Mathematical analysis 3. November-December 2023, Harbin

Lecturer Aleksei Savelev

1 Integrals Dependent on a Parameter

1.1 Basic Theorems

When dealing with functions of "two variables", i.e., with functions defined on the direct product of two sets, we need to answer the question: do the operations for different variables commute? We will study this question in the case where one of the operations is integration.

Our goal is to study the properties of an "integral dependent on a parameter", i.e., a function J defined by an equation of the form

$$J(y) = \int_X f(x, y) d\mu(x) \quad (y \in Y)$$

Here μ is a measure defined on a σ-algebra of subsets of a set X, the function $x \mapsto f(x, y)$ is summable on X for every $y \in Y$, and Y is a subset of a metrizable topological space $\widetilde{Y}$. If X is a topological space, then we always assume that the measure μ is defined on all Borel sets (and, consequently, all continuous functions on X are measurable).

First of all, we are interested in the continuity and (in the case where $Y \subset \mathbb{R}^m$) in the differentiability of the function J. Actually, this is a question about the validity of interchanging the integration with respect to the first variable with other analytical operations (passage to the limit, differentiation) with respect to the second variable. We encountered such a situation previously when we discussed the passage to the limit under the integral sign and the index of a function played the role of a parameter. These results will serve as a basis for subsequent reasoning.

It is also natural to study the problem of integration with respect to a parameter. However, there is no need to touch on this here, since to a great extent the problem is solved by Fubini's theorem.

1.1.1

For the first step, we restate three theorems from previous parts for the case of a "continuous parameter". In all three statements, a is a limit point [1] of the set Y in the space $\widetilde{Y}$ and $\varphi(x) = \lim_{y\to a} f(x,y)$, where f and φ are functions (in general, complex-valued) defined on $X \times Y$ and X, respectively. We present conditions under which the following relation holds:

$$J(y) = \int_X f(x,y)d\mu(x) \underset{y\to a}{\longrightarrow} \int_X \varphi(x)d\mu(x) \tag{1}$$

i.e.,

$$\lim_{y\to a} \int_X f(x,y)d\mu(x) = \int_X \left(\lim_{y\to a} f(x,y)\right) d\mu(x)$$

Theorem 1

If $\mu(X) < +\infty$ and the convergence $f(x,y) \underset{y\to a}{\longrightarrow} \varphi(x)$ is uniform with respect to $x \in X$, then the function φ is summable on X and relation (1) holds.

Proof

Since the space $\tilde{Y}$ is metrizable, we can argue "in terms of sequences". We must prove that

$$J\left(y_n\right) = \int_X f\left(x, y_n\right) d\mu(x) \underset{n\to\infty}{\longrightarrow} \int_X \varphi(x)d\mu(x)$$

for every sequence $\{y_n\}$ of points $y_n \in Y\backslash\{a\}, n \in \mathbb{N}$, converging to a. This fact and the summability of φ follows directly from Theorem 1 of "Interchange of Limits and Integration" all conditions of which are fulfilled with $f_n(x) = f\left(x, y_n\right)$.□

1.1.2

For the next step we discuss here modifications of Lebesgue's theorems and the corollary to Vitali's theorem for the case of a "continuous parameter".

Theorem 2

Let $\varphi(x) = \lim_{y\to a} f(x,y)$ for almost all $x \in X$. Assume that there exist a neighborhood U of a and a function $g : X \to \mathbb{R}$ such that the following conditions hold:

$$(L_{loc}) \begin{cases} \text{(a) for almost all } x \in X \text{ and every } y \in (Y \cap U)\backslash\{a\} \\ \qquad \text{the inequality } |f(x,y)| \leqslant g(x) \text{ holds,} \\ \text{(b) the function } g \text{ is summable on } X. \end{cases}$$

Then the function φ is summable on X and relation (1) holds.

[1] In particular, if $\tilde{Y} = [-\infty, +\infty]$, then the cases $a = \pm\infty$ are possible.

Condition (L_{loc}) can be weakened by requiring that the inequality $|f(x,y)| \leqslant g(x)$ be valid for each $y \in Y$ only on a set of full measure possibly depending on y. The above proof of the theorem remains valid for this generalization of condition ($L_{1\ \text{loc}}$). However, in the sequel, (see Theorems 5 and 6) we need the exact formulation of condition (L_{loc}) given in Theorem 2.

Proof As in Theorem 1, we consider the sequence of functions $f_n(x) = f(x, y_n)$, where $y_n \to a, y_n \in (Y \cap U)\backslash\{a\}$ and apply Lebesgue's theorem.□

In the case where $X = \mathbb{N}$ and μ is the counting measure, the integral $\int_X f(x,y)d\mu(x)$ becomes the sum of the (absolutely convergent) series $\sum_{n=1}^{\infty} f(n,y)$, and condition ($L_{\text{loc}}$) coincides with the condition in the Weierstrass M-test for uniform convergence of a series in a neighborhood of a. It follows from Theorem 2 that the limit of the sum can be found termwise.

If $\mu(X) < +\infty$ and the function f is bounded, then condition (L_{loc}) obviously holds for every limit point of Y.□

In the case of finite measure, condition (L_{loc}) can be replaced by a modification of condition (V) of previously discussed **Corollary***

(**Corollary*** *Let $\mu(X) < +\infty$, and let $\{f_n\}_n \geqslant 1$ be a sequence of measurable functions that converges in measure to a function f. If there exist $p > 1$ and $C > 0$ such that*

$$\int_X |f_n|^p \, d\mu \leqslant C \quad \text{for all } n$$

then the functions f_n and f are summable and $\int_X |f_n - f| \, d\mu \underset{n\to\infty}{\longrightarrow} 0$.)

Theorem 3

Let $\mu(X) < +\infty$ and $\varphi(x) = \lim_{y\to a} f(x,y)$ for almost all $x \in X$. If there exists a neighborhood U of a and numbers $s > 1$ and $C > 0$ such that

$$\int_X |f(x,y)|^s d\mu(x) \leqslant C \quad \textit{for all } y \in (Y \cap U)\backslash\{a\}, \qquad (V_{loc})$$

then the function φ is summable on X and relation (1) holds.

Proof

The proof of this theorem is the same as the proof of the preceding one, the only difference being that now we refer to **Corollary*** instead of Lebesgue's theorem.□

In some cases, condition (V_{loc}) is a useful alternative to condition (L_{loc}). For example, if $X = Y$ is a ball in $\mathbb{R}^m$, μ is Lebesgue measure, and $f(x,y) = \frac{1}{\|x-y\|^p}$, where $p < m$, then condition (V_{loc}) (for $1 < s < m/p$) is fulfilled at an arbitrary point $a \in Y$, and, therefore, the function J is continuous on Y. At the same time, condition (L_{loc}) cannot hold at any $a \in Y$, since we have

$$\sup_{y\in U\backslash\{x\}} f(x,y) = +\infty \quad \text{for all } x \in U$$

in each neighborhood U of a.

1.1.3

The following theorem is obviously a special case of Theorem 2 of previous Section.

Theorem 4

If a function f satisfies condition (L_{loc}) at a point $y_0 \in Y$ and is continuous with respect to the second variable at almost all $x \in X$, i.e.,

$$f(x,y) \underset{y\to y_0}{\longrightarrow} f(x,y_0) \qquad \text{for almost all } x \in X \tag{2}$$

then the function J is continuous at y_0 :

$$J(y) = \int_X f(x,y)d\mu(x) \underset{y\to y_0}{\longrightarrow} J(y_0) = \int_X f(x,y_0)\,d\mu(x)$$

We remark that condition (2) is certainly fulfilled if X is a topological space and the function f is continuous on $X \times Y$.

Corollary

If X is a compact space with a finite measure and $Y \subset \mathbb{R}$ is an arbitrary interval, then the continuity of f on $X \times Y$ implies the continuity of the integral $J(y) = \int_X f(x,y)d\mu(x)$ on this interval.

Proof

Indeed, every point of the interval has a relative neighborhood U whose closure is a compact set contained in the interval. By the Weierstrass theorem, the function f is bounded on the product $X \times U$, which guarantees the fulfillment of condition (L_{loc}).□

It is clear that the corollary is valid not only for an interval but for every locally compact space Y; in particular, it is valid if Y is an open or closed subset of a Euclidean space.

1.1.4

We consider two examples. We prove that the functions H and K defined by the equations

$$H(y) = \int_0^\infty \frac{\cos xy}{1+x^2}dx \qquad \text{for } y \in \mathbb{R}$$
$$K(y) = \int_0^\infty e^{-xy}\sin x dx \qquad \text{for } y \in (0,+\infty)$$

are continuous.

In the first case, we have $f(x,y) = \frac{\cos xy}{1+x^2}$. Since

$$\left|\frac{\cos xy}{1+x^2}\right| \leqslant \frac{1}{1+x^2} \qquad \text{for all } x,y \in \mathbb{R} \qquad \text{and} \int_0^\infty \frac{dx}{1+x^2} < +\infty$$

we see that the function f satisfies condition (L_{loc}) at every point $y \in \mathbb{R}$. It remains to refer to Theorem 2.

In the second case, we have $f(x,y) = e^{-xy}\sin x$. In contrast to the preceding example, there is now no majorant g_0 common for all $y \in Y$ and summable on $(0,+\infty)$ for which the inequality $|f(x,y)| \leqslant g_0(x)$ holds for all $x, y > 0$. Nevertheless, condition (L_{loc}) still holds for every point $y \in (0,+\infty)$, but now, for every $y > 0$, we must choose a neighborhood and a summable majorant depending on the neighborhood. Indeed, let $y_0 > 0$ and $U = (y_0/2, +\infty)$. Then

$$\left|e^{-xy}\sin x\right| \leqslant e^{-\frac{xy_0}{2}} \quad \text{for all } x > 0, y \in U, \quad \text{and} \quad \int_0^\infty e^{-\frac{xy_0}{2}}\,dx < +\infty.$$

The second of the above examples can be handled in a different way by direct calculation. Indeed, integrating by parts twice, we obtain

$$\begin{aligned} K(y) &= -\,e^{-xy}\cos x\Big|_0^\infty - y\int_0^\infty e^{-xy}\cos x dx \\ &= 1 - y\left(e^{-xy}\sin x\Big|_0^\infty + y\int_0^\infty e^{-xy}\sin x dx\right) = 1 - y^2K(y) \end{aligned}$$

Consequently, $K(y) = 1/\left(1+y^2\right)$ for every $y > 0$.

The first solution, based on the general scheme, is presented here for two reasons. First, it is typical for such problems. For example, in the same way, we can prove that the integral $\int_0^\infty e^{-xy}h(x)dx$ is continuous with respect to the parameter y for every bounded function h. Secondly, even if we know how to calculate the integral, we must sometimes check condition (L_{loc}) for the integrand (we will discuss this idea later in Ex. 2 Sect 1.1.6).

1.1.5

Theorem 2 allows us easily obtain conditions not only for the continuity but also for the differentiability of an integral depending on a parameter.

Theorem 5

Let $Y \subset \mathbb{R}$ be an arbitrary interval. Assume that:

(a) the derivative

$$f'_y(x,y) = \lim_{h\to 0}\frac{f(x,y+h) - f(x,y)}{h}$$

exists for almost all $x \in X$ and every $y \in Y$;

(b) *the function f'_y satisfies condition (L_{loc}) at a point $y_0 \in Y$.*

Then the function J is differentiable at y_0 and

$$J'(y_0) = \int_X f'_y(x,y_0)\,d\mu(x) \tag{3}$$

This formula is called the Leibniz rule.

Proof Let $x \in X, y_0 + h \in Y, h \neq 0$, and

$$F(x,h) = \frac{f(x, y_0 + h) - f(x, y_0)}{h}.$$

Since

$$\frac{J(y_0 + h) - J(y_0)}{h} = \int_X \frac{f(x, y_0 + h) - f(x, y_0)}{h} d\mu(x) = \int_X F(x,h) d\mu(x), \tag{4}$$

we see that the existence of a finite derivative $J'(y_0)$ and Eq. (3) is immediately obtained by passing to the limit as $h \to 0$ under the integral sign in Eq. (4). We can justify the passage to the limit by **Theorem 3** if we prove that the function F satisfies condition (L_{loc}) at the point $h = 0$. Let us check this. Since the function f'_y satisfies condition (L_{loc}), there exist a positive number δ and a function g summable on X such that

$$\left|f'_y(x,y)\right| \leqslant g(x) \quad \text{for almost all } x \in X \text{ and for } y \in Y, 0 < |y - y_0| < \delta.$$

The Lagrange mean value theorem applied to the function $y \mapsto f(x,y)$ on the interval with endpoints y_0 and $y_0 + h$ gives the relation $F(x,h) = f'_y(x, y_0 + \theta h)$, where θ is a number in the interval $(0,1)$. Therefore, $|F(x,h)| \leqslant g(x)$ for almost all $x \in X$ and $0 < |h| < \delta$. Consequently, condition (L_{loc}) is fulfilled for F.□

Usually, when using the theorem proved above, there is no doubt as to the existence of the partial derivative f'_y and it only remains to check that it satisfies condition (L_{loc}). The situation is even simpler in the case where $X = [p,q]$, $Y = \langle a,b\rangle$, and the functions f and f'_y are continuous in the rectangle $X \times Y$. Then the function $J(y) = \int_p^q f(x,y)dx$ is continuously differentiable on $\langle a,b\rangle$ and $J'(y) = \int_p^q f'_y(x,y)dx$.

Remark Theorem 5 obviously also remains valid in the more general setting where Y is a subset of a multi-dimensional space and the derivative $J'(y)$ is replaced by the partial derivative with respect to one of the coordinates.

1.1.6

Now we consider some applications of the results obtained. First of all, we apply them to calculate two important integrals of functions whose primitives cannot be expressed in terms of elementary functions.

Example 1

Calculate the integral

$$J(y) = \int_0^\infty e^{-x^2} \cos yx dx \quad \text{for } y \in \mathbb{R}$$

By the theorem on differentiation of an integral with respect to a parameter (all conditions of this theorem are obviously met), this is a smooth function and

$$J'(y) = -\int_0^\infty xe^{-x^2}\sin yx dx$$

Integrating by parts, we obtain

$$J'(y) = \frac{1}{2}e^{-x^2}\sin yx\Big|_0^\infty - \frac{y}{2}\int_0^\infty e^{-x^2}\cos yx dx = -\frac{y}{2}\int_0^\infty e^{-x^2}\cos yx dx$$

Therefore, $J'(y) + \frac{y}{2}J(y) = 0$. Consequently, $\left(e^{y^2/4}J(y)\right)' = 0$. Thus, $J(y) = Ce^{-y^2/4}$. Since $C = J(0) = \frac{\sqrt{\pi}}{2}$ (discussed in Integral Theory), we come to the required result,

$$J(y) = \int_0^\infty e^{-x^2}\cos yx dx = \frac{\sqrt{\pi}}{2}e^{-y^2/4}$$

Example 2
We consider the integral

$$J(y) = \int_0^\infty e^{-xy}\frac{\sin x}{x}dx \quad \text{for } y \in (0, +\infty)$$

We prove that the function J is differentiable and use this fact to find $J(y)$. It is clear that, in our case, we have $f'_y(x, y) = -e^{-xy}\sin x$ for all $x, y > 0$. As proved in Sect. 1.1.4, the function f'_y satisfies condition (L_{loc}) at every point of the semiaxis $(0, +\infty)$. Therefore, we may use the Leibniz rule,

$$J'(y) = -\int_0^\infty e^{-xy}\sin x dx \quad \text{for } y > 0$$

The last integral was calculated in **Example 1.4** (we remark that the knowledge of this integral does not spare us the necessity of using condition (L_{loc}) for justification of the above equation). Consequently,

$$J'(y) = -\frac{1}{1+y^2} \quad \text{and} \quad J(y) = C - \arctan y \quad \text{for all } y > 0$$

where C is a constant. To determine the constant, we observe that $J(y) \underset{y\to+\infty}{\longrightarrow} 0$ since $|J(y)| \leqslant \int_0^\infty e^{-xy}dx = \frac{1}{y}$. Therefore, $C = \frac{\pi}{2}$, and thus

$$J(y) = \frac{\pi}{2} - \arctan y \text{ for all } y > 0 \tag{5}$$

Up to now, we have considered the integral $J(y)$ only for $y > 0$. However, the integrand also makes sense for $y = 0$. Moreover, we must know that, although the function $f(x, 0) = \frac{\sin x}{x}$ is not summable, the improper integral $\int_0^\infty \frac{\sin x}{x}dx$ nevertheless converges. Therefore, it is natural to define the integral $J(y)$ also

for $y = 0$ by $J(0) = \int_0^\infty \frac{\sin x}{x} dx$. This naturally raises the question of whether the integral $J(y)$ thus defined is continuous at zero. It is clear that

$$e^{-xy} \frac{\sin x}{x} \underset{y \to 0}{\longrightarrow} \frac{\sin x}{x} \text{ for all } x > 0$$

The justification of the passage to the limit $J(y) \to J(0)$ is complicated by the fact that the integrand $J(0)$ is not summable. Therefore, we cannot use **Theorem 2** here, the conditions of which guarantee the summability of the limiting function. In **Sect. 1.4**, we obtain general theorems allowing us to verify the continuity of an improper integral depending on a parameter, but now we prove that the function J is continuous at zero directly. We verify that the difference

$$J(y) - J(0) = \int_0^\infty \left(e^{-yx} - 1\right) \frac{\sin x}{x} dx$$

tends to zero as $y \to 0$. To this end, we estimate the integral over the intervals $[0, t]$ and $[t, +\infty)$ separately; here $t > 0$ is an auxiliary parameter which will be specified later. The integral over the interval $[0, t]$ can be coarsely estimated: since $0 \leqslant 1 - e^{-yx} \leqslant yx$, we have

$$\left| \int_0^t \left(e^{-yx} - 1\right) \frac{\sin x}{x} dx \right| \leqslant \int_0^t xy \frac{1}{x} dx = yt$$

Integrating by parts in the second integral, we obtain

$$\begin{aligned} \int_t^\infty \left(e^{-yx} - 1\right) \frac{\sin x}{x} dx &= \int_t^\infty \left(e^{-yx} - 1\right) \frac{d(-\cos x)}{x} \\ &= -\left(1 - e^{-ty}\right) \frac{\cos t}{t} + \int_t^\infty \cos x \left(\frac{e^{-yx} - 1}{x}\right)_x' dx \end{aligned}$$

Consequently,

$$\begin{aligned} \left| \int_t^\infty \left(e^{-yx} - 1\right) \frac{\sin x}{x} dx \right| &\leqslant \frac{1}{t} + \int_t^\infty \left(\frac{1}{x^2} + \frac{y}{x} e^{-yx}\right) dx \\ &\leqslant \frac{2}{t} + \frac{y}{t} \int_t^\infty e^{-yx} dx < \frac{3}{t} \end{aligned}$$

Thus, $|J(y) - J(0)| \leqslant yt + \frac{3}{t}$ for all positive y and t. Putting $t = \frac{1}{\sqrt{y}}$, we see that $|J(y) - J(0)| \leqslant 4\sqrt{y}$, which implies the continuity of $J(y)$ as $y \to 0$.

Taking into account (5), we obtain the value of the important integral

$$\int_0^\infty \frac{\sin x}{x} dx = \frac{\pi}{2}$$

1.1.7

Theorem 5 also remains valid in the case of differentiability with respect to a complex parameter.

Theorem 6

Let Y be an open subset of the complex plane. If the conditions:

(a) *the function $y \mapsto f(x,y)$ is holomorphic in Y for almost all $x \in X$;*

(b) *the partial derivative f'_y satisfies condition (L_{loc}) at a point $y_0 \in Y$, are fulfilled, then the integral $J(y) = \int_X f(x,y)d\mu(x)$ is differentiable at y_0 and*

$$J'(y_0) = \int_X f'_y(x, y_0)\, d\mu(x)$$

Proof

The proof of **Theorem 5** can be repeated verbatim, the only difference being that now, in the case where the disk $\bar{B}(y_0, |h|)$ lies in Y, we use the estimate

$$|F(x,h)| = \left| \int_0^1 f'_y(x, y_0 + th)\, dt \right| \leqslant \max_{0 \leqslant t \leqslant 1} \left| f'_y(x, y_0 + th) \right|$$

instead of the Lagrange mean value theorem. By condition (L_{loc}), for h sufficiently small in absolute value, the right-hand side of the above inequality has a summable majorant independent of h. Knowing this, we can finish the proof as in the case of a real parameter.□

It follows from the above theorem that if the function φ is summable on a finite interval $[a,b]$, then the function

$$F(z) = \int_a^b \varphi(t) e^{zt} dt$$

is holomorphic on the entire complex plane. Thus, the Laplace and Fourier transforms of a summable function with compact support, i.e., the integrals

$$\mathcal{L}(z) = \int_{\mathbb{R}_+} \varphi(t) e^{-zt} dt \quad \text{and} \quad \mathcal{F}(z) = \int_{\mathbb{R}} \varphi(t) e^{-izt} dt$$

are entire functions.

Example 1

We find the Laplace transform of a power function. Let $a > 0, z \in \mathbb{C}$, $x = \mathcal{R}e(z) > 0$, and

$$\mathcal{L}(z) = \int_0^\infty t^{a-1} e^{-zt} dt$$

Obviously, $|f(t,z)| = \left|t^{a-1}e^{-zt}\right| = t^{a-1}e^{-xt}$, and, therefore, the integrand is summable for every $z, \mathcal{R}e(z) > 0$. The derivative f'_z satisfies the condition (L_{loc}) at every point in the right half-plane. Therefore,

$$\mathcal{L}'(z) = -\int_0^\infty t^a e^{-zt} dt = \frac{1}{z} t^a e^{-tz} \Big|_{t=0}^{\infty} - \frac{a}{z} \int_0^\infty t^{a-1} e^{-zt} dt = -\frac{a}{z} \mathcal{L}(z)$$

This equation can be represented in the form $(z^a\mathcal{L}(z))' \equiv 0$, which implies that $z^a\mathcal{L}(z) \equiv \text{const}$. We will assume that z^a is the branch of the power function equal to 1 at $z = 1$. Then $\mathcal{L}(z) = \frac{\mathcal{L}(1)}{z^a}$, and it remains to recall the definition of the gamma function to complete the calculation,

$$\mathcal{L}(1) = \int_0^\infty t^{a-1}e^{-t}dt = \Gamma(a)$$

Thus, $\mathcal{L}(z) = \frac{\Gamma(a)}{z^a}$.

Example 2

Let X be a closed subset of the complex plane, let G be the complement of X, and let h be a function summable on X with respect to the measure μ (we recall that according to our agreement at the beginning of the section, a measure on a topological space is defined at least for all Borel subsets). We define a function J on G by the equation

$$J(z) = \int_X \frac{h(\zeta)}{\zeta - z}d\mu(\zeta) \quad (z \in G)$$

The function J is called an integral of Cauchy type.

We verify that this function is holomorphic in G and its derivatives can be calculated by differentiation with respect to the parameter under the integral sign, i.e.,

$$J^{(n)}(z) = n!\int_X \frac{h(\zeta)}{(\zeta - z)^{n+1}}d\mu(\zeta) \quad \text{for all } z \in G, n \in \mathbb{N}$$

In our case, we have $f(\zeta, z) = h(\zeta)/(\zeta - z)$ and $f'_z(\zeta, z) = h(\zeta)/(\zeta - z)^2$ for $\zeta \in X$, $z \in G$. In a neighborhood of z_0, the denominator $\zeta - z$ is separated from zero. Indeed, if the disk $B(z_0, 2r)$ is contained in G, then the inequality $|\zeta - z| \geqslant r$ holds for $|z - z_0| < r$ and $\zeta \in X$. Therefore, the function $\zeta \mapsto f(\zeta, z)$ is summable on X for every $z \in G$ and

$$|f'_z(\zeta, z)| = \left|\frac{h(\zeta)}{(\zeta - z)^2}\right| \leqslant \frac{|h(\zeta)|}{r^2} \quad \text{for all } \zeta \in X, |z - z_0| < r$$

The last estimate shows that the function f'_z satisfies condition (L_{loc}) at z_0. Since z_0 is arbitrary, we obtain by Theorem 1.1.7 that the function J is holomorphic in G and

$$J'(z) = \int_X \frac{h(\zeta)}{(\zeta - z)^2}d\mu(\zeta) \quad \text{for all } z \in G$$

The higher order derivatives are calculated similarly.

1.2 The Gamma function

In the present section, we consider an important example of an integral depending on a parameter. Here we are talking about the gamma function introduced by Euler, or the Euler integral of the second kind, which is of the same fundamental significance as the elementary functions. We have already used the gamma function to calculate the volume of the m-dimensional ball.

1.2.1

We recall that the gamma function is defined for $x > 0$ by the formula

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \tag{1}$$

It is easy to verify that the derivative f'_x of the integrand $f(t, x) = t^{x-1}e^{-t}$ satisfies condition (L_{loc}) in a neighborhood of every point $x_0 > 0$. By Theorem 5, the gamma function is differentiable and

$$\Gamma'(x) = \int_0^\infty t^{x-1} e^{-t} \ln t dt$$

Similarly, we can prove that the gamma function has a derivative of an arbitrary order and find a formula for it. In particular,

$$\Gamma''(x) = \int_0^\infty t^{x-1} e^{-t} \ln^2 t dt > 0.$$

Therefore, the gamma function is a convex function of class $C^\infty((0, +\infty))$.

Integrating by parts, we can easily verify that Γ satisfies the functional equation

$$\Gamma(x + 1) = x\Gamma(x) \quad \text{for } x > 0 \tag{2}$$

We evaluate Γ for positive integers. It is clear that $\Gamma(1) = 1$. By Eq. (2) and induction, we obtain $\Gamma(n + 1) = n!$ for all $n \in \mathbb{N}$. Thus, the gamma function is a continuation of the function $n!$ to the positive real axis (at first sight, the function $n!$ is intimately connected only with positive integers).

By the change of variable $t = u^2$, the integral $\int_0^\infty t^{-1/2} e^{-t} dt = \Gamma(1/2)$ can be reduced to the Euler-Poisson integral $I = \int_{-\infty}^\infty e^{-u^2} du$, which was calculated early. Thus, $\Gamma(1/2) = I = \sqrt{\pi}$. Based on this result and functional equation (2), we can find the values of Γ at half-integers,

$$\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\sqrt{\pi} \quad (n \in \mathbb{N})$$

Equation (2) enables us to study the behavior of Γ in the vicinity of zero,

$$\Gamma(x) = \frac{1}{x}\Gamma(x + 1) \sim \frac{1}{x} \quad \text{as } x \to +0.$$

For large x, the values $\Gamma(x)$ are large, since

$$\Gamma(1+x) = \int_0^\infty t^x e^{-t} dt \geqslant \int_x^\infty t^x e^{-t} dt \geqslant x^x \int_x^\infty e^{-t} dt = \left(\frac{x}{e}\right)^x$$

This simple estimate describes well the growth of Γ at infinity. Later, we obtain the precise asymptotic behavior of $\Gamma(x)$ as $x \to +\infty$.

The functional equation (2) suggests a natural continuation of Γ to the negative semi-axis. Indeed, we should take the formula $\Gamma(x) = \frac{1}{x}\Gamma(x+1)$ as the definition of Γ on the interval $(-1, 0)$. Then the values of Γ on $(-1, 0)$ are negative and the onesided limits at the points 0 and -1 are infinite. Using the definition of Γ on $(-1, 0)$, we can define it on the interval $(-2, -1)$. Proceeding in this way, we define $\Gamma(x)$ for all $x < 0, x \neq -1, -2, \ldots$ We see that $(-1)^n\Gamma(x) > 0$ if $x \in (-n, -n+1)$, and $|\Gamma(x)| \underset{x \to -n}{\longrightarrow} +\infty (n = 1, 2, \ldots)$. Now it is clear that Eq. (2) can be generalized as follows:

$$\Gamma(x+1) = x\Gamma(x) \quad \text{for } x \in \mathbb{R}\backslash\{0, -1, -2, \ldots\} \qquad (2')$$

The properties of the gamma function obtained above allows us to sketch the graph of Γ (see Fig. 1.1). We remark that since $\Gamma(2) = 1 = \Gamma(1)$, Rolle's theorem implies that there is a (unique) critical point of Γ in the interval $(1, 2)$. At this point the function assumes a local minimum. Moreover, every interval $(-n, -n+1)$, $n \in \mathbb{N}$, contains a unique critical point of Γ.

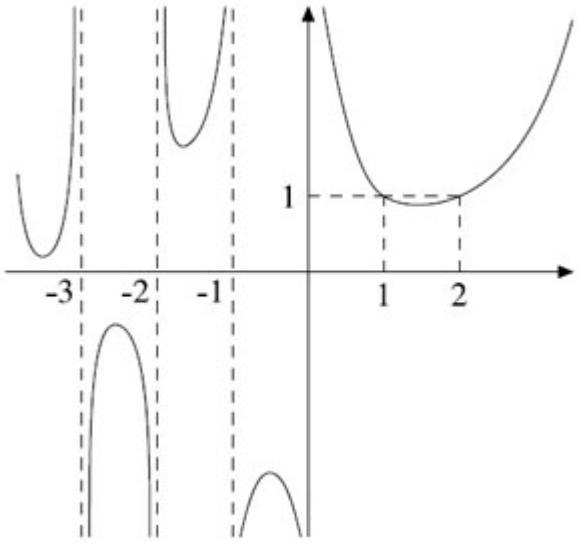

Fig. 1.1 Graph of the gamma function

Replacing x by a complex number z in Eq. (1) (and regarding t^{z-1} as $e^{(z-1)\ln t}$), we see that this equation allows us to define Γ not only at $x > 0$

but also at complex z provided $\mathrm{Re}(z) > 0$, i.e., in the right complex half-plane. It follows from Theorem 6 that Γ is holomorphic in this half-plane. Moreover, the identity $\Gamma(z+1) = z\Gamma(z)$ remains valid and can be used to define Γ in the entire complex plane except at the points $0, -1, -2, \ldots$ in the same way as for Γ on the semi-axis $(-\infty, 0)$. However, we content ourselves with the study of the gamma function only on the real axis.

1.2.2

Now we obtain very important formulas for the gamma function.

First of all, we recall the formula connecting the functions B and Γ. The function B is defined as $\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$, where $x, y > 0$. As proved at last semester, we have $\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, i.e.,

$$\int_0^1 t^{x-1}(1-t)^{y-1}dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{3}$$

From this equation, we derive the following asymptotic relation:

$$\Gamma(x+a) \sim x^a\Gamma(x) \text{ for } x \to +\infty \tag{4}$$

By virtue of the functional equation for Γ it is sufficient to prove this for $a > 0$. By (3), we obtain

$$\frac{\Gamma(x)\Gamma(a)}{\Gamma(x+a)} = \int_0^1 t^{a-1}(1-t)^{x-1}dt \quad (a, x > 0)$$

For convenience, we replace x with $x+1$. Using the change of variables $t = u/x$, we obtain

$$\frac{\Gamma(x+1)\Gamma(a)}{\Gamma(x+a+1)} = \frac{1}{x^a}\int_0^x u^{a-1}\left(1-\frac{u}{x}\right)^x du$$

Since $1 - t \leqslant e^{-t}$, we see that $1 - \frac{u}{x} \leqslant e^{-u/x}$ and $\left(1-\frac{u}{x}\right)^x \leqslant e^{-u}$ for $0 \leqslant u \leqslant x$. Consequently, for every x, the integrand in the last integral (we assume that this function is zero for $u > x$) has the majorant $u^{a-1}e^{-u}$ summable on $(0, +\infty)$. Therefore, Theorem 2 implies

$$x^a\frac{\Gamma(x+1)\Gamma(a)}{\Gamma(x+a+1)} = \int_0^x u^{a-1}\left(1-\frac{u}{x}\right)^x du \underset{x\to+\infty}{\longrightarrow} \int_0^\infty u^{a-1}e^{-u}du = \Gamma(a)$$

Dividing by $\Gamma(a)$, we can represent this in a form equivalent to (4):

$$x^a\frac{x\Gamma(x)}{(x+a)\Gamma(x+a)} \underset{x\to+\infty}{\longrightarrow} 1$$

1.2.3

The following formula makes it possible to find the values of Γ without integration:

$$\Gamma(x) = \lim_{n\to\infty} \frac{n^x n!}{x(x+1)\cdots(x+n-1)(x+n)} \quad \text{for } x \in \mathbb{R}, x \neq 0, -1, -2, \ldots$$

This formula is similar to Euler's definition of Γ and is known as the Euler-Gauss formula.

For the proof we observe that $\Gamma(x+n) = (x+n-1)\cdots(x+1)x\Gamma(x)$. Therefore,

$$\frac{n^x n!}{x(x+1)\cdots(x+n)} = \frac{n}{x+n}\cdot\frac{n^x(n-1)!}{x(x+1)\cdots(x+n-1)} = \frac{n}{x+n}\cdot\Gamma(x)\cdot\frac{n^x\Gamma(n)}{\Gamma(x+n)}.$$

It remains to use relation (4).

For $x = \frac{1}{2}$, the Euler-Gauss formula essentially coincides with the Wallis formula. Indeed, for $x = \frac{1}{2}$ we obtain

$$\sqrt{\pi} = \Gamma\left(\frac{1}{2}\right) = \lim_{n\to\infty} \frac{\sqrt{n}n!}{\frac{1}{2}\cdot\frac{3}{2}\cdots\left(\frac{1}{2}+n\right)} = 2\lim_{n\to\infty}\sqrt{n}\frac{(2n)!!}{(2n+1)!!}$$

which is equivalent to the Wallis formula.

To obtain one more famous formula connected with the gamma function, we recall the asymptotic behavior of the partial sums of the harmonic series: there exists a γ (the Euler constant) such that

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \ln n + \gamma + o(1)$$

This follows from the convergence of the series $\sum_{k=1}^{\infty}\left(\frac{1}{k} - \ln\left(1+\frac{1}{k}\right)\right)$, since its nth partial sum is equal to $1 + \frac{1}{2} + \cdots + \frac{1}{n} - \ln(n+1)$.

We will use this result to obtain a beautiful expansion of the function $1/\Gamma$ in an infinite product. We recall that by the infinite product of a numerical sequence $a_1, a_2, \ldots$, we mean the limit $\lim_{n\to\infty}\prod_{k=1}^{n} a_k$, which is denoted by $\prod_{k=1}^{\infty} a_k$.

We prove that

$$\frac{1}{\Gamma(x)} = xe^{\gamma x}\prod_{k=1}^{\infty}\left(1+\frac{x}{k}\right)e^{-\frac{x}{k}} \quad (x \in \mathbb{R}) \tag{5}$$

(since $|\Gamma(x)| \to +\infty$ as $x \to 0, -1, -2, \ldots$, it is natural to assume that the quotient $1/\Gamma$ is zero at these points). The relation obtained is called the Weierstrass formula.

For the proof, we rewrite the Euler-Gauss formula as

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

Now, after elementary transformations, we obtain

$$\frac{1}{\Gamma(x)} = x \lim_{n\to\infty} n^{-x} \prod_{k=1}^{n} \left(1+\frac{x}{k}\right) = x \lim_{n\to\infty} e^{x\left(1+\frac{1}{2}+\cdots+\frac{1}{n}-\ln n\right)} \prod_{k=1}^{n} \left(1+\frac{x}{k}\right) e^{-\frac{x}{k}}$$

Since $1+\frac{1}{2}+\cdots+\frac{1}{n}-\ln n \to \gamma$, we see that the $\lim_{\lim_{n\to\infty}} \prod_{k=1}^{n} \left(1+\frac{x}{k}\right) e^{-\frac{x}{k}}$ exists and the Weierstrass formula is valid.

1.2.4

Equation (3) enables us to obtain Legendre's (duplication) formula, also simply called the duplication formula:

$$\Gamma(x)\Gamma\left(x+\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^{2x-1}}\Gamma(2x) \quad (x>0)$$

To this end, we transform the right-hand side of the equation

$$\frac{\Gamma^2(x)}{\Gamma(2x)} = \int_0^1 t^{x-1}(1-t)^{x-1}dt$$

We have

$$\frac{\Gamma^2(x)}{\Gamma(2x)} = \int_0^1 \left(t-t^2\right)^{x-1} dt = \int_0^1 \left(\frac{1}{4}-\left(\frac{1}{2}-t\right)^2\right)^{x-1} dt = 2\int_0^{\frac{1}{2}} \left(\frac{1}{4}-s^2\right)^{x-1} ds$$

Substituting $u = 4s^2$, we obtain by (3)

$$\frac{\Gamma^2(x)}{\Gamma(2x)} = 2^{1-2x}\int_0^1 u^{-\frac{1}{2}}(1-u)^{x-1}du = 2^{1-2x}\frac{\Gamma\left(\frac{1}{2}\right)\Gamma(x)}{\Gamma\left(x+\frac{1}{2}\right)}$$

Since $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, we come to the required formula.

As follows from $(2')$, the formula proved above is valid not only for positive x but also for all real x such that $2x \neq 0, -1, -2, \ldots$

1.2.5

Now we obtain one of the most important formulas connected with the gamma function. This is Euler's reflection formula

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x} \quad \text{for } x \in \mathbb{R}\backslash\mathbb{Z}$$

We prove that the product $\theta(x) = \frac{\sin \pi x}{\pi}\Gamma(x)\Gamma(1-x)$ is constant on $\mathbb{R}\backslash\mathbb{Z}$. It follows from Eq. $(2')$ that the function θ has period 1 . Indeed,

$$\theta(x+1) = -\frac{\sin(\pi x)}{\pi}\Gamma(x+1)\Gamma(-x) = -\frac{\sin(\pi x)}{\pi}x\Gamma(x)\frac{\Gamma(1-x)}{-x} = \theta(x).$$

Moreover,

$$\theta(x) = \frac{\sin(\pi x)}{\pi x}\Gamma(x+1)\Gamma(1-x)$$

Hence, extending θ by the formula $\theta(n) = 1$ for $n \in \mathbb{Z}$, we obtain a 1 -periodic function infinitely differentiable in a neighborhood of zero and, consequently, on the entire real axis. It is clear that $\theta > 0$ on $\mathbb{R}$.

For $x > 0$, Legendre's formula implies (as the reader can easily verify) the relation

$$\theta\left(\frac{x}{2}\right)\theta\left(\frac{1+x}{2}\right) = \theta(x)$$

Taking logarithms, we see that

$$g\left(\frac{x}{2}\right) + g\left(\frac{x+1}{2}\right) = g(x), \tag{6}$$

where $g = \ln\theta$. Consequently, the continuous and 1-periodic function g'' satisfies the identity

$$g''\left(\frac{x}{2}\right) + g''\left(\frac{1+x}{2}\right) = 4g''(x).$$

For $M = \max|g''|$, we obtain that $2M \geqslant 4M$. Since $0 \leqslant M < +\infty$, this means that $M = 0$, i.e., $g = \ln\theta$ is a linear function. Taking into account that $g(0) = g(1) = 0$, we obtain $g \equiv 0$, i.e., $\theta \equiv 1$.

The reflection formula can be used to obtain Euler's famous factorization of the sine function into "simple factors" just as polynomials can be represented in a similar form.

Since the sine function has infinitely many zeros, we have to use infinite products. Euler's result is as follows:

$$\sin\pi x = \pi x\prod_{n=1}^{\infty}\left(1 - \frac{x^2}{n^2}\right) \qquad \text{for each } x \in \mathbb{R} \tag{7}$$

Rejecting the trivial case, we may assume that $x \notin \mathbb{Z}$. Multiplying the Weierstrass formulas for $\Gamma(x)$ and $\Gamma(-x)$, we obtain

$$\frac{1}{\Gamma(x)\Gamma(-x)} = -x^2\prod_{n=1}^{\infty}\left(1 - \frac{x^2}{n^2}\right)$$

It remains to apply the reflection formula,

$$\sin\pi x = \frac{\pi}{\Gamma(x)\Gamma(1-x)} = \frac{\pi}{(-x)\Gamma(x)\Gamma(-x)} = \pi x\prod_{n=1}^{\infty}\left(1 - \frac{x^2}{n^2}\right)$$

We remark that, as seen from the above proof, the reflection formula can in turn be derived from the Weierstrass formula and Eq. (7).

1.2.6

Now we turn to a more substantial study of the asymptotic behavior of $\Gamma(x)$ as $x \to +\infty$. The asymptotic behavior is described by Stirling's formula

$$\Gamma(x) \underset{x\to+\infty}{\sim} \sqrt{2\pi} x^{x-\frac{1}{2}} e^{-x} \tag{8}$$

In the Next Section we obtain this result as a particular case of a more general statement, but now we use a different approach based on our knowledge of the gamma function and allowing us to obtain a sharpening of asymptotic formula (8).

First of all, we replace the rapidly decreasing gamma function by its logarithm. The next step is to find the asymptotic behavior of the second derivative of $\ln \Gamma(x)$.

Taking logarithms in Eq. (5) for $x > 0$, we obtain

$$-\ln \Gamma(x) = \ln x + \gamma x + \sum_{n=1}^{\infty} \left(\ln\left(1 + \frac{x}{n}\right) - \frac{x}{n} \right)$$

Differentiating twice, we obtain

$$(\ln \Gamma(x))'' = \sum_{n=0}^{\infty} \frac{1}{(x+n)^2} \tag{9}$$

The termwise differentiation is legal since the series obtained converges uniformly on every closed interval lying in $(0, +\infty)$.

The general method that, in particular, makes it possible to obtain arbitrarily precise asymptotic representation of the sum of series (9) as $x \to +\infty$ is provided by the Euler-Maclaurin formula. However, we will not use it, instead obtaining the first several terms of the asymptotic of $(\ln \Gamma(x))''$ directly. The principal term of the asymptotic can easily be found since the sum of series (9) is close to the integral. We obtain

$$\frac{1}{x} = \int_0^{\infty} \frac{dt}{(x+t)^2} \leqslant \sum_{n=0}^{\infty} \frac{1}{(x+n)^2} \leqslant \frac{1}{x^2} + \int_0^{\infty} \frac{dt}{(x+t)^2} = \frac{1}{x} + \frac{1}{x^2}$$

Thus,

$$(\ln \Gamma(x))'' = \frac{1}{x} + O\left(\frac{1}{x^2}\right)$$

(from here to the end of this section, we assume that $x > 0$ and that the symbol O refers to $x \to +\infty$ without saying it explicitly). The trick that we will use here is as follows. We will successively sharpen the asymptotic formula obtained, extracting the principal parts by series whose sums can easily be found. First, we represent $\frac{1}{x}$ in the form

$$\frac{1}{x} = \sum_{n=0}^{\infty} \left(\frac{1}{x+n} - \frac{1}{x+n+1} \right) = \sum_{n=0}^{\infty} \frac{1}{(x+n)(x+n+1)}$$

and subtract it from (9). We obtain

$$\begin{aligned} (\ln \Gamma(x))'' - \frac{1}{x} &= \sum_{n=0}^{\infty} \left(\frac{1}{(x+n)^2} - \frac{1}{(x+n)(x+n+1)} \right) \\ &= \sum_{n=0}^{\infty} \frac{1}{(x+n)^2(x+n+1)}. \end{aligned} \tag{10}$$

Again, comparing the series obtained with the corresponding integral, we see that

$$\begin{aligned} \frac{1}{2(x+1)^2} = \int_0^{\infty} \frac{dt}{(x+t+1)^3} \leqslant (\ln \Gamma(x))'' - \frac{1}{x} &= \sum_{n=0}^{\infty} \frac{1}{(x+n)^2(x+n+1)} \\ \leqslant \frac{1}{x^3} + \int_0^{\infty} \frac{dt}{(x+t)^3} &= \frac{1}{2x^2} + \frac{1}{x^3} \end{aligned}$$

Consequently,

$$(\ln \Gamma(x))'' - \frac{1}{x} = \frac{1}{2x^2} + h(x), \text{ where } h(x) = O\left(\frac{1}{x^3}\right) \tag{11}$$

This result is already sufficient to prove (8). Indeed, it is clear that

$$\int_1^x h(t)dt = \int_1^{\infty} h(t)dt - \int_x^{\infty} h(t)dt = \text{const} + O\left(\frac{1}{x^2}\right)$$

Therefore, integrating expansion (11) from 1 to x, we obtain the equation

$$(\ln \Gamma(x))' = A + \ln x - \frac{1}{2x} + O\left(\frac{1}{x^2}\right)$$

One more integration gives the relation

$$\ln \Gamma(x) = B + Ax + x \ln x - x - \frac{1}{2} \ln x + O\left(\frac{1}{x}\right).$$

To find A and B, it is convenient to write this equation (slightly coarsening it) as the equivalence

$$\Gamma(x) \underset{x \to +\infty}{\sim} C x^{x - \frac{1}{2}} e^{(A-1)x}$$

where $C = e^B$. To determine A, we use the functional equation, which implies that

$$\Gamma(x) = \frac{\Gamma(x+1)}{x} \underset{x\to+\infty}{\sim} \frac{C}{x}(x+1)^{x+\frac{1}{2}} e^{(A-1)(x+1)}.$$

Taking the ratio of the right-hand sides of these equivalencies, we obtain

$$\left(1+\frac{1}{x}\right)^{x+\frac{1}{2}} e^{A-1} \underset{x\to+\infty}{\longrightarrow} 1$$

which is possible only if $A = 0$. The constant B can be found similarly by means of Legendre's formula, which implies that

$$C^2 x^{x-\frac{1}{2}} e^{-x} \left(x+\frac{1}{2}\right)^{x} e^{-x-\frac{1}{2}} \underset{x\to+\infty}{\sim} \frac{\sqrt{\pi}}{2^{2x-1}} C (2x)^{2x-\frac{1}{2}} e^{-2x}.$$

Dividing by $Cx^{2x-\frac{1}{2}}e^{-2x}$, we see that $C\left(1+\frac{1}{2x}\right)^{x} e^{-\frac{1}{2}} \underset{x\to+\infty}{\longrightarrow} \sqrt{2\pi}$, which implies the equality $C = \sqrt{2\pi}$. Thus,

$$\ln\Gamma(x) = \left(x-\frac{1}{2}\right)\ln x - x + \frac{1}{2}\ln(2\pi) + O\left(\frac{1}{x}\right),$$

i.e.,

$$\Gamma(x) = \sqrt{2\pi} x^{x-\frac{1}{2}} e^{-x}\left(1+O\left(\frac{1}{x}\right)\right). \qquad (8')$$

The above relations, as well as Eq. (8), are also called Stirling's formulas.
To sharpen the asymptotic, we represent $\frac{1}{x^2}$ in the form

$$\frac{1}{x^2} = \sum_{n=0}^{\infty}\left(\frac{1}{(x+n)^2} - \frac{1}{(x+n+1)^2}\right) = \sum_{n=0}^{\infty}\frac{2(x+n)+1}{(x+n)^2(x+n+1)^2}$$

and, multiplying by $\frac{1}{2}$, we subtract it from (10). We obtain

$$(\ln\Gamma(x))'' - \frac{1}{x} - \frac{1}{2x^2} = \frac{1}{2}\sum_{n=0}^{\infty}\frac{1}{(x+n)^2(x+n+1)^2} \qquad (12)$$

Since

$$\begin{aligned}\frac{1}{6(x+1)^3} &= \frac{1}{2}\int_0^{\infty}\frac{dt}{(x+t+1)^4}\\ &\leqslant (\ln\Gamma(x))'' - \frac{1}{x} - \frac{1}{2x^2} = \frac{1}{2}\sum_{n=0}^{\infty}\frac{1}{(x+n)^2(x+n+1)^2}\\ &\leqslant \frac{1}{2x^4} + \frac{1}{2}\int_0^{\infty}\frac{dt}{(x+t)^4} = \frac{1}{2x^4} + \frac{1}{6x^3},\end{aligned}$$

we see that

$$(\ln\Gamma(x))'' - \frac{1}{x} - \frac{1}{2x^2} = \frac{1}{6x^3} + O\left(\frac{1}{x^4}\right)$$

The further sharpening of the asymptotic can be performed repeatedly, but we make only one more step. Applying the trick used twice, we represent $\frac{1}{x^3}$ in the form

$$\frac{1}{x^3} = \sum_{n=0}^{\infty}\left(\frac{1}{(x+n)^3} - \frac{1}{(x+n+1)^3}\right) = \sum_{n=0}^{\infty}\frac{3(x+n)^2+3(x+n)+1}{(x+n)^3(x+n+1)^3}$$

Multiplying by $\frac{1}{6}$ and subtracting from (12), we obtain

$$(\ln\Gamma(x))'' - \frac{1}{x} - \frac{1}{2x^2} - \frac{1}{6x^3} = -\frac{1}{6}\sum_{n=0}^{\infty}\frac{1}{(x+n)^3(x+n+1)^3} \equiv -\frac{1}{6}s(x),$$

where

$$\begin{aligned}\frac{1}{5(x+1)^5} = \int_0^{\infty}\frac{dt}{(x+t+1)^6} \leqslant s(x) &= \sum_{n=0}^{\infty}\frac{1}{(x+n)^3(x+n+1)^3}\\ &< \frac{1}{x^6} + \int_0^{\infty}\frac{dt}{(x+t)^6} = \frac{1}{5x^5} + \frac{1}{x^6}.\end{aligned}$$

It can easily be verified that $\frac{1}{5x^5} - \frac{1}{x^6} < \frac{1}{5(x+1)^5}$. Therefore, $\left|s(x) - \frac{1}{5x^5}\right| < \frac{1}{x^6}$. Thus,

$$(\ln\Gamma(x))'' = \frac{1}{x} + \frac{1}{2x^2} + \frac{1}{6x^3} - \frac{1}{30x^5} + \frac{\theta}{6x^6}, \quad |\theta| < 1$$

After integration we obtain the following sharpening of formula $(8')$:

$$\Gamma(x) = \sqrt{2\pi}x^{x-\frac{1}{2}}e^{-x}e^{\frac{1}{12x}-\frac{1}{360x^3}+\frac{\theta}{120x^4}}, \quad |\theta| < 1 \qquad (8'')$$

1.2.7

We generalize Legendre's formula and verify that the relation (the Gauss multiplication theorem)

$$\Gamma(x)\Gamma\left(x+\frac{1}{p}\right)\cdots\Gamma\left(x+\frac{p-1}{p}\right) = \frac{(2\pi)^{\frac{p-1}{2}}}{p^{px-\frac{1}{2}}}\Gamma(px) \quad (px \neq 0, -1, -2, \ldots)$$

is valid for every $p = 2, 3, 4, \ldots$. For the proof, we use the Euler-Gauss formula and represent the left-hand side in the form

$$\prod_{k=0}^{p-1}\Gamma\left(x+\frac{k}{p}\right)=\lim_{n\to\infty}\prod_{k=0}^{p-1}\frac{n!n^{x+\frac{k}{p}}}{\prod_{j=0}^{n}\left(x+\frac{k}{p}+j\right)}=\lim_{n\to\infty}\frac{(n!)^p n^{px+\frac{p-1}{2}}p^{(n+1)p}}{\prod_{k=0}^{p-1}\prod_{j=0}^{n}(px+pj+k)}$$

It can easily be seen that the arising product is equal to the product of factors of the form $px+i$ for $0\leqslant i\leqslant pn+p-1$. Replacing the last $p-1$ factors by the equivalent quantities pn (as $n\to\infty$), we obtain that the product is equivalent to

$$(pn)^{p-1}\prod_{i=0}^{pn}(px+i)$$

Therefore,

$$\begin{aligned}\prod_{k=0}^{p-1}\Gamma\left(x+\frac{k}{p}\right)&=\lim_{n\to\infty}\frac{(n!)^p n^{px+\frac{p-1}{2}}p^{(n+1)p}}{(pn)^{p-1}\prod_{i=0}^{pn}(px+i)}\\&=p^{-px}\lim_{n\to\infty}\frac{(n!)^p p^{np+1}}{(np)!n^{\frac{p-1}{2}}}\cdot\lim_{n\to\infty}\frac{(np)^{px}(np)!}{px(px+1)\cdots(px+pn)}\end{aligned}$$

By the Gauss formula, the second limit is $\Gamma(px)$. It remains to observe that the first limit (independent of x) is equal to $(2\pi)^{\frac{p-1}{2}}\sqrt{p}$. This can easily be proved by Stirling's formula and is left to the reader. Thus, we arrive at the required result.

1.2.8

We now pause to discuss one more property of the gamma function. It will be shown that this property along with functional equation (2) characterizes the gamma function up to a constant factor. We speak of the logarithmic convexity. A positive function f is called logarithmically convex if $\ln f$ is a convex function.

The convexity of $\ln\Gamma$ certainly follows from formula (9) demonstrating that $(\ln\Gamma)''>0$. However, the logarithmic convexity of Γ can be proved directly from the definition of Γ. Indeed, the logarithmic convexity of Γ is obviously equivalent to the fact that $\Gamma(\alpha x+(1-\alpha)y)\leqslant\Gamma^{\alpha}(x)\Gamma^{1-\alpha}(y)$ for all positive x,y and $\alpha\in(0,1)$. The last inequality follows from Hölder's inequality. Indeed,

$$\begin{aligned}\Gamma(\alpha x+(1-\alpha)y)&=\int_0^{\infty}\left(t^{x-1}e^{-t}\right)^{\alpha}\left(t^{y-1}e^{-t}\right)^{1-\alpha}dt\\&\leqslant\left(\int_0^{\infty}t^{x-1}e^{-t}dt\right)^{\alpha}\left(\int_0^{\infty}t^{y-1}e^{-t}dt\right)^{1-\alpha}=\Gamma^{\alpha}(x)\Gamma^{1-\alpha}(y).\end{aligned}$$

The gamma function is not a unique solution of the functional equation $f(x+1)=xf(x)$. For example, other solutions can be obtained by multiplying

the gamma function by 1-periodic functions. Thus, this equation does not determine the gamma function uniquely. The state of affairs changes radically if we seek solutions in the class of logarithmically convex functions. In this class the equation in question has a unique (up to a positive coefficient) solution.

In other words, the following statement is true.

Theorem

If a logarithmically convex function f on $(0,+\infty)$ satisfies the functional equation $f(x+1)=xf(x)$, then $f(x)=f(1)\Gamma(x)$.

Proof

We verify that the quotient f/Γ is constant. To this end, we consider the function $M=\ln(f/\Gamma)$, which is continuous on $(0,+\infty)$ as the difference of two convex functions. Moreover, M is one-sided 1-periodic, i.e., $M(x+1)=M(x)$ for all $x>0$. Assuming that M is not constant, we consider a point $x_0\in(1,2]$ at which M attains its maximum value. In this case, for some $h\in(0,1)$, the second difference $\Delta_h^2M(x)=M(x+h)-2M(x)+M(x-h)$ is negative, $\Delta_h^2M\left(x_0\right)=\delta<0$. At the same time, $\Delta_h^2(\ln f(x))\geqslant 0$ since f is logarithmically convex. However, for each n, the one-sided periodicity implies

$$0\leqslant\Delta_h^2\left(\ln f\left(x_0+n\right)\right)=\Delta_h^2M\left(x_0+n\right)+\Delta_h^2\left(\ln\Gamma\left(x_0+n\right)\right)=$$
$$=\delta+\Delta_h^2\left(\ln\Gamma\left(x_0+n\right)\right).$$

It follows from (4) that $\Delta_h^2(\ln\Gamma(x))\to 0$ for $x\to+\infty$. Therefore, passing to the limit as $n\to\infty$ in the inequality $0\leqslant\delta+\Delta_h^2\left(\ln\Gamma\left(x_0+n\right)\right)$, we obtain $0\leqslant\delta<0$, a contradiction.□

1.3 Existence Conditions and Basic Properties of Convolution

We will assume that all functions considered in the present section are, in general, complex-valued and measurable on $\mathbb{R}^m$ (in the wide sense), and a measure will mean Lebesgue measure. As before, let $B(r)$ be the ball of radius r with center at the origin.

1.3.1

We introduce the main concept to which this Section is devoted.

Definition Let f and g be functions measurable on $\mathbb{R}^m$. If

$$\int_{\mathbb{R}^m} |f(x-y)g(y)|dy < +\infty \quad \text{for almost all } x \in \mathbb{R}^m \tag{1}$$

then the function h defined almost everywhere by the equation

$$h(x) = \int_{\mathbb{R}^m} f(x-y)g(y)dy \tag{2}$$

is called the convolution of f and g and is denoted by $f * g$.

Condition (1) will be called the convolution existence condition. By the change of variable $y \to z = x - y$, we can easily verify that the above condition is equivalent to the condition

$$\int_{\mathbb{R}^m} |f(z)g(x-z)|dz < +\infty \quad \text{for almost all } x \in \mathbb{R}^m$$

in which case equation (2) implies $h(x) = \int_{\mathbb{R}^m} f(z)g(x-z)dz$. Therefore, the convolutions $f * g$ and $g * f$ exist simultaneously and are equal. Thus, convolution is commutative, i.e., $f * g = g * f$ (if at least one of the convolutions exists).

We see that the properties of convolution are similar to those of multiplication of numbers. The convolution is not only commutative, but, obviously, also distributive, i.e., $f * (g_1 + g_2) = f * g_1 + f * g_2$. Without going deeply into this analogy (see Exercise 1), we will use the terminology invoked by this association. In particular, we call the functions f and g the convolution factors.

We also mention that convolution commutes with a shift: if f_h is a shift of f, i.e., $f_h(x) = f(x-h)$, then it follows directly from the definition of convolution that $(f * g)_h = f_h * g = f * g_h$.

Besides pure mathematical questions (among them, as we will see in the next section, are approximation problems) the concept of convolution has its origins in applied problems. For example, convolution arises as a natural mathematical model of a real device that transforms incoming signals. Let us discuss it in more detail. Suppose we have a device ("black box") reacting to signals, which will be regarded as functions of time with compact support. It is natural to assume that the reaction of the device (its "response") to the signal f_h coming with a delay h differs from its response to the signal f only in the corresponding

delay in time. In other words, the transformation performed by the device that takes an incoming signal f to its response $\widetilde{f}$ commutes with the shift in time: $\widetilde{f}_h = (\widetilde{f})_h$. Furthermore, we assume that the device takes a linear combination of signals to a linear combination of the responses. The main characteristic of such a device (or, as is often said, the system function) is its reaction to a pulse action δ_α, which can be regarded as a function with unit integral (the "pulse energy") constant on a very small interval $\Delta_\alpha = [0, \alpha)$ and equal zero outside it. In other words, $\delta_\alpha = \frac{1}{\alpha}\chi_\alpha$, where χ_α is the characteristic function of the interval Δ_α. For sufficiently small α, the reaction of the device to the signals δ_α does not practically depend on α. Therefore, replacing δ_α by the "limit function", we can regard an instantaneous unit pulse as the Dirac function δ, which has the following properties:

$$\delta(t) = 0 \quad \text{if } t \neq 0, \quad \delta(0) = +\infty, \quad \int_{-\infty}^{\infty} \delta(t)dt = 1$$

The response E to a signal $\delta \approx \delta_\alpha$ is called the system function of the device. Representing an arbitrary signal f as a linear combination of step functions constant on the intervals $[n\alpha, (n+1)\alpha)$ with required accuracy, we obtain that

$$f(t) \approx \sum_n f(n\alpha)\chi_\alpha(t - n\alpha) \approx \alpha \sum_n f(n\alpha)\delta_\alpha(t - n\alpha)$$

Because the transformation performed by the device is linear and commutes with a shift, we obtain

$$\tilde{f}(s) \approx \alpha \sum_n f(n\alpha)E(s - n\alpha)$$

This sum is nothing but an integral sum for the integral

$$\int_{-\infty}^{\infty} f(t)E(s - t)dt$$

Taking into account the fact that the above approximation becomes arbitrarily accurate for sufficiently small α, we may assume that $\widetilde{f}(s) = \int_{-\infty}^{\infty} f(u)E(s - u)du$. Thus, the response of the device to a signal f coincides with the convolution of f and the system function of the device. For that reason convolution is of essential importance in the theoretical foundations of optics and radio engineering.

1.3.2

First, we establish an auxiliary statement.

Lemma *If f and g are measurable functions on $\mathbb{R}^m$ satisfying condition (1), then their convolution $f * g$ is also measurable on $\mathbb{R}^m$.*

Proof It is sufficient to proof the theorem for real-valued functions. In this case, we can use Lemma (**Remind** Let f be a measurable function defined on $\mathbb{R}^m$. Then the functions $(x, y) \mapsto f(x - y)$ and $(x, y) \mapsto f(x + y)$ are measurable on the space $\mathbb{R}^{2m}$.), which implies that the integrand in Eq. (2) is not only

measurable for almost all $x \in \mathbb{R}^m$ as a function of y, but also measurable with respect to the "totality" of the variables x and y (i.e., the function $(x, y) \mapsto F(x, y) = f(x - y)g(y)$ is measurable on $\mathbb{R}^m \times \mathbb{R}^m$). Therefore, to prove the lemma, it remains to refer to Corollary 2 of Tonelli's theorem.□

(**Remind! Corollary 2** If the function f is measurable on $X \times Y$, then:

(1) for almost all $x \in X$, the function f_x is measurable on Y;

(2) if $\int_Y |f_x(y)|\, d\nu(y) < +\infty$ for almost all $x \in X$, then the function $x \mapsto \int_Y f(x, y)dv(y)$ is measurable on X in the wide sense.).

Theorem *The convolution of functions f and g summable on $\mathbb{R}^m$ is defined almost everywhere on $\mathbb{R}^m$, is summable, and satisfies the inequality*

$$\int_{\mathbb{R}^m} |(f * g)(x)|dx \leqslant \int_{\mathbb{R}^m} |f(x)|dx \int_{\mathbb{R}^m} |g(x)|dx.$$

Proof We put $H(x) = \int_{\mathbb{R}^m} |f(x - y)g(y)|dy$. Since (**Remind** Let f be a measurable function defined on $\mathbb{R}^m$. Then the functions $(x, y) \mapsto f(x - y)$ and $(x, y) \mapsto f(x + y)$ are measurable on the space $\mathbb{R}^{2m}$.) the function $(x, y) \mapsto f(x - y)g(y)$ is measurable on $\mathbb{R}^m \times \mathbb{R}^m$, it follows from Tonelli's theorem that

$$\int_{\mathbb{R}^m} H(x)dx = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^m} |f(x - y)|dx \right) |g(y)|dy.$$

The change of variable $x \mapsto x - y$ shows that the inner integral is equal to $\int_{\mathbb{R}^m} |f(x)|dx$ for every y. Therefore,

$$\int_{\mathbb{R}^m} H(x)dx = \int_{\mathbb{R}^m} |f(x)|dx \int_{\mathbb{R}^m} |g(y)|dy < +\infty$$

Consequently, the function H summable and, therefore, $H(x) < +\infty$ almost everywhere. Thus, condition (1) is fulfilled, and the convolution $(f*g)(x)$ exists. Its measurability is established in the lemma, and the summability follows from the inequality $|(f * g)(x)| \leqslant H(x)$. Moreover,

$$\int_{\mathbb{R}^m} |(f * g)(x)|dx \leqslant \int_{\mathbb{R}^m} H(x)dx = \int_{\mathbb{R}^m} |f(x)|dx \int_{\mathbb{R}^m} |g(y)|dy$$

□

1.3.3

We supplement Theorem 1.3.2 and obtain alternative conditions sufficient for the existence of a convolution. We consider a wider class of functions than $\mathscr{L}(\mathbb{R}^m)$, the class of measurable functions, which is frequently encountered in function theory and in other branches of mathematics.

Definition *A measurable function f in $\mathbb{R}^m$ is called locally summable in $\mathbb{R}^m$ if it is summable on every bounded set, i.e., if*

$$\int_{B(R)} |f(x)|dx < +\infty \quad \text{for every } R > 0$$

The set of all functions locally summable in $\mathbb{R}^m$ will be denoted by $\mathscr{L}_{\text{loc}}\ (\mathbb{R}^m)$. Obviously, every locally summable function is almost everywhere finite and $C\,(\mathbb{R}^m) \subset \mathscr{L}_{\text{loc}}\ (\mathbb{R}^m)$.

We remind that the closure of the set $\{x \in \mathbb{R}^m \mid f(x) \neq 0\}$ is called the support of a function f and is denoted by $\operatorname{supp}(f)$. By $A + B$, where $A, B \subset \mathbb{R}^m$, we denote the set $\{a + b \mid a \in A, b \in B\}$.

Theorem *If $f \in \mathscr{L}_{\text{loc}}\,(\mathbb{R}^m)$ and g is a summable function with compact support, then the convolution $f * g$ exists and*

$$\operatorname{supp}(f * g) \subset \operatorname{supp}(f) + \operatorname{supp}(g) \tag{3}$$

Proof Let $\operatorname{supp}(g) \subset B(r)$. As in Theorem 1.3.2, we put

$$H(x) = \int_{\mathbb{R}^m} |f(x-y)g(y)|dy = \int_{B(r)} |f(x-y)g(y)|dy$$

and prove that $H(x) < +\infty$ almost everywhere. For this, we check that $H \in \mathscr{L}_{\text{loc}}\ (\mathbb{R}^m)$, i.e., that $\int_{B(R)} H(x)dx < +\infty$ for every $R > 0$. Indeed, since $\operatorname{supp}(g) \subset B(r)$, we have

$$\begin{aligned}
\int_{B(R)} H(x)dx &= \int_{B(R)} \left(\int_{B(r)} |f(x-y)g(y)|dy \right) dx \\
&= \int_{B(r)} |g(y)| \left(\int_{B(R)} |f(x-y)|dx \right) dy \\
&\leqslant \int_{B(r)} |g(y)| \left(\int_{B(r+R)} |f(u)|du \right) dy \\
&= \int_{B(r+R)} |f(u)|du \cdot \int_{B(r)} |g(y)|dy < +\infty.
\end{aligned}$$

The last inequality is valid since f is locally summable. Thus, the function H is finite almost everywhere in the ball $B(R)$, and, consequently, almost everywhere on $\mathbb{R}^m$. Therefore, condition (1) is fulfilled, and the convolution $f * g$ exists.

To prove inclusion (3), we remark that if $f(x-y)g(y) \neq 0$, then $x - y \in \operatorname{supp}(f)$ and $y \in \operatorname{supp}(g)$, and so, $x = (x-y) + y \in \operatorname{supp}(f) + \operatorname{supp}(g)$. Therefore, $f(x-y)g(y) \equiv 0$ in the case where $x \notin \operatorname{supp}(f) + \operatorname{supp}(g)$. Consequently, $f * g = 0$ outside the set $\operatorname{supp}(f) + \operatorname{supp}(g)$, i.e.,

$$\{x \in \mathbb{R}^m \mid (f * g)(x) \neq 0\} \subset \operatorname{supp}(f) + \operatorname{supp}(g).$$

Since $\operatorname{supp}(g)$ is compact, the set on the right-hand side of this inclusion is closed (we leave it to the reader to prove this independently), which implies that

$$\operatorname{supp}(f * g) = \overline{\{x \in \mathbb{R}^m \mid (f * g)(x) \neq 0\}} \subset \operatorname{supp}(f) + \operatorname{supp}(g)$$

Corollary *The convolution of two summable functions with compact supports has a compact support.*

1.3.4

We now discuss differential properties of convolution. First, we prove an auxiliary result.

Lemma (Truncation lemma) *Let $f, \tilde{f} \in \mathscr{L}_{\text{loc}}(\mathbb{R}^m)$, and let a function g be bounded and satisfy the inclusion $\operatorname{supp}(g) \subseteq B(r)$. If f coincides with $\widetilde{f}$ in the ball $B(R+r)$, then the convolutions $f*g$ and $\widetilde{f}*g$ coincide in the ball $B(R)$.*

Proof Let $\|x\| < R$. Then $\|x-y\| < R+r$ for $\|y\| < r$. Therefore,

$$(f*g)(x) = \int_{B(r)} f(x-y)g(y)dy = \int_{B(r)} \tilde{f}(x-y)g(y)dy = (\tilde{f}*g)(x).$$

Theorem *Let $f \in \mathscr{L}_{\text{loc}}(\mathbb{R}^m)$, and let g be a bounded function with compact support. Then:*

(1) if at least one of the functions f or g is continuous, then the convolution $f*g$ is continuous;

(2) if at least one of the functions f or g is continuously differentiable, then the convolution is continuously differentiable and its derivatives can be calculated by the formula ($k = 1, \dots, m$)

$$\frac{\partial(f*g)}{\partial x_k}(x) = \begin{cases} \left(f * \frac{\partial g}{\partial x_k}\right)(x) & \text{if } g \in C^1(\mathbb{R}^m) \\ \left(\frac{\partial f}{\partial x_k} * g\right)(x) & \text{if } f \in C^1(\mathbb{R}^m) \end{cases} \tag{4}$$

Remark The first assertion of the theorem admits an essential sharpening. As we will see in the sequel, the convolution of a locally summable function and a bounded summable function is continuous without any additional assumptions.

Proof We will assume that $\operatorname{supp}(g) \subset B(r)$.

(1) If g is continuous and f is summable, then the integral $\int_{\mathbb{R}^m} f(y)g(x-y)dy$ is continuous with respect to the parameter by Theorem 4 (section 1.1.3). If f is not summable, then we use the obvious fact that it is sufficient to prove the continuity of the convolution in an arbitrary ball $B(R)$. We can also use the truncation lemma and replace f by a summable function $\widetilde{f}$ that has a compact support and coincides with f on a ball $B(R+r)$. The same method can be applied if f is continuous because, in this case, we may assume that $\widetilde{f}$ is continuous.

(2) Turning to the proof of the smoothness of convolution, we first assume that the function g is smooth. It is obvious that if $x_0 \in \mathbb{R}^m$ and $\|x - x_0\| < 1$, then

$$(f*g)(x) = \int_{\mathbb{R}^m} f(y)g(x-y)dy = \int_{B(x_0, r+1)} f(y)g(x-y)dy$$

Applying the Leibniz rule (see Theorem 1.1.5) to the right-hand side of this equation, we immediately obtain the required result,

$$\frac{\partial(f*g)}{\partial x_k}(x)=\int_{B(x_0,r+1)} f(y)\frac{\partial g}{\partial x_k}(x-y)dy=\left(f*\frac{\partial g}{\partial x_k}\right)(x)$$

We verify that, in the case in question, the application of the Leibniz rule is legal. For this, we must check that the partial derivative $\frac{\partial}{\partial x_k}(f(y)g(x-y))=f(y)\frac{\partial g}{\partial x_k}(x-y)$ satisfies condition (L_{loc}) at x_0.

This is indeed the case because

$$\left|f(y)\frac{\partial g}{\partial x_k}(x-y)\right|\leqslant M|f(y)|\chi_{B(x_0,r+1)}(y)\quad\text{ for all } x\in B\left(x_0,r+1\right),$$

where $M=\max_x\left|\frac{\partial g(x)}{\partial x_k}\right|$.

Now assume that f is continuously differentiable. To prove that the convolution is differentiable on $B(R)$, we should replace f by a function $\widetilde{f}$ that has a compact support and coincides with f on a sufficiently large ball, as we did in the proof of the continuity of convolution, the only difference being that the function $\widetilde{f}$ must now be smooth. For example, we can multiply f by a smooth function that has a compact support and is equal to 1 on a ball $B(R+r)$. Then, interchanging the roles of g and $\widetilde{f}$ and using the formula proved above, we find that

$$\frac{\partial(f*g)}{\partial x_k}(x)=\frac{\partial(g*\tilde{f})}{\partial x_k}(x)=\left(g*\frac{\partial\tilde{f}}{\partial x_k}\right)(x)=\left(g*\frac{\partial f}{\partial x_k}\right)(x)=\left(\frac{\partial f}{\partial x_k}*g\right)(x)$$

for $\|x\|<R$. Since R is arbitrary, this proves the theorem.

Corollary *The convolution of a locally summable function f and a bounded function φ with compact support is infinitely differentiable if at least one of the functions f or φ is infinitely differentiable.*

Proof The assertion should be proved by induction using (4).

In particular, it follows from the corollary that a linear differential operator with constant coefficients commutes with convolution.

1.3.5

The concept of convolution has different generalizations and modifications. We mention some of them.

In the case where the functions in question are periodic on the real line, the convolution is defined in the same way as above with the only difference that now the integral over $\mathbb{R}$ is replaced by the integral over an interval with length equal to the period (no matter which interval is used). For definiteness, we will assume that the period is 2π. It is clear that the convolution of periodic functions is also periodic. We leave it to the reader to verify independently that

an analog of Theorem 1.3.2 is valid for the convolution of periodic functions. One simply repeats the proof of this theorem, changing the domain of integration appropriately.

The above applies in full to functions defined on $\mathbb{R}^m$ and 2π-periodic with respect to each variable. Their convolution is defined by the equation

$$(f * g)(x) = \int_{(-\pi,\pi)^m} f(x-y)g(y)dy$$

One more version of the definition of convolution can be obtained as follows. If a function g is summable and non-negative, then the integral $\int_{\mathbb{R}^m} f(x-y)g(y)dy$ can be regarded as an integral with respect to the measure v having density g with respect to Lebesgue measure, $(f * g)(x) = \int_{\mathbb{R}^m} f(x-y)d\nu(y)$. The right-hand side of this equation will be used as the definition of the convolution of a function and a measure. To guarantee the existence of the convolution, we will assume that all measures are finite and all functions are bounded.

Definition *Let v be a finite Borel measure on $\mathbb{R}^m$, and let f be a measurable bounded function on $\mathbb{R}^m$. The convolution $f * v$ is defined by the equation*

$$(f * v)(x) = \int_{\mathbb{R}^m} f(x-y)dv(y) \quad (x \in \mathbb{R}^m)$$

One more version of convolution can be considered if μ is the counting measure defined on the integer lattice $\mathbb{Z}^m$. In this case, instead of functions, we speak of multiple sequences. By analogy with (2), the convolution of such sequences $f = \{f_n\}_{n\in\mathbb{Z}^m}$ and $g = \{g_n\}_{n\in\mathbb{Z}^m}$ is defined by the formula

$$(f * g)_k = \int_{\mathbb{Z}^m} f(k-n)g(n)d\mu(n) = \sum_{n\in\mathbb{Z}^m} f_{k-n}g_n \quad (k \in \mathbb{Z}^m)$$

1.4 Approximate Identities

1.4.1

If one of the convolution factors is non-negative and its integral is 1 , then the convolution can be regarded as the mean value of the other factor. Indeed, if $g \geqslant 0$ and $\int_{\mathbb{R}^m} g(y)dy = 1$, then $\inf_{\mathbb{R}^m} f \leqslant (f * g)(x) \leqslant \sup_{\mathbb{R}^m} f$ for $x \in \mathbb{R}^m$. If the support of g is contained in a ball $B(r)$, then the estimate can be sharpened as follows:

$$\inf_{B(x,r)} f \leqslant (f * g)(x) \leqslant \sup_{B(x,r)} f \quad (x \in \mathbb{R}^m)\,.$$

Therefore, if f is continuous, then the convolution must be close to f for a small r. At the same time, the convolution often has higher degree of smoothness than the function f itself. In particular, as we will prove in Example 1 of Sect. 1.4.2, the convolution of an arbitrary locally summable function and the characteristic function of an arbitrary ball is continuous. Thus, we may hope that convolution can be used to obtain a method of approximating functions by smoother ones.

Since the convolution of a locally summable function and a characteristic function of a ball is continuous, we obtain that there is no locally summable function playing the role of identity for convolution; in other words, there is no locally summable function the convolution with which would not change the other convolution factor. At the same time, a convolution with measure δ_0 generated by a unit point mass concentrated at zero has this property,

$$(f * \delta_0)(x) = \int_{\mathbb{R}^m} f(x-y)d\delta_0(y) = f(x) \quad \text{for all } x \text{ in } \mathbb{R}^m$$

The measure δ_0 certainly does not have a density with respect to Lebesgue measure. However, avoiding integration with respect to the measure δ_0, the famous physicist Paul Dirac actually suggested to assume that such a density nevertheless exists. He introduced a "function" δ (now known as the Dirac delta function) having the following properties:

$$\text{I. } \delta(x) = \begin{cases} 0 & \text{for } x \neq 0, \\ +\infty & \text{for } x = 0, \end{cases} \quad \text{and}$$

$$\text{II. } \int_{\mathbb{R}^m} \delta(x)dx = 1.$$

From this he concluded that, for every continuous function f on $\mathbb{R}^m$, the relation $f(x) = \int_{\mathbb{R}^m} f(x-y)\delta(y)dy$ is valid, i.e., that δ is an identity for convolution in the class of continuous functions. It is this fact that plays a crucial role. Properties I and II characterizing the Dirac delta function are clearly incompatible. However, if we regard the integral $\int_{\mathbb{R}^m} f(x-y)\delta(y)dy$ simply as a new notation for the integral $\int_{\mathbb{R}^m} f(x-y)d\delta_0(y)$, then the calculation involving the function δ becomes legal.

As has already been said, the measure δ_0 has no density with respect to Lebesgue measure, and a function satisfying properties I and II does not exist. In this connection there is a problem of approximating δ_0 by measures having densities, i.e., by measures of the form $\omega(x)dx$. In a wide range of cases, we will see that, on the one hand, a convolution with a measure of this form causes little change in the function (since the measure is close to δ_0) and, on the other hand, it results in a function smoother than the initial one. This opens possibilities for approximating arbitrary functions by smooth ones. The character of an approximation may be different and requires clarification.

First of all we define a family of functions by which the measure δ_0 is approximated.

Definition *Let $T \subset (0, +\infty)$, and let t_0 be a limit point of T $(0 \leqslant t_0 \leqslant +\infty)$. A family of functions $\{\omega_t\}_{t\in T}$ defined on $\mathbb{R}^m$ is called an approximate identity in $\mathbb{R}^m$ (as $t \to t_0$) if*

(a) $\omega_t \geqslant 0$,

(b) $\int_{\mathbb{R}^m} \omega_t(x)dx = 1$,

(c) $\int_{\|x\|>\delta} \omega_t(x)dx \underset{t\to t_0}{\longrightarrow} 0$ *for every* $\delta > 0$.

Remarks

(1) Taking into account equation (b), we can restate condition (c) in the following form:

$$\int_{\|x\|<\delta} \omega_t(x)dx \underset{t\to t_0}{\longrightarrow} 1 \quad \text{for every } \delta > 0$$

Thus, the main contribution to the integral $\int_{\mathbb{R}^m} \omega_t(x)dx$ comes from the integral over an arbitrarily small neighborhood of zero. This property of an approximate identity is sometimes called the localization property. It says that, for t close to t_0, the graph of ω_t can schematically be displayed as a "narrow and tall hump". Such functions are sometimes called δ-images.

(2) Sometimes the positivity condition for ω_t is lifted and condition a) is replaced by the less restrictive assumption

$$\text{(a')} \qquad \int_{\mathbb{R}^m} |\omega_t(x)|\, dx \leqslant C \quad \text{for some } C > 0 \text{ and all } t \in T$$

(and the function ω_t in condition (c) is replaced by $|\omega_t|$).

Because of equation (b), condition (a') is automatically fulfilled for nonnegative functions. Many of the results obtained below also remain valid in a more general setting, but we will not dwell on this.

1.4.2

We consider some examples of approximate identities. In all cases the families under consideration are approximate identities as $t \to +0, T = (0, +\infty)$, and the convolution factors are assumed to be locally summable on $\mathbb{R}^m$.

Example 1 (Steklov averages) Let $\omega_t = \frac{1}{v(t)}\chi_{B(t)}$, where $v(t)$ is the volume (Lebesgue measure) of a ball $B(t)$ in $\mathbb{R}^m$. Obviously, this family is an approximate identity. The value of the convolution $f * \omega_t$ at a point x is the average of f over the ball $B(x,t)$:

$$(f * \omega_t)(x) = \int_{\mathbb{R}^m} f(y)\frac{1}{v(t)}\chi_{B(t)}(x-y)dy = \frac{1}{v(t)}\int_{B(x,t)} f(y)dy.$$

This average has systematically been used by Steklov, and the convolutions $f_t = f * \omega_t$ are called Steklov averages of f. They are continuous if the function is locally summable. Indeed, assume that $\|x - x_0\| < 1$. For such x, the symmetric difference e_x of the balls $B(x_0, t)$ and $B(x,t)$ lies in the ball $B(x_0, 1+t)$. Since the function f is summable on $B(x_0, 1+t)$, and $\lambda(e_x) \to 0$ as $x \to x_0$, it follows from the absolute continuity of the integral that

$$|f_t(x) - f_t(x_0)| \leqslant \frac{1}{v(t)}\int_{e_x} |f(y)|dy \underset{x\to x_0}{\longrightarrow} 0$$

Example 2 The example considered above fits into a general scheme allowing one to construct different approximate identities. The scheme is as follows.

Let ψ be a non-negative summable function on $\mathbb{R}^m$, and let

$$C = \int_{\mathbb{R}^m} \psi(x)dx > 0$$

We put

$$\omega_t(x) = \frac{1}{Ct^m}\psi\left(\frac{x}{t}\right) \quad (x \in \mathbb{R}^m)$$

The family $\{\omega_t\}_{t>0}$ is an approximate identity as $t \to 0$. Condition (a) in the definition of an approximate identity is obviously fulfilled, and the fact that condition (b) is valid can be verified by the change of variable $y = x/t$:

$$\int_{\mathbb{R}^m} \omega_t(x)dx = \frac{1}{C}\int_{\mathbb{R}^m} \psi(y)dy = 1$$

At the same time, condition (c) is also fulfilled since we have

$$\int_{\|x\|<\delta} \omega_t(x)dx = \frac{1}{C}\int_{\|y\|<\delta/t} \psi(y)dy \underset{t\to 0}{\longrightarrow} 1$$

for every $\delta > 0$.

In Example 1, the characteristic function of the ball $B(1)$ plays the role of ψ.

It is especially convenient to use approximate identities obtained by the method described above in the case where ψ is a function of class C^∞ and its support lies in the unit ball. Such approximate identities were first systematically used by Sobolev, and we call them Sobolev approximate identities.

1.4.3

Now we state the main result concerning approximate identities.

Theorem *Let f be a bounded measurable function on $\mathbb{R}^m$, and let $\{\omega_t\}_{t\in T}$ be an approximate identity as $t \to t_0$, $f_t = f * \omega_t$. Then:*

(1) if the limit $L = \lim_{x\to x_0} f(x)$ exists and is finite for a point x_0 in $\mathbb{R}^m$, then $f_t(x_0) \underset{t\to t_0}{\longrightarrow} L$

(2) if $f \in C(\mathbb{R}^m)$, then $f_t \underset{t\to t_0}{\rightrightarrows} f$ on every bounded set.

Proof By definition

$$f_t(x_0) = \int_{\mathbb{R}^m} f(x_0 - y)\,\omega_t(y)dy$$

Multiplying the equation

$$1 = \int_{\mathbb{R}^m} \omega_t(y)dy$$

by L and subtracting the equation obtained from the preceding one, we obtain

$$f_t(x_0) - L = \int_{\mathbb{R}^m} (f(x_0 - y) - L)\,\omega_t(y)dy$$

We prove that the right-hand side of this relation tends to zero as $t \to t_0$. By assumption, we have $|f| \leqslant C$ everywhere. Therefore, the inequality

$$\begin{aligned} |f_t(x_0) - L| &\leqslant \int_{\mathbb{R}^m} |f(x_0 - y) - L|\,\omega_t(y)dy = \int_{\|y\|<\delta} \cdots + \int_{\|y\|>\delta} \cdots \\ &\leqslant \sup_{0<\|z-x_0\|<\delta} |f(z) - L| \int_{\|y\|<\delta} \omega_t(y)dy + 2C \int_{\|y\|>\delta} \omega_t(y)dy \end{aligned} \tag{1}$$

holds for every $\delta > 0$. Since $\int_{\|y\|<\delta} \omega_t(y)dy \leqslant \int_{\mathbb{R}^m} \omega_t(y)dy = 1$, it follows that

$$|f_t(x_0) - L| \leqslant \sup_{0<\|z-x_0\|<\delta} |f(z) - L| + 2C \int_{\|y\|>\delta} \omega_t(y)dy$$

Now, we can make the first summand on the right-hand side of the inequality arbitrarily small by an appropriate choice of δ, and then, fixing δ, we can make the second summand small by condition (c).

The proof of the second assertion of the theorem will repeat the proof of the first one if we replace x_0 by x, L by $f(x)$, and take into account that, for every bounded set E, we can choose the same δ for all $x \in E$ since f is uniformly continuous on every bounded set.

Remark It can be seen from the proof of the theorem that if f is uniformly continuous on the entire space, then $f_t \underset{t\to t_0}{\rightrightarrows} f$ on $\mathbb{R}^m$.

Corollary *If g is a bounded function continuous at zero, then*

$$\int_{\mathbb{R}^m} g(y)\omega_t(y)dy \underset{t\to t_0}{\longrightarrow} g(0)$$

Proof This is a particular case of the statement of the theorem where $x_0 = 0$,$f(x) = g(-x)$ and $L = g(0)$.

The corollary reinforces our motivation to introduce approximate identities. It follows from the corollary that the measures v_t with densities ω_t converge to the measure δ_0 generated by the unit load concentrated at zero in the sense that, for every bounded continuous function g, we have

$$\int_{\mathbb{R}^m} g(x)dv_t(x) \underset{t\to t_0}{\longrightarrow} g(0) = \int_{\mathbb{R}^m} g(x)d\delta_0(x)$$

If we also assume that ω_t are functions with compact supports contracting to zero, then this statement is valid for every (possibly unbounded) continuous function.

1.4.4

We consider an important application of approximate identities and prove Weierstrass' famous approximation theorem stating that every continuous function on a closed bounded interval can be approximated by a polynomial as closely as desired. The method of proof we use here is that we first replace the given function, with small error, by the convolution with some "nice" function, and then construct a polynomial approximation for the convolution. This method works equally well for functions of one variable and for functions of several variables.

Following Weierstrass, we will consider the convolutions of a given function and functions of the form

$$W_t(x) = \frac{1}{t^m} e^{-\pi \frac{\|x\|^2}{t^2}} \quad (x \in \mathbb{R}^m, t > 0)$$

This family is an approximate identity as $t \to +0$. Condition (a) of the definition of an approximate identity is obviously fulfilled; using the value of the multidimensional Euler integral, we can easily verify condition (b). We leave the verification of the localization property to the reader.

Theorem 1 *(Weierstrass approximation theorem) Let $f \in C(\mathbb{R}^m)$. Then for any $R > 0$ and $\varepsilon > 0$ there exists a polynomial P in m variables such that*

$$|f(x) - P(x)| < \varepsilon \quad \text{for all } x \text{ in } \bar{B}(R).$$

Proof First we assume that the support of f is a compact set lying in the ball $\bar{B} \equiv \bar{B}(R)$ (otherwise we can increase the radius R). We put $f_t = f * W_t$. As pointed out in the remark to Theorem 1.4.3, $f_t \rightrightarrows f$ as $t \to 0$. We fix a t such that

$$|f(x) - f_t(x)| < \varepsilon \quad \text{for each } x \text{ in } \mathbb{R}^m. \tag{2}$$

Now we show that every function f_t can be uniformly approximated by a polynomial in the ball $\bar{B}$. Since f is zero outside $\bar{B}$, we obtain

$$f_t(x) = \int_{B(R)} f(y) W_t(x-y) dy$$

We assume that $x \in \bar{B}$, which implies that $x - y \in B(2R)$ in the last integral.

The next idea is to find a good polynomial approximation for the function W_t in the ball $B(2R)$ and use the fact that the convolution of a function with compact support and a polynomial is again a polynomial. To verify the last assertion, we consider an arbitrary polynomial Q. It is clear that $Q(x-y)$ is also a polynomial in the coordinates of x with coefficients dependent on y. After multiplying by the function $f(y)$ with compact support, we obtain that the coefficients become summable. Integrating them, we obtain certain numbers, and, therefore, the convolution is a polynomial.

Now we turn our attention to approximating the function W_t by a polynomial. By Taylor's formula (with the Lagrange form of the remainder) we obtain $e^{-u} = T_{n-1}(u) + r_n(u)$, where T_{n-1} is a polynomial of degree $n-1, r_n(u) = \frac{1}{n!} e^{-\theta u}(-u)^n$, $0 < \theta < 1$. It is clear that $|r_n(u)| \leqslant u^n/n\ !$ for $u \geqslant 0$. By the definition of W_t, we obtain

$$W_t(x) = \frac{1}{t^m} T_{n-1}\left(\pi \frac{\|x\|^2}{t^2}\right) + \frac{1}{t^m} r_n\left(\pi \frac{\|x\|^2}{t^2}\right) = P_n(x) + \rho_n(x),$$

where P_n is a polynomial (as the composition of T_{n-1} and the polynomial $\pi \frac{\|x\|^2}{t^2}$) and ρ_n satisfies the estimate

$$|\rho_n(x)| = \left| \frac{1}{t^m} r_n\left(\pi \frac{\|x\|^2}{t^2}\right)\right| \leqslant \frac{1}{t^m n!}\left(\pi \frac{\|x\|^2}{t^2}\right)^n \leqslant \frac{1}{t^m n!}\left(\frac{4\pi R^2}{t^2}\right)^n \qquad (3)$$

for $\|x\| \leqslant 2R$. Since $f(x) = 0$ outside the ball $\bar{B}$, we see that the convolution $f * \rho_n$ satisfies the inequality

$$|(f * \rho_n)(x)| = \left| \int_{B(R)} f(y) \rho_n(x-y) dy \right| \leqslant M \int_{B(R)} |\rho_n(x-y)|\, dy$$

where $M = \max_x |f(x)|$. Since the inequality $\|x - y\| \leqslant 2R$ is valid for $\|x\| \leqslant R$, we can use (3) to estimate the integral on the right-hand side of the above inequality. We obtain

$$|(f * \rho_n)(x)| \leqslant M v(R) \frac{1}{t^m n!}\left(\frac{4\pi R^2}{t^2}\right)^n$$

where $v(R)$ is the m-dimensional volume of the ball $B(R)$. Now we fix an n so that the right-hand side of the last inequality is less than ε. Then we obviously obtain

$$|f_t(x) - (f * P_n)(x)| = |(f * \rho_n)(x)| < \varepsilon \tag{4}$$

for $\|x\| \leqslant R$. This inequality together with (2) shows that $|f(x) - (f * P_n)(x)| < 2\varepsilon$ for $x \in \bar{B}$. This completes the proof of the theorem for a function with compact support because, as noted above, the convolution $f * P_n$ is a polynomial.

In the general case, it is sufficient to replace f by a continuous function f_1 that has a compact support and coincides with f in the ball $\bar{B}$. Constructing a polynomial that approximates f_1 in $\bar{B}$, we also find an approximation for f.

Corollary 1 *Let f be a continuous function on a compact set $K \subset \mathbb{R}^m$. Then for every $\varepsilon > 0$ there exists a polynomial P such that $|f(x) - P(x)| < \varepsilon$ for all $x \in K$.*

Proof By the Tietze-Urysohn theorem (*Every function f_0 continuous on a closed subset of a metrizable space X is the restriction of a function from $C(X)$. If $|f_0| \leq C$, one may assume that the extended function also satisfies this inequality.*), every continuous function on a closed subset of the space $\mathbb{R}^m$ can be extended to a continuous function defined on the entire space. Therefore, it is sufficient to apply the theorem to the extended function, assuming that R is so large that $K \subset B(R)$.

We shall mention one more consequence of the Weierstrass approximation theorem.

Corollary 2 *Let f be a continuous function on $\mathbb{R}^m$. Assume that f has a compact support. Then, for every $\varepsilon > 0$, there exists an infinitely differentiable function g with a compact support such that $|f(x) - g(x)| < \varepsilon$ for all $x \in \mathbb{R}^m$.*

Proof Assume that f vanishes outside the ball $B(R)$, and let P be a polynomial approximating f with accuracy ε in the ball $B(R+1)$. We obtain the required function g if we multiply P by a function φ of class C^∞ such that $0 \leqslant \varphi \leqslant 1$, $\varphi(x) = 1$ for $x \in B(R)$, and φ vanishes outside $B(R+1)$.

Remark If the function f in Corollary 2 is non-negative, then we may assume that the function g is also non-negative.

Indeed, otherwise, the function g can be replaced by $\varphi \cdot (g + \varepsilon)$, which, obviously, is non-negative and approximates f with accuracy 2ε.

Generalizing Theorem 1, we prove that a smooth function together with its derivatives can be approximated by a polynomial. In the next theorem, the letter k denotes a multi-index $\left(k \in \mathbb{Z}_+^m\right)$, and the symbol $D^k f$, where $k = (k_1, \ldots, k_m)$, denotes the derivative of f of order $|k| = k_1 + \cdots + k_m$ such that the differentiation with respect to the j th coordinate is carried out k_j times.

Theorem 2 *Let $f \in C^r(\mathbb{R}^m)$ $(r \in \mathbb{N})$. Then, for all $R > 0$ and $\varepsilon > 0$, there exists a polynomial P in m variables such that*

$$\left|D^k f(x) - D^k P(x)\right| < \varepsilon \quad \text{for all } x \text{ in } \bar{B}(R) \text{ and all } k, 0 \leqslant |k| \leqslant r.$$

Proof As in Theorem 1, we may assume without loss of generality that $\operatorname{supp}(f) \subset \bar{B}(R)$. Since, by properties of convolution, we have $D^k(f_t) = \left(D^k f\right) * W_t$, we can choose the parameter $t > 0$ so that inequality (2) and similar inequalities for $D^k f$ are valid for $|k| \leqslant r$ and all $x \in \mathbb{R}^m$. We put $M = \max_{x, |k| \leqslant r} \left|D^k f(x)\right|$.

Then, for an appropriate choice of n, inequality (4) turns out to be valid not only for the function f, but also for all its derivatives up to order r inclusive.

1.4.5

Here, relying on the concept of the convolution of periodic functions, we define a periodic approximate identity and prove Weierstrass' theorem on approximation by trigonometric polynomials. We can easily change a period by contraction. Therefore, we may assume without loss of generality that all functions considered in the present section are 2π-periodic with respect to each variable (and only such functions will be called periodic).

For the case of periodic functions, the definition of an approximate identity from Sect. 1.4.1 can easily be modified as follows (below, $Q = [-\pi, \pi]^m$): a family of periodic functions $\{\omega_t\}_{t\in T}$ is called a periodic approximate identity (as $t \to t_0$) if:

(a) $\omega_t \geqslant 0$,

(b) $\int_Q \omega_t(x)dx = 1$,

(c) $\int_{Q\setminus B(\delta)} \omega_t(x)dx \underset{t\to t_0}{\longrightarrow} 0$ for each $\delta \in (0, \pi)$.

We also introduce the following strong version of the localization property:

$$(\mathrm{c}') \quad \omega_t(x)dx \underset{t\to t_0}{\rightrightarrows} 0 \quad \text{on } Q\setminus B(\delta) \text{ for each } \delta \in (0, \pi)$$

An almost verbatim repetition of the proof of Theorem 1.4.3 verifies the following approximative properties for the periodic convolution $f_t = f * \omega_t$.

Theorem *Let a periodic function f be measurable and bounded on the cube Q. Then:*

(a) if the limit $L = \lim_{x\to x_0} f(x)$ exists and is finite at a point $x_0, x_0 \in \mathbb{R}^m$, then $f_t(x_0) \underset{t\to t_0}{\longrightarrow} L$

(b) if $f \in C(\mathbb{R}^m)$, then $f_t \underset{t\to t_0}{\rightrightarrows} f$ on $\mathbb{R}^m$.

Remark If an approximate identity satisfies condition (c′), then assertion (a) remains valid for every periodic function summable on Q. For the proof, we replace inequality (1) by the inequality

$$\begin{aligned} |f_t(x_0) - L| &\leqslant \int_Q |f(x_0 - y) - L|\, \omega_t(y)dy = \int_{B(\delta)} \cdots + \int_{Q\setminus B(\delta)} \cdots \\ &\leqslant \sup_{y\in B(\delta)} |f(x_0 - y) - L| + \sup_{y\in Q\setminus B(\delta)} \omega_t(y) \int_Q |f(x_0 - y) - L|\, dy, \end{aligned}$$

after which the proof can be completed as in Theorem 1.4.3: first, by a choice of δ, we make the first summand on the right-hand side of the inequality small, and then make the second summand small with the help of condition (c′).

Leaning on the last theorem, we now obtain a periodic version of the Weierstrass approximation theorem. Since the convolution of a summable function

with a trigonometric polynomial is again a trigonometric polynomial, it is sufficient for us to construct an approximate identity consisting of such polynomials.

We begin with the one-dimensional case and consider the trigonometric polynomial

$$\Theta_n(x) = \frac{1}{c_n} \cos^{2n} \frac{x}{2} = \frac{1}{c_n} \left(\frac{1 + \cos x}{2} \right)^n ,$$

where the coefficient c_n is such that $\int_{-\pi}^{\pi} \Theta_n(x) dx = 1$, i.e., $c_n = \int_{-\pi}^{\pi} \cos^{2n} \frac{x}{2} dx$. These integrals were calculated previously, but in what follows it is important only that they tend to zero not too fast,

$$c_n = 4 \int_0^{\frac{\pi}{2}} \cos^{2n} y dy \geqslant 4 \int_0^{\frac{\pi}{2}} \sin y \cos^{2n} y dy = \frac{4}{2n+1} > \frac{1}{n}$$

It follows that the sequence of functions Θ_n have the strong localization property (c′) stated at the beginning of the present section,

$$\sup_{\delta < |x| < \pi} \Theta_n(x) \leqslant n \cos^{2n} \frac{\delta}{2} \underset{n \to \infty}{\longrightarrow} 0 \quad \text{for each } \delta \in (0, \pi)$$

Using the periodic approximate identity Θ_n of one variable constructed above, we can easily construct its multi-dimensional counterpart,

$$\omega_n(x) = \Theta_n(x_1) \cdots \Theta_n(x_m) \quad \text{for } x = (x_1, \ldots, x_m) \in \mathbb{R}^m$$

This sequence of trigonometric polynomials satisfies conditions (a)-(c) and, therefore, the theorem is valid for the sequence with $T = \mathbb{N}$ and $t_0 = +\infty$. Since $f_n = f * \omega_n$ is a trigonometric polynomial, we come to a periodic version of the Weierstrass approximation theorem.

Corollary (Weierstrass) *Let f be everywhere continuous periodic function. Then there exists a sequence of trigonometric polynomials converging to f uniformly on $\mathbb{R}^m$.*

2 Surface Integrals

In this chapter, our main aim is to give an exact meaning to the notion of the area of a smooth surface and to develop a means of its calculation. Undoubtedly, everybody has an intuitive idea of the area of a curved surface that one uses in everyday life (for instance, when one estimates the amount of paint consumption). At the same time, the evaluation of a curved surface area is quite a difficult problem compared with the analogous problem in the case of a plane figure. It is possible to reduce the first problem to the second one by elementary means only in the cases of conic and cylindrical surfaces: it is sufficient to "unroll" them. At school, one adds to this the calculation of the area of a sphere or its parts as a result of some unobvious argumentation.

Before we proceed to the computational side of the problem, it is necessary to overcome the principal difficulty, that is, to define the area of a surface.

We do not restrict ourselves to two-dimensional surfaces, so in what follows we discuss the construction of the measure (surface area) on smooth manifolds of arbitrary dimension. For this purpose, we need some results from the theory of smooth maps. For the convenience of the reader, the required material is collected in the auxiliary first section of the chapter.

2.1 Auxiliary Notions

Here we remind the reader of the principal notions and facts of the theory of smooth manifolds and fix the related notation and terminology.

We denote by the symbol $C^r(\mathcal{O}, \mathbb{R}^m)\,(1 \leqslant r \leqslant +\infty)$ the set of r times continuously differentiable maps defined on the set $\mathcal{O} \subset \mathbb{R}^k$, which is always assumed to be open, taking values in $\mathbb{R}^m$. We call C^1-maps smooth maps. If we talk about a map which is smooth on an arbitrary (non-open) set, we will always mean that it is defined and continuously differentiable on some neighborhood of this set, i.e., on a wider open set.

We denote by $d_a\Phi$ the differential of the map $\Phi \in C^1(\mathcal{O}, \mathbb{R}^m)$ at a point $a \in \mathcal{O}$, and by $\Phi'(a)$ the corresponding matrix (in the canonical bases of the spaces $\mathbb{R}^k$ and $\mathbb{R}^m$), i.e., the Jacobian matrix. This $m \times k$ matrix (with m rows and k columns) is formed, as is well known, by the partial derivatives $\frac{\partial \varphi_j}{\partial t_i} (1 \leqslant j \leqslant m, 1 \leqslant i \leqslant k)$ of the coordinate functions $\varphi_1, \ldots, \varphi_m$ of the map Φ. If $m = k$, then the Jacobian matrix is square. Its determinant $\det \left\| \frac{\partial \varphi_j}{\partial t_i} \right\|$ (called the Jacobian of the map Φ) is also denoted by $\frac{D(\varphi_1, \ldots, \varphi_k)}{D(t_1, \ldots, t_k)}$.

2.1.1

We introduce a concept that is essential in what follows.

Definition *A set $M, M \subset \mathbb{R}^m$, is called a simple k-dimensional manifold if it is homeomorphic to an open subset $\mathcal{O}$ of the set $\mathbb{R}^k (k \leqslant m)$. The homeomorphism $\Phi : \mathcal{O} \xrightarrow{on} M$ is called a parametrization of the manifold M. If for some $r = 1, 2, \ldots, +\infty$, $\Phi \in C^r(\mathcal{O}, \mathbb{R}^m)$ and $\operatorname{rank} d_a\Phi = k$ at every point $a \in \mathcal{O}$, then*

the parametrization Φ is said to be smooth of class C^r. A simple manifold that has such a parametrization is also called smooth (of class C^r).

We emphasize that, by definition, the domain of the parametrization is always an open set.

Since the position of a point $p = \Phi(t)$ on a manifold is uniquely determined by the parameter t, its coordinates $t_1, \dots, t_k$ are often called the curvilinear coordinates of the point p. In particular cases they often have a simple geometrical meaning that simplifies the solution of the problem.

The simplest example of a k-dimensional manifold is a k-dimensional vector subspace. Its parametrization can be obtained, for example, in the following way. Fix an arbitrary basis $\tau_1, \dots, \tau_k$ in the subspace and set

$$\Phi(t) = t_1\tau_1 + \dots + t_k\tau_k \quad \left(t = (t_1, \dots, t_k) \in \mathbb{R}^k\right).$$

Clearly, the map Φ satisfies all the requirements. The "curvilinear coordinates" of a vector in the subspace are simply its coordinates in the basis $\tau_1, \dots, \tau_k$.

Definition *(The first definition of a smooth manifold) A set $M, M \subset \mathbb{R}^m$, is called a k-dimensional manifold of class C^r if every point p from M has a neighborhood U such that the intersection $U \cap M$ is a simple k-dimensional manifold of class C^r. Its parametrization is called a local parametrization of the manifold M in the vicinity of the point p.*

A k-dimensional manifold of class C^0 is defined in an analogous way: it is a set M which is, locally, a simple k-dimensional manifold (without any smoothness conditions). Each of its points has a neighborhood U such that the intersection $U \cap M$ is a simple k-dimensional manifold. The number k is called the dimension of the manifold M and is denoted by the symbol $\dim M$. The difference $m - \dim M$ is called the codimension of the manifold. A manifold of codimension 1 is called a surface.

Contrary to local parametrization, a parametrization of a simple manifold is also called a global parametrization.

If the value of the parameter $r \geqslant 1$ is not important (in most cases, it is sufficient to take $r = 1$), we call the set M a smooth k-dimensional manifold, or a smooth manifold, and sometimes simply a manifold since otherwise the character of the manifold is indicated explicitly.

If a point p belongs to a manifold $M \subset \mathbb{R}^m$, then by its M-neighborhood, or relative neighborhood, we mean the intersection of the neighborhood of p in $\mathbb{R}^m$ with the manifold M. It is clear that every point of the manifold has a base of M neighborhoods whose closures lie in M. A coordinate neighborhood is a relative neighborhood which is a simple manifold, i.e., it admits a parametrization (and, consequently, curvilinear coordinates can be introduced).

In simple and important cases we come across examples of "almost smooth" manifolds (consider, for example, the boundary of a square, or a cube, etc.). Therefore, we expand the definition of a smooth manifold as follows: a piecewise smooth k-dimensional manifold is a union of a smooth k-dimensional manifold (possibly non-connected) and a set of zero k-dimensional Hausdorff measure. It

is clear that the boundaries of polyhedral bodies are piecewise smooth surfaces according to this definition.

Speaking formally, when we consider a smooth manifold in $\mathbb{R}^m$, we do not exclude the possibility that $\dim M$ equals m. In this case, as follows from the definition, M is simply an open subset of $\mathbb{R}^m$. Therefore, in what follows, we consider only manifolds whose dimension is less than the dimension of the enveloping space unless otherwise stated. At the same time, we admit the possibility that $\dim M = 1$. In this case we use the term "curve" instead of the term "manifold". A connected simple curve is also called a simple arc.

Another definition of a smooth manifold will be of use.

Definition *(The second definition of a smooth manifold) A set $M \subset \mathbb{R}^m$ is called a k-dimensional $(1 \leqslant k < m)$ manifold of class C^r if for every point $p \in M$ there exist a neighborhood U and functions $F_1, \ldots, F_{m-k}$ of class C^r defined on it such that:*

(1) $x \in M \cap U$ if and only if

$$F_1(x) = 0, \quad \ldots, \quad F_{m-k}(x) = 0 \tag{5}$$

and

(2) the vectors

$$\operatorname{grad} F_1(p), \quad \ldots, \quad \operatorname{grad} F_{m-k}(p) \tag{6}$$

are linearly independent.

In particular, a smooth surface (a manifold of codimension 1) is locally a level set of some smooth function with non-zero gradient. As is seen from the latter definition, locally every manifold lies in some surface.

2.1.2

Related to the notion of a smooth manifold are the important notions of tangent vector and tangent space. Recall that a path (in $\mathbb{R}^m$) is any continuous map from some segment into $\mathbb{R}^m$. A path is called smooth if its coordinate functions are smooth, and piecewise smooth if it is defined on a union $\bigcup_{j=0}^{n-1} [c_j, c_{j+1}]$ and its restrictions to the segments $[c_j, c_{j+1}]$ are smooth paths.

Definition *Let M be a smooth manifold in $\mathbb{R}^m$. A vector $\tau \in \mathbb{R}^m$ is called a tangent vector to M at a point $p, p \in M$, if there exists a smooth path $\gamma : [a, b] \mapsto \mathbb{R}^m$ such that $\gamma(t) \in M$ for $t \in [a, b]$, and for some $c \in (a, b)$ we have $\gamma(c) = p$ and $\gamma'(c) = \tau$.*

If some M-neighborhood of a point $p = \gamma(c)$ lies in a level set of a smooth function F, then $F(\gamma(t)) \equiv \text{const}$ for t close to c. Therefore, $\langle \operatorname{grad} F(p),\ \gamma'(c) \rangle = 0$, i.e., the tangent vector at the point p is orthogonal to the vector $\operatorname{grad} F(p)$.

Let Φ be a local parametrization of the manifold M in the vicinity of a point $p = \Phi(a)$. "Freezing" all coordinates of a point $a = (a_1, \ldots, a_k)$, except the j-th one, and making the latter change in the vicinity of a_j, we get a path that parametrizes the curve that passes through the point p. This curve is called a coordinate line. The vector tangent to this curve at the point p that corresponds

to the mentioned parametrization is the j-th column of the matrix $\Phi'(a)$; we denote it by $D_j\Phi(a)$ or $\tau_j = \tau_j(a)$. Since rank $d_a\Phi = k$, the vectors $\tau_1, \ldots, \tau_k$ are linearly independent. It is clear that $\tau_j(a) = d_a\Phi\,(e_j)$ (where the vectors $e_1, \ldots, e_k$ form the canonical basis in $\mathbb{R}^k$). We call them the canonical tangent vectors related to the parametrization Φ. The set of all vectors tangent to the manifold M at the point p is called the tangent space and is denoted by $T_p(M)$, or T_p for short. Note that this term needs validation, that is, one must check that T_p is actually a vector space.

Lemma *T_p is a k-dimensional subspace of the space $\mathbb{R}^m$.*

In the case where $k = m - 1$, we also call the subspace T_p a tangent plane.

Proof Assume that in the vicinity of the point p the manifold M is given by the Eqs. (1) and that the vectors (2) are linearly independent.

We check that, along with any two vectors τ_1, τ_2, the set T_p contains their linear combination $\tau = \alpha_1\tau_1 + \alpha_2\tau_2$. We may assume that $\tau \neq 0$, otherwise it suffices to take a constant path. Augmenting the vector system (2), which is orthogonal to the vector τ, with vectors $h_1, \ldots, h_{k-1}$ to a basis in the orthogonal complement to τ, we consider the system of equations

$$\begin{aligned} &F_1(x) = 0, \quad \ldots, \quad F_{m-k}(x) = 0, \quad \langle x - p, h_1\rangle = 0, \quad \ldots, \\ &\langle x - p, h_{k-1}\rangle = 0 \end{aligned} \tag{7}$$

According to the second definition of a smooth manifold, this system defines a smooth one-dimensional manifold in the vicinity of the point p, i.e., a smooth curve that obviously lies in M and passes through p.

Let γ be some parametrization of this curve in the vicinity of the point p. Without loss of generality, we may assume that $p = \gamma(0)$. Then the (non-zero!) vector $\gamma'(0)$ is orthogonal to the gradients (at the point p) of all functions appearing in system (3). Therefore, it is proportional to the vector τ.

For an appropriate choice of the coefficient θ, the vector tangent to the path $\widetilde{\gamma}(t) = \gamma(\theta t), |t| \leqslant \delta$, at $t = 0$ coincides with τ, i.e., $\tau \in T_p$.

Thus, we have proved that T_p is a vector subspace of the space $\mathbb{R}^m$. Its dimension does not exceed k since all vectors in it are orthogonal to the vectors of system (2). Moreover, it contains k linearly independent vectors that are tangent to the coordinate lines. Therefore, $\dim T_p = k$.□

Remark If the surface M is defined by the equation $F(x) = 0$ in the vicinity of the point p and $\operatorname{grad} F(p) \neq 0$, then, as noted before the lemma, the vectors tangent to it at the point p are orthogonal to the vector $\operatorname{grad} F(p)$. Therefore, the tangent space to M at the point p is the plane that consists of the vectors orthogonal to $\operatorname{grad} F(p)$, i.e., it is defined by the equation $\langle x, \operatorname{grad} F(p)\rangle = 0$.

We note that since the canonical tangent vectors corresponding to the parametrization Φ of the manifold M are linearly independent, they form a basis in the tangent space. The linearity of the map $d_a\Phi : \mathbb{R}^k \to \mathbb{R}^m$ implies that for every vector $t = (t_1, \ldots, t_k)$ in $\mathbb{R}^k$ the equality

$$d_a\Phi(t) = \sum_{j=1}^{k} t_j\tau_j$$

holds. Therefore, the differential of the parametrization maps $\mathbb{R}^k$ onto the tangent space isomorphically.

Sometimes it is more geometrically clear to consider the affine tangent space L_p instead of the tangent space T_p; it is the shift of T_p by the vector $p : L_p = p + T_p$. Since $p + d\Phi_a(t-a) \in L_p$, where $p = \Phi(a)$, and

$$\Phi(t) = p + d\Phi_a(t-a) + o(\|t-a\|) \quad \text{as } t \to a,$$

the point $x = \Phi(t)$ satisfies the relation

$$\begin{aligned} \operatorname{dist}(x, L_p) = \operatorname{dist}(\Phi(t), L_p) &\leqslant \|\Phi(t) - (p + d\Phi_a(t-a))\| \\ &= o(\|t-a\|) \quad \text{as } t \to a. \end{aligned}$$

By Corollary 1 from Sect. 2.1.4, the map Φ^{-1} satisfies the Lipschitz condition

$$\|t-a\| = \left\|\Phi^{-1}(x) - \Phi^{-1}(p)\right\| \leqslant C\|x-p\|$$

in the vicinity of the point p, and therefore

$$\operatorname{dist}(x, L_p) = o(\|x-p\|) \quad \text{as } x \to p, x \in M$$

Thus, when we substitute points in the subspace L_p for points in the manifold, the relative error tends to zero, i.e., the manifold M is "almost flat" in the small. The latter relation is a formalization of our intuitive idea of the tangent space as the space "tight-fitting" to the manifold. It can be proved that this property of the affine tangent space uniquely determines it.

2.1.3

We give some examples.

Example 1 An important example of a surface is the graph of a smooth function f defined on an open subset of the space $\mathbb{R}^{m-1}$. By definition of the graph, it is the set

$$\Gamma_f = \{(x_1, \ldots, x_{m-1}, y) \in \mathbb{R}^m \mid (x_1, \ldots, x_{m-1}) \in \mathcal{O}, y = f(x_1, \ldots, x_{m-1})\}$$

The map

$$\mathcal{O} \ni x = (x_1, \ldots, x_{m-1}) \mapsto \Phi(x) = (x_1, \ldots, x_{m-1}, f(x))$$

is, obviously, a global parametrization of the graph. We call this parametrization canonical.

The graph of the function f may be considered as a zero level set of the function $F(x_1, \ldots, x_{m-1}, y) = y - f(x_1, \ldots, x_{m-1})$ defined on the set $\mathcal{O}' = \mathcal{O} \times \mathbb{R}$. We note that $\operatorname{grad} F \neq 0$ everywhere in the set $\mathcal{O}'$ and, in particular, in Γ_f. As follows from the remark after the proof of Lemma 2.1.2, the affine tangent plane at the point $p = (a_1, \ldots, a_{m-1}, f(a))$, where $a = (a_1, \ldots, a_{m-1}) \in \mathcal{O}$, is given by the equation

$$y - f(a) = \langle \operatorname{grad} f(a), x - a \rangle = \sum_{j=1}^{m-1} f'_{x_j}(a)\,(x_j - a_j)$$

A set that can be obtained from the graph by changing the order of coordinates (so that the "dependent" coordinate does not occur in the last position) is also called a graph, or, more precisely, a graph in a wider sense. Clearly, a set $M \subset \mathbb{R}^m$ such that $M \cap U$ is a graph (in this wide sense) for some neighborhood U of each of its points is a surface.

Example 2 Consider the sphere

$$S(R) = \left\{(x_1, \ldots, x_m) \in \mathbb{R}^m \mid x_1^2 + \cdots + x_m^2 = R^2\right\}$$

in $\mathbb{R}^m$. We check that it is a surface. For every point $p = (p_1, \ldots, p_m)$ in $S(R)$, at least one coordinate is non-zero. Assume, for the sake of definiteness, that $p_m > 0$. Then the point p belongs to the upper hemisphere $S_+(R) = \{x \in S(R) \mid x_m > 0\}$ which is simply the graph of the function $f\,(x_1, \ldots, x_{m-1}) = \sqrt{R^2 - x_1^2 - \cdots - x_{m-1}^2}$ defined in a ball of the space $\mathbb{R}^{m-1}$. This function is of class C^∞. Therefore, the hemisphere $S_+(R)$, and thus all the sphere $S(R)$, are C^∞ surfaces.

It is intuitively clear that the sphere has no global parametrization. At the same time, one can easily give a map that parametrizes almost all of the sphere. We restrict ourselves to the most obvious particular case of the two-dimensional sphere in $\mathbb{R}^3$ (see the discussion of the general case in Exercise 5). Recall the geographical coordinates, the longitude φ and the latitude θ of a point on the surface of the Earth. For $\varphi \in [-\pi, \pi]$ and $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, we set

$$\Phi(\varphi, \theta) = (R\cos\varphi\cos\theta, R\sin\varphi\cos\theta, R\sin\theta) \tag{8}$$

(the corresponding coordinate lines are parallels and meridians; the eastern hemisphere corresponds to the positive values of φ, the western to the negative values, the northern hemisphere is determined by the inequality $\theta > 0$, whereas the southern hemisphere corresponds to $\theta < 0$). In the given example we deal with a rather typical situation. Speaking formally, the map Φ is defined for any φ and θ, but we are interested only in its restrictions to some subsets that are convenient for our considerations. It is clear that Φ is an infinitely differentiable map, but it is not bijective since $\Phi\left(\varphi, \pm\frac{\pi}{2}\right) = (0, 0, \pm R)$ for all values of φ (there is no natural way to ascribe a longitude to the north or south pole). Moreover, for any θ we lose injectivity for $\varphi = \pm\pi$ since the angles $\varphi = \pi$ and $\varphi = -\pi$ correspond to the same point on the sphere. These values of the parameter φ correspond to the meridian on the Earth, called the International Date Line, where the date changes as a ship or aeroplane travels east or west across it. Deleting it (together with the poles), we get the "cut sphere", i.e., the C^∞-surface that has a global parametrization (4) defined on an open rectangle $|\varphi| < \pi, |\theta| < \frac{\pi}{2}$. The condition $\operatorname{rank} d\Phi \equiv 2$, i.e., the linear independence of the tangent vectors

$$\tau_1 = D_1\Phi(\varphi,\theta) = (-R\sin\varphi\cos\theta, R\cos\varphi\cos\theta, 0)$$
$$\tau_2 = D_2\Phi(\varphi,\theta) = (-R\cos\varphi\sin\theta, -R\sin\varphi\sin\theta, R\cos\theta)$$

is a consequence of their orthogonality (τ_1 is tangent to a parallel and τ_2 to a meridian), since $\|\tau_1\| = R\cos\theta \neq 0$ and $\|\tau_2\| = R$.

As we will see later, the deletion of a meridian is inessential when one integrates over a sphere.

It should also be noted that in analysis the angle θ' between the radius vector and the positive direction of the OZ axis is often used instead of the latitude. It varies in the interval $[0,\pi]$ and is related to the latitude θ by the relation $\theta + \theta' = \pi/2$.

Example 3 Consider the torus, the surface in $\mathbb{R}^3$ created by rotating the circle $(R-x)^2 + z^2 = r^2 (0 < r < R)$ around the axis OZ. As can easily be seen, the torus may be defined by the equation $\left(R - \sqrt{x^2+y^2}\right)^2 + z^2 = r^2$. No global parametrization of the torus exists (we do not dwell on the proof of this fact). The position of a point on this surface is determined by two angles φ and θ (the analogs of the longitude and the latitude on the sphere) by the relations

$$x = (R + r\cos\theta)\cos\varphi, \quad y = (R + r\cos\theta)\sin\varphi, \quad z = r\sin\theta$$

The infinitely differentiable map (defined in $\mathbb{R}^2$)

$$\Phi(\varphi,\theta) = ((R + r\cos\theta)\cos\varphi, (R + r\cos\theta)\sin\varphi, r\sin\theta)$$

maps the square $[-\pi,\pi]^2$ onto the torus. It is not bijective due to the 2π-periodicity of trigonometric functions. Deleting two circles corresponding to the angles $\varphi = \pm\pi$ and $\theta = \pm\pi$, we obtain "the torus with two cuts", a surface of the class C^∞, for which the restriction of Φ to the square $(-\pi,\pi)^2$ is a global parametrization. The condition $\operatorname{rank} d\Phi \equiv 2$ is fulfilled because the tangent vectors

$$\tau_1 = D_1\Phi(\varphi,\theta) = (-(R + r\cos\theta)\sin\varphi, (R + r\cos\theta)\cos\varphi, 0),$$
$$\tau_2 = D_2\Phi(\varphi,\theta) = (-r\cos\varphi\sin\theta, -r\sin\varphi\sin\theta, r\cos\theta)$$

are linearly independent: they are orthogonal and $\|\tau_1\| = R + r\cos\theta > 0$, $\|\tau_2\| = r > 0$.

In order to check that not only the torus with the cuts, but also the torus in the whole, is a smooth surface, we need to show that every point $p = \Phi(\varphi_0,\theta_0)$ has a neighborhood in the torus that admits a global parametrization. This parametrization can be obtained if we change the definition domain of the mapping Φ. We leave it to the reader to check that the square $(\varphi_0 - \pi, \varphi_0 + \pi) \times (\theta_0 - \pi, \theta_0 + \pi)$ may be regarded as such a domain. The corresponding neighborhood of the point p on the torus is the torus with the cuts along the circles $\varphi = \varphi_0 \pm \pi$ and $\theta = \theta_0 \pm \pi$.

We note that in the limit case where $r = R$, we rotate the circle $(R-x)^2 + z^2 = R^2$ around the OZ-axis. The set M thus obtained is not a smooth surface

since there is no M-neighborhood of the origin that is a simple surface. The reader can check however that the set $M\backslash\{0\}$ is a smooth surface of class C^∞, and so M is a piecewise smooth surface.

Example 4 Consider a manifold of minimal dimension, i.e., a curve. Its parametrization in a vicinity of an arbitrary point is a smooth vector function defined on an interval of a real line. It is a homeomorphism with non-zero derivative. It is clear that the graph of a function of one variable defined on an interval is a smooth flat curve, i.e., a curve in $\mathbb{R}^2$.

Another well-known curve is a circle. To get a more general example, recall that according to the second definition of a smooth manifold, the level set of a smooth function of two variables with non-zero gradient is a smooth curve. The example of the lemniscate of Bernoulli, a plane set consisting of the points (x, y) such that

$$\left(x^2 + y^2\right)^2 - \left(x^2 - y^2\right) = 0,$$

demonstrates that the hypothesis about the gradient is important: the point $(0, 0)$, where the gradient of the function $F(x, y) = \left(x^2 + y^2\right)^2 - \left(x^2 - y^2\right)$ is equal to zero, has no relative neighborhood which is homeomorphic to an interval. Near the origin, the lemniscate may be viewed as a union of two graphs, i.e., "a selfintersecting curve". Note that if we delete the origin from this set, we get a (disconnected) smooth curve. This shows that the lemniscate is a piecewise smooth curve.

Example 5 Consider the group $O(n)$ of orthogonal $n \times n$ matrices. We regard it as a subset of the n^2-dimensional Euclidean space which we identify with the set of all $n \times n$ matrices $U = \{u_{i,j}\}_{i,j=1}^n$ with elements u_{ij}. This subset is defined by the system of equations

$$\begin{array}{ll} u_{i,1}^2 + \cdots + u_{i,n}^2 = 1, & 1 \leqslant i \leqslant n, \\ u_{i,1}u_{k,1} + \cdots + u_{i,n}u_{k,n} = 0, & 1 \leqslant i < k \leqslant n. \end{array}$$

The gradients of the functions $F_i(U) = u_{i,1}^2 + \cdots + u_{i,n}^2$ and $F_{ik}(U) = u_{i,1}u_{k,1} + \cdots + u_{i,n}u_{k,n}$ evaluated at the points of $O(n)$ are linearly independent. To convince ourselves that this is true, represent these gradients as the matrices made up of the derivatives over $u_{i,j}$ placed at the intersection of the i th row and the j th column. Then each row of a matrix which represents a linear combination of the gradients contains only a linear combination of (pairwise orthogonal) rows of the matrix U, whence the needed property easily follows. Thus, $O(n)$ is a smooth manifold of dimension $n^2 - n - n(n-1)/2 = n(n-1)/2$. It is natural to call the map $U \mapsto U_0U$ (or $U \mapsto UU_0$), where $U \in \mathbb{R}^{n^2}$ and U_0 is a certain element in $O(n)$, the left (correspondingly, the right) shift in the set of all matrices. The shift preserves the Euclidean distance between matrices since, as is easily seen, the Euclidean norms of the matrices U and U_0U (UU_0), considered as elements of the space $\mathbb{R}^{n^2}$, are the same. Therefore, the shift in $O(n)$ is an isometry relative to the metric in $O(n)$ induced from the enveloping n^2-dimensional space.

2.1.4

In what follows, it is important that a parametrization of a k-dimensional manifold in $\mathbb{R}^m$ can be viewed, locally, as a restriction to the subspace $\mathbb{R}^k$ of a diffeomorphism defined on an open subspace of the space $\mathbb{R}^m$. More precisely, we can consider a canonical embedding of $\mathbb{R}^k$ in $\mathbb{R}^m$ where the vectors $(x_1, \ldots, x_k)$ in $\mathbb{R}^k$ are identified with the vectors $(x_1, \ldots, x_k, 0, \ldots, 0)$ in $\mathbb{R}^m$. Then the following statement about the extension of a parametrization to a diffeomorphism is true.

Lemma *Let $\mathcal{O}$ be an open subset of the space $\mathbb{R}^k$ and a be a point in $\mathcal{O}$. For a smooth parametrization Φ of the set $\Phi(\mathcal{O}) \subset \mathbb{R}^m$, a neighborhood $V \subset \mathbb{R}^m$ of the point $a = (a_1, \ldots, a_k, 0, \ldots, 0)$ and a diffeomorphism F defined on it can be given such that Φ and F coincide on $V \cap \mathbb{R}^k$.*

Proof Since the rank of the Jacobian matrix $\Phi'(a)$ is k, it has a $k \times k$ non-zero minor. Without loss of generality, we can assume that it is formed by the first rows of the matrix. Then $\frac{D(\varphi_1, \ldots, \varphi_k)}{D(t_1, \ldots, t_k)}(a) \neq 0$, where $\varphi_1, \ldots, \varphi_k$ are the coordinate functions of the map Φ.

We consider the map Θ from $\mathcal{O} \times \mathbb{R}^{m-k}$ to $\mathbb{R}^m$ defined by the formula

$$\Theta(t_1, \ldots, t_k, t_{k+1}, \ldots, t_m) = \Phi(t_1, \ldots, t_k) + (0, \ldots, 0, t_{k+1}, \ldots, t_m)$$

where $(t_1, \ldots, t_k) \in \mathcal{O}$ and $(t_{k+1}, \ldots, t_m) \in \mathbb{R}^{m-k}$. It is clear that Θ is a smooth map and extends Φ to $\mathcal{O} \times \mathbb{R}^{m-k}$. Moreover, $\operatorname{rank} d_a\Theta = m$ since $\det \Theta'(a) = \frac{D(\varphi_1, \ldots, \varphi_k)}{D(t_1, \ldots, t_k)}(a) \neq 0$.

By the local invertibility theorem, the restriction of Θ to a (sufficiently small) neighborhood V of the point a is a diffeomorphism. This restriction should be taken for F.□

If Φ is a local parametrization of a k-dimensional manifold M in the vicinity of a point p and F is the diffeomorphism described in the above lemma, then Φ^{-1} and F^{-1} coincide on some M-neighborhood of the point p, more precisely, on the set $\Phi(V_0)$, where $V_0 = V \cap \mathbb{R}^k$. Thus, the next two statements follow from this lemma.

Corollary 1 *In a sufficiently small M-neighborhood of a point p, the map Φ^{-1} satisfies the Lipschitz condition, i.e.,*

$$\left\|\Phi^{-1}(x) - \Phi^{-1}(y)\right\| \leqslant C\|x - y\| \quad \text{for } x, y \in M$$

and some C.

To prove the result, it suffices to note that the smooth map F^{-1} satisfies the Lipschitz condition in every closed ball in the domain of the map.

Corollary 2 *Let $\mathcal{O}$ and $\mathcal{O}'$ be open sets in $\mathbb{R}^k$ and $\Phi \in C^1(\mathcal{O}, \mathbb{R}^m)$ be a parametrization of the manifold $M, M \subset \mathbb{R}^m$. If $\Psi \in C^1(\mathcal{O}', \mathbb{R}^m)$ and $\Psi(\mathcal{O}') \subset M$, then the composition $\Phi^{-1} \circ \Psi$ is a smooth map.*

Indeed, for any point $t_0 \in \mathcal{O}'$ the map Φ^{-1} coincides with the smooth map F^{-1} in some M-neighborhood of the point $\Psi(t_0)$. Therefore, in a sufficiently small neighborhood of the point t_0, the map $\Phi^{-1} \circ \Psi = F^{-1} \circ \Psi$ is a composition of smooth maps.

2.1.5

We will use a simple geometrical fact based upon the following observation.

Every open subset G of the space $\mathbb{R}^m$ is a union of balls in G for which the radii and the coordinates of the centers are rational numbers.

Since every point x of the set G may be regarded as the center of a ball $B(x,r)$ in G, it suffices to note that $x \in B(y,\rho) \subset B(x,r)$ for $0 < \rho < r/2$ and $\|x-y\| < \rho$. Clearly, the number ρ and the coordinates of the vector y may be chosen to be rational.

Theorem (Lindelöf) *For any family $\{G_\alpha\}_{\alpha\in A}$ of sets that are open in $\mathbb{R}^m$ there exists an at most countable subfamily $\{G_\alpha\}_{\alpha\in A_0}$ (the set $A_0, A_0 \subset A$, is at most countable) with the same union:*

$$\bigcup_{\alpha\in A} G_\alpha = \bigcup_{\alpha\in A_0} G_\alpha$$

Proof Consider arbitrary balls, with rational radii and rational center coordinates, that are contained in at least one of the sets G_α. The collection of such balls is countable. Let $\{B_n\}_{n\in\mathbb{N}}$ be an enumeration of this collection. By the choice of these balls, for any $n \in \mathbb{N}$ there exists an index $\alpha_n \in A$ such that $B_n \subset G_{\alpha_n}$. Moreover, each set G_α is exhausted by the balls chosen this way:

$$G_\alpha \subset \bigcup_{n\in\mathbb{N}} B_n \quad \text{for any index } \alpha \in A$$

Consequently,

$$\bigcup_{\alpha\in A} G_\alpha \subset \bigcup_{n\in\mathbb{N}} B_n \subset \bigcup_{n\in\mathbb{N}} G_{\alpha_n}$$

Due to the evident inclusion $\bigcup_{n\in\mathbb{N}} G_{\alpha_n} \subset \bigcup_{\alpha\in A} G_\alpha$, one can take the set A_0 as the set of indices α_n.□

Corollary 1 *A smooth manifold can be represented as a union of an at most countable family of simple manifolds.*

Since the range of curvilinear coordinates is a countable union of compact sets, the following statement is true.

Corollary 2 *A smooth manifold is an at most countable union of compact sets such that each such set is a subset of a simple manifold.*

The following corollary is an immediate consequence of the preceding one.

Corollary 3 *A smooth manifold in $\mathbb{R}^m$ is a Borel subset of this space.*

Since a smooth surface is locally a graph (in the broad sense) of a smooth function, we also have the following.

Corollary 4 *A smooth surface is an at most countable union of graphs of smooth functions.*

2.1.6

Now we prove a useful fact that allows us to represent a smooth function as a sum of smooth functions with small supports. It often leads to important technical

simplifications due to the "localization" of the problem (see Sect. 2.6.5). Recall that the support of a function $\varphi : \mathbb{R}^m \to \mathbb{R}$, denoted by the symbol $\operatorname{supp}(\varphi)$, is the closure of the set $\{x \mid \varphi(x) \neq 0\}$.

Theorem (On a smooth partition of unity) *For every $\varepsilon > 0$ there exists a nonnegative function φ_ε of class $C^\infty(\mathbb{R}^m)$ such that $\operatorname{supp}(\varphi_\varepsilon) = [-\varepsilon, \varepsilon]^m$ and*

$$\sum_{n \in \mathbb{Z}^m} \varphi_\varepsilon(x - \varepsilon n) = 1 \quad \text{for any } x \text{ in } \mathbb{R}^m$$

Note that the number of non-zero summands of this sum is finite near every point $a \in \mathbb{R}^m$. More precisely, if $x \in a + (-\varepsilon, \varepsilon)^m$ and $\varphi_\varepsilon(x - \varepsilon n) \neq 0$, then $n \in \frac{1}{\varepsilon}a + (-2, 2)^m$.

Proof We use the following well-known example of a function of class $C^\infty(\mathbb{R})$:

$$\Psi(t) = \begin{cases} 0 & \text{if } t \leqslant 0 \\ e^{-1/t} & \text{if } t > 0 \end{cases}$$

The existence of its derivatives of all orders is evident for non-zero t, and at zero is a consequence of the easy-to-check representation of $\Psi^{(n)}(t)$, for $t > 0$, in the form $\Psi^{(n)}(t) = P_n(1/t)e^{-1/t}$, where P_n is a polynomial.

Set $\psi(x) = \prod_{k=1}^m \Psi\left(1 - x_k^2\right)$ where $x = (x_1, \ldots, x_m)$. Clearly, ψ is a class $C^\infty(\mathbb{R}^m)$ function which is positive in the cube $(-1, 1)^m$ and equal to zero outside it. Therefore, each function $x \mapsto \psi(x - n)$ is positive in the shifted cube $n + (-1, 1)^m$ $(n \in \mathbb{Z}^m)$. Since every point x belongs to at least one such cube, the sum

$$\Phi(x) = \sum_{n \in \mathbb{Z}^m} \psi(x - n)$$

is positive. It is the sum of only a finite number of infinitely-differentiable functions (see the remark that follows the statement of the theorem). Consequently, $\Phi \in C^\infty(\mathbb{R}^m)$. Take $\varphi_1(x) = \frac{\psi(x)}{\Phi(x)}$. It is clear that this function satisfies the hypothesis of the theorem for $\varepsilon = 1$. To construct the function φ_ε for arbitrary ε, it suffices, via a scaling, to set $\varphi_\varepsilon(x) = \varphi_1\left(\frac{1}{\varepsilon}x\right)$.□

2.1.7

We show how one can construct a smooth approximation of characteristic functions using a partition of unity. It is intuitively clear that outside the set E, the values of its characteristic function can be altered gradually, without sudden jumps, decreasing to zero. It is also plausible that such a "descent" can be effectuated in the vicinity of E, without overstepping the limits of its arbitrarily small ε-neighborhood. Recall that the ε-neighborhood of the set E is the set

$$E_\varepsilon = \{y \in \mathbb{R}^m \mid \operatorname{dist}(y, E) < \varepsilon\} = \bigcup_{x \in E} B(x, \varepsilon)$$

We also show that this smoothing can be made without a steep drop of the smoothing function, i.e., we can control the norm of its gradient so that, under the circumstances considered, it is of the smallest possible order.

Theorem (On a smooth descent) *For every set $E \subset \mathbb{R}^m$ and every $\varepsilon > 0$ there exists a function θ_ε of class $C^\infty(\mathbb{R}^m)$ such that:*

(a) $0 \leqslant \theta_\varepsilon \leqslant 1$ on $\mathbb{R}^m$;

(b) $\theta_\varepsilon(x) = 1$ if $x \in E$;

(c) $\theta_\varepsilon(x) = 0$ outside E_ε;

(d) $\|\operatorname{grad} \theta_\varepsilon\| \leqslant \frac{c_m}{\varepsilon}$ on $\mathbb{R}^m$, where c_m is a coefficient that depends only on the dimension.

Proof Take $\delta = \varepsilon/(2\sqrt{m})$ and let φ_δ be the function constructed in previous Theorem, $Q = (-1, 1)^m$. We keep only those summands in the sum

$$\sum_{n \in \mathbb{Z}^m} \varphi_\delta(x - \delta n) = 1$$

for which the cube $\delta(n + Q)$ intersects E, and set

$$\theta_\varepsilon(x) = \sum_{\substack{n \in \mathbb{Z}^m \cdot \\ \delta(n+Q) \cap \dot{E} \neq \varnothing}} \varphi_\delta(x - \delta n)$$

It is clear that $0 \leqslant \theta_\varepsilon(x) \leqslant 1$ everywhere and $\theta_\varepsilon(x) = 1$ if $x \in E$. Moreover, since $\operatorname{diam}(Q) = 2\sqrt{m}$, we have $\theta_\varepsilon(x) = 0$ outside E_ε. Therefore, the function θ_ε satisfies conditions (a)-(c). Now we check the condition (d). Since the sum that defines θ_ε consists of a finite number of summands near every point (as was mentioned after the statement of the previous Theorem), it can be differentiated termwise. It is clear that

$$\|\operatorname{grad} \theta_\varepsilon(x)\| \leqslant \sum_{n \in \mathbb{Z}^m} \|\operatorname{grad} \varphi_\delta(x - \delta n)\|$$

Since every point x belongs to at most 2^m cubes $\delta(n + Q)$, it follows that

$$\begin{aligned} \|\operatorname{grad} \theta_\varepsilon(x)\| &\leqslant 2^m \max_x \|\operatorname{grad} \varphi_\delta(x)\| = \frac{2^m}{\delta} \max_x \left\|\operatorname{grad} \varphi_1\left(\frac{1}{\delta} x\right)\right\| \\ &= \frac{2^m}{\delta} L = \frac{2^{m+1}\sqrt{m}}{\varepsilon} L, \end{aligned}$$

where $L = \max_y \|\operatorname{grad} \varphi_1(y)\|$ does not depend on ε (but does depend on m).□

2.1.8

We conclude this section with a modification of the theorem on the partition of unity. First, we prove a useful geometric fact.

Lemma *Let K be a compact set in the space $\mathbb{R}^m$ and let $\{G_\alpha\}_{\alpha \in A}$ be its open cover. Then there exists a number $\delta > 0$ such that every set e that intersects K*

and has the property $\operatorname{diam}(e) < \delta$ *is contained in at least one set belonging to the cover.*

Proof Assume that the statement of the lemma is false. Then for every $n \in \mathbb{N}$ there exists a set e_n that is not contained in any set G_α and at the same time

$$e_n \cap K \neq \varnothing, \quad \operatorname{diam}(e_n) < \frac{1}{n}$$

Fix a point x_n in each $e_n \cap K$. Without loss of generality, one can consider that $x_n \to x_0$ for some $x_0 \in K$ (otherwise, one may pass to a subsequence). The point x_0 belongs to some set in the family G_α, say, to G_{α_0}. Therefore, $B(x_0, r) \subset G_{\alpha_0}$ for some $r > 0$. If n is large enough, then $\|x_n - x_0\| < r/2$ and $\operatorname{diam}(e_n) < r/2$, whence $e_n \subset B(x_0, r) \subset G_{\alpha_0}$. This shows that the sets e_n with large indices are contained in the set G_{α_0}, contrary to the choice of e_n.□

It is convenient to use the following theorem in those situations where one needs to replace an arbitrary function by functions with supports lying in the prescribed sets.

Theorem (On a partition of unity subordinate to a cover) *Let K be a compact subset of the space $\mathbb{R}^m$ and let $\{G_\alpha\}_{\alpha \in A}$ be its open cover. Then there exists a finite family of non-negative finitary functions $\psi_1, \ldots, \psi_N$ of class $C^\infty(\mathbb{R}^m)$ such that*

$$\sum_{j=1}^{N} \psi_j \leqslant 1 \quad \text{on } \mathbb{R}^m, \quad \sum_{j=1}^{N} \psi_j(x) = 1 \quad \text{if } x \in K$$

and the support of ψ_j is contained in one of the sets that make up the cover for all j.

The family of functions $\psi_1, \ldots, \psi_N$ is called a partition of unity for K subordinate to the cover $\{G_\alpha\}_{\alpha \in A}$.

Proof Let δ be a number from the lemma above corresponding to the given cover. Consider the partition of unity $1 = \sum_{n \in \mathbb{Z}^m} \varphi_\varepsilon(x - \varepsilon n)$ constructed in Theorem (On a smooth partition of unity) taking ε so small that $\operatorname{diam}(\operatorname{supp}(\varphi_\varepsilon)) < \delta$. Keeping only those summands $\varphi_\varepsilon(x - \varepsilon n)$ in the partition of unity whose supports intersect K, we obviously get a finite family, as required, and it only remains to enumerate it.□

2.2 Surface Area

2.2.1

By a k-dimensional area in $\mathbb{R}^m (1 \leqslant k \leqslant m)$ we will understand a Borel measure satisfying properties similar to the properties of the Lebesgue measure λ_k. In particular, on subsets of k-dimensional affine subspaces, this measure must coincide with the Lebesgue measure. This allows us to speak about the area of sets consisting of planar parts, in particular, in the case $k = m - 1$, the faces of polyhedra. However, this does not, of course, suffice for a reasonable definition

of the area of "curvilinear figures", and we must specify some property of area that would allow us to compare its values on non-planar sets (i.e., sets that do not lie in k-dimensional affine subspaces). In our approach, the role of such a condition is played by the following intuitively clear requirement: the area does not increase under a weak contraction. Since the image of a Borel set under a weak contraction (and even under a projection) may not be a Borel set, we will assume that the latter condition applies only to compact sets. So, we adopt the following definition.

Definition *Let $k, m \in \mathbb{N}, 1 \leqslant k \leqslant m$. A measure σ_k defined on the σ-algebra $\mathfrak{B}^m$ of all Borel subsets of $\mathbb{R}^m$ is called a k-dimensional area (in $\mathbb{R}^m$) if it satisfies the following two axioms:*

(I) on every k-dimensional affine subspace L of $\mathbb{R}^m$, the measure σ_k coincides with the restriction of the Lebesgue measure λ_k to the σ-algebra of Borel subsets of L;

(II) on compact sets, σ_k does not increase under weak contractions: if Φ is a weak contraction of a compact set Q, then

$$\sigma_k(\Phi(Q)) \leqslant \sigma_k(Q)$$

We have agreed to call manifolds of codimension 1 surfaces. Hence it is natural to call an $(m-1)$-dimensional area a surface area. However, by abuse of language, we will use this term for a k-dimensional area for arbitrary k. As we will soon see, a k-dimensional area in $\mathbb{R}^m$ exists for all $k = 1, \ldots, m$.

As for every Borel measure, a surface area enjoys the regularity property:

$$\text{if } \sigma_k(E) < +\infty, \quad \text{then } \sigma_k(E) = \sup\{\sigma_k(Q) \mid Q \text{ is a compact set, } Q \subset E\}. \tag{1}$$

By axiom (I), σ_m coincides with the Lebesgue measure λ_m (more precisely, with its restriction to $\mathfrak{B}^m$). Property (II) holds for the Lebesgue measure.

Note that axiom (II) and condition (1) imply that a surface area is invariant under any isometry, since both an isometry and the map inverse to an isometry are weak contractions. In particular, it is invariant under translations and rotations, so that the areas of congruent sets are equal. Under an orthogonal projection (which is, obviously, a weak contraction), the area of a compact set does not increase.

Let us establish another important property of σ_k.

Theorem *The area of a Borel set of finite area does not decrease under an expanding map.*

Proof Let $E \subset \mathbb{R}^m$ be a Borel set of finite area and Θ be an expanding map on E. We already know that, the image $E' = \Theta(E)$ is again a Borel set. If E is a compact set, we can apply axiom (II) to the map Θ^{-1} (since it is a weak contraction), whence $\sigma_k(E) \leqslant \sigma_k(E')$. In the general case, use condition (1).□

We complement the theorem with a simple, but important result. It provides a two-sided bound on the area of a set that has a Lipschitz parametrization. This property will be repeatedly used in what follows.

Lemma *Let $E \subset \mathbb{R}^m$ be a Borel set and Ψ be a map from E to $\mathbb{R}^k$. If there exists a $C > 1$ such that*

$$\frac{1}{C}\|x-y\| \leqslant \|\Psi(x)-\Psi(y)\| \leqslant C\|x-y\| \quad \text{for } x, y \in E$$

then

$$\frac{1}{C^k}\lambda_k(\Psi(E)) \leqslant \sigma_k(E) \leqslant C^k\lambda_k(\Psi(E))$$

Proof Obviously, it suffices to prove the desired inequality for bounded sets. Thus we assume that the set E and, consequently, its image E' are bounded. It follows from the assumptions of the lemma that $H = C\Psi$ and $\Theta : u \mapsto \Psi^{-1}(Cu)$ are expanding maps on E and $\frac{1}{C}E'$, respectively. As we have established, each of these maps sends a Borel set to a Borel set. Since the weak contraction H^{-1} is uniformly continuous, we may assume that it is defined on the compact set $Q = C\overline{E'}$, whose area (coinciding with the Lebesgue measure) is finite. Therefore,

$$\sigma_k(E) = \sigma_k\left(H^{-1}\left(CE'\right)\right) \leqslant \sigma_k\left(H^{-1}(Q)\right) \leqslant \sigma_k(Q) = \lambda_k(Q) < +\infty$$

This allows us to apply the theorem to the expanding map H and obtain the required upper bound:

$$\sigma_k(E) \leqslant \sigma_k(H(E)) = \lambda_k(C\Psi(E)) = C^k\lambda_k(\Psi(E)).$$

On the other hand,

$$\sigma_k(E) = \sigma_k\left(\Theta\left(\frac{1}{C}\Psi(E)\right)\right) \geqslant \sigma_k\left(\frac{1}{C}\Psi(E)\right) = \lambda_k\left(\frac{1}{C}\Psi(E)\right) = \frac{1}{C^k}\lambda_k(\Psi(E)),$$

which yields the lower bound.□

Remark As can be seen from the proof, the lemma remains valid if the codomain of Ψ is not $\mathbb{R}^k$, but a k-dimensional (linear or affine) subspace of a space of arbitrary dimension, where a k-dimensional area coincides with the Lebesgue measure by axiom(I).

2.2.2

The question of the existence of a surface area has been essentially solved. Indeed, the Hausdorff measure μ_k satisfies axiom (II) and is proportional to the Lebesgue measure on k-dimensional subspaces. The proportionality coefficient is equal to the volume α_k of the unit ball in $\mathbb{R}^k$. Therefore, the function $\alpha_k\mu_k$ regarded on all Borel subsets of $\mathbb{R}^m$ is a k-dimensional area. Thus the following theorem holds.

Theorem *For every positive integer $k, 1 \leqslant k \leqslant m$, there is a k-dimensional area in $\mathbb{R}^m$.*

One can show that a Borel measure satisfying conditions (I) and (II) is not unique. The discussion of related subtle results is beyond the scope of this book. Note, however, that the non-uniqueness of area may manifest itself only on quite complicated sets. We will soon see that the area of Borel sets satisfying some natural geometric conditions is uniquely determined.

2.2.3

Now we turn to the one-dimensional case and consider the problem of computing the measure σ_1 on simple arcs, i.e., homeomorphic images of intervals. For brevity, we will also use the term "arc" as a synonym. Of course, it is natural to call $\sigma_1(L)$ the length of an arc L. However, using this term now may cause a certain ambiguity. Indeed, way back in school the reader learned, in the example of a circle, the definition of the length of a curve as the limit of the lengths of inscribed polygonal chains. A natural generalization of this definition leads to the classical definition of arc length. Thus it is desirable to ensure that the measure σ_1 agrees with this definition. Before proceeding to the solution of this problem, we first introduce the notion of the length of a path.

Consider an arbitrary path $\gamma : [a, b] \to \mathbb{R}^m$. Given a partition τ of the interval $[a, b]$ formed by points $t_0 = a < t_1 < \cdots < t_n = b$, set

$$S_\tau = \sum_{i=0}^{n-1} \|\gamma(t_{i+1}) - \gamma(t_i)\|$$

By definition, the length of γ is equal to $s(\gamma) = \sup_\tau S_\tau$. A path is called rectifiable if it has finite length. Note that $s(\gamma) \geqslant \|\gamma(b) - \gamma(a)\|$, since $S_\tau \geqslant \|\gamma(b) - \gamma(a)\|$ by the triangle inequality.

If $\varphi_1, \ldots, \varphi_m$ are the coordinate functions of γ, then, obviously, for any $i = 1, \ldots, m$ and $k = 1, \ldots, n$,

$$|\varphi_i(t_{k+1}) - \varphi_i(t_k)| \leqslant \|\gamma(t_{k+1}) - \gamma(t_k)\| \leqslant \sum_{j=1}^{m} |\varphi_j(t_{k+1}) - \varphi_j(t_k)|$$

Hence, the variations of the functions φ_k satisfy the inequality

$$\mathbf{V}_a^b(\varphi_i) \leqslant s(\gamma) \leqslant \sum_{j=1}^{m} \mathbf{V}_a^b(\varphi_j) \tag{1}$$

Thus a path is rectifiable if and only if all its coordinate functions are of bounded variation. One can see that the path length is additive, i.e., $s(\gamma) = s(\gamma_1) + s(\gamma_2)$, where γ_1, γ_2 are the restrictions of γ to the intervals $[a, c]$ and $[c, b]$, respectively $(a < c < b)$.

To define the classical arc length, we need an easy auxiliary result. For a path, which is a homeomorphism of a line segment onto an arc, we keep the term "parametrization", even though it is defined not on an open interval, as it should be according to Definition 2.1.1, but on a closed interval.

Lemma *The lengths of two parametrizations of a simple arc coincide.*

Proof Let $\gamma : [a,b] \to \mathbb{R}^m, \gamma_1 : [p,q] \to \mathbb{R}^m$ be two parametrizations of a simple arc L. Set $\omega(x) = \gamma_1^{-1}(\gamma(x))(a \leqslant x \leqslant b)$. Then the function ω is continuous, one-to-one, and, consequently, strictly monotone. Therefore, every partition $\tau = \{x_0, \ldots, x_n\}$ of the interval $[a,b]$ gives rise to a partition of the interval $[p,q]$, which is formed by the points $\omega(x_0), \ldots, \omega(x_n)$ if ω is increasing, and by the points $\omega(x_n), \ldots, \omega(x_0)$ if ω is decreasing. Furthermore, $\gamma(x) = \gamma_1(\omega(x))$. Hence

$$S_\tau = \sum_{k=0}^{n-1} \|\gamma(x_{k+1}) - \gamma(x_k)\| = \sum_{k=0}^{n-1} \|\gamma_1(\omega(x_{k+1})) - \gamma_1(\omega(x_k))\| \leqslant s(\gamma_1)$$

But τ is arbitrary, whence $s(\gamma) \leqslant s(\gamma_1)$. Since the parametrizations γ and γ_1 are interchangeable, this means that $s(\gamma) = s(\gamma_1)$.□

Now we define the length of an arc as the common value of the lengths of all its parametrizations. The length of an arc L will be denoted by $s(L)$. Thus $s(L) = s(\gamma)$ if γ is a parametrization of L (using the symbol s both for the length of an arc and the length of its parametrization does not lead to a contradiction). An arc is called rectifiable if it has finite length. Since the length of a path is not less than the distance between its endpoints, the arc length satisfies the clear geometric principle "a line segment is the shortest arc connecting two given points":

$$s(L) \geqslant \|B - A\| \quad \text{if } L \text{ contains } A \text{ and } B. \tag{2}$$

It follows from axiom (II) that this principle can be extended to (any) measure σ_1 :

$$\sigma_1(L) \geqslant \|B - A\| \quad \text{if } L \text{ contains } A \text{ and } B \tag{3}$$

(since the projection of L to the line passing through A and B contains the whole segment connecting them).

Note also that if a path γ is rectifiable, then the function $x \mapsto \theta(x) = s\left(\gamma|_{[a,x]}\right)$ is continuous. Indeed, if $a \leqslant x < y \leqslant b$, then, in view of (1) (hereafter $\varphi_1, \ldots, \varphi_m$ are the coordinate functions of the parametrization γ),

$$|\theta(y) - \theta(x)| = s\left(\gamma|_{[x,y]}\right) \leqslant \sum_{j=1}^{m} \mathbf{V}_x^y(\varphi_j)$$

where $\mathbf{V}_x^y(\varphi_j) \to 0$ as $x - y \to 0$.

The continuity of θ allows us to introduce a new parametrization of a rectifiable arc L. Since the set of values of θ coincides with $[0, S]$ where $S = s(L)$, the map $u \mapsto \delta(u) = \gamma\left(\theta^{-1}(u)\right)$ is defined on $[0, S]$, continuous, one-toone, and satisfies $\delta([0, S]) \subset L$. The reader can easily check that $\delta([0, S]) = L$. Thus δ is a parametrization of L. It follows from the definition of θ that for $0 \leqslant u \leqslant S$ we always have $s(\delta([0, u])) = u$. Furthermore, by the additivity of length, this

parametrization also has the following property: if $0 \leqslant u_1 < u_2 \leqslant S$, then $s(\delta([u_1, u_2])) = u_2 - u_1$. Thus the parameter u in δ has a simple geometric interpretation: the difference of two values u_1, u_2 (where $u_1 < u_2$) is equal to the length of the arc corresponding to the interval $[u_1, u_2]$. This parametrization of a simple arc is called natural. It is a weak contraction on $[0, S]$, since $u_2 - u_1 = s(\delta([u_1, u_2])) \geqslant \|\delta(u_2) - \delta(u_1)\|$ by (2).

Theorem *For every simple arc $L, \sigma_1(L) = s(L)$.*

Proof First we check that $s(L) \leqslant \sigma_1(L)$. Let γ be a parametrization of L defined on $[a, b]$. Consider an arbitrary partition τ of $[a, b]$ formed by points $t_0 = a < t_1 < \cdots < t_n = b$ and set $L_k = \gamma([t_k, t_{k+1}])\,(0 \leqslant k < n)$. Since $\|\gamma(t_{k+1}) - \gamma(t_k)\| \leqslant \sigma_1(L_k)$ by $(2')$, it follows that

$$S_\tau \equiv \sum_{k=0}^{n-1} \|\gamma(t_{k+1}) - \gamma(t_k)\| \leqslant \sum_{k=0}^{n-1} \sigma_1(L_k) = \sigma_1(L)$$

But τ is arbitrary, whence $s(L) \leqslant \sigma_1(L)$.

When proving the reverse inequality, we may assume that L is rectifiable, i.e., $S = s(L) < +\infty$. As we have already noticed, the natural parametrization of L is a weak contraction on $[0, S]$. Hence $\sigma_1(L) \leqslant \sigma_1([0, S]) = S = s(L)$.□

This theorem allows us to call σ_1 the length.

2.2.4

As we have seen in the previous subsection, the σ_1 measure of any simple arc L is equal to the supremum of the lengths of polygonal chains inscribed into L. "Common sense" suggests that in order to compute the area of a curved surface M, we should apply a similar procedure: consider polyhedral surfaces with vertices on M (polyhedra inscribed in M), compute the sum of the areas of their faces, and then take the limit as the faces get smaller and smaller. However, simple analysis shows that this approach cannot lead to a reasonable result even for a cylinder. We briefly sketch the construction of the corresponding classical counterexample, the so-called "Schwarz lantern".

Consider a right circular cylinder of radius R and height H and inscribe into it a polyhedral surface constructed as follows. Cut the cylinder into m equal small cylinders by planes perpendicular to its axis. Into the top and the bottom of each small cylinder inscribe a regular n-gon so that the vertices of the top polygon lie above the middles of the arcs subtended by the sides of the bottom polygon. In other words, the top polygon is rotated by $\frac{\pi}{n}$ with respect to the bottom one. Join each vertex of each polygon with the closest vertices of the polygons one level up or down by line segments. The pair of such segments going from a given vertex to a neighboring polygon, along with the corresponding side of this polygon, forms an isosceles triangle. These triangles together form a polyhedral surface that resembles a Chinese lantern (see Fig. 2.1).

Between two neighboring cross sections there are $2n$ triangular faces (half of them are based on the bottom n-gon, and the other half, on the top n-gon). Thus our polyhedral surface consists of $2mn$ equal triangular faces. Clearly, the

lengths of the sides of these faces tend to zero as $m, n \to \infty$. Observe that the planes of the faces are almost perpendicular to the axis of the cylinder provided that the height of the levels is small compared to the side length of the polygons.

The area s_{mn} of each face is easy to compute:

$$s_{mn} = R \sin \frac{\pi}{n} \sqrt{\left(2R \sin^2 \frac{\pi}{2n}\right)^2 + \left(\frac{H}{m}\right)^2}$$

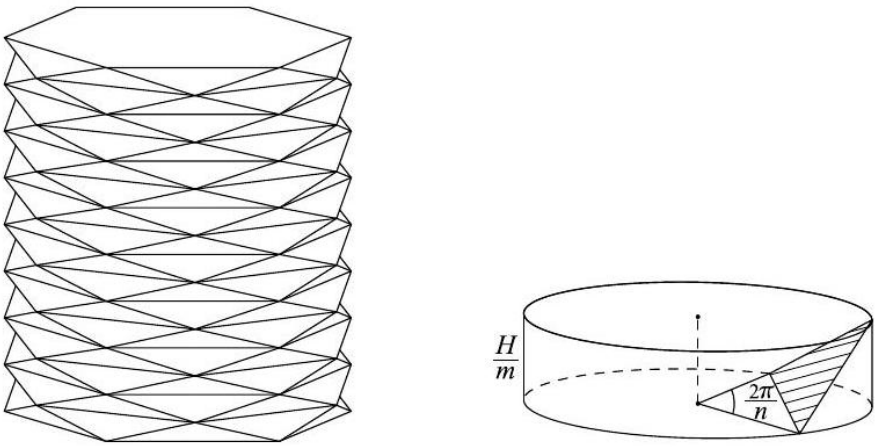

Fig. 2.1 "Schwarz lantern"

Discarding the second term under the square root sign, we have (taking into account that $\sin \varphi \geqslant \frac{2}{\pi}\varphi$ for $\varphi \in \left[0, \frac{2}{\pi}\right]$)

$$s_{mn} \geqslant R \sin \frac{\pi}{n} \cdot 2R \sin^2 \frac{\pi}{2n} \geqslant \frac{4R^2}{n^3}$$

Hence the total area of the polyhedral surface, i.e., $2mns_{mn}$, is not less than $8R^2 \frac{m}{n^2}$. Therefore, we can inscribe into the cylinder a polyhedral surface of arbitrarily large area (taking $m \gg n^2$), though its faces are triangles with arbitrarily small sides. Note that the limit is equal to the area of the cylinder, i.e., $2\pi RH$, only if $\frac{m}{n^2} \to 0$.

The above construction makes it clear why the approach which is effective when computing arc lengths fails if we need to compute areas. The reason is simple: the segments of a polygonal chain inscribed into a smooth curve are almost tangent to the curve provided that their lengths are sufficiently small. For surfaces, the situation is quite different: arbitrarily small faces of a polyhedron inscribed into a curved surface may be almost orthogonal to the surface (the inscribed surface may be "rugged"). Thus, when computing the area of a surface, we should abandon the naive approach related to inscribed polyhedra.

2.3 Properties of the Surface Area of a Smooth Manifold

In this section, all manifolds and parametrizations are assumed smooth by default. Manifolds of codimension one are called surfaces.

2.3.1

Our immediate goal is to obtain a formula for computing the area of Borel subsets of a simple smooth k-dimensional manifold. Then, by the countable additivity of the area, we will be able to compute the areas of countable unions of such sets, in particular, the areas of subsets of arbitrary smooth manifolds.

Let us discuss geometric considerations that suggest what form the desired formula should have. Let $\Phi \in C^1(\mathcal{O}, \mathbb{R}^m)$ be a parametrization of a simple manifold M with $\dim M = k$, and let $\widetilde{\Phi}(t) = \Phi(a) + d_a\Phi(t-a)$ be the linearization of Φ at a point $a \in \mathcal{O}$. The set of values of $\widetilde{\Phi}$ is the k-dimensional affine tangent space, on which the Lebesgue measure λ_k is defined. Consider a cubic cell $Q_h \subset \mathcal{O}$ with a vertex at a and edge length h. Its image under Φ is a "curved parallelotope" $R_h = \Phi(Q_h)$. For small h, it is almost isometric to the corresponding "scale" $\widetilde{R}_h$, the image of Q_h under the linearized map $\widetilde{\Phi}$ (see the lemma in the next section). Hence we should expect that the area of the "curved parallelotope" R_h is close to the Lebesgue measure of the set $\widetilde{R}_h$. It can be obtained by a translation from the parallelotope $d_a\Phi\left([0,h)^k\right)$ lying in the tangent space. Therefore, $\lambda_k\left(\widetilde{R}_h\right) = \lambda_k\left(d_a\Phi\left([0,h)^k\right)\right) = h^k\lambda_k\left(d_a\Phi\left([0,1)^k\right)\right)$. We will call $C_a = d_a\Phi\left([0,1)^k\right)$ the accompanying parallelotope. Thus $\sigma_k(R_h)$ must be close to $h^k\lambda_k(C_a)$, in the sense that their ratio tends to one as h decreases. This suggests that the surface area of a simple smooth manifold M is just a weighted image of the k dimensional Lebesgue measure under Φ, with the weight $\omega_\Phi : t \mapsto \lambda_k(C_t)$ equal to the volume of the accompanying parallelotope.

2.3.2

In order to verify that our heuristic considerations are correct, we first prove a two-sided bound on the deviation of a point of a manifold from the tangent subspace. One of the possible "straightening" maps, which is of a simple geometric nature, could be obtained by orthogonally projecting a sufficiently small neighborhood of the tangency point to the tangent space. However, it is technically more convenient to associate a separate straightening map (close to a projection) with every parametrization.

Lemma *Let $\Phi \in C^1(\mathcal{O}, \mathbb{R}^m)$ be a local parametrization of a manifold M and $\widetilde{\Phi}(t) = \Phi(a) + d_a\Phi(t-a)$ be its linearization at a point $a \in \mathcal{O}$. Then:*

(1) the map $\Psi = \widetilde{\Phi} \circ \Phi^{-1}$ is almost isometric in the vicinity of the point $p = \Phi(a)$: for every $C > 1$, in a sufficiently small M-neighborhood U of p,

$$\frac{1}{C}\|x-y\| \leqslant \|\Psi(x) - \Psi(y)\| \leqslant C\|x-y\| \quad (x, y \in U)$$

(2) for every set $A \in \mathfrak{A}^k$,

$$\lambda_k(\widetilde{\Phi}(A)) = \omega_\Phi(a)\lambda_k(A) \tag{1}$$

Proof It suffices to check that in the vicinity of p the inequality

$$\|(x-y)-(\Psi(x)-\Psi(y))\| \leqslant \left(1-\frac{1}{C}\right)\|x-y\|$$

holds. Since the map Φ^{-1} locally satisfies the Lipschitz condition (see Corollary 1 in Sect. 2.1.4), it suffices to check that for every $\varepsilon > 0$ there exists a small M neighborhood U of the point p such that

$$\|(x-y)-(\Psi(x)-\Psi(y))\| \leqslant \varepsilon \left\|\Phi^{-1}(x)-\Phi^{-1}(y)\right\| \quad (x,y \in U).$$

Setting $s = \Phi^{-1}(x)$ and $t = \Phi^{-1}(y)$, we see that this inequality is equivalent to the condition

$$\|(\Phi(s)-\Phi(t))-d_a\Phi(s-t)\| \leqslant \varepsilon\|s-t\| \quad \text{in the vicinity of } a \tag{2}$$

The latter follows from the smoothness of Φ. Indeed, let r be so small that $\|d_u\Phi - d_a\Phi\| \leqslant \varepsilon$ for every u from the k-dimensional ball $B(a,r)$. Then, by Corollary from Lagrange's inequality (we are not discuss it on our course),

$$\|(\Phi(s)-\Phi(t))-d_a\Phi(s-t)\| \leqslant \sup_{u\in B(a,r)} \|d_u\Phi - d_a\Phi\|\,\|s-t\|$$

which implies (2). Thus, as a desired M-neighborhood U of p, we can take $\Phi(B(a,r))$ provided that the radius r is sufficiently small.

To prove (1), observe that the measure $A \mapsto \lambda_k(\widetilde{\Phi}(A))$ is translation-invariant and hence proportional to λ_k. Since $\lambda_k\left(\widetilde{\Phi}\left([0,1)^k\right)\right) = \lambda_k\left(C_a\right) = \omega_\Phi(a)$, the proportionality coefficient is equal to $\omega_\Phi(a)$.□

Now we are in a position to prove a formula for computing the area of a set lying on a smooth manifold. The idea of the proof is the same as in Theorem for Change of Variable in a Multiple Integral.

Theorem *For every Borel set E contained in a simple smooth manifold M,*

$$\sigma_k(E) = \int_{\Phi^{-1}(E)} \omega_\Phi(t)dt \tag{3}$$

where Φ is an arbitrary smooth parametrization of M.

As we mentioned before, the axioms of area do not uniquely determine it on all Borel sets. In contrast, the above theorem shows that on sufficiently "good" sets—smooth manifolds and their Borel subsets—all areas coincide. Thus the difference between various areas may manifest itself only on Borel sets of quite a complicated nature.

Proof Let $\mathcal{O}$ be the open set on which the parametrization Φ is defined, and consider the measure $v(A) = \sigma_k(\Phi(A))$ on Borel subsets of $\mathcal{O}$. We will verify that it satisfies the condition

$$\inf_A \omega_\Phi \lambda_k(A) \leqslant v(A) \leqslant \sup_A \omega_\Phi \lambda_k(A) \tag{4}$$

this implies that $v(A) = \int_A \omega_\Phi(t)dt$, which is equivalent to the desired assertion.

If these inequalities hold for sets forming an increasing sequence, then they obviously hold for the union of these sets. Hence it suffices to prove (4) assuming that A is a bounded set whose closure is contained in $\mathcal{O}$. Both inequalities (4) are proved in the same way, so we will prove only the upper bound, leaving the reader to carry out a similar argument for the lower bound.

If the right inequality in (4) is false, then for some $C_0 > 1$ we have

$$\nu(A) > C_0 \sup_A \omega_\Phi \lambda_k(A) \tag{5}$$

Divide A into finitely many parts with the diameter of each part at most diam $(A)/2$. Then (5) must hold for one of these parts, which we denote by A_1. Replacing A by A_1 and repeating the argument, we obtain a set A_2, etc. By induction, we will construct a sequence of nested sets A_n with diameters tending to zero. Take a point a in the intersection $\bigcap_{n=1}^{\infty} \bar{A}_n$. By the construction of the sets A_n, they satisfy (5):

$$v\left(A_n\right) > C_0 \sup_{A_n} \omega_\Phi \lambda_k\left(A_n\right) \tag{6}$$

Let us show that this leads to a contradiction. By the lemma, for every $C > 1$ (to be specified later) there exists a neighborhood V of a such that for $x, y \in U = \Phi(V)$ we have

$$\frac{1}{C}\|x - y\| \leqslant \|\Psi(x) - \Psi(y)\| \leqslant C\|x - y\|$$

where, as in the lemma, $\Psi = \widetilde{\Phi} \circ \Phi^{-1}$ and $\widetilde{\Phi}$ is the linearization of Φ (i.e., $\widetilde{\Phi}(t) = \Phi(a) + d_a\Phi(t-a)\Big)$. If n is so large that $A_n \subset V$, then $\Phi\left(A_n\right) \subset U$. Hence, by Lemma 8.2.1, for the set $E_n = \Phi\left(A_n\right)$ we have $\sigma_k\left(E_n\right) \leqslant C^k \lambda_k\left(\Psi\left(E_n\right)\right)$. Since $\Psi\left(E_n\right) = \widetilde{\Phi}\left(A_n\right)$ and, according to (1), $\lambda_k\left(\widetilde{\Phi}\left(A_n\right)\right) = \omega_\Phi(a)\lambda_k\left(A_n\right)$, it follows that

$$\begin{aligned} v\left(A_n\right) &= \sigma_k\left(\Phi\left(A_n\right)\right) = \sigma_k\left(E_n\right) \leqslant C^k \lambda_k\left(\Psi\left(E_n\right)\right) = C^k \lambda_k\left(\widetilde{\Phi}\left(A_n\right)\right) \\ &\leqslant C^k \sup_{A_n} \omega_\Phi \lambda_k\left(A_n\right). \end{aligned}$$

From (6) and the last inequality we see that $1 < C_0 \leqslant C^k$. However, this is not possible if C is chosen sufficiently close to 1 . Therefore, our assumption is false, and the theorem follows.□

A special case of this result (for $k = m$) is Theorem on the behavior of the Lebesgue measure under a diffeomorphism, since in this case $\omega_\Phi = J_\Phi \equiv$

$|\det(\Phi')|$. The proofs of these theorems are similar, but here we have used the properties of area, which has allowed us not to keep track of the measures of the images of small cubic cells.

2.3.3

Here we will discuss the basic properties of the area σ_k in the space $\mathbb{R}^m$, always assuming that $k < m$. For brevity, we will call it just the area, omitting the reference to the dimension.

It is clear that the properties of σ_k substantially differ in some respects from the familiar properties of the Lebesgue measure. For example, since every m-dimensional cube contains a continuum of congruent pairwise disjoint k dimensional cubes, in $\mathbb{R}^m$ there are compact sets of infinite area. It is also clear that every non-empty open set has infinite area. This fact immediately implies that the area is not σ-finite and cannot be a regular measure. Thus, when studying the properties of σ_k in more detail, we will rather consider not the area "as a whole", but its restrictions to subsets contained in a fixed manifold. To be more precise, we introduce the following notation related to a smooth k-dimensional manifold $M \subset \mathbb{R}^m$ (with $k < m$). By $\mathfrak{B}_M$ we denote the system of all Borel sets contained in M, and by σ_M the restriction of the k-dimensional area to $\mathfrak{B}_M$. Since M itself is a Borel set, $\mathfrak{B}_M$ is a σ-algebra and σ_M is a measure.

Formula (3) allows us to compute the area of a Borel subset of a simple manifold. It is obviously valid also for subsets of a coordinate neighborhood of an arbitrary, not necessarily simple, manifold provided that Φ is the corresponding parametrization.

Let us establish several important properties of the area.

(1) The area of a compact subset of a smooth manifold is finite.

For a compact subset of a coordinate neighborhood, this follows from (3), since its inverse image under every parametrization is a compact set and the weight ω_Φ is continuous. An arbitrary compact subset of a manifold can be covered by finitely many coordinate neighborhoods of finite area.

(2) The measure σ_M is σ-finite.

This property follows from the previous one and Corollary 2 from Sect. 2.1.5. Since Borel sets of zero measure may have non-Borel subsets, the measure σ_M is not complete. To obtain a complete measure, we should consider its Carathéodory extension. The σ-algebra on which it is defined will be denoted by $\mathfrak{A}_M$, and the elements of $\mathfrak{A}_M$ will be called Lebesgue measurable or, in short, measurable. The extension of σ_M to $\mathfrak{A}_M$ is unique. For this extension we keep the old notation σ_M and still call it the area (of the manifold M). Every measurable set can be approximated from the inside and from the outside by Borel sets of the same measure.

If Φ is a parametrization of a simple manifold M and $E \subset M$ is an arbitrary Lebesgue measurable set, then formula (3), established for Borel sets, remains valid. Thus:

(3) the measure σ_M on a simple k-dimensional manifold M is a weighted image of the k-dimensional Lebesgue measure with respect to an arbitrary parametrization Φ. A subset of a simple manifold is measurable if and only if its inverse image is measurable.

Indeed, a measurable set E can be written in the form $E = A \cup e$, where A is a Borel set and $\sigma_M(e) = 0$. Besides, $e \subset e'$, where e' is a Borel set of zero area. Hence $\Phi^{-1}(e) \subset \Phi^{-1}(e')$ and, moreover, $\lambda_k\left(\Phi^{-1}(e')\right) = 0$. The latter holds, since $0 = \sigma_M(e') = \int_{\Phi^{-1}(e')} \omega_\Phi(t)dt$ and $\omega_\Phi > 0$. By the completeness of the Lebesgue measure, the set $\Phi^{-1}(e)$ is measurable (and has zero measure). Therefore,

$$\sigma_M(E) = \sigma_M(A) = \int_{\Phi^{-1}(A)} \omega_\Phi(t)dt = \int_{\Phi^{-1}(E)} \omega_\Phi(t)dt$$

In a similar way one can show that the measurability of $\Phi^{-1}(E)$ implies the measurability of E.

(4) The measure σ_M is regular, i.e.

$$\sigma_M(E) = \inf_{\substack{E \subset G \subset M \\ G \text{ is open in } M}} \sigma_M(G) = \sup_{\substack{E \supset F \\ F \text{ is compact set}}} \sigma_M(F)$$

for every measurable set $E, E \subset M$.

If M is a simple manifold, then this property is an immediate consequence of the regularity of the Lebesgue measure and formula (3). We leave the reader to prove it in the case of an arbitrary manifold.

(5) Let L be an arbitrary smooth manifold, $\dim L < k$. *Then $\sigma_k(L) = 0$.*

Let us show that every point from L has a neighborhood U such that the intersection $L \cap U$ has zero area (this suffices because, by Lindelöf's theorem, L can be covered by countably many such neighborhoods).

According to the second definition of a smooth manifold, U can be chosen in such a way that $L \cap U$ is contained in a simple smooth manifold M of dimension k. Moreover, we may assume without loss of generality that $\dim L = k - 1$.

Let Φ be a parametrization of M. Then $\Phi^{-1}(L)$ is a smooth surface in $\mathbb{R}^k$, which can be written as a countable union of graphs of smooth functions (see Corollary 4 in Sect. 2.1.5). Since the k-dimensional volume of every such graph vanishes, we have $\lambda_k\left(\Phi^{-1}(L)\right) = 0$. Thus $\sigma_k(L) = \int_{\Phi^{-1}(L)} \omega_\Phi(t)dt = 0$.

It follows, for example, that when computing the area of a subset of a sphere, we may discard manifolds of smaller dimension. This allows us to assume without loss of generality that the set under consideration is contained in the "cut" sphere and use the corresponding parametrization and formula (3).

(6) Under the homothety with ratio $a > 0$, the measure of a set contained in a k-dimensional manifold M is multiplied by a^k : $\sigma_k(aE) = a^k\sigma_k(E)$ if $E \in \mathfrak{A}_M$.

Indeed, the area is proportional to the Hausdorff measure, which has the desired property.

Note that in the case of an arbitrary linear transformation of a manifold M, the measures on M and on its image do not have such a simple relation. To see

this, it suffices to consider compressing a circle: a simple calculation shows that the length of an arc of an ellipse with eccentricity $\varepsilon \neq 1$ can be expressed via the elliptic integral $\int_0^{\varphi} \sqrt{1-\varepsilon^2 \sin^2 t}\, dt$, which is not an elementary function.

In conclusion, we state the property of the area mentioned immediately after Definition 2.2.1.

(7) The area is invariant under isometries.

In particular, the area on a sphere is rotation-invariant.

2.3.4

In order to compute the areas of manifolds via formula (3), we need explicit expressions for the weight $\omega_\Phi(t)$ equal to the measure of the accompanying parallelotope C_t. For $k = 1, C_t$ is just the line segment with endpoints 0 and $\Phi'(t)$. Hence $\omega_\Phi(t) = \lambda_1(C_t) = \|\Phi'(t)\|$, so that, in order to compute, for example, the length of an arc $L = \Phi([a,b])$, we should integrate the length of the tangent vector: $\sigma_1(L) = \int_a^b \|\Phi'(t)\|\, dt$.

In the general case, the parallelotope C_t is spanned by the canonical tangent vectors $\tau_1 = \tau_1(t), \ldots, \tau_k = \tau_k(t)$ corresponding to the parametrization Φ. Since they are linearly independent, the volume of C_t is positive, i.e., we always have $\omega_\Phi(t) > 0$. The value $\omega_\Phi(t)$ can be computed via the Gram determinant:

$$\omega_\Phi(t) = \lambda_k(C_t) = \sqrt{\Gamma(\tau_1, \ldots, \tau_k)} = \sqrt{\det\left(\left(\langle \tau_i, \tau_j \rangle\right)_{i,j=1}^k\right)}$$
$$= \sqrt{\det\left[\left(\Phi'(t)\right)^T \Phi'(t)\right]}.$$

It follows from the Binet-Cauchy formula that

$$\omega_\Phi^2(t) = \det\left[\left(\Phi'(t)\right)^T \Phi'(t)\right] = \sum_{j_1 < j_2 < \cdots < j_k} \left(\frac{D(\varphi_{j_1}, \varphi_{j_2}, \ldots, \varphi_{j_k})}{D(t_1, t_2, \ldots, t_k)}(t)\right)^2$$

where $\varphi_1, \varphi_2, \ldots, \varphi_m$ are the coordinate functions of Φ.

For a surface, i.e., in the case $k = m-1$, the expression for $\omega_\Phi(t)$ provided by the Binet-Cauchy formula simplifies:

$$\omega_\Phi^2(t) = \sum_{j=1}^m \left(\frac{D(\varphi_1, \ldots, \widehat{\varphi_j}, \ldots, \varphi_m)}{D(t_1, \ldots, t_{m-1})}(t)\right)^2$$

(the symbol $\widehat{\ }$ indicates that the corresponding function is omitted)

The right-hand side has a simple geometric interpretation. Let $e_1, \ldots, e_m$ be the canonical basis in $\mathbb{R}^m$. Consider the vector

$$N_\Phi(t) = \sum_{j=1}^m (-1)^{j+1} \frac{D(\varphi_1, \ldots, \widehat{\varphi_j}, \ldots, \varphi_m)}{D(t_1, \ldots, t_{m-1})}(t) \cdot e_j$$

It can be written via the symbolic determinant

$$N_\Phi(t) = \begin{vmatrix} e_1 & e_2 & \dots & e_m \\ \frac{\partial \varphi_1(t)}{\partial t_1} & \frac{\partial \varphi_2(t)}{\partial t_1} & \dots & \frac{\partial \varphi_m(t)}{\partial t_1} \\ \frac{\partial \varphi_1(t)}{\partial t_2} & \frac{\partial \varphi_2(t)}{\partial t_2} & \dots & \frac{\partial \varphi_m(t)}{\partial t_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \varphi_1(t)}{\partial t_{m-1}} & \frac{\partial \varphi_2(t)}{\partial t_{m-1}} & \dots & \frac{\partial \varphi_m(t)}{\partial t_{m-1}} \end{vmatrix}$$

whose rows, except for the first one, consist of tangent vectors.

The vector $N_\Phi(t)$ is orthogonal to each tangent vector $\tau_j(t)$, since the inner product $\langle N_\Phi(t), \tau_j(t)\rangle$ can be written as a determinant with two equal rows. Thus $N_\Phi(t)$ is a normal to M at the point $\Phi(t)$. It will be called the normal corresponding to the parametrization Φ. Obviously, the length of $N_\Phi(t)$ is equal to $\omega_\Phi(t)$.

For $m = 3, k = 2$, we see that $N_\Phi(t)$ is simply the vector product of the tangent vectors: $N_\Phi(t) = \tau_1(t) \times \tau_2(t)$. Clearly,

$$\omega_\Phi^2 = \begin{vmatrix} \langle \tau_1, \tau_1\rangle & \langle \tau_1, \tau_2\rangle \\ \langle \tau_2, \tau_1\rangle & \langle \tau_2, \tau_2\rangle \end{vmatrix} = EG - F^2$$

where E, F, G are the coefficients of the first fundamental form of the surface, i.e., $E = \|\tau_1\|^2, G = \|\tau_2\|^2$ and $F = \langle \tau_1, \tau_2\rangle$. If the tangent vectors τ_1 and τ_2 are orthogonal, then $\omega_\Phi = \|\tau_1\| \cdot \|\tau_2\|$.

If $M = \Gamma_f$ is the graph of a function f from $C^1(\mathcal{O}, \mathbb{R})$, then the map

$$\mathcal{O} \ni x = (x_1, \dots, x_{m-1}) \mapsto \Phi(x) = (x, f(x))$$

is its canonical parametrization, and

$$\frac{D(\varphi_1, \dots, \widehat{\varphi}_j, \dots, \varphi_m)}{D(x_1, \dots, x_{m-1})}(x) = (-1)^{m+j+1} \frac{\partial f}{\partial x_j}(x)$$

for $1 \leqslant j \leqslant m-1$. If $j = m$, then this determinant is equal to one. Hence $N_\Phi(x) = (-1)^m \left(f'_{x_1}(x), \dots, f'_{x_{m-1}}(x), -1\right)$ and $\omega_\Phi(x) = \sqrt{1 + \|\operatorname{grad} f(x)\|^2}$.

The density ω_Φ corresponding to the canonical parametrization can easily be computed directly from geometric considerations, without using the general formula. Indeed, the tangent vectors corresponding to the canonical parametrization of the graph are $\tau_j = \left(0, \dots, 0, 1, 0, \dots, 0, f'_{x_j}(x)\right)$. The projection to $\mathbb{R}^{m-1}$ sends the accompanying parallelotope C_x spanned by these vectors to the unit cube $[0,1]^{m-1}$. Hence $\lambda_{m-1}(C_x) = \frac{1}{|\cos\theta(x)|}$, where $\theta(x)$ is the angle between the last coordinate axis and an arbitrary normal to the tangent plane $T_{(x,f(x))}$. As is well known (see Example 1 in Sect. 2.1.3), one such normal is the vector $v(x) = \left(-f'_{x_1}(x), \dots, -f'_{x_{m-1}}(x), 1\right)$ (observe that the normals $N_\Phi(x)$ and $v(x)$ coincide for odd m and are opposite for even m). Therefore,

$$|\cos\theta(x)| = \frac{\langle v(x), e_m\rangle}{\|v(x)\| \, \|e_m\|} = \frac{1}{\|v(x)\|}$$

Hence

$$\omega_\Phi(x) = \lambda_{m-1}(C_x) = \frac{1}{|\cos\theta(x)|} = \|v(x)\| = \sqrt{1 + \|\operatorname{grad} f(x)\|^2}$$

Thus the area of every set E contained in the graph Γ_f can be computed by the formula

$$\sigma_{\Gamma_f}(E) = \int_{\Phi^{-1}(E)} \frac{dx}{|\cos\theta(x)|} = \int_{P(E)} \sqrt{1 + \|\operatorname{grad} f(x)\|^2}\, dx \tag{7}$$

where $P(E)$ is the orthogonal projection of E to $\mathbb{R}^{m-1}$.

In particular, if $f(x) = \sqrt{R^2 - \|x\|^2}$ for $x \in B^{m-1}(R)$, then Γ_f is the upper hemisphere $S_+^{m-1}(R)$. Since the radius vector of a point on the sphere $S^{m-1}(R)$ is a normal to $S^{m-1}(R)$, we have $\cos\theta = f(x)/R$. Hence for the area of a measurable set E lying on the hemisphere $S_+^{m-1}(R)$ we obtain the formula

$$\sigma_{\Gamma_f}(E) = \int_{P(E)} \frac{1}{|\cos\theta(x)|} dx = \int_{P(E)} \frac{R}{\sqrt{R^2 - \|x\|^2}} dx. \tag{8}$$

2.3.5

Now we consider several examples.

Example 1 (The area of subsets of a two-dimensional sphere) According to Property (5) from Sect. 2.3.3, when computing the areas of subsets of the sphere $S^2(R) = \{(x, y, z) \mid x^2 + y^2 + z^2 = R^2\}$, one may discard smooth curves. We introduce spherical coordinates φ, θ, i.e., consider the map

$$(\varphi, \theta) \mapsto \Phi(\varphi, \theta) = (R\cos\varphi\cos\theta, R\sin\varphi\cos\theta, R\sin\theta) \in S^2$$
$$\left(|\varphi| < \pi, |\theta| < \frac{\pi}{2}\right)$$

It provides a parametrization of the sphere cut along the meridian $\varphi = \pm\pi$. The coordinate lines for this parametrization are meridians and parallels. This allows one to easily compute the approximate area of the small quadrilateral bounded by the coordinate lines corresponding to the angles $\varphi, \varphi + h$ and $\theta, \theta + h (h > 0)$. Since the meridians and the parallels are orthogonal, the accompanying parallelotope is a rectangle whose side lengths for small h are almost equal to the lengths of the arcs bounding the curved quadrilateral. The latter are circular arcs of radius R (along the meridian) and $R\cos\theta$ (along the parallel). Hence the area of the accompanying parallelogram is approximately equal to $\left(R^2\cos\theta\right)h^2$. This suggests that the weight ω_Φ corresponding to the parametrization in question is equal to $R^2\cos\theta$. We leave the reader to check that the obtained result is correct by finding the tangent vectors and computing the corresponding Gram determinant.

Knowing the weight, one can easily find the area of a set lying on the sphere. For simplicity, consider a spherical quadrilateral Q bounded by parallels and meridians:

$$Q = \left\{ \Phi(\varphi, \theta) \mid -\pi \leqslant \alpha_1 < \varphi < \alpha_2 \leqslant \pi, -\frac{\pi}{2} \leqslant \beta_1 < \theta < \beta_2 \leqslant \frac{\pi}{2} \right\}.$$

Obviously,

$$\sigma_2(Q) = \int_{\alpha_1}^{\alpha_2} \int_{\beta_1}^{\beta_2} R^2 \cos\theta d\varphi d\theta = R^2 \left(\sin\beta_2 - \sin\beta_1\right)\left(\alpha_2 - \alpha_1\right)$$

For the extreme values of $\alpha_1, \alpha_2, \beta_1, \beta_2$, we obtain the area of the whole sphere: $\sigma_2\left(S^2(R)\right) = 4\pi R^2$.

Example 2 (The area of subsets of a two-dimensional torus) According to Property (5) from Sect. 2.3.3, when computing the areas of subsets of the torus $T^2 = \left\{ (x, y, z) \mid \left(R - \sqrt{x^2 + y^2}\right)^2 + z^2 = r^2 \right\}$, we may discard smooth curves. Let us find the weight corresponding to the parametrization of the cut torus considered in Example 3 from Sect. 2.1.3. Recall that this parametrization Φ is as follows:

$$\Phi(\varphi, \theta) = ((R + r\cos\theta)\cos\varphi, (R + r\cos\theta)\sin\varphi, r\sin\theta)$$

where $\varphi, \theta \in (-\pi, \pi)$.

As in the case of a sphere, the coordinate lines (the analogs of parallels and meridians) form two families of orthogonal circles. Computing, as in Example 1, the area of a small curved quadrilateral bounded by coordinate lines, we arrive at the plausible conclusion that the weight corresponding to the chosen parametrization has the form $\omega_\Phi(\varphi, \theta) = r(R + r\cos\theta)$. The reader can easily perform all necessary formal calculations.

It is clear that the area of the curved quadrilateral on the torus bounded by "meridians" $\varphi = \alpha_1, \varphi = \alpha_2$ and "parallels" $\theta = \beta_1, \theta = \beta_2$ is equal to

$$\int_{\alpha_1}^{\alpha_2} \int_{\beta_1}^{\beta_2} r(R + r\cos\theta) d\varphi d\theta = r\left(R\left(\beta_2 - \beta_1\right) + r\left(\sin\beta_2 - \sin\beta_1\right)\right)\left(\alpha_2 - \alpha_1\right)$$

For the extreme values of $\alpha_1, \alpha_2, \beta_1, \beta_2$, we obtain the area of the whole torus: $\sigma_2\left(T^2\right) = 4\pi^2 rR$.

Example 3 (The area of subsets of a conic surface) Let us find out how the area of a set E lying on the conic surface $\left\{(x, y) \mid x \in \mathbb{R}^{m-1}, y = c\|x\|\right\}$ is related to the area of its projection $P(E)$ to the plane $y = 0$. If $0 \notin E$, then E lies on a smooth surface, the graph of the function $f(x) = c\|x\|$ defined on $\mathbb{R}^{m-1}\backslash\{0\}$. One can easily see that $\|\operatorname{grad} f(x)\| = |c|$ everywhere (the angle between the tangent plane and the plane $y = 0$ is constant). Hence

$$\sigma_{m-1}(E) = \int_{P(E)} \sqrt{1 + \|\operatorname{grad} f(x)\|^2} dx = \sqrt{1 + c^2}\lambda_{m-1}(P(E)).$$

We now consider several examples related to the multi-dimensional sphere.

Example 4 (The area of a multi-dimensional sphere) To compute the area of a sphere $S^{m-1}(R)$ of arbitrary radius, it suffices to compute the area of the unit sphere S^{m-1}, since $S^{m-1}(R) = RS^{m-1}$ and, consequently (see Property (6) in Sect. 2.3.3), $\sigma_{m-1}\left(S^{m-1}(R)\right) = R^{m-1}\sigma_{m-1}\left(S^{m-1}\right)$. The area of the unit sphere has already been computed using the fact that it consists of two hemispheres each of which is the graph of a smooth function. We will not reproduce these calculations, but only write down the formula they lead to:

$$\sigma_{m-1}\left(S^{m-1}(R)\right) = \frac{2\pi^{m/2}}{\Gamma(m/2)}R^{m-1}$$

The right-hand side is equal to $m\alpha_m R^{m-1} = \left(\alpha_m R^m\right)'_R$. Hence the area of a sphere is equal to the derivative (with respect to the radius) of the volume of the corresponding ball. Later (see Remark (3) in Sect. 2.4.3) we will discuss this question in more detail.

Let us also compute the area of the spherical "cap" cut from the sphere $S^{m-1}(R)$ by a plane at distance $H(0 \leqslant H < R)$ from its center (for $H = 0$, we thus obtain a hemisphere). Since the area is rotation-invariant, we may say that the set in question is

$$S_H(R) = \left\{(x, y) = (x_1, \ldots, x_{m-1}, y) \in S^{m-1}(R) \mid y > H\right\}$$

It is obviously the part of the graph of the function $x \mapsto f(x) = \sqrt{R^2 - \|x\|^2}$ that lies above the ball $B^{m-1}\left(\sqrt{R^2 - H^2}\right)$. Formula $(7')$ yields

$$\sigma_{m-1}\left(S_H(R)\right) = \int_{B^{m-1}\left(\sqrt{R^2-H^2}\right)} \frac{R}{\sqrt{R^2 - \|x\|^2}}dx$$

Now we use the formula for computing the integral of a radial function:

$$\begin{aligned}\sigma_{m-1}\left(S_H(R)\right) &= (m-1)\alpha_{m-1}R\int_0^{\sqrt{R^2-H^2}} \frac{u^{m-2}du}{\sqrt{R^2-u^2}} \\ &= (m-1)\alpha_{m-1}R^{m-1}\int_0^{\sqrt{R^2-H^2}/R} \frac{v^{m-2}dv}{\sqrt{1-v^2}}\end{aligned} \tag{9}$$

Here $\alpha_{m-1} = \lambda_{m-1}\left(B^{m-1}\right) = \pi^{(m-1)/2}/\Gamma((m+1)/2)$. Setting $H = \delta R$, we can rewrite (9) as

$$\begin{aligned}\sigma_{m-1}\left(S_{\delta R}(R)\right) &= (m-1)\alpha_{m-1}R^{m-1}\int_0^{\sqrt{1-\delta^2}} \frac{v^{m-2}dv}{\sqrt{1-v^2}} \\ &= (m-1)\alpha_{m-1}R^{m-1}\int_{\arcsin\delta}^{\pi/2} \cos^{m-2}t\,dt\end{aligned}$$

Now we find out what part of the multi-dimensional sphere falls into the spherical cap as the dimension m increases indefinitely while the distance from the plane cutting off the cap to the center of the sphere remains constant (to

simplify the formulas, we consider spheres in $\mathbb{R}^{m+1}$ rather than in $\mathbb{R}^m$). It is particularly instructive to consider this question in the two cases where the sphere radius is equal to one (in all dimensions) and where it is proportional to $\sqrt{m}$.

First, we consider the case of the unit sphere.

Example 5 In a space of large dimension, we have the "concentration of measure" phenomenon: almost all of the area of a sphere is concentrated in an arbitrarily narrow zone near the "equator". More precisely, for the caps $S_\delta = S_\delta(1)$, the ratio $\sigma_m(S_\delta)/\sigma_m(S_0)$ decays rapidly as the dimension grows:

$$\frac{\sigma_m(S_\delta)}{\sigma_m(S_0)} = \frac{\int_{\arcsin\delta}^{\pi/2} \cos^{m-1} t dt}{\int_0^{\pi/2} \cos^{m-1} t dt} < 3e^{-\frac{m\delta^2}{2}} \tag{10}$$

To prove (10), let us first estimate the denominator.

$$W_m \equiv \int_0^{\pi/2} \cos^m t dt = \frac{(m-1)!!}{m!!} v_m$$

where v_m is equal to 1 or $\pi/2$ depending on the parity of m. Hence $W_m W_{m+1} = \frac{\pi}{2m+2}$. Since the integrals W_m decrease, it follows that

$$\sqrt{\frac{\pi}{2m+2}} < W_m < \sqrt{\frac{\pi}{2m}}$$

To estimate the numerator, apply the inequalities $\delta \leqslant \arcsin\delta$ and $\cos t \leqslant e^{-t^2/2}$:

$$\begin{aligned} W_m(\delta) \equiv \int_{\arcsin\delta}^{\pi/2} \cos^m t dt &\leqslant \int_\delta^{\pi/2} e^{-\frac{m}{2}t^2} dt < \int_0^\infty e^{-\frac{m}{2}(t+\delta)^2} dt \\ &\leqslant \sqrt{\frac{\pi}{2m}} e^{-\frac{m}{2}\delta^2} \end{aligned}$$

Thus for $m > 1$ we have

$$\frac{\sigma_m(S_\delta)}{\sigma_m(S_0)} = \frac{W_{m-1}(\delta)}{W_{m-1}} < \sqrt{\frac{m}{m-1}} e^{-\frac{m-1}{2}\delta^2} \leqslant \sqrt{\frac{me}{m-1}} e^{-\frac{m}{2}\delta^2} < 3e^{-\frac{m\delta^2}{2}}.$$

Observe that, having replaced $\arcsin\delta$ with δ, we have estimated the area of the cap determined by the inequality $y \geqslant \sin\delta$, which is a little larger than S_δ. Such a set appears naturally if we replace the Euclidean metric on the sphere by the stronger geodesic metric. In the latter metric, the distance between two points of the sphere is equal to the angle between their radius vectors. It is clear that the larger cap consists of the points for which the deviation from the equator (the intersection of the sphere with the plane $y = 0$) is at least δ in the geodesic metric.

Example 6 Now let us discuss the second of the questions posed above: as m grows, how does the ratio of the area of the cap $S_H(R)$ to the area of the

ambient sphere $S^m(R)$ behave under the condition that $R = R_m = \theta\sqrt{m}$, where $\theta > 0$. Set $P_m(H) = \sigma_m\left(S_H(R)\right)/\sigma_m\left(S^m(R)\right)$. This value may be regarded as the probability that a point "picked at random" from the sphere falls into the cap. One may also say that $P_m(H)$ is the probability that the last coordinate of a point picked at random from the sphere is greater than H.

It follows from (9) (with m replaced by $m+1$) that

$$P_m(H) = \frac{m\alpha_m}{(m+1)\alpha_{m+1}} \int_0^{\sqrt{1-H^2/(m\theta^2)}} \frac{t^{m-1}dt}{\sqrt{1-t^2}}$$

Since $\Gamma\left(x+\frac{1}{2}\right) \sim \sqrt{x}\Gamma(x)$ as $x \to +\infty$ (see formula (4) in Sect. 7.2.2), we have

$$\frac{m\alpha_m}{(m+1)\alpha_{m+1}} = \frac{m}{m+1}\frac{\pi^{m/2}}{\Gamma\left(\frac{m+2}{2}\right)}\frac{\Gamma\left(\frac{m+3}{2}\right)}{\pi^{\frac{m+1}{2}}} \underset{m\to\infty}{\sim} \sqrt{\frac{m}{2\pi}}$$

whence

$$P_m(H) \underset{m\to\infty}{\sim} \sqrt{\frac{m}{2\pi}} \int_0^{\sqrt{1-H^2/(m\theta^2)}} \frac{t^{m-1}}{\sqrt{1-t^2}}dt$$

Let us see how the last integral behaves, setting for brevity $\delta = H/\theta$. Making the substitution $v = \sqrt{m}\sqrt{1-t^2}$, we obtain

$$\sqrt{m}\int_0^{\sqrt{1-\delta^2/m}} \frac{t^{m-1}}{\sqrt{1-t^2}}dt = \int_\delta^{\sqrt{m}} \left(1-\frac{v^2}{m}\right)^{\frac{m-2}{2}} dv \underset{m\to\infty}{\longrightarrow} \int_\delta^\infty e^{-v^2/2}dv$$

Therefore,

$$P_m(H) \underset{m\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int_{H/\theta}^\infty e^{-v^2/2}dv = \frac{1}{\theta\sqrt{2\pi}}\int_H^\infty e^{-t^2/\left(2\theta^2\right)}dt$$

This limit may be interpreted as the probability that the Gaussian random variable with density $\frac{1}{\theta\sqrt{2\pi}}e^{-t^2/\left(2\theta^2\right)}$ takes a value greater than H. This result, sometimes called Poincaré's or Maxwell's lemma, means that with respect to the normalized surface area of the m-dimensional sphere of radius $\theta\sqrt{m}$, the distribution of the coordinates is "almost Gaussian".

2.3.6

Now we are going to discuss a more special question, namely, the behavior of the area under a bending. By a bending of a manifold M one usually means a transformation under which the lengths of curves lying on M do not change. For our purposes, this sense is too wide, since under such a map a smooth manifold may transform into a set that is not a manifold. For instance, an interval of the real axis can be bent into a "figure-of-eight" (the continuous map $t \mapsto ((1-\cos t)\operatorname{sign} t, \sin t)$ transforms the interval $(-2\pi, 2\pi)$ into a pair

of touching circles; it is one-to-one, but not homeomorphic). In addition, we continue to restrict ourselves to smooth maps. Thus it would be wise to impose additional conditions on bending transformations.

Definition *A bending of a smooth manifold M lying in $\mathbb{R}^m$ is a smooth map $\Theta : M \to \mathbb{R}^d$ satisfying the following conditions:*

(1) Θ is a homeomorphism between M and $\Theta(M)$;

(2) Θ preserves the lengths of smooth curves: $\sigma_1(L) = \sigma_1(\Theta(L))$ for every smooth curve L contained in M.

Recall that the smoothness of Θ on M means that this map is smooth on an open set containing M.

It is intuitively clear that a bending does not change the area of a set lying on the surface. This observation underlies, for example, the computation of the areas of a cone and a cylinder known from school. Let us discuss the first of these examples in more detail (for the second one, see Exercise 5).

Example Consider a cone K in $\mathbb{R}^3$ formed by rays starting at the origin (the vertex of the cone). Such a cone is uniquely determined by its intersection with the unit sphere, i.e., the set $\ell = K \cap S^2$. Obviously, the smoothness of the surface $K\backslash\{0\}$ means that ℓ is a smooth curve. Adopting this assumption, we also assume that the length Δ of ℓ is finite and Ψ is the natural parametrization defined on $(0, \Delta)$. Let us "unfold" the cone in such a way that the curve ℓ turns into an arc of the unit circle of the same length; more precisely, a point $\Psi(s)$ maps to $z(s) = (\cos s, \sin s)$, and the ray passing through $\Psi(s)$ maps to the ray passing through $z(s)$ (all rays are assumed to start at the origin). We also assume that $\Delta < 2\pi$ (otherwise ℓ should be divided into several parts).

To formally verify that the described map is a bending, it is convenient to use the inverse map Θ. We will assume that it is defined in an angle C with the vertex at the origin whose points are determined by their polar angles lying in the interval $(0, \Delta)$. To obtain an analytic expression for Θ, consider the smooth map $P : C \to (0, \Delta) \times (0, +\infty)$ that associates with each point z of C its polar coordinates $\varphi(z)$ and $r(z)$. Then $\Theta(z) = r(z)\Psi(\varphi(z))$. Let us check that Θ is indeed a bending.

Let $\gamma(t)(t \in (a, b))$ be a parametrization of a smooth curve L lying in C. It generates the parametrization $\Phi = \Theta \circ \gamma$ of the curve $\Theta(L) \subset K$. We must show that the lengths $\sigma_1(L)$ and $\sigma_1(\Theta(L))$ of these curves coincide. Set $P(\gamma(t)) = (\omega(t), \rho(t))$. Note that $\sigma_1(L) = \int_a^b \sqrt{(\rho'(t))^2 + \rho^2(t) \cdot (\omega'(t))^2}dt$. Since $\Phi(t) = \rho(t)\Psi(\omega(t))$, the tangent vector Φ' breaks into two terms: $\Phi' = \rho'\Psi + \rho\omega'\Psi'$. Since $\|\Psi\| = \|\Psi'\| = 1$ and $\Psi' \perp \Psi$, it follows that $\|\Phi'\|^2 = (\rho')^2 + \rho^2 \cdot (\omega')^2$. Hence

$$\sigma_1(\Theta(L)) = \int_a^b \|\Phi'(t)\|\, dt = \int_a^b \sqrt{(\rho'(t))^2 + \rho^2(t) \cdot (\omega'(t))^2}dt = \sigma_1(L)$$

as required.

Thus Θ is a bending. By Theorem 8.3.6 (see below), it preserves the area. In particular, the area of the part of the cone K lying in the ball of radius R

centered at the vertex of the cone is equal to the area of the circular sector $C \cap B(0,R)$.

Now let us study under what conditions a smooth map is a bending.

Lemma *Let $M \subset \mathbb{R}^m$ be a smooth manifold and $\Theta : M \to \mathbb{R}^d$ be a smooth homeomorphism. It preserves the lengths of curves lying in M if and only iffor every point p in M, the map $d_p\Theta$ is an isometry of the tangent space $T_p(M)$ onto $\mathbb{R}^d$.*

Proof Let L ($L \subset M$) be a smooth curve passing through a point p, and let $t \mapsto \gamma(t)$ ($t \in (\alpha, \beta)$) be a (smooth) parametrization of L with $p = \gamma(t_0)$. Then $\delta = \Theta \circ \gamma$ is a parametrization of the curve $\widetilde{L} = \Theta(L)$. It is clear that $\delta'(t) = d_{\gamma(t)}\Theta(\gamma'(t))$, and the lengths of the arcs $\ell = \gamma((t_0,t)), \widetilde{\ell} = \delta((t_0,t))$ are given by the formulas

$$\sigma_1(\ell) = \int_{t_0}^{t} \|\gamma'(u)\|\, du, \quad \sigma_1(\tilde{\ell}) = \int_{t_0}^{t} \|\delta'(u)\|\, du \tag{11}$$

If Θ is a bending, then these lengths are equal. Differentiating with respect to t, we see that $\|\delta'(t_0)\| = \|\gamma'(t_0)\|$. This means that $\|d_p\Theta(\gamma'(t_0))\| = \|\gamma'(t_0)\|$, i.e., $\|d_p\Theta(x)\| = \|x\|$ for every vector x that can be written in the form $x = \gamma'(t_0)$. By the definition of the tangent space, every vector from $T_p(M)$ can be written in this form. Thus it follows from the preservation of the lengths of curves that $\|d_p\Theta(x)\| = \|x\|$ for $x \in T_p(M)$, i.e., $d_p\Theta$ is an isometry on the tangent space $T_p(M)$. If this condition is satisfied for every $p \in M$, then $\|\delta'(t)\| = \|\gamma'(t)\|$ for every t, so that the right-hand sides of equalities (11) coincide, which implies that Θ is a bending.□

It follows from the lemma that if Θ is a bending, then $\operatorname{rank}\Theta' = \dim M$ and, consequently, the set $\widetilde{M} = \Theta(M)$ is a smooth manifold of the same dimension as M.

Now we can easily prove that a bending preserves the area.

Theorem *Let M be a smooth k-dimensional manifold, Θ be a bending of M, and $\widetilde{M} = \Theta(M)$. Then for every set $E \in \mathfrak{A}_M$, its image $\widetilde{E} = \Theta(E)$ has the same area:*

$$\sigma_k(E) = \sigma_k(\widetilde{E})$$

Proof Obviously, it suffices to prove the assertion of the theorem for a set lying in some coordinate neighborhood $U \subset M$. Let Φ be a parametrization of U and $\widetilde{\Phi} = \Theta \circ \Phi$. The differential $d_{\Phi(t)}\Theta$ is an isometry of the accompanying parallelotope corresponding to Φ onto the parallelotope corresponding to $\widetilde{\Phi}$. Since isometries preserve Lebesgue measure, the measures of these parallelotopes coincide, i.e., the weights ω_Φ and $\omega_{\widetilde{\Phi}}$ are equal. Hence the areas of sets contained in U do not change.□

Note that a bending may also be an expanding map; for instance, the "straightening" of a circular arc, the "unfolding" of a cylinder into a plane, etc. These examples show that under an expanding map that strictly increases the distance between some points, the length and the area do not always strictly increase.

2.3.7

We have already observed that the area of a sphere is rotation-invariant (see Property (7) in Sect. 2.3.3). Let us discuss another example of an invariant measure. Consider the measure $\sigma = \sigma_{n(n-1)/2}$ on the group $O(n)$ of orthogonal $n \times n$ matrices with the metric induced from $\mathbb{R}^{n^2}$ (see Sect. 2.1.3, Example 5). Since this set is compact, the measure σ is finite. As we have established in Sect. 2.1.3, a translation on the group $O(n)$ (the multiplication on the left or on the right by a fixed element U_0 from $O(n)$) maps $O(n)$ isometrically onto itself. Since the area is invariant under isometries, the measure σ is invariant under translations on $O(n)$. In particular, for every summable function f on $O(n)$ and every V in $O(n)$, we have

$$\int_{O(n)} f(UV)d\sigma(U) = \int_{O(n)} f(VU)d\sigma(U) = \int_{O(n)} f(U)d\sigma(U). \qquad (12)$$

Using the existence of an invariant measure on the group $O(n)$, one can prove both the uniqueness of such a measure and the uniqueness of a rotation-invariant measure on the sphere. More precisely, the following theorem holds.

Theorem

(1) A finite Borel rotation-invariant measure on S^{m-1} is unique up to a multiplicative constant.

(2) A finite Borel measure on $O(n)$ invariant under an arbitrary right or left translation (i.e., under left or right multiplication by an element of the group $O(n)$) is unique up to a multiplicative constant.

Proof (1) Let v be a Borel rotation-invariant measure on S^{m-1} and σ be the area on $O(m)$, which we know to be translation-invariant. Assume that $v\left(S^{m-1}\right) = \sigma_{m-1}\left(S^{m-1}\right)$; we are going to prove that the measures v and σ_{m-1} coincide.

Consider the measure μ on $O(m)$ obtained by normalizing the measure σ (i.e., $\mu = \frac{1}{\sigma(O(m))}\sigma\Big)$, and let E be a Borel subset of the sphere S^{m-1}. First let us show that the value $\int_{O(m)} \chi_E\left(Ux_0\right) d\mu(U)$ does not depend on the choice of a point x_0 from S^{m-1}. Indeed, for every vector $x \in S^{m-1}$ there is an orthogonal transformation V such that $x = Vx_0$. Hence, by (12) with $f(U) = \chi_E\left(Ux_0\right)$,

$$\int_{O(m)} \chi_E(Ux)d\mu(U) = \int_{O(m)} \chi_E\left(UVx_0\right) d\mu(U) = \int_{O(m)} \chi_E\left(Ux_0\right) d\mu(U),$$

as required. Since, by the invariance of v,

$$v(E) = \int_{S^{m-1}} \chi_E(x)dv(x) = \int_{S^{m-1}} \chi(Ux)dv(x)$$

for every U in $O(m)$, integrating this equality with respect to the (normalized) measure μ and changing the order of integration yields

$$\nu(E) = \int_{O(m)} \nu(E) d\mu(U) = \int_{O(m)} \left(\int_{S^{m-1}} \chi_E(Ux) d\nu(x) \right) d\mu(U)$$
$$= \int_{S^{m-1}} \left(\int_{O(m)} \chi_E(Ux) d\mu(U) \right) d\nu(x) = \nu\left(S^{m-1}\right) \int_{O(m)} \chi_E(Ux) d\mu(U),$$

where the right-hand side, as we have established above, does not depend on x. Obviously, a similar equality can be written with v replaced by σ_{m-1}. The righthand sides of these equalities are equal. Therefore, the left-hand sides also coincide, as required.

When changing the order of integration (and, consequently, using Tonelli's theorem), we have assumed that the function $(x, U) \mapsto \varphi(x, U) \equiv \chi_E(Ux)$ is measurable on $S^{m-1} \times O(m)$. This is indeed the case, since φ is the characteristic function of the set $\{(x, U) \mid Ux \in E\} = \Psi^{-1}(E)$, where $\Psi(x, U) = Ux$. Since the map Ψ is obviously continuous and E is a Borel set, its inverse image $\Psi^{-1}(E)$ is also a Borel set.

The proof of Claim (2) is completely analogous.□

2.4 Integration over a Smooth Manifold

2.4.1

The computation of an integral with respect to the surface area of a smooth manifold, or, in short, of a surface integral, can be reduced to the computation of a multiple integral with respect to the Lebesgue measure. This transition does not require additional efforts, since, by Property (3) from Sect. 2.3.3, the area of a simple manifold is a weighted image of the Lebesgue measure. Hence we may apply the general Theorem on the computation of an integral with respect to a weighted image of a measure, which in the case under consideration leads to the following result.

Theorem *Let M be a simple smooth manifold in $\mathbb{R}^m$, $\dim M = k$, and f be a nonnegative measurable function on M. Then*

$$\int_M f(x) d\sigma_k(x) = \int_{\Phi^{-1}(M)} f(\Phi(t)) \omega_\Phi(t) dt$$

for every parametrization Φ of M.

This formula also holds for every summable function on M.

Recall that $\omega_\Phi(t)$ has a simple geometric interpretation: this is the volume of the accompanying parallelotope. For $k = 1$, it is equal to the length of the tangent vector $\Phi'(t)$, and for $k = m - 1$, to the length of the normal $N_\Phi(t)$ corresponding to the parametrization Φ (see Sect. 2.3.4).

As in the general situation, a similar formula holds for functions defined not on the whole manifold M, but only on a measurable subset of M. If the manifold M is not simple, then an integral over M can be computed by considering a

partition of M into at most countably many sets each of which is contained in a coordinate neighborhood.

In the case $\dim M = m$, the assertion of the theorem is just the change of variables formula for a diffeomorphism.

Observe the important special case when the manifold is the graph of a smooth function φ defined on an open subset $\mathcal{O}$ of $\mathbb{R}^{m-1}$. Consider the canonical parametrization of the graph $\mathcal{O} \ni x \mapsto \Phi(x) = (x, \varphi(x))$. Then (see Sect. 2.3.4) $\omega_\Phi(x) = \sqrt{1 + \|\operatorname{grad}\varphi(x)\|^2}$ and $\Phi^{-1}(E) = P(E)$, where P is the orthogonal projection to $\mathbb{R}^{m-1}$. Hence for every measurable set $E \subset M = \Gamma_\varphi$,

$$\int_E f d\sigma_{m-1} = \int_{P(E)} f(x, \varphi(x)) \sqrt{1 + \|\operatorname{grad}\varphi(x)\|^2} dx \tag{1}$$

Example 1 Let $\Sigma_m = S^{m-1} \cap \mathbb{R}^m_+$ be the part of the unit sphere S^{m-1} of $\mathbb{R}^m$ lying in the "first octant". Assuming that the sphere is homogeneous, we are going to find the center of mass C of the surface Σ_m. By symmetry, all coordinates of this vector are equal: $C = (c, \dots, c)$. As we have established, they are given by the formula

$$c = \frac{1}{\sigma_{m-1}(\Sigma_m)} \int_{\Sigma_m} x_m d\sigma_{m-1}(x) = \frac{2^m}{m\alpha_m} \int_{\Sigma_m} x_m d\sigma_{m-1}(x)$$

To compute this integral, observe that Σ_m is a subset of the graph of the function $\varphi(t) = \sqrt{1 - \|t\|^2}$ defined on the unit ball B^{m-1} and the projection Σ_m coincides with the intersection $A = B^{m-1} \cap \mathbb{R}^{m-1}_+$. Applying (1) with $f(x) = x_m$, we obtain

$$\begin{aligned} c &= \frac{2^m}{m\alpha_m} \int_A \varphi(t) \sqrt{1 + \|\operatorname{grad}\varphi(t)\|^2} dt = \frac{2^m}{m\alpha_m} \int_A 1 dt = \frac{2^m}{m\alpha_m} \lambda_{m-1}(A) \\ &= \frac{2\alpha_{m-1}}{m\alpha_m} = \frac{\Gamma\left(\frac{m}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{m+1}{2}\right)}. \end{aligned}$$

In particular,

$$\|C\| = \frac{\sqrt{m}\Gamma\left(\frac{m}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{m+1}{2}\right)}$$

As $m \to \infty$, these norms tend to $\sqrt{\frac{2}{\pi}}$. One can show that they decrease. Note that the center of mass C' of the part of the unit ball lying in $\mathbb{R}^m_+$ has the coordinates $c' = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m+1}{2}\right)} \frac{m}{m+1} = \frac{m}{m+1} c$ and $\|C'\|$ tends to the same limit as $\|C\|$, but increases rather than decreases.

Example 2 Let M be a smooth k-dimensional manifold in $\mathbb{R}^m, \sigma_k(M) < +\infty$, and $x_0 \in M$. Let us find out for which $p > 0$ the integral

$$I_0 = \int_M \frac{d\sigma_k(x)}{\|x - x_0\|^p}$$

is finite. First we will find a necessary condition. Consider a parametrization Φ of an M-neighborhood of the point $x_0 = \Phi(a)$. In some ball $B(a,\rho) \subset \mathbb{R}^k$, Φ satisfies the Lipschitz condition: $\|\Phi(t) - \Phi(a)\| \leqslant L\|t-a\|$, where L is a fixed positive number. If ρ is sufficiently small, then $\omega_\Phi(t) \geqslant \frac{1}{2}\omega_\Phi(a)$ for $t \in B(a,\rho)$. Hence

$$I_0 \geqslant \int_{B(a,\rho)} \frac{\omega_\Phi(t)dt}{\|\Phi(t)-\Phi(a)\|^p} \geqslant \frac{\omega_\Phi(a)}{2L^p} \int_{B(a,\rho)} \frac{dt}{\|t-a\|^p}$$

If I_0 is finite, then the integral on the right-hand side is also finite, which is possible only for $p < k$.

Now we will show that the condition $p < k$ is not only necessary, but also sufficient for I_0 to be finite. In order to prove at once a somewhat stronger result, we introduce the integral

$$I(y) = \int_M \frac{d\sigma_k(x)}{\|x-y\|^p} \quad (y \in \mathbb{R}^m)$$

Obviously, $I_0 = I(x_0)$. We will prove that for $p < k$, the integral I is bounded in the vicinity of x_0. Note that in general the condition $y \notin M$ is not sufficient for $I(y)$ to be finite if $y \in \bar{M}\backslash M$.

We still assume that Φ is a parametrization of an M-neighborhood of x_0 and $x_0 = \Phi(a)$. Recall that in the vicinity of a the parametrization Φ is the restriction of some diffeomorphism P (see Lemma 2.1.4; we assume that the space $\mathbb{R}^k$, on a subset of which Φ is defined, is canonically embedded into $\mathbb{R}^m$). In a sufficiently small ball $B(x_0,r)$, the map F^{-1} satisfies the Lipschitz condition with some constant C :

$$\left\|F^{-1}(x) - F^{-1}(y)\right\| \leqslant C\|x-y\| \quad \text{for } x,y \in B(x_0,r) \tag{2}$$

Assuming that r is so small that $\omega_\Phi\left(\Phi^{-1}(x)\right) \leqslant 2\omega_\Phi(a)$ for all x in $M_r = M \cap B(x_0,r)$, we will prove that the integral I is bounded on the ball $B(x_0,r)$.

Taking an arbitrary point y from this ball, write the integral $I(y)$ in the form

$$I(y) = \int_{M_r} \frac{d\sigma_k(x)}{\|x-y\|^p} + \int_{M\backslash M_r} \frac{d\sigma_k(x)}{\|x-y\|^p} = I_1(y) + I_2(y).$$

It is clear that

$$I_2(y) \leqslant \frac{1}{r^p} \int_{M\backslash M_r} d\sigma_k(x) \leqslant \frac{1}{r^p}\sigma_k(M) < +\infty$$

It remains to estimate the integral $I_1(y)$ over the simple manifold M_r :

$$I_1(y) = \int_{\Phi^{-1}(M_r)} \frac{\omega_\Phi(t)dt}{\|\Phi(t)-y\|^p} \leqslant 2\omega_\Phi(a) \int_{\Phi^{-1}(M_r)} \frac{dt}{\|\Phi(t)-y\|^p}$$

Let us estimate the norm $\|\Phi(t)-y\|$ from below. Let $x = \Phi(t)$ and $s = F^{-1}(y)$. It follows from (2) that for $x, y \in B(x_0,r)$,

$$\|\Phi(t)-y\| = \|x-y\| \geqslant \frac{1}{C}\|t-s\| \geqslant \frac{1}{C}\|t-u\|,$$

where u is the projection of s to $\mathbb{R}^k$. Since the points $x = \Phi(t)$ and y lie in the ball $B(x_0, r)$, we have $\|\Phi(t)-y\| < 2r$ and, consequently, $\|t-u\| \leqslant C\|\Phi(t)-y\| < 2\mathrm{Cr}$. Hence

$$\int_{\Phi^{-1}(M_r)} \frac{dt}{\|\Phi(t)-y\|^p} \leqslant \int_{\|t-u\|<2Cr} \left(\frac{C}{\|t-u\|}\right)^p dt = C^p \int_{\|v\|<2Cr} \frac{dv}{\|v\|^p}$$
$$= C^p \frac{k\alpha_k}{k-p}(2Cr)^{k-p}.$$

So, for $y \in B(x_0, r)$,

$$I(y) \leqslant 2C^k \omega_\Phi(a) \frac{k\alpha_k}{k-p}(2r)^{k-p} + \frac{1}{r^p}\sigma_k(M)$$

(the parameters C and r depend on the manifold M and the point x_0, but not on the exponent p).

Example 3 (Integrals similar to a simple-layer potential) Let M be a smooth k-dimensional manifold in $\mathbb{R}^m$, E be a compact subset of M, and $w \in C(E)$. Let us check that for $p < k$, the function

$$y \mapsto I(y) = \int_E \frac{w(x)}{\|x-y\|^p} d\sigma_k(x) \quad (y \in \mathbb{R}^m)$$

is continuous in the whole space and infinitely differentiable in $\mathbb{R}^m \backslash E$.

The smoothness of I outside E follows from the fact that for $y_0 \notin E$ the norm $\|y-x\|$ is bounded away from zero if $x \in E$ and y lies in a sufficiently small neighborhood of y_0. Hence the integrand, as well as all its partial derivatives with respect to the coordinates $y_1, \ldots, y_m$, are bounded in the vicinity of y_0. Thus at y_0 the condition (L_{loc}) is satisfied and we can apply the Leibniz rule.

To prove that I is continuous at a point $y_0 \in E$, we use Theorem 3 of Sect. 1.1.2. Fix a number $s > 1$ such that $sp < k$ and put $C = \max_E |w|$. As we have established in the previous example, the integral $\widetilde{I}(y) = \int_E \frac{C\, d\sigma_k(x)}{\|x-y\|^{sp}}$ is bounded in the vicinity of y_0, and this, by Theorem 3 of Sect. 1.1.2, suffices for I to be continuous at this point.

2.4.2

In this section we will obtain a generalization of Fubini's theorem to the case where an open subset of $\mathbb{R}^m (m \geqslant 2)$ stratifies not into affine subspaces, but into the level surfaces of a smooth function. In the special case where the level surfaces are concentric spheres. Indeed, in this theorem it is proved that for every function f summable in the ball $B(0, r) \subset \mathbb{R}^m$,

$$\int_{B(0,r)} f(x)dx = \int_0^r t^{m-1} \left(\int_{S^{m-1}} f(t\xi) d\sigma_{m-1}(\xi) \right) dt$$

Since both the area σ_{m-1} and the volume λ_m are translation-invariant, we may consider spheres with arbitrary center. Furthermore, using the equality $\sigma_{m-1}(tE) = t^{m-1}\sigma_{m-1}(E)$ for $t > 0$ (see Property (6) in Sect. 2.3.3) and the change of variables theorem, we can write the next assertion:

$$\int_{B(a,r)} f(x)dx = \int_0^r \left(\int_{S(a,t)} f(x) d\sigma_{m-1}(x) \right) dt \tag{3}$$

The theorem we are going to consider next is a far-reaching generalization of this result.

Theorem *(Kronrod - Federer) Let $\mathcal{O}$ be an open subset of $\mathbb{R}^m$, $F \in C^1(\mathcal{O})$ and* $\operatorname{grad} F \neq 0$ *in $\mathcal{O}$. Then for every function f summable in $\mathcal{O}$,*

$$\int_{\mathcal{O}} f(x)dx = \int_{-\infty}^{\infty} \left(\int_{M(t)} \frac{f(x)}{\| \operatorname{grad} F(x) \|} d\sigma_{m-1}(x) \right) dt \tag{4}$$

where $M(t) = \{x \in \mathcal{O} \mid F(x) = t\}$.

Proof Let us first prove a local version of this theorem: every point $x_0 \in \mathcal{O}$ has a small neighborhood U such that (4) holds for every function f vanishing outside U.

We assume without loss of generality that $F(x_0) = 0$. Moreover, applying if necessary a translation and an orthogonal transformation, we may assume that $x_0 = 0$ and that the tangent plane to $M(0)$ at the origin coincides with the coordinate subspace $x_m = 0$. To simplify formulas, denote by u the projection of x to this subspace, and let v be the last coordinate x_m, so that $x = (u, v), u = (u_1, \ldots, u_{m-1}) \in \mathbb{R}^{m-1}, v \in \mathbb{R}$. Then $F'_{u_k}(0) = 0$ for $1 \leqslant k < m$ and $F'_v(0) \neq 0$. Consider the map $T : \mathcal{O} \to \mathbb{R}^m$ that "straightens" the level surfaces: $T(x) = (u, F(x))$. It transforms the level surfaces into planes parallel to the subspace $\mathbb{R}^{m-1}$. The Jacobian J_T of this map at $x = 0$ does not vanish, since $J_T(0) = F'_v(0) \neq 0$. Hence the restriction of T to some neighborhood U of the origin is a diffeomorphism. Let us assume that U is projected into a ball of radius δ and lies between level surfaces $M(-\varepsilon)$ and $M(\varepsilon)$, i.e.,

$$U = \{x = (u, v) | \|u\| < \delta, |F(x)| < \varepsilon\}$$

where δ and ε are sufficiently small positive numbers. Then for $|t| < \varepsilon$ the set $T(M(t) \cap U)$ is contained in the affine subspace $v = t$, and $T(U)$ coincides with the Cartesian product $W = B^{m-1}(0, \delta) \times (-\varepsilon, \varepsilon)$. Clearly, the map Φ inverse to the restriction of T to U, as well as T itself, affects only the last coordinate of the argument, so that it has the form

$$\Phi(u, t) = (u, \varphi(u, t)), \qquad \text{where } \|u\| < \delta, |t| < \varepsilon$$

(φ is the last coordinate function of the map $\Phi, \varphi \in C^1(W)$).

Thus, for $|t| < \varepsilon$, the part of $M(t)$ lying in U is just the graph of the smooth function $u \mapsto \varphi_t(u) \equiv \varphi(u,t)$. Since $F(u, \varphi_t(u)) \equiv t$, it is easy to establish a relation between the gradients of the functions φ_t and F :

$$F'_{u_j}(u, \varphi_t(u)) + F'_v(u, \varphi_t(u)) \frac{\partial \varphi(u,t)}{\partial u_j} \equiv 0 \quad \text{for } 1 \leqslant j < m$$

Hence

$$\frac{1}{|F'_v|} = \frac{\sqrt{1 + \|\operatorname{grad} \varphi_t\|^2}}{\|\operatorname{grad} F\|} \tag{5}$$

Moreover, using the identity $F'_v(u, \varphi(u,t)) \frac{\partial \varphi(u,t)}{\partial t} \equiv 1$, we can compute the Jacobian of Φ :

$$J_\Phi(u,t) = \frac{\partial \varphi(u,t)}{\partial t} = \frac{1}{F'_v(\Phi(u,t))}$$

Assuming that f vanishes outside U, we obtain, making a substitution, that

$$\int_U f(x)dx = \int_W \frac{f(\Phi(u,t))}{|F'_v(\Phi(u,t))|} du dt = \int_{-\varepsilon}^{\varepsilon} \left(\int_{\|u\|<\delta} \frac{f(\Phi(u,t))}{|F'_v(\Phi(u,t))|} du \right) dt.$$

Let us write the inner integral as an integral over the graph of φ_t (see (1)). In view of (5), we have

$$\begin{aligned} \int_{\|u\|<\delta} \frac{f(\Phi(u,t))}{|F'_v(\Phi(u,t))|} du &= \int_{\|u\|<\delta} \frac{f(u, \varphi_t(u))}{\|\operatorname{grad} F(u, \varphi_t(u))\|} \sqrt{1 + \|\operatorname{grad} \varphi_t(u)\|^2} du \\ &= \int_{M'(t)} \frac{f(x)}{\|\operatorname{grad} F(x)\|} d\sigma_{m-1}(x), \end{aligned} \tag{6}$$

where $M'(t)$ is the graph of φ_t, i.e., the part of $M(t)$ contained in U. Thus

$$\begin{aligned} \int_U f(x)dx &= \int_{-\varepsilon}^{\varepsilon} \left(\int_{M(t)\cap U} \frac{f(x)}{\|\operatorname{grad} F(x)\|} d\sigma_{m-1}(x) \right) dt \\ &= \int_{-\infty}^{\infty} \left(\int_{M(t)\cap U} \frac{f(x)}{\|\operatorname{grad} F(x)\|} d\sigma_{m-1}(x) \right) dt \end{aligned}$$

Since $f = 0$ outside U, it follows that (4) holds for functions that do not vanish only in a sufficiently small neighborhood of x_0.

It follows from the obtained local version of the theorem that (4) holds for a summable function f supported by a compact subset of $\mathcal{O}$. Indeed, for each point $x \in \mathcal{O}$ choose a neighborhood $U_x \subset \mathcal{O}$ such that (4) holds for functions vanishing outside U_x. By Theorem 2.1.8, there exists a partition of unity $\varphi_1, \ldots, \varphi_N$ on the set $\operatorname{supp}(f)$ subordinate to the family $\{U_x\}_{x\in\mathcal{O}}$. Write (4) for $f\varphi_k$:

$$\int_{\mathcal{O}} f(x)\varphi_k(x)dx = \int_{-\infty}^{\infty} \left(\int_{M(t)} f(x)\varphi_k(x)d\sigma_{m-1}(x) \right) dt.$$

Adding these equalities, we obtain the desired result for compactly supported functions. To prove it for a non-negative function with arbitrary support (obviously, this suffices for proving the theorem in full strength), we should exhaust the set $\mathcal{O}$ by an increasing sequence of compact subsets K_n and apply Levi's theorem to both sides of (4) with f replaced by $f\chi_{K_n}$.□

Remark If f is continuous, then the function

$$t \mapsto \int_{M(t)} \frac{f(x)}{\| \operatorname{grad} F(x)\|} d\sigma_{m-1}(x)$$

is also continuous. If the support of f is small, this follows from (6). The general case can be proved using a partition of unity.

Observe also that the theorem remains valid if the smoothness of F is violated at a closed set E satisfying the conditions

$$\lambda_m(E) = 0, \quad \sigma_{m-1}(M(t) \cap E) = 0 \quad \text{for every } t$$

To prove this, it suffices to apply the theorem to $\mathcal{O}\backslash E$ replacing $M(t)$ with $M(t)\backslash E$.

2.4.3

We now make a few more remarks about the obtained result.

(1) Obviously, formula (3) follows from the above theorem with $\mathcal{O} = \mathbb{R}^m\backslash\{a\}$ and $F(x) = \|x - a\|$ (note that $\| \operatorname{grad} F(x)\| \equiv 1$).

(2) Replacing f with $f\| \operatorname{grad} F\|$, we can rewrite (4) in the form

$$\int_{\mathcal{O}} f(x)\| \operatorname{grad} F(x)\|dx = \int_{-\infty}^{\infty} \left(\int_{M(t)} f(x)d\sigma_{m-1}(x) \right) dt \qquad (4')$$

If F is sufficiently smooth, the condition $\operatorname{grad} F \neq 0$ can be dropped, since, by Sard's theorem, the set of critical values of the function $F \in C^m(\mathcal{O})$ has zero measure. Indeed, let $\widetilde{\mathcal{O}} = \{x \in \mathcal{O} \mid \operatorname{grad} F(x) \neq 0\}$, and let $E \subset \mathbb{R}$ be the set of critical values of F, so that $\lambda_1(E) = 0$. If $t \notin E$, then F does not take the value t on $\mathcal{O}\backslash\widetilde{\mathcal{O}}$, and, consequently, $M(t) \cap \widetilde{\mathcal{O}} = M(t)$. Thus we can obtain the desired result by applying (4′) to $\widetilde{\mathcal{O}}$.

(3) Let $V(u) = \lambda_m(\mathcal{O}(F < u))(u \in \mathbb{R})$. Assuming that $V(u) < +\infty, f \equiv 1$ and applying (4) to $\mathcal{O}(F < u)$, we have (taking into account that $M(t) = \varnothing$ for $t > u$)

$$V(u) = \int_{-\infty}^{u} \left(\int_{M(t)} \frac{d\sigma_{m-1}(x)}{\| \operatorname{grad} F(x)\|} \right) dt$$

Since the function $t \mapsto \int_{M(t)} \frac{d\sigma_{m-1}(x)}{\|\operatorname{grad} F(x)\|}$ is continuous (see Remark 8.4.2), differentiating the last equality, we arrive at the following result:

$$V'(u) = \int_{M(u)} \frac{d\sigma_{m-1}(x)}{\|\operatorname{grad} F(x)\|} \tag{7}$$

In the special case where $F(x) = \|x\|$ (and, correspondingly, $\|\operatorname{grad} F(x)\| \equiv 1$), the obtained formula leads to the equality $\left(\lambda_m(B(u))\right)' = \sigma_{m-1}\left(S^{m-1}(u)\right)$, which we have already encountered (see Sect. 2.3.5, Example 4).

2.4.4

Now we use formula (7) to relate the area of a surface and its Minkowski area. One can prove that for a compact set A contained in a smooth surface $S \subset \mathbb{R}^m$,

$$\sigma_{m-1}(A) = \lim_{\varepsilon \to 0} \frac{\lambda_m\left(A_\varepsilon \backslash A\right)}{2\varepsilon}$$

where A_ε is the ε-neighborhood of A. We will prove a similar formula not for an arbitrary compact subset of a smooth surface, but for the boundary of a compact Lebesgue set of a smooth function.

Theorem *Let $\mathcal{O}$ be an open subset of $\mathbb{R}^m, F \in C^2(\mathcal{O}), K = \mathcal{O}(F \leqslant C)$ and $M = \partial K$. If the set K is compact and $\operatorname{grad} F \neq 0$ on M, then the area of M coincides with the Minkowski area, i.e.,*

$$\sigma_{m-1}(M) = \lim_{\varepsilon \to 0} \frac{\lambda_m\left(K_\varepsilon \backslash K\right)}{\varepsilon}$$

Proof First assume that $\|\operatorname{grad} F\| \equiv 1$ on M; we may also assume without loss of generality that $C = 0$. Fix $\delta > 0$ such that $\bar{K}_\delta \subset \mathcal{O}$. Let ω be the modulus of continuity of the map $x \mapsto \operatorname{grad} F(x)$ on the set $K_\delta \backslash K$. We will assume that δ is so small that $\omega(\delta) < 1/2$. Let us show that for small $\varepsilon > 0$ the sets $V(\varepsilon) = \{x \in K_\delta \mid F(x) \leqslant \varepsilon\}$ are close to the ε-neighborhoods of K. For this, keeping the above notation, we will show that the following lemma holds.

Lemma *Let $\varepsilon < \delta/2, \varepsilon' = \varepsilon(1 - \omega(2\varepsilon))$ and $\varepsilon'' = \varepsilon(1 + \omega(2\varepsilon))$; then*

$$V\left(\varepsilon'\right) \subset K_\varepsilon \subset V\left(\varepsilon''\right)$$

Proof of the lemma If $x \in K_\delta \backslash K$ and x_0 is the point of M that is closest to x, then

$$\begin{aligned} F(x) = F(x) - F\left(x_0\right) &\leqslant \max_{z \in [x, x_0]} \|\operatorname{grad} f(z)\| \, \|x - x_0\| \\ &\leqslant \left(1 + \omega\left(\|x - x_0\|\right)\right) \|x - x_0\| . \end{aligned}$$

Hence for every point x in K_ε, for $\varepsilon < \delta$ we have $F(x) \leqslant (1 + \omega(\varepsilon))\varepsilon \leqslant \varepsilon''$ and, consequently, $K_\varepsilon \subset V\left(\varepsilon''\right)$.

Now we will prove that $V\left(\varepsilon'\right) \subset K_\varepsilon$. Let $x \in V(\varepsilon) \backslash K$, and let x_0 be the point of M that is closest to x. By the definition of $V(\varepsilon)$, we have $\|x - x_0\| < \delta$.

One can easily see that the vectors $x - x_0$ and $\operatorname{grad} F(x_0)$ are proportional: $x - x_0 = \|x - x_0\| \operatorname{grad} F(x_0)$. Hence for some $z \in [x_0, x]$ we have

$$\begin{aligned} \varepsilon \geqslant & F(x) - F(x_0) = \langle \operatorname{grad} F(z), x - x_0 \rangle = \langle \operatorname{grad} F(x_0), x - x_0 \rangle \\ & + \langle \operatorname{grad} F(z) - \operatorname{grad} F(x_0), x - x_0 \rangle \\ \geqslant & \|x - x_0\| - \omega(\|x - x_0\|) \|x - x_0\| \end{aligned} \tag{8}$$

Since $\|x - x_0\| < \delta$, we obtain $\omega(\|x - x_0\|) \leqslant \omega(\delta) \leqslant 1/2$, whence $\|x - x_0\| \leqslant 2\varepsilon$. Returning to (8), we see that

$$\varepsilon \geqslant \|x - x_0\| - \omega(\|x - x_0\|) \|x - x_0\| \geqslant \|x - x_0\| - \omega(2\varepsilon) \|x - x_0\|,$$

i.e., $\|x - x_0\| \leqslant t = \varepsilon/(1 - \omega(2\varepsilon))$. It follows that $V(\varepsilon) \subset K_t$. Since $t > \varepsilon$, we have $\omega(2\varepsilon) \leqslant \omega(2t)$, whence $\varepsilon = t(1 - \omega(2\varepsilon)) > t(1 - \omega(2t))$. Therefore,

$$V(t(1 - \omega(2t))) \subset V(\varepsilon) \subset K_t,$$

which, replacing t by ε, can be rewritten in the form

$$V(\varepsilon') = V(\varepsilon(1 - \omega(2\varepsilon))) \subset K_\varepsilon$$

□

Let us return to the proof of the theorem. It follows from the lemma that

$$(1 - \omega(2\varepsilon)) \frac{\lambda_m(V(\varepsilon') \setminus K)}{\varepsilon'} \leqslant \frac{\lambda_m(K_\varepsilon \setminus K)}{\varepsilon} \leqslant (1 + \omega(2\varepsilon)) \frac{\lambda_m(V(\varepsilon'') \setminus K)}{\varepsilon''}.$$

By (7), as $\varepsilon \to 0$, the outermost parts of this inequality tend to $\int_M \frac{d\sigma_{m-1}(x)}{\|\operatorname{grad} F(x)\|} = \sigma_{m-1}(M)$, which completes the proof of the theorem under the additional assumption made above. Note that at this stage of the proof, we haven't yet used the C^2 smoothness of F, but have used only the C^1-smoothness.

In the general case (still assuming that $C = 0$), we introduce an auxiliary function H by the formula

$$H(x) = \frac{F(x)}{\sqrt{\|\operatorname{grad} F(x)\|^2 + F^2(x)}}.$$

It is obvious that $H \in C^1(\mathcal{O}), K = \mathcal{O}(H \leqslant 0)$, and

$$\operatorname{grad} H(x) = \frac{\operatorname{grad} F(x)}{\sqrt{\|\operatorname{grad} F(x)\|^2 + F^2(x)}} + F(x) \operatorname{grad} \frac{1}{\sqrt{\|\operatorname{grad} F(x)\|^2 + F^2(x)}}.$$

Hence $\|\operatorname{grad} H(x)\| = 1$ for $x \in M$, and the desired equality holds by the first part of the proof.□

2.4.5

Using the isoperimetric inequality and Theorem 2.4.4, we can obtain the main special case of the Gagliardo-Nirenberg-Sobolev inequality. For $p = 1$, it takes the form

$$\left(\int_{\mathbb{R}^m} |F(x)|^{\frac{m}{m-1}} dx\right)^{\frac{m-1}{m}} \leqslant \frac{1}{2}\int_{\mathbb{R}^m} \|\operatorname{grad} F(x)\| dx$$

where $F \in C_0^1(\mathbb{R}^m)$. We will prove it and, in passing, reduce the coefficient on the right-hand side. Since every smooth function, along with all its derivatives, can be uniformly approximated by functions of the class C_0^∞ (see Theorem 2 in Sect. 1.4.4), in what follows we assume that $F \in C_0^\infty(\mathbb{R}^m)$. Then formula (4') with $f \equiv 1$ yields

$$\int_{\mathbb{R}^m} \|\operatorname{grad} F(x)\| dx = \int_{-\infty}^{\infty} \sigma_{m-1}(M(t)) dt \tag{9}$$

where $M(t)$ is the boundary of the set $V(t) = \{x \in \mathbb{R}^m \mid F(x) \geqslant t\}$. Since, by Sard's theorem, the set of critical values of the function F has zero measure, this equality obviously remains valid if we integrate $\|\operatorname{grad} F(x)\|$ not over the whole space $\mathbb{R}^m$, but only over the set $\mathcal{O} = \{x \in \mathbb{R}^m \mid F(x) \neq 0, \operatorname{grad} F(x) \neq 0\}$. In this case we may assume that $F \geqslant 0$, since otherwise F can be replaced by $|F|$.

By Theorem 2.4.4, for non-critical values $t \in \mathbb{R}$,

$$\sigma_{m-1}(M(t)) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \lambda_m\left((V(t))_\varepsilon \backslash V(t)\right)$$

by the isoperimetric inequality, the right-hand side is not less than $m\alpha_m^{\frac{1}{m}} \lambda_m^{\frac{m-1}{m}}(V(t))$, where α_m is the volume of the unit ball. Thus (9) implies that

$$\int_{\mathbb{R}^m} \|\operatorname{grad} F(x)\| dx \geqslant m\alpha_m^{\frac{1}{m}} \int_0^\infty \lambda_m^{\frac{m-1}{m}}(V(t)) dt \tag{10}$$

To estimate the last integral, we need the following lemma.

Lemma *If a non-negative function ψ does not increase on $[0, +\infty)$, then for any $r > 1$ and $s > 0$,*

$$\left(\int_0^s \psi^r(t) dt^r\right)^{\frac{1}{r}} \leqslant \int_0^s \psi(t) dt$$

Proof of the lemma Denote the left- and right-hand sides of the inequality by $I(s)$ and $J(s)$, respectively. Since the function ψ does not increase, $I(s) \geqslant \psi(s)\left(\int_0^s dt^r\right)^{\frac{1}{r}} = s\psi(s)$. Hence for almost all $s > 0$ we have

$$I'(s) = \frac{1}{r} I^{1-r}(s) r s^{r-1} \psi^r(s) \leqslant (s\psi(s))^{1-r} s^{r-1} \psi^r(s) = \psi(s) = J'(s).$$

The lemma follows, since the functions I, J are absolutely continuous and $I(0) = J(0)(= 0)$.$\square$

Applying the lemma to the function $\psi(t) = \lambda_m^{\frac{1}{r}}(V(t))$, we obtain

$$\int_0^\infty \lambda_m^{\frac{1}{r}}(V(t))dt \geqslant \left(r\int_0^\infty t^{r-1}\lambda_m(V(t))dt\right)^{\frac{1}{r}} = \left(\int_{\mathbb{R}^m} |F(x)|^r dx\right)^{\frac{1}{r}}$$

(at the end, we have used Proposition 6.4.3). For $r = \frac{m}{m-1}$, this inequality together with (10) yields the desired bound:

$$\left(\int_{\mathbb{R}^m} |F(x)|^{\frac{m}{m-1}} dx\right)^{\frac{m-1}{m}} \leqslant \frac{1}{m\alpha_m^{1/m}} \int_{\mathbb{R}^m} \|\operatorname{grad} F(x)\| dx$$

Note that $m\alpha_m^{1/m} \geqslant 2\sqrt{m}$, because

$$\alpha_m^{\frac{1}{m}} = \lambda_m^{\frac{1}{m}}(B(1)) \geqslant \lambda_m^{\frac{1}{m}}\left(\left(-\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{m}}\right)^m\right) = \frac{2}{\sqrt{m}}$$

2.5 Integration of Vector Fields

2.5.1

In problems of mechanics and physics, one often encounters integrals of the form

$$\int_M \langle V(x), \theta(x)\rangle d\sigma(x)$$

where M is a smooth manifold, $V(x)$ are vectors corresponding to the problem under consideration, $\theta(x)$ is a unit vector related only to M, and σ is the surface area of M. Note that for the integrand to be summable it suffices that all coordinates of the vector $V(x)$ be summable on M, which is equivalent to the condition $\int_M \|V(x)\| d\sigma(x) < +\infty$. In what follows, we assume that this condition is satisfied.

We will restrict ourselves to the discussion of two extreme cases, which are of special interest.

(I) M is a one-dimensional manifold and $\theta(x)$ is a unit tangent vector to M at x.

(II) M is a manifold of codimension 1 (surface) and $\theta(x)$ is a unit normal to M at x.

We introduce several terms that will allow us to clarify the physical interpretation of the arising integrals.

Let us regard a continuous map $V : E \to \mathbb{R}^m$, where $E \subset \mathbb{R}^m$, as the family of vectors $\{V(x)\}_{x\in E}$ and call it a vector field on E. As a rule, we assume that the set E is open and the field is smooth (the latter means that the map V is C^1-smooth).

We can interpret $V(x)$ as the force applied at the point x and speak about a force field. We may also imagine that in the set E there is a steady-state

flow of matter (fluid or gas) such that the velocity of the particle that at time t is at position $x \in E$ does not depend on the time and is equal to $V(x)$. In this case, one says that in E there is a stationary flow and V is its velocity field. We will abide by these mechanical interpretations. However, one should bear in mind that in applications an important role is also played by vector fields of another nature, for instance, the electric or magnetic fields appearing in Maxwell's equations.

2.5.2

Integration over an Oriented Curve. Let us first discuss case I. For simplicity, we assume that the vector field V is defined in a domain $\mathcal{O} \subset \mathbb{R}^m$. It makes sense to change the notation, in order to emphasize the one-dimensional nature of the problem under consideration. A one-dimensional manifold will be called a curve and denoted by L. The measure $\sigma = \sigma_1$ will be called the length, as usual.

Denote a unit tangent vector to L at a point x by $\tau(x)$. Clearly, there exist only two such vectors: $\tau(x)$ and $-\tau(x)$. An orientation on a smooth curve L is a continuous family of unit tangent vectors defined on L. In other words, a continuous family $\tau = \{\tau(x)\}_{x\in L}$ is an orientation on L if $\|\tau(x)\| = 1$ and $\tau(x)$ is a tangent vector to L at x for all $x \in L$. A curve equipped with an orientation, i.e., the pair (L, τ), is called an oriented curve.

Using the coordinate functions $V_1, \ldots, V_m$ of the field V, the line integral of $\langle V, \tau\rangle$ can be written in the form

$$\int_L \langle V(x), \tau(x)\rangle d\sigma(x) = \int_{(L,\tau)} V_1(x)dx_1 + \cdots + V_m(x)dx_m \tag{1}$$

It is also denoted by $\int_L V_1(x)dx_1 + \cdots + V_m(x)dx_m$; the latter notation does not explicitly indicate the orientation (which is assumed to be given).

Clearly, reversing the orientation from $\tau = \{\tau(x)\}_{x\in L}$ to $\{-\tau(x)\}_{x\in L}$ changes the sign of the line integral.

On a connected curve there are only two opposite orientations. Indeed, if $\{\tilde\tau(x)\}_{x\in L}$ is an orientation on L, then the function $x \mapsto \langle\tau(x), \tilde\tau(x)\rangle$ is continuous on L and takes only the values \pm 1 . By the connectedness, this function is constant on L, which implies that $\tilde\tau$ coincides either with τ or with the opposite orientation. Note that in order to define an orientation on a connected curve, it suffices to define a tangent vector only at one point.

Using a smooth parametrization $\gamma : (a, b) \mapsto \mathbb{R}^m$ of a simple curve L, one can easily construct an orientation on L which we will call the orientation corresponding to γ. It is defined by the formula

$$\tau(x) = \frac{\gamma'\left(\gamma^{-1}(x)\right)}{\|\gamma'\left(\gamma^{-1}(x)\right)\|} \quad (x \in L)$$

This implies (see Theorem 8.4.1) a formula for computing the integral (1):

$$\int_{(L,\tau)} V_1(x)dx_1 + \cdots + V_m(x)dx_m = \int_a^b \langle V(\gamma(t)), \gamma'(t)\rangle\, dt$$

This leads to a useful generalization. Let γ be a piecewise smooth path in $\mathcal{O}$ defined on $[a, b]$. The integral over γ of a vector field V is the integral on the right-hand side of the last formula. It is denoted by $\int_\gamma V_1(x)dx_1 + \cdots + V_m(x)dx_m$.

Now we will explain how one can interpret the integral (1) over an oriented curve (L, τ). Assume that V is a force field and L is a curve contained in $\mathcal{O}$. Consider a small L-neighborhood U of a point $x \in L$. In view of the smallness of U, we may assume that this piece of the curve is almost straight and the field V is almost constant on it. Hence the work done by the force V along U must be approximately equal to the work done by the constant force $V(x)$ in moving the particle by the vector $\sigma(U)\tau(x)$. The latter work is equal to $\langle V(x), \tau(x)\rangle\sigma(U)$. Thus it is natural to assume that the work $A(e, \tau)$ done by the force V along an arbitrary segment e of the oriented curve satisfies the estimates

$$\inf_{x\in e}\langle V(x), \tau(x)|\sigma(e) \leqslant A(e, \tau) \leqslant \sup_{x\in e}\langle V(x), \tau(x)|\sigma(e)$$

In addition, $A(e, \tau)$ depends additively on e. Under these assumptions, using the general scheme, we see that the work done by the force V in moving a particle along the oriented curve (L, τ) is given by the integral (1). It is clear that the integral over a piecewise smooth path lying in $\mathcal{O}$ has the same interpretation.

Definition *A vector field $V = (V_1, \ldots, V_m)$ defined in a domain $\mathcal{O}$ is called potential if there exists a smooth function F on $\mathcal{O}$ (a potential of V) such that $V(x) = \operatorname{grad} F(x)$ for all points $x \in \mathcal{O}$.*

In the case of a potential field, the integral $\int_\gamma V_1(x)dx_1 + \cdots + V_m(x)dx_m$ satisfies the so-called gradient theorem, or the fundamental theorem of calculus for line integrals.

Proposition 1 *Let F be a potential of a vector field $V = (V_1, \ldots, V_m)$ defined in a domain $\mathcal{O}$, and let γ be a piecewise smooth path in $\mathcal{O}$ starting at A and ending at B. Then*

$$\int_\gamma V_1(x)dx_1 + \cdots + V_m(x)dx_m = F(B) - F(A)$$

Proof It suffices to prove the assertion only for a smooth path γ. We assume that it is defined on an interval $[a, b]$, so that $A = \gamma(a)$ and $B = \gamma(b)$. It is easy to check that $\langle V(\gamma), \gamma'\rangle = (F(\gamma))'$. Hence

$$\begin{aligned}\int_\gamma V_1(x)dx_1 + \cdots + V_m(x)dx_m = \int_a^b \langle V(\gamma(t)), \gamma'(t)\rangle\, dt = \int_a^b (F(\gamma(t)))'dt \\ = F(\gamma(b)) - F(\gamma(a)) = F(B) - F(A).\end{aligned}$$

□

Thus the work done by a potential field along a path depends only on the values of the potential at its endpoints and is equal to the increment of the potential. In this case, one says that the integral is path-independent. Crucially, the converse is also true.

Proposition 2 *If a line integral is path-independent, then the corresponding vector field is potential.*

Proof Note that any two points A and B of a domain $\mathcal{O}$ can be joined by a piecewise smooth path lying in $\mathcal{O}$. Let V be a vector field in $\mathcal{O}$ for which all integrals along such paths coincide. Denote their common value by $\int_A^B V_1(z)dz_1 + \cdots + V_m(z)dz_m$. Fixing a point $A \in \mathcal{O}$, consider the "integral with variable upper limit"

$$F(x) = \int_A^x V_1(z)dz_1 + \cdots + V_m(z)dz_m \quad (x \in \mathcal{O}).$$

It is easy to see that $F(y) - F(x) = \int_x^y V_1(z)dz_1 + \cdots + V_m(z)dz_m$ (to check this, write $F(y)$ as the integral over a path passing through x). Let us show that F is a potential of V. Fix x and consider an arbitrary vector e_j of the canonical basis. For a sufficiently small real t, we have

$$\begin{aligned} F\left(x + te_j\right) - F(x) &= \int_x^{x+te_j} V_1(z)dz_1 + \cdots + V_m(z)dz_m \\ &= \int_0^t \left\langle V\left(x + se_j\right), e_j\right\rangle ds \\ &= \int_0^t V_j\left(x + se_j\right) ds \\ &= t\int_0^1 V_j\left(x + tue_j\right) du. \end{aligned}$$

Therefore, $F\left(x + te_j\right) - F(x) = t\left(V_j(x) + o(1)\right)$ as $t \to 0$, i.e., $\frac{\partial F}{\partial x_j}(x) = V_j(x)$.□

In the case of path-independence, the integral over a closed path vanishes (recall that a path is called closed if both its endpoints coincide). However, in the general case, this is not true.

Example Consider the force field $V(x, y) = \left(-\frac{y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$ defined in the "punctured" plane $\mathbb{R}^2\backslash\{0\}$. Let us compute its work A along a circle, or, more precisely, along the closed path $\gamma(t) = (\cos t, \sin t)$, where $t \in [0, 2\pi]$:

$$A = \int_\gamma -\frac{y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy = \int_0^{2\pi} dt = 2\pi$$

This example shows that the work done by a non-potential field along a closed path may be non-zero. Note that the restriction of the field under consideration to every half-plane that does not contain the origin is potential. In particular, its restriction to the half-plane $x > 0$ is the gradient field of the function $F(x, y) = \arctan\frac{y}{x}$.

For smooth fields, one can easily establish a simple and important necessary condition for potentiality. Indeed, if F is a potential of a smooth field $V = (V_1, \ldots, V_m)$ in a domain $\mathcal{O}$, then, by the mixed derivatives (or Clairaut's) theorem, we have

$$\frac{\partial V_k}{\partial x_j}(x) = \frac{\partial^2 F}{\partial x_j \partial x_k}(x) = \frac{\partial^2 F}{\partial x_k \partial x_j}(x) = \frac{\partial V_j}{\partial x_k}(x)$$

□

Thus the equalities

$$\frac{\partial V_k}{\partial x_j}(x) = \frac{\partial V_j}{\partial x_k}(x) \quad (x \in \mathcal{O}, j, k = 1, \ldots, m)$$

are necessary conditions for the field V to be potential. As the above example shows, in the general case, these conditions are not sufficient. However, in "good" domains, they are. Leaving aside the thorough investigation of this problem, we restrict ourselves to a special case of the result known as the Poincaré lemma.

Proposition 3 *A smooth field $V = (V_1, \ldots, V_m)$ defined in a convex domain $\mathcal{O}$ and satisfying condition (2) is potential.*

Proof To simplify formulas, we assume that $\mathcal{O}$ contains the origin. Then for every point $x = (x_1, \ldots, x_m)$ in $\mathcal{O}$, the straight path $\gamma_x(t) = tx, t \in [0, 1]$, lies in $\mathcal{O}$. Set $F(x) = \int_{\gamma_x} V_1(z)dz_1 + \cdots + V_m(z)dz_m$; we will show that F is a potential of V. Indeed,

$$F(x) = \int_0^1 \langle V(\gamma_x(t)), \gamma_x'(t) \rangle \, dt = \sum_{k=1}^m x_k \int_0^1 V_k(tx)dt$$

Differentiating with respect to x_j and using (2), we obtain

$$\begin{aligned}
\frac{\partial F}{\partial x_j}(x) &= \int_0^1 V_j(tx)dt + \sum_{k=1}^m x_k \int_0^1 t \frac{\partial V_k}{\partial x_j}(tx)dt \\
&= \int_0^1 \left(V_j(tx) + t \sum_{k=1}^m x_k \frac{\partial V_j}{\partial x_k}(tx) \right) dt \\
&= \int_0^1 (tV_j(tx))_t' \, dt = V_j(x).
\end{aligned}$$

□

Remark The proof does not use the convexity of the domain in full strength. In particular, it remains valid for star domains ($\mathcal{O}$ is a star domain if there is a point $x_0 \in \mathcal{O}$ such that the line segment $\{x_0 + t(x - x_0) \mid t \in [0, 1]\}$ lies in $\mathcal{O}$ for every $x \in \mathcal{O}$).

We will say that a vector field defined in a domain $\mathcal{O}$ is locally potential if every point of $\mathcal{O}$ has a neighborhood in which the field has a potential. Proposition 3 implies an obvious but useful corollary.

Corollary *A smooth field is locally potential if and only if it satisfies condition (2).*

The above example shows that a locally potential field may not be globally potential, and the integral of a locally potential field over a closed path may be non-zero.

2.5.3

Side of a Surface and the Flow of a Vector Field. Now we proceed to case II. Consider integrals of the form $\int_M \langle V(x), v(x)\rangle d\sigma(x)$, which often appear in problems of physics and mechanics. Here M is a smooth surface, $v(x)$ is a unit normal to M at a point x, and $V(x)$ is a vector corresponding to the problem under investigation.

Recall that a normal to a smooth surface $M \subset \mathbb{R}^m$ at a point $x \in M$ is a non-zero vector orthogonal to the tangent space T_x. A unit normal is a normal of unit length. At every point of a surface there exist only two (opposite) normals.

A side of a smooth surface M is a continuous family of unit normals defined on M. In other words, a continuous family $\{\nu(x)\}_{x\in M}$ is a side of M if $\|v(x)\| = 1$ and $v(x)$ is a normal to M at x for every $x \in M$.

Using our intuitive notion of the surface area as a value proportional to the amount of paint needed to paint it (as mentioned at the beginning of this chapter), we may now say that a side of a surface may be thought of as the surface together with a coat of paint, or as the collection of all positions of the paintbrush. There is also a more widely used everyday interpretation, that of the "visible side". This is determined by the part of the surface that is "visible" to an observer, or, more exactly, by the part on which the normals are oriented "towards" the visual ray (making an obtuse angle with it).

Now we proceed from an informal discussion to necessary elaborations related to the notion of a side of a surface. If $v = \{v(x)\}_{x\in M}$ is a side of a surface M, then, obviously, the opposite family $\{-v(x)\}_{x\in M}$ is also a side of M. On a connected surface, there are no other sides (to prove this, it suffices to reproduce almost literally the argument used when considering an orientation on a curve). With this in mind, surfaces on which there is a side are called two-sided. To indicate a side of a connected surface, it suffices to define a normal at least at one point.

Clearly, if a smooth surface M has a global parametrization Φ, then one can easily construct a side of M using the vector N_Φ (see Sect. 2.3.4):

$$v(x) = \frac{N_\Phi\left(\Phi^{-1}(x)\right)}{\|N_\Phi\left(\Phi^{-1}(x)\right)\|} \quad (x \in M)$$

We say that this side is generated by Φ, or corresponds to Φ.

The graph Γ_φ of a smooth function φ is a two-sided surface. Its canonical parametrization generates the side

$$x = (u, \varphi(u)) \mapsto v(x) = \frac{(-\operatorname{grad}\varphi(u), 1)}{\sqrt{1 + \|\operatorname{grad}\varphi(u)\|^2}}$$

Note that all vectors of this side make acute angles with the x_m axis. Hence we will say that it is the upper side of the graph and the opposite side is the lower side.

Another important example of a two-sided surface, the boundary of a "sufficiently good" compact set, will be considered in the next section.

We see from the above that every sufficiently small M-neighborhood of every point of a smooth surface M has a side. However, this does not mean that the whole surface M also has a side. A counterexample is the surface called the Möbius strip, which can be obtained by "giving a half-twist" to a rectangle and then "gluing together" its opposite sides. Speaking more formally, given a rectangle $[-a, a]\times\ (-b, b)$, we identify the centrally symmetric points lying at the vertical sides (note that by identifying the points symmetric with respect to the y axis, we will obtain an ordinary cylindric surface, which is obviously two-sided). It can be proved that one cannot define a side on the Möbius strip. We encourage the reader to experiment by painting the surface obtained by gluing a twisted narrow rectangular strip of paper. Smooth surfaces on which one cannot define a side are called one-sided.

Having chosen a side $\{v(x)\}_{x\in M}$ of a two-sided surface lying in a domain where a vector field V is defined, we can consider the surface integral

$$\int_M \langle V(x), v(x)\rangle d\sigma(x) \tag{2}$$

(reversing the side obviously changes the sign of the integral). If the chosen side is generated by a parametrization $\Phi \in C^1(G)$, then the computation of this integral reduces to the computation of a multiple integral (see Theorem 2.4.1 and the formula for N_Φ in Sect. 2.3.4):

$$\begin{aligned}\int_M \langle V(x), v(x)\rangle d\sigma(x) &= \int_G \langle V(\Phi(u)), N_\Phi(u)\rangle\, du \\ &= \int_G \begin{vmatrix} V_1(\Phi(u)) & \dots & V_m(\Phi(u)) \\ \frac{\partial\varphi_1(u)}{\partial u_1} & \dots & \frac{\partial\varphi_m(u)}{\partial u_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial\varphi_1(u)}{\partial u_{m-1}} & \dots & \frac{\partial\varphi_m(u)}{\partial u_{m-1}} \end{vmatrix} du.\end{aligned}$$

Now we turn to a physical problem leading to the integral (4). Assume that in a domain $\mathcal{O} \subset \mathbb{R}^m$ there is a vector field V, which we regard as the velocity field of a stationary fluid flow. How can one compute the amount of fluid flowing through a smooth two-sided surface $M \subset \mathcal{O}$ per unit time? When solving this problem, one should bear in mind that particles of fluid can traverse the surface in different directions "moving from one side to the other". If the surface bounds a body, this means that the fluid may flow out of it as well as into it. Hence, to make our problem more definite, we fix a side $\{v(x)\}_{x\in M}$ of M.

Consider a small M-neighborhood U of a point $x \in M$. Then we may assume that this piece of the surface is almost planar and the velocity V is

almost constant on it. Hence the fluid flowing through U per unit time fills a curved parallelotope

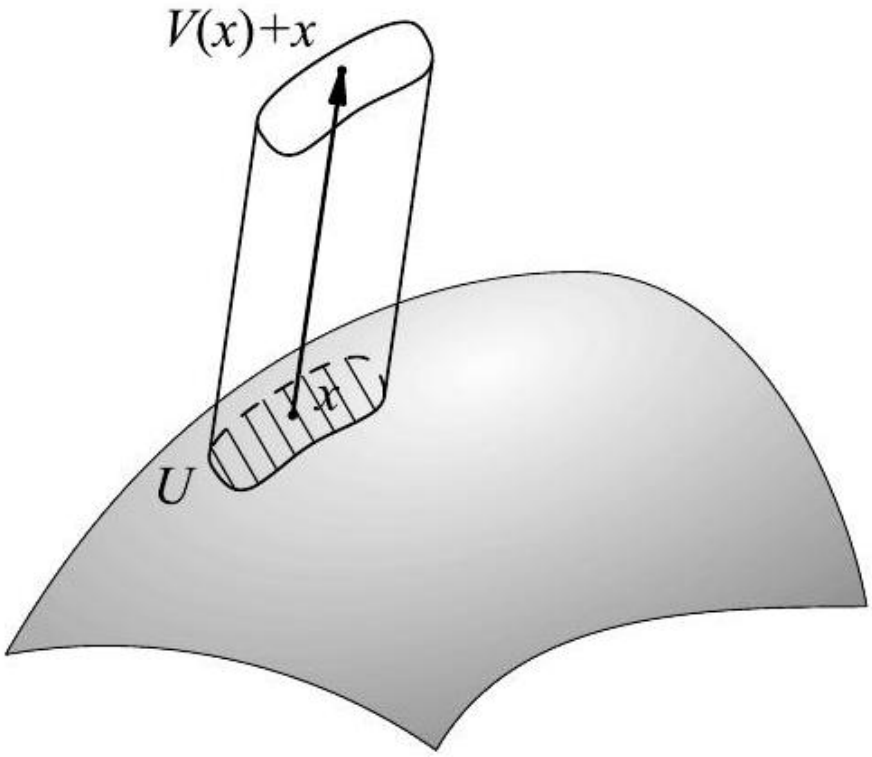

Fig. 2.2 The parallelotope with base U and edges equal to $V(x)$

close to the parallelotope with base U and edges equal to $V(x)$. The volume of the latter is equal to $\sigma(U)|\langle V(x), v(x)\rangle|$ (see Fig. 2.2).

The inner product $\langle V(x), v(x)\rangle$ is positive if the vectors $V(x)$ and $v(x)$ make an acute angle, i.e., if the fluid traverses M "in the direction $v(x)$ ", and is negative otherwise. Hence the absolute value of the integral $\int_U \langle V(x), \nu(x)\rangle d\sigma(x)$ is equal to the amount of fluid flowing through U per unit time. Its sign depends on the choice of a side of the surface and characterizes the direction of the fluid motion. In view of these considerations, the integral (4) is called the flow of the vector field V through M in the given direction.

Example Let f be a smooth function on a domain $\mathcal{O} \subset \mathbb{R}^m$ that has no critical points. Set

$$v(x) = \frac{1}{\| \operatorname{grad} f(x)\|} \operatorname{grad} f(x), \quad V(x) = \frac{1}{\| \operatorname{grad} f(x)\|} v(x) \quad (x \in \mathcal{O})$$

It is clear that the family $\{v(x)\}_{x \in M_C}$ is a side of the level surface $M_C = \{x \in \mathcal{O} \mid f(x) = C\}$. The flow of v in this direction is just the area of M_C.

The flow of V through M_C also has a simple geometric interpretation: it is the derivative at $u = C$ of the volume of the set $\mathcal{O}_u = \{x \in \mathcal{O} \mid f(x) \leqslant u\}$ (see Remark (3) in Sect. 2.4.3).

2.5.4

Having discussed integration of vector fields over manifolds of minimal (Sect. 2.5.2) and maximal (Sect. 2.5.3) dimension, a few words are in order regarding integration over plane curves. In this situation, the maximal and the minimal dimensions coincide (both are equal to 1). Hence a plane curve L has not only a direction, but also a side. Formally, we obtain two types of line integrals of a vector field $V = (V_1, V_2)$ over L. First, the integral over an oriented curve

$$\int_{(L,\tau)} V_1(x,y)dx + V_2(x,y)dy = \int_L \langle V(x,y), \tau(x,y)\rangle d\sigma_1(x,y)$$

second, the integral over L corresponding to a side $v = \{v(x,y)\}_{(x,y)\in L}$:

$$\int_L \langle V(x,y), v(x,y)\rangle d\sigma_1(x,y)$$

One can easily see that there is a close relation between these two integrals. To make it precise, consider the orthogonal transformation $z = (x,y) \mapsto U(z) = (-y,x)$ that rotates a vector z by $\pi/2$ "counterclockwise" (identifying $\mathbb{R}^2$ with the set of complex numbers $\mathbb{C}$, we can write it simply as $U(z) = iz$). Since by rotating a normal by a right angle we obtain a tangent vector, every side v of L gives rise to an orientation $\tau = U(\nu)$. Conversely, every orientation τ on L gives rise to the side $v = U^{-1}(\tau)$. Given an orientation and a side related by the formula $\tau = U(\nu)$, we say that they agree with each other. Obviously, $\langle V, v\rangle = \langle \bar{V}, \tau\rangle$, where $\bar{V} = U(V)$, and hence the flow of V in the direction v is equal to the integral of the field $\bar{V} = U(V)$ over the oriented curve (L,τ), where $\tau = U(\nu)$:

$$\int_L \langle V, v\rangle d\sigma_1 = \int_L \langle \bar{V}, \tau\rangle d\sigma_1$$

This equality can be rewritten in the form

$$\int_L \langle V, v\rangle d\sigma_1 = \int_{(L,\tau)} -V_2(x,y)dx + V_1(x,y)dy \tag{3}$$

We will use this in the next section when discussing Green's formula (Sect. 2.6.7).

2.6 The Gauss-Ostrogradski Formula

2.6.1

The classical integral calculus is based on the Newton-Leibniz formula

$$\int_a^b f'(x)dx = f(b) - f(a)$$

which expresses the integral of the derivative in terms of the values of the function at the endpoints of the interval of integration.

What is an analog of this formula in the multi-dimensional case? It is natural to assume that, for a function of several variables, one should replace f' by a partial derivative, the interval by a compact set K, and consider the integral $\int_K \frac{\partial f}{\partial x_j} d\lambda_m$. Is it possible to express it by a formula including the values of f on the boundary of the integration domain only? The goal of this section is to show that the answer to this question is affirmative under very wide assumptions on the structure of the boundary of the set K.

The simplest version of the formula we seek can be obtained using Fubini's theorem. Integrating the partial derivative $\frac{\partial f}{\partial x_m}$ of the function f that is smooth on the parallelepiped $P = Q \times [a, b]$, where $Q \subset \mathbb{R}^{m-1}$, we get

$$\int_P \frac{\partial f}{\partial x_m}(x)dx = \int_Q \left(\int_a^b \frac{\partial f}{\partial x_m}(u, v)dv \right) du = \int_Q f(u, b)du - \int_Q f(u, a)du$$

(we identify a point x from P with the pair $(u, v), u \in Q, v \in [a, b]$). The integrals on the right-hand side are simply the integrals of f over the top and the bottom bases of the parallelepiped P. Denoting these parts of the boundary of P by the symbols $\partial P_+$and ∂P_-, one can, obviously, write

$$\int_P \frac{\partial f}{\partial x_m}(x)dx = \int_{\partial P_+} f(x)d\sigma_{m-1}(x) - \int_{\partial P_-} f(x)d\sigma_{m-1}(x). \qquad (1)$$

The next step is crucial for our argument. In the situation that arose above, one should ponder over the fact that we have to consider the difference and not, say, the sum of the integrals over $\partial P_+$and ∂P_-. It is desirable to find an explanation of this phenomenon that would allow one to get rid of the "asymmetry" between the integrals over $\partial P_+$and ∂P_-. This can be done using the notion of the outer normal, which will enable us to rewrite the right-hand side of the equality (1) as an integral over the boundary of the parallelepiped P.

To do this, consider the outer normal v on ∂P. We postpone the precise definition of this notion until the next subsection. Still, using intuitive considerations, one can say that the outer normal coincides with the vector e_m on ∂P_+, coincides with $(-e_m)$ on ∂P_-, and is orthogonal to e_m on the rest of ∂P. Thus, the formula (1) can be rewritten as

$$\int_P \frac{\partial f}{\partial x_m}(x)dx = \int_{\partial P_+} f(x)\,\langle v(x), e_m\rangle\, d\sigma_{m-1}(x) + \int_{\partial P_-} f(x)\,\langle v(x), e_m\rangle\, d\sigma_{m-1}(x)$$
$$= \int_{\partial P_+\cup\partial P_-} f(x)\,\langle v(x), e_m\rangle\, d\sigma_{m-1}(x)$$
$$= \int_{\partial P} f(x)\,\langle v(x), e_m\rangle\, d\sigma_{m-1}(x)$$

Taking into account that the partial derivative $\frac{\partial f}{\partial x_m}$ is the directional derivative in the direction e_m, it is useful to transform this equality into

$$\int_P \frac{\partial f}{\partial e_m}(x)dx = \int_{\partial P} f(x)\,\langle v(x), e_m\rangle\, d\sigma_{m-1}(x) \tag{2}$$

emphasizing the connection of the integrand on the right with the direction of the differentiation on the left.

Thus, we have obtained the simplest version of the classical Gauss-Ostrogradski formula, which is precisely the generalization of the Newton-Leibniz formula we are aiming at.

Note that in the one-dimensional case, the Newton-Leibniz formula can be interpreted as a special case of the formula (2) if one considers the interval $[a, b]$ as a parallelepiped P, and defines the measure σ_0 on its boundary as the sum of two unit point masses at the points a and b and the "unit normals" at these points as the vectors $-e$ and e correspondingly where e is the unit vector on the real line.

Even now, the reader can easily check that in the equality (2), one can replace $\frac{\partial f}{\partial e_m}$ by the partial derivative with respect to any other coordinate or, more generally, the directional derivative in any direction. It is much harder to prove that the formula we obtained is valid not only for parallelepipeds, but also for more general compact sets. The description of such sets together with the verification of the corresponding equality are the main topics of this section. The final result will be obtained as the outcome of the process of the gradual extension of the class of admissible sets.

Everywhere in this section, we assume that $m > 1$. The surface area and the Lebesgue measure λ_m will be denoted by the letters σ and λ respectively without specifying the dimension explicitly.

2.6.2

Let $A \subset \mathbb{R}^m$ and let $p \in \partial A$. If near the point p the boundary ∂A coincides with a smooth surface M, then a normal $N(p)$ to M at the point p is called a normal to the boundary of A.

The normal $N(p)$ is called an outer normal to ∂A at the point p if $p+tN(p) \notin A$ and $p - tN(p) \in \operatorname{Int} A$ for all sufficiently small positive t. The side of M consisting of outer normals is called the outer side of M, and the opposite side is called the inner side. In the case when the set A can be defined by an

inequality near the point $p \in \partial A$, i.e., $A \cap B(p,\delta) = \{x \in B(p,\delta) \mid F(x) \leqslant 0\}$ where F is a function smooth in the ball $B(p,\delta)$ with non-vanishing gradient, the corresponding part of the boundary ∂A is nothing but the zero level set of the function F. In this case $\operatorname{grad} F(p)$ is an outer normal to ∂A because the function F strictly increases in the direction of the gradient.

Later we shall need some special sets closely related to the subgraph of a smooth function. To define them, we will identify the space $\mathbb{R}^m$ with the Cartesian product $\mathbb{R}^{m-1} \times \mathbb{R}$ and will write a point x from $\mathbb{R}^m$ as $x = (u, v)$ where $u \in \mathbb{R}^{m-1}$ and $v \in \mathbb{R}$. Let φ be a function smooth on the closed cube $Q \subset \mathbb{R}^{m-1}$. The sets

$$\{(u,v) \mid u \in Q, c \leqslant v \leqslant \varphi(u)\} \quad \text{and} \quad \{(u,v) \mid u \in Q, \varphi(u) \leqslant v \leqslant d\}$$

where $c < \varphi$ and, respectively, $\varphi < d$, will be called the lower and the upper beams. The sets obtained from them by reenumerations of the coordinates will be called beams.

The points (u, v) of the graph of the function φ whose projections u lie in the interior of the cube Q form the non-trivial part of the beam boundary. The remaining points on the boundary form its trivial part (for the lower beam, this is contained in the boundary of the infinite parallelepiped $Q \times [c, +\infty)$).

If the function φ is not constant, the non-trivial part of the beam boundary is determined uniquely. Otherwise the non-trivial part should be specified explicitly (for example, every face of a cubic beam can be its non-trivial part).

It is clear that the beam boundary consists of finitely many compact subsets of smooth surfaces. The outer normal $v(x)$ to the boundary exists at every point x of smoothness, i.e., almost everywhere with respect to σ.

Completing the definition from Sect. 2.5.3, we will call the family of outer normals $\{v(x)\}_{x \in \partial\mathcal{B}}$ the outer side of the boundary of the beam $\mathcal{B}$. Note that, according to this definition, the outer side is defined not everywhere but only almost everywhere on $\partial\mathcal{B}$. Also, for every vector $e \in \mathbb{R}^m$, the function $x \mapsto \langle v(x), e\rangle$ is continuous almost everywhere on $\partial\mathcal{B}$ and, therefore, is measurable.

Since, near each point of the non-trivial part of the boundary, the lower beam is described by the inequality $F(u,v) = v - \varphi(u) \leqslant 0$, the gradient $\operatorname{grad} F(u,v) = (-\operatorname{grad}\varphi(u), 1)$ is an outer normal. In other words, the outer side of the non-trivial part of the lower beam is the upper side of the graph Γ_φ, and the unit outer normal at the point $x = (u, \varphi(u))$ belonging to this part of the boundary is equal to

$$v(x) = \frac{(-\operatorname{grad}\varphi(u), 1)}{\sqrt{1 + \|\operatorname{grad}\varphi(u)\|^2}} \tag{3}$$

The outer side of the non-trivial part of the boundary of the upper beam is the lower side of the graph consisting of the opposite normals.

2.6.3

The next theorem constitutes the first step in the generalization of the formula (2). The symbol v will denote the outer side of the beam boundary.

Theorem *Assume that a function f smooth on the beam $\mathcal{B} \subset \mathbb{R}^m$ vanishes on the trivial part of its boundary. Then for every unit vector $e \in \mathbb{R}^m$, the equality*

$$\int_{\mathcal{B}} \frac{\partial f}{\partial e}(x)dx = \int_{\partial\mathcal{B}} f(x)\langle v(x), e\rangle d\sigma(x) \tag{4}$$

holds.

Proof Changing the enumeration of the coordinates, if necessary, we may assume that $\mathcal{B}$ is either a lower beam, or an upper one. Since the arguments are essentially the same in these two cases, we will restrict ourselves to the consideration of the lower beam. By definition, it is a set of the form

$$\mathcal{B} = \{(u, v) \mid u \in Q, c \leqslant v \leqslant \varphi(u)\},$$

where φ is a smooth function on the closed cube $Q \subset \mathbb{R}^{m-1}$ satisfying the condition $\varphi > c$. Note that, since the directional derivative is a linear combination of partial derivatives, it suffices to prove the equality (4) for the case when e is one of the vectors $e_1, \ldots, e_m$ in the canonical basis of $\mathbb{R}^m$.

First, consider the case $e = e_m$. By Fubini's theorem, we have

$$\int_{\mathcal{B}} \frac{\partial f}{\partial e_m}(x)dx = \int_Q \left(\int_c^{\varphi(u)} \frac{\partial f}{\partial v}(u, v)dv \right) du = \int_Q (f(u, \varphi(u)) - f(u, c))du$$

In addition, $f(u, c) = 0$ because the function f vanishes on the trivial part of the beam boundary. Therefore,

$$\begin{aligned} \int_{\mathcal{B}} \frac{\partial f}{\partial e_m}(x)dx &= \int_Q f(u, \varphi(u))du \\ &= \int_Q \frac{f(u, \varphi(u))}{\sqrt{1 + \| \operatorname{grad} \varphi(u)\|^2}} \sqrt{1 + \| \operatorname{grad} \varphi(u)\|^2} du. \end{aligned}$$

In view of equality (3), this means that

$$\int_{\mathcal{B}} \frac{\partial f}{\partial e_m}(x)dx = \int_{\Gamma_\varphi} f(x)\, \langle v(x), e_m\rangle\, d\sigma(x) = \int_{\partial\mathcal{B}} f(x)\, \langle v(x), e_m\rangle\, d\sigma(x)$$

(in the last equality, we again used the assumption $f \equiv 0$ on $\partial\mathcal{B}\backslash\Gamma_\varphi$).

Let now $e = e_k, 1 \leqslant k < m$. The proof is the same for all such k, so we may consider $k = m - 1$ only. We may assume that $m \geqslant 3$. In the two-dimensional case, the argument, which the reader can easily check himself, is much simpler.

Represent the cube Q as the product $Q = R \times [a, b]$, where R is a cube in $\mathbb{R}^{m-2}$. We will write a point u from Q as $u = (s, t)$ where $s \in R$ and $a \leqslant t \leqslant b$. Using this notation, we get

$$\int_{\mathcal{B}} \frac{\partial f}{\partial e_{m-1}}(x)dx = \int_R \left(\int_a^b \left(\int_c^{\varphi(s,t)} \frac{\partial f}{\partial t}(s, t, v)dv \right) dt \right) ds \tag{5}$$

by Fubini's theorem. To transform the inner integral (to swap the integration with respect to v and the differentiation with respect to t), we shall need a generalization of the Leibniz rule for differentiation of an integral depending on a parameter. This generalization is given by the following lemma.

Lemma *Let $\psi \in C^1([a, b])$ and $c < \psi(t)$ for $a \leqslant t \leqslant b$. If the function f is smooth in a neighborhood of the curvilinear trapezoid*

$$T = \left\{ (t, v) \in \mathbb{R}^2 \mid t \in [a, b], c \leqslant v \leqslant \psi(t) \right\},$$

then

$$\int_c^{\psi(t)} \frac{\partial f}{\partial t}(t, v)dv = \frac{d}{dt} \left(\int_c^{\psi(t)} f(t, v)dv \right) - f(t, \psi(t))\psi'(t)$$

Proof of Lemma Put $F(t, \theta) = \int_c^\theta f(t, v)dv$ for (t, θ) in a sufficiently small neighborhood of the trapezoid T. Since $F'_\theta(t, \theta) = f(t, \theta)$ and, by the Leibniz rule, $F'_t(t, \theta) = \int_c^\theta f'_t(t, v)dv$, we can differentiate the composition $F(t, \psi(t))$ to get

$$\begin{aligned} \frac{d}{dt} \left(\int_c^{\psi(t)} f(t, v)dv \right) &= (F(t, \psi(t)))'_t = F'_t(t, \psi(t)) + F'_\theta(t, \psi(t))\psi'(t) \\ &= \int_c^{\psi(t)} \frac{\partial f}{\partial t}(t, v)dv + f(t, \psi(t))\psi'(t) \end{aligned}$$

which is equivalent to the equality we sought to prove.□

Let us return to the proof of the theorem. Take $\psi(t) = \varphi(s, t)$ and apply the lemma to the inner integral on the right-hand side of Eq. (5):

$$\int_c^{\varphi(s,t)} \frac{\partial f}{\partial t}(s, t, v)dv = \frac{\partial}{\partial t} \left(\int_c^{\varphi(s,t)} f(s, t, v)dv \right) - f(s, t, \varphi(s, t)) \frac{\partial \varphi}{\partial t}(s, t).$$

Integrating this equality with respect to t, we obtain

$$\begin{aligned} &\int_a^b \left(\int_c^{\varphi(s,t)} \frac{\partial f}{\partial t}(s, t, v)dv \right) dt \\ &\quad = \int_c^{\varphi(s,t)} f(s, t, v)dv \Bigg|_{t=a}^{t=b} - \int_a^b f(s, t, \varphi(s, t)) \frac{\partial \varphi}{\partial t}(s, t)dt \end{aligned}$$

Since the points (s, a, v) and (s, b, v) belong to the trivial part of the beam boundary on which $f \equiv 0$, the substitution term vanishes. This allows us to rewrite Eq. (5) as

$$\int_{\mathcal{B}} \frac{\partial f}{\partial e_{m-1}}(x)dx = -\int_R \left(\int_a^b f(s,t,\varphi(s,t)) \frac{\partial \varphi}{\partial t}(s,t)dt \right) ds$$
$$= \int_Q f(u,\varphi(u)) \left\langle (-\operatorname{grad}\varphi(u), 1), e_{m-1} \right\rangle du.$$

Taking Eq. (3) into account, we can represent the resulting integral as an integral with respect to the measure σ :

$$\int_{\mathcal{B}} \frac{\partial f}{\partial e_{m-1}}(x)dx = \int_Q f(u,\varphi(u)) \left\langle v(u,\varphi(u)), e_{m-1} \right\rangle \sqrt{1 + \|\operatorname{grad}\varphi(u)\|^2} du$$
$$= \int_{\Gamma_\varphi} f(x) \left\langle v(x), e_{m-1} \right\rangle d\sigma(x) = \int_{\partial\mathcal{B}} f(x) \left\langle v(x), e_{m-1} \right\rangle d\sigma(x)$$

(here we used the condition $f \equiv 0$ on $\partial\mathcal{B}\backslash\Gamma_\varphi$ again). Thus, the Gauss-Ostrogradski formula for the beam $\mathcal{B}$ and the unit vectors $e_1, \ldots, e_{m-1}$ is proved, which also proves Eq. (4).□

The assumption that $f \equiv 0$ on the trivial part of $\partial\mathcal{B}$ is, of course, superfluous. It has been made only to simplify the proof of the preliminary version of the GaussOstrogradski formula. The final version (see Sect. 2.6.5) contains no such assumption.

2.6.4 Standard Compact Sets.

We will now introduce the compact sets that will be used in the general Gauss-Ostrogradski formula. We shall obtain the formula for compacts sets with smooth boundary without using the results of this subsection (see the first stage of the proof of Theorem 2.6.5). Since our goal is to prove the Gauss-Ostrogradski formula for more general compact sets, not only for compact sets with smooth boundaries (on which all notions of surface area coincide), we will now abandon the consideration of arbitrary surface areas and instead use only the area proportional to the Hausdorff measure μ_{m-1}. This area will still be denoted by the letter σ.

Definition *A compact set $K \subset \mathbb{R}^m$ is called a standard compact set if its boundary can be represented as $\partial K = M \cup E$ where:*

(a) for every point $p \in M$, there exists a ball B_p centered at p and a function $F \in C^1(B_p)$ such that $F > 0$ on $B_p\backslash K$, $F \leqslant 0$ on $B_p \cap K$, and $\operatorname{grad} F(p) \neq 0$;

(b) $\sigma(M) < +\infty$;

(c) E is a compact set and $\sigma(E) = 0$.

Condition (a) implies that M is a smooth surface. We shall call M the regular part of the boundary of the compactum K, and E its singular part. Condition

(c) allows us to ignore the integral over the singular part when integrating over ∂K because it vanishes.

It is obvious that a beam is a standard compact set. As a rule, compact bounded by one or several smooth surfaces (e.g., a ball, a torus, or an m-dimensional annulus) are also standard compact sets. All bounded domains studied in school geometry (a polyhedron, a truncated cone, etc.) are also standard compact sets.

It is clear that the function F from condition (a) of the definition vanishes on $B_p \cap M$. As has already been pointed out in Sect. 8.6.2, $\operatorname{grad} F(p)$ is an outer normal to ∂K at the point $p \in M$. Therefore an outer normal exists at every point of M. The mapping that sends an arbitrary point of M to the unit outer normal at that point is continuous on M because, locally in a neighborhood of a point $p \in M$, the unit outer normals coincide with the normalized gradients of the function F. Thus, the family of unit outer normals forms a side of the surface M, which, according to the definition of Sect. 2.6.2, is an outer side of M. Since in a neighborhood of the point p, the level set $F(x) = 0$ coincides with the graph of some smooth function, there exists a sufficiently small open parallelepiped $P \subset B_p$ containing p for which the intersection $\bar{P} \cap K = \mathcal{B}_p$ is a beam. Obviously, the non-trivial part of its boundary coincides with $M \cap P$, and on this part, the outer normals to M are also outer normals to $\partial \mathcal{B}_p$.

It is important to have some sufficiently simple conditions that would allow us to check the equality $\sigma(E) = 0$ in condition (c). In particular, this is so if E is a subset of a smooth manifold L of codimension greater than 1 because, by property (5) from Sect. 2.3.3, we have $\sigma(L) = 0$. In what follows, we shall also use another condition that ensures the equality $\sigma(E) = 0$. To state it, let us remind the reader that the ε neighborhood of a set E is the open set $E_\varepsilon = \bigcup_{x \in E} B(x, \varepsilon)$ (it consists of all points $y \in \mathbb{R}^m$ for which $\operatorname{dist}(y, E) < \varepsilon$.

Definition *A set $E \subset \mathbb{R}^m$ is called negligible in $\mathbb{R}^m$ if the volume of its ε neighborhood E_ε satisfies $\lambda(E_\varepsilon) = o(\varepsilon)$ as $\varepsilon \to 0$.*

It is obvious that every negligible set is bounded. Since a set and its closure have the same ε-neighborhoods, the closure of every negligible set is negligible.

On the line, only the empty set is negligible. On the plane every finite set is negligible, but not every discrete set (see Example 6). The reader can easily check that the union of a finite family of negligible sets is negligible. In the space $\mathbb{R}^m$, $m \geqslant 3$, every bounded subset of an affine subspace L is negligible if $\dim L \leqslant m - 2$. The next proposition is useful when verifying condition (c) in the definition of a standard compact set.

Proposition *Every negligible subset of the space $\mathbb{R}^m$ has zero area.*

Proof Let $E \subset \mathbb{R}^m$ be a negligible set. As we have already mentioned, it is bounded. Let us check that $\sigma(E) = \alpha_{m-1}\mu_{m-1}(E) = 0$.

Fix an arbitrary $\varepsilon > 0$. We will call the points x and y ε-distinguishable if $\|x - y\| \geqslant \varepsilon$. Obviously, a bounded set can contain only finitely many pairwise ε-distinguishable points. Consider a set A consisting of the maximal possible number of ε-distinguishable points belonging to E. We have

$$E \subset \bigcup_{x \in A} B(x, \varepsilon)$$

because otherwise the set A could be augmented by a point from $E \setminus \bigcup_{x \in A} B(x, \varepsilon)$, which would contradict its maximality. Furthermore, the balls $B(x, \varepsilon/2)$ and $B(y, \varepsilon/2)$ centered at two different points of A are disjoint because the points in this set have distances at least ε between them. Since the balls $B(x, \varepsilon)(x \in A)$ form a cover of the set E, according to the definition of $\mu_{m-1}(E, \varepsilon)$, we get

$$\begin{aligned}\mu_{m-1}(E, \varepsilon) &\leqslant \sum_{x \in A} \varepsilon^{m-1} = \frac{2^m}{\alpha_m \varepsilon} \sum_{x \in A} \lambda(B(x, \varepsilon/2)) = \frac{2^m}{\alpha_m \varepsilon} \lambda \left(\bigcup_{x \in A} B(x, \varepsilon/2) \right) \\ &\leqslant \frac{2^m}{\alpha_m \varepsilon} \lambda (E_\varepsilon) \underset{\varepsilon \to 0}{\longrightarrow} 0.\end{aligned}$$

Hence $\mu_{m-1}(E) = \lim_{\varepsilon \to 0} \mu_{m-1}(E, \varepsilon) = 0$.□

For compact subsets of smooth surfaces, the converse statement also holds.

Lemma *If a compact subset E of a smooth surface M has zero area, it is negligible.*

Proof Every point of the surface has an M-neighborhood whose closure is contained in the graph of some smooth function. It is clear that the set E can be covered by finitely many such neighborhoods $U_n : E \subset \bigcup_{n=1}^N U_n$. Put

$$E_n = E \cap \bar{U}_n \quad (n = 1, \ldots, N)$$

Obviously, the sets E_n are compact and $E = \bigcup_{n=1}^N E_n$. Therefore, it suffices to prove the statement of the lemma for the sets E_n, which allows us to assume in what follows that M is the graph of a smooth function $\varphi \in C^1(G)$ where G is an open subset of the space $\mathbb{R}^{m-1}$.

As before, we will represent a point x of the space $\mathbb{R}^m$ as $x = (u, v)$ where $u \in \mathbb{R}^{m-1}, v \in \mathbb{R}$, identifying $\mathbb{R}^{m-1}$ with the plane $v = 0$. Let H be the projection of the set E to $\mathbb{R}^{m-1}$. By the δ-neighborhood of H, we will mean the δ-neighborhood in the space $\mathbb{R}^{m-1}$, preserving the notation H_δ for it. Choose $\delta > 0$ so small that H_δ is contained in G together with its closure $\overline{H_\delta}$, and put $L = \max_{u \in \overline{H_\delta}} \| \operatorname{grad} \varphi(u) \|$.

Since the canonical parametrization Φ of the graph Γ_φ is an expansion and $E = \Phi(H)$, we have $\lambda_{m-1}(H) \leqslant \sigma(\Phi(H)) = \sigma(E) = 0$. Therefore, $\lambda_{m-1}(H) = 0$. Since $\bigcap_{\varepsilon > 0} H_\varepsilon = H$, the upper semicontinuity of measure implies that

$$\lambda_{m-1} (H_\varepsilon) \underset{\varepsilon \to 0}{\longrightarrow} 0 \tag{6}$$

Consider the layer

$$A(\varepsilon) = \{(u, v) \in \mathbb{R}^m \,|\, u \in H_\varepsilon, |\, v - \varphi(u) \,| < (L + 1)\varepsilon\}$$

around the graph of φ over H_ε with $0 < \varepsilon < \delta$. Since $\lambda(A(\varepsilon)) = 2(L + 1)\varepsilon \lambda_{m-1} (H_\varepsilon)$, by (6), we have $\lambda(A(\varepsilon)) = o(\varepsilon)$ as $\varepsilon \to 0$. Hence, to prove that

E is negligible, it suffices to show that $E_\varepsilon \subset A(\varepsilon)$. Let $x = (u, v) \in E_\varepsilon$. Let us check that $x \in A(\varepsilon)$, that is, that $u \in H_\varepsilon$ and $|v - \varphi(u)| < (L + 1)\varepsilon$. By the definition of the ε-neighborhood, there exists a point $x' = (u', v') \in E \subset \Gamma_\varphi$ such that $\|x - x'\| < \varepsilon$. Since $\|u - u'\| \leqslant \|x - x'\| < \varepsilon$ and $u' \in H$, we have $u \in H_\varepsilon$. Furthermore, $v' = \varphi(u')$ and $|v - v'| \leqslant \|x - x'\|$. Thus

$$|v - \varphi(u)| \leqslant |v - v'| + |\varphi(u') - \varphi(u)| \leqslant \|x - x'\| + L\|u - u'\| < (L + 1)\varepsilon,$$

whence $x \in A(\varepsilon)$.□

Note that one cannot relax the conditions of the lemma by replacing compactness with boundedness. The lemma also fails if one assumes that E is contained not in one but in the union of two smooth surfaces.

We will now present a simple but useful corollary to this lemma.

Corollary *A compact subset E of a smooth manifold of codimension greater than 1 is negligible.*

Proof

Indeed, the area of such a manifold equals zero (see property (5) in Sect. 2.3.3). Moreover, locally it is contained in a manifold of codimension 1, i.e., in a surface (see the end of Sect. 2.1.1). Thus E can be covered by finitely many compact sets, each of which is contained in a smooth surface and has zero area. It remains to use the lemma.□

2.6.5

In this subsection, we will generalize the preliminary version of the Gauss-Ostrogradski formula obtained in Sect. 2.6.3 replacing beams by an arbitrary standard compact set. We will also call the outer side v of the surface M the outer side of ∂K. Thus, the outer side is defined and continuous almost everywhere on ∂K. Fixing an arbitrary vector $e \in \mathbb{R}^m$, we conclude that the function $x \mapsto \langle v(x), e\rangle$ is continuous almost everywhere on ∂K (with respect to the measure σ) and, thereby, measurable.

Theorem (The Gauss-Ostrogradski formula) *Let f be a function smooth on a standard compact set $K \subset \mathbb{R}^m$. Then for every unit vector $e \in \mathbb{R}^m$, one has*

$$\int_K \frac{\partial f}{\partial e}(x)dx = \int_{\partial K} f(x)\langle v(x), e\rangle d\sigma(x).$$

Before we start the proof, let us note that it will be carried out in three stages. For compact with smooth boundaries (a ball, a torus, etc.) the result that will be obtained at the first stage is enough. If the boundary of the compactum contains a singular part, in most cases, it is a negligible set (as it is for a polyhedron, a half-ball, a cone, etc.). This case will be covered at the second stage of the proof.

Proof By the definition of a standard compact set, $\partial K = E \cup M$, where M is the regular part of ∂K, and $\mu_{m-1}(E) = 0$.

(I) Assume that $f \equiv 0$ on an open set G containing E (this assumption on the function f is vacuous in the smooth boundary case, i.e., when $E = \varnothing$). We will construct a cover of K of a special form. For each point $x \in \operatorname{Int} K$, choose an open cube $Q_x \subset \operatorname{Int} K$ centered at x. For each point $p \in M$, choose an open parallelepiped R_p containing p such that the intersection $\overline{R_p} \cap K$ is a beam lying in the interior of K except for the closure of the non-trivial part of its boundary, which is contained in M. Such a parallelepiped exists by the definition of a standard compact set.

The sets $G, \{Q_x\}_{x \in \operatorname{Int} K}$ and $\{R_p\}_{p \in M}$ form an open cover of the compactum K. Let $G, Q_{x_1}, \ldots, Q_{x_J}$ and $R_{p_1}, \ldots, R_{p_N}$ be a finite subcover. Consider the partition of unity subordinate to this subcover (see Theorem 2.1.8). It consist of the smooth functions $\omega, \psi_1, \ldots, \psi_J$ and $\theta_1, \ldots, \theta_N$ ($\omega \equiv 0$ outside $G, \psi_j \equiv 0$ outside Q_{x_j} and $\theta_n \equiv 0$ outside R_{p_n} for $j = 1, \ldots, J, n = 1, \ldots, N$). We have

$$1 = \omega(x) + \sum_{j=1}^{J} \psi_j(x) + \sum_{n=1}^{N} \theta_n(x) \quad \text{for all } x \in K$$

Due to the condition $f \equiv 0$ on G, it follows that

$$f(x) = \sum_{j=1}^{J} \psi_j(x) f(x) + \sum_{n=1}^{N} \theta_n(x) f(x) \quad \text{for all } x \in K \tag{7}$$

Hence

$$\int_K \frac{\partial f}{\partial e}(x)dx = \sum_{j=1}^{J} \int_K \frac{\partial (\psi_j f)}{\partial e}(x)dx + \sum_{n=1}^{N} \int_K \frac{\partial (\theta_n f)}{\partial e}(x)dx$$

Taking into account that $\psi_j \equiv 0$ outside Q_{x_j} and $\theta_n \equiv 0$ outside R_{p_n}, we obtain

$$\int_K \frac{\partial f}{\partial e}(x)dx = \sum_{j=1}^{J} \int_{Q_{x_j}} \frac{\partial (\psi_j f)}{\partial e}(x)dx + \sum_{n=1}^{N} \int_{K \cap R_{p_n}} \frac{\partial (\theta_n f)}{\partial e}(x)dx \tag{8}$$

By Theorem 2.6.3, all terms in the first sum are equal to zero (because $\psi_j \equiv 0$ on the entire boundary of the cube Q_{x_j}). Let us transform the integrals in the second sum. To this end, note that $\theta_n \equiv 0$ on ∂R_{p_n} and, therefore, the function $\theta_n f$ vanishes on the trivial part of the boundary of the beam $K \cap \bar{R}_{p_n}$. Thus we can apply the Gauss-Ostrogradski formula for beams to these integrals too (see Theorem 2.6.3):

$$\int_{K \cap R_{p_n}} \frac{\partial (\theta_n f)}{\partial e}(x)dx = \int_{\partial (K \cap R_{p_n})} \theta_n(x) f(x) \langle v_n(x), e \rangle \, d\sigma(x)$$

where v_n is the unit outer normal to $\partial (K \cap R_{p_n})$. Since $\theta_n(x) \neq 0$ only on the nontrivial part of the boundary of the beam $K \cap \bar{R}_{p_n}$, i.e., on $M \cap R_{p_n}$, and since on that part v_n coincides with the unit outer normal v to M, we have

$$\int_{K\cap R_{p_n}} \frac{\partial(\theta_n f)}{\partial e}(x)dx = \int_{M\cap R_{p_n}} \theta_n(x)f(x)\,\langle v_n(x), e\rangle\, d\sigma(x)$$
$$= \int_M \theta_n(x)f(x)\langle v(x), e\rangle d\sigma(x)$$

(in the end, we have taken into account that $\theta_n \equiv 0$ outside R_{p_n}). Thus, Eq. (8) implies that

$$\int_K \frac{\partial f}{\partial e}(x)dx = \sum_{n=1}^N \int_M \theta_n(x)f(x)\langle v(x), e\rangle d\sigma(x)$$
$$= \int_M f(x)\sum_{n=1}^N \theta_n(x)\langle v(x), e\rangle d\sigma(x).$$

Since the functions $\psi_1, \ldots, \psi_J$ vanish on M, if follows from Eq. (7) that $f(x)\sum_{n=1}^N \theta_n(x) = f(x)$ for $x \in M$. Thus,

$$\int_K \frac{\partial f}{\partial e}(x)dx = \int_M f(x)\langle v(x), e\rangle d\sigma(x) = \int_{\partial K} f(x)\langle v(x), e\rangle d\sigma(x)$$

(II) Let us now turn to the case where the singular part E is negligible. We will verify that the difference

$$\Delta = \int_K \frac{\partial f}{\partial e}(x)dx - \int_{\partial K} f(x)\langle v(x), e\rangle d\sigma(x)$$

between the left- and the right-hand sides of the formula we wish to prove is arbitrarily small. To this end, fix an arbitrary positive number ε and apply Theorem 2.1.7 on a smooth descent to the set E_ε (it is easy to see that its ε-neighborhood coincides with $E_{2\varepsilon}$). We conclude that there exists a function $\theta \in C^\infty(\mathbb{R}^m)$ such that:

(a) $0 \leqslant \theta \leqslant 1$ on the entire space $\mathbb{R}^m$;

(b) $\theta = 1$ on E_ε;

(c) $\theta = 0$ outside $E_{2\varepsilon}$;

(d) $\|\operatorname{grad}\theta\| \leqslant \frac{C}{\varepsilon}$ everywhere on $\mathbb{R}^m$, where C is some constant depending only on the dimension m.

Put $L_0 = \max_K |f|$ and $L_1 = \max_K \left|\frac{\partial f}{\partial e}\right|$.

Since the function $(1-\theta)f$ vanishes on E_ε, we can apply to it the already proven part of the theorem. Therefore

$$\int_K \frac{\partial f}{\partial e}(x)dx = \int_K \frac{\partial(\theta f)}{\partial e}(x)dx + \int_K \frac{\partial(f-\theta f)}{\partial e}(x)dx$$
$$= \int_K \frac{\partial(\theta f)}{\partial e}(x)dx + \int_{\partial K}(1-\theta(x))f(x)\langle v(x), e\rangle d\sigma(x)$$

Hence,

$$\Delta = \int_K \frac{\partial(\theta f)}{\partial e}(x)dx - \int_{\partial K} \theta(x)f(x)\langle v(x), e\rangle d\sigma(x)$$

Due to property (c), one can reduce the sets of integration in both integrals to their intersections with $E_{2\varepsilon}$. This gives us the inequality

$$|\Delta| \leqslant \int_{K\cap E_{2\varepsilon}} \left(\left|\frac{\partial\theta}{\partial e}(x)\right| |f(x)| + |\theta(x)| \left|\frac{\partial f}{\partial e}(x)\right| \right) dx + \int_{E_{2\varepsilon}\cap\partial K} |\theta(x)f(x)| d\sigma(x).$$

So,

$$\begin{aligned} |\Delta| &\leqslant \int_{K\cap E_{2\varepsilon}} \left(\frac{C}{\varepsilon}L_0 + L_1\right) dx + \int_{E_{2\varepsilon}\cap\partial K} L_0 d\sigma(x) \\ &\leqslant \left(\frac{C}{\varepsilon}L_0 + L_1\right) \lambda\left(E_{2\varepsilon}\right) + L_0\sigma\left(E_{2\varepsilon}\cap\partial K\right). \end{aligned}$$

Since E is a negligible set, the term $\left(\frac{C}{\varepsilon}L_0 + L_1\right)\lambda\left(E_{2\varepsilon}\right)$ gets arbitrarily small as $\varepsilon \to 0$. The same can be said about the second term. Indeed,

$$\sigma\left(E_{2\varepsilon}\cap\partial K\right) \to \sigma(E) \quad \text{ as } \varepsilon \to 0$$

due to the upper semicontinuity of the measure σ (it is here that we use the finiteness of the area of the boundary of a standard compact set). It remains to recall that $\sigma(E) = 0$. Thus, $|\Delta| = 0$.

(III) Consider now the general case. We will carry out the proof as followings. We will start by improving the set K somewhat by expanding it and applying the Gauss-Ostrogradski formula to the expanded set, and then we will pass to the limit contracting the auxiliary set back to K.

Since $\mu_{m-1}(E) = 0$, one can fix an arbitrarily small positive number ε and choose the balls $B_j = B\left(x_j, r_j\right)$ so that

$$E \subset \bigcup_{j=1}^{\infty} B_j, \quad \sum_{j=1}^{\infty} r_j^{m-1} < \varepsilon^{m-1}$$

We will assume ε to be so small that the function f is continuously differentiable in $K_{2\varepsilon}$. Taking into account the compactness of the set E, one can assume that $E \subset \bigcup_{j=1}^{N} B_j \subset E_{2\varepsilon}$. Note also that, when intersecting a surface of finite area with concentric spheres, we will get sets of zero areas except, perhaps, for a countable set of radii because the family $\{\sigma(M\cap\partial B(a,r))\}_{r>0}$ is summable. Thus, without loss of generality, we may assume that

$$\sigma\left(M\cap\partial B_j\right) = 0 \quad (j = 1, \ldots, N)$$

Now, introduce the set $K(\varepsilon) = K \cup \bigcup_{j=1}^{N} \bar{B}_j$. Obviously, its boundary is disjoint with E and consists only of points belonging to the regular part of ∂K

or to the spheres $\partial B_1, \ldots, \partial B_N$. The boundary of $K(\varepsilon)$ can lose its smoothness only on the intersections of spheres or on the intersection of the set M with the spheres. Therefore the area of the singular part of the boundary of $K(\varepsilon)$ equals zero. This part consists of finitely many compact sets, each of which is negligible (by Lemma 2.6.4). Thus, their union is negligible too. Therefore, the set $K(\varepsilon)$ is a standard compact set whose boundary has a negligible singular part, so the Gauss-Ostrogradski formula is valid for this set:

$$\int_{K(\varepsilon)} \frac{\partial f}{\partial e}(x)dx = \int_{\partial K(\varepsilon)} f(x)\langle v(x), e\rangle d\sigma(x) \tag{9}$$

Putting

$$M'(\varepsilon) = \partial K \setminus \bigcup_{j=1}^{N} B_j, \quad M''(\varepsilon) = \partial K(\varepsilon) \setminus M'(\varepsilon)$$

and separating the integrals over K and M in Eq. (9) from the rest, we can rewrite it in the following form:

$$\begin{aligned} &\int_K \frac{\partial f}{\partial e}(x)dx + \int_{K(\varepsilon)\setminus K} \frac{\partial f}{\partial e}(x)dx \\ &\quad = \int_{M'(\varepsilon)} f(x)\langle v(x), e\rangle d\sigma(x) + \int_{M''(\varepsilon)} f(x)\langle v(x), e\rangle d\sigma(x) \\ &\quad = \int_M \cdots - \int_{M\setminus M'(\varepsilon)} \cdots + \int_{M''(\varepsilon)} \cdots \end{aligned} \tag{10}$$

Since $K(\varepsilon)\setminus K \subset E_{2\varepsilon}$ and $M\setminus M'(\varepsilon) \subset M \cap E_{2\varepsilon}$, we have

$$\lambda_m(K(\varepsilon)\setminus K) \leqslant \lambda_m\left(E_{2\varepsilon}\right), \quad \sigma\left(M\setminus M'(\varepsilon)\right) \leqslant \sigma\left(M \cap E_{2\varepsilon}\right)$$

Furthermore,

$$\sigma\left(M''(\varepsilon)\right) \leqslant \sum_{j=1}^{N} \sigma\left(\partial B_j\right) = m\alpha_m \sum_{j=1}^{N} r_j^{m-1} < m\alpha_m \varepsilon^{m-1}$$

The right-hand sides of these three inequalities tend to zero as $\varepsilon \to 0$. Thus, passing to the limit as $\varepsilon \to 0$ in Eq. (10), we obtain the desired formula. $\square$

Example The Gauss-Ostrogradski formula allows one to express the volume of a body as an integral over its boundary. For instance, applying this formula to the function $f(x) = \langle x, e\rangle$, we get

$$\lambda(K) = \int_{\partial K} \langle x, e\rangle\langle e, v(x)\rangle d\sigma(x)$$

In other words, the volume of the body K is equal to the flux of the vector field $V(x) = \langle x, e\rangle e$ through its boundary "outwards".

This result can be generalized as follows. Let L be a subspace of $\mathbb{R}^m$, and let P be the orthogonal projection to L. Then $\dim L \cdot \lambda(K) = \int_{\partial K} \langle P(x), \nu(x)\rangle d\sigma(x)$,

i.e., the flux of the projection P through the outer side is proportional to the volume of the compactum, the proportionality coefficient being equal to the dimension of the subspace to which one projects. In particular, for $L = \mathbb{R}^m$, we get $\lambda(K) = \frac{1}{m}\int_{\partial K}\langle x, v(x)\rangle d\sigma(x)$.

2.6.6

Now we transform the Gauss-Ostrogradski formula to clarify its physical meaning.

Let $\mathcal{O}$ be an open set in $\mathbb{R}^m$. Let $\{V(x)\}_{x\in\mathcal{O}}$ be a smooth vector field with the coordinate functions $V_1, \ldots, V_m$. According to the Gauss-Ostrogradski formula,

$$\begin{aligned}\int_{\partial K}\langle V(x), v(x)\rangle d\sigma(x) &= \sum_{j=1}^{m}\int_{\partial K} V_j(x)\,\langle e_j, v(x)\rangle\, d\sigma(x)\\ &= \sum_{j=1}^{m}\int_K \frac{\partial V_j}{\partial x_j}(x)dx = \int_K\left(\sum_{j=1}^{m}\frac{\partial V_j}{\partial x_j}(x)\right)dx\end{aligned}$$

where v is the outer side of the standard compact set $K \subset \mathcal{O}$. The left-hand side of this equation is the flux of the vector field V through the outer side of the boundary ∂K. The integrand $\sum_{j=1}^{m}\frac{\partial V_j}{\partial x_j}$ on the right-hand side is called the divergence of the vector field V and denoted $\operatorname{div} V$. Using this notation, the obtained formula can be rewritten in the following form (the so-called "vector form" of the GaussOstrogradski formula, or the divergence formula):

$$\int_K \operatorname{div} V(x)dx = \int_{\partial K}\langle V(x), v(x)\rangle d\sigma(x) \tag{11}$$

Note that $\operatorname{div} V(x)$ is simply the trace of the Jacobian matrix $\left(\frac{\partial V_j}{\partial x_k}(x)\right)_{j,k=1}^{m}$, i.e., the trace of the operator $d_x V$. Since the trace does not depend on the choice of the basis, when computing the divergence, one can use any orthonormal coordinate system, not only the canonical one.

The last result can also be established in another way. According to the mean value theorem, for every $a \in \mathcal{O}$ and for every sufficiently small $\varepsilon > 0$, one has

$$\frac{1}{\alpha_m \varepsilon^m}\int_{B(a,\varepsilon)} \operatorname{div} V(x)dx = \operatorname{div} V\left(x_\varepsilon\right), \qquad \text{where } x_\varepsilon \in B(a,\varepsilon)$$

Therefore,

$$\begin{aligned}\operatorname{div} V(a) &= \lim_{\varepsilon\to 0}\frac{1}{\alpha_m\varepsilon^m}\int_{B(a,\varepsilon)}\operatorname{div} V(x)dx\\ &= \lim_{\varepsilon\to 0}\frac{1}{\alpha_m\varepsilon^m}\int_{\|x-a\|=\varepsilon}\langle V(x), v(x)\rangle d\sigma(x)\end{aligned}$$

It can be seen from this that the value $\operatorname{div} V(a)$ does not depend on the choice of the coordinate system.

If one views V as an incompressible fluid velocity field, then the flux through the boundary of a body can be non-zero only if the body contains some sources (if the flux is positive) or sinks (if the flux is negative).

The quantity $\frac{1}{\alpha_m \varepsilon^m} \int_{\|x-a\|=\varepsilon} \langle V(x), \nu(x) \rangle d\sigma(x)$ on the right-hand side of the last equality characterizes the average intensity of the sources (sinks) in the ball $B(a, \varepsilon)$, and its limit $\operatorname{div} V(a)$ can be interpreted as the intensity of the source (sink) at the point a.

Example (The law of Archimedes) Let us show how the Gauss-Ostrogradski formula can be used to derive the law of Archimedes from Pascal's law. Let us remind the reader that, according to Pascal's law, the pressure a liquid exerts on a submersed flat area is directed along the normal to the area and is equal to the weight of the pillar of the liquid whose base is congruent to the submersed area and whose height is the submersion depth. Let us compute the Archimedes force acting on a body $K \subset \mathbb{R}^3$ submersed into a liquid. To this end, introduce the Cartesian coordinates for which the OXY-plane coincides with the liquid surface and the axis OZ is directed downward. At each point $(x, y, z) \in \partial K$, the body K is subjected to the pressure force $F(x, y, z) = -g\rho z v(x, y, z)$, where $v(x, y, z)$ is the unit outer normal to ∂K, ρ is liquid density and g is acceleration of gravity. The resultant, i.e., the Archimedean force $\iint_{\partial K} (-g\rho z) v(x, y, z) d\sigma(x, y, z)$ has the coordinates

$$
\begin{aligned}
F_x &= -g\rho \iint_{\partial K} z \, \langle v(x, y, z), e_1 \rangle \, d\sigma(x, y, z), \\
F_y &= -g\rho \iint_{\partial K} z \, \langle v(x, y, z), e_2 \rangle \, d\sigma(x, y, z), \\
F_z &= -g\rho \iint_{\partial K} z \, \langle v(x, y, z), e_3 \rangle \, d\sigma(x, y, z).
\end{aligned}
$$

Rewriting the first of these equalities as $F_x = \iint_{\partial K} \langle V, v \rangle d\sigma$, where $V(x, y, z) = (-g\rho z, 0, 0)$, and, using the Gauss-Ostrogradski formula, we obtain

$$
F_x = \iiint_K \operatorname{div} V(x, y, z) dx dy dz = \iiint_K 0 dx dy dz = 0.
$$

Similarly, it can be shown that $F_y = 0$. The vertical component of the Archimedean force can be expressed in terms of the divergence of the vector field $\widetilde{V}(x, y, z) = (0, 0, -g\rho z)$, so it equals

$$
\begin{aligned}
F_z &= \iint_{\partial K} \langle \widetilde{V}(x, y, z), v(x, y, z) \rangle d\sigma(x, y, z) = \iiint_K \operatorname{div} \widetilde{V}(x, y, z) dx dy dz \\
&= \iiint_K (-g\rho) dx dy dz = -g\rho \lambda_3(K).
\end{aligned}
$$

Thus, a buoyancy force numerically equal to the weight of the fluid forced out by the body acts on this body in a vertical direction.

2.6.7

Green's Formula. Let us single out the two-dimensional case of the Gauss-Ostrogradski formula. Let K be a standard compact set in $\mathbb{R}^2$ and let v be its outer side. On the regular part L of the boundary ∂K, the side v agrees with the direction $\tau = U(\nu)$ (see Sect. 2.5.4). The pair $(\partial K, \tau)$ will be called an oriented boundary of the planar standard compact set K and denoted by the symbol $\partial^+ K$.

For a vector field $V = (V_1, V_2)$ that is smooth in some neighborhood of the compactum K, the vector form (11) of the Gauss-Ostrogradski formula yields

$$\iint_K \operatorname{div} V(x,y)dxdy = \int_{\partial K} \langle V(x,y), v(x,y)\rangle d\sigma_1(x,y).$$

By Eq. (5) from Sect. 2.5.4, this equality can be rewritten as

$$\iint_K \left(\frac{\partial V_1}{\partial x}(x,y) + \frac{\partial V_2}{\partial y}(x,y)\right) dxdy = \int_{\partial^+ K} -V_2(x,y)dx + V_1(x,y)dy.$$

Putting $P = -V_2$ and $Q = V_1$, we arrive at an important result known as Green's formula:

$$\iint_K \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dxdy = \int_{\partial^+ K} P(x,y)dx + Q(x,y)dy$$

In particular, Green's formula allows one to express the area of a standard compact set as an integral over its boundary: taking the functions $P(x,y) \equiv 0$, $Q(x,y) \equiv x$ or $P(x,y) \equiv -y, Q(x,y) \equiv 0$, we obtain

$$\lambda_2(K) = \int_{\partial^+ K} xdy = -\int_{\partial^+ K} ydx = \frac{1}{2}\int_{\partial^+ K} -ydx + xdy$$

In Sect. 2.5, it was noted (for an example, see Sect. 2.5.2) that the integral of a locally potential field over a closed oriented curve can be non-zero. On the other hand, it is obvious that Green's formula implies the following.

Corollary 1 *Let $V = (P, Q)$ be a locally potential vector field that is smooth in some domain $\mathcal{O} \subset \mathbb{R}^2$. Let $K \subset \mathcal{O}$ be a standard compact set with oriented boundary. Then*

$$\int_{\partial^+ K} P(x,y)dx + Q(x,y)dy = 0.$$

This corollary gives a simple geometric condition for the integral of a locally potential vector field over a closed curve to vanish. This is the case if the curve "bounds a set in $\mathcal{O}$ ", i.e., coincides with the boundary of some standard compact set contained in $\mathcal{O}$. Otherwise (for example, if the curve "surrounds" a point that does not belong to the domain) it is easy to find a smooth locally potential vector field in $\mathcal{O}$ that has a non-zero integral over this curve (for an example, see Sect. 2.5.2).

Let us point out one important special case of Corollary 1 related to holomorphic functions. Let $L \subset \mathbb{C}$ be a piecewise smooth oriented curve that lies in the domain of a continuous complex-valued function f. Let $g = \mathcal{R}ef$ and $h = \mathcal{I}mf$. Guided by the formal multiplication identity $f(z)dz = (g + ih)(dx + idy) =$

$(gdx - hdy) + i(hdx + gdy)$, we define the integral $\int_L f(z)dz$ as the sum $\int_L gdx - hdy + i\int_L hdx + gdy$. It is easy to see that $\int_L f(z)dz = \int_a^b f(\gamma(t))\gamma'(t)dt$ for every smooth parametrization γ that agrees with the orientation of the curve L.

Corollary 2 (The Cauchy theorem) *If a function f has a continuous derivative $\frac{df}{dz}$ in a domain $\mathcal{O} \subset \mathbb{C}$ and $K \subset \mathcal{O}$ is a standard compact set, then $\int_{\partial^+ K} f(z)dz = 0$.*

Proof By our assumptions, the functions $g = \mathcal{R}ef$ and $h = \mathcal{I}mf$ belong to the class $C^1(\mathcal{O})$. Moreover, the Cauchy-Riemann conditions $g'_x = h'_y, h'_x = -g'_y$ hold, which ensures that the vector fields $(g, -h)$ and (h, g) are locally potential (see Sect. 8.5.2, the corollary to Proposition 3). Thus, the equalities

$$\int_{\partial^+ K} gdx - hdy = 0 \quad \text{and} \quad \int_{\partial^+ K} hdx + gdy = 0$$

follow from Corollary 1.