1
Introduction

1.1 Historical remarks

The history of the integral is both long and interesting. A monograph could easily be devoted to it. Here we make only a few remarks in order to set our topic into a proper historical perspective and refer the interested reader to several excellent books; see Hawkins [13], Medvedev [33], Pesin [36] and van Dalen and Monna [7] for instance. The roots of integration can be traced to Archimedes but the real story of integration starts with Newton and Leibniz. Even today, if $F : [a,b] \mapsto \mathbb{R}$ and $F'(x) = f(x)$ for every $x \in [a,b]$ we say that $F(b) - F(a)$ is the definite *Newton's integral* of f from a to b, in symbols

$$F(b) - F(a) = \mathcal{N}\int_a^b f$$

or briefly

$$F(b) - F(a) = \int_a^b f.$$

We also refer to the function F as the Newton indefinite integral of f. The Newton definition today looks much more solid than the Leibniz definition of an integral as a sum of infinitely many infinitesimal quantities. This is because the concept of derivative is firmly entrenched in our mind as a solidly defined mathematical entity. In Newton's time, however, the concepts of limit and derivative were somewhat nebulous. Despite the logical shortcomings of the beginning of calculus the early masters of calculus, e.g. the Bernoulli brothers and Euler, were able to make wonderful discoveries with the new-found tool. Of all the various definitions that would survive a modern critical scrutiny, by far the simplest and most intuitive is that which was given at the beginning of the modern era by Cauchy (1789–1857) and completed and fully investigated by Riemann (1826–1866). In fact, it is the Riemann theory that is still today taught at universities to physicists, engineers and others who need to know integration. A brief account of some finer points of Riemann integration is given in Sections 1.3—1.5. This we do because the main topic of this book is indebted to Riemann, and we wanted to give the reader an opportunity to compare results in Riemann integration with the theory expounded in this book. However, Sections 1.3—1.5 require some mathematical maturity and are not intended for a student's first reading of the book and are typeset in a smaller font. Apart from Section 1.2 containing notation, the rest of the book is independent of Chapter 1.

Among non-specialists there is an almost universal identification of the integral with the Riemann integral and this is surprising for two reasons. Firstly the Riemann integral, despite its wide use and its intuitive appeal, has serious shortcomings, as we shall see later. Secondly over eighty years ago Lebesgue (1875–1941) gave another definition of what is now known as the Lebesgue integral. This integral turns out to be the correct one for almost all uses and is the one used almost exclusively by professional mathematicians. In 1914 O. Perron proposed yet another definition, which had an additional advantage over the Lebesgue definition: it included the Newton integral and all improper integrals as well. All indications are that the Lebesgue (or Perron) theory is not popular with non-mathematicians, the reason most likely being the level of mathematical sophistication required for understanding it. In 1957 Kurzweil [20], in connection with research in differential equations, gave an elementary definition of the integral equivalent to the Perron one. For Kurzweil's own presentation of the theory see [20]. Henstock later [14] independently rediscovered Kurzweil's approach and advanced it further [15, 16, 18, 17]. The great advantages of the Kurzweil–Henstock theory are that it preserves the intuitive geometrical background of the Riemann theory, it is so simple that it can be presented in introductory courses, and it has the power of the Lebesgue theory. A further essential contribution was made by McShane. He recaptured Lebesgue integration in the Kurzweil–Henstock framework and by doing so made it accessible to non-specialists (see [31], [32]). In the second chapter we strive for the most elementary presentation of the Kurzweil–Henstock theory, suitable as an introductory course replacing the usual one on Riemann integration. The third chapter could serve as a first course on the theory of the integral. The rest of the book is devoted to more advanced topics and surrounding ideas.

1.2 Notation and the Riemann definition

The sets of integers, positive integers, rationals, reals and positive reals are denoted by $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{R}_+$, respectively. The positive or negative part of a real number a will be denoted by a^+ or a^-, respectively; i.e. $a^+ = (|a| + a)/2$, $a^- = (|a| - a)/2$. Unless something is specified to the contrary the word function means a real valued function. For a real valued function f then the meaning of f^+ and f^- is clear. Generally speaking, operations with functions are understood pointwise, for instance $f + g : x \mapsto f(x) + g(x)$, $\mathrm{Max}(f, g) : x \mapsto \mathrm{Max}(f(x), g(x))$ etc. Similarly with relations, $f \le g$ means $f(x) \le g(x)$ for every x from the common domain of definition of f and g. Likewise the inequality $f \le K$ means $f(x) \le K$ on the domain of f. The inverse function† to f is denoted by f_{-1}. If S is a set then $\mathbf{1}_S$ will denote the characteristic function of S, i.e. $\mathbf{1}_S(x) = 1$ for $x \in S$ and $\mathbf{1}_S(x) = 0$ for $x \notin S$. The sequence $n \mapsto c_n; n \in \mathbb{N}$ will be abbreviated to $\{c_n\}$, with a similar convention for sequences of functions. We use the term increasing (decreasing) in the wider sense, i.e. an increasing function might take the same value twice; an increasing (decreasing) function which is one-to-one will be called strictly increasing (decreasing). We shall use the symbol $\sup\{f;\ M\}$ for the supremum‡ of a function f over a set M and employ a similar notation for the infimum§. Sometimes we might write a defining relation for the set M instead of M itself, for instance $\sup\{a_n;\ n \ge N\}$ denotes $\sup\{a_N, a_{N+1}, \dots\}$. An interval $[a, b]$ will always be closed and (a, b) open. We shall use $|I|$ for the length of a bounded interval¶ I. Important for our further development are the concepts of a *division of an interval,* and that of a *partition of an interval.* By a division D of a compact interval $[a, b]$ we mean a set of intervals $[x_i,\ x_{i+1}]$ such that

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b. \tag{1.1}$$

The points x_i are called the points of the division D. A function

$$\varphi : [a,\ b] \to \mathbb{R}$$

† Many authors use the notation f^{-1}; we reject that since f^{-1} could also legitimately denote $1/f$.
‡ The supremum of a set is its least upper bound.
§ greatest lower bound
¶ We accept as intervals also the sets $[a,\ a]$, consisting of one point, and $(a,\ a) = \emptyset$ the empty set. For these so-called degenerate intervals I the length is zero, $|I| = 0$.

is called *a step function* if there is a division (1.1) such that φ is constant on every interval (x_{i-1}, x_i). A partition of a compact interval $[a, b]$ is a set of couples (ξ_k, I_k) such that the points $\xi_k \in I_k$, the closed intervals I_k are non-overlapping† and

$$\bigcup_1^n I_k = [a, b]. \tag{1.2}$$

We shall call the point ξ_k the *tag* of I_k. Often it will be convenient to have the intervals, $I_k = [u_k, v_k]$, ordered; hence for a partition

$$\pi \equiv \{(\xi_k, [u_k, v_k]); k = 1, 2, \ldots, n\} \tag{1.3}$$

we have

$$a = u_1 \le \xi_1 \le v_1 = u_2 \le \xi_2 \le v_2 \le \cdots \le v_n = b \tag{1.4}$$

The letters π and Π (possibly with subscripts) will denote partitions. A partition

$$\{(\xi_i, [u_i, v_i]); i = 1, 2, \ldots, n\}$$

can be abbreviated to $\{(\xi_i, [u_i, v_i])\}$ or even to $\{(\xi, [u, v])\}$ if the range of subscripts i is clear from the context or is not particularly important. If $\delta > 0$ then a partition π for which

$$\xi_i - \delta < u_i \le \xi_i \le v_i < \xi_i + \delta \tag{1.5}$$

for all i with $1 \le i \le n$ is called a δ-fine partition of $[a, b]$. It is obvious that a partition π is δ-fine if and only if the length of the largest interval of π, which we denote by $n(\pi)$, is less than 2δ. Similarly as with $n(\pi)$ we denote by $n(D)$ the length of the largest interval of the division D.

Given a function $f : [a, b] \to \mathbb{R}$ then a partition (1.4) has an associated Riemann sum‡

$$\sum_\pi f = \sum_{i=1}^n f(\xi_i)(v_i - u_i), \tag{1.6}$$

which we shall also abbreviate as $\sum_\pi f(\xi)(v - u)$. If the partition σ is given by (y_k, J_k) with $k = 1, 2, \ldots, m$, then, for the Riemann sum, we shall naturally use the notation

$$\sum_\sigma f = \sum_{k=1}^m f(y_k)|J_k| = \sum_\sigma f(y)|J|. \tag{1.7}$$

† i.e. they do not have any interior points in common.
‡ See Figure 1.1.

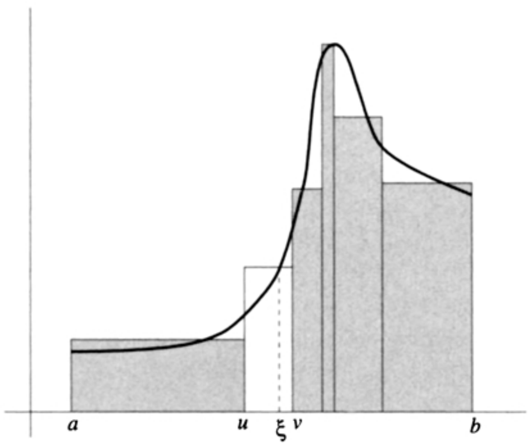

Fig. 1.1. Riemann sum

We extend our shorthand to similar sums, e.g. we shall denote by $\sum_\pi f(u, v)$ the sum $\sum_{i=1}^{n}[f(v_i) - f(u_i)]$.

The Riemann integral, $\int_a^b f$, is defined as the limit of Riemann sums. More precisely we define:

DEFINITION 1.2.1 *A number A is the Riemann integral of f from a to b (or on $[a, b]$) if for every positive ε there is a positive number δ such that for every δ-fine partition π*

$$|\sum_\pi f - A| < \varepsilon. \tag{1.8}$$

We denote the Riemann integral A as usual by $\int_a^b f$ or by $\int_a^b f(x)dx$. If we wish to distinguish the integral from another integral, e.g. the Newton integral, or if we wish to emphasize that the integral is to be understood in the sense of Definition 1.2.1, then we write $\mathcal{R}\int_a^b f$. We shall often abbreviate Riemann integral to R-integral and if there is no danger of confusion just to integral. The function f is called Riemann integrable, or briefly R-integrable, if the Riemann integral of f exists.

It is a consequence of Definition 1.2.1 that an R-integrable function f must be bounded. We choose $\varepsilon = 1$ and find a corresponding δ from Definition 1.2.1. We divide the interval into n equal intervals $[u_i, v_i]$ with $n^{-1}(b-a) < \delta$ and choose a number $C > \text{Max}(|f(u_j)|, j = 1, 2, \ldots, n)$. Let $x \in [a, b]$ be arbitrary; it lies in some $[u_\alpha, v_\alpha]$ and inequality (1.8) for the partition (1.3) with $\xi_j = u_j$ $(j \neq \alpha)$ and $\xi_\alpha = x$ yields

$$|f(x)(v_\alpha - u_\alpha)| < |A| + 1 + \sum_{i=1, i\neq\alpha}^{n} |f(u_i)|(v_i - u_i),$$

and consequently

$$|f(x)| < \frac{n[|A| + 1 + C(b-a)]}{b-a}.$$

Since x was arbitrary this shows f is bounded. Unless something is specified to the contrary we shall assume for the rest of this chapter that all functions appearing are bounded.†

1.3 Basic theorems, upper and lower integrals

For a division D as given in (1.1) we introduce the Darboux upper and lower sums $S(D)$ and $s(D)$ defined by

$$S(D) = \sum_{i=0}^{n-1} M_i(x_{i+1} - x_i), \tag{1.9}$$

$$s(D) = \sum_{i=0}^{n-1} m_i(x_{i+1} - x_i), \tag{1.10}$$

where $M_i = \sup\{f, [x_i, x_{i+1}]\}$ and $m_i = \inf\{f, [x_i, x_{i+1}]\}$. Obviously there is a division D_π naturally associated with a partition π; in the spirit of our shorthand writing we shall denote the the upper Darboux sum by $S(\pi)$ or by $S(D_\pi)$, in symbols

$$S(\pi) = S(D_\pi) = \sum_\pi M(v - u) = \sum_{i=1}^{n} M_i(v_i - u_i),$$

where now M_i stands for the supremum of f on $[u_i, v_i]$. Of course, we use the same convention for lower Darboux sums and other similar structured sums. A division $\tilde{D}$ is a *refinement* of D if all the points of D are also points of $\tilde{D}$. Adding points to a division increases the lower sums and decreases the upper sums,

$$S(D) \geq S(\tilde{D}) \text{ and } s(D) \leq s(\tilde{D}). \tag{1.11}$$

We prove only the first of these inequalities and it is clearly sufficient to prove it if $\tilde{D}$ has only one additional point c. (In the general case we can move from D to $\tilde{D}$ by adding points one by one.) If $c \in (x_i, x_{i+1})$ then the contribution of the intervals $[x_i, c]$ and $[c, x_{i+1}]$ to $S(\tilde{D})$ is

$$\sup\{f, [x_i, c]\}(c - x_i) + \sup\{f, [c, x_{i+1}]\}(x_{i+1} - c) \leq M_i(x_{i+1} - x_i),$$

† The reader not interested in the Riemann theory can now start reading Chapter 2.

and the inequality $S(D) \geq S(\tilde{D})$ follows. By comparing an upper sum $S(D_1)$ and a lower sum $s(D_2)$ with their common refinement we obtain

$$S(D_1) \geq s(D_2). \tag{1.12}$$

It is convenient to define the upper and lower Riemann integrals of f on $[a, b]$, in symbols $\overline{\int_a^b} f$ and $\underline{\int_a^b} f$, by

$$\overline{\int_a^b} f = \inf\{S(D) : D \text{ a division of } [a, b]\}, \tag{1.13}$$
$$\underline{\int_a^b} f = \sup\{s(D) : D \text{ a division of } [a, b]\}. \tag{1.14}$$

In view of relation (1.12) the upper and lower integrals always exist† and

$$\overline{\int_a^b} f \geq \underline{\int_a^b} f.$$

It follows easily that

$$\overline{\int_a^b} (-f) = -\underline{\int_a^b} f.$$

The connections between the upper, lower and Riemann integral are stated in the following theorem.

THEOREM 1.3.1 *The following statements are equivalent:*

(i) *The function f is Riemann integrable on $[a, b]$.*
(ii) *For every positive ε there is a positive δ such that*

$$S(D) - s(D) < \varepsilon, \tag{1.15}$$

whenever $n(D) < \delta$.
(iii) *For every positive ε there is a division D such that inequality* (1.15) *holds.*
(iv) *The upper and lower integrals of f are equal.*

Proof The implications (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) are fairly obvious. Since every Riemann sum $\sum_\pi f$ lies between the corresponding Darboux sums $S(D_\pi)$ and $s(D_\pi)$ it is clear that

$$\sum_\pi f - \overline{\int_a^b} f > s(D_\pi) - S(D_\pi),$$
$$\overline{\int_a^b} f - \sum_\pi f < S(D_\pi) - s(D_\pi).$$

† According to the convention adopted f is bounded.

This shows (ii) $\Rightarrow$ (i) with $A = \overline{\int_a^b} f$. To prove (i)$\Rightarrow$ (ii) we find a positive δ such that for every partition π with $n(\pi) < \delta$

$$|\sum_\pi f(\xi)(v-u) - A| < \frac{\varepsilon}{4}.$$

On each interval $[u, v]$ we choose ξ so that

$$m + \frac{\varepsilon}{4(b-a)} > f(\xi).$$

It follows that $s(D_\pi) > A - \varepsilon/2$. Similarly $S(D_\pi) < A + \varepsilon/2$ and inequality (1.15) follows. To complete the proof it suffices to show (iv) $\Rightarrow$ (ii). This implication becomes obvious with the following lemma. •

LEMMA 1.3.2 *For every positive ε there exists a positive δ such that*

$$S(D) - \overline{\int_a^b} f \quad < \varepsilon, \tag{1.16}$$

$$\underline{\int_a^b} f - s(D) \quad < \varepsilon, \tag{1.17}$$

whenever

$$n(D) < \delta. \tag{1.18}$$

Proof Only (1.16) needs proof since (1.17) follows from (1.16) applied to $-f$. For the proof of (1.16) we denote by A the value of the upper integral and find a division D_ε such that

$$S(D_\varepsilon) < A + \frac{\varepsilon}{2}. \tag{1.19}$$

Let N be the number of dividing points of D_ε and $|f| \le K$. We show that the number $\varepsilon/4NK$ serves as the required δ. Let D be any division satisfying condition (1.18) and $\bar{D}$ a common refinement of D_ε and D. By inequalities (1.11) and (1.19)

$$S(\bar{D}) < A + \frac{\varepsilon}{2}. \tag{1.20}$$

We now estimate $S(D) - S(\bar{D})$. Every interval $[u, v]$ which is common to D and $\bar{D}$ makes the same contribution to both $S(D)$ and $S(\bar{D})$. An interval $[y, z]$ of D which is not an interval of $\bar{D}$ contains at least one point of D_ε. Hence there are at most N intervals $[y, z]$. The part of difference $S(D) - S(\bar{D})$ restricted to $[y, z]$ is at most $[M - (-M)](y - z)$ and that does not exceed $2M\delta$. Consequently $S(D) - S(\bar{D}) \le N.2M\delta = \varepsilon/2$. This together with (1.20) establishes (1.16). •

Every upper Darboux sum $S(D)$ defines a step function φ_D such that $S(D) = \int_a^b \varphi_D$ and $\varphi_D \ge f$. This step function can be modified into a continuous piecewise linear function H with a trapezoidal graph such that $\varphi_D \le H$ and $\int_a^b H - S(D) < \varepsilon$. See Figure 1.2. The next lemma easily follows.

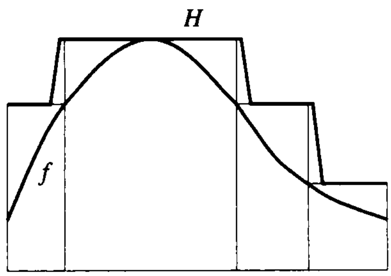

Fig. 1.2. Trapezoidal approximation from above

LEMMA 1.3.3 *For every positive* ε *there exist continuous functions* h, H *such that* $h \le f \le H$ *and*

$$\underline{\int_a^b} f \quad < \quad \int_a^b h + \varepsilon, \tag{1.21}$$

$$\overline{\int_a^b} f \quad > \quad \int_a^b H - \varepsilon. \tag{1.22}$$

The basic properties of the Riemann integral follow easily from Definition 1.2.1 and Theorem 1.3.1. For ease of reference we state them here.

Homogeneity If $c > 0$ then

$$\overline{\int_a^b} cf = c \overline{\int_a^b} f. \tag{1.23}$$

The same equation holds for the lower integral and if f is integrable then it is valid for any c and for the R-integral.

Preservation of inequalities If $f \le g$ then

$$\overline{\int_a^b} f \le \overline{\int_a^b} g. \tag{1.24}$$

In particular, if $m \le f(x) \le M$ for all $x \in [a, b]$ then

$$m(b-a) \le \overline{\int_a^b} f \le M(b-a). \tag{1.25}$$

Consequently, if $|f| \le K$ then

$$\left| \overline{\int_a^b} f \right| \le K(b-a). \tag{1.26}$$

All these inequalities hold with the upper integral replaced by the lower integral and for an R-integrable f with $\int_a^b$ substituted for $\overline{\int_a^b}$.

Absolute value For any bounded f

$$\left|\underline{\int}_a^b f\right| \leq \underline{\int}_a^b |f|, \tag{1.27}$$

with a similar inequality holding for the upper integral. If f is integrable then so is $|f|$ and the above inequality holds with the lower integral replaced by the integral.

Integral as an additive function of intervals For any bounded f we have

$$\overline{\int}_a^b f = \overline{\int}_a^c f + \overline{\int}_c^b f, \tag{1.28}$$

with the same relation holding for the lower integrals. It follows that a function f is integrable on $[a,b]$ if and only if it is integrable on $[a,c]$ and $[c,b]$ with $a < c < b$ and then equation (1.28) holds with $\overline{\int}_a^b$ replaced by $\int_a^b$.

Additivity The upper integral is subadditive, the lower integral superadditive which means

$$\overline{\int}_a^b f + \overline{\int}_a^b g \quad \geq \quad \overline{\int}_a^b (f+g), \tag{1.29}$$

$$\underline{\int}_a^b f + \underline{\int}_a^b g \quad \leq \quad \underline{\int}_a^b (f+g). \tag{1.30}$$

It follows that if f and g are integrable then so is $f+g$ and

$$\int_a^b (f+g) = \int_a^b f + \int_a^b g. \tag{1.31}$$

1.4 Differentiability, continuity and integrability

We are assuming that f is bounded; let $|f| \leq K$. The functions

$$F(x) = \int_a^x f, \tag{1.32}$$

$$U(x) = \overline{\int}_a^x f,$$

$$L(x) = \underline{\int}_a^x f, \tag{1.33}$$

which we shall call the indefinite integral, the indefinite upper integral and the indefinite lower integral, will play an important rôle in this section. By inequality (1.26) and similar inequalities for the integral and for the lower integral, and by using equation (1.28) for the additivity of the (upper, lower) integral, we see that F, U, L are *Lipschitz continuous*† *with the constant* K*;*

† A function F is said to be Lipschitz continuous on the interval I with the constant L if $|F(x) - F(y)| \leq L|x-y|$ for x, y in I. Instead of Lipschitz continuous one often says just Lipschitz or merely L.