# Generalized Riemann Integral

Karel Hrbacek
The City College of New York, New York
khrbacek@sci.ccny.cuny.edu

July 27, 2014

These notes present the theory of generalized Riemann integral, due to R. Henstock and J. Kurzweil, from a nonstandard point of view. The key notion we use, that of $a$-ultrasmall numbers, is due to B. Benninghofen and M. M. Richter, *A general theory of superinfinitesimals*, Fund. Math. 123: 199–215, 1987, who call them "superinfinitesimals." E. Gordon, *Nonstandard Methods in Commutative Harmonic Analysis*, Amer. Math. Society, 1997, developed an approach to relative standardness that is different from that of Y. Péraire; in particular, his relative infinitesimals are the superinfinitesimals. Here we have combined the two techniques. B. Benninghofen presented an approach to the generalized Riemann integral using superinfinitesimals in *Superinfinitesimals and the calculus of the generalized Riemann integral*, in *Models and Sets*, G. H. Müller and M. M. Richter, eds., Lecture Notes in Math. 1103, Springer, Berlin, 1984, pp. 9 - 52. Our development of the generalized Riemann integral follows the excellent exposition in R. Bartle, *A Modern Theory of Integration*, Amer. Math. Society, 2001, to which the reader is referred for further study of this topic.

AUN refers to K. Hrbacek, O. Lessmann and R. O'Donovan, "*Analysis with Ultrasmall Numbers.*"

## 1 $a$-Ultrasmall Numbers

The fundamental problem of calculus is to determine the function $f$ from its derivative $f'$. In AUN, Chapter 4 we solved this problem under the assumption that $f'$ is continuous; that is, we learned to integrate continuous functions. In AUN, Chapter 9 the theory of integration is developed for a larger class of functions, those that are Riemann integrable. Among Riemann integrable functions there are some that are not continuous, yet it turns out that the Riemann integrable functions are precisely those functions that are continuous "almost everywhere" (see Theorem 31). Because of this and other reasons, Riemann theory is not sufficiently general for many applications in analysis.

In order to see how to go about formulating a general theory of integration, let us revisit the procedure used in AUN, Section 4.1 to recover the original function $f$ from its derivative. The assumption of continuity of $f'$ allowed us to use the uniform version of the increment equation, with $dx$ ultrasmall relative to $f$, independent of $x_i$. This in turn motivated the definition of Riemann integral in terms of fine partitions, that is, partitions where each $dx_i$ is ultrasmall relative to $f$, independent of the tag $t_i$. Without the assumption of continuity of $f'$, the increment equation requires $dx_i$ to be ultrasmall relative to $f$ and $x_i$. This suggests the following definition:

A tagged partition $(\mathcal{P}, \mathcal{T})$ is **superfine** (relative to the level of $f$) if each $dx_i$ is ultrasmall relative to $f$ and $t_i$.

It is an easy exercise to verify that most of the arguments in AUN, Chapter 9 would go through if one would replace "fine partitions" by "superfine partitions" in the definition of the Riemann integral. The class of integrable functions would become much larger and, in particular, the fundamental theorem of calculus would hold for all differentiable functions.

Unfortunately, in our framework for relative analysis it is not possible to prove that any partitions superfine in this strong sense exist. (See KH, *Relative set theory: Some external issues*, Journ. Logic and Analysis 2:8, 2010, 1–37.) The strong version of stability, as well as of other principles, postulated in our theory requires numbers ultrasmall relative to a given context to be too small to make up a superfine partition. There is a coarser notion of ultrasmall numbers that does not have such nice uniform properties, and therefore is not as suitable for development of analysis in general, but is tailor-made for the special purpose of generalizing the theory of integration. We develop this notion in the rest of the current section and return to generalized Riemann integral in Sections 2 and 3.

Several arguments in these notes use the Principle of Idealization discussed in the Appendix to AUN. The readers should take a look at this material before proceeding further, or as needed.

**Definition 1**
*Given a context and a real number $a$:*

*(1) A real number $r$ is $a$-**accessible** if $r = \varphi(a)$ for some observable function $\varphi : \mathbb{R} \to \mathbb{R}$.*

*(2) A real number $x$ is $a$-**ultralarge** if $|x| > r$ for all $a$-accessible $r > 0$.*

*(3) A real number $h \neq 0$ is $a$-**ultrasmall** if $|h| \leq r$ for all $a$-accessible $r > 0$.*

**Remark** $a$-accessible, $a$-ultrasmall and $a$-ultralarge are external concepts, and the conventions from AUN, Section 1.5 apply to them.

We say that a function $\varphi$ is **positive** if $\varphi(x) > 0$ for all $x$ in its domain. It follows immediately from these definitions that $x$ is not $a$-ultralarge if and only if $|x| \le \varphi(a)$ for some observable positive $\varphi : \mathbb{R} \to \mathbb{R}$, and $h \neq 0$ is $a$-ultrasmall if and only if $|h| \le \varphi(a)$ for all observable positive $\varphi : \mathbb{R} \to \mathbb{R}$.

We also note that the above statements remain true if one requires only that $\varphi$ be defined at $a$, in place of being defined on all of $\mathbb{R}$: if $\varphi$ is any observable positive function defined at $a$, then $\overline{\varphi} : \mathbb{R} \to \mathbb{R}$ defined by

$$\overline{\varphi}(x) = \begin{cases} \varphi(x) & \text{if defined,} \\ 1 & \text{otherwise,} \end{cases}$$

is an observable positive function defined for all $x$, and $\overline{\varphi}(a) = \varphi(a)$.

**Theorem 1**

*(1) If $r$ is $a$-accessible relative to $p_1, \ldots, p_k$, then $r$ is observable relative to $a, p_1, \ldots, p_k$.*

*(2) If $h$ is $a$-ultrasmall relative to $p_1, \ldots, p_k$, then $h$ is ultrasmall relative to $p_1, \ldots, p_k$.*

*(3) If $h$ is ultrasmall relative to $a, p_1, \ldots, p_k$, then $h$ is $a$-ultrasmall relative to $p_1, \ldots, p_k$.*

**Proof:**

(1) If $\varphi$ is observable relative to $p_1, \ldots, p_k$, then $\varphi(a)$ is observable relative to $a, p_1, \ldots, p_k$ by Closure.

(2) For an observable $r > 0$ let $\varphi_r$ be the constant function with value $r$, $\varphi_r : x \mapsto r$ for all $x \in \mathbb{R}$. Then $\varphi_r$ is an observable positive function. Hence by definition we have $|h| \le \varphi_r(a) = r$, which shows that $h$ is ultrasmall.

(3) follows from (1).

□

If $|h| < \varphi(a)$ for all observable positive $\varphi$, then in particular $|h| < \frac{|a|}{n}$ for any observable $n$, hence $h$ is of the form $\delta \cdot a$ for some $\delta \simeq 0$.

It is clear that if $a$ is observable relative to $p_1, \ldots, p_k$, then $h$ is $a$-ultrasmall if and only if $h$ is ultrasmall. In particular $h$ is not observable relative to $p_1, \ldots, p_k$. The important fact is that if $a$ is not observable relative to $p_1, \ldots, p_k$, then numbers $a$-ultrasmall relative to $p_1, \ldots, p_k$ can actually be observable relative to $a, p_1, \ldots, p_k$, as we prove in the following theorem.

**Theorem 2**

*For any real number $a$ not observable relative to $p_1, \ldots, p_k$ there are some numbers $a$-ultrasmall relative to $p_1, \ldots, p_k$ and observable relative to $a, p_1, \ldots, p_k$.*

**Proof:** We use the Idealization Principle. Let $\{\varphi_1, \ldots, \varphi_k\}$ be a finite set of positive functions, observable relative to $p_1, \ldots, p_k$. Then $h = \min(\varphi_1(a), \ldots, \varphi_k(a)) > 0$ and $h$ is observable relative to $a, p_1, \ldots, p_k$. In other words, there exists $h > 0$ observable relative to $a, p_1, \ldots, p_k$ such that $h \leq \varphi_i(a)$ for $i = 1, \ldots, k$. By Idealization, there exists $h > 0$ observable relative to $a, p_1, \ldots, p_k$ such that $h \leq \varphi(a)$ holds for all positive functions observable relative to $p_1, \ldots, p_k$; any such $h$ is $a$-ultrasmall relative to $p_1, \ldots, p_k$. □

**Exercise 1** (Answer page 33)
Show the following statements.

(1) If $r$ and $s$ are $a$-accessible, then $r \pm s$, $r \cdot s$ and $r/s$ (if $s \neq 0$) are $a$-accessible.

(2) If $x$ and $y$ are not $a$-ultralarge, then $x \pm y$ and $x \cdot y$ are not $a$-ultralarge.

(3) If $h, k$ are $a$-ultrasmall and $x$ is not $a$-ultralarge, then $h \pm k$ and $x \cdot h$ are $a$-ultrasmall or 0.

We need a version of the Closure Principle for $a$-accessible numbers.

**Theorem 3 ($a$-Closure Principle)**
*Given a statement $\mathcal{P}(y, a, b, p_1, \ldots, p_k)$ of traditional mathematics, and $a, b$ such that $b$ is $a$-accessible relative to a given context, where $p_1, \ldots, p_k$ are observable: If there exists a number $y$ for which the statement is true, then there exists an $a$-accessible number $y$ for which the statement is true.*

**Proof:** Let $b = \varphi(a)$, for an observable $\varphi : \mathbb{R} \to \mathbb{R}$. We consider the statement $\mathcal{P}(y, x, \varphi(x))$ with variable $x$. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a function such that, for all $x \in \mathbb{R}$, if there is some $y$ for which $\mathcal{P}(y, x, \varphi(x))$ is true, then $\psi(x)$ is one such $y$, i.e., $\mathcal{P}(\psi(x), x, \varphi(x))$ holds. We omit the detailed justification of the existence of such a function (it follows easily from the axioms of Separation, Replacement, and Choice; see the Appendix to AUN for these). By Closure, we can assume that $\psi$ is observable. Then $y = \psi(a)$ is $a$-accessible, and $\mathcal{P}(\psi(a), a, \varphi(a))$ is true.
□

It is easy to modify the statement and proof of the $a$-closure principle to allow a finite list $b_1, \ldots, b_\ell$ of parameters in place of the single parameter $b$.

We next show that numbers $a$-ultrasmall relative to $f$ can replace numbers ultrasmall relative to $f$ and $a$ in the definition of $\lim_{x\to a} f(x)$ (and hence also in the definition of continuity of $f$ at $a$ and derivative of $f$ at $a$).

**Theorem 4**
*The following statements are equivalent:*

*(1)* $\lim\limits_{x\to a} f(x) = L$

*(2) Relative to a context where $f$ is observable: $L$ is $a$-accessible and $f(a+h) - L$ is $a$-ultrasmall (or 0), for all $a$-ultrasmall $h$.*

**Proof:** We work in a context where $f$ is observable.

(1) implies (2): The limit $L$ is observable, hence $a$-accessible. Let $\varepsilon > 0$ be $a$-accessible. By the epsilon-delta definition of limit, there exists $\delta > 0$ such that

$$0 < |x - a| < \delta \text{ implies } |f(x) - L| < \varepsilon.$$

Using the $a$-Closure Principle, we can take $\delta$ to be $a$-accessible. If now $h$ is $a$-ultrasmall, we set $x = a + h$ and have $|x - a| = |h| < \delta$, hence $|f(a+h) - L| < \varepsilon$. As $\varepsilon$ is an arbitrary positive $a$-accessible number, $f(a+h) - L$ is $a$-ultrasmall.

(2) implies (1): We assume (1) is false and prove that (2) is false. So let $\lim_{x\to a} f(x) \neq L$, i.e., there exists $\varepsilon > 0$ such that

$(*)$ for every $\delta > 0$ there exists $x$ such that $0 < |x - a| < \delta$ and $|f(x) - L| \geq \varepsilon$.

By $a$-Closure Principle we can assume that $\varepsilon$ is $a$-accessible. Let $\{\varphi_1, \ldots, \varphi_k\}$ be an observable finite set of positive functions; we define $\varphi$ by

$$\varphi : x \mapsto \min\{\varphi_1(x), \ldots, \varphi_k(x)\}.$$

Notice that $\varphi$ is observable. Let $\delta = \varphi(a)$ in (*). We get that there exists $x$ such that $0 < |x - a| < \varphi_i(a)$ and $|f(x) - L| \geq \varepsilon$ is true for $i = 1, \ldots, k$. Applying Idealization, we obtain $x$ such that

$$0 < |x - a| < \varphi(a) \quad \text{and} \quad |f(x) - L| \geq \varepsilon$$

is true for all observable positive $\varphi$. Then $h = x - a$ is $a$-ultrasmall and we have $|f(a+h) - L| \geq \varepsilon$, so $f(a+h) - L$ is not $a$-ultrasmall. Hence (2) fails. □

We immediately deduce the following $a$-version of the Increment Equation.

**Theorem 5**
*Relative a context where $f$ is observable: Suppose that $f$ is differentiable at $a$. Let $dx$ be $a$-ultrasmall. Then there is $\varepsilon$ which is $a$-ultrasmall or 0, such that*

$$f(a + dx) = f(a) + f'(a) \cdot dx + \varepsilon \cdot dx.$$

We leave the corresponding straddle version of the $a$-Increment Equation as an exercise.

**Exercise 2** (**Straddle version**) (Answer page 33)
Relative a context where $f$ is observable: Suppose that $f$ is differentiable at $a$. Let $x_1 \leq a \leq x_2$ be such that $a - x_1$ and $x_2 - a$ are $a$-ultrasmall or 0. Show that there is $\varepsilon$ which is $a$-ultrasmall or 0, such that

$$f(x_2) - f(x_1) = f'(a)(x_2 - x_1) + \varepsilon \cdot (x_2 - x_1).$$

**Definition 2**
*A tagged partition $(\mathcal{P}, \mathcal{T})$ is **superfine** if each $dx_i$ is $t_i$-ultrasmall.*

Existence of superfine partitions follows from a classical lemma.

**Definition 3**
*Let $\varphi$ be a positive function defined on $[a, b]$. We say that a tagged partition $(\mathcal{P}, \mathcal{T})$ of $[a, b]$ is* ***subordinate*** *to $\varphi$ if $dx_i < \varphi(t_i)$, for all $i = 0, \ldots, n-1$.*

**Theorem 6 (Cousin's Lemma)**
*If $\varphi$ is a positive function defined on $[a, b]$, then there is a tagged partition $(\mathcal{P}, \mathcal{T})$ of $[a, b]$ subordinate to $\varphi$.*

**Proof:** We proceed by contradiction and assume that there is no tagged partition of $I_0 = [a, b]$ subordinate to $\varphi$. Let $c = (a + b)/2$ be the midpoint of the interval $I_0$. Either the interval $[a, c]$ or the interval $[c, b]$ has no partition subordinate to $\varphi$; otherwise, we could combine them and obtain a partition of $I_0$ subordinate to $\varphi$. In the first case we let $I_1 = [a, c]$; otherwise, $I_1 = [c, b]$; $I_1$ has no partition subordinate to $\varphi$ and the length of $I_1$ is $(b - a)/2$. Continuing in this manner, we construct a nested sequence of closed intervals $I_0, I_1, \ldots . I_n, \ldots$, none of which has a partition subordinate to $\varphi$, and such that the length of $I_n$ is $(b - a)/2^n$; in particular, the length of $I_n$ converges to 0. By the nested interval theorem, there is a number $c$ that belongs to every $I_n$. Let $n$ be such that $(b - a)/2^n < \varphi(c)$. Then the trivial partition of $I_n$ (that is, $x_0$ is the left endpoint of $I_n$, $x_1$ is the right endpoint of $I_n$), tagged by $t_0 = c$, is subordinate to $\varphi$, a contradiction.

□

**Theorem 7**
*For every $a < b$ there exists a superfine partition of $[a, b]$.*

**Proof:** We again use Idealization. Let $\{\varphi_1, \ldots, \varphi_k\}$ be a set of positive functions defined on $[a, b]$ and observable relative to $a$ and $b$; then

$$\varphi(x) = \min(\varphi_1(x), \ldots, \varphi_k(x))$$

is a positive function, and $\varphi \leq \varphi_i$ for all $i = 1, \ldots, k$. Applying Idealization, we obtain a positive function $\overline{\varphi}$ such that $\overline{\varphi} \leq \varphi$ for all observable positive $\varphi$. By Cousin's Lemma, there exists a tagged partition of $[a, b]$ subordinate to $\overline{\varphi}$. It is clear that this partition is superfine. □

We conclude with two technical results about superfine partitions.

Let $x_0 < x_1 < \cdots < x_n$ be a fine partition of $[a, b]$; then for each $i$ there is at most one element in $[x_i, x_{i+1}]$ which is observable relative to $a$ and $b$. We now show that for superfine partitions we necessarily take that element as a tag.

**Theorem 8**
*Let $(\mathcal{P}, \mathcal{T})$ be a superfine partition of $[a, b]$. Every observable real number $c \in [a, b]$ belongs to $\mathcal{T}$.*

**Proof:** The context is specified by $a, b$. Let $c \in [a, b]$ be observable. Let $\varphi$ be the function defined by

$$\varphi(x) = \begin{cases} |x - c| & \text{if } x \neq c; \\ 1 & \text{if } x = c. \end{cases}$$

Then $\varphi$ is positive and observable. Let $i$ be such that $c \in [x_i, x_{i+1}]$. We show that $t_i = c$. If not, then since $dx_i$ is $t_i$-ultrasmall, we must have $dx_i < \varphi(t_i) = |t_i - c|$, i.e., $|x_{i+1} - x_i| < |t_i - c|$, so $c \notin [x_i, x_{i+1}]$, a contradiction. This shows that $c \in \mathcal{T}$. □

Let $(\mathcal{P}, \mathcal{T})$ be a tagged partition. We define a new tagged partition $(\mathcal{P}^*, \mathcal{T}^*)$ as follows: whenever $x_i < t_i < x_{i+1}$, we split the interval $[x_i, x_{i+1}]$ into $[x_i, t_i]$ and $[t_i, x_{i+1}]$, and let $t_i$ be the tag for both. We note that $f(t_i)(x_{i+1} - x_i) = f(t_i)(x_{i+1} - t_i) + f(t_i)(t_i - x_i)$, so $\sum(f; \mathcal{P}, \mathcal{T}) = \sum(f; \mathcal{P}^*, \mathcal{T}^*)$. The partition $(\mathcal{P}^*, \mathcal{T}^*)$ has the property that the tag for each subinterval is either the left or the right endpoint. If $(\mathcal{P}, \mathcal{T})$ is superfine, then $(\mathcal{P}^*, \mathcal{T}^*)$ is superfine and, by the previous proposition, every $c \in [a, b]$ from the context is specified by one of the points $x_0, \ldots, x_n$ of the partition $\mathcal{P}^*$.

**Theorem 9**
*Let $a, b \in \mathbb{R}$ and a system of open intervals $\{I_k\}_{k=1}^{\infty}$ appear at the observation level. If $(\mathcal{P}, \mathcal{T})$ is a superfine partition of $[a, b]$, then for each $t_i \in \bigcup_{k=1}^{\infty} I_k$ there is some $k$ such that $[x_i, x_{i+1}] \subseteq I_k$.*

**Proof:** Let us write $I_k = (a_k, b_k)$, for some $a_k < b_k$. We define $\varphi : [a, b] \to \mathbb{R}$ by

$$\varphi(x) = \begin{cases} \min(x - a_k, b_k - x) & \text{where } k \text{ is least such that } t_i \in (a_k, b_k); \\ 1 & \text{if no such } k \text{ exists.} \end{cases}$$

Notice that $\varphi$ is well-defined. It is a positive function at the observation level.

Let $t_i \in \bigcup_{k=1}^{\infty} (a_k, b_k)$ and let $k$ be the least index such that $t_i \in (a_k, b_k)$. The partition is superfine, so using the definition of $\varphi$ we have

$$x_{i+1} - x_i = dx_i < \varphi(t_i) \leq \min(t_i - a_k, b_k - t_i).$$

It follows that $x_i, x_{i+1} \in (a_k, b_k)$, i.e., $[x_i, x_{i+1}] \subseteq I_k$. □

# 2 The generalized Riemann integral

The generalization of Riemann integral that we present here was developed independently by Ralph Henstock and Jaroslav Kurzweil; it is sometimes called Henstock-Kurzweil integral.

**Definition 4**
*A function $f$ defined on $[a, b]$ is **generalized Riemann integrable on** $[a, b]$ (or simply **integrable on** $[a, b]$) if there is an observable number $R$ such that*

$$\sum(f; \mathcal{P}, \mathcal{T}) \simeq R,$$

*for all superfine tagged partitions $(\mathcal{P}, \mathcal{T})$ of $[a, b]$.*
*In this is the case, we write*

$$\int_a^b f(x) \cdot dx = R.$$

Superfine partitions are fine, so it follows immediately that all Riemann integrable functions are integrable in the new, generalized sense. In particular, all continuous functions and all monotone functions defined on $[a, b]$ are integrable.

The first two theorems are analogs of AUN Theorems 132 and 133 for the Riemann integral. They are proved by replacing the word "fine" with "superfine" in the proofs from AUN, Chapter 9.

**Theorem 10 (Linearity)**
*Let $f$ and $g$ be integrable on $[a, b]$ and let $\lambda, \mu$ be real numbers. Then $\lambda \cdot f + \mu \cdot g$ is integrable on $[a, b]$ and*

$$\int_a^b (\lambda \cdot f + \mu \cdot g)(x) \cdot dx = \lambda \int_a^b f(x) \cdot dx + \mu \int_a^b g(x) \cdot dx.$$

**Theorem 11 (Monotonicity)**
*Let $f$ and $g$ be integrable on $[a, b]$. Assume that $f(x) \le g(x)$, for all $a \le x \le b$. Then*

$$\int_a^b f(x) \cdot dx \le \int_a^b g(x) \cdot dx.$$

**Theorem 12 (Cauchy Test)**
*Let $f$ be defined on $[a, b]$. Then $f$ is integrable on $[a, b]$ if and only if*

$$\sum(f; \mathcal{P}, \mathcal{T}) \simeq \sum(f; \mathcal{P}', \mathcal{T}'),$$

*for all superfine tagged partitions $(\mathcal{P}, \mathcal{T}), (\mathcal{P}', \mathcal{T}')$ of $[a, b]$.*

**Proof:** If $f$ is integrable, then $\sum(f; \mathcal{P}, \mathcal{T}) \simeq \int_a^b f(x) \cdot dx \simeq \sum(f; \mathcal{P}', \mathcal{T}')$, so $f$ has the Cauchy property.

For the converse, assume that $f$ has the Cauchy property. The context is specified by $f$, $a$ and $b$. It suffices to show that the numbers $\sum(f; \mathcal{P}, \mathcal{T})$ are not ultralarge; we can then let $R$ be the observable neighbor of $\sum(f; \mathcal{P}, \mathcal{T})$, and the Cauchy property implies that $f$ is integrable and $\int_a^b f(x) \cdot dx = R$.

We fix one superfine partition $(\mathcal{P}_0, \mathcal{T}_0)$ and let

$$\widetilde{m} = \sum(f; \mathcal{P}_0, \mathcal{T}_0) - 1 \quad \text{and} \quad \widetilde{M} = \sum(f; \mathcal{P}_0, \mathcal{T}_0) + 1.$$

For every superfine partition $(\mathcal{P}, \mathcal{T})$ we have

$$\sum(f; \mathcal{P}, \mathcal{T}) \simeq \sum(f; \mathcal{P}_0, \mathcal{T}_0)$$

and hence $\widetilde{m} < \sum(f; \mathcal{P}, \mathcal{T}) < \widetilde{M}$. The following statement is therefore true:

"There exist $m, M$ such that $m < \sum(f; \mathcal{P}, \mathcal{T}) < M$, for all superfine partitions."

The statement is internal and its parameters are $f, a, b$; therefore, by Closure,

"There exist observable $m, M$ such that $m < \sum(f; \mathcal{P}, \mathcal{T}) < M$, for all superfine partitions."

This is precisely the assertion that no $\sum(f;\mathcal{P},\mathcal{T})$ is ultralarge. □

**Theorem 13 (Additivity)**
*Let $a<c<b$; then $f$ is integrable on $[a,b]$ if and only if $f$ is integrable on $[a,c]$ and on $[c,b]$. If this is the case, then*

$$\int_a^b f(x)\cdot dx = \int_a^c f(x)\cdot dx + \int_c^b f(x)\cdot dx.$$

**Proof:** Assume that $f$ is integrable on $[a,b]$. Let $(\mathcal{P}_1,\mathcal{T}_1)$ and $(\mathcal{P}_2,\mathcal{T}_2)$ be two superfine partitions of $[a,c]$. We extend them to superfine partitions $(\mathcal{P}_1',\mathcal{T}_1')$ and $(\mathcal{P}_2',\mathcal{T}_2')$ of $[a,b]$ in such a way that they coincide on $[c,b]$. We then have

$$\sum(f;\mathcal{P}_1,\mathcal{T}_1)-\sum(f;\mathcal{P}_2,\mathcal{T}_2)=\sum(f;\mathcal{P}_1',\mathcal{T}_1')-\sum(f;\mathcal{P}_2',\mathcal{T}_2')\simeq 0,$$

because $f$ is integrable on $[a,b]$. Hence $f$ has the Cauchy property on $[a,c]$ and is integrable on $[a,c]$.

Assume that $f$ is integrable on $[a,c]$ and $[c,b]$. Let $(\mathcal{P},\mathcal{T})$ be a superfine partition of $[a,b]$; by the comment following Theorem 8 we can assume without loss of generality that $c$ is a partition point of $\mathcal{P}$. Let $(\mathcal{P}_1,\mathcal{T}_1)$ and $(\mathcal{P}_2,\mathcal{T}_2)$ be the restrictions of $(\mathcal{P},\mathcal{T})$ to $[a,c]$ and $[c,b]$, respectively. Then

$$\sum(f;\mathcal{P},\mathcal{T})=\sum(f;\mathcal{P}_1,\mathcal{T}_1)+\sum(f;\mathcal{P}_2,\mathcal{T}_2)\simeq\int_a^c f(x)\cdot dx+\int_c^b f(x)\cdot dx,$$

so $f$ is integrable on $[a,b]$ and

$$\int_a^b f(x)\cdot dx = \int_a^c f(x)\cdot dx + \int_c^b f(x)\cdot dx.$$

□

**Theorem 14 (Fundamental Theorem of Calculus)**
*If $f$ is differentiable on $[a,b]$, then $f'$ is integrable on $[a,b]$ and*

$$\int_a^b f'(x)\cdot dx = f(b)-f(a).$$

**Proof:** Let $(\mathcal{P},\mathcal{T})$ be a superfine partition of $[a,b]$. By the straddle version of the $a$-Increment Equation (Exercise 2), $f(x_{i+1})-f(x_i)=f'(t_i)\cdot dx_i+\varepsilon_i\cdot dx_i$ with $\varepsilon_i$ which is $t_i$-ultrasmall or 0. Hence

$$\sum_{i=0}^{n-1} f'(t_i)\cdot dx_i=\sum_{i=0}^{n-1}(f(x_{i+1})-f(x_i))-\sum_{i=0}^{n-1}\varepsilon_i\cdot dx_i.$$

But $\sum_{i=0}^{n-1}(f(x_{i+1})-f(x_i)) = f(x_n)-f(x_0) = f(b)-f(a)$. Moreover, all $\varepsilon_i$ are ultrasmall (or 0) and $dx_i$ are ultrasmall with $\sum_{i=0}^{n-1} dx_i = b-a$, so $\sum_{i=0}^{n-1}\varepsilon_i\cdot dx_i\simeq 0$, as usual. This shows that $f'$ is integrable and

$$\int_a^b f'(x)\cdot dx = f(b)-f(a).$$

□

Theorem 14 fulfills our goal of solving the fundamental problem of calculus for all differentiable functions. The reader will notice that, up to this point, the proofs in this section closely resemble those of the analogous results in AUN, Chapter 9. However, there are powerful theorems about the generalized Riemann integral that go far beyond Chapter 9. Their proofs tend to be more subtle and, in some cases, involve a combination of "epsilon–delta" and "ultrasmall" arguments. We derive several such results in the rest of this section.

**Theorem 15**
*Let $f$ be a function and $R$ a real number. The following statements are equivalent:*

*(1) $f$ is integrable on $[a, b]$ and $\int_a^b f(x) \cdot dx = R$.*

*(2) For every $\varepsilon > 0$ there is a positive function $\delta$ such that*

$$|\sum(f; \mathcal{P}, \mathcal{T}) - R| < \varepsilon,$$

*for all tagged partitions $(\mathcal{P}, \mathcal{T})$ subordinate to $\delta$.*

**Proof:** The context is specified by $f$, $a$ and $b$.

(1) implies (2): Let $\varepsilon > 0$ be observable. In the proof of Theorem 7 we showed that there is a positive function $\overline{\varphi}$ with the property that every partition $(\mathcal{P}, \mathcal{T})$ subordinate to $\overline{\varphi}$ is superfine, and hence $|\sum(f; \mathcal{P}, \mathcal{T}) - R| < \varepsilon$. Letting $\delta = \overline{\varphi}$ proves (2) for observable $\varepsilon > 0$. By Closure, (2) is true for all $\varepsilon > 0$.

(2) implies (1): Let $(\mathcal{P}, \mathcal{T})$ be superfine, and let $\varepsilon > 0$ be observable. By Closure there is an observable positive function $\delta$ with the property in (2). As $(\mathcal{P}, \mathcal{T})$ is subordinate to $\delta$, we have $|\sum(f; \mathcal{P}, \mathcal{T}) - R| < \varepsilon$. This is true for all observable $\varepsilon > 0$, so $\sum(f; \mathcal{P}, \mathcal{T}) \simeq R$.

□

**Definition 5**
*A **partially tagged partition** of $[a, b]$ is a partition $\mathcal{P} = \{x_0, x_1, \ldots, x_n\}$ and a* partial tagging *$\mathcal{T} = \{t_j : j \in J\}$, where $J \subseteq \{0, 1, \ldots, n-1\}$ and $t_j \in [x_j, x_{j+1}]$ for all $j \in J$.*

As for tagged partitions, we use the notation $\sum(f; \mathcal{P}, \mathcal{T})$ for $\sum_{j \in J} f(t_j) \cdot dx_j$.

**Definition 6**
*We say that a partially tagged partition is **subordinate to** $\varphi$ if*

$$dx_j < \varphi(t_j), \qquad \text{for all } j \in J.$$

*We call it **superfine** if*

$$dx_j \text{ is } t_j\text{-ultrasmall}, \qquad \text{for all } j \in J.$$

The next key lemma shows that, for integrable functions, the Riemann sums give a good approximation of the integral not only over the whole interval $[a, b]$, but over any collection of subintervals from the partition as well.

**Theorem 16 (Saks-Henstock Lemma)**
*Let $f$ be integrable on $[a, b]$ and let $\varepsilon > 0$. There exists a positive function $\delta$ such that*

$$\Big|\sum_{j\in J} f(t_j)\cdot dx_j - \sum_{j\in J}\int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big| < \varepsilon,$$

*for every partially tagged partition $(\mathcal{P}, \mathcal{T})$ subordinate to $\delta$. In fact, for any such partially tagged partition we also have*

$$\sum_{j\in J}\Big|f(t_j)\cdot dx_j - \int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big| < 2\varepsilon,$$

*and even*

$$\Big|\sum_{j\in J}|f(t_j)\cdot dx_j| - \sum_{j\in J}\Big|\int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big|\Big| < 2\varepsilon.$$

**Proof:** Since $f$ is integrable, we can find a positive function $\delta$ in $\mathbf{V}(f, a, b, \varepsilon)$ such that

$$\Big|\sum(f;\mathcal{P},\mathcal{T}) - \int_a^b f(x)\cdot dx\Big| < \varepsilon,$$

for every tagged partition $(\mathcal{P}, \mathcal{T})$ subordinate to $\delta$.

Consider now a partially tagged partition $(\mathcal{P}, \mathcal{T})$ subordinate to $\delta$. The context contains $f, a, b, \varepsilon$, and this partially tagged partition $(\mathcal{P}, \mathcal{T})$. In particular, $J$ is observable. Let $i \notin J$. Since $f$ is integrable over $[x_i, x_{i+1}]$, we have

$$\int_{x_i}^{x_{i+1}} f(x)\cdot dx \simeq \sum(f;\mathcal{P}_i,\mathcal{T}_i)$$

for any superfine partition $(\mathcal{P}_i, \mathcal{T}_i)$ of $[x_i, x_{i+1}]$ (relative to the context augmented by $x_i, x_{i+1}$). Select one such partition for each $i \notin J$ (this is justified by the Principle of Finite Choice, see the Appendix in AUN.) The union of the partitions $(\mathcal{P}_i, \mathcal{T}_i)$ together with $(\mathcal{P}, \mathcal{T})$ is a partition of $[a, b]$ subordinate to $\delta$. Therefore,

$$\Big|\int_a^b f(x)\cdot dx - \sum_{j\in J} f(t_j)\cdot dx_j - \sum_{i\notin J}\sum(f;\mathcal{P}_i,\mathcal{T}_i)\Big| < \varepsilon.$$

On the other hand, by the additivity property of the integral we have

$$\int_a^b f(x)\cdot dx = \sum_{j\in J}\int_{x_j}^{x_{j+1}} f(x)\cdot dx + \sum_{i\notin J}\int_{x_i}^{x_{i+1}} f(x)\cdot dx.$$

Substituting this into the previous inequality, we get

$$\Big|\sum_{j\in J}\int_{x_j}^{x_{j+1}} f(x)\cdot dx - \sum_{j\in J} f(t_j)\cdot dx_j + \sum_{i\notin J}\Big[\int_{x_i}^{x_{i+1}} f(x)\cdot dx - \sum(f;\mathcal{P}_i,\mathcal{T}_i)\Big]\Big| < \varepsilon.$$

As $\{i : i \notin J\}$ is finite and observable, and for each $i$ we have $\int_{x_i}^{x_{i+1}} f(x)\cdot dx \simeq \sum(f;\mathcal{P}_i,\mathcal{T}_i)$, the quantity between the square brackets is ultraclose to 0. It follows that

$$\Big|\sum_{j\in J}\int_{x_j}^{x_{j+1}} f(x)\cdot dx - \sum_{j\in J} f(t_j)\cdot dx_j\Big| < \varepsilon.$$

For the second claim, we consider the two partially tagged partitions determined by

$$J^+ = \{j \in J : f(t_j)\cdot dx_j \geq \int_{x_j}^{x_{j+1}} f(x)\cdot dx\}$$

and

$$J^- = \{j \in J : j \notin J^+\}.$$

Applying the first result separately to $J^+$ and $J^-$, we get

$$\sum_{j\in J^+}\Big|f(t_j)\cdot dx_j - \int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big| = \sum_{j\in J^+}\left(f(t_j)\cdot dx_j - \int_{x_j}^{x_{j+1}} f(x)\cdot dx\right) < \varepsilon$$

and

$$\sum_{j\in J^-}\Big|f(t_j)\cdot dx_j - \int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big| = \sum_{j\in J^-}\left(\int_{x_j}^{x_{j+1}} f(x)\cdot dx - f(t_j)\cdot dx_j\right) < \varepsilon.$$

The second claim follows by adding the two lines.

For the final claim, we use the triangle inequality

$$|f(t_j)\cdot dx_j| \leq \Big|f(t_j)\cdot dx_j - \int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big| + \Big|\int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big|,$$

so

$$\begin{aligned}\sum_{j\in J}|f(t_j)\cdot dx_j| &\leq \sum_{j\in J}\Big|f(t_j)\cdot dx_j - \int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big| + \sum_{j\in J}\Big|\int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big| \\ &< \sum_{j\in J}\Big|\int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big| + 2\varepsilon,\end{aligned}$$

and similarly, from

$$\Big|\int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big| \leq \Big|f(t_j)\cdot dx_j\Big| + \Big|\int_{x_j}^{x_{j+1}} f(x)\cdot dx - f(t_j)\cdot dx_j\Big|$$

one deduces that

$$\sum_{j\in J}\Big|\int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big| < \sum_{j\in J}|f(t_j)\cdot dx_j| + 2\varepsilon.$$

□

**Exercise 3** (Answer page 33)
Show that if $f$ is integrable on $[a,b]$ and $(\mathcal{P},\mathcal{T})$ is a superfine partially tagged partition of $[a,b]$, then

$$\sum_{j\in J} f(t_j)\cdot dx_j \simeq \sum_{j\in J}\int_{x_j}^{x_{j+1}} f(x)\cdot dx.$$

Moreover

$$\sum_{j\in J} |f(t_j)\cdot dx_j| \simeq \sum_{j\in J}\Big|\int_{x_j}^{x_{j+1}} f(x)\cdot dx\Big|.$$

We are ready to prove the central result in the theory of generalized Riemann integral.

**Theorem 17 (Monotone Convergence Theorem)**
*Let $(f_n)_{n=1}^{\infty}$ be a monotone sequence of integrable functions on $[a,b]$ such that*

$$f(x) = \lim_{n\to\infty} f_n(x) \quad \textit{exists for each } x\in[a,b].$$

*If the sequence $(\int_a^b f_n(x)\cdot dx)_{n=1}^{\infty}$ is bounded, then $f$ is integrable on $[a,b]$ and*

$$\int_a^b f(x)\cdot dx = \lim_{n\to\infty}\int_a^b f_n(x)\cdot dx.$$

**Proof:** We give a proof for the case when $(f_n)_{n=1}^{\infty}$ is increasing.

The context is specified by the sequence $(f_n)_{n=1}^{\infty}$, $a$, and $b$. Hence the function $f$ is observable.

Let $k\in\mathbb{N}$. Since $f_k$ is integrable, there is a positive function $\delta_k$ such that

$$\sum_{i\in J}\Big|f_k(t_i)\cdot dx_i - \int_{x_i}^{x_{i+1}} f_k(x)\cdot dx\Big| \le \frac{1}{2^k}, \tag{†}$$

for every partially tagged partition subordinate to $\delta_k$. Since this is a traditional mathematical statement, we can choose one such $\delta_k$ for each $k$, and by Closure, we may assume that the sequence $(\delta_k)_{k=1}^{\infty}$ is observable.

By the monotonicity property of the integral, the sequence $(\int_a^b f_n(x)\cdot dx)_{n=1}^{\infty}$ is increasing. By assumption, it is bounded, so by the Monotone Convergence Theorem for sequences, this sequence converges. Let an observable $R$ be such that

$$R = \lim_{n\to\infty}\int_a^b f_n(x)\cdot dx.$$

Let $(\mathcal{P},\mathcal{T})$ be a superfine partition. By the Local Stability Principle (see the Appendix to these notes), there is an ultralarge $N$ such that $(\mathcal{P},\mathcal{T})$ remains superfine relative to the context augmented by $N$.

Since $N$ is ultralarge, we have

$$R \simeq \int_a^b f_N(x)\cdot dx.$$

For each $x \in [a, b]$, let $k(x) \geq N$ be the least such that

$$|f_{k(x)}(x) - f(x)| \leq \frac{1}{N}.$$

In particular, $f_{k(t_i)}(t_i) \simeq f(t_i)$, for $i = 0, \ldots, n$. Notice that the function $x \mapsto k(x)$ is observable relative to the context augmented by $N$.

We define

$$\delta(x) = \delta_{k(x)}(x),$$

so $\delta$ is a positive function observable relative to the context augmented by $N$.

Since $(\mathcal{P}, \mathcal{T})$ is superfine relative to the context augmented by $N$, it is subordinate to $\delta$.

We need to show that

$$\sum_{i=0}^{n-1} f(t_i) \cdot dx_i \simeq R.$$

First, notice that

$$\sum_{i=0}^{n-1} f(t_i) \cdot dx_i \simeq \sum_{i=0}^{n-1} f_{k(t_i)}(t_i) \cdot dx_i, \tag{*}$$

since $f(t_i) = f_{k(t_i)}(t_i) + \varepsilon_i$, with $\varepsilon_i \simeq 0$, and $\sum_{i=0}^{n-1} \varepsilon_i \cdot dx_i \simeq 0$.

Second, using the Saks-Henstock lemma, we must have

$$\sum_{i=0}^{n-1} f_{k(t_i)}(t_i) \cdot dx_i \simeq \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f_{k(t_i)}(x) \cdot dx. \tag{**}$$

To see this, define

$$J_p = \{i : k(t_i) = p\}, \qquad \text{for each } p \in \mathbb{N}.$$

(Of course, $J_p = \emptyset$ for all but finitely many values of $p$.) We consider the partially tagged partition $(\mathcal{P}, \mathcal{T}_p)$ obtained from $(\mathcal{P}, \mathcal{T})$ by restricting $\mathcal{T}$ to $J_p$. Now since $(\mathcal{P}, \mathcal{T})$ is subordinate to $\delta$, if $i \in J_p$ we have

$$dx_i < \delta(t_i) = \delta_{k_i(t)}(t_i) = \delta_p(t_i),$$

by definition of $\delta$. This shows that $(\mathcal{P}, \mathcal{T}_p)$ is subordinate to $\delta_p$. By (†) we deduce that

$$\sum_{i \in J_p} \left| f_{k(t_i)}(t_i) \cdot dx_i - \int_{x_i}^{x_{i+1}} f_{k(t_i)}(x) \cdot dx \right| \leq \frac{2}{2^p} = \frac{1}{2^{p-1}}.$$

As each $k(t_i)$ is equal to some $p \geq N$, adding these inequalities gives

$$
\begin{aligned}
\Big|\sum_{i=0}^{n-1} f_{k(t_i)}(t_i)\cdot dx_i - \sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}} f_{k(t_i)}(x)\cdot dx\Big| \\
&\le \sum_{i=0}^{n-1}\Big|f_{k(t_i)}(t_i)\cdot dx_i - \int_{x_i}^{x_{i+1}} f_{k(t_i)}(x)\cdot dx\Big| \\
&= \sum_{p\ge N}\sum_{i\in J_p}\Big|f_{k(t_i)}(t_i)\cdot dx_i - \int_{x_i}^{x_{i+1}} f_{k(t_i)}(x)\cdot dx\Big| \\
&\le \sum_{p\ge N}\frac{1}{2^{p-1}} \simeq 0,
\end{aligned}
$$

since $N$ is ultralarge.

Finally, we show that

$$
\sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}} f_{k(t_i)}(x)\cdot dx \simeq R. \qquad (***)
$$

Let $K = \max\{k(t_i) : 0 \le i < n\}$. Then $K \ge N$ and

$$
f_N(x) \le f_{k(t_i)}(x) \le f_K(x), \quad \text{for all } x \in [a,b].
$$

Hence

$$
\int_{x_i}^{x_{i+1}} f_N(x)\cdot dx \le \int_{x_i}^{x_{i+1}} f_{k(t_i)}(x)\cdot dx \le \int_{x_i}^{x_{i+1}} f_K(x)\cdot dx,
$$

and

$$
R \simeq \int_a^b f_N(x)\cdot dx \le \sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}} f_{k(t_i)}(x)\cdot dx \le \int_a^b f_K(x)\cdot dx \simeq R.
$$

(The last step holds because $K$ is ultralarge.) This establishes (***).

Now putting together (*), (**), and (***) yields the conclusion.

□

We derive a few consequences of Monotone Convergence Theorem here, and several more in the next section. Given an interval $I = (a,b)$, we denote by $\ell(I)$ the length of $I$; i.e., $\ell(I) = b - a$.

**Definition 7**

*Let $A \subseteq \mathbb{R}$. We say that $A$ is a **null set** if there are open intervals $I_n$ such that*

$$
A \subseteq \bigcup_{n=1}^{\infty} I_n \quad \text{and} \quad \sum_{n=1}^{\infty} \ell(I_n) \simeq 0.
$$

**Example**

*Let $C = \{c_n\}_{n=1}^{\infty}$ be a countable set. Fix $\varepsilon \simeq 0$ and, for each $n$, let*

$$
I_n = \left(c_n - \frac{\varepsilon}{2^n},\; c_n + \frac{\varepsilon}{2^n}\right).
$$

*Then* $\ell(I_n) = 2 \cdot \frac{\varepsilon}{2^n}$ *and* $\sum_{n=1}^{\infty} \ell(I_n) = 2 \cdot \varepsilon \simeq 0$. *Thus, all countable sets are null sets.*

*The Cantor set is an example of an uncountable null set.*

We say that a statement about real numbers is true **almost everywhere** if the set of those $x$ for which the statement is false is a null set.

**Theorem 18**
*If $f$ is integrable on $[a, b]$ and $f(x) = g(x)$ almost everywhere, then $g$ is integrable on $[a, b]$ and*

$$\int_a^b g(x) \cdot dx = \int_a^b f(x) \cdot dx.$$

**Proof:** Let $h(x) = g(x) - f(x)$. Then there is a null set $E$ such that $h(x) = 0$ for all $x \notin E$. It suffices to prove that $h$ is integrable and

$$\int_a^b h(x) \cdot dx = 0.$$

The context contains $h$, $a$, $b$, and $E$. By Closure, we can find an observable $\{I_k\}_{k=1}^{\infty}$ such that $E \subseteq \bigcup_k I_k$ and $\sum \ell(I_k) \simeq 0$.

We first show the result in the case when $h$ is bounded. By Closure, there is then an observable $M$ such that $|h(x)| \leq M$, for all $x \in [a, b]$. If $(\mathcal{P}, \mathcal{T})$ is superfine relative to this level, then by Theorem 9, for every $t_i \in E$, we have $[x_i, x_{i+1}] \subseteq I_{k_i}$ for some $k_i$. If $t_i \notin E$, then $h(t_i) = 0$. These observations give

$$\begin{aligned} |\sum(h; \mathcal{P}, \mathcal{T})| &= |\sum_{t_i \in E} h(t_i) \cdot dx_i + \sum_{t_i \notin E} h(t_i) \cdot dx_i| \\ &\leq \sum_{t_i \in E} M \cdot dx_i \\ &\leq M \cdot \sum_{k=1}^{\infty} \ell(I_k) \simeq 0. \end{aligned}$$

(In the last step we used the fact that the intervals $[x_i, x_{i+1}]$ are non-overlapping, so $\sum_{k_i = k} dx_i \leq \ell(I_k)$.) Hence $h$ is integrable and $\int_a^b h(x) \cdot dx = 0$.

Now assume that $h \geq 0$, but possibly unbounded. For each $n \in \mathbb{N}$ we define

$$h_n(x) = \min(h(x), n), \quad \text{for } x \in [a, b].$$

Clearly $h_n(x) = 0$ for all $x \notin E$ and $h_n$ is bounded. By the first paragraph, $h_n$ is integrable and $\int_a^b h_n(x) \cdot dx = 0$. As $\{h_n\}_{n=1}^{\infty}$ is increasing and $\lim_{n\to\infty} h_n(x) = h(x)$, the Monotone Convergence Theorem shows that $h$ is integrable and

$$\int_a^b h(x) \cdot dx = \lim_{n\to\infty} \int_a^b h_n(x) \cdot dx = 0.$$

Now let $h$ be arbitrary. Write $h = h^+ - h^-$; clearly $h^+(x)$, $h^-(x)$ equal 0 except for $x \in E$, and $h^+ \geq 0$, $h^- \geq 0$. Thus both $h^+$ and $h^-$ are integrable and $\int_a^b h^+(x)\cdot dx = \int_a^b h^-(x)\cdot dx = 0$ by the previous two paragraphs. It follows by additivity that $h$ is integrable and

$$\int_a^b h(x)\cdot dx = \int_a^b h^+(x)\cdot dx - \int_a^b h^-(x)\cdot dx = 0.$$

□

**Example**

*The Dirichlet function $f : [0,1] \to \mathbb{R}$ is defined by*

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q}; \\ 0 & \text{otherwise.} \end{cases}$$

*As all countable sets are null sets, it follows that the Dirichlet function is generalized Riemann integrable and $\int f(x)\cdot dx = 0$. We saw in AUN, Chapter 9 that the Dirichlet function is not Riemann integrable. We now give a classical example showing that the monotone convergence theorem fails for Riemann integrable functions. Fix an enumeration of $\mathbb{Q} = \{q_n : n \in \mathbb{N}\}$. Define $f_n : [0,1] \to \mathbb{Q}$ by*

$$f_n(x) = \begin{cases} 1 & \text{if } x = q_k, \text{ for some } k \le n; \\ 0 & \text{otherwise.} \end{cases}$$

*Then the sequence $(f_n)$ is increasing and its limit is the Dirichlet function$f$. Moreover, $f_n$ is continuous everywhere except on a finite set, so $f_n$ is Riemann integrable and $\int_a^b f_n(x)\cdot dx = 0$. Thus, $\lim_{n\to\infty}\int_a^b f_n(x)\cdot dx = 0$, but the limit function $f$ is not Riemann integrable.*

The final result of this section is a substantially strengthened version of the Fundamental Theorem of Calculus, which allowed us to recover the function from its derivative. It states essentially that if a continuous function is differentiable everywhere except on a countable set, then one can still recover the function from its derivative. (Recall that if a function is differentiable everywhere on $[a,b]$, then it is continuous.)

**Theorem 19 (Fundamental Theorem of Calculus, Strong Version)**
*Let $f$ be continuous on $[a,b]$. Let $C \subseteq [a,b]$ be a countable set. Let $g$ be a function defined on $[a,b]$ such that $g(x) = f'(x)$ for all $x \notin C$. Then $g$ is integrable and*

$$\int_a^b g(x)\cdot dx = f(b) - f(a).$$

**Proof:** The context is specified by $f, g, a, b$ and an enumeration $\{c_k\}_{k=1}^{\infty}$ of $C$. The set $C$ is a null set, and hence, by Theorem 18, we may assume without loss of generality that $g(x) = 0$ for all $x \in C$. Let $(\mathcal{P}, \mathcal{T})$ be a superfine partition of $[a,b]$. If $t_i \notin C$, then $g(t_i)\cdot dx_i = f'(t_i)\cdot dx_i = f(x_{i+1}) - f(x_i) - \varepsilon_i\cdot dx_i$ where $\varepsilon_i \simeq 0$, by the straddle version of the Increment Equation. If $t_i \in C$, then $g(t_i)\cdot dx_i = 0$. Hence

$$\sum_{i=0}^{n-1} g(t_i)\cdot dx_i = \sum_{t_i\notin C}(f(x_{i+1}) - f(x_i) - \varepsilon_i\cdot dx_i) \simeq \sum_{t_i\notin C}(f(x_{i+1}) - f(x_i)).$$

It remains to prove that $\sum_{t_i \in C}(f(x_{i+1}) - f(x_i)) \simeq 0$.

Let $\varepsilon > 0$ be observable. From the assumption that $f$ is continuous at $c_k$ it follows that there is a $\delta_k > 0$ such that

$$|x - c_k| < \delta_k \quad \text{implies} \quad |f(x) - f(c_k)| < \frac{\varepsilon}{2^k}.$$

We define the function $\varphi$ as follows

$$\varphi(x) = \begin{cases} \delta_k & \text{if } x = c_k; \\ 1 & \text{otherwise.} \end{cases}$$

Then $\varphi$ is an observable positive function. Suppose that $t_i = c_k$. Then, since the partition is superfine we have

$$dx_i < \varphi(c_k) = \delta_k.$$

Hence

$$|f(x_{i+1}) - f(x_i)| \leq |f(x_{i+1}) - f(c_k)| + |f(c_k) - f(x_i)| \leq \frac{\varepsilon}{2^k} + \frac{\varepsilon}{2^k} = \frac{\varepsilon}{2^{k-1}}.$$

Each $c_k$ can be equal to at most two tags ($t_i$ and $t_{i+1}$, if it so happens that $t_i = c_k = t_{i+1}$), so we have

$$\sum_{t_i \in C} |f(x_{i+1}) - f(x_i)| \leq \sum_{k=1}^{\infty} 2 \cdot \frac{\varepsilon}{2^{k-1}} = 4\varepsilon.$$

As $\varepsilon$ was an arbitrary observable positive number, we have

$$\sum_{t_i \in C} |f(x_{i+1}) - f(x_i)| \simeq 0.$$

□

# 3 The Lebesgue integral

The most popular advanced theory of integration is due to Henri Lebesgue. We show that the main theorems about Lebesgue integral follow from the results of the preceding section.

**Definition 8**
*A function $f : [a, b] \to \mathbb{R}$ is **Lebesgue integrable** on $[a, b]$ if both $f$ and $|f|$ are generalized Riemann integrable on $[a, b]$.*

**Exercise 4** (Answer page 33)
Show that Lebesgue integrability is equivalent to the generalized Riemann integrability of $f^+$ and $f^-$.

We will show later in this section that the function $f$ defined by

$$f(x) = \begin{cases} \frac{1}{x} \sin(\frac{1}{x}) & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$$

is integrable on $[0,1]$, but not Lebesgue integrable on $[0,1]$.

**Theorem 20 (Comparison Test)**
*Let $f$ and $g$ be integrable on $[a,b]$. If $|f| \le g$, then $f$ is Lebesgue integrable on $[a,b]$.*

**Proof:** The context is specified by $f$, $g$, $a$, and $b$. Let $(\mathcal{P},\mathcal{T})$ be a superfine partition of $[a,b]$. From $-g \le f \le g$ we derive that

$$-\int_{x_i}^{x_{i+1}} g(x)\cdot dx \le \int_{x_i}^{x_{i+1}} f(x)\cdot dx \le \int_{x_i}^{x_{i+1}} g(x)\cdot dx,$$

i.e.,

$$\Big|\int_{x_i}^{x_{i+1}} f(x)\cdot dx\Big| \le \int_{x_i}^{x_{i+1}} g(x)\cdot dx.$$

Hence, $\sum_{i=0}^{n-1} |\int_{x_i}^{x_{i+1}} f(x)\cdot dx| \le \int_a^b g(x)\cdot dx$ is not ultralarge. Let $R$ be the observable neighbor of

$$\sum_{i=0}^{n-1}\Big|\int_{x_i}^{x_{i+1}} f(x)\cdot dx\Big|.$$

Now

$$|\sum(|f|;\mathcal{P},\mathcal{T}) - R| \le \Big|\sum_{i=0}^{n-1}|f(t_i)|\cdot dx_i - \sum_{i=0}^{n-1}|\int_{x_i}^{x_{i+1}} f(x)\cdot dx|\Big| + \Big|\sum_{i=0}^{n-1}|\int_{x_i}^{x_{i+1}} f(x)\cdot dx| - R\Big|.$$

The first term is ultraclose to 0 by the Saks-Henstock lemma and the second also, by definition of $R$. We conclude that $|f|$ is integrable. □

**Theorem 21 (Linearity)**
*Let $f$ and $g$ be Lebesgue integrable on $[a,b]$ and let $\lambda,\mu$ be real numbers. Then $\lambda\cdot f+\mu\cdot g$ is Lebesgue integrable on $[a,b]$.*

**Proof:** The function $\lambda\cdot f+\mu\cdot g$ is integrable by Theorem 10. Furthermore, $|\lambda\cdot f(x)+\mu\cdot g(x)| \le |\lambda|\cdot|f(x)|+|\mu|\cdot|g(x)|$ and the function on the right is integrable. The Comparison Test implies that $|\lambda\cdot f(x)+\mu\cdot g(x)|$ is integrable. □

**Exercise 5** (**Additivity**) (Answer page 33)
Let $a < c < b$; show that $f$ is Lebesgue integrable on $[a,b]$ if and only if $f$ is Lebesgue integrable on $[a,c]$ and on $[c,b]$.

**Theorem 22**
*Let $f,g$ be integrable and $f \le g$. Then $f$ is Lebesgue integrable if and only if $g$ is Lebesgue integrable.*

**Proof:** Suppose $f$ and $g$ are integrable and $f$ is Lebesgue integrable. Then $g = (g - f) + f$ where $g - f$ is integrable and nonnegative; so $g - f$ is Lebesgue integrable. Hence $g$ is Lebesgue integrable by Theorem 21. The proof when $g$ is supposed to be Lebesgue integrable is similar. $\square$

**Exercise 6** (Answer page 33)
In the Monotone Convergence Theorem, show that if one of the functions $f_n$ is Lebesgue integrable, then $f = \lim f_n$ is Lebesgue integrable.

**Exercise 7** (Answer page 34)
Show that if $f$ is Lebesgue integrable and $f(x) = g(x)$ almost everywhere, then $g$ is Lebesgue integrable.

**Theorem 23**
*If $f, g, h$ are integrable and $h \le f$, $h \le g$, then the functions $\min\{f(x), g(x)\}$ and $\max\{f(x), g(x)\}$ are integrable.*

**Proof:** We first observe that

$$\min\{f(x), g(x)\} = \frac{1}{2}(f(x) + g(x) - |f(x) - g(x)|)$$

and

$$\max\{f(x), g(x)\} = \frac{1}{2}(f(x) + g(x) + |f(x) - g(x)|)$$

(check separately the cases $f(x) \ge g(x)$ and $f(x) \le g(x)$). Next, from the first equation, $|f(x) - g(x)| = f(x) + g(x) - 2 \cdot \min(f(x), g(x)) \le f + g - 2h$, so $|f(x) - g(x)|$ is integrable by the Comparison Test. It follows that $\min\{f(x), g(x)\}$ and $\max\{f(x), g(x)\}$ are integrable. $\square$

The next theorem is a cornerstone of Lebesgue integration. Suppose that $(f_k)$ is a sequence of functions bounded below, i.e., there is a function $h$ such that $f_k \ge h$ for all $k$. Then $h_k(x) = \inf_{\ell \ge k} f_\ell(x)$ exists, and the sequence $(h_k(x))$ is increasing, for all $x \in [a, b]$. We define the function $\liminf f_k$ by

$$(\liminf f_k)(x) = \liminf f_k(x) = \lim_{k \to \infty} h_k(x),$$

for each $x$ where the limit exists (i.e., where the sequence $(h_k(x))$ is bounded).

**Theorem 24 (Fatou's Lemma)**
*Let $f_k$ be integrable functions, for $k = 1, 2, \ldots$. Let $h$ be integrable and such that $h \le f_k$, for all $k$. Suppose that $\liminf f_k$ is defined for all $x \in [a, b]$. If $\liminf \int_a^b f_k(x)dx < \infty$, then $\liminf f_k$ is integrable and*

$$\int_a^b \liminf f_k(x) \cdot dx \le \liminf \int_a^b f_k(x) \cdot dx.$$

*If $h$ is Lebesgue integrable, then $\liminf f_k$ is Lebesgue integrable.*

**Proof:** For fixed $k$ and $n \geq k$, let

$$g_n^k = \min(f_k, \dots, f_n).$$

Then $g_n^k \geq h$. An easy induction using the previous theorem shows that $g_n^k$ is integrable. Moreover, $g_n^k \geq g_{n+1}^k$, so the sequence $(g_n^k)_{n\geq 1}$ is decreasing. But

$$\int_a^b h(x) \cdot dx \leq \int_a^b g_n^k(x) \cdot dx \leq \int_a^b g_k^k(x) \cdot dx,$$

so $(\int_a^b g_n^k(x) \cdot dx)_{n\geq k}$ is bounded. By the Monotone Convergence Theorem, we have that

$$h_k = \inf_{\ell \geq k} f_\ell = \lim_{n\to\infty} g_n^k \text{ is integrable.}$$

The sequence $(h_k)_{k\geq 1}$ is increasing and $h \leq h_k \leq f_k$, so

$$\int_a^b h(x) \cdot dx \leq \int_a^b h_k(x) \cdot dx \leq \int_a^b f_k(x) \cdot dx, \quad \text{for each } k.$$

Hence $\int_a^b h(x) \cdot dx \leq \liminf \int_a^b h_k(x) \cdot dx \leq \liminf \int_a^b f_k(x) \cdot dx < \infty$. But the sequence $(\int_a^b h_k(x) \cdot dx)_{k\geq 1}$ is increasing, so

$$\liminf \int_a^b h_k(x) \cdot dx = \lim_{k\to\infty} \int_a^b h_k(x) \cdot dx < \infty$$

and $(\int_a^b h_k(x) \cdot dx)_{k\geq 1}$ is bounded. Applying the Monotone Convergence Theorem once again, we deduce

$$\int_a^b \liminf f_k(x) \cdot dx = \int_a^b \lim_{k\to\infty} h_k(x) \cdot dx = \lim_{k\to\infty} \int_a^b h_k(x) \cdot dx$$
$$\leq \liminf \int_a^b f_k(x) \cdot dx.$$

The last claim follows from Exercise 6. □

**Theorem 25 (Dominated Convergence Theorem)**
*Let $f_n$ be integrable functions for $n = 1, 2, \dots$ and let $f(x) = \lim_{n\to\infty} f_n(x)$ exist for all $x \in [a, b]$. Suppose that there are integrable functions $h_1, h_2$ such that*

$$h_1(x) \leq f_n(x) \leq h_2(x), \quad \text{for all } x \in [a, b].$$

*Then the function $f$ is integrable and*

$$\int_a^b f(x) \cdot dx = \lim_{n\to\infty} \int_a^b f_n(x) \cdot dx.$$

*If at least one of $h_1, h_2$ is Lebesgue integrable, then $f$ is Lebesgue integrable.*

**Proof:** The assumptions of Fatou's lemma are satisfied, so we deduce that $f$ is integrable and

$$\int_a^b \lim f_n(x) \cdot dx \leq \liminf \int_a^b f_n(x) \cdot dx.$$

Similarly,

$$-\int_a^b \lim f_n(x) \cdot dx = \int_a^b \lim(-f_n(x)) \cdot dx$$
$$\leq \liminf \int_a^b (-f_n(x)) \cdot dx = -\limsup \int_a^b f_n(x) \cdot dx,$$

i.e.,

$$\limsup \int_a^b f_n(x) \cdot dx \leq \int_a^b \lim f_n(x) \cdot dx.$$

It follows that $\liminf \int_a^b f_n(x) \cdot dx = \limsup \int_a^b f_n(x) \cdot dx$, i.e., $\lim \int_a^b f_n(x) \cdot dx$ exists and equals $\int_a^b \lim f_n(x) \cdot dx$. □

**Exercise 8 (Mean Convergence Theorem)** (Answer page 34)
Under the assumptions of the Dominated Convergence Theorem, show that the functions $f - f_n$ are Lebesgue integrable and

$$\lim_{n\to\infty} \int_a^b |f(x) - f_n(x)| \cdot dx = 0.$$

For the next example we need a corollary of the Dominated Convergence Theorem.

**Theorem 26**
*Let $f : [a, b] \to \mathbb{R}$ be such that*

$$|f(x)| \leq h(x), \quad \text{for all } x \in [a, b],$$

*for some $h$ integrable on $[a, b]$. Suppose that $f$ is integrable on $[r, b]$, for every $a < r < b$. Then $f$ is integrable on $[a, b]$ and*

$$\int_a^b f(x) \cdot dx = \lim_{r\to a+} \int_r^b f(x) \cdot dx.$$

*Moreover, if $b = r_0 > r_1 > r_2 > \ldots$ and $\lim_{k\to\infty} r_k = a$, then*

$$\int_a^b f(x) \cdot dx = \sum_{k=0}^{\infty} \int_{r_{k+1}}^{r_k} f(x) \cdot dx.$$

**Proof:** We define $f_k$ by

$$f_k(x) = \begin{cases} f(x) & \text{for } r_k \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

Then each $f_k$ is integrable on $[a, b]$, also $|f_k(x)| \leq h(x)$, and $\lim_{k\to\infty} f_k(x) = f(x)$, for all $x \in [a, b]$. By the dominated convergence theorem, $f$ is integrable on $[a,b]$ and

$$\begin{aligned}\int_a^b f(x)dx &= \lim_{k\to\infty}\int_a^b f_k(x)dx \\ &= \lim_{k\to\infty}\int_{r_k}^b f(x)\cdot dx \\ &= \lim_{k\to\infty}\sum_{i=0}^{k-1}\int_{r_{i+1}}^{r_i} f(x)\cdot dx \\ &= \sum_{i=0}^{\infty}\int_{r_{i+1}}^{r_i} f(x)\cdot dx.\end{aligned}$$

□

**Example**

*We show in this example that the function $f$ defined by*

$$f(x) = \begin{cases} \frac{1}{x}\sin(\frac{1}{x}) & \textit{if } x \neq 0; \\ 0 & \textit{if } x = 0\end{cases}$$

*is integrable on $[0,1]$, but not Lebesgue integrable on $[0,1]$.*

*We first show that $f$ is not Lebesgue integrable on $[0,1]$. Assume, for a contradiction, that $|f|$ is integrable on $[0,1]$. Let $a_k = \frac{1}{k\pi}$ and note that*

$$[0,1] = \{0\} \cup \bigcup_{k=1}^{\infty}(a_{k+1}, a_k] \cup (a_1, 1].$$

*Let $n \geq 1$. By monotonicity and additivity, we have*

$$\int_0^1 \frac{1}{x}\left|\sin\left(\frac{1}{x}\right)\right|\cdot dx \geq \int_{a_n}^{a_1}\frac{1}{x}\left|\sin\left(\frac{1}{x}\right)\right|\cdot dx = \sum_{k=1}^{n-1}\int_{a_{k+1}}^{a_k}\frac{1}{x}\left|\sin\left(\frac{1}{x}\right)\right|\cdot dx.$$

*We use the substitutions $u = \frac{1}{x}$ and $v = u - k\pi$ to deduce*

$$\begin{aligned}\int_{a_{k+1}}^{a_k}\frac{1}{x}\left|\sin\left(\frac{1}{x}\right)\right|\cdot dx &= \int_{(k+1)\pi}^{k\pi} u|\sin(u)|\left(-\frac{1}{u^2}\right)\cdot du \\ &= \int_0^{\pi}\frac{1}{v+k\pi}\sin(v)\cdot dv \\ &\geq \frac{1}{(k+1)\pi}\int_0^{\pi}\sin(v)\cdot dv = \frac{2}{(k+1)\pi}.\end{aligned}$$

*Hence*

$$\int_0^1 \frac{1}{x}\left|\sin\left(\frac{1}{x}\right)\right|\cdot dx \geq \frac{2}{\pi}\sum_{k=1}^{n-1}\frac{1}{k+1}, \quad \textit{for each } n \in \mathbb{N}.$$

*But $\sum_{k=1}^{\infty}\frac{1}{k+1} = \infty$, so $\int_0^1 \frac{1}{x}\left|\sin(\frac{1}{x})\right|\cdot dx = \infty$, a contradiction.*

*We now show that $f$ is integrable: The function $t \mapsto \frac{1}{t}\sin(\frac{1}{t})$ is continuous on $(0,1]$, and integration by parts gives*

$$\int_x^1 \frac{1}{t}\sin\left(\frac{1}{t}\right)\cdot dt = \int_x^1 (-t)\left(-\frac{1}{t^2}\sin\left(\frac{1}{t}\right)\right)\cdot dt = t\cos\left(\frac{1}{t}\right)\bigg|_x^1 - \int_x^1 \cos\left(\frac{1}{t}\right)\cdot dt,$$

*for each $0 < x \leq 1$. As $t \mapsto \cos(\frac{1}{t})$ is bounded by the constant function with value $1$, which is integrable on $[0,1]$, it follows from the corollary to the dominated convergence theorem that $t \mapsto \cos(\frac{1}{t})$ is integrable on $[0,1]$ and $\int_0^1 \cos(\frac{1}{t})\cdot dt = \lim_{x\to 0+}\int_x^1 \cos(\frac{1}{t})\cdot dt$. The function $G(x) = -\int_x^1 \cos(\frac{1}{t})\cdot dt$ is thus continuous at $0$. As $x \mapsto \cos(\frac{1}{x})$ is continuous on $(0,1]$, we have $G'(x) = \cos\frac{1}{x}$, for all $x \in (0,1]$, by the Fundamental Theorem of Calculus for continuous functions. We now let $F$ be defined by*

$$F(x) = \begin{cases} x\cos(\frac{1}{x}) + G(x) & \text{for } 0 < x \leq 1, \\ 0 & \text{if } x = 0. \end{cases}$$

*Note that $F$ is continuous on $[0,1]$ and*

$$F'(x) = \frac{1}{x}\sin\left(\frac{1}{x}\right) = f(x), \qquad \text{for } 0 < x \leq 1.$$

*So $f$ is integrable by the strong version of the Fundamental Theorem of Calculus (Theorem 19).*

With this example we end the systematic development of the theory of integration. Our goal in the next few pages is to define some concepts that play a key role in the traditional expositions of Lebesgue integral.

**Definition 9**
*A collection $\Sigma$ of subsets of a set $S$ is called an* ***algebra*** *if*

*(1) $S \in \Sigma$;*

*(2) If $X, Y \in \Sigma$, then $X \cup Y \in \Sigma$;*

*(3) If $X \in \Sigma$, then $S \setminus X \in \Sigma$.*

**Exercise 9** (Answer page 34)
Let $\Sigma$ be an algebra of subsets of $S$ and $X_1, \ldots, X_n \in \Sigma$. Show that

(1) $X_1 \cup \ldots \cup X_n \in \Sigma$,

(2) $X_1 \cap \ldots \cap X_n \in \Sigma$,

(3) If $X, Y \in \Sigma$, then $X \setminus Y \in \Sigma$.

**Definition 10**
*An algebra $\Sigma$ is called a $\sigma$-**algebra** if*

$$\bigcup_{n\geq 1} X_n \in \Sigma,$$

*for every sequence $(X_n)_{n\geq 1}$ of sets from $\Sigma$.*

**Exercise 10** (Answer page 34)
Let $\Sigma$ be a $\sigma$-algebra. Show that if $X_n \in \Sigma$ $(n = 1, 2, \dots)$, then $\bigcap_{n\geq 1} X_n \in \Sigma$.

**Definition 11**
*The **characteristic function** $\chi_A$ of a set $A$ is the function defined by*

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

**Definition 12**
*A set $A \subseteq [a, b]$ is **Lebesgue measurable** if $\chi_A$ is integrable on $[a, b]$.*
*We denote by $\mathcal{M}[a, b]$ the collection of all Lebesgue measurable subsets of $[a, b]$.*

With these definitions we have:

**Theorem 27**
*Let $[a, b]$ be a closed interval. Then*

*(1) $\mathcal{M}[a, b]$ is a $\sigma$-algebra.*

*(2) For every $a \leq c \leq d \leq b$, the closed interval $[c, d]$ is Lebesgue measurable.*

*(3) All null subsets of $[a, b]$ are Lebesgue measurable.*

**Proof:** If $X, Y \in \mathcal{M}$, then $X \cup Y \in \mathcal{M}$: $\chi_{X\cup Y} = \max(\chi_X, \chi_Y)$, so it is integrable by Theorem 23.

If $X_k \in \mathcal{M}$ for $k = 1, 2, \dots$, we let $Y_n = X_1 \cup \dots \cup X_n$. By the above and induction, $Y_n \in \mathcal{M}$. We note that $(\chi_{Y_n})_{n\geq 1}$ is increasing and $\chi_X = \lim_{n\to\infty} \chi_{Y_n}$. So $\chi_X$ is integrable, by monotone convergence theorem.

The remaining claims are left as an exercise. □

**Definition 13**
*Let $\Sigma$ be an algebra of subsets of $S$. A function $\mu : \Sigma \to [0, \infty]$ defined for all sets in $\Sigma$ is called a **finitely additive measure on** $\Sigma$ if*

*(1) $\mu(X) \geq 0$ for all $X \in \Sigma$;*

*(2) $\mu(\emptyset) = 0$, $\mu(S) > 0$;*

*(3) $\mu(X \cup Y) = \mu(X) + \mu(Y)$ whenever $X$ and $Y$ are disjoint.*

*If, in addition, $\Sigma$ is a $\sigma$-algebra and*

*(4) $\mu(\bigcup_{k\geq 1} X_k) = \sum_{k\geq 1} \mu(X_k)$ holds for every sequence $(X_k)$ of pairwise disjoint sets from $\Sigma$,*

*then $\mu$ is called a **$\sigma$-additive measure on** $\Sigma$.*

Observe that an additive measure is monotone: If $X \subseteq Y$, with $X, Y \in \Sigma$ then $\mu(X) \leq \mu(Y)$. To see this, write $Y = X \cup (Y \setminus X)$ (recall that $Y \setminus X \in \Sigma$), so $\mu(Y) = \mu(X) + \mu(Y \setminus X) \geq \mu(X)$.

**Definition 14**
*The **Lebesgue measure** is the function $m : \mathcal{M}[a, b] \to \mathbb{R}$ defined by*

$$m(A) = \int_a^b \chi_A(x)dx.$$

The next theorem justifies the name. The proof is left as an easy exercise.

**Theorem 28**
*Let $[a, b]$ be a closed interval. Then*

*(1) The Lebesgue measure $m$ is a $\sigma$-additive measure on $\mathcal{M}[a, b]$.*

*(2) For $a \leq c < d \leq b$, we have $m([c, d]) = d - c$.*

*(3) If $E$ is a null set, then $m(E) = 0$.*

**Theorem 29**
*For every set $A \in \mathcal{M}[a, b]$ and every $\varepsilon > 0$ there is an open set $O$ such that $A \subseteq O$ and $m(O) - m(A) \leq \varepsilon$.*

**Proof:** Let $\varphi$ be a positive function. We show first that there is a (finite or infinite) sequence $(I_k)$ of non-overlapping closed intervals, and a sequence of tags $(t_k)$, such that each

$$t_k \in I_k \cap A, \quad \ell(I_k) < \varphi(t_k), \quad \text{and} \quad A \subseteq \bigcup I_k.$$

We start with $I = [a, b]$ and split it into two subintervals, $I_0 = [a, m]$ and $I_1 = [m, b]$, where $m = (a + b)/2$ is the midpoint of $I$. Similarly, we split $I_0$ into $I_{00}$ and $I_{01}$ using the midpoint of $I_0$, and $I_1$ into $I_{10}$ and $I_{11}$. Continuing in this manner, we construct a countably infinite system of intervals $I_\eta$ ($\eta$ is a finite sequence of 0 and 1) such that

$$[a,b] = \bigcup_{\ell(\eta)=n} I_\eta \quad \text{and} \quad \ell(I_\eta) = \frac{b-a}{2^n}, \quad \text{if } \eta \text{ has length } n.$$

For each $x \in A$, we can find $n$ such that $(b-a)/2^n < \varphi(x)$, and therefore we can find a sequence $\eta = \eta(x)$ of length $n$ such that $I_\eta \subseteq (x - \varphi(x),\ x + \varphi(x))$. We let $t_\eta = x$. Let $(\eta_k)$ be an enumeration of the sequences $\eta$ we have used, making sure that no sequence $\eta_k$ extends $\eta_n$, for $k > n$. Then the intervals $I_k = I_{\eta_k}$ are non-overlapping (by removing the extensions, we made sure that no interval $I_k$ in the list is contained in an interval $I_n$ for $n < k$) and their union contains $A$. Letting $t_k = t_{\eta_k}$, all the requirements are satisfied.

The context is specified by $A$. Let $\overline{\varphi}$ be a positive function as in the proof of Theorem 7, i.e., every partition $(\mathcal{P}, \mathcal{T})$ subordinate to $\overline{\varphi}$ is superfine, and hence

$$\sum(\chi_A; \mathcal{P}, \mathcal{T}) \simeq \int_a^b \chi_A(x) \cdot dx = m(A).$$

Let $(I_k)$ and $(t_k)$ correspond to $\overline{\varphi}$ as in the first paragraph. It is clear that, for each $n \in \mathbb{N}$, $(I_k)_{k=1}^n$ and $(t_k)_{k=1}^n$ determine a partially tagged subpartition $(\mathcal{P}_n, \mathcal{T}_n)$ of $[a,b]$ subordinate to $\overline{\varphi}$ (hence superfine). We write $I_k = [x_k, x_{k+1}]$.

Then

$$\begin{aligned}
\sum_{k=1}^n \ell(I_k) &= \sum(\chi_A; \mathcal{P}_n, \mathcal{T}_n) && \text{(since each } t_k \in A) \\
&\simeq \sum_{k=1}^n \int_{x_k}^{x_{k+1}} \chi_A(x) \cdot dx && \text{(Exercise 3)} \\
&= \sum_{k=1}^n m(A \cap I_k) \le m(A).
\end{aligned}$$

Now $A \subseteq \bigcup I_k$, and therefore

$$\sum \ell(I_k) \simeq m(A).$$

Let $\varepsilon > 0$ be observable. To complete the proof, it remains only to replace each closed interval $I_k = [x_k, x_{k+1}]$ by a slightly larger open interval $J_k = (x_k - \frac{\varepsilon}{2^{k+2}},\ x_{k+1} + \frac{\varepsilon}{2^{k+2}})$, and let $O = \bigcup J_k$. Then $A \subseteq O$ and

$$m(O) \le \sum m(J_k) = \sum \ell(I_k) + \sum \frac{\varepsilon}{2^{k+1}} \le m(A) + \varepsilon.$$

This is true for each observable $\varepsilon > 0$, so for each $\varepsilon > 0$ by closure. □

**Theorem 30**
*If $m(N) = 0$, then $N$ is a null set. If $\{N_k\}_{k=1}^\infty$ are null sets, then $\bigcup_{k=1}^\infty N_k$ is a null set.*

**Proof:** This is a corollary of the preceding theorem. □

This result establishes that our measurable sets coincide with the usual ones, and that null sets are precisely the sets of measure 0.

We give one more definition.

**Definition 15**
*A function $f : [a, b] \to \mathbb{R}$ is (Lebesgue)* ***measurable*** *if, for all $c < d$, the set*

$$\{x \in [a, b] : c \le f(x) \le d\}$$

*is Lebesgue measurable.*

Every nonnegative measurable function is a limit of an increasing sequence of **simple** measurable functions, i.e., functions of the form

$$\sum_{i=1}^{n} a_i \cdot \chi_{A_i},$$

where $A_i \subseteq [a, b]$ are measurable, and $a_i$ are real numbers (for the proof, see e.g. H. R. Royden, *Real Analysis, Third Edition,* Prentice Hall, 1988.) It is an easy exercise to show that simple functions are Lebesgue integrable and

$$\int_a^b \left( \sum_{i=1}^{n} a_i \cdot \chi_{A_i}(x) \right) \cdot dx = \sum_{i=1}^{n} a_i \cdot m(A_i).$$

By the Monotone Convergence Theorem it follows that a nonnegative measurable function is integrable if and only if there is an increasing sequence $(f_k)_{k \ge 1}$ of simple measurable functions such that

$$\lim_{k \to \infty} f_k = f \quad \text{and} \quad \left( \int_a^b f_k(x) dx \right)_{k \ge 1} \text{ is bounded.}$$

Readers familiar with the usual approaches to Lebesgue integral should be able to conclude from the above observations that functions Lebesgue integrable according to the traditional definitions are precisely the functions Lebesgue integrable according to our definition, and the values of the integrals are the same.

In this book we restricted ourselves to the theory of integration for functions defined on a bounded interval $[a, b]$. It is important to be able to integrate functions defined on $\mathbb{R}$ (or a subset of $\mathbb{R}$), but this is now easy. We give only the key definitions.

**Definition 16**
*A function $f : \mathbb{R} \to \mathbb{R}$ is* ***integrable on*** $\mathbb{R}$ *if there is an observable number $R$ such that*

$$\int_r^s f(x) \cdot dx \simeq R, \quad \text{for all } r \simeq -\infty,\ s \simeq \infty.$$

*A set $A \subseteq \mathbb{R}$ is* ***measurable*** *if*

$$A \cap [r, s] \text{ is measurable}, \quad \text{for all } r \simeq -\infty,\ s \simeq \infty.$$

We leave it as a challenging exercise for the reader to formulate and prove analogs of the results of sections 2 and 3 for $\mathbb{R}$ in place of $[a, b]$. The main difference is that bounded measurable functions on $\mathbb{R}$ are not necessarily integrable. There are even measurable sets $A$ for which the characteristic function is not integrable (for example, $A = \mathbb{R}$).

# 4 Lebesgue Theorem

We finish with a classical result.

**Theorem 31 (Lebesgue Theorem)**
*A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if it is continuous almost everywhere.*

We first introduce a characterization of continuous functions that is used in the proof.

**Definition 17**
*Let $f : A \to \mathbb{R}$ be a function. Let $a$ be a limit point of $A$ and let $L$ be a real number. We say that $L$ is a* ***limit value*** *of $f$ at $a$ if $L \simeq f(x)$ for some $x \simeq a$, $x \neq a$, $x \in A$.*

The context of the previous definition is $a, A, f, L$.

**Theorem 32**
*Let $f : A \to \mathbb{R}$ be a function, $a$ be a limit point of $A$, and $L$ be a real number. The following conditions are equivalent:*

*(1)* $\lim_{x\to a} f(x) = L$.

*(2) $L$ is the unique limit value of $f$ at $a$ and $f(x)$ is limited for all $x \simeq a$, $x \neq a$, $x \in A$.*

**Proof:**
We show (1) implies (2). The context is given by $f, a, A$, and $L$. Assume $\lim_{x\to a} f(x) = L$. Then $L \simeq f(x)$ for all $x \simeq a$, $x \neq a$, $x \in A$. But such $x$ exist since $a$ is a limit point of $A$. Hence, $L$ is a limit value of $f$ at $a$ and $L$ is limited. Let $L'$ be another limit value. Consider an extended context $f, a, A, L, L'$, and write $\overset{+}{\simeq}$ when working relative to it. By definition, there is $x' \overset{+}{\simeq} a$, $x' \neq a$, such that $f(x') \overset{+}{\simeq} L'$. But $f(x) \overset{+}{\simeq} L$ also, by Closure, so $L \overset{+}{\simeq} L'$. This shows that $L = L'$ since $L, L'$ are observable relative to $f, a, A, L, L'$.

To see that (2) implies (1), notice that the assumed unique limit value $L$ of $f$ at $a$ is observable relative to $f, a, A$, by Closure. If $x \simeq a$, $x \neq a$, $x \in A$, then $f(x)$ is limited, and so the observable neighbor of $f(x)$ is defined and is also a limit value of $f$ at $a$, i.e., it is equal to $L$. Hence $L \simeq f(x)$, which shows that $\lim_{x\to a} f(x) = L$. □

**Definition 18**
*Let $f : A \to \mathbb{R}$ be a function and $a$ a limit point of $A$. We let $\mathcal{L}_a$ denote the set of all limit values of $f$ at $a$.*

The set $\mathcal{L}_a$ is observable relative to $f, a, A$.

**Theorem 33**
*Let $f : A \to \mathbb{R}$ be a function and $a$ be a limit point of $A$. The set $\mathcal{L}_a$ is closed.*

**Proof:** The context is given by $f, a, A$. Let $M$ be observable and such that $M \simeq L$ for $L \in \mathcal{L}_a$. Consider the extended context $f, a, A, L$ and write $\overset{+}{\simeq}$ when working relative to it. By definition $L \overset{+}{\simeq} f(x)$ for some $x \overset{+}{\simeq} a$, $x \neq a$, $x \in A$. But then $M \simeq f(x)$ where $x \simeq a$, $x \neq a$, $x \in A$, so $M \in \mathcal{L}_a$. □

**Definition 19**
*Let $I$ be an interval and let $f : I \to \mathbb{R}$ be bounded.*

*(1) We say that the function $f^{\uparrow} : I \to \mathbb{R}$ defined by*

$$f^{\uparrow}(x) = \max(\mathcal{L}_x \cup \{f(x)\}), \quad \textit{for each } x \in I,$$

*is the **upper envelope** of $f$.*

*(2) We say that the function $f^{\uparrow} : I \to \mathbb{R}$ defined by*

$$f^{\uparrow}(x) = \min(\mathcal{L}_x \cup \{f(x)\}), \quad \textit{for each } x \in I,$$

*is the **lower envelope** of $f$.*

*(3) We say that the function $\omega : I \to \mathbb{R}$ defined by*

$$\omega(x) = f^{\uparrow}(x) - f^{\downarrow}(x), \quad \textit{for each } x \in I,$$

*is the **variation** of $f$.*

We note that the functions $f^{\uparrow}, f^{\downarrow}$ and $\omega$ are defined on $I$ because the sets $\mathcal{L}_x$ are closed and bounded. Further, all these functions and observable relative to $f, I$.

**Theorem 34**
*A bounded function $f : I \to \mathbb{R}$ is continuous at $a \in I$ if and only if $\omega(a) = 0$.*

**Proof:** If $f$ is continuous at $a$, then $f(a)$ is the only limit value of $f$ at $a$, and $f^{\uparrow}(a) = f^{\downarrow}(a) = f(a)$, so $\omega(a) = 0$. Conversely, if $f^{\uparrow}(a) = f^{\downarrow}(a)$, then $f(a)$ is the only limit value of $f$ at $a$. (Note that $f$ has at least one limit value at $a$, because $f(x)$ is limited, for any $x \simeq a$.) As $f$ is bounded, $\lim_{x \to a} f(x) = f(a)$. □

**Proof of Lebesgue Theorem:**
Let $D = \{x \in [a, b] : f \text{ is not continuous at } x\}$. This set is observable.
Assume first that $f$ is Riemann integrable. Let

$$D_k = \{x \in [a, b] : \omega(x) \geq \frac{1}{k}\}, \quad \text{for } k \in \mathbb{N}.$$

By Theorem 34, $D = \bigcup_{k=1}^{\infty} D_k$, so it suffices to show that each $D_k$ is a null set. The context contains $f, a, b$ and $k$.

Recall (AUN, Chapter 9) that, if $f$ is Riemann integrable on $[a, b]$, then there exists a fine partition $\mathcal{P}$ such that

$$S(\mathcal{P}) - s(\mathcal{P}) = \sum_{i=0}^{n-1} (F_i - f_i) \cdot dx_i \simeq 0,$$

where $F_i$ ($f_i$, resp.) is the supremum (infimum, resp.) of $f$ on the interval $[x_i, x_{i+1}]$. Let $K = \{i : F_i - f_i \geq \frac{1}{k}\}$; as

$$S(\mathcal{P}) - s(\mathcal{P}) \geq \sum_{i \in K} (F_i - f_i) \cdot dx_i \geq \frac{1}{k} \cdot \sum_{i \in K} dx_i,$$

and $\frac{1}{k}$ is observable, it follows that $\sum_{i \in K} dx_i \simeq 0$. Finally, we note that $D_k \subseteq \bigcup_{i \in K} (x_i, x_{i+1}) \cup \{x_0, \ldots, x_n\}$. It follows that $D_k$ is a null set.

For the converse, assume that $D$ is a null set. Let $\{J_k\}_{k=1}^{\infty}$ be a sequence of open intervals such that

$$D \subseteq \bigcup_{k=1}^{\infty} J_k \quad \text{and} \quad \sum \ell(J_k) \simeq 0.$$

We now augment the context by the sequence $\{J_k\}_{k=1}^{\infty}$, and let $(\mathcal{P}, \mathcal{T})$ be any partition of $[a, b]$ superfine relative to this augmented context. In the notation of the first part of this proof, we will show that

$$\sum_{i=0}^{n-1} f_i \cdot dx_i \simeq \sum_{i=0}^{n-1} F_i \cdot dx_i \tag{*}$$

This is enough (see the direction Darboux ⇒ Riemann of AUN, Theorem 139). Let

$$I_1 = \{i : t_i \in D\} \quad \text{and} \quad I_2 = \{i : t_i \notin D\}.$$

By Theorem 9, for each $i \in I_1$ there is $k_i$ such that $[x_i, x_{i+1}] \subseteq J_{k_i}$. As the intervals of the partition are non-overlapping, we conclude that $\sum_{i \in I_1} dx_i \leq \sum_{k=1}^{\infty} \ell(J_k) \simeq 0$ (see the argument in the proof of Theorem 19), and hence $\sum_{i \in I_1} (F_i - f_i) \cdot dx_i \simeq 0$, i.e., $\sum_{i \in I_1} F_i \cdot dx_i \simeq \sum_{i \in I_1} f_i \cdot dx_i$. Now for $i \in I_2$ the function $f$ is continuous at $t_i$, hence $f(x) - f(t_i)$ is $t_i$-ultrasmall (or 0) for every $x \in [x_i, x_{i+1}]$. This implies that $F_i \simeq f_i$ for $i \in I_2$ and thus $\sum_{i \in I_2} F_i \cdot dx_i \simeq \sum_{i \in I_2} f_i \cdot dx_i$ as usual. Together we have (*). □

# Appendix

The axioms of RBST do not give a *complete* description of the universe of relative set theory. (See KH, *Relative set theory: Internal view*, Journ. Logic and Analysis 1:8, 2009, 1 -108.) Other axioms of some practical usefulness can be added to RBST. We give an example that is used in the proof of Theorem 17.

**Local Transfer Principle.**
Let $\mathcal{P}(x_1, \ldots, x_k; \mathbf{S}_\alpha)$ be any statement in the $\in$–$\mathbf{S}$-language.
If $\mathcal{P}(x_1, \ldots, x_k; \mathbf{S}_\alpha)$ holds, then there exists $\gamma \sqsupset \alpha$ such that $\mathcal{P}(x_1, \ldots, x_k; \mathbf{S}_\beta)$ holds for all $\beta$ with $\alpha \sqsubseteq \beta \sqsubseteq \gamma$.

The point is that $x_1, \ldots, x_k$ are arbitrary; they do not have to be observable relative to $\alpha$!

Also, for every $\alpha \sqsubset \beta$ there exists $N \in \mathbb{N}$ which is ultralarge relative to $\alpha$ and observable relative to $\beta$.

# Answers to Exercises.

**Answer to Exercise 1, page 4**

(1) Let $r$ and $s$ be $a$-accessible. Let $f, g$ be observable and such that $r = f(a)$ and $s = g(a)$. Then $r \pm s = (f+g)(a)$, $r \cdot s = (f \cdot g)(a)$ and $r/s = (f/g)(a)$ (if $s = g(a) \neq 0$). But $f \pm g$, $f \cdot g$, and $f/g$ are observable by Closure, so $r \pm s$, $r \cdot s$, and $r/s$ are $a$-accessible.

(2) Let $x$ and $y$ be $a$-limited. Let $r$, $s$ be positive $a$-accessible numbers such that $|x| \leq r$ and $|y| \leq s$. Then $|x \pm y| \leq |x| + |y| \leq r + s$ and $r + s$ is $a$-accessible so $x \pm y$ is $a$-limited. Also $|x \cdot y| = |x| \cdot |y| \leq r \cdot s$, and $r \cdot s$ is $a$-accessible so $x \cdot y$ is $a$-limited.

(3) Let $h, k$ be $a$-ultrasmall. Let $r > 0$ be $a$-accessible. Then $r/2$ is $a$-accessible (since 2 is $a$-accessible). Thus $|h \pm k| \leq |h| + |k| \leq r/2 + r/2 = r$. So $h \pm k$ is $a$-ultrasmall or 0. Let $x$ be $a$-limited. Let $r > 0$ be $a$-accessible such that $|x| \leq r$. Let $s > 0$ be $a$-accessible. Then $r/s > 0$ is $a$-accessible by (1). Now $|x \cdot h| = |x| \cdot |h| \leq r \cdot r/s = r$. This shows that $x \cdot h$ is $a$-ultrasmall or 0.

**Answer to Exercise 2, page 5**

The proof is identical to the proof of the straddle version from the Increment Equation, but one uses the operations on $a$-ultrasmalls instead.

**Answer to Exercise 3, page 13**

The context is given by $f, a, b$. Let $(\mathcal{P}, \mathcal{T})$ be a superfine partially tagged partition of $[a, b]$. Let $\varepsilon > 0$ be observable. By Closure, we can find a positive observable $\delta$ like in the Saks-Henstock lemma. Since $(\mathcal{P}, \mathcal{T})$ is superfine, it is subordinate to $\delta$ so

$$\Big| \sum_{j \in J} f(t_j) \cdot dx_j - \sum_{j \in J} \int_{x_j}^{x_{j+1}} f(x) \cdot dx \Big| < \epsilon.$$

Since $\epsilon > 0$ was arbitrary, the quantity between the absolute values is ultrasmall or 0 and so $\sum_{j \in J} f(t_j) \cdot dx_j \simeq \sum_{j \in J} \int_{x_j}^{x_{j+1}} f(x) \cdot dx$. The second claim is proved similarly.

**Answer to Exercise 4, page 18**

Assume that $f^+$ and $f^-$ are generalized Riemann integrable. Then by linearity, $f = f^+ - f^-$ and $|f| = f^+ - f^-$ are generalized Riemann integrable.

For the converse, notice that $f^+ = \frac{|f|+f}{2}$ and $f^- = \frac{|f|-f}{2}$. Hence, if $|f|$ and $f$ are generalized Riemann integrable, then so are $f^+$ and $f^-$, by linearity.

**Answer to Exercise 5, page 19**

This is clear since $f$ and $|f|$ are integrable on $[a, b]$ if and only if they are integrable on $[a, c]$ and $[c, b]$.

**Answer to Exercise 6, page 20**

If one of the functions $f_n$ is Lebesgue integrable then they all are by Theorem 22, so in particular $f_n^+$ and $f_n^-$ are integrable, so also $-f_n^-$. But $f^+(x) = \lim_{n\to\infty} f_n^+(x)$ and $-f^-(x) = \lim_{x\to\infty} -f_n^-(x)$. By the Monotone Convergence Theorem applied to $f^+$ and $-f^-$ we have that $f^+$ and $-f^-$ are integrable, so $f^+$ and $f^-$ are integrable. Hence $f$ is Lebesgue integrable.

**Answer to Exercise 7, page 20**

If $f(x) = g(x)$ almost everywhere, then $|f(x)| = |g(x)|$ almost everywhere, so the theorem follows from the corresponding theorem for generalized Riemann integral, applied to $f$, $g$, and to $|f|$, $|g|$.

**Answer to Exercise 8, page 22**

Since $f$ is Lebesgue integrable, then $f - f_n$ is Lebesgue integrable for every $n$. Let $g_n = |f - f_n|$, for $n = 1, 2, \ldots$. Then each $g_n$ is integrable and the sequence tends to 0. Also $0 \leq g_n(x) \leq \max\{h_1(x), h_2(x)\}$, and both 0 and $\max\{h_1, h_2\}$ are integrable. Hence, by the Dominated Convergence Theorem, we have

$$\lim_{n\to\infty} \int_a^b |f - f_n| \cdot dx = \lim_{n\to\infty} \int_a^b g_n(x) \cdot dx = \int_a^b 0 \cdot dx = 0.$$

**Answer to Exercise 9, page 24**

(1) and (2) are simple inductions on the definition. For (3) notice that $X \setminus Y = X \cap (S \setminus Y)$. But $S \setminus Y \in \Sigma$ if $Y \in \Sigma$, and since $X \in \Sigma$, the intersection $X \cap (S \setminus Y) \in \Sigma$ also.

**Answer to Exercise 10, page 25**

It is enough to show that the complement of $\bigcap_{n=1}^{\infty} X_n$ belongs to $\Sigma$. But $S \setminus \bigcap_{n=1}^{\infty} X_n = \bigcup_{n=1}^{\infty} (S \setminus X_n)$. Now $S \setminus X_n \in \Sigma$ if $X_n \in \Sigma$, and so $\bigcup_{n=1}^{\infty} (S \setminus X_n) \in \Sigma$ since $\Sigma$ is a $\sigma$-algebra.