$, то из существования одного повторного интеграла следует ее $\mu \otimes \nu$-интегрируемость.

Однако без предположения измеримости даже великие итальянцы бессильны (в качестве задачи построить пример, но эта задача сложнее предыдущей).

3.6.7. Пример. Если функции f и g интегрируемы на $\mathbb{R}^n$, то из теорем Тонелли и Фубини следует интегрируемость функции $(x,y) \mapsto f(x)g(y)$ на $\mathbb{R}^n \times \mathbb{R}^n$ и равенство

$$\iint f(x)g(y)\,dx\,dy = \int f(x)\,dx \int g(y)\,dy.$$

По индукции определено конечное произведение мер $\mu_1 \otimes \cdots \otimes \mu_n$. Можно ввести произведение любого набора вероятностных мер. Можно также распространить конструкцию на произведение нескольких σ-конечных мер, т.е. бесконечных мер, сосредоточенных на счетных объединениях множеств конечной меры (как мера Лебега на $\mathbb{R}^n$). В частности, определено произведение $\mu \otimes \lambda$, где λ — мера Лебега на прямой. С помощью этого произведения докажем такой полезный факт.

3.6.8. Предложение. *Пусть $f \geqslant 0$ — μ-интегрируемая функция. Тогда*

$$\int_X f\,d\mu = \int_0^{+\infty} \mu(x\colon f(x) \geqslant t)\,dt = \mu \otimes \lambda(\{(x,t)\colon 0 \leqslant t \leqslant f(x)\}).$$

Доказательство. Подграфик

$$P = \{(x,t)\colon 0 \leqslant t \leqslant f(x)\}$$

является $\mu\otimes\lambda$-измеримым множеством, что следует из измеримости функции $(x,t)\mapsto f(x)-t$ на $X\times\mathbb{R}$, вытекающей из измеримости $(x,t)\mapsto f(x)$ и $(x,t)\mapsto t$. Теперь $\mu\otimes\lambda(P)$ вычислим по теореме Фубини двумя способами. Сначала при фиксированном x замечаем, что P_x есть отрезок $[0,f(x)]$ длины $f(x)$, что дает интеграл от f после интегрирования по x. Затем при фиксированном t видим, что P^t есть множество таких x, что $f(x)\geqslant t$. □

Отметим, что при почти всех t верно равенство

$$\mu(x\colon f(x)\geqslant t) = \mu(x\colon f(x) > t).$$

Это равенство верно при всех t с $\mu(f^{-1}(t)) = 0$, но точек f с прообразом положительной меры может быть лишь счетное число, ибо эти прообразы дизъюнктны (точек с мерой прообраза более $1/k$ лишь конечное число при каждом k).

3.6.9. Пример. Обозначим через V_n объем единичного шара в $\mathbb{R}^n$. Тогда объем шара радиуса R равен V_nR^n. Вычисляя V_n с помощью теоремы Фубини и замечая, что при фиксированном значении $x_n\in[-1,1]$ множество $\{(x_1,\ldots,x_{n-1})\colon\ x_1^2+\cdots+x_{n-1}^2 = 1-x_n^2\}$ является $(n-1)$-мерным шаром радиуса $(1-x_n^2)^{1/2}$, получаем рекуррентное соотношение

$$V_n = V_{n-1}\int_{-1}^{1}(1-x_n^2)^{(n-1)/2}\,dx_n = 2V_{n-1}\int_0^1(1-t^2)^{(n-1)/2}\,dt =$$
$$= V_{n-1}\int_0^1 s^{-1/2}(1-s)^{(n-1)/2}\,ds = V_{n-1}B(1/2,(n+1)/2) =$$
$$= V_{n-1}\frac{\Gamma(1/2)\Gamma((n+1)/2)}{\Gamma(1+n/2)},$$

откуда с учетом равенства $V_2=\pi$ находим ответ

$$V_n = \frac{\pi^{n/2}}{\Gamma(1+n/2)},$$

который можно преобразовать с помощью формулы понижения. Надо отдельно рассмотреть случаи четного и нечетного n, что дает на первый взгляд более «явное» выражение

$$V_{2m} = \frac{\pi^m}{m!},\quad V_{2m+1} = \frac{2(2\pi)^m}{(2m+1)!!},\quad (2m+1)!! = 1\cdot3\cdot5\cdots(2m+1)$$

хотя во многих приложениях формула с гамма-функцией оказывается полезней.

3.7. Неравенства для интегралов

В теории интеграла важную роль играют различные неравенства, среди которых особое место занимают неравенство Чебышёва (уже обсуждавшееся), неравенства Йенсена, Коши–Буняковского, Гёльдера и Минковского, о которых сейчас пойдет речь. Эти пять неравенства надо знать.

Напомним, что функция V на промежутке (a,b) (возможно, неограниченном) называется выпуклой, если

$$V(tx+(1-t)y) \leqslant tV(x)+(1-t)V(y) \quad \forall\, t\in[0,1],\ x,y\in(a,b).$$

Для дважды дифференцируемой функции V выпуклость равносильна неравенству $V'' \geqslant 0$.

По индукции проверяется, что

$$V(t_1x_1+\cdots+t_nx_n) \leqslant t_1V(x_1)+\cdots+t_nV(x_n)$$

при $x_i\in(a,b)$, $t_i\geqslant 0$, $\sum_i t_i = 1$. s

Из выпуклости V вытекает, что для всякого $u_0\in(a,b)$ найдется такое число α, что

$$V(u)\geqslant V(u_0)+\alpha(u-u_0) \quad \forall\, u\in(a,b).$$

В случае дифференцируемой функции можно взять $\alpha = V'(u_0)$.

Основные примеры выпуклых функций: x, $|x|$, $|x|^p$ при $p\geqslant 1$, e^x, e^{-x}, $e^{|x|}$, $x\ln x$ при $x>0$.

3.7.1. Теорема. (Неравенство Йенсена) *Пусть μ — вероятностная мера на $(X,\mathcal{A})$ и f — μ-интегрируемая функция со значениями в интервале U (возможно, неограниченном), на котором задана выпуклая функция V, причем функция $V(f)$ интегрируема. Тогда*

$$V\biggl(\int_X f(x)\,\mu(dx)\biggr) \leqslant \int_X V\bigl(f(x)\bigr)\,\mu(dx). \tag{3.7.1}$$

В частности, если $\exp f$ интегрируема, то

$$\exp\biggl(\int_X f(x)\,\mu(dx)\biggr) \leqslant \int_X \exp f(x)\,\mu(dx),$$

а если $|f|^p$ интегрируема при каком-то $p>1$, то

$$\biggl(\int_X |f|\,d\mu\biggr)^p \leqslant \int_X |f|^p\,d\mu.$$

Доказательство. Для упрощения будем считать, что $U = \mathbb{R}$, в качестве упражнения следует восстановить детали в общей случае. Как указано выше, есть такое число α, что $V(u) \geqslant V(u_0) + \alpha(u - u_0)$ для всех u, где в качестве u_0 берем интеграл от f. Поэтому

$$V\big(f(x)\big) \geqslant V(u_0) + \alpha\big[f(x) - u_0\big].$$

Проинтегрировав это неравенство по x, заметим, что интеграл от второго слагаемого в правой части равен нулю по определению u_0. □

Пусть $(X, \mathcal{A}, \mu)$ — пространство с конечной неотрицательной мерой μ и $p \in [1, \infty)$. Обозначим через $\mathcal{L}^p(\mu)$ множество всех μ-измеримых функций f, для которых $|f|^p$ — μ-интегрируемая функция. В частности, $\mathcal{L}^1(\mu)$ — множество всех μ-интегрируемых функций.

При $1 \leqslant p < \infty$ положим

$$\|f\|_p := \|f\|_{L^p(\mu)} := \left(\int_X |f|^p \, d\mu\right)^{1/p}, \quad f \in \mathcal{L}^p(\mu).$$

3.7.2. Теорема. (Неравенство Гёльдера) *Пусть* $f \in \mathcal{L}^p(\mu)$, $g \in \mathcal{L}^q(\mu)$, *где* $1 < p < \infty$, $q = p(p-1)^{-1}$. *Тогда* $fg \in \mathcal{L}^1(\mu)$ *и* $\|fg\|_1 \leqslant \|f\|_p \|g\|_q$, *т. е.*

$$\int_X |fg| \, d\mu \leqslant \left(\int_X |f|^p \, d\mu\right)^{1/p} \left(\int_X |g|^q \, d\mu\right)^{1/q}. \tag{3.7.2}$$

Доказательство. Функция fg определена п.в. и измерима. Нетрудно показать, что для всех неотрицательных чисел a и b справедливо неравенство $ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}$ (умеете?). Тогда

$$\frac{|f(x)|}{\|f\|_p} \frac{|g(x)|}{\|g\|_q} \leqslant \frac{1}{p} \frac{|f(x)|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g(x)|^q}{\|g\|_q^q}.$$

Правая часть этого неравенства интегрируема, причем ее интеграл равен 1, поэтому левая часть также интегрируема и ее интеграл не превосходит 1, что равносильно (3.7.2). □

Последнее неравенство из теоремы с неравенством Йенсена — частный случай неравенства Гёльдера.

Непосредственным следствием неравенства Гёльдера является приводимое ниже *неравенство Коши–Буняковского*, которое, однако, можно легко доказать непосредственно.

3.7.3. Следствие. *Если* $f, g \in \mathcal{L}^2(\mu)$, *то* $fg \in \mathcal{L}^1(\mu)$ *и*

$$\int_X fg\, d\mu \leqslant \left(\int_X |f|^2\, d\mu\right)^{1/2} \left(\int_X |g|^2\, d\mu\right)^{1/2}. \qquad (3.7.3)$$

3.7.4. Теорема. (Неравенство Минковского) *Предположим, что* $p \in [1, +\infty)$ *и* $f, g \in \mathcal{L}^p(\mu)$. *Тогда* $f + g \in \mathcal{L}^p(\mu)$, *причем*

$$\left(\int_X |f+g|^p\, d\mu\right)^{1/p} \leqslant \left(\int_X |f|^p\, d\mu\right)^{1/p} + \left(\int_X |g|^p\, d\mu\right)^{1/p}. \qquad (3.7.4)$$

Доказательство. Функция $f + g$ определена п.в. и измерима. При $p = 1$ неравенство (3.7.4) очевидно. При $p > 1$ выполнена оценка $|f+g|^p \leqslant 2^p(|f|^p + |g|^p)$, поэтому $|f+g|^p \in \mathcal{L}^1(\mu)$. Заметим, что

$$|f+g|^p \leqslant |f+g|^{p-1}|f| + |f+g|^{p-1}|g|. \qquad (3.7.5)$$

Поскольку $|f+g|^{p-1} \in \mathcal{L}^{p/(p-1)}(\mu) = \mathcal{L}^q(\mu)$, то в силу неравенства Гёльдера имеем

$$\int_X |f+g|^{p-1}|f|\, d\mu \leqslant \left(\int_X |f+g|^p\, d\mu\right)^{1/q} \left(\int_X |f|^p\, d\mu\right)^{1/p}.$$

Оценив аналогичным образом интеграл от второго слагаемого в правой части (3.7.5), приходим к оценке

$$\int_X |f+g|^p\, d\mu \leqslant$$
$$\leqslant \left(\int_X |f+g|^p\, d\mu\right)^{1/q} \left[\left(\int_X |f|^p\, d\mu\right)^{1/p} + \left(\int_X |g|^p\, d\mu\right)^{1/p}\right].$$

Заметив, что $1 - 1/q = 1/p$, получаем $\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$. $\square$

3.8. Связь интеграла с производной и замена переменных

В теории интеграла Римана для непрерывных функций одним из ваднейших фактов является формула Ньютона – Лейбница

$$f(x) - f(a) = \int_a^x f'(t)\, dt$$

для функции с непрерывной производной. При этом для непрерывной функции верно равенство

$$f(x) = \frac{d}{dx}\int_a^x f(t)\, dt.$$

Будет ли первое равенство верно для интеграла Лебега без предположения о непрерывности производной? Как обобщить второе равенство? Конечно, при обобщении сразу ясно, что оно не может сохраниться во всех точках. Даже для функции $f = I_{[0,1]}$ в нуле равенство нарушается. При этом функция $f(x) = |x|$ не имеет производной в нуле, но первая формула для нее верна.

Сначала приведем ответ на первый вопрос.

3.8.1. Теорема. *Пусть функция f на $[a, b]$ непрерывна и на интервале (a, b) имеет интегрируемую по Лебегу производную. Тогда верна формула Ньютона–Лейбница*

$$f(x) - f(a) = \int_a^x f'(t)\,dt.$$

Однако даже если f имеет производную во всех точках прямой, эта производная не обязана быть интегрируемой на каждом отрезке. Вот пример: $f(x) = x^2 \sin(x^{-3})$ при $x \neq 0$, $f(0) = 0$.

Вычислите производную и проверьте, что она не интегрируема по Лебегу на $[0, 1]$.

Поэтому в предыдущей теореме нельзя убрать условие интегрируемости f'. Кроме того, недостаточно вместо дифференцируемости всюду иметь дифференцируемость лишь почти всюду. В качестве примера рассмотрите лестницу Кантора H: это непрерывная возрастающая функция на $[0, 1]$, сначала задаваемая на интервалах дополнения ($1/2$ на средней трети, $1/4$ и $3/4$ на средних третях оставшегося и т.д.). Для нее вообще $H'(x) = 0$ во всех точках дополнения, т.е. почти всюду, но функция ведь непостоянная.

Чтобы разобраться с этими вопросами, полезно ввести два определения.

3.8.2. Определение. *Функция f на $[a, b]$ имеет ограниченную вариацию, если*

$$V_a^b f := \sup \sum_{i=1}^n |f(t_{i+1} - f(t_i)| < \infty,$$

где $\sup$ *взят по всем конечным разбиениям $a = t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n = b$.*

Основной пример: возрастающая функция f, для нее

$$V_a^b f = f(b) - f(a).$$

Нетрудно проверить также, что разность возрастающих функций имеет ограниченную вариацию. Оказывается, не бывает других функций ограниченной вариации. В качестве задачи докажите, что если f имеет ограниченную вариацию, то функции $V(t) = V_a^t f$ и $W(t) = V(t) - f(t)$ — возрастающие. Поэтому $f = V - W$.

Важное значение имеет следующая теорема Лебега.

3.8.3. Теорема. *Пусть f — возрастающая функция на $[a,b]$. Тогда почти всюду существуют производная $f'(x)$, причем она интегрируема и верно неравенство*

$$\int_a^b f'(x)\,dx \leqslant f(b) - f(a).$$

Значит, почти всюду дифференцируема всякая функция ограниченной вариации, причем ее производная интегрируема.

Как показывает пример лестницы Кантора, равенства может не быть.

Введем еще более узкий класс функций.

3.8.4. Определение. *Функция F на $[a,b]$ абсолютно непрерывна, если для каждого $\varepsilon > 0$ найдется такое $\delta > 0$, что для всякого конечного набора попарно непересекающихся интервалов $(a_1, b_1), \ldots, (a_n, b_n)$ в $[a,b]$ с $\sum_{i=1}^n |b_i - a_i| < \delta$ верно неравенство*

$$\sum_{i=1}^n |F(b_i) - F(a_i)| < \varepsilon.$$

Это свойство влечет непрерывность, но сильнее ее (хотя напоминает равномерную непрерывность для тех, кто знал, что это такое). Скажем, если взять кусочно-линейную функцию F так, что $F(n-1) = (-1)^n n^{-1}$, то из-за расходимости гармонического ряда для всякого δ можно набрать конечный набор дизъюнктных интервалов суммарной длины менее δ и суммой модулей приращений более 1.

Важный конструктивный пример абсолютно непрерывной функции — липшицева функция, т.е. функция f, для которой

$$|f(x) - f(y)| \leqslant L|x - y|$$

с некоторым числом L. Для такой функции в качестве δ из определения годится $\varepsilon/(L+1)$.

Заметим, что абсолютно непрерывная функция имеет ограниченную вариацию. В самом деле, взяв δ для $\varepsilon = 1$, разделим $[a,b]$ на N промежутков длины не более δ. Тогда $V_a^b F \leqslant N$. Из предыдущей теоремы следует, что $F'(x)$ существует почти всюду и интегрируема. Это заключение совершенно нетривиально даже для липшицевых функций.

Важнейшее значение в теории интеграла имеет следующая теорема Лебега.

3.8.5. Теорема. *Функция F на $[a,b]$ абсолютно непрерывна в точности тогда, когда найдется такая интегрируемая функция f, что*

$$F(x) = F(a) + \int_a^x f(y)\,dy.$$

При этом $F'(x) = f(x)$ почти всюду.

В качестве упражнения докажите, что если функции f и g абсолютно непрерывны, то такова и fg. Поэтому приходим к формуле интегрирования по частям:

$$\int_a^b fg'\,dx = fg|_a^b - \int_a^b f'g\,dx.$$

Она следует из формулы Ньютона – Лейбница для fg.

С этим результатом связана следующая формула замены переменных в интеграле Лебега.

Пусть φ — строго возрастающая абсолютно непрерывная функция на отрезке $[a,b]$, f — ограниченная борелевская функция на отрезке $[\varphi(a),\varphi(b)]$. Тогда верно равенство

$$\int_a^b f(\varphi(x))\varphi'(x)\,dx = \int_{\varphi(a)}^{\varphi(b)} f(y)\,dy.$$

Доказывается эта формула так. Функция

$$F(y) = \int_{\varphi(a)}^{y} f(t)\,dt.$$

Функция F липшицева, причем $F'(t) = f(y)$ почти всюду. Композиция $F(\varphi)$ тоже абсолютно непрерывна. Кроме того, можно показать, что $(F(\varphi(x))' = f(\varphi(x))\varphi'(x)$ почти всюду. Если функция f непрерывна, то это очевидно, ибо тогда $F'(y) = f(y)$ всюду. Значит, в левой части доказываемой формулы стоит $F(\varphi(b)) - F(\varphi(a))$, что совпадает с правой частью.

Гораздо труднее доказывается следующий многомерный факт. В нем используется замечательная и сама по себе теорема Радемахера:

всякое липшицево отображение $\varphi\colon \mathbb{R}^n \to \mathbb{R}^k$ почти всюду дифференцируемо.

3.8.6. Теорема. *Пусть* $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ *— локально липшицево отображение, инъективное на ограниченном измеримом по Лебегу множестве* $E \subset \mathbb{R}^n$, f *— ограниченная борелевская функция на множестве* $\varphi(E)$. *Тогда множество* $\varphi(E)$ *измеримо по Лебегу и*

$$\int_{\varphi(E)} f(y)\,dy = \int_E f(\varphi(x))|\det \varphi'(x)|\,dx.$$