THE RIEMANN INTEGRAL

TSOGTGEREL GANTUMUR

Abstract. The Riemann integral in $\mathbb{R}^n$.

Contents

1. Problem of antidifferentiation

Consider the problem of finding F satisfying

$$F' = f, \tag{1}$$

where f is a given function. Given f, finding F is called *antidifferentiation*, and F is called an *antiderivative of* f, cf. Appendix A. We know that the antiderivative of a given function can only be found *up to an additive constant*, and that if we know one antiderivative of a given function, all other antiderivatives are found by adding an arbitrary constant to it.

We also have some useful rules for antidifferentiation, that allow us, for instance, to construct an antiderivative of $f + g$, given that antiderivatives of f and of g are available. These rules guarantee that certain elementary functions admit antiderivatives, but the resulting set of functions is not big enough; They cannot even guarantee that all smooth functions admit antiderivatives. In modern mathematics how we deal with this problem is to construct an antiderivative F as the limit of some infinite process, known as *integration*. Specifically, we will study the *Riemann integral*, which is one of the simplest yet powerful approach to integration.

We proceed by identifying some crucial properties of antidifferentiation, which will then guide us in constructing the Riemann integral. Given $f : \mathbb{R} \to \mathbb{R}$, whose antiderivative $F : \mathbb{R} \to \mathbb{R}$ is *assumed* to exist, we introduce the notation

$$I_{a,x}(f) = F(x) - F(a), \tag{2}$$

where $a, x \in \mathbb{R}$. The reason for subtraction is of course to remove the ambiguity given by the fact that antiderivatives are unique only up to an additive constant. In other words, $I_{a,x}(f)$ is the value $F(x)$ of the antiderivative of f, under the normalization that $F(a) = 0$. Immediate corollaries of the definition are

$$I_{a,a}(f) = 0, \tag{3}$$

Date: 2016/12/12 13:44 (EST). Incomplete draft.

and the following *domain additivity* property

$$I_{a,x}(f) = I_{a,b}(f) + I_{b,x}(f), \qquad a, b, x \in \mathbb{R}. \tag{4}$$

Then we have *linearity*, cf. Lemma A.7, expressed by

$$I_{a,x}(\alpha f) = \alpha I_{a,x}(f), \tag{5}$$

for $\alpha \in \mathbb{R}$, and

$$I_{a,x}(f + g) = I_{a,x}(f) + I_{a,x}(g). \tag{6}$$

Another important property is *monotonicity*: If $a \leq x$ and if $f \leq g$ in $[a, x]$, then

$$I_{a,x}(f) \leq I_{a,x}(g). \tag{7}$$

Finally, since $(x)' = 1$, we have

$$I_{a,x}(1) = x - a. \tag{8}$$

Remark 1.1. The aforementioned simple properties will be enough to pin down a good notion of integral. Note that these properties have been derived under the assumption that all involved functions admit antiderivatives, and what we would like now is to give a general construction that satisfies those properties, and hope that it will enable us to show that for instance, smooth functions admit antiderivatives. The basic idea is as follows. Given a general function f, we want to find a sequence of functions $g_1, g_2, \ldots$, such that each g_i is simple enough so that $I_{a,b}(g_i)$ is well-defined, and that g_i approximates f better and better as $i \to \infty$. Then we would define $I_{a,b}(f)$ as the limit of $I_{a,b}(g_i)$ as $i \to \infty$. In order for this plan to work, we need to address the following issues.

- Choose the pool of "simple" functions from which the sequence $\{g_i\}$ to be picked.
- Clarify the notion "g_i approximates f better and better as $i \to \infty$."
- Check if $I_{a,b}(g_i)$ converges to some number $\xi \in \mathbb{R}$ as $i \to \infty$.
- Verify that the limit ξ does not depend on the particular sequence $\{g_i\}$ we used.

For the first item, we use piecewise constant functions, also known as *step functions*. These functions have the form

$$g(x) = \sum_{j=1}^{k} A_j \chi_{J_j}(x), \tag{9}$$

where $A_j \in \mathbb{R}$, and χ_{J_j} is the characteristic function of some interval J_j, such as $J_j = (a_j, b_j)$, $J_j = [a_j, b_j)$, $J_j = (a_j, b_j]$, or $J_j = [a_j, b_j]$. For later reference, we call $J_1, \ldots, J_k$ the *support intervals* of g. Recall that in general, the *characteristic function* of a set $B \subset \mathbb{R}^n$ is defined by

$$\chi_B(x) = \begin{cases} 1 & \text{for } x \in B, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

For a nonempty *closed* interval $J = [\alpha, \beta] \subset [a, b]$, we can compute

$$I_{a,b}(\chi_J) = I_{a,\alpha}(0 \cdot 1) + I_{\alpha,\beta}(1) + I_{\beta,b}(0 \cdot 1) = 0 + (\beta - \alpha) + 0 = \beta - \alpha, \tag{11}$$

where we have used domain additivity, linearity, and (8). On the other hand, for a nonempty *open* interval $J = (\alpha, \beta) \subset [a, b]$, we have

$$I_{a,b}(\chi_J) = I_{a,b}(\chi_{[\alpha,\beta]}) - I_{a,b}(\chi_{[\alpha,\alpha]}) - I_{a,b}(\chi_{[\beta,\beta]}) = (\beta - \alpha) - 0 - 0 = \beta - \alpha. \tag{12}$$

Similarly, we infer $I_{a,b}(\chi_{[\alpha,\beta)}) = I_{a,b}(\chi_{(\alpha,\beta]}) = \beta - \alpha$, as long as $a \leq \alpha \leq \beta \leq b$. To conclude, for any $J \subset [a, b]$ of the form $J = [\alpha, \beta]$, $J = [\alpha, \beta)$, $J = (\alpha, \beta]$, or $J = (\alpha, \beta)$, we have

$$I_{a,b}(\chi_J) = |J|, \tag{13}$$

where $|J| = \beta - \alpha$ is by definition the *length* of J. Once this is known, we can compute $I_{a,b}(g)$ for step functions (9) by linearity, as

$$I_{a,b}(g) = \sum_{j=1}^{k} A_j I_{a,b}(\chi_{J_j}) = \sum_{j=1}^{k} A_j |J_j|, \tag{14}$$

where we have assumed that $J_j \subset [a,b]$ for all j.

Remark 1.2. We have restricted ourselves to the case $a \leq b$, even though including the case $a > b$ would not have posed any principal difficulties. However, in order not to clutter the arguments, we will postpone dealing with the case $a > b$ to Section 3.

The next item in the list of Remark 1.1 is to clarify when it would be useful to say that a sequence of step functions g_i approximates f well as $i \to \infty$. Since we want to eventually define $I_{a,b}(f) = \lim\limits_{i\to\infty} I_{a,b}(g_i)$, it would be ideal if we introduce closeness of g_i to f by

$$g_i \text{ is close to } f \qquad \Longleftrightarrow \qquad g_i \leq f \quad \text{and} \quad I_{a,b}(f) - I_{a,b}(g_i) \text{ is small.} \tag{15}$$

However, this would not work, for the simple reason that we have not defined $I_{a,b}(f)$ yet. A way out of this situation is to introduce a second sequence of step functions $\{h_i\}$, bounding f from above, and require that $I_{a,b}(h_i) - I_{a,b}(g_i)$ is small, as follows.

$$g_i \text{ and } h_i \text{ are close to } f \qquad \Longleftrightarrow \qquad g_i \leq f \leq h_i \quad \text{and} \quad I_{a,b}(h_i) - I_{a,b}(g_i) \text{ is small.} \tag{16}$$

Intuitively, if $\{g_i\}$ and $\{h_i\}$ satisfy the preceding condition, then both $I_{a,b}(g_i)$ and $I_{a,b}(h_i)$ will be close to the yet-to-be-defined quantity $I_{a,b}(f)$.

Example 1.3. Let us illustrate what we have discussed by a concrete example where we know $I_{a,b}(f)$ explicitly. We let $f(x) = x$ and $[a,b] = [0,b]$. For $m \in \mathbb{N}$, subdivide $[0,b]$ into m equal subintervals, by using the *node points* $x_k = \delta k$, $k = 0, \ldots, m$, where $\delta = \frac{b}{m}$. Let

$$g_m(x) = \delta k \qquad \text{and} \qquad h_m(x) = \delta k + \delta, \qquad \text{for} \quad x \in [\delta k, \delta k + \delta). \tag{17}$$

The point $x = b$ is missing in this definition, but it will be inconsequential. For concreteness, we set $g_m(b) = 0$ and $h_m(b) = 1$. In other words, we have

$$g_m = \sum_{k=0}^{m-1} \delta k \chi_{[\delta k, \delta k+\delta)}, \qquad h_m = \sum_{k=0}^{m-1} (\delta k + \delta) \chi_{[\delta k, \delta k+\delta)} + \chi_{[b,b]}. \tag{18}$$

By construction, we have $g_m(x) \leq f(x) \leq h_m(x)$ for all $x \in [0,b]$, and

$$\begin{aligned} I_{0,b}(h_m) - I_{0,b}(g_m) = I_{0,b}(h_m - g_m) &= \sum_{k=0}^{m-1} \delta I_{0,b}\big(\chi_{[\delta k, \delta k+\delta)}\big) + I_{0,b}\big(\chi_{[b,b]}\big) \\ &= m\delta \cdot \delta = \delta, \end{aligned} \tag{19}$$

which can be made arbitrarily small by choosing m large. This we have bounded f from above and below by two step functions, whose integrals are arbitrarily close to each other. Finally, we compute

$$I_{0,b}(g_m) = \sum_{k=0}^{m-1} \delta k I_{0,b}\big(\chi_{[\delta k, \delta k+\delta)}\big) = \delta^2 \sum_{k=0}^{m-1} k = \frac{b^2}{m^2} \cdot \frac{m(m-1)}{2} \to \frac{b^2}{2} \quad \text{as } m \to \infty. \tag{20}$$

This is consistent with the fact that $(x^2/2)' = x$ and so $I_{0,b}(f) = b^2/2$.

Remark 1.4. Based on the foregoing discussion, we make some observations.

(a) The support intervals in the definition of step functions (9) may be assumed to be non-overlapping, since a step function with two overlapping support intervals can be written as a step function with 3 non-overlapping support intervals.

(b) Similarly, if g and h are step functions, then by subdividing the support intervals if necessary, we can always assume that g and h have the same support intervals.
(c) Furthermore, all support intervals in (9) may be assumed to be open, because any interval J with $|J| > 0$ is the union of an open interval and one or two zero-length intervals, and zero-length support intervals have no contribution to the integral (14).
(d) One cannot cover $[a, b]$ by disjoint open intervals. So once we assume that all support intervals in (9) are open and disjoint, there will necessarily be at least a finitely many points $x_0, \ldots, x_k$ that are not covered by the support intervals, and hence any step function g with the same set of support intervals will satisfy $g(x_j) = 0$ for $j = 0, \ldots, k$. This means that the condition $g_i \leq f \leq h_i$ in $[a, b]$ from (16) cannot be satisfied in general. However, we can replace it by the condition that $g_i \leq f \leq h_i$ *in each support interval*, because we could always add terms such as $f(x_j)\chi_{[x_j,x_j]}$ to g_i and h_i to ensure that $g_i \leq f \leq h_i$ in $[a, b]$, but these additional terms would not affect the value of the integral (14) anyway.

We now introduce some new terminology. A *grid* on $[a, b]$ is a collection

$$P = \{(x_0, x_1), (x_1, x_2), \ldots, (x_{k-1}, x_k)\}, \tag{21}$$

of disjoint open intervals, such that $[x_0, x_1] \cup [x_1, x_2] \cup \ldots \cup [x_{k-1}, x_k] = [a, b]$. The points $x_0, \ldots, x_k$ are called the *nodes* of the grid. We call $g : [a, b] \to \mathbb{R}$ a *step function subordinate to* P if g can be written in the form

$$g = \sum_{J \in P} A_J \chi_J, \tag{22}$$

where P is a grid on $[a, b]$, and $A_J \in \mathbb{R}$ for $J \in P$. For step functions, we define

$$I_{a,b}(g) = \sum_{J \in P} A_J |J|. \tag{23}$$

A *refinement* of a grid is a grid that is obtained by adding nodes to the original grid. For example, $\tilde{P} = \{(0, 1), (1, 2), (2, 4)\}$ is a refinement of $P = \{(0, 2), (2, 4)\}$. If g is a step function subordinate to a grid P, and if $\tilde{P}$ is a refinement of P, then there is a step function $\tilde{g}$ subordinate to $\tilde{P}$, such that $\tilde{g} = g$ in each $J \in \tilde{P}$. We get $\tilde{g}$ simply by replacing the characteristic function $\chi_{(x_i,x_{i+1})}$ in (22) by the sum $\chi_{(x_i,x')} + \chi_{(x',x_{i+1})}$ every time a new node $x' \in (x_i, x_{i+1})$ is introduced. In particular, we have $I_{a,b}(\tilde{g}) = I_{a,b}(g)$. We shall consider the two step functions g and $\tilde{g}$ as being the same.

Remark 1.5. Let $f : [a, b] \to \mathbb{R}$, and suppose that we have two sequences $\{g_i\}$ and $\{h_i\}$, satisfying the following conditions.

(i) g_i and h_i are step functions subordinate to some grid P_i, for each i.
(ii) $g_i \leq f \leq h_i$ in each $J \in P_i$, for each i.
(iii) $I_{a,b}(h_i - g_i) \to 0$ as $i \to \infty$.

Let $\alpha_i = I_{a,b}(g_i)$ and $\beta_k = I_{a,b}(h_k)$. Then from (ii), since $g_i \leq f \leq h_k$ (in each support interval) for any i and k, we have $\alpha_i \leq \beta_k$ for all i and k. Let $\varepsilon > 0$ be arbitrary, and let N be such that $i \geq N$ implies $\beta_i - \alpha_i \leq \varepsilon$. Then for any $i \geq N$ and $k \geq N$, we have

$$\alpha_i - \alpha_k \leq \beta_k - \alpha_k \leq \varepsilon, \tag{24}$$

implying that $\{\alpha_i\}$ is a Cauchy sequence. Hence there exists $\xi \in \mathbb{R}$ such that $\alpha_i \to \xi$ as $i \to \infty$. Since $\beta_i - \alpha_i \to 0$ as $i \to \infty$, we also get $\beta_i \to \xi$ as $i \to \infty$, and so ξ would be a very good candidate for the value of $I_{a,b}(f)$ that is yet to be defined.

Exercise 1.6. Continuing the preceding remark, let $\{\tilde{g}_i\}$ and $\{\tilde{h}_i\}$ be two sequences satisfying the conditions (i)-(iii). Show that the limit $\tilde{\xi}$ of $I_{a,b}(\tilde{g}_i)$ as $i \to \infty$ is equal to ξ.

Remark 1.5 and Exercise 1.6 show that as long as $\{g_i\}$ and $\{h_i\}$ are two sequences satisfying the conditions (i)-(iii) of Remark 1.5, the limit of $I_{a,b}(g_i)$ as $i \to \infty$ exists, and does not depend on the choice of the sequences $\{g_i\}$ and $\{h_i\}$. Thus the following definition makes sense.

$$I_{a,b}(f) := \lim_{i\to\infty} I_{a,b}(g_i). \tag{25}$$

We could have also used the upper sequence $\{h_i\}$ in the defintion, since we have

$$\lim_{i\to\infty} I_{a,b}(g_i) = \lim_{i\to\infty} I_{a,b}(h_i). \tag{26}$$

This completes the program outlined in Remark 1.1. What remains now is to verify if the procedure (25) would be able to solve the antidifferentiation problem. Before addressing that question, however, we would like to extend the foregoing construction to higher dimensions.

2. The Riemann integral

A *grid* on a rectangle $Q = [a_1, b_1] \times \ldots \times [a_n, b_n] \subset \mathbb{R}^n$ is a collection

$$G = \{I_1 \times \ldots \times I_n : I_j \in P_j,\ j = 1, \ldots, n\}, \tag{27}$$

where P_j is a grid on the interval $[a_j, b_j]$, for $j = 1, \ldots, n$. We call $g : Q \to \mathbb{R}$ a *step function subordinate to* G if g can be written in the form

$$g = \sum_{q\in G} A_q \chi_q, \tag{28}$$

where G is a grid on Q, and $A_q \in \mathbb{R}$ for $q \in G$. The (n-dimensional) *volume* of a rectangle $q = I_1 \times \ldots \times I_n$ is $|q| = |I_1| \cdots |I_n|$.

Definition 2.1. If g is a step function subordinate to a grid G on Q, then its *integral over* Q is defined as

$$\int_Q g = \sum_{q\in G} A_q |q|. \tag{29}$$

Note that in 1 dimension, we have

$$\int_{[a,b]} g = I_{a,b}(g), \tag{30}$$

where $I_{a,b}(g)$ is the notation from the preceding section.

Remark 2.2. (a) If f and g are step functions subordinate to G, and $\lambda, \mu \in \mathbb{R}$, then we have

$$\int_Q (\lambda f + \mu g) = \lambda \int_Q f + \mu \int_Q g. \tag{31}$$

(b) If f and g are step functions subordinate to G, satisfying $f \leq g$, then we have

$$\int_Q f \leq \int_Q g. \tag{32}$$

To extend the definition of integral to more general functions, we will use approximation of general functions by step functions.

Definition 2.3. A function $f : Q \to \mathbb{R}$ is called *(Riemann) integrable* if for any $\varepsilon > 0$, there exist a grid G_ε on Q, and simple functions g_ε and h_ε subordinate to G_ε, such that

$$g_\varepsilon \leq f \leq h_\varepsilon \qquad \text{in each} \quad q \in G_\varepsilon, \tag{33}$$

and

$$\int_Q (h_\varepsilon - g_\varepsilon) \leq \varepsilon. \tag{34}$$

If f is integrable, we call

$$\int_Q f = \lim_{\varepsilon\to 0} \int_Q g_\varepsilon, \tag{35}$$

the *integral of f over Q*.

Exercise 2.4. Show that the limit in (35) exists, and does not depend on the choice of $\{g_\varepsilon\}$ and $\{h_\varepsilon\}$. Show also that

$$\lim_{\varepsilon\to 0} \int_Q g_\varepsilon = \lim_{\varepsilon\to 0} \int_Q h_\varepsilon. \tag{36}$$

Hint: Remark 1.5 and Exercise 1.6.

Remark 2.5. One useful observation is that given a grid G on Q, the step functions

$$g = \sum_{q\in G} A_q \chi_q, \qquad \text{and} \qquad h = \sum_{q\in G} B_q \chi_q, \tag{37}$$

given by

$$A_q = \inf_q f \quad \text{and} \quad B_q = \sup_q f \qquad \text{for} \quad q \in G, \tag{38}$$

minimize the quantity $\int_Q (h-g)$, while satisfying $g \leq f \leq h$ in each $q \in G$. Thus we can always assume that (37) and (38) hold.

Theorem 2.6. *Any continuous function $f : Q \to \mathbb{R}$ is integrable.*

Proof. In Theorem 2.8 below, we will prove that f is uniformly continuous, meaning that for any $\varepsilon > 0$, there exists $\delta > 0$, such that $|f(x) - f(y)| \leq \varepsilon$ whenever $x, y \in Q$ satisfies $|x - y|_\infty < \delta$. Let $\varepsilon > 0$ be arbitrary, and pick $\delta > 0$ according to the preceding property. Then we choose an integer $m \in \mathbb{N}$ so large that $\max\{b_1 - a_1, \ldots, b_n - a_n\} < m\delta$, and construct a grid G by subdividing each side of $Q = [a_1, b_1] \times \ldots \times [a_n, b_n]$ into m equal subintervals. This will ensure that the length of each side of any cell $q \in G$ is less than δ, and hence

$$|f(x) - f(y)| \leq \varepsilon \qquad \text{whenever} \quad x, y \in q, \quad q \in G. \tag{39}$$

Finally, we define the step functions g and h by (37) and (38). By construction, we have

$$g \leq f \leq h \qquad \text{in each} \quad q \in G. \tag{40}$$

Moreover, from (39) we have $B_q - A_q \leq \varepsilon$ for all $q \in G$, and thus

$$\int_Q (h-g) = \sum_{q\in G} (B_q - A_q)|q| \leq \varepsilon \sum_{q\in G} |q| = |Q|\varepsilon. \tag{41}$$

As $\varepsilon > 0$ was arbitrary, we conclude that f is integrable. $\square$

To complete the preceding proof, we need to show that the function f in the statement of the theorem is uniformly continuous.

Definition 2.7. A function $u : \Omega \to \mathbb{R}$ with $\Omega \subset \mathbb{R}^n$ is called *uniformly continuous* if for any $\varepsilon > 0$, there exists $\delta > 0$, such that $|u(x) - u(y)| \leq \varepsilon$ whenever $x, y \in \Omega$ satisfies $|x - y|_\infty < \delta$.

Theorem 2.8 (Heine-Cantor)**.** *Let $K \subset \mathbb{R}^n$ be closed and bounded set, and let $f : K \to \mathbb{R}$ be a continuous function. Then f is uniformly continuous.*

Proof. Let $P = K \times K \subset \mathbb{R}^{2n}$, and let $F : P \to \mathbb{R}$ be defined by $F(x, y) = |f(x) - f(y)|$. Obviously, F is continuous. Suppose that f is *not* uniformly continuous. This means that for any $\delta > 0$ there exist $x, y \in K$ such that $|x - y|_\infty < \delta$ and $|f(x) - f(y)| > \varepsilon$, where $\varepsilon > 0$ is some constant. In other words, there exists a sequence $(x_i, y_i) \in P$, $i = 1, 2, \ldots$, such that $|x_i - y_i|_\infty \to 0$ as $i \to \infty$ and $F(x_i, y_i) > \varepsilon$ for all i. Since P is bounded, by considering a large cube Q containing P, and subdividing Q into smaller cubes, we can extract a converging subsequence of $\{(x_i, y_i)\}$. Thus without loss of generality, we can assume that $(x_i, y_i) \to (x, y)$ as $i \to \infty$, for some $(x, y) \in \mathbb{R}^{2n}$. As K is closed, we have $(x, y) \in P$. Moreover, $|x_i - y_i|_\infty \to 0$ as $i \to \infty$ implies that $x = y$. By continuity of F, we have $F(x_i, y_i) \to F(x, x) = 0$, which contradicts the assumption that $F(x_i, y_i) > \varepsilon$ for all i. Hence, f must be uniformly continuous. □

Exercise 2.9. Prove the following properties of the Riemann integral.

(a) If f and g are integrable, and $\lambda, \mu \in \mathbb{R}$, then $\lambda f + \mu g$ is integrable, and

$$\int_Q (\lambda f + \mu g) = \lambda \int_Q f + \mu \int_Q g. \tag{42}$$

(b) If f and g are integrable functions satisfying $f \leq g$, then we have

$$\int_Q f \leq \int_Q g. \tag{43}$$

(c) If f is integrable, then $|f|$ is integrable, and

$$\Big| \int_Q f \Big| \leq \int_Q |f|. \tag{44}$$

(d) If f is integrable, then f^2 is integrable. In addition, if g is integrable, then $fg = \frac{1}{2}(f + g)^2 - \frac{1}{2}f^2 - \frac{1}{2}g^2$ is integrable.

3. The fundamental theorem of calculus

By using the Riemann integral, in this section we are going to show that all continuous functions admit antiderivatives. That is, if $f : [a, b] \to \mathbb{R}$ is a continuous function, then $F : [a, b] \to \mathbb{R}$ defined by the integrals

$$F(x) = \int_{[a,x]} f, \qquad x \in [a, b], \tag{45}$$

satisfies $F' = f$ in (a, b). This gives a satisfactory answer to the antidifferentiation problem posed in Section 1, namely the problem of constructing F satisfying the differential equation $F' = f$, where f is a given function.

In the following, we will use the notation

$$\int_a^b f = \int_{[a,b]} f. \tag{46}$$

Lemma 3.1. *If $f : [a, b] \to \mathbb{R}$ is an integrable function, then for any $c \in [a, b]$, the restrictions $f|_{[a,c]}$ and $f|_{[c,b]}$ are integrable over $[a, c]$ and $[c, b]$, respectively, and we have*

$$\int_a^c f + \int_c^b f = \int_a^b f. \tag{47}$$

Proof. Let $\varepsilon > 0$, and let g and h be step functions subordinate to a grid G, satisfying $g \leq f \leq h$ in each $q \in G$, and $\int_a^b (h - g) \leq \varepsilon$. Without loss of generality, we can assume that c is a node in G, meaning that G can be decomposed as $G = G_1 \cup G_2$, where G_1 and G_2 are grids on $[a, c]$ and $[c, b]$, respectively. Then we have

$$\int_a^c (h - g) \leq \int_a^b (h - g) \leq \varepsilon, \tag{48}$$

which shows that the restriction $f|_{[a,c]}$ is integrable over $[a, c]$. Similarly, $f|_{[c,b]}$ is integrable over $[c, b]$. Hence we have

$$\int_a^c g + \int_c^b g \to \int_a^c f + \int_c^b f \qquad \text{as} \quad \varepsilon \to 0. \tag{49}$$

On the other hand, we have

$$\int_a^c g + \int_c^b g = \int_a^b g \to \int_a^b f \qquad \text{as} \quad \varepsilon \to 0, \tag{50}$$

which completes the proof. □

Theorem 3.2 (Fundamental theorem of calculus). *Let $f : [a, b] \to \mathbb{R}$ be an integrable function, and let*

$$F(x) = \int_a^x f, \qquad x \in [a, b]. \tag{51}$$

Then there exists a constant $M \geq 0$ such that

$$|F(x) - F(y)| \leq M|x - y|, \qquad x, y \in [a, b]. \tag{52}$$

In addition, if f is continuous at $x \in (a, b)$, then F is differentiable at x with

$$F'(x) = f(x). \tag{53}$$

Proof. Let g and h be step functions subordinate to some grid G of $[a, b]$, such that $g \leq f \leq h$ in each $q \in G$. Then for any $x, y \in [a, b]$ with $x < y$, we have

$$(y - x) \min_{[a,b]} g \leq \int_x^y g \leq \int_x^y f \leq \int_x^y h \leq (y - x) \max_{[a,b]} h, \tag{54}$$

implying that

$$\Big| \int_x^y f \Big| \leq M(y - x), \tag{55}$$

for some constant M independent of x and y. Since we have

$$F(y) - F(x) = \int_x^y f, \tag{56}$$

this proves the first part of the theorem.

For $x, y \in [a, b]$ with $x < y$, we have

$$(y - x) \inf_{[x,y]} f \leq \int_x^y f \leq (y - x) \sup_{[x,y]} f, \tag{57}$$

meaning that

$$\inf_{[x,y]} f \leq \frac{F(y) - F(x)}{y - x} \leq \sup_{[x,y]} f. \tag{58}$$

If f is continuous at x, then both $\inf_{[x,y]} f$ and $\sup_{[x,y]} f$ tend to $f(x)$ as $y \to x$, and if f is continuous at y, then both $\inf_{[x,y]} f$ and $\sup_{[x,y]} f$ tend to $f(y)$ as $x \to y$. Therefore, $F'(x)$ exists and is equal to $f(x)$ whenever f is continuous at x. □

Exercise 3.3. In the setting of the preceding theorem, show that if f is continuous at a, then F is differentiable at a, in the sense that

$$\frac{F(x) - F(a)}{x - a} \to f(a) \qquad \text{as} \quad (a, b) \ni x \to a. \tag{59}$$

Remark 3.4. Let $f \in \mathscr{C}([a,b])$, and let F be as in Theorem 3.2. Suppose that $G \in \mathscr{C}([a,b])$ is differentiable and $G' = f$ in (a,b). Since $F' = f$ in (a,b), this implies that $F(x) = G(x) + C$ for some constant C. Moreover, taking into account that $F(a) = 0$, we infer

$$F(x) = G(x) - G(a). \tag{60}$$

This gives us a way to compute the integral $F(x) = \int_a^x f$, provided that we have some antiderivative G of f available.

Example 3.5. For $f(x) = \cos x$, we have

$$\int_a^x f = \sin x - \sin a, \tag{61}$$

since $G(x) = \sin x$ is an antiderivative of f.

Remark 3.6. In Section 1, we started with the definition $I_{a,x}(f) = F(x) - F(a)$, where f and F are related by $F' = f$, and derived some simple properties of $I_{a,x}(f)$, such as linearity, monotonicity, and domain additivity. Then in Section 2, using those simple properties as a guiding light, we introduced the Riemann integral $\int_{[a,x]} f$ for $a \leq x$, and showed that it is well-defined at least when f is continuous. We have also shown that the Riemann integral $\int_{[a,x]} f$ satisfies the aforementioned simple properties of $I_{a,x}(f)$. Finally, in the current section (Theorem 3.2), we proved that when f is continuous, the function $F(x) = \int_{[a,x]} f$ satisfies $F' = f$. Thus by employing the Riemann integral, we have *constructed* antiderivatives for continuous functions. Now, as promised in Remark 1.2, we shall remove the restriction $a \leq x$. In fact, for $x \leq y$, we simply define

$$\int_y^x f = -\int_{[x,y]} f, \tag{62}$$

whenever f is integrable on $[x,y]$. It follows from Lemma 3.1 that if $f : [a,b] \to \mathbb{R}$ is an integrable function, then for any $x, y, z \in [a,b]$, we have

$$\int_x^y f + \int_y^z f = \int_x^z f. \tag{63}$$

Moreover, by Theorem 3.2, if $f : [a,b] \to \mathbb{R}$ is continuous and if $y \in (a,b)$, then

$$F(x) = \int_y^x f, \qquad x \in [a,b], \tag{64}$$

satisfies $F' = f$ in (a,b).

Exercise 3.7. Give detailed proofs of (63) and (64).

4. Fubini's theorem

In practice, higher dimensional integrals are often computed by reduction to repeated one dimensional integrals, which is justified by Fubini's theorem that we will prove below. Let $Q = [a_1, b_1] \times \ldots \times [a_n, b_n] \subset \mathbb{R}^n$, and let $f : Q \to \mathbb{R}$. Let $R = [a_1, b_1] \times \ldots \times [a_m, b_m]$ for some $m < n$, and for some fixed t, let $g : R \to \mathbb{R}$ be given by $g(x) = f(x,t)$. We can call g the restriction of f to a slice of Q defined by $t = \text{const}$. Then we write

$$\int_R f(x,t)\, \mathrm{d}^m x = \int_R g. \tag{65}$$

For $m = 1$, we simply write $\mathrm{d}x$ instead of $\mathrm{d}^1 x$.

Theorem 4.1 (Fubini)**.** *Let $Q' \subset \mathbb{R}^{n-1}$ be a rectangle, and let $Q = Q' \times [a,b]$. Suppose that $f : Q \to \mathbb{R}$ is integrable, and*

$$F(t) = \int_{Q'} f(x,t)\,\mathrm{d}^{n-1}x, \tag{66}$$

exists for each $t \in [a,b]$. Then $F : [a,b] \to \mathbb{R}$ is integrable, and

$$\int_Q f = \int_{[a,b]} F. \tag{67}$$

Proof. Let $\varepsilon > 0$, and let g and h be step functions subordinate to a grid Γ, satisfying $g \leq f \leq h$ in each $q \in \Gamma$, and $\int_Q (h-g) \leq \varepsilon$. For each $t \in [a,b]$, let

$$G(t) = \int_{Q'} g(x,t)\,\mathrm{d}^{n-1}x, \qquad \text{and} \qquad H(t) = \int_{Q'} h(x,t)\,\mathrm{d}^{n-1}x. \tag{68}$$

Let P be the grid on $[a,b]$, that is induced by the grid Γ on $Q = Q' \times [a,b]$. Then we have $G \leq F \leq H$ on each interval $I \in P$. Moreover, for any $q = q' \times I \in \Gamma$, we have

$$\int_a^b \int_{Q'} \chi_q(x,t)\,\mathrm{d}^{n-1}x\,\mathrm{d}t = \int_a^b \chi_I(t)|q'|\,\mathrm{d}t = |I||q'| = |q| = \int_Q \chi_q, \tag{69}$$

and hence

$$\int_a^b \int_{Q'} \sum_{q\in\Gamma} A_q\chi_q(x,t)\,\mathrm{d}^{n-1}x\,\mathrm{d}t = \sum_{q\in\Gamma} A_q \int_a^b \int_{Q'} \chi_q(x,t)\,\mathrm{d}^{n-1}x\,\mathrm{d}t = \int_Q \sum_{q\in\Gamma} A_q\chi_q. \tag{70}$$

This means in particular that

$$\int_Q g = \int_{[a,b]} G, \qquad \text{and} \qquad \int_Q h = \int_{[a,b]} H, \tag{71}$$

and

$$\int_{[a,b]} (H-G) = \int_Q (h-g) \leq \varepsilon. \tag{72}$$

Since $\varepsilon > 0$ is arbitrary, we infer that F is integrable. Furthermore, invoking

$$\int_{[a,b]} G \leq \int_{[a,b]} F \leq \int_{[a,b]} H \qquad \text{and} \qquad \int_Q g \leq \int_Q f \leq \int_Q h, \tag{73}$$

we get

$$-\,\varepsilon \leq \int_Q g - \int_{[a,b]} H \leq \int_Q f - \int_{[a,b]} F \leq \int_Q h - \int_{[a,b]} G \leq \varepsilon, \tag{74}$$

and using again the fact that $\varepsilon > 0$ is arbitrary, (67) is obtained. □

Example 4.2. In light of Theorem 2.6, provided that $f : Q \to \mathbb{R}$ is continuous, all conditions of the preceding theorem are satisfied. For example, we can compute a multidimensional integral as follows.

$$\begin{aligned} \int_{[0,1]^3} x_1x_2e^{x_3}\mathrm{d}^3x &= \int_0^1 \int_{[0,1]^2} x_1x_2e^{x_3}\mathrm{d}^2(x_1,x_2)\,\mathrm{d}x_3 = \int_0^1\int_0^1\int_0^1 x_1x_2e^{x_3}\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 \\ &= \int_0^1\int_0^1 x_2e^{x_3}\Big(\frac{x_1^2}{2}\Big|_0^1\Big)\,\mathrm{d}x_2\,\mathrm{d}x_3 = \frac{1}{2}\int_0^1\int_0^1 x_2e^{x_3}\mathrm{d}x_2\,\mathrm{d}x_3 \\ &= \frac{1}{2}\int_0^1 e^{x_3}\Big(\frac{x_2^2}{2}\Big|_0^1\Big)\,\mathrm{d}x_3 = \frac{1}{4}\int_0^1 e^{x_3}\mathrm{d}x_3 = \frac{e}{4}. \end{aligned} \tag{75}$$

5. Negligible sets

Definition 5.1. A set $B \subset Q$ is called *negligible*, or of *volume zero*, if for any $\varepsilon > 0$, there exist a grid Γ on Q and a finite collection $q_1, \ldots, q_k \in \Gamma$, such that

$$B \subset \bar{q}_1 \cup \ldots \cup \bar{q}_k, \qquad \text{and} \qquad |q_1| + \ldots + |q_k| \leq \varepsilon. \tag{76}$$

Lemma 5.2. *Let $Q' \subset \mathbb{R}^{n-1}$ be a closed rectangle, and let $h : Q' \to \mathbb{R}$ be a continuous function. Then the graph $B = \{(x, h(x)) : x \in Q'\}$ is negligible as a subset of $Q = Q' \times [a, b]$, for any $a \leq \min_{Q'} h$ and $b \geq \max_{Q'} h$.*

Proof. Let $\varepsilon > 0$, and let $\delta > 0$ be such that $|h(x) - h(y)| < \varepsilon$ whenever $x, y \in Q'$ satisfy $|x - y|_\infty \leq \delta$. This is possible since h is uniformly continuous in Q'. Then we construct a grid G in Q' by subdividing each side of Q' into m equal subintervals, with m so large that the length of each side of any cell $r \in G$ is less than δ. Thus we have

$$|h(x) - h(y)| \leq \varepsilon \qquad \text{whenever} \quad x, y \in r, \quad r \in G. \tag{77}$$

For $r \in G$, we let

$$\alpha_r = \min_{x \in \bar{r}} h(x), \qquad \beta_r = \max_{x \in \bar{r}} h(x), \tag{78}$$

and pick a grid P on the interval $[a, b]$ such that α_r and β_r $(r \in G)$ are among the nodes of P. Finally, we construct the grid Γ in $Q = Q' \times [a, b]$, by combining the grid G along Q' and the grid P along the interval $[a, b]$. Then by construction, for each $r \in G$, there exist finitely many $p_1, \ldots, p_i \in \Gamma$ such that $\bar{p}_1 \cup \ldots \cup \bar{p}_i = \bar{r} \times [\alpha_r, \beta_r]$, whose combined volume is bounded by $|p_1| + \ldots + |p_i| = |r|(\beta_r - \alpha_r) \leq |r|\varepsilon$. This means that the set $\bigcup\{\bar{r} \times [\alpha_r, \beta_r] : r \in G\}$ can be covered by finitely many $q_1, \ldots, q_k \in \Gamma$ with the combined volume

$$|q_1| + \ldots + |q_i| \leq \sum_{r \in G} |r|\varepsilon = |Q'|\varepsilon. \tag{79}$$

Since $B \subset \bigcup\{\bar{r} \times [\alpha_r, \beta_r] : r \in G\}$, the proof is established. □

Remark 5.3. It is obvious that any subset of a negligible set is negligible. Moreover, the union of two negligible sets is negligible.

Exercise 5.4. Show that the surface of a sphere, and the surface of a cube are negligible.

Theorem 5.5. *Let $B \subset Q$ be a negligible set, and let $f : Q \to \mathbb{R}$ be a bounded function, which is continuous in $Q \setminus B$. Then f is integrable.*

Proof. Let $\varepsilon > 0$. Then by definition, there exist a grid Γ on Q and a finite collection $q_1, \ldots, q_k \in \Gamma$, such that

$$B \subset \bar{q}_1 \cup \ldots \cup \bar{q}_k, \qquad \text{and} \qquad |q_1| + \ldots + |q_k| \leq \varepsilon. \tag{80}$$

From the complementary perspective, there exists a finite collection $p_1, \ldots, p_j \in \Gamma$, such that

$$p_1 \cup \ldots \cup p_j \subset Q \setminus B, \qquad \text{and} \qquad |p_1| + \ldots + |p_j| \geq |Q| - \varepsilon. \tag{81}$$

In other words, we have $\Gamma = \{q_1, \ldots, q_k, p_1, \ldots, p_j\}$. Now, by adding more grid nodes in each direction, and shrinking each p_i slightly, we can assume that

$$K = \bar{p}_1 \cup \ldots \cup \bar{p}_j \subset Q \setminus B, \qquad \text{and} \qquad |K| \geq |Q| - 2\varepsilon. \tag{82}$$

Note that the latter is equivalent to $|Q \setminus K| \leq 2\varepsilon$. Since K is closed and bounded, f is uniformly continuous in K by the Heine-Cantor theorem (Theorem 2.8). Let $\delta > 0$ be such that $|f(x) - f(y)| < \varepsilon$ whenever $x, y \in K$ satisfy $|x - y|_\infty \leq \delta$. Then we refine the grid Γ so that the length of each side of any cell $q \in \Gamma$ is less than δ. Thus with $\Gamma_K = \{q \in \Gamma : q \subset K\}$, we have

$$|f(x) - f(y)| \leq \varepsilon \qquad \text{whenever} \quad x, y \in q, \quad q \in \Gamma_K. \tag{83}$$

Finally, we construct the step functions

$$g = \sum_{q\in\Gamma} A_q \chi_q, \qquad \text{and} \qquad h = \sum_{q\in\Gamma} B_q \chi_q, \tag{84}$$

with the coefficients given by

$$A_q = \inf_q f \quad \text{and} \quad B_q = \sup_q f \qquad \text{for} \quad q \in \Gamma_K, \tag{85}$$

and

$$A_q = -M \quad \text{and} \quad B_q = M \qquad \text{for} \quad q \in \Gamma \setminus \Gamma_K. \tag{86}$$

where $M \in \mathbb{R}$ is such that $|f(x)| \leq M$ for all $x \in Q$. By construction, we have $g \leq f \leq h$ in each $q \in \Gamma$. Moreover, we have

$$\int_Q (h-g) = \sum_{q\in\Gamma} (B_q - A_q)|q| \leq \sum_{q\in\Gamma_K} \varepsilon|q| + \sum_{q\in\Gamma\setminus\Gamma_K} 2M|q| \leq \varepsilon|Q| + 2M\cdot 2\varepsilon, \tag{87}$$

where in the last step we have invoked $|Q \setminus K| \leq 2\varepsilon$. □

Let us prove one more useful criterion on negligibility.

Lemma 5.6. *Let $Q \subset \mathbb{R}^m$ be a rectangle, and let $\phi : Q \to \mathbb{R}^n$ be a Lipschitz continuous function, in the sense that*

$$|\phi(x) - \phi(y)|_\infty \leq \lambda |x-y|_\infty, \qquad \textit{for all} \quad x, y \in Q, \tag{88}$$

with some constant $\lambda \geq 0$. If $m < n$, then $\phi(Q)$ is negligible. If $m = n$ and $B \subset Q$ is negligible, then $\phi(B)$ is negligible.

Proof. We prove the case $m = n$ first. Since $\phi(B)$ is negligible, by definition, for any $\varepsilon > 0$, there exists a finite collection $q_1, \ldots, q_k$ of rectangles, such that

$$B \subset \bar{q}_1 \cup \ldots \cup \bar{q}_k, \qquad \text{and} \qquad |q_1| + \ldots + |q_k| \leq \varepsilon. \tag{89}$$

By introducing a very fine grid whose elements are cubes, and replacing each q_i by a collection of cubes covering it, we can assume that $q_1, \ldots, q_k$ are cubes, with $|q_1| + \ldots + |q_k| \leq 2\varepsilon$. Let $2r_i$ be the side length of q_i, and let $x_i \in \mathbb{R}^n$ be the centre of of q_i. Then for any $x \in \bar{q}_i$, the Lipschitz condition gives

$$|\phi(x) - \phi(x_i)|_\infty \leq \lambda |x - x_i|_\infty \leq \lambda r_i, \tag{90}$$

meaning that $\phi(\bar{q}_i)$ is contained in $\bar{p}_i$, where $p_i = \{y \in \mathbb{R}^n : |y - \phi(x_i)|_\infty < \lambda r_i\}$. We have

$$\phi(B) \subset \bar{p}_1 \cup \ldots \cup \bar{p}_k, \tag{91}$$

and

$$|p_1| + \ldots + |p_k| = 2^n \lambda^n (r_1^n + \ldots + r_k^n) = \lambda^n (|q_1| + \ldots + |q_k|) \leq 2\lambda^n \varepsilon. \tag{92}$$

As $\varepsilon > 0$ is arbitrary, this shows that $\phi(B)$ is negligible.

We now turn to the case $m < n$. Let $P = Q \times [0,1]^{n-m} \subset \mathbb{R}^n$, and let $\psi : P \to \mathbb{R}^n$ be defined by $\psi(x_1, \ldots, x_n) = \phi(x_1, \ldots, x_m)$ for $x \in P$. Then ψ is Lipschitz continuous, and we have $\psi(B) = \phi(Q)$ with $B = Q \times \{0\} \subset Q \times [0,1]^{n-m}$. Since $B \subset P$ is negligible, we conclude by the previous paragraph that $\phi(Q)$ is negligible. □

Exercise 5.7. Let $U \subset \mathbb{R}^m$ be open, and let $\phi : U \to \mathbb{R}^n$ be a continuously differentiable map. Let $K \subset U$ be a closed and bounded set. Then show that the restriction $\phi : K \to \mathbb{R}^n$ is Lipschitz continuous. Furthermore, show that for $k < n$, any k-dimensional bounded manifold in $\mathbb{R}^n$ is negligible.

6. Jordan sets

Definition 6.1. A set $K \subset Q$ is called a *Jordan (measurable) set* if the characteristic function χ_K is integrable, and

$$|K| = \int_Q \chi_K, \tag{93}$$

is called the *Jordan content* or the *volume of* K. For a Jordan set $K \subset Q$, we say that $f : K \to \mathbb{R}$ is *integrable over* K, if $f\chi_K$ is integrable, and

$$\int_K f = \int_Q f\chi_K, \tag{94}$$

is called the *integral of* f *over* K.

Example 6.2. Let us find the area of the disk $D = \{(x, y) : x^2 + y^2 < 1\}$. For any fixed $x \in (-1, 1)$, the restriction $g(y) = \chi_D(x, y)$ as a function of $y \in (-1, 1)$ is equal to the characteristic function of the interval $(-\sqrt{1-x^2}, \sqrt{1-x^2})$. Thus we compute

$$|D| = \int_{[-1,1]^2} \chi_D = \int_{-1}^{1}\int_{-1}^{1} \chi_D(x, y)\,\mathrm{d}y\,\mathrm{d}x = \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \mathrm{d}y\,\mathrm{d}x = \int_{-1}^{1} 2\sqrt{1-x^2}\,\mathrm{d}x. \tag{95}$$

An antiderivative of $2\sqrt{1-x^2}$ for $x \in (-1, 1)$ can be found by the substitution $x = \cos t$, as

$$\int 2\sqrt{1-x^2}\,\mathrm{d}x = -\int 2\sin^2 t\,\mathrm{d}t = \int (\cos 2t - 1)\mathrm{d}t = \frac{\sin 2t}{2} - t + C, \tag{96}$$

yielding

$$|D| = \int_{-1}^{1} 2\sqrt{1-x^2}\,\mathrm{d}x = \Big(\frac{\sin 2t}{2} - t\Big)\Big|_{\pi}^{0} = \pi. \tag{97}$$

Exercise 6.3. Let $K \subset Q$ be a Jordan set. Prove the following.

(a) If f and g are integrable over K, and $\lambda, \mu \in \mathbb{R}$, then $\lambda f + \mu g$ is integrable over K, and

$$\int_K (\lambda f + \mu g) = \lambda \int_K f + \mu \int_K g. \tag{98}$$

(b) If f and g are integrable over K satisfying $f \le g$, then we have

$$\int_K f \le \int_K g. \tag{99}$$

(c) If f is integrable over K, then $|f|$ is integrable, and

$$\Big|\int_K f\Big| \le \int_K |f|. \tag{100}$$

Exercise 6.4. Let $K \subset Q$ and $L \subset Q$ be Jordan sets, and suppose that $f : Q \to \mathbb{R}$ is integrable over both K and L. Show that f is integrable over both $K \cup L$ and $K \cap L$, with

$$\int_{K\cup L} f + \int_{K\cap L} f = \int_K f + \int_L f. \tag{101}$$

7. Linear transformations

Let f be a continuous function defined in $\mathbb{R}$, and let F be its antiderivative (which can be constructed by using the Riemann integral). Let ϕ be a continuously differentiable function defined in $\mathbb{R}$. Then by applying the chain rule, we get

$$\frac{\mathrm{d}}{\mathrm{d}x} F(\phi(x)) = F'(\phi(x))\phi'(x) = f(\phi(x))\phi'(x), \tag{102}$$

which shows that $F \circ \phi$ is an antiderivative of $(f \circ \phi)\phi'$. Since f and $(f \circ \phi)\phi'$ are both continuous, in light of the fundamental theorem of calculus, we also have

$$\int_{\phi(a)}^{\phi(b)} f = F(\phi(b)) - F(\phi(a)) = \int_a^b (f \circ \phi)\phi', \tag{103}$$

which is called the substitution rule, or the change of variables formula for integrals. In this and the next sections, we want to extend the substitution rule to integrable functions (as opposed to continuous functions only), and to higher dimensional Riemann integrals.

We start with (homogeneous) linear transformations in one dimension. Thus we let $\lambda \in \mathbb{R}$ and $\phi(x) = \lambda x$. Under the mapping ϕ, the interval $[a, b]$ is transformed to

$$P = \phi([a, b]) = \begin{cases} [\lambda a, \lambda b] & \text{if } \lambda \geq 0, \\ [\lambda b, \lambda a] & \text{if } \lambda < 0. \end{cases} \tag{104}$$

If $\lambda = 0$, then $P = \{0\}$ and $\phi' = 0$, meaning that the substitution rule (103) becomes a triviality. Hence in the following, we assume that $\lambda \neq 0$. Suppose that f is integrable on P. Then for any given $\varepsilon > 0$, there exist step functions g and h subordinate to a grid Γ, such that $g \leq f \leq h$ in each $J \in \Gamma$ and

$$\int_P (h - g) \leq \varepsilon. \tag{105}$$

The grid Γ induces a grid on $[a, b]$, as $\Gamma' = \{\phi^{-1}(J) : J \in \Gamma\}$. Then the functions $\tilde{g} = g \circ \phi$ and $\tilde{h} = h \circ \phi$ are step functions subordinate to Γ', and $\tilde{g} \leq f \circ \phi \leq \tilde{h}$ in each $I \in \Gamma'$. Furthermore, we have

$$\int_a^b (\tilde{h} - \tilde{g}) = \sum_{J \in \Gamma} (h - g)\big|_J |\phi^{-1}(J)| = \sum_{J \in \Gamma} (h - g)\big|_J \frac{|J|}{|\lambda|} = |\lambda|^{-1} \int_P (h - g) \leq \frac{\varepsilon}{|\lambda|}, \tag{106}$$

implying that $f \circ \phi$ is integrable over $[a, b]$, with

$$\int_a^b f \circ \phi = \lim_{\varepsilon \to 0} \int_a^b \tilde{g} = \lim_{\varepsilon \to 0} |\lambda|^{-1} \int_P g = |\lambda|^{-1} \int_P f. \tag{107}$$

We can write it as

$$\int_P f = |\lambda| \int_{[a,b]} f \circ \phi, \tag{108}$$

which gives the substitution rule (103) in the particular case $\phi(x) = \lambda x$ with $\lambda > 0$, because in this case we have $\phi'(x) = \lambda = |\lambda|$ and $P = [\phi(a), \phi(b)]$. In fact, for $\lambda < 0$, we have

$$\int_{\phi(a)}^{\phi(b)} f = -\int_{\phi(b)}^{\phi(a)} f = -\int_P f = -|\lambda| \int_{[a,b]} f \circ \phi = \lambda \int_a^b f \circ \phi, \tag{109}$$

meaning that (103) holds regardless of the sign of λ.

Remark 7.1 (Oriented intervals)**.** In order to generalize (109) to higher dimensions, we need to discuss what $\int_a^c f$ with $c < a$ would become in higher dimensions. Note that the integral $\int_a^c f$ with $c < a$ may be considered as an integral of f over $[c, a]$, but the interval $[c, a]$ is equipped with a negative length (or a negative density). Thinking along these lines leads to the notion of oriented intervals, which yields a form of (109) that is more amenable to generalization. An *oriented interval* is an interval, together with a choice of a sign (that is, $+1$ or -1). Thus the interval $I = [0, 1]$ may have *positive orientation*, meaning that we choose $+1$ as the orientation of I, or it may have *negative orientation*, where we choose -1 as the orientation of I. In this context, the interval $[0, 1]$ with negative orientation will be denoted by $[1, 0]$. More generally, when $J = [a, b]$ is declared to be an oriented interval, we will assume that the orientation of J is equal to the sign of $b - a$, where of course $a > b$ is allowed. We sometimes write $-J = [b, a]$ when $J = [a, b]$, to indicate orientation reversing. Furthermore, we define

$$\int_{[b,a]} f = -\int_{[a,b]} f, \tag{110}$$

for $a < b$. Finally, we orient the image $\phi([a, b])$ of $[a, b]$ under $\phi(x) = \lambda x$ according to the following rule. If $\lambda > 0$, then $\phi([a, b])$ and $[a, b]$ have the same orientation, and if $\lambda < 0$, then $\phi([a, b])$ and $[a, b]$ have the opposite orientations. In other words, we set the orientation of $\phi([a, b])$ as $\operatorname{sgn}(\lambda)$ times the orientation of $[a, b]$. We now can write (109) as

$$\int_{\phi([a,b])} f = \lambda \int_{[a,b]} f \circ \phi. \tag{111}$$

The reader may verify that this is valid even for $a > b$.

In the rest of this section, we want to generalize (111) to maps $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ of the form $\Phi(x) = Ax$, where $A \in \mathbb{R}^{n\times n}$. This will be achieved in several steps. As a first step, let us assume that A is diagonal, that is, $\Phi(x) = (\lambda_1 x_1, \dots, \lambda_n x_n)$ for some constants $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Let $Q \subset \mathbb{R}^n$ be a rectangle, and let $f : \Phi(Q) \to \mathbb{R}$ be an integrable function, where $\Phi(Q) = \{\Phi(x) : x \in Q\}$. We expect that the eventual formula will be of the form

$$\int_{\Phi(Q)} f = c \int_Q f \circ \Phi, \tag{112}$$

with some constant $c = c(\lambda_1, \dots, \lambda_n)$. It is easy to see that $\Phi(Q) \subset \mathbb{R}^n$ is a rectangle, possibly degenerate (meaning that $|\Phi(Q)| = 0$). The degeneracy occurs if $\lambda_k = 0$ for some k. This will turn out to be a trivial case, and from now on, we assume that $\lambda_1 \cdots \lambda_n \neq 0$, that is, the map Φ is invertible. For any rectangle $q \subset \mathbb{R}^n$, the image $\Phi^{-1}(q)$ is again a rectangle, with volume

$$|\Phi^{-1}(q)| = \frac{|q|}{|\lambda_1 \cdots \lambda_n|}. \tag{113}$$

We can now basically repeat what we did in the one dimensional case. For any given $\varepsilon > 0$, there exist step functions g and h subordinate to a grid Γ of $\Phi(Q)$, such that $g \leq f \leq h$ in each $q \in \Gamma$ and

$$\int_{\Phi(Q)} (h - g) \leq \varepsilon. \tag{114}$$

The grid Γ induces a grid on Q, as $\Gamma' = \{\Phi^{-1}(q) : q \in \Gamma\}$. Then the functions $\tilde{g} = g \circ \Phi$ and $\tilde{h} = h \circ \Phi$ are step functions subordinate to Γ', and $\tilde{g} \leq f \circ \Phi \leq \tilde{h}$ in each $q \in \Gamma'$. Furthermore, we have

$$\begin{aligned} \int_Q (\tilde{h} - \tilde{g}) &= \sum_{q\in\Gamma} (h-g)\big|_q |\Phi^{-1}(q)| = \sum_{q\in\Gamma} (h-g)\big|_q \frac{|q|}{|\lambda_1 \cdots \lambda_n|} \\ &= |\lambda_1 \cdots \lambda_n|^{-1} \int_P (h - g) \leq \frac{\varepsilon}{|\lambda_1 \cdots \lambda_n|}, \end{aligned} \tag{115}$$

implying that $f \circ \Phi$ is integrable over Q, with

$$\int_Q f \circ \Phi = \lim_{\varepsilon\to 0} \int_Q \tilde{g} = \lim_{\varepsilon\to 0} |\lambda_1 \cdots \lambda_n|^{-1} \int_{\Phi(Q)} g = |\lambda_1 \cdots \lambda_n|^{-1} \int_{\Phi(Q)} f. \tag{116}$$

We can write it as

$$\int_{\Phi(Q)} f = |\lambda_1 \cdots \lambda_n| \int_Q f \circ \Phi, \tag{117}$$

which is the substitution rule we have been looking for. We can also check that if $\lambda_k = 0$ for some k, then both sides of the preceding equality vanish, and therefore it is trivially true.

Remark 7.2 (Oriented domains)**.** It is now a matter of introducing new definitions to extend (117) to the "oriented" setting. We define an *oriented domain* as a set $K \subset \mathbb{R}^n$, together with a choice of a sign (that is, $+1$ or -1). If K is an oriented domain, the same domain with the opposite orientation is denoted by $-K$. For positively oriented domains, the integral $\int_Q f$ is defined to be the same integral, where Q is considered to be a subset of $\mathbb{R}^n$, that is, we may simply forget about the orientation. For negatively oriented domains, we define

$$\int_{-Q} f = -\int_Q f, \tag{118}$$

where Q is positively oriented. Furthermore, if Q is an oriented domain and if $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $\Phi(x) = (\lambda_1 x_1, \ldots, \lambda_n x_n)$, then the orientation of $\Phi(Q)$ is set to be $\operatorname{sgn}(\lambda_1 \cdots \lambda_n)$ times the orientation of Q. With these new concepts at hand, it is not difficult to derive

$$\int_{\Phi(Q)} f = \lambda_1 \cdots \lambda_n \int_Q f \circ \Phi, \tag{119}$$

which now holds even when Q is an oriented rectangle. As there is a slight danger of confusion between (117) and (119), one must be careful to specify whether the integrals are in the "oriented" setting as in (119), or in the "plain vanilla" setting as in (117).

To extend the substitution rule to general linear transformations $\Phi(x) = Ax$, we need to prove a couple of preliminary lemmata. The first lemma generalizes the integrability criterion to allow approximation by integrable functions, as opposed to step functions only.

Lemma 7.3. *Let $f : Q \to \mathbb{R}$, and suppose that for any $\varepsilon > 0$, there exist integrable functions g_ε and h_ε, such that $g_\varepsilon \leq f \leq h_\varepsilon$ and*

$$\int_Q (h_\varepsilon - g_\varepsilon) \leq \varepsilon. \tag{120}$$

Then f is integrable, and

$$\int_Q f = \lim_{\varepsilon \to 0} \int_Q g_\varepsilon. \tag{121}$$

Proof. Let $\varepsilon > 0$ be arbitrary, and let g_ε and h_ε be as in the statement. Since g_ε and h_ε are integrable, there exist step functions $\tilde{g}_\varepsilon$ and $\tilde{h}_\varepsilon$, subordinate to some grid Γ, such that $\tilde{g}_\varepsilon \leq g_\varepsilon$ and $h_\varepsilon \leq \tilde{h}_\varepsilon$ in each $q \in \Gamma$, and that

$$\int_Q (g_\varepsilon - \tilde{g}_\varepsilon) \leq \varepsilon, \qquad \text{and} \qquad \int_Q (\tilde{h}_\varepsilon - h_\varepsilon) \leq \varepsilon. \tag{122}$$

This implies that $\tilde{g}_\varepsilon \leq f \leq \tilde{h}_\varepsilon$ in each $q \in \Gamma$, and that

$$\int_Q (\tilde{h}_\varepsilon - \tilde{g}_\varepsilon) \leq \int_Q (\tilde{h}_\varepsilon - h_\varepsilon) + \int_Q (h_\varepsilon - g_\varepsilon) + \int_Q (g_\varepsilon - \tilde{g}_\varepsilon) \leq 3\varepsilon. \tag{123}$$

As $\varepsilon > 0$ is arbitrary, we conclude that f is integrable. □

The following lemma reduces the computation of the volume $|\Phi(Q)|$ for an arbitrary rectangle $Q \subset \mathbb{R}^n$ to that of $|\Phi(Q_1)|$, where $Q_1 = [0, 1]^n$ is the unit cube.

Lemma 7.4. *Let $Q \subset \mathbb{R}^n$ be a rectangle, and let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be defined by $\Phi(x) = Ax$, with some $A \subset \mathbb{R}^{n\times n}$. Then we have $|\Phi(Q)| = |Q||\Phi(Q_1)|$, where $Q_1 = [0, 1]^n$ is the unit cube.*

Proof. If A is singular, then $\dim \Phi(\mathbb{R}^n) < n$, and hence $|\Phi(Q)| = |\Phi(Q_1)| = 0$. In the following, we assume that A is invertible.

For $\lambda > 0$, let $Q_\lambda = [0, \lambda]^n$. Then we have $Q_\lambda = \Lambda(Q_1)$, where $\Lambda : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $\Lambda(x) = \lambda x$. Since $\Phi(\lambda x) = \lambda\Phi(x)$, we have $\Phi(Q_\lambda) = \Lambda(\Phi(Q_1))$, and hence

$|\Phi(Q_\lambda)| = \lambda^n|\Phi(Q_1)|$. Now, by using the nodal points $0, \lambda, 2\lambda, \ldots$ in each coordinate direction, we construct a grid Γ in a rectangle R containing $Q = [0, b_1] \times \ldots \times [0, b_n]$. Thus, each $q \in \Gamma$ is a cube with side-length λ, and the region R must be strictly larger than Q whenever any of $b_1, \ldots, b_n$ is a non-integer multiple of λ. More precisely, let $\ell_k \in \mathbb{N}$ be such that $\lambda\ell_k < b_k$ and $\lambda(\ell_k + 1) \geq b_k$, for $k = 1, \ldots, n$, and we set $R = [0, \lambda(\ell_1 + 1)] \times \ldots \times [0, \lambda(\ell_n + 1)]$. Next, we set $h = \chi_R$, and define g to be the characteristic function of $R' = [0, \lambda\ell_1] \times \ldots \times [0, \lambda\ell_n]$. It is clear that $g \leq \chi_Q \leq h$, and that

$$\int_R (h - g) = |R \setminus R'| \leq C\lambda, \tag{124}$$

for some constant C independent of λ. We let $\tilde{g} = g \circ \Phi^{-1}$ and $\tilde{h} = h \circ \Phi^{-1}$. Then we have $\tilde{g} \leq \chi_{\Phi(Q)} \leq \tilde{h}$ and

$$\int_P (\tilde{h} - \tilde{g}) = \sum_{q \in \Gamma \setminus \Gamma'} |\Phi(q)| = \sum_{q \in \Gamma \setminus \Gamma'} \lambda^n|\Phi(Q_1)| = |R \setminus R'||\Phi(Q_1)| \leq C|\Phi(Q_1)|\lambda, \tag{125}$$

where $P \subset \mathbb{R}^n$ is some large rectangle containing $\Phi(Q)$, and $\Gamma' = \{q \in \Gamma : q \subset R'\}$. As $\lambda > 0$ is arbitrary, Lemma 7.3 guarantees that the integral of $\chi_{\Phi(Q)}$ can be computed by taking the limit $\lambda \to 0$ in

$$\int_P \tilde{g} = \sum_{q \in \Gamma'} |\Phi(q)| = \sum_{q \in \Gamma'} \lambda^n|\Phi(Q_1)| = |R'||\Phi(Q_1)|. \tag{126}$$

Since $|R'| \to |Q|$ as $\lambda \to 0$, we have $|\Phi(Q)| = |Q||\Phi(Q_1)|$. □

To compute the volume $|\Phi(Q_1)|$, we consider it as a function of the matrix A, or equivalently, as a function of the columns $A_1, \ldots, A_n$ of A, as

$$\omega(A) = \omega(A_1, \ldots, A_n) := |\Phi(Q_1)|. \tag{127}$$

Geometrically, $\Phi(Q_1)$ is an n-dimensional parallelogram whose sides are given by the columns of A. First of all, we know that $\omega(A) = 0$ if A is singular. In particular, we have

$$\omega(\ldots, V, \ldots, V, \ldots) = 0 \qquad \text{for} \quad V \in \mathbb{R}^n. \tag{128}$$

Let us scale one of the columns of A by a factor $\lambda \geq 0$, and see how ω changes. We can write Φ in terms of the columns of A as $\Phi(x) = x_1A_1 + \ldots + x_nA_n$. If we scale the first columns of A by λ, we get the new map $\Phi_\lambda(x) = x_1\lambda A_1 + x_2A_2 + \ldots + x_nA_n$. The parallelogram $\Phi_\lambda(Q_1)$ is obtained by scaling $\Phi(Q_1)$ along the side represented by A_1. We have

$$\begin{aligned} \Phi_\lambda(Q_1) &= \{x_1\lambda A_1 + x_2A_2 + \ldots + x_nA_n : 0 \leq x_k \leq 1,\, k = 1, \ldots, n\} \\ &= \{x_1A_1 + x_2A_2 + \ldots + x_nA_n : 0 \leq x_1 \leq \lambda,\, 0 \leq x_k \leq 1,\, k = 2, \ldots, n\} \\ &= \Phi([0, \lambda] \times [0, 1] \times \ldots \times [0, 1]), \end{aligned} \tag{129}$$

and since the volume of $[0, \lambda] \times [0, 1] \times \ldots \times [0, 1]$ is λ, Lemma 7.4 gives

$$\omega(\lambda A_1, A_2, \ldots, A_n) = |\Phi_\lambda(Q_1)| = \lambda|\Phi(Q_1)| = \lambda\omega(A_1, A_2, \ldots, A_n). \tag{130}$$

If we pick $\lambda < 0$, then the same argument yields

$$\Phi_\lambda(Q_1) = \Phi([\lambda, 0] \times [0, 1] \times \ldots \times [0, 1]), \tag{131}$$

and therefore, for $\lambda \in \mathbb{R}$, we have

$$\omega(\lambda A_1, A_2, \ldots, A_n) = |\lambda|\omega(A_1, A_2, \ldots, A_n). \tag{132}$$

Now, we consider the map $\Psi_\beta(x) = x_1(A_1+\beta A_2)+x_2A_2+\ldots+x_nA_n$ where $\beta \in \mathbb{R}$, that is, we add a multiple of the second column to the first column of A. The parallelogram $\Psi_\beta(Q_1)$ is obtained by applying a shear transformation to $\Phi(Q_1)$. In this case, we have

$$\begin{aligned}\Psi_\beta(Q_1) &= \{x_1(A_1+\beta A_2)+x_2A_2+\ldots+x_nA_n : 0 \leq x_k \leq 1,\, k=1,\ldots,n\}\\ &= \{x_1A_1+(x_2+\beta x_1)A_2+\ldots+x_nA_n : 0 \leq x_k \leq 1,\, k=1,\ldots,n\}\\ &= \Phi(S_\beta),\end{aligned} \tag{133}$$

where S_β would be a shear transformation of Q_1, and is defined by

$$S_\beta = \{(x_1, x_2+\beta x_1, x_3, \ldots, x_n) : x \in Q_1\}. \tag{134}$$

It is not difficult to show that $|S_\beta| = 1$, and hence Lemma 7.4 yields

$$\omega(A_1+\beta A_2, A_2, \ldots, A_n) = |\Psi_\beta(Q_1)| = |\Phi(S_\beta)| = |S_\beta||\Phi(Q_1)| = \omega(A_1, A_2, \ldots, A_n). \tag{135}$$

We conclude that shear transformations do not change volume.

Exercise 7.5. Show that $|S_\beta| = 1$.

Supposing that A is invertible, any vector $V \in \mathbb{R}^n$ can be written as

$$V = \beta_1A_1+\ldots+\beta_nA_n. \tag{136}$$

Then by using the transformation properties (130) and (135), for $\beta_1 \geq 0$, we have

$$\begin{aligned}\omega(A_1+V, A_2, \ldots, A_n) &= \omega(A_1+\beta_1A_1, A_2, \ldots, A_n) = (1+\beta_1)\omega(A_1, A_2, \ldots, A_n)\\ &= \omega(A_1, A_2, \ldots, A_n) + \beta_1\omega(A_1, A_2, \ldots, A_n)\\ &= \omega(A_1, A_2, \ldots, A_n) + \omega(\beta_1A_1, A_2, \ldots, A_n)\\ &= \omega(A_1, A_2, \ldots, A_n) + \omega(V, A_2, \ldots, A_n).\end{aligned} \tag{137}$$

For $-1 \leq \beta_1 < 0$, we have

$$\begin{aligned}\omega(A_1+V, A_2, \ldots, A_n) &= (1+\beta_1)\omega(A_1, A_2, \ldots, A_n)\\ &= \omega(A_1, A_2, \ldots, A_n) - (-\beta_1)\omega(A_1, A_2, \ldots, A_n)\\ &= \omega(A_1, A_2, \ldots, A_n) - \omega(-\beta_1A_1, A_2, \ldots, A_n)\\ &= \omega(A_1, A_2, \ldots, A_n) - \omega(-V, A_2, \ldots, A_n).\end{aligned} \tag{138}$$

Finally, for $\beta_1 < -1$, we have

$$\begin{aligned}\omega(A_1+V, A_2, \ldots, A_n) &= -(1+\beta_1)\omega(-A_1, A_2, \ldots, A_n)\\ &= -\omega(-A_1, A_2, \ldots, A_n) + (-\beta_1)\omega(-A_1, A_2, \ldots, A_n)\\ &= -\omega(-A_1, A_2, \ldots, A_n) + \omega(\beta_1A_1, A_2, \ldots, A_n)\\ &= -\omega(-A_1, A_2, \ldots, A_n) + \omega(V, A_2, \ldots, A_n).\end{aligned} \tag{139}$$

We notice from the preceding equalities that all three cases *would* be special cases of

$$\omega(A_1+V, A_2, \ldots, A_n) = \omega(A_1, A_2, \ldots, A_n) + \omega(V, A_2, \ldots, A_n), \tag{140}$$

if we had the property $\omega(-A_1, A_2, \ldots, A_n) = -\omega(A_1, A_2, \ldots, A_n)$. This idea leads to the proof of the following result.

Lemma 7.6. *We have* $\omega(A) = |\det(A)|$.

Proof. We introduce

$$\tilde{\omega}(A) = \operatorname{sgn}(\det(A))\omega(A) = \begin{cases}\omega(A) & \text{for } \det(A) \geq 0,\\ -\omega(A) & \text{for } \det(A) < 0.\end{cases} \tag{141}$$

Then (132) implies

$$\tilde{\omega}(\lambda A_1, A_2, \ldots, A_n) = \lambda\tilde{\omega}(A_1, A_2, \ldots, A_n), \qquad \lambda \in \mathbb{R}, \tag{142}$$

and an inspection of (137)-(139) reveals that the additivity property (140) holds for $\tilde{\omega}$. Of course, (140) and (142) hold not only for the first argument, but for each argument of $\tilde{\omega}$. We also have (128) for $\tilde{\omega}$, meaning that $\tilde{\omega}$ is a totally antisymmetric n-linear function. Now linear algebra tells us that

$$\tilde{\omega}(A) = \tilde{\omega}(I)\det(A) = \det(A). \tag{143}$$

Finally, we conclude that $\omega(A) = \operatorname{sgn}(\det(A))\tilde{\omega}(A) = \operatorname{sgn}(\det(A))\det(A) = |\det(A)|$. □

In combination with Lemma 7.4, the preceding lemma implies that $|\Phi(Q)| = |Q||\det(A)|$ for any rectangle $Q \subset \mathbb{R}^n$ and a linear map $\Phi(x) = Ax$ with some $A \subset \mathbb{R}^{n\times n}$.

Theorem 7.7. *Let $Q \subset \mathbb{R}^n$ be a rectangle, and let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be defined by $\Phi(x) = Ax$, with some $A \subset \mathbb{R}^{n\times n}$. Let $f : \Phi(Q) \to \mathbb{R}$ be a bounded function such that $f \circ \Phi$ is integrable over Q. Then f is integrable over $\Phi(Q)$, with*

$$\int_{\Phi(Q)} f = |\det(A)| \int_Q f \circ \Phi. \tag{144}$$

Proof. We assume that A is invertible, since otherwise the theorem is trivial. For any given $\varepsilon > 0$, there exist step functions g and h subordinate to a grid Γ of Q, such that $g \leq f\circ\Phi \leq h$ in each $q \in \Gamma$ and

$$\int_Q (h - g) \leq \varepsilon. \tag{145}$$

Let $B = Q \setminus \bigcup\{q : q \in \Gamma\}$ be the space between the grid rectangles. We can modify g and h in B, such that $g \leq f \circ \Phi \leq h$ holds everywhere in Q. Then $\tilde{g} = g \circ \Phi^{-1}$ and $\tilde{h} = h \circ \Phi^{-1}$ are integrable functions satisfying $\tilde{g} \leq f \leq \tilde{h}$, and

$$\begin{aligned} \int_{\Phi(Q)} (\tilde{h} - \tilde{g}) &= \sum_{q\in\Gamma} (h-g)\big|_q |\Phi(q)| = \sum_{q\in\Gamma} (h-g)\big|_q |\det(A)||q| \\ &= |\det(A)| \int_Q (h-g) \leq |\det(A)|\varepsilon. \end{aligned} \tag{146}$$

Invoking Lemma 7.3, we infer that f is integrable over $\Phi(Q)$, with

$$\int_{\Phi(Q)} f = \lim_{\varepsilon\to 0} \int_{\Phi(Q)} \tilde{g} = \lim_{\varepsilon\to 0} |\det(A)| \int_Q g = |\det(A)| \int_Q f \circ \Phi. \tag{147}$$

This completes the proof. □

Remark 7.8. Here we extend Theorem 7.7 to the setting where the domains are equipped with orienations. If Q is an oriented domain and if $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation $\Phi(x) = Ax$, then the orientation of $\Phi(Q)$ is *defined* to be $\operatorname{sgn}(\det(A))$ times the orientation of Q. In other words, transformations with $\det(A) > 0$ preserve the orientation of Q, and those with $\det(A) < 0$ flip the orientation of Q. Then (144) in Theorem 7.7 should be replaced by

$$\int_{\Phi(Q)} f = \det(A) \int_Q f \circ \Phi, \tag{148}$$

where it is now understood that the integrals take into account the orientation of their domains, cf. (118) in Remark 7.2.

8. Change of variables

In this section, we will extend the substitution rule from the preceding section to a general class of differentiable mappings. The main device that makes everything work is local approximation of differentiable mappings by linear maps.

Let $U \subset \mathbb{R}^n$ be an open set, and let $\Phi : U \to \mathbb{R}^n$ be a continuously differentiable map. Let $Q \subset U$ be a closed rectangle. We assume that the restriction $\Phi : Q \to \mathbb{R}^n$ is injective, and that the inverse $\Phi^{-1} : \Phi(Q) \to Q$ is continuously differentiable. Let $f : \Phi(Q) \to \mathbb{R}$ be a bounded function such that $f \circ \Phi$ is integrable over Q. Following the proof of Theorem 7.7, for any given $\varepsilon > 0$, there exist step functions g and h subordinate to a grid Γ of Q, such that $g \leq f \circ \Phi \leq h$ in each $q \in \Gamma$ and

$$\int_Q (h - g) \leq \varepsilon. \tag{149}$$

Let $B = Q \setminus \bigcup\{q : q \in \Gamma\}$. We can modify g and h in B, such that $g \leq f \circ \Phi \leq h$ holds everywhere in Q. By Lemma 5.6, the set $\Phi(B)$ is negligible, and hence $\tilde{g} = g \circ \Phi^{-1}$ and $\tilde{h} = h \circ \Phi^{-1}$ are integrable functions satisfying $\tilde{g} \leq f \leq \tilde{h}$.

- If a rectangle $q \in \Gamma$ is very small, then $\Phi(x) \approx \Phi(x_q) + D\Phi(x_q)(x - x_q)$ for $x \in q$, where $x_q \in q$ is some point. Hence we expect that $|\Phi(q)| \approx |\det(D\Phi(x_q))|$, and
$$\begin{aligned}\int_{\Phi(Q)} (\tilde{h} - \tilde{g}) &= \sum_{q\in\Gamma} (h - g)\big|_q |\Phi(q)| \approx \sum_{q\in\Gamma} (h - g)\big|_q |\det(D\Phi(x_q))||q| \\ &\leq \max_Q |\det(D\Phi)| \int_Q (h - g) \leq C\varepsilon.\end{aligned} \tag{150}$$
If we can justify this expectation, then by Lemma 7.3 we would infer that f is integrable over $\Phi(Q)$.
- The next step would be to take the limit $\varepsilon \to 0$ in
$$\int_{\Phi(Q)} \tilde{g} = \sum_{q\in\Gamma} g\big|_q |\Phi(q)| \approx \sum_{q\in\Gamma} g\big|_q \,|\det(D\Phi(x_q))||q| \approx \int_Q g\,|\det(D\Phi)|, \tag{151}$$
which suggests the guess
$$\int_{\Phi(Q)} f = \int_Q f \circ \Phi\,|\det(D\Phi)|. \tag{152}$$

We shall now rigorously justify the aforementioned heuristic reasoning. The following result will be of fundamental importance.

Theorem 8.1. *For any $\delta > 0$, there exists $\rho > 0$ such that whenever $q \subset Q$ is a rectangle with diameter not exceeding ρ, we have*

$$(1 - \gamma\delta)|\det(D\Phi(x_q))||q| \leq |\Phi(q)| \leq (1 + \gamma\delta)|\det(D\Phi(x_q))||q|, \tag{153}$$

where x_q is the centre of q, and γ is the ratio between the longest and the shortest sides of q.

Proof. Let $\delta > 0$ be a small number, and let $q = [a_1, b_1] \times \ldots \times [a_n, b_n] \subset Q$ be a rectangle. Without loss of generality, we assume that q is closed, and that q is nondegenerate in the sense that $|q| > 0$. Let $\Lambda : \mathbb{R}^n \to \mathbb{R}^n$ be defined by $\Lambda(x) = \big(\frac{a_1+b_1+(b_1-a_1)x_1}{2}, \ldots, \frac{a_n+b_n+(b_n-a_n)x_n}{2}\big)$. Note that $\Lambda(Q_1) = q$, where $Q_1 = [-1, 1]^n$, and $\Lambda(0) = x_q$. We can write $\Lambda(x) = x_q + Ax$ with a diagonal matrix $A \in \mathbb{R}^{n\times n}$. Next, we define $F : Q_1 \to \mathbb{R}^n$ by

$$F(x) = \Lambda^{-1}(D\Phi(x_q)^{-1}\Phi(\Lambda(x))). \tag{154}$$

By Lemma 8.2 below, for all $x, y \in Q_1$, we have

$$|F(y) - F(x) - DF(0)(y - x)|_\infty \leq n|y - x|_\infty \sup_{z\in Q_1} |DF(z) - DF(0)|_\infty. \tag{155}$$

We compute

$$DF(x) = A^{-1}D\Phi(x_q)^{-1}D\Phi(\Lambda(x))A, \tag{156}$$

which implies that $DF(0) = I$ and

$$DF(z) - DF(0) = A^{-1}D\Phi(x_q)^{-1}(D\Phi(\Lambda(z)) - D\Phi(x_q))A. \tag{157}$$

Taking into account that A is diagonal, we infer

$$\begin{aligned} |DF(z) - DF(0)|_\infty &\le \gamma|D\Phi(x_q)^{-1}(D\Phi(\Lambda(z)) - D\Phi(x_q))|_\infty \\ &\le \gamma n \sup_Q |(D\Phi)^{-1}|_\infty |D\Phi(\Lambda(z)) - D\Phi(x_q))|_\infty \end{aligned} \tag{158}$$

By the uniform continuity of $D\Phi$, we can choose $\rho > 0$ such that

$$n^2 \sup_Q |(D\Phi)^{-1}|_\infty \sup_{z\in q} |D\Phi(z) - D\Phi(x_q)|_\infty \le \delta, \tag{159}$$

whenever $\operatorname{diam}(q) \le \rho$, yielding

$$|F(y) - F(x) - (y - x)|_\infty \le \gamma\delta|y - x|_\infty, \tag{160}$$

for all $x, y \in Q_1$. From this, we have

$$|F(y) - F(x)|_\infty \le |y - x|_\infty + |F(y) - F(x) - (y - x)|_\infty \le (1 + \gamma\delta)|y - x|_\infty, \tag{161}$$

and putting $y = 0$, we get

$$|F(x)|_\infty \le (1 + \gamma\delta)|x|_\infty, \tag{162}$$

meaning that $F(Q_1) \subset (1 + \gamma\delta)Q_1 := \{(1 + \gamma\delta)x : x \in Q_1\}$. A similar reasoning gives

$$|F(y) - F(x)|_\infty \ge |y - x|_\infty - |F(y) - F(x) - (y - x)|_\infty \ge (1 - \gamma\delta)|y - x|_\infty, \tag{163}$$

implying that F is injective if γ is suitably controlled.

Our next goal is to show that $(1 - \gamma\delta)Q_1 \subset F(Q_1)$, that is, to show that the equation $F(x) = z$ can be solved for the unknown $x \in Q_1$ for any given $z \in (1 - \gamma\delta)Q_1$. Suppose that $z \in (1 - \gamma\delta)Q_1$, and define the map $\psi : Q_1 \to \mathbb{R}^n$ by $\psi(x) = z + x - F(x)$. We have $\psi(x) = x$ if and only if $F(x) = z$. Moreover, from (160) we get

$$|\psi(x) - \psi(y)|_\infty = |x - y + F(y) - F(x)|_\infty \le \gamma\delta|y - x|_\infty, \tag{164}$$

and

$$|\psi(x)|_\infty \le |z|_\infty + |x - 0 + F(0) - F(x)|_\infty \le |z|_\infty + \gamma\delta|x|_\infty \le (1 - \gamma\delta) + \gamma\delta = 1, \tag{165}$$

meaning that $\psi(Q_1) \subset Q_1$ and ψ is a contraction. Therefore, by the contraction mapping principle, there is a unique $x \in Q_1$ satisfying $\psi(x) = x$. This shows that $F(x) = z$ has a solution for any given $z \in (1 - \gamma\delta)Q_1$, and hence $(1 - \gamma\delta)Q_1 \subset F(Q_1)$.

At this point, we can write

$$(1 - \gamma\delta)Q_1 \subset F(Q_1) \subset (1 + \gamma\delta)Q_1. \tag{166}$$

Taking into account the definition $F(x) = \Lambda^{-1}(D\Phi(x_q)^{-1}\Phi(\Lambda(x)))$, and unwinding the above inclusion, we get

$$D\Phi(x_q)(\Lambda((1 - \gamma\delta)Q_1)) \subset \Phi(q) \subset D\Phi(x_q)(\Lambda((1 + \gamma\delta)Q_1)), \tag{167}$$

which completes the proof. □

Lemma 8.2. *Let $\Omega \subset \mathbb{R}^n$ be an open set, and let $F : \Omega \to \mathbb{R}^n$ be differentiable in Ω. Suppose that $DF(x^*)$ is invertible for some $x^* \in K$, where $K \subset \Omega$ is a convex set. Then we have*

$$|F(y) - F(x) - DF(x^*)(y - x)|_\infty \le n|y - x|_\infty \sup_{z\in K} |DF(z) - DF(x^*)|_\infty, \tag{168}$$

for any $x, y \in K$.

Proof. Let $x, y \in K$, and let $g(t) = F(x + tV)$, where $V = y - x$. Then we have

$$F(y) - F(x) = g(1) - g(0), \tag{169}$$

and

$$g(t) = D_V F(x + tV) = DF(x + tV)V. \tag{170}$$

By the mean value theorem, for each k, there exists $0 < t_k < 1$ such that $g_k(1) - g_k(0) = g_k'(t_k)$, that is,

$$F_k(x) - F_k(y) = D_V F_k(x + t_k V) = DF_k(x + t_k V)V = DF_k(x + t_k V)(y - x). \tag{171}$$

This implies that

$$F_k(y) - F_k(x) - DF_k(x^*)(y - x) = (DF_k(x + t_k V) - DF_k(x^*))(y - x), \tag{172}$$

and hence

$$|F_k(y) - F_k(x) - DF_k(x^*)(y - x)| \leq \sum_{i=1}^{n} |\partial_i F_k(x + t_k V) - \partial_i F_k(x^*)||y_i - x_i|. \tag{173}$$

Note that $\partial_i F_k$ is simply an entry in the Jacobian matrix DF, and so we have

$$|\partial_i F_k(x + t_k V) - \partial_i F_k(x^*)| \leq \sup_{z \in K} |DF(z) - DF(x^*)|_\infty, \tag{174}$$

because $x + t_k V \in K$ by convexity. This completes the proof. □

Finally, we are ready to state and prove the general change of variables formula.

Theorem 8.3. *Let $U \subset \mathbb{R}^n$ be an open set, and let $\Phi : U \to \mathbb{R}^n$ be a continuously differentiable map. Let $Q \subset U$ be a closed rectangle. We assume that the restriction $\Phi : Q \to \mathbb{R}^n$ is injective, and that the inverse $\Phi^{-1} : \Phi(Q) \to Q$ is continuously differentiable. Let $f : \Phi(Q) \to \mathbb{R}$ be a bounded function such that $f \circ \Phi$ is integrable over Q. Then f is integrable over $\Phi(Q)$, with*

$$\int_{\Phi(Q)} f = \int_Q f \circ \Phi \, |\det(D\Phi)|. \tag{175}$$

Proof. Without loss of generality, we assume $f \geq 0$, since otherwise, we can decompose $f = \max\{f, 0\} - \max\{-f, 0\}$ and work with two positive functions. For any given $\varepsilon > 0$, there exist step functions g and h subordinate to a grid Γ of Q, such that $0 \leq g \leq f \circ \Phi \leq h$ in each $q \in \Gamma$ and

$$\int_Q (h - g) \leq \varepsilon. \tag{176}$$

Let $B = Q \setminus \bigcup\{q : q \in \Gamma\}$. We can modify g and h in B, such that $g \leq f \circ \Phi \leq h$ holds everywhere in Q. By Lemma 5.6, the set $\Phi(B)$ is negligible, and hence $\tilde{g} = g \circ \Phi^{-1}$ and $\tilde{h} = h \circ \Phi^{-1}$ are integrable functions satisfying $\tilde{g} \leq f \leq \tilde{h}$. Without loss of generality, we can assume that the diameter ρ and the degeneracy parameter γ of $q \in \Gamma$ are suitably controlled, so that Theorem 8.1 can be applied to any $q \in \Gamma$, regardless of the choice of ε, with $\delta > 0$ that can be chosen at will. Thus we have

$$\begin{aligned}\int_{\Phi(Q)} (\tilde{h} - \tilde{g}) &= \sum_{q \in \Gamma} (h - g)\big|_q |\Phi(q)| \leq (1 + \gamma\delta) \sum_{q \in \Gamma} (h - g)\big|_q |\det(D\Phi(x_q))||q| \\ &\leq (1 + \gamma\delta) \max_Q |\det(D\Phi)| \int_Q (h - g) \leq C\varepsilon,\end{aligned} \tag{177}$$

implying that f is integrable over $\Phi(Q)$, cf. Lemma 7.3.

Now we need to take the limit $\varepsilon \to 0$ in

$$\int_{\Phi(Q)} \tilde{g} = \sum_{q \in \Gamma} g\big|_q |\Phi(q)|, \qquad \text{and/or} \qquad \int_{\Phi(Q)} \tilde{h} = \sum_{q \in \Gamma} h\big|_q |\Phi(q)|. \tag{178}$$

We have

$$\sum_{q\in\Gamma} g\big|_q |\Phi(q)| \geq (1-\gamma\delta) \sum_{q\in\Gamma} g\big|_q \, |\det(D\Phi(x_q))||q|, \tag{179}$$

$$\sum_{q\in\Gamma} h\big|_q |\Phi(q)| \leq (1+\gamma\delta) \sum_{q\in\Gamma} h\big|_q \, |\det(D\Phi(x_q))||q|, \tag{180}$$

which yield

$$(1-\gamma\delta)\sum_{q\in\Gamma} G_q|q| \leq \sum_{q\in\Gamma} g\big|_q |\Phi(q)|, \qquad \sum_{q\in\Gamma} h\big|_q |\Phi(q)| \leq (1+\gamma\delta)\sum_{q\in\Gamma} H_q|q|, \tag{181}$$

where

$$G_q = \inf_q \left(g\big|_q \, |\det(D\Phi)| \right), \qquad \text{and} \qquad H_q = \sup_q \left(h\big|_q \, |\det(D\Phi)| \right). \tag{182}$$

Then for

$$G = \sum_{q\in\Gamma} G_q \chi_q, \qquad \text{and} \qquad H = \sum_{q\in\Gamma} H_q \chi_q, \tag{183}$$

we have $G \leq (f\circ\Phi)\,|\det(D\Phi)| \leq H$, and

$$\int_Q (f\circ\Phi)\,|\det(D\Phi)| = \lim_{\varepsilon\to 0}\int_Q G = \lim_{\varepsilon\to 0}\int_Q H. \tag{184}$$

Since we have

$$(1-\gamma\delta)\int_Q G \leq \int_{\Phi(Q)} \tilde{g} \leq \int_{\Phi(Q)} \tilde{h} \leq (1+\gamma\delta)\int_Q H, \tag{185}$$

and $\delta > 0$ can be made arbitrarily small while keeping γ bounded, we infer

$$\lim_{\varepsilon\to 0}\int_{\Phi(Q)} \tilde{g} = \lim_{\varepsilon\to 0}\int_{\Phi(Q)} \tilde{h} = \int_Q (f\circ\Phi)\,|\det(D\Phi)|, \tag{186}$$

which completes the proof. □

Appendix A. Antidifferentiation in elementary terms

Consider the problem of finding F satisfying

$$F' = f, \tag{187}$$

where f is a given function. Given f, finding F is called *antidifferentiation*, and F is called an *antiderivative of* f.

Remark A.1. Suppose that f is a function defined on (a,b), and let $F' = G' = f$ on (a,b), that is, let F and G be antiderivatives of f. Then $(F-G)' = F' - G' = 0$ on (a,b), and invoking the mean value theorem we infer that

$$F(x) = G(x) + C, \qquad x \in (a,b), \tag{188}$$

for some constant $C \in \mathbb{R}$. On the other hand, if $G' = f$ on (a,b), and if $C \in \mathbb{R}$, then a new function F defined by (188) is also an antiderivetive of f, because

$$F'(x) = (G(x)+C)' = G'(x) + 0 = f(x), \qquad x \in (a,b). \tag{189}$$

What this means is that the antiderivative of a given function can only be found *up to an additive constant*, and that if we know one antiderivative of a given function, all other antiderivatives are found by adding an arbitrary constant to it.

Definition A.2. Let G be an antiderivative of g on some interval (a, b), i.e., let $G'(x) = g(x)$ for $x \in (a, b)$. Then the set of all antiderivatives of g is denoted by

$$\int g(x)\mathrm{d}x = \{G + C : C \in \mathbb{R}\}, \tag{190}$$

which is called the *indefinite integral of* g. Alternatively and more informally, it is a standard practice to think of the indefinite integral as a notation for infinitely many functions (one function for each value of $C \in \mathbb{R}$), and write

$$\int g(x)\mathrm{d}x = G(x) + C, \tag{191}$$

where $C \in \mathbb{R}$ is considered to be an "arbitrary constant."

Example A.3. (a) We have $G' = 0$ for the zero function $G(x) = 0$, i.e., $G \equiv 0$ is an antiderivative of $g \equiv 0$. Hence we can write

$$\int 0\,\mathrm{d}x = 0 + C = C. \tag{192}$$

(b) More generally, for $\alpha \in \mathbb{R}$ we have $(x^\alpha)' = \alpha x^{\alpha-1}$ at each $x > 0$, i.e., $G(x) = \frac{1}{\alpha}x^\alpha$ is an antiderivative of $g(x) = x^{\alpha-1}$ on the interval $(0, \infty)$, for each $\alpha \in \mathbb{R} \setminus \{0\}$. Hence we have

$$\int x^{\alpha-1}\mathrm{d}x = \frac{x^\alpha}{\alpha} + C, \tag{193}$$

for $\alpha \in \mathbb{R} \setminus \{0\}$. Note that if $\alpha \in \mathbb{N}$, the relation $G'(x) = g(x)$ is true for $x \in \mathbb{R}$, and if $\alpha \in \{-1, -2, -3, \ldots\}$, it is true for $x \in \mathbb{R} \setminus \{0\}$.

(c) As for the case $\alpha = 0$, we recall $(\log x)' = \frac{1}{x}$ for $x > 0$, which leads to

$$\int \frac{\mathrm{d}x}{x} = \log x + C. \tag{194}$$

(d) Since $(e^x)' = e^x$ for $x \in \mathbb{R}$, we have

$$\int e^x\mathrm{d}x = e^x + C. \tag{195}$$

(e) From $(\sin x)' = \cos x$ and $(\cos x)' = -\sin x$ for $x \in \mathbb{R}$, we infer

$$\int \cos x\,\mathrm{d}x = \sin x + C, \qquad \text{and} \qquad \int \sin x\,\mathrm{d}x = -\cos x + C. \tag{196}$$

(f) Similarly, we have

$$\int \frac{\mathrm{d}x}{1 + x^2} = \arctan x + C, \qquad \text{and} \qquad \int \frac{\mathrm{d}x}{\sqrt{1 - x^2}} = \arcsin x + C, \tag{197}$$

where the former is valid for $x \in \mathbb{R}$, and the latter is for $x \in (-1, 1)$.

Exercise A.4. By a direct guess, find the indefinite integrals of the following functions.

(a) $g(x) = (\cos x)^{-2}$.
(b) $g(x) = 2\sin x \cos x$.
(c) $g(x) = e^{2x}$.
(d) $g(x) = 2^x$.

Example A.5. (a) Since $(2x^3 + e^x)' = 2(x^3)' + (e^x)' = 6x^2 + e^x$, we have

$$\int (6x^2 + e^x)\mathrm{d}x = 2x^3 + e^x + C. \tag{198}$$

(b) Let $\alpha \in \mathbb{R}$ be a constant. Then we have $(\sin(\alpha x))' = \alpha \cos(\alpha x)$, and hence

$$\int \cos(\alpha x)\mathrm{d}x = \frac{\sin(\alpha x)}{\alpha} + C \qquad \text{for} \quad \alpha \neq 0. \tag{199}$$

(c) We have $(\log\log x)' = \frac{1}{\log x} \cdot \frac{1}{x}$, which means that

$$\int \frac{\mathrm{d}x}{x \log x} = \log\log x + C. \tag{200}$$

Remark A.6. Starting with the power functions x^a ($a \in \mathbb{R}$), the exponential, logarithm, trigonometric and inverse trigonometric functions, and by combining them by using finitely many addition, subtraction, multiplication, quotient, and composition operations, we can generate a wide variety of functions. Let us call these functions *elementary functions*. Then the derivative of an elementary function is an elementary function, because we have differentiation rules that tell us how to compute $(f+g)'$, $(fg)'$, $(f\circ g)'$, etc., based on the knowledge of f' and g'. Each differentiation rule can be applied "in reverse" to compute antiderivatives of a large number of elementary functions. However, these "antidifferentiation rules" cannot give antiderivatives of all elementary functions, because as discovered by Joseph Liouville around 1840, there are elementary functions whose antiderivatives are not elementary. As a reflection, for example, there is no useful formula that gives an antiderivative of fg, based on antiderivatives of f and g. This makes antidifferentiation of elementary functions somewhat of a challenge, as opposed to differentiation, which is completely straightforward. Nevertheless, there exist algorithms, such as the Risch algorithm, that can decide whether an elementary function is the derivative of an elementary function, and if so, compute the antiderivative. More generally, by using the *Riemann integral*, we can construct an antiderivative of, say, any continuous function as the limit of a sequence of functions, and thus demonstrate that continuous functions admit antiderivatives.

In this appendix, we will develop a few useful antidifferentiation rules.

Lemma A.7. *a) Let $F' = f$ on (a,b), and let $\alpha \in \mathbb{R}\setminus\{0\}$ be a nonzero constant. Then we have $(\alpha F)' = \alpha f$ on (a,b), that is,*

$$\int \alpha f(x)\mathrm{d}x = \alpha \int f(x)\mathrm{d}x. \tag{201}$$

b) Let $F' = f$ on (a,b), and let $\alpha \in \mathbb{R}$. Then we have

$$(F(\alpha x))' = \alpha F'(\alpha x) = \alpha f(\alpha x) \qquad \textit{for} \quad x \in (a,b). \tag{202}$$

In other words, for $\alpha \neq 0$ we have

$$\int f(\alpha x)\mathrm{d}x = \frac{1}{\alpha}F(x) + C = \frac{1}{\alpha}\int f(x)\mathrm{d}x. \tag{203}$$

c) Let $F' = f$ and $G' = g$ on (a,b). Then we have $(F+G)' = f+g$ on (a,b), that is,

$$\int (f(x)+g(x))\mathrm{d}x = \int f(x)\mathrm{d}x + \int g(x)\mathrm{d}x. \tag{204}$$

Exercise A.8. Give a detailed proof of the preceding lemma. Is (201) true for $\alpha = 0$?

The chain rule of differentiation leads to the following rule for antidifferentiation.

Theorem A.9 (Substitution)**.** *Let $F' = f$ on (a,b), and let $\phi : (c,d) \to (a,b)$ be a differentiable function. Then $F\circ\phi$ is an antiderivative of $(f\circ\phi)\phi'$ on (c,d), that is, we have*

$$\int f(\phi(x))\phi'(x)\mathrm{d}x = F(\phi(x)) + C = \Big(\int f(y)\mathrm{d}y\Big)\Big|_{y=\phi(x)}. \tag{205}$$

Proof. Taking into account that $F' = f$, we get

$$\frac{\mathrm{d}}{\mathrm{d}x}F(\phi(x)) = F'(\phi(x))\phi'(x) = f(\phi(x))\phi'(x) \qquad \text{for} \quad x \in (c,d), \tag{206}$$

which shows that $F \circ \phi$ is an antiderivative of $(f \circ \phi)\phi'$ on (c,d). □

Recognizing if a given integral is amenable to substitution is the same as deciding if the expression under the integral can be written in the form $f(\phi(x))\phi'(x)$.

Example A.10. We have

$$\begin{aligned}\int \frac{x}{\sqrt{1+x^2}}\,\mathrm{d}x &= \frac{1}{2}\int \frac{(x^2)'}{\sqrt{1+x^2}}\,\mathrm{d}x = \frac{1}{2}\int \frac{(1+x^2)'}{\sqrt{1+x^2}}\,\mathrm{d}x = \frac{1}{2}\Big(\int \frac{1}{\sqrt{y}}\,\mathrm{d}y\Big)\Big|_{y=1+x^2} \\ &= \frac{1}{2}\Big(\int y^{-\frac{1}{2}}\mathrm{d}y\Big)\Big|_{y=1+x^2} = y^{\frac{1}{2}}\Big|_{y=1+x^2} + C = \sqrt{1+x^2} + C.\end{aligned} \tag{207}$$

Exercise A.11. Compute the following indefinite integrals.

$$\text{(a)} \quad \int \cos^2 x \sin x\,\mathrm{d}x \qquad \text{(b)} \quad \int (8x+2)e^{2x^2+x}\,\mathrm{d}x \qquad \text{(c)} \quad \int \frac{\sin\log x}{x}\,\mathrm{d}x$$

There is no "product rule" for antidifferentiation, and the following statement is basically the best we can do, in the sense that it is the most useful antidifferentiation rule that can be derived from the product rule of differentiation. In practical terms, this rule allows us to replace fg' by $f'g$ under integration.

Theorem A.12 (Integration by parts)**.** *Let f and g be functions differentiable on (a,b), and let $F' = f'g$ on (a,b). Then $fg - F$ is an antiderivative of fg' on (a,b), that is, we have*

$$\int f(x)g'(x)\mathrm{d}x = f(x)g(x) - \int f'(x)g(x)\mathrm{d}x. \tag{208}$$

Proof. By a direct computation, we infer

$$(fg - F)' = f'g + fg' - F' = f'g + fg' - f'g = fg', \tag{209}$$

which shows that $fg - F$ is an antiderivative of fg' on (a,b). □

Example A.13. We have

$$\begin{aligned}\int \log x\,\mathrm{d}x &= \int \log x \cdot (x)'\,\mathrm{d}x = x\log x - \int (\log x)' \cdot x\,\mathrm{d}x = x\log x - \int \frac{1}{x}\cdot x\,\mathrm{d}x \\ &= x\log x - \int \mathrm{d}x = x\log x - x + C.\end{aligned} \tag{210}$$

Exercise A.14. Compute the following indefinite integrals.

$$\text{(a)} \quad \int xe^x\mathrm{d}x \qquad \text{(b)} \quad \int x^2\cos x\,\mathrm{d}x \qquad \text{(c)} \quad \int e^x \sin x\,\mathrm{d}x$$