Notes on Integration
Version date: 3/11/19

Contents

6 Riemann Integration

Integration (here meaning *definite* integration, not antidifferentiation[1]) is always about "adding stuff up". The "stuff" may be not be a finite, or even countable, set of numbers; the "sum" may be something like amount of space between the graph of a positive function $f : [a, b] \to \mathbf{R}$ ("area under a curve"), or the total mass or (electric) charge of a solid object for which we know the "mass density" or "charge density" at each point. Generally, the "stuff" is described in some way by a real-valued or vector-valued function on some subset of $\mathbf{R}^n$. Regardless of the dimension of the domain, or whether the function is real-valued vs. vector-valued, or which theory of integration is used (there are several progessively more general theories), the core idea that *integration is about adding stuff up* is always there, even when its presence isn't obvious. But to rigorize the vague "adding stuff up", one starts first with the simplest theory of integration, the subject of this chapter.

[1]In higher mathematics, "definite integration" is the default meaning of "integration", unless context makes it clear that antidifferentiation is meant.

This chapter develops the theory of the Riemann integral of a real-valued (and, later, a vector-valued) function f on a closed, bounded interval $[a, b] \subset \mathbf{R}$. The approach we use is very intuitive, rigorizing the "limit of Riemann sums" idea that's presented in Calculus 1. This approach, which uses the same definition of "Riemann integrable" as in Rosenlicht's textbook [5], has an additional advantage: it provides the most natural generalization to integration of vector-valued functions. For integration of real-valued functions, however, it is not the most efficient approach. A more efficient approach that (non-obviously) turns out to be essentially equivalent is discussed in Section 6.4, as optional reading for the interested student.

Throughout this chapter, when we use notation of the form "$[a, b]$", it is understood that $a, b \in \mathbf{R}$ and that $a < b$.

The symbol "▲" is used in these notes to mark the end of a definition, remark, or example. When all that is being defined is notation specific to these notes, sometimes we label that definition as "Notation 6.x" to avoid giving the impression that this notation is standard among mathematicians.

6.0 Overview of the chapter

In this chapter, first we will define *(Riemann-)integrable real-valued function* f on an interval $[a, b]$, and the integral $\int_a^b f(x)\,dx$ (Section 6.1). From the definitions, based on *Riemann sums*, it is not obvious which functions on $[a, b]$ are integrable (for example, it is not obvious that continuous functions are integrable). Rather than trying to apply the definitions directly in many examples, which would be quite time-consuming, we first establish several general properties (starting in Section 6.2), such as the fact that the set $\mathcal{R}([a, b])$ of integrable functions on $[a, b]$ is a vector space and that the map $\mathcal{R}([a, b]) \to \mathbf{R}$ given by $f \mapsto \int_a^b f(x)\,dx$ is linear. In Section 6.3 we show that step-functions (Definition 6.42) are integrable, and develop an integrability criterion (for any $f : [a, b] \to \mathbf{R}$) based on step-functions. We use this in Section 6.5 to establish that continuous functions on $[a, b]$ are integrable. (As mentioned earlier, the intervening Section 6.4, optional reading, relates the approach taken in Sections 6.1–6.3 to a different approach preferred by many mathematicians. This approach involves something called *upper* and *lower integrals*)

The methods presented in Section 6.3 rely on the concept of *upper* and *lower sums* introduced there, as does the step-function-related integrability-criterion mentioned above (and therefore, also, our proof that "continuous implies integrable" in Section 6.5). However, *every important result that we prove using upper and lower sums can be proven without them.* We have used them in Sections 6.3 and 6.5 for two reasons: (1) they are very helpful for developing a visual understanding of integration of real-valued functions, and (2) upper and lower *sums* provide a bridge connecting the definitions of "integrable" and "integral" in Section 6.1 to those based on upper and lower *integrals* in Section 6.4. The approach to integration discussed in Section 6.4 relies *critically* on upper and lower sums; there they are not simply a dispensable convenience.

Section 6.6 establishes an additivity property reflecting the principle that "integration is about adding stuff up": if $a < c < b$ and f is integrable on $[a, b]$, then f is integrable on the subintervals $[a, c]$ and $[c, b]$, and $\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$ (the "amount of stuff" between a and b is the amount of stuff between a and c plus the amount of stuff between c and b). This is used in Section 6.7 to prove several theorems that go by the name "The Fundamental Theorem of Calculus", one of which is then used in Section 6.8 to prove the validity of the change-of-variable formula learned in Calculus 1.

In Section 6.9 we generalize from integrating real-valued functions on $[a, b]$ to integrating vector-valued functions on $[a, b]$. In this detour from real-valued functions, the generality and strengths of the Riemann-sum approach to the Riemann integral—i.e. the approach taken in Sections 6.1–6.3, as opposed to the approach taken in Section 6.4—become more evident: for any complete normed vector space $(V, \|\ \|)$, we can define "integrable function $f : [a, b] \to V$" and $\int_a^b f(x)\,dx$ *exactly* the way we did for real-valued functions (modulo obvious notational changes). In particular this applies to *any* finite-dimensional vector space with *any* norm. When $V = \mathbf{R}^n$, this definition of the vector-valued integral agrees with the one taught in Calculus 3, but now we see the Calc-3 definition as a corollary of something much more general. Our new definition of the vector-valued integral makes no reference to a basis of V (which the Calc-3 definition does implicitly or explicitly). We do not need V to be $\mathbf{R}^n$, or even to be finite-dimensional. Several of our results for real-valued functions generalize, without any change in the proofs, to V-valued functions, e.g. linearity of the map $f \mapsto \int_a^b f(x)\,dx$. In addition, once we establish an integrability criterion (Proposition 6.85) whose "$V = \mathbf{R}$" case eliminates any need for upper and lower sums (and, therefore, any need for the step-function-related integrability-criterion we used for real-valued functions) we are able to prove generalized versions, for V-valued functions, of other earlier results, e.g. the "triangle inequality for integrals" (equations (6.31) and (6.72)) and the integrability of continuous functions.

Finally, in Section 6.10 we return to the real-valued case, and use our earlier results to define the natural logarithm function and derive its properties. We use these to define a^r for all $a > 0$ and $r \in \mathbf{R}$, using a unified definition that does not depend in any way on whether r is positive or negative, is an integer, is rational, or is irrational. We see that this elegant (albeit nonintuitive) definition reduces to the usual definition for rational r, and implies that for irrational exponents, the "intuitive"l definition of a^r actually works—i.e. that a^r can be defined *unambiguously* (if inelegantly) as a limit obtained by approaching r through rational exponents. (It's very unlikely that the student was shown in high school, or wherever he/she first encountered irrational exponents, that this definition is *unambiguous*, i.e. that the value of the limit is does not depend on *which* of the uncountably many rational sequences approaching r is used.) We also show that all the usual algebraic "rules of exponents" follow, and show the functions $x \mapsto a^x$ (for any $a > 0$) and $x \mapsto x^r$ (for any $r \in \mathbf{R}$) are differentiable and have the "expected" formulas for their derivatives. For the function $x \mapsto x^r$ with r irrational, this would be extraordinary difficult using only the "intuitive" definition of x^r, but with our unified definition the derivative computations are identical for *all* r.

6.1 Definitions and first examples

Definition 6.1 (Partitions) A *partition* P of a closed, bounded interval $[a, b]$ is a finite set $\{x_0, x_1, \ldots, x_N\}$, where $a = x_0 < x_1 < \cdots < x_N = b$. (Thus the number of points in P is $N + 1 \geq 2$.) ▲

Remark 6.2 This use of the word "partition" is special to intervals. The student may be used to "partition of a set S" meaning a disjoint collection of subsets of S whose union is S. In the setting of Definition 6.1, the interval $[a, b]$ is the union of the subintervals $[x_{j-1}, x_j]$, $1 \leq j \leq N$, but these subintervals are not disjoint (unless $N = 1$): for $0 < j < N$, the point x_j lies in two of these subintervals, as the right endpoint of one and the left endpoint of another. We *could* express $[a, b]$ as the disjoint union of $N-1$ half-open intervals and one closed interval, e.g. $[x_0, x_1) \cup [x_1, x_2), \cup \cdots \cup [x_{N-2}, x_{N-1}) \cup [x_{N-1}, x_N]$ or $[x_0, x_1] \cup (x_1, x_2] \cup \cdots \cup (x_{N-2}, x_{N-1}] \cup (x_{N-1}, x_N]$, but the choice of which interval should include a given x_j (other than x_0 and x_N) would be asymmetric and artificial. For purposes of integration, it turns out that 1-point overlaps are irrelevant (we will see why later), so we allow ourselves to speak of $[a, b]$ as being "partitioned" into the closed subintervals $\{[x_{j-1}, x_j]\}$ even though this is not quite consistent with the set-theoretic notion. Finally, since the set of these closed subintervals is completely determined by the set of their endpoints, and vice-versa, the exceptional meaning of "partition" in the context of intervals doesn't cause a problem once one gets used to it.

Notation 6.3 For each partition $P = \{x_0, x_1, \ldots, x_N\}$ of an interval $[a, b]$, we define $\Delta_j(P) = x_j - x_{j-1}$, $1 \leq j \leq N$. When a single partition P is understood from context, we write simply Δ_j rather than $\Delta_j(P)$. ▲

Observe that for any partition P of $[a, b]$, we have

$$\sum_j \Delta_j(P) = b - a. \tag{6.1}$$

Definition 6.4 (Pointed partitions and Riemann sums) Let $a, b \in \mathbf{R}$ be given, with $a < b$, and let $P = \{x_0, x_1, \ldots, x_N\}$ be a partition of $[a, b]$.

1. We define the *width* of P (also called the *mesh* of P), denoted $\text{wid}(P)$ in these notes, to be $\max\{\Delta_j : 1 \leq j \leq N\}$.

2. A *pointing* T of P is a set $T = \{t_1, \ldots t_N\}$ such that $t_j \in [x_{j-1}, x_j]$ for each $j \in \{1, \ldots, N\}$. We call the pair (P, T) a *pointed partition* (of $[a, b]$). We define the width of (P, T) to be the width of P.

3. Given $f : [a, b] \to \mathbf{R}$ and a pointing $T = \{t_1, \dots t_N\}$ of the partition P, the *Riemann sum* for f corresponding to the pointed partition (P, T) is

$$S(f; P, T) = \sum_{j=1}^{N} f(t_j)\Delta_j \,. \tag{6.2}$$

4. Given $f : [a, b] \to \mathbf{R}$, we will write

$$\begin{aligned} \mathcal{S}(f; P) &= \{S(f; P, T) : T \text{ is a pointing of } P\} \\ &= \text{the set of all Riemann sums of } f \text{ associated with the partition } P. \end{aligned}$$

▲

Note that every partition $\{x_0, \dots, x_N\}$ has a pointing (in fact, uncountably many); e.g. we can take $t_j = x_j$ for $1 \le j \le N$ (the "right-endpoint pointing").

For a general pointing of a partition, with notation as in Definition 6.4, we think of the point t_j as a "sample point" within the interval $[x_{j-1}, x_j]$, providing a "sample value" of f on this interval.

Remark 6.5 Observe that any interval $[a, b]$ as above has partitions of arbitrarily small width: given $\delta > 0$, let N be any positive integer such that $\Delta := \frac{b-a}{N} < \delta$, let $x_j = a + j\Delta$ for $0 \le j \le N$, and let P be the partition $\{x_0, x_1, \dots, x_N\}$; we then have $\text{wid}(P) < \delta$. Hence there also always exist pointed partitions of $[a, b]$ arbitrarily small width.

Definition 6.6 (Integrability) A function $f : [a, b] \to \mathbf{R}$ is *Riemann integrable* if there is a real number A such that for every $\epsilon > 0$ there exists $\delta > 0$ such that if (P, T) is any pointed partition of $[a, b]$ of width less than δ, then $|S(f; P, T) - A| < \epsilon$. More generally, if f is a real-valued function whose domain includes $[a, b]$, we say that f *is Riemann integrable on* $[a, b]$ (or *over* $[a, b]$) if $f|_{[a,b]}$ is Riemann integrable. ▲

Notation 6.7 For $a, b \in \mathbf{R}$ with $a < b$, we will let $\mathcal{R}([a, b])$ denote the set of all real-valued functions on $[a, b]$ that are Riemann-integrable. ▲

With notation as in Definition 6.6, suppose that A, A' are two real numbers satisfying the condition required of A in the definition. Let $\epsilon > 0$ be given, and $\delta > 0$ be such that for every pointed partition of $[a, b]$ of width less than δ, we have $|S(f; P, T) - A| < \epsilon$ and $|S(f; P, T) - A'| < \epsilon$. Let (P_δ, T_δ) be a pointed partition of $[a, b]$ of width $< \delta$; such (P, T) exists by Remark 6.5. Then $|A' - A| \le |A' - S(f; P_\delta, T_\delta)| + |S(f; P_\delta, T_\delta) - A| < 2\epsilon$. Since this is true for all $\epsilon > 0$, it follows that $A' - A = 0$, hence that $A' = A$. Therefore *if f is integrable on $[a, b]$, then there is a* unique *number A satisfying the condition in Definition 6.6.* Thus we can make the following definition:

Definition 6.8 (the Riemann integral) Let $f : [a, b] \to \mathbf{R}$ be a Riemann-integrable function. We define the *Riemann integral of* f to be the unique real number A satisfying the condition given in Definition 6.6. This number is denoted

$$\int_a^b f(x)\,dx,\ \int_a^b f(t)\,dt,\ \text{etc.}; \tag{6.3}$$

any letter not reserved with another meaning can be used in place of the "dummy" variable x, t, etc., in the sample notation above. Since the name of the dummy variable does not affect the value of the integral, in these notes we will also use the notation

$$\int_a^b f \tag{6.4}$$

in place of (6.3). More generally, if f is a real-valued function on a domain that includes $[a, b]$, and f is integrable on $[a, b]$, we use the same notation (6.3), (6.4) for the Riemann integral of $f|_{[a,b]}$, and refer to the value of this integral as the *Riemann integral of* f *over* $[a, b]$.

Any conclusion of the form $\int_a^b f =$ [specific number] implicitly means "f is integrable on $[a, b]$ and $\int_a^b f =$ [that number]," if the integrability of f has not already been stated explicitly.

Finally, we define the phrase "$\int_a^b f$ exists" (or "$\int_a^b f(x)\,dx$ exists", etc. for any dummy variable), to mean that f is integrable on $[a, b]$. ▲

Definitions 6.6 and 6.8 give precise meaning to the notion that a definite integral is a "limit of Riemann sums". It is tempting to write, suggestively, that the value of $\int_a^b f(x)\,dx$ is "$\lim_{\text{wid}(P)\to 0} S(f; P, T)$", but this limit-notation cannot be interpreted literally. The quantity $S(f; P, T)$ is not the value of a function of $\text{wid}(P)$, or even the value of a function of P. For every $\delta > 0$, there are infinitely many partitions of $[a, b]$ of width δ, and for every partition there are infinitely many pointings. Thus for every value of $\text{wid}(P)$, there can be (and usually are) *infinitely many* values of Riemann sums associated with partitions of this width. In the notation "$\lim_{x\to x_0} g(x)$" for the limit at x_0 of a function $g : U \setminus \{x_0\} \to \mathbf{R}$, where $U \subset \mathbf{R}$, for each x there is *one and only one* number $g(x)$. However, there are a few ways to write the integral of an integrable function as a *true* limit. The following exercise gives one of these; Theorem 6.32, later in these notes, gives another.

Exercise 6.1 Let $f : [a, b] \to \mathbf{R}$ be given.

(a) Prove that if f is integrable on $[a, b]$, then for any sequence $((P_n, T_n))_{n=1}^{\infty}$ of pointed partitions for which $\text{wid}(P_n) \to 0$ as $n \to \infty$,

$$\lim_{n\to\infty} S(f; P_n, T_n) = \int_a^b f(x)\ dx. \tag{6.5}$$

(Hence the integral can be evaluated by taking such a limit, *if you know ahead of time that f is integrable.*)

(b) Assume that for every sequence $((P_n, T_n))$ of pointed partitions for which $\mathrm{wid}(P_n) \to 0$ as $n \to \infty$, $\lim_{n\to\infty} S(f; P_n, T_n)$ exists. Prove that f is integrable on $[a, b]$, and that for every such sequence $((P_n, T_n))$, the equality (6.5) holds.

Proposition 6.9 ("Integrable implies bounded") *If $f : [a, b] \to \mathbf{R}$ is Riemann integrable, then f is bounded.*

Proof: Let $f : [a, b] \to \mathbf{R}$ be a Riemann-integrable function, and let $A = \int_a^b f$. Let $\delta > 0$ be such that for each pointed partition (P, T) of $[a, b]$ of width less than δ, $|S(f; P, T) - A| < 1$; equivalently

$$A - 1 < S(f; P, T) < A + 1 \tag{6.6}$$

(such δ exists by the definition of integrability). Fix a partition $P = \{x_0, \dots x_N\}$ of $[a, b]$ of width less than δ. Then (6.6) holds for every pointing T of P.

Assume that f is unbounded from above. Then f is unbounded from above on at least one of the intervals $I_j := [x_{j-1}, x_j]$, since there are only finitely many such intervals. Let $j_0 \in \{1, \dots, N\}$ be such that f is unbounded from above on I_{j_0}. For each $n \in \mathbf{N}$, choose $z_n \in I_{j_0}$ such that $f(z_n) > n$. For each $j \in \{1, \dots, N\}$ with $j \neq j_0$, fix any number $t_j \in [x_{j-1}, x_j]$, let $T^{(n)}$ be the pointing $\{t_1^{(n)}, \dots, t_N^{(n)}\}$ of P for which $t_j^{(n)} = \begin{cases} t_j & \text{if } j \neq j_0, \\ z_n & \text{if } j = j_0\,, \end{cases}$ and let $A' = \sum_{j \neq j_0} f(t_j)\Delta_j$. Then

$$S(f; P, T^{(n)}) = A' + f(z_n)\Delta_{j_0} > A' + n\Delta_{j_0}$$

For n sufficiently large, we have $A' + n\Delta_{j_0} > A + 1$, contradicting the second inequality in (6.6). Hence f is bounded from above.

If f is unbounded from below, a similar argument shows that the first inequality in (6.6) is contradicted. Hence f is bounded from below as well as from above, and is therefore bounded. ■

An argument similar to the one preceding Definition 6.8 leads to a useful necessary criterion for integrability:

Proposition 6.10 If $f : [a, b] \to \mathbf{R}$ is Riemann integrable, then for every $\epsilon > 0$ there exists $\delta > 0$ such that if (P, T) and (Q, T') are pointed partitions of $[a, b]$ of width less than δ, we have $|S(f; P, T) - S(f; Q, T')| < \epsilon$.

We omit the proof here, since this proposition is part of a more powerul result we will prove later (Theorem 6.32), and we want to get quickly to some simple examples of integrable and non-integrable functions. In the latter case, we will use Proposition 6.10 in its contrapositive form: for any given $f : [a, b] \to \mathbf{R}$, if there exists $\epsilon_0 > 0$ such that, for all $\delta > 0$, there exist pointed pointed partitions $(P, T), (Q, T')$ of $[a, b]$ of width less than δ for which $|S(f; P, T) - S(f; Q, T')| \geq \epsilon_0$, then f is not Riemann integrable.

Example 6.11 (an integrable function) For any $c \in \mathbf{R}$, the constant function $f : [a, b] \to \mathbf{R}$ given by $f(x) = c$ is integrable, and

$$\int_a^b c\,dx = c(b - a).$$

This follows from the fact that, as the student may check, every Riemann sum for f has the value $c(b - a)$. ▲

In particular, $\mathcal{R}([a, b])$ is nonempty!

Example 6.12 (a non-integrable function) Define $f : [a, b] \to \mathbf{R}$ by $f(x) = 1$ if $x \in \mathbf{Q}$ and $f(x) = 0$ if $x \notin \mathbf{Q}$. Let $P = \{x_0, \ldots, x_N\}$ be a partition of $[a, b]$. For $1 \leq j \leq N$ choose $t_j, t_j' \in [x_{j-1}, x_j]$ such that $t_j \in \mathbf{Q}$ and $t_j' \notin \mathbf{Q}$. Let $T = \{t_1, \ldots, t_N\}, T' = \{t_1', \ldots, t_N'\}$. Then the Riemann sums of f corresponding to the pointed partitions T, T' respectively are Then, the corresponding Riemann sum is

$$S(f; P, T) = \sum_j f(t_j)\Delta_j = \sum_j \Delta_j = b - a$$

and

$$S(f; P, T') = \sum_j f(t_j')\Delta_j = \sum_j 0 = 0.$$

Hence $S(f; P, T) - S(f; P, T') = b - a$. Since this is true regardless of how small $\mathrm{wid}(P)$ is, it follows that f is not Riemann integrable. (In the contrapositive form of Proposition 6.10 that we stated above, take $\epsilon_0 = b - a$, take δ arbitrary, and take $Q = P$.) ▲

Definitions 6.6 and 6.8 are very intuitive, and, as we shall see later, generalize naturally to the integration of vector-valued functions (functions $[a, b] \to V$, where V is a complete normed vector space). However, these definitions can be unwieldy at times; it can be a chore to show the integrability of functions that are any more complicated than the constant function in Example 6.11. In the interests of efficiency, we postpone presenting other examples until we have developed equivalent definitions that are (often) easier to work with. For now, however, we introduce some notation that will allow us to rewrite the definition of "$\int_a^b f = A$" more succinctly than in Definition 6.6:

Notation 6.13 (a) Let $\mathcal{P}([a,b])$ denote the set of partitions of $[a,b]$, and, for each $\delta > 0$, let $\mathcal{P}_\delta([a,b]) \subset \mathcal{P}([a,b])$ denote the set of partitions of $[a,b]$ of width less than δ.

(b) For each function $f : [a,b] \to \mathbf{R}$ and each $\delta > 0$, let

$$\begin{aligned}\mathcal{S}_\delta(f) &= \bigcup_{P \in \mathcal{P}_\delta([a,b])} \mathcal{S}(f;P) \\ &= \{\text{all Riemann sums of } f \text{ associated to partitions of width less than } \delta\}.\end{aligned}$$

▲

Remark 6.14 The definition of "$\int_a^b f = A$" can now be rewritten simply as: for each $\epsilon > 0$ there exists $\delta > 0$ such that, for all $S \in \mathcal{S}_\delta(f)$, we have $|S - A| < \epsilon$. Even more simply: $\int_a^b f = A$ if (and only if) for each $\epsilon > 0$ there exists $\delta > 0$ such that $\mathcal{S}_\delta(f) \subset B_\epsilon(A)$. ▲

For simplicity, **henceforth in these notes we will say simply that f is *integrable* on $[a,b]$ if f is Riemann integrable on $[a,b]$, and refer to $\int_a^b f$ as the *integral* of f over $[a,b]$.** The student is cautioned that there are more general types of integrability—in particular, a type called *Lebesgue integrability*—and that usually when mathematicians say to each other (or to graduate students) that a function on $[a,b]$ is integrable, they mean Lebesgue-integrable. The analog of Proposition 6.9 is *false* for Lebesgue-integrable functions, and false even for functions for which we define an *improper integral* as in Calculus 2. Indeed, the fact that no unbounded function is Riemann integrable is viewed as a *weakness* of Riemann integration compared to Lebesgue integration. Nonetheless, studying Lebesgue integration without first studying Riemann integration can interfere with developing an intuitive understanding of *any* form of integration.

6.2 Linearity and order properties of the integral

Proposition 6.15 *Let $f, g : [a,b] \to \mathbf{R}$ and $c \in \mathbf{R}$ be given. If both f and g are integrable, then so are $f+g$ and cf, and the following equalities hold:*

$$\int_a^b (f+g) = \int_a^b f + \int_a^b g. \tag{6.7}$$

$$\int_a^b cf = c\int_a^b f. \tag{6.8}$$

Proof: From the definition (6.2), we easily see that, for any pointed partition (P,T) of $[a,b]$, we have $S(f+g;P,T) = S(f;P,T) + S(g;P,T)$ and $S(cf;P,T) = cS(f;P,T)$.

Assume now that f and g are integrable, and let $A = \int_a^b f, C = \int_a^b g$. Let $\epsilon > 0$ be given, and let $\delta_1, \delta_2 > 0$ be such that $\mathcal{S}_{\delta_1}(f) \subset B_\epsilon(A)$ and $\mathcal{S}_{\delta_2}(g) \subset B_\epsilon(C)$. Then for any pointed partition (P,T) of $[a,b]$ of width less than $\min\{\delta_1, \delta_2\}$, we have

$$\begin{aligned} |S(f+g;P,T) - (A+C)| &= |(S(f;P,T) - A) + (S(g;P,T) - C)| \\ &\leq |S(f;P,T) - A| + |S(g;P,T) - C| \\ &< 2\epsilon. \end{aligned}$$

It follows that $f+g$ is integrable and that (6.7) holds. Similarly, for any pointed partition (P,T) of $[a,b]$ of width less than δ_1,

$$|S(cf;P,T) - cA| = |cS(f;P,T) - cA| = |c|\,|S(f;P,T) - A| \leq |c|\epsilon,$$

from which the integrability of cf and the equality (6.8) follow. ■

Remark 6.16 The proof of Proposition 6.15 illustrated something that comes up in innumerable proofs. As you may have learned in MAA 4211, in proofs we are quite often in a situation of a form like the following: Statement 1 is true for all $x \in (0, \delta_1)$ (or for all $n > N_1$), statement 1 is true for all $x \in (0, \delta_2)$ (or for all $n > N_2$), ..., statement k is true for all $x \in (0, \delta_k)$ (or for all $n > N_k$). We then say "Let $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_k\}$" (or "Let $N = \max\{N_1, \ldots, N_k\}$"), and are then guaranteed that all k statements are true for all $x \in (0, \delta)$ (or for all $n \geq N$). As long as there are only finitely many statements involved (typically there are only two), this device always works. Once the student has had sufficient experience, he/she should not have trouble following proofs in which several of these intermediate steps are omitted. For example, in the proof of Proposition 6.15, we could have replaced the third and fourth sentences with, "Let $\epsilon > 0$ be given, and let $\delta > 0$ be such that $\mathcal{S}_\delta(f) \subset B_\epsilon(A)$ and $\mathcal{S}_\delta(g) \subset B_\epsilon(C)$. Then for any pointed partition (P,T) of width less than δ, we have ...." By the end of MAA 4211, students should definitely have had enough experience to be comfortable with such arguments (*but should always know how to justify them the "long way"*). So, **henceforth in these notes, we will use this device to shorten arguments whenever we can**.

Observe that, since $\mathcal{R}([a,b])$ is nonempty, Proposition 6.15 can be phrased alternatively as follows:

Proposition 6.17 (Linearity of the integral) *The set $\mathcal{R}([a,b])$ is a vector space, and the map $f \mapsto \int_a^b f$ is a linear map $\mathcal{R}([a,b]) \to \mathbf{R}$.*

The integration-map $f \mapsto \int_a^b f$ also has the following "non-negativity" property:

Proposition 6.18 *Assume that* $f : [a, b] \to \mathbf{R}$ *is integrable and that* $f(x) \geq 0$ *for all* $x \in [a, b]$*. Then* $\int_a^b f \geq 0$.

Exercise 6.2 Prove Proposition 6.18.

Corollary 6.19 (Order property of the integral) *Assume that* $f, g : [a, b] \to \mathbf{R}$ *are integrable and that* $f(x) \geq g(x)$ *for all* $x \in [a, b]$*. Then* $\int_a^b f \geq \int_a^b g$.

Proof: Let $h = f - g$. Then $h \in \mathcal{R}([a, b])$ (by Proposition 6.17) and $h(x) \geq 0$ for all $x \in [a, b]$. Hence

$$0 \leq \int_a^b h = \int_a^b (f - g) = \int_a^b f - \int_a^b g,$$

and the result follows. ■

Remark 6.20 Note that Proposition 6.18 does *not* imply that if f is integrable over $[a, b]$ and $f(x) > 0$ for all $x \in [a, b]$, then $\int_a^b f > 0$. A similar observation applies to Corollary 6.19. Changing a non-strict inequality to a strict one in *hypotheses* does not mean that inequalities in *conclusions* become strict. (For example, given a convergent real-valued sequence $(a_n)_{n=1}^{\infty}$, it is true that if $a_n \geq 0$ for all n then $\lim_{n\to\infty} a_n \geq 0$, but it is *not* true that if $a_n > 0$ for all n then $\lim_{n\to\infty} a_n > 0$.) We will see later (Remark 6.54 following Exercise 6.6) that if f in Proposition 6.18 is assumed *continuous*, and $f(x) > 0$ for all $x \in [a, b]$, then indeed $\int_a^b f > 0$.

It is natural to ask whether pointwise-positivity of f *does* imply positivity of the integral if we assume only that f is integrable, not that f is continuous. In these notes, *we will leave this question open*. Students are invited to try either to prove that the integral is positive under these hypotheses, or to find a counterexample in which the hypotheses are met but $\int_a^b f = 0$. ▲

6.3 Upper and lower sums

Since unbounded functions are not (Riemann-)integrable (Proposition 6.9), we will simplify some parts of the presentation below by restricting attention to bounded functions.

Notation 6.21 We will write $\mathcal{B}([a, b])$ for the set of bounded real-valued functions on $[a, b]$. ▲

Thus, Proposition 6.9 can be written succinctly as: $\mathcal{R}([a, b]) \subset \mathcal{B}([a, b])$.

Definition 6.22 For each function $f : [a,b] \to \mathbf{R}$, $P \in \mathcal{P}([a,b])$, and $\delta > 0$, we define

$$\begin{aligned} U(f;P) &= \sup(\mathcal{S}(f;P)) \\ L(f;P) &= \inf(\mathcal{S}(f;P)), \\ U_\delta(f) &= \sup(\mathcal{S}_\delta(f)), \\ L_\delta(f) &= \inf(\mathcal{S}_\delta(f)). \end{aligned}$$

The quantities $U(f;P), L(f;P)$ are called, respectively, the *upper* and *lower sums of* f *with respect to* P. ▲

Observe that, trivially, in the setting of Definition 6.22 we have

$$\begin{aligned} L(f;P) &\leq U(f;P) \\ \text{and} \qquad L_\delta(f) &\leq U_\delta(f). \end{aligned} \tag{6.9}$$

Remark 6.23 Recall that, in general, the supremum of a nonempty set of real numbers can be ∞, and the infimum can be $-\infty$. A consequence of the upcoming Proposition 6.27 is that for $f \in \mathcal{B}([a,b])$, the upper and lower sums of f with respect to any partition are *finite* (i.e real numbers, never $\pm\infty$), and for any $\delta > 0$ so are the numbers $U_\delta(f)$ and $L_\delta(f)$. ▲

Lemma 6.24 *Let* $\{X_\alpha : \alpha \in A\}$ *be a collection of nonempty subsets* X_α *of* $\mathbf{R}$ *indexed by a nonempty set* A. *Then*

$$\begin{aligned} \sup\left(\bigcup\{X_\alpha : \alpha \in A\}\right) &= \sup\{\sup(X_\alpha) : \alpha \in A\} \\ \textit{and} \qquad \inf\left(\bigcup\{X_\alpha : \alpha \in A\}\right) &= \inf\{\inf(X_\alpha) : \alpha \in A\}. \end{aligned}$$

Exercise 6.3 Prove Lemma 6.24.

In the setting of Definition 6.22, applying Lemma 6.24 to the indexed collection $\{\mathcal{S}(f;P) : P \in \mathcal{P}_\delta([a,b])\}$, we have

$$\begin{aligned} U_\delta(f) &= \sup\{\sup(\mathcal{S}(f;P)) : P \in \mathcal{P}_\delta([a,b])\} \\ &= \sup\{U(f;P) : P \in \mathcal{P}_\delta([a,b])\}, \end{aligned} \tag{6.10}$$

and similarly

$$L_\delta(f) = \inf\{L(f;P) : P \in \mathcal{P}_\delta([a,b])\}. \tag{6.11}$$

Example 6.25 Let $f : [0,1] \to \mathbf{R}$ be the squaring function: $f(x) = x^2$. For each positive integer N, let $P_N = \{x_j := \frac{j}{N} : 0 \leq j \leq N\}$, a partition of $[0,1]$. Consecutive points of this partition are equally spaced: $\Delta_j(P_N) = \frac{1}{N}$ for each $j \in \{1, 2 \ldots, N\}$.[2] Let $T = \{t_1, \ldots, t_N\}$ be a pointing of P_N. Then

$$S(f; P_N, T) = \sum_{j=1}^{N} t_j^2 \ \Delta_j(P_N) = \sum_{j=1}^{N} t_j^2 \, \frac{1}{N} \ .$$

For the j^{th} term in the sum, we have $\frac{j-1}{N} = x_{j-1} \leq t_j \leq x_j = \frac{j}{N}$, implying $\frac{(j-1)^2}{N^2} \leq t_j^2 \leq \frac{j^2}{N^2}$. Hence

$$\sum_{j=1}^{N} \frac{(j-1)^2}{N^2} \frac{1}{N} \ \leq \ S(f; P_N, T) \ \leq \ \sum_{j=1}^{N} \frac{j^2}{N^2} \frac{1}{N} \ . \tag{6.12}$$

Morever, if we take T to be the "right-endpoint pointing" of P_N (i.e. $t_j = x_j$ for $1 \leq j \leq N$) then the value of $S(f; P_N, T)$ is exactly the rightmost sum in (6.12), while if we take T to be the "left-endpoint pointing" of P_N (i.e. $t_j = x_{j-1}$ for $1 \leq j \leq N$) then the value of $S(f; P_N, T)$ is exactly the leftmost sum in (6.12). Hence, using the fact that $\sum_{j=1}^{N} j^2 = \frac{N(N+1)(2N+1)}{6}$ (which is easily proven by induction, and which you may have learned in high school), it follows from (6.12) that

$$\begin{aligned} U(f; P_N) = \frac{1}{N^3} \sum_{j=1}^{N} j^2 \quad &= \quad \frac{1}{N^3} \frac{N(N+1)(2N+1)}{6} \\ &= \quad \frac{1}{3} + \frac{1}{2N} + \frac{1}{6N^2} \end{aligned}$$

and that

$$\begin{aligned} L(f; P_N) = \frac{1}{N^3} \sum_{j=1}^{N} (j-1)^2 = \frac{1}{N^3} \sum_{j=1}^{N-1} j^2 \quad &= \quad \frac{1}{N^3} \frac{(N-1)N(2N-1)}{6} \\ &= \quad \frac{1}{3} - \frac{1}{2N} + \frac{1}{6N^2} \, . \end{aligned}$$

▲

Remark 6.26 In Example 6.25, the fact that the supremum $U(f; P_N) = \sup(\mathcal{S}(f; P_N))$ and infimum $L(f; P_N) = \inf(\mathcal{S}(f; P_N))$ were achieved by, respectively, the right-endpoint and left-endpoint pointings of P_N, *was a consequence of having chosen the function f in this example to be monotone-increasing* on the interval of interest, $[0, 1]$. *In this example,* $U(f; P_N)$ and $L(f; P_N)$ turned out to be the *maximal* and *minimal* Riemann sums of f for

[2] A partition with equally-spaced points is sometimes called a *regular partition*.

this partition. For a general function $f : [a, b] \to \mathbf{R}$ (bounded or otherwise), and partition P (with or without equally-spaced points) the values $U(f; P)$ and $L(f; P)$ may not be achieved by *any* pointings of P, let alone by the left-endpoint or right-endpoint pointings; there be no maximal or minimal Riemann sums. Never forget that "sup" and "inf" are *more general concepts* than "max" and "min", and that you cannot replace "sup" by "max" (or "inf" by "min") unless you have shown that the supremum (or infimum) of the set in question *lies in that set*. ▲

Proposition 6.27 *Let $f \in \mathcal{B}([a, b]$ be given, and let $M = \sup\{f(x) : x \in [a, b]\}$ and $m = \inf\{f(x) : x \in [a, b]\}$. Then for each $P \in \mathcal{P}([a, b])$ and $S \in \mathcal{S}(f; P)$, we have*

$$m(b - a) \le S \le M(b - a). \tag{6.13}$$

Hence

$$\mathcal{S}(f, P) \subset [m(b - a), M(b - a)] \tag{6.14}$$

and

$$m(b - a) \le L(f; P) \le U(f; P) \le M(b - a). \tag{6.15}$$

Consequently, for each $\delta > 0$ and $P \in \mathcal{P}_\delta([a, b])$,

$$m(b - a) \ \le\ L_\delta(f) \ \le\ L(f; P) \ \le U(f; P) \ \le\ U_\delta(f) \ \le\ M(b - a). \tag{6.16}$$

Proof: Let $P \in \mathcal{P}([a, b])$ and let $S \in \mathcal{S}(f; P)$. Then $S = S(f; P, T)$ for some pointing T of P, so (6.13) follows from the Riemann-sum definition (6.2) and the fact that $\sum_j \Delta_j = b-a$. The first and third inequalities in (6.15) follow immediately from (6.13) and the definitions of $L(f; P)$ and $U(f; P)$, and the middle inequality is simply the trivially-true inequality (6.9). The inequalities $L_\delta(f) \le L(f; P)$ and $U(f; P) \le U_\delta(f)$ in (6.16) follow from (6.11) and (6.10), respectively, and the remaining inequalities follow from (6.15). ■

Proposition 6.28 *Let $f \in \mathcal{B}([a, b])$ be given. Define functions $h_1, h_2 : (0, \infty) \to \mathbf{R}$ by*

$$\begin{aligned} h_1(\delta) &= L_\delta(f) \\ \text{and} \qquad h_2(\delta) &= U_\delta(f). \end{aligned}$$

Then h_1 is monotone decreasing, h_2 is monotone increasing, and both functions are bounded.

Proof: For $\delta_1, \delta_2 \in \mathbf{R}$ with $\delta_1 < \delta_2$, every partition of width less than δ_1 also has width less than δ_2. Hence $\mathcal{P}_{\delta_1}([a, b]) \subset \mathcal{P}_{\delta_2}([a, b])$, implying that $\mathcal{S}_{\delta_1}(f) \subset \mathcal{S}_{\delta_2}(f)$. But for any nonempty subsets A, B of $\mathbf{R}$ with $A \subset B$, we have $\inf(A) \ge \inf(B)$ and $\sup(A) \le \sup(B)$. Hence $L_{\delta_1}(f) \ge L_{\delta_2}(f)$ and $U_{\delta_1}(f) \le U_{\delta_2}(f)$.

This proves the asserted monotonicity. Boundedness follows from (6.16). (In the notation of (6.16), the ranges of both h_1 and h_2 lie in $[m(b-a), M(b-a)]$.) ■

We will soon be considering certain limits as $\delta \to 0$. Notice that for a decreasing function h_1 on $(0, \infty)$, the value $h_1(\delta)$ *increases* as $\delta \to 0$; similarly, for an increasing function h_2 on $(0, \infty)$, the value $h_2(\delta)$ *decreases* as $\delta \to 0$.

Lemma 6.29 *Let $I \subset \mathbf{R}$ be an interval bounded from below, and let c be the left endpoint of $\bar{I}$ (the closure of I); equivalently, let $c = \inf(I)$.*

(i) If $h : I \setminus \{c\} \to \mathbf{R}$ is an increasing function that is bounded from below, then

$$\lim_{u \to c} h(u) = \inf(\text{range}(h)). \tag{6.17}$$

(ii) If $h : I \setminus \{c\} \to \mathbf{R}$ is a decreasing function that is bounded from above, then

$$\lim_{u \to c} h(u) = \sup(\text{range}(h)). \tag{6.18}$$

In particular, under the indicated hypotheses, the limits above exist.

Proof: Let $u_1 \in I \setminus \{c\}$ be such that $h(u_1) < \alpha + \epsilon$; such u_1 exists since (by definition of "inf") $\alpha + \epsilon$ is not a lower bound of range(h). Let $r = u_1 - c$; thus $r > 0$ and $u_1 = c + r$. Then for all u with $c < u < c + r$ we have $\alpha \le h(u) \le h(u_1) < \alpha + \epsilon$. Thus for all $u \in I \setminus \{c\}$ for which $|u - c| < r$, we have $|h(u) - \alpha| = h(u) - \alpha < \epsilon$. Since ϵ was arbitrary, this establishes that $\lim_{u \to c} h(u) = \alpha$.

This proves (i). Statement (ii) can be deduced by applying (i) to the function $-h$.
■

Corollary 6.30 *Let $f \in \mathcal{B}([a, b])$ be given. Then $\lim_{\delta \to 0} L_\delta(f)$ and $\lim_{\delta \to 0} U_\delta(f)$ both exist, and*

$$\lim_{\delta \to 0} L_\delta(f) \ \le \ \lim_{\delta \to 0} U_\delta(f). \tag{6.19}$$

Proof: Let $h_1, h_2 : (0, \infty) \to \mathbf{R}$ be the functions defined in Proposition 6.28. By the Proposition, each of these functions is monotone and bounded, so Lemma 6.29 implies that the limits in (6.19) exist. Since both these limits exist, and $L_\delta(f) \le U_\delta(f)$ for each $\delta > 0$, the inequality (6.19) follows. ■

Another general lemma that will be used shortly is the following:

Lemma 6.31 *Let $X \subset \mathbf{R}$ be a nonempty, bounded set. Let $r > 0$ and assume that for all $x_1, x_2 \in X$ we have $|x_1 - x_2| \le r$. Then $\sup(X) - \inf(X) \le r$.*

Proof: Let $\epsilon > 0$, and let $x_1, x_2 \in X$ be such that $x_1 > \sup(X) - \epsilon$ and $x_2 < \inf(X) + \epsilon$; such x_1, x_2 exist since $\sup(X)$ and $\inf(X)$ are, respectively, the *least* upper bound and *greatest* lower bound of X. Then $\sup(X) < x_1 + \epsilon$ and $\inf(X) > x_2 - \epsilon$, so $\sup(X) - \inf(X) < x_1 + \epsilon - (x_2 - \epsilon) = x_1 - x_2 + 2\epsilon \le r + 2\epsilon$. Since ϵ was arbitrary, the result follows. ■

We can now recast integrability in terms of the limits in Corollary 6.30:

Theorem 6.32 *For each $f \in \mathcal{B}([a,b])$, the following are equivalent:*

(i) *f is integrable over $[a,b]$.*

(ii) $\lim_{\delta\to 0} L_\delta(f) = \lim_{\delta\to 0} U_\delta(f)$.

(iii) $\lim_{\delta\to 0}(U_\delta(f) - L_\delta(f)) = 0$.

(iv) *For every $\epsilon > 0$ there exists $\delta > 0$ such that for all $S_1, S_2 \in \mathcal{S}_\delta(f)$, $|S_2 - S_1| < \epsilon$.*

In the integrable case,

$$\int_a^b f = \lim_{\delta\to 0} L_\delta(f) = \lim_{\delta\to 0} U_\delta(f). \tag{6.20}$$

Proof: Let $f \in \mathcal{B}([a,b])$ be given, and recall from Remark 6.14 that the definition of "$\int_a^b f = A$" can be rewritten as: for all $\epsilon > 0$ there exists $\delta > 0$ such that $\mathcal{S}_\delta(f) \subset (A - \epsilon, A + \epsilon)$. We will establish the equivalence of (i), (ii), (iii) by showing that each of (i) and (iii) is equivalent to (ii).

(i) $\Longrightarrow$ (ii), plus last sentence of Proposition.

Assume that f is integrable over $[a,b]$, and let $A = \int_a^b f$. Let $\epsilon > 0$ be given, and let $\delta > 0$ be such that $\mathcal{S}_\delta(f) \subset (A-\epsilon, A+\epsilon)$. Then the infimum $L_\delta(f)$ of $\mathcal{S}_\delta(f)$ and supremum $U_\delta(f)$ of $\mathcal{S}_\delta(f)$ lie in $[A - \epsilon, A + \epsilon]$, implying that $|L_\delta(f) - A| \le \epsilon$ and $|U_\delta(f) - A| \le \epsilon$. Since ϵ was arbitrary, it follows that $\lim_{\delta\to 0} L_\delta(f) = A = \lim_{\delta\to 0} U_\delta(f)$. This implies both statement (ii) and the last sentence of the Proposition.

(ii) $\Longleftrightarrow$ (iii)

By Corollary 6.30, both $\lim_{\delta\to 0} U_\delta(f)$ and $\lim_{\delta\to 0} L_\delta(f)$ exist. Hence $\lim_{\delta\to 0}(U_\delta(f) - L_\delta(f)) = \lim_{\delta\to 0} U_\delta(f) - \lim_{\delta\to 0} L_\delta(f)$. The equivalence of (ii) and (iii) is immediate from this equality.

(ii) $\implies$ both (i) and (iv)

Assume the equality in (ii) holds, and let A be the (common) value of the indicated limits. Let $\epsilon > 0$ be given. Let $\delta > 0$ be such that $|L_\delta(f) - A| < \epsilon$ and $|U_\delta(f) - A| < \epsilon$; such δ exists since $\lim_{\delta\to 0} L_\delta(f) = A = \lim_{\delta\to 0} U_\delta(f)$. Then for all $S \in \mathcal{S}_\delta(f)$,

$$A - \epsilon < L_\delta(f) \le S \le U_\delta(f) < A + \epsilon,$$

implying that $\mathcal{S}_\delta(f) \subset (A - \epsilon, A + \epsilon)$. Hence $\int_a^b f = A$, so (i) is true. Furthermore, for all $S_1, S_2 \in \mathcal{S}_\delta(f)$ we have $|S_2 - S_1| < 2\epsilon$. Since ϵ was arbitrary, this establishes (iv).

(iv) $\implies$ (iii)

Assume that (iv) holds. Let $\delta_0 > 0$ be such that for all $S_1, S_2 \in \mathcal{S}_{\delta_0}(f)$, $|S_2 - S_1| < \epsilon$. Then by Lemma 6.31, $0 \le U_{\delta_0}(f) - L_{\delta_0}(f) = \sup(\mathcal{S}_{\delta_0}(f)) - \inf(\mathcal{S}_{\delta_0}(f) \le \epsilon$. The monotonicities of the functions $\delta \mapsto L_\delta(f), \delta \mapsto U_\delta(f)$ established in Proposition 6.28 imply that for all $\delta \in (0, \delta_0]$, we have

$$L_{\delta_0}(f) \le L_\delta(f) \le U_\delta(f) \le U_{\delta_0}(f).$$

Hence for all such δ, $0 \le U_\delta(f) - L_\delta(f) \le U_{\delta_0}(f) - L_{\delta_0}(f) \le \epsilon$. Since, by Corollary 6.30, $\lim_{\delta\to 0} U_\delta(f)$ and $\lim_{\delta\to 0} L_\delta(f)$ both exist, so does $\lim_{\delta\to 0}(U_\delta(f) - L_\delta(f))$, and by the basic order-property established in MAA 4211 for limits of real-valued functions,

$$0 \le \lim_{\delta\to 0}(U_\delta(f) - L_\delta(f)) \le \epsilon. \tag{6.21}$$

Since ϵ was arbitrary, (6.21), this implies that $\lim_{\delta\to 0}(U_\delta(f) - L_\delta(f)) = 0$. ■

Remark 6.33 Since every Riemann-integrable function is bounded, the (previously unproven) Proposition 6.10 amounts to the "(i) $\implies$ (iv)" implication in Theorem 6.32. ▲

Remark 6.34 Statement (iv) in Theorem 6.32 can be thought of, loosely, as a "Cauchy criterion for the convergence of Riemann sums" (with "convergence of Riemann sums" interpreted heuristically, since the set of Riemann sums of a function f on $[a, b]$ is not a sequence). ▲

Remark 6.35 The equivalence of (i) and (iv) in Theorem 6.32 can be proven without any use of upper and lower sums. We will give such a proof later, when we discuss integration of vector-valued functions. ▲

The following characterization of upper and lower sums, worthwhile for its own sake, simplifies our work when we apply Theorem 6.32 to compute integrals or prove integrability.

Proposition 6.36 *Let $f \in \mathcal{B}([a,b])$ be given, and let $P = \{x_0, \ldots, x_N\}$ be a partition of $[a,b]$. For $1 \le j \le N$, let*

$$m_j = \inf\{f(x) : x_{j-1} \le x \le x_j\} \quad \textit{and} \quad M_j = \sup\{f(x) : x_{j-1} \le x \le x_j\}.$$

Then

$$L(f;P) = \sum_{j=1}^{N} m_j \Delta_j \quad \textit{and} \quad U(f;P) = \sum_{j=1}^{N} M_j \Delta_j\,. \tag{6.22}$$

Proof: Since $f(x) \le M_j$ for all $x \in [x_{j-1}, x_j], 1 \le j \le N$, it is clear that for any pointing T of P we have $S(f;P,T) \le \sum_j M_j \Delta_j$, so $\sum_{j=1}^{N} M_j \Delta_j$ is an upper bound for $\mathcal{S}(f;P)$.

Now let $\epsilon > 0$ be given. For each $j \in \{1, 2, \ldots, N\}$, let $t_j \in [x_{j-1}, x_j]$ be such that $f(t_j) > M_j - \frac{\epsilon}{b-a}$; such t_j exists by the definition of M_j. Let $T = \{t_1, \ldots, t_N\}$. Then T is a pointing of P, and

$$\begin{aligned} S(f;P,T) = \sum_{j=1}^{N} f(t_j)\Delta_j > \sum_j \left(M_j - \frac{\epsilon}{b-a}\right)\Delta_j \quad &= \left(\sum_j M_j \Delta_j\right) - \frac{\epsilon}{b-a}\sum_j \Delta_j \\ &= \left(\sum_j M_j \Delta_j\right) - \epsilon. \end{aligned}$$

Hence no number smaller than $\sum_j M_j \Delta_j$ is an upper bound for $\mathcal{S}(f;P)$. Thus $\sum_j M_j \Delta_j$ is the least upper bound (= supremum) of $\mathcal{S}(f;P)$, yielding the second equality in (6.22). A similar argument (left to the student) establishes the first equality. ■

Definition 6.37 Let A be a set and let $B \subset A$. The *characteristic function*[3] *of* B *(viewed as a subset* A*)* is the function $\chi_B : A \to \mathbf{R}$ defined by

$$\chi_B(p) = \begin{cases} 1 \text{ if } p \in B, \\ 0 \text{ if } p \notin B. \end{cases}$$

▲

For example, the function in Example 6.12 is simply the restriction of $\chi_{\mathbf{Q}}$ to $[a,b]$ (regarding $\mathbf{Q}$ as a subset of $\mathbf{R}$). The characteristic function of any interval (including a one-point interval) is an example of a *step function*; see Definition 6.42.

The next example and our proof of the next proposition illustrate how Theorem 6.32 can be used.

[3]In some areas of mathematics, such as probability, characteristic functions are called *indicator functions*, and the notation $\mathbf{1}_B$ is used instead of χ_B.

Example 6.38 Fix $c \in [a, b]$ and let $f = \chi_{\{c\}} : [a, b] \to \mathbf{R}$. Let $\delta > 0$ and let $P = \{x_0, \ldots x_N\})$ be a partition of $[a, b]$ of width less than δ. With notation as in Proposition 6.36, $m_j = 0$ for every j, and $M_j = 0$ unless $c \in [x_{j-1}, x_j]$; the latter can happen for at most two values of j. When nonzero, the value of M_j is 1. Hence, using (6.22), we have

$$0 = L(f; P) \leq U(f; P) < 2\delta.$$

Since this holds for all $P \in \mathcal{P}_\delta([a, b])$, it follows that

$$0 \leq L_\delta(f) \leq U_\delta(f) \leq 2\delta.$$

Using the Squeeze Theorem and Theorem 6.32, we conclude that

$$\int_a^b f = 0.$$

▲

Proposition 6.39 Suppose that $a \leq c < d \leq b$. Then the characteristic function $\chi_{(c,d)} : [a, b] \to \mathbf{R}$ is integrable, and

$$\int_a^b \chi_{(c,d)} = d - c.$$

Proof: To streamline notation in this proof, let $f = \chi_{(c,d)}$. We will show that $\lim_{\delta \to 0} L_\delta(f) = \lim_{\delta \to 0} U_\delta(f) = d - c$. For this, it suffices to restrict attention to δ less than any fixed, positive number; in particular, to $\delta < d - c$.

Let $\delta \in (0, d - c)$ be given, let $P = \{x_0, \ldots, x_N\}$ be a partition of $[a, b]$ of width less than δ, and let j_1, j_2 be the unique indices in $\{1, \ldots, N\}$ such that $c \in [x_{j_1-1}, x_{j_1})$ and $d \in (x_{j_2-1}, x_{j_2}]$. Then $x_{j_1-1} \leq c < d \leq x_{j_2}$, so $j_1 - 1 < j_2$; equivalently, $j_1 \leq j_2$. If $j_2 = j_1$ then $\delta > x_{j_1} - x_{j_1-1} = x_{j_2} - x_{j_1-1} > \delta$, a contradiction, so in fact we have $j_1 < j_2$; equivalently, $j_1 \leq j_2 - 1$. Hence $c < x_{j_1} \leq x_{j_2-1} < d$, so x_j lies in (c, d) if $j_1 \leq j < j_2$.

For $j \in \{1, \ldots, N\}$, let m_j, M_j be as in Proposition 6.36; observe that each of these numbers is either 0 or 1. Then for each $j \in \{1, \ldots, N\}$ we have the following:

If $j < j_1$ or $j > j_2$ then $[x_{j-1}, x_j] \cap (c, d) = \emptyset$, so $m_j = M_j = 0$.
If $j_1 < j < j_2$ then $[x_{j-1}, x_j] \subset (c, d)$, so $m_j = M_j = 1$.
If $j = j_1$ or $j = j_2$ then $0 \leq m_j \leq M_j \leq 1$.

(As will be seen, more precise information about the indices j_1, j_2 is unnecessary, so we do not waste time on that.) Therefore, using (6.22),

$$\begin{aligned} U(f; P) \leq \sum_{j=j_1}^{j_2} \Delta_j(P) = x_{j_2} - x_{j_1-1} &= (x_{j_2} - d) + (d - c) + (c - x_{j_1-1}) \\ &< (x_{j_2} - x_{j_2-1}) + (d - c) + (x_{j_1} - x_{j_1}) \\ &< \delta + (d - c) + \delta \\ &= d - c + 2\delta, \end{aligned}$$

and, since $x_{j_2-1} > x_{j_2} - \delta \geq d - \delta$ and $x_{j_1} < x_{j_1-1} + \delta \leq c + \delta$,

$$L(f;P) \geq \sum_{j=j_1+1}^{j_2-1} \Delta_j = x_{j_2-1} - x_{j_1} \geq (d-\delta) - (c+\delta) = (d-c) - 2\delta.$$

(For the case in which $j_1 + 1 > j_2 - 1$—which can happen only if $j_1 + 1 = j_2$, since $j_1 < j_2$—recall that the notation "$\sum_{j=m}^{n}$" means "the sum over all j satisfying $m \leq j \leq n$ if this index-set is nonempty, and 0 if this index-set is empty.")

Since the above inequalities hold for all $P \in \mathcal{P}_\delta([a,b])$, it follows that

$$(d-c) - 2\delta \leq L_\delta(f) \leq U_\delta(f) \leq (d-c) + 2\delta.$$

The result now follows from the Squeeze Theorem and Theorem 6.32. ■

Proposition 6.40 *Let $f, g : [a,b] \to \mathbf{R}$. Assume that f is integrable and that g differs from f at only finitely many points (i.e. that there are only finitely many $x \in [a,b]$ for which $g(x) \neq f(x)$). Then g is integrable, and $\int_a^b g = \int_a^b f$.*

Proof: Let $x_1, \ldots, x_n$ be the values of x for which $g(x) \neq f(x)$ (we may assume there is at least one such value, since otherwise $g = f$ and the conclusion is trivial). Let $h = g - f$. Then $h(x) = 0$ for all $x \notin \{x_1, \ldots, x_n\}$, so h is a linear combination of the functions $\chi_{\{x_1\}}, \ldots, \chi_{\{x_n\}}$; specifically, $h = \sum_i c_i \chi_{\{x_i\}}$ where $c_i = h(x_i)$. By Proposition 6.17 and Example 6.38, h is integrable and

$$\int_a^b h = \sum_i c_i \int_a^b \chi_{\{x_i\}} = \sum_i c_i \cdot 0 = 0.$$

But $g = f + h$, so g is the sum of two integrable functions. Using Proposition 6.17, the conclusion follows. ■

Corollary 6.41 *Let $I \subset [a,b]$ be an interval, and let $c \leq d$ be the left and right endpoints, respectively, of $\bar{I}$. Then $\chi_I : [a,b] \to \mathbf{R}$ is integrable and $\int_a^b \chi_I = d - c$.*

Corollary 6.41 follows easily from Proposition 6.39, Example 6.38, and "linearity of the integral" (Proposition 6.17). This corollary is also a special case of the upcoming Proposition 6.44 (and is proven essentially the same way), so we omit writing a separate proof here.

Definition 6.42 A function $f : [a, b] \to \mathbf{R}$ is a *step function* if there exists a partition $\{x_0, \ldots x_N\}$ of $[a, b]$ such that for each $j \in \{1, \ldots, N\}$, f is constant on the open interval (x_{j-1}, x_j). ▲

Lemma 6.43 *If $f : [a, b] \to \mathbf{R}$ is a step-function, then f is a linear combination of characteristic functions of intervals.*

Proof: Let $f : [a, b] \to \mathbf{R}$ be a step-function, let $P = \{x_0, \ldots x_N\} \in \mathcal{P}([a, b])$ be such that for each $j \in \{1, \ldots, N\}$, $f|_{(x_{j-1}, x_j)}$ is constant, and for each such j let c_j denote the (constant) value of $f|_{(x_{j-1}, x_j)}$. Then, as is easily verified,

$$f = \sum_{j=1}^{N} c_j \chi_{(x_{j-1}, x_j)} + \sum_{j=0}^{N} f(x_j) \chi_{\{x_j\}}, \tag{6.23}$$

a linear combination of characteristic functions of intervals. ■

Exercise 6.4 Prove the converse of Lemma 6.43. (Note that in the phrase "linear combination of characteristic functions of intervals", it is not given that the intervals do not overlap.)

Proposition 6.44 *Let $f : [a, b] \to \mathbf{R}$ be a step-function, and let P, N and $c_1, \ldots, c_N$ be as in Lemma 6.43. Then f is integrable and*

$$\int_a^b f = \sum_{j=1}^{N} c_j \Delta_j(P).$$

Proof: This follows from equation 6.23, Proposition 6.17, and Example 6.38. ■

Proposition 6.45 ("Step-function lemma") *A function $f \in \mathcal{B}([a, b])$ is integrable if and only if for each $\epsilon > 0$, there exists a partition P of $[a, b]$ such that*

$$U(f; P) - L(f; P) < \epsilon. \tag{6.24}$$

Remark 6.46 The strength of Proposition 6.45 is that *there is no reference to the width of the partition P*. This makes the Proposition much simpler to apply than many of our results up till now. ▲

Proof of Proposition 6.45:

($\Longrightarrow$) Assume that f is integrable. Then by Theorem 6.32, $\lim_{\delta\to 0}(U_\delta(f)-L_\delta(f)) = 0$. Let $\epsilon > 0$ be given, and let $\delta > 0$ be such that $U_\delta(f) - L_\delta(f) < \epsilon$; such δ exists since the above limit is 0. Let P be any partition of width less than δ. Then, by (6.16), we have $U(f; P) - L(f; P) \le U_\delta(f) - L_\delta(f) < \epsilon$.

($\Longleftarrow$) Assume that for each $\epsilon > 0$, there exists a partition P of $[a, b]$ such that $U(f; P) - L(f; P) < \epsilon$.

Let $\epsilon > 0$ be given, and let $P = \{x_0, \ldots, x_N\} \in \mathcal{P}([a, b])$ be such that $U(f; P) - L(f; P) < \epsilon$. For $1 \le j \le N$ let M_j and m_j be as in Proposition 6.36. Let $m = \inf\{f(x) : x \in [a, b]\}$ and $M = \sup\{f(x) : x \in [a, b]\}$ Define functions $f_1, f_2 : [a, b] \to \mathbf{R}$ by

$$\begin{aligned} f_1 &= \sum_{j=1}^{N} m_j \chi_{(x_{j-1},x_j)} + m\chi_P = \sum_{j=1}^{N} m_j \chi_{(x_{j-1},x_j)} + \sum_{j=0}^{N} m\chi_{\{x_j\}} \\ f_2 &= \sum_{j=1}^{N} M_j \chi_{(x_{j-1},x_j)} + M\chi_P = \sum_{j=1}^{N} M_j \chi_{(x_{j-1},x_j)} + \sum_{j=0}^{N} M\chi_{\{x_j\}}. \end{aligned}$$

Observe also that

$$f_1(x) \le f(x) \le f_2(x) \tag{6.25}$$

for every $x \in [a, b]$. Furthermore, f_1 and f_2 are step functions, hence are integrable, and from Proposition 6.44 we have

$$\int_a^b f_1 = \sum_{j=1}^{N} m_j \Delta_j(P) = L(f; P) \tag{6.26}$$

and

$$\int_a^b f_2 = \sum_{j=1}^{N} M_j \Delta_j(P) = U(f; P). \tag{6.27}$$

Let $\delta > 0$ be such that

$$\mathcal{S}_\delta(f_1) \subset (L(f; P) - \epsilon, L(f; P) + \epsilon) \tag{6.28}$$

and

$$\mathcal{S}_\delta(f_2) \subset (U(f; P) - \epsilon, U(f; P) + \epsilon); \tag{6.29}$$

such δ exists by (6.26)–(6.27) (see Remark 6.14). Let (Q, T) be any pointed partition of $[a, b]$ of width less than δ. From (6.25) and the definition of "Riemann sum", it is immediate that $S(f_1; Q, T) \le S(f; Q, T) \le S(f_2; Q, T)$. But $S(f_1; Q, T) \in \mathcal{S}_\delta(f_1)$ and $S(f_2; Q, T) \in \mathcal{S}_\delta(f_2)$. Thus, using (6.28)–(6.29), we have

$$L(f; P) - \epsilon < S(f_1; Q, T) \le S(f; Q, T) \le S(f_2; Q, T) < U(f; P) + \epsilon. \tag{6.30}$$

Now let $S_1, S_2 \in \mathcal{S}_\delta(f)$. By (6.30), both S_1 and S_2 lie in the interval $(L(f;P) - \epsilon, U(f;P) + \epsilon)$. Hence

$$|S_2 - S_1| \leq (U(f;P) + \epsilon) - (L(f;P) - \epsilon) = (U(f;P) - L(f;P)) + 2\epsilon < 3\epsilon$$

(using our initial hypothesis). Theorem 6.32 (specifically, the implication "(iii) $\Longrightarrow$ (i)") therefore implies that f is integrable. ■

Remark 6.47 Our proof shows that Proposition 6.45 is equivalent to the more easily visualized statement: A function $f \in \mathcal{B}([a,b])$ is integrable if and only if for each $\epsilon > 0$, there exist step-functions $f_1, f_2 : [a,b] \to \mathbf{R}$ such that $f_1(x) \leq f(x) \leq f_2(x)$ for all $x \in [a,b]$ (f is "squeezed" between f_1 and f_2) such that $\int_a^b (f_2 - f_1) = \int_a^b f_2 - \int_a^b f_1 < \epsilon$. ▲

Exercise 6.5 For any real-valued function f, the *positive part of* f, denoted f_+, and *negative part of* f, denoted f_-, are defined by $f_+(x) = \max\{f(x), 0\}$ and $f_-(x) = -\min\{f(x), 0\}$. (Thus both f_+ and f_- are non-negative, and $f = f_+ - f_-$ [why?].)

Parts (a) and (b) below can be done in either order: whichever part you do first, you may use to help you do the other part quickly. (But, obviously, you may not resort to circular reasoning.) To see how the result of (b) can be used to help with (a), compare the function f_+ with $|f| + f$.

(a) Prove that if f is integrable on $[a,b]$, then so are f_+ and f_-.

(b) Prove that if f is integrable on $[a,b]$ then so is $|f|$ (the function $x \mapsto |f(x)|$), and

$$\left| \int_a^b f \right| \leq \int_a^b |f|. \tag{6.31}$$

6.4 Upper and lower integrals

This section is optional reading.

Definition 6.48 For any $f : [a,b] \to \mathbf{R}$, we define the *lower* and *upper Riemann integrals* of f over $[a,b]$ to be

$$\underline{\int_a^b} f = \sup\{L(f;P) : P \in \mathcal{P}([a,b])\},$$
$$\overline{\int_a^b} f = \inf\{U(f;P) : P \in \mathcal{P}([a,b])\},$$

respectively. We will frequently omit "Riemann" from this terminology, and may write the lower and upper integrals using dummy-variable notation, e.g. "$\underline{\int_a^b} f(x)\,dx$" for "$\underline{\int_a^b} f$." ▲

In words: the lower integral is the *supremum* of *lower sums*, while the the upper integral is the *infimum* of *upper sums*

Now consider any fixed, arbitrary, $f \in \mathcal{B}([a,b])$. Using Proposition 6.28 and Lemma 6.29, we can express the limits of $U_\delta(f)$ and $L_\delta(f)$ (as $\delta \to 0$) as follows:

$$\begin{aligned}\lim_{\delta\to 0} U_\delta(f) &= \inf_{\delta>0}(\sup\{U(f;P) : P \in \mathcal{P}_\delta([a,b])\}); \\ \lim_{\delta\to 0} L_\delta(f) &= \sup_{\delta>0}(\{\inf\{U(f;P) : P \in \mathcal{P}_\delta([a,b])\}).\end{aligned}$$

For each $\delta > 0$ and $P \in \mathcal{P}_\delta([a,b])$, we have $U(f;P) \geq L(f;P)$. Hence, using Lemma 6.24,

$$\begin{aligned}\inf_{\delta>0}(\sup\{U(f;P) : P \in \mathcal{P}_\delta([a,b])\}) &\geq \inf_{\delta>0}(\inf\{U(f;P) : P \in \mathcal{P}_\delta([a,b])\}) \\ &= \inf\left(\bigcup_{\delta>0}\{\{U(f;P) : P \in \mathcal{P}_\delta([a,b])\}\right) \\ &= \inf\{U(f;P) : P \in \mathcal{P}([a,b]\} \\ &= \overline{\int_a^b} f.\end{aligned}$$

Thus

$$\lim_{\delta\to 0} U_\delta(f) \geq \overline{\int_a^b} f, \tag{6.32}$$

and similarly

$$\lim_{\delta\to 0} L_\delta(f) \leq \underline{\int_a^b} f. \tag{6.33}$$

We will show that the inequalities (6.32)–(6.33) can be sharpened to equalities when f is integrable, but some preliminary work is needed first.

Definition 6.49 Let P and Q denote partitions of $[a,b]$. We say Q is a *refinement* of P, or that Q *refines* P, if $P \subset Q$. The *common refinement* of P and Q is $P \cup Q$. ▲

Lemma 6.50 *Let $f : [a,b] \to \mathbf{R}$ be given.*

(i) Let P and Q be partitions of $[a,b]$, and assume that Q refines P. Then

$$L(f;P) \leq L(f;Q) \leq U(f;Q) \leq U(f;P). \tag{6.34}$$

(ii) For any *partitions P, Q of $[a,b]$,*

$$L(f;P) \leq U(f;Q). \tag{6.35}$$

(iii) The upper and lower integrals satisfy

$$\underline{\int_a^b} f(x)\,dx \;\leq\; \overline{\int_a^b} f(x)\,dx. \tag{6.36}$$

Sketch of proof. (i) The middle inequality in (6.34) is simply (6.9). The first and third inequalities can be reduced to the case in which $P = \{a,b\}$ and $Q = \{a,c,b\}$, where the result is quickly established by comparing the Riemann sums associated with Q with those associated with P.

(ii) Let P and Q be partitions of $[a,b]$, and let R be their common refinement. Applying (i) twice, we obtain

$$L(f;P) \leq L(f;R) \leq U(f;R) \leq U(f;Q).$$

Hence (6.35) holds.

(iii) For each partition Q, taking the supremum over all partitions P in (6.35) yields

$$\underline{\int_a^b} f(x)\,dx \leq U(f;Q).$$

Taking the infimum over Q then yields (6.36). ■

Theorem 6.51 *A bounded function $f : [a,b] \to \mathbf{R}$ is integrable if and only if*

$$\underline{\int_a^b} f \;=\; \overline{\int_a^b} f. \tag{6.37}$$

In the integrable case,

$$\underline{\int_a^b} f \;=\; \overline{\int_a^b} f = \int_a^b f. \tag{6.38}$$

Proof: Let $f \in \mathcal{B}([a,b])$.

First suppose that f is integrable. Then by Proposition 6.9, f is bounded, so our analysis leading to (6.32)–(6.33) applies. These inequalities, together with (6.36), yield

$$\lim_{\delta\to 0} L_\delta(f) \leq \underline{\int_a^b} f \;\leq\; \overline{\int_a^b} f \;\leq\; \lim_{\delta\to 0} U_\delta(f). \tag{6.39}$$

But from Theorem 6.32, since f is integrable, the leftmost and rightmost expressions in (6.39) are equal to each other and to $\int_a^b f$. Hence

$$\int_a^b f = \lim_{\delta\to 0} L_\delta(f) = \underline{\int_a^b} f \;=\; \overline{\int_a^b} f \;=\; \lim_{\delta\to 0} U_\delta(f).$$

This proves that the upper and lower integrals are equal, and establishes (6.38).

Conversely, suppose that $\underline{\int_a^b} f = \overline{\int_a^b} f$, and let A denote the value of these quantities. Let $\epsilon > 0$ be given. Let P, Q be partitions of $[a.b]$ such that $L(f;P) > A - \epsilon$ and $U(f;Q) < A + \epsilon$; such partitions exist by the definition of lower and upper integrals. Let R be the common refinement of P and Q. Then, as in the proof of Lemma 6.50(ii), we have $L(f;P) \le L(f;R) \le U(f;R) \le U(f;Q)$. Hence

$$A - \epsilon < L(f;R) \le U(f;R) < A + \epsilon,$$

implying $U(f;R) - L(f;R) < 2\epsilon$. Since ϵ was arbitrary, Proposition 6.45 then implies that f is integrable. ∎

Among the implications of Theorem 6.51 is that *if* $f : [a,b] \to \mathbf{R}$ is integrable, then the inequalities in (6.32) and (6.33) can be replaced by equalities. The student may well wonder whether equality holds in (6.32) and (6.33) even without the assumption of integrability. The answer is yes (this is one of several unrelated results each of which is sometimes given the name "Darboux's Theorem"), but the proof is not obvious, and we do not give it in these notes. We refer the interested student to [10, Section 18.2,Theorem VIII].

Remark 6.52 (Two approaches to the Riemann integral) Because Theorem 6.51 is true, equation (6.37) can be taken as the *definition* of "a bounded function f is (Riemann) integrable on $[a,b]$", in place of Definition 6.6, without changing either the set of functions being called "integrable" or the values of their integrals. If we use (6.37) to *define* what "integrable" means, then Theorem 6.51 yields the second sentence of Definition 6.6 as a *theorem* rather than a definition. Many (probably most) textbooks use this alternate definition of integrability, often phrased without any mention of Riemann sums (taking (6.22) to be the *definition* of $L(f;P)$ and $U(f;P)$). This approach has several advantages—for example, the definition of integrability is *much* simpler (there is no ϵ or δ; the width of a partition is never even mentioned), and many proofs can be done more efficiently.

However, there are also disadvantages[4] of using (6.37) instead of Definition 6.6 to define integrability. The chief *mathematical* disadvantage of the approach based on (6.37)

[4]Most of the "disadvantages" referred to here are pedagogical in nature, so should properly be called "disadvantages *in the opinion of the author of these notes*", but repeating such a mouthful in place

is that the generalization to integrals of vector-valued functions is less natural (especially for functions taking values in an infinite-dimensional vector space). The other potential disadvantages are primarily pedagogical. One is that unless a proof of Theorem 6.51 is provided, the notion of "$\int_a^b f(x)\,dx$" in this approach does not clearly reduce to the notion that students learn in Calculus 1 (and again in Calculus 3, generalized to definite integrals of functions of two or three variables)—a notion that is completely correct, but that is usually not given a precise statement in Calculus 1-2-3 because students are not yet equipped to understand or appreciate the precise statement. Definition 6.6 is *exactly* the Calculus 1-2-3 notion of integrability, just defined precisely. It is this notion, rather than equation (6.37), on which all quantities defined through integrals in physics and other sciences are based.[5] Without Theorem 6.51, it is not clear that the "upper integral = lower integral" definition of integrability leads to the same notions of integration, or values of integrals, conceptualized in Calculus 1 (whether or not (6.22) is used to define upper and lower sums). Thus, some mathematicians find presentations of the Riemann integral that take (6.37) as definition, but do not include a proof of Theorem 6.51 (e.g. the presentation in [7]), to be unsatisfying. But when presentations that take (6.37) as definition *do* include a proof of Theorem 6.51 (as in [6, Theorem 6.14] and [10, Section 18.2]), some of the efficiency initially gained from the upper-integral/lower-integral definition is lost.

When (6.22) is used to define upper and lower sums, in addition to using (6.37) to define integrability, there is another efficiency-gain (the need to prove Proposition 6.36 is avoided), but offsetting are additional pedagogical disadvantages. One is that all connection to Riemann sums has been removed (unless prominent mention is made elsewhere in the presentation), putting even more distance between the integral defined this way and the integral as conceptualized in Calculus 1-2-3 and in the sciences. Another is that, using (6.22), we cannot even *define* upper and lower sums (and therefore upper and lower integrals), even within the extended reals, without restricting attention to functions that are at least *semi-bounded*: bounded above or bounded below. (In contrast, upper and lower sums as defined in Definition 6.22 *always* exist in the extended reals; no boundedness assumptions are needed.) Usually, to simplify presentations based on (6.22) and (6.37), a restriction is made to functions that are *bounded*, not just semi-bounded. Thus one loses Proposition 6.9. Instead of the non-integrability of unbounded functions being a consequence of the *concept* of the Riemann integral, unbounded functions are simply removed from consideration from the start (and students may reasonably wonder,

of "disadvantages" everywhere would have made this discussion hard to read. The judgment of what is pedagogically better or worse is highly subjective. Every instructor is forced to make pedagogical choices, some aspect of which would be deemed disadvantageous by instructors making different choices. No criticism is intended of anyone whose pedagogical choices are different from those of the author of these notes.

[5] It is exactly these quantities in physics, including vector-valued integrals, for which the notion of "integral" was originally developed; calculus was invented in order to provide a mathematical description of physical laws. Examples of quantities in physics that, to this day, are understood by starting with Riemann sums, include work, centers of mass, moments of inertia, hydrostatic force, all line integrals in electricity and magnetism (E&M), and all flux integrals in E&M and in fluid dynamics.

"Why?"). Thus one of the chief deficiencies of the Riemann integral (as compared with the Lebesgue integral) cannot be *demonstrated*; it's been defined away. ▲

6.5 The integrability of continuous functions

Theorem 6.53 (Continuous functions are integrable) *If $f : [a, b] \to \mathbf{R}$ is continuous, then f is integrable on $[a, b]$.*

Proof: Let $f : [a, b] \to \mathbf{R}$ be a continuous function. Since $[a, b]$ is compact, f is bounded. Therefore, by Proposition 6.45, to prove that f is integrable it suffices to show that for each $\epsilon > 0$, there exists a partition P of $[a, b]$ such that $U(f; P) - L(f; P) < \epsilon$.

Let $\epsilon > 0$ be given. Recall from MAA 4211 that every continuous function on a compact space is uniformly continuous. Since $[a, b]$ is compact and f is continuous, it follows that f is uniformly continuous. Let $\delta > 0$ be such that if $x, y \in [a, b]$ and $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon/(b - a)$; such δ exists since f is uniformly continuous. Let $P = \{x_0, \dots, x_N\} \in \mathcal{P}_\delta([a, b])$; such P exists by Remark 6.5.

Recall also from MAA 4211 that every continuous real-valued function on a compact space attains a maximum value and a minimum value. In particular, this applies to $f|_{[x_{j-1},x_j]}$ for each $j \in \{1, \dots, N\}$. For each such j let m_j, M_j denote, respectively, the minimum and maximum values of $f|_{[x_{j-1},x_j]}$, and let $x_j', x_j'' \in [x_{j-1}, x_j]$ be such that $f(x_j') = m_j$ and $f(x_j'') = M_j$. Then, for each $j \in \{1, \dots, N\}$, we have $|x_j' - x_j''| \leq \mathrm{wid}(P) < \delta$, so

$$M_j - m_j = f(x_j'') - f(x_j') < \frac{\epsilon}{b - a} \ .$$

But by Proposition 6.36, $L(f; P) = \sum_j m_j \Delta_j$ and $U(f; P) = \sum_j M_j \Delta_j$ (where $\Delta_j = \Delta_j(P)$). Hence

$$U(f; P) - L(f; P) = \sum_j (M_j - m_j)\Delta_j < \sum_j \frac{\epsilon}{b - a}\Delta_j = \frac{\epsilon}{b - a}\sum_j \Delta_j = \epsilon.$$

Since ϵ was arbitrary, we conclude from Proposition 6.45 that f is integrable. ■

Exercise 6.6 (a) Assume that $f : [a, b] \to \mathbf{R}$ is continuous, that $f(x) \geq 0$ for all $x \in [a, b]$, and that $f(x) > 0$ for some $x \in [a, b]$. Prove that $\int_a^b f > 0$.

(b) Assume that $f, g : [a, b] \to \mathbf{R}$ are continuous, that $f(x) \geq g(x)$ for all $x \in [a, b]$, and that $f(x) > g(x)$ for some $x \in [a, b]$. As a corollary of part (a), prove that $\int_a^b f > \int_a^b g$.

Remark 6.54 In particular, if $f : [a, b] \to \mathbf{R}$ is continuous and $f(x) > 0$ for all $x \in [a, b]$, then $\int_a^b f > 0$. (See Remark 6.20.) ▲

6.6 Additivity of the integral

The Riemann integral has an additivity property (unrelated to linearity) expressed by the following proposition.

Proposition 6.55 (Additivity of the integral) *Suppose $a < c < b$. A function $f : [a, b] \to \mathbf{R}$ is integrable on $[a, b]$ if and only if it is integrable on both $[a, c]$ and $[c, b]$. In the integrable case,*

$$\int_a^b f = \int_a^c f + \int_c^b f. \tag{6.40}$$

Remark 6.56 As mentioned in Section 6.0, equation (6.40) reflects the principle that "integration is about adding stuff up": the "amount of stuff" between a and b is the "amount of stuff" between a and c plus the "amount of stuff" between c and b.

Proof of Proposition 6.55: Let $f_1 = f|_{[a,c]}$ and $f_2 = f|_{[c,b]}$.

First suppose that f_1 and f_2 are integrable, and let A and C, respectively, denote their integrals. Since f_1 and f_2 are integrable, they are bounded; hence so is f. Let $M > 0$ be such that $|f(x)| \le M$ for all $x \in [a, b]$.

Let $\epsilon > 0$. Let $\delta_0 > 0$ be such $\mathcal{S}_{\delta_0}(f_1) \subset B_\epsilon(A)$ and $\mathcal{S}_{\delta_0}(f_2) \subset B_\epsilon(C)$; such δ_0 exists by the assumed integrability of f_1 and f_2. Let $\delta = \min\{\delta_0, \frac{\epsilon}{4M}\}$. Then $\mathcal{S}_\delta(f_1) \subset \mathcal{S}_\epsilon(f_1) \subset B_\epsilon(A)$ and $\mathcal{S}_\delta(f_2) \subset \mathcal{S}_\epsilon(f_2) \subset B_\epsilon(C)$ (since $\delta \le \delta_0$) and $4M\delta \le \epsilon$, facts we will use later.

Let $(P, T) = (\{x_0, \dots, x_N\}, \{t_1, \dots, t_N\})$ be a pointed partition of $[a, b]$ of width less than δ. Define

$$\begin{aligned} j' &= \max\{j \in \{1, \dots, N\} : x_j < c\}, && (6.41)\\ j'' &= \min\{j \in \{1, \dots, N\} : x_j > c\} && (6.42) \end{aligned}$$

(thus the value of $j'' - j'$ is either 2 or 1, accordingly as c is or is not an element of P). Define partitions P', P'' of $[a, c], [c, b]$, respectively, by

$$\begin{aligned} P' = (P \cap [a, c]) \cup \{c\} &= \{x_0, \dots, x_{j'}, c\}, && (6.43)\\ P'' = \{c\} \cup (P \cap [c, b]) &= \{c, x_{j''}, \dots, x_N\}; && (6.44) \end{aligned}$$

observe that $\mathrm{wid}(P')$ and $\mathrm{wid}(P'')$ are at most $\mathrm{wid}(P)$, hence are less than δ. Define pointings T', T'' of P', P'', respectively, by

$$\begin{aligned} T' &= \{t_1, \dots, t_{j'}, c\},\\ T'' &= \{c, t_{j''}, \dots, t_N\}. \end{aligned}$$

For $j \le j'$ and for $j > j''$, the j^{th} term in the sum defining $S(f; P, T)$ is a term in either the sum defining $S(f_1; P', T')$ or the sum defining $S(f_2; P'', T'')$ (but not both). Similarly, every term except possibly the last (respectively, first) in the sum defining $S(f_1; P', T')$ (resp., $S(f_2; P'', T'')$) is a term in the sum defining $S(f; P, T)$. Hence

$$\begin{aligned} & S(f;P,T) - (S(f_1;P',T') + S(f_2;P'',T'')) \\ & = \begin{cases} \begin{array}{c} f(t_{j''})(x_{j''} - x_{j'}) - f(c)(c - x_{j'}) - f(c)(x_{j''} - c) \\ = (f(t_{j''}) - f(c))(x_{j''} - x_{j'}), \end{array} & \text{if } c \notin P \\ \\ (f(t_{j'+1}) - f(c))(c - x_{j'}) + (f(t_{j''}) - f(c))(x_{j''} - c) & \text{if } c \in P. \end{cases} \end{aligned} \tag{6.45}$$

The numbers $(c - x_{j'})$ and $(x_{j''} - c)$ lie in the interval $(0, \delta)$, as does $(x_{j''} - x_{j'})$ in the "$c \notin P$" case above. Since $|f(x) - f(y)| \leq |f(x)| + |f(y)| \leq M$ for all $x, y \in [a, b]$, and $2M < 4M$, the triangle inequality then shows that whichever line of (6.45) applies, we have

$$|S(f;P,T) - (S(f_1;P',T') + S(f_2;P'',T''))| \leq 4M\delta.$$

By the triangle inequality and the definition of δ, this implies that

$$\begin{aligned} |S(f;P,T) - (A + C)| &\leq |S(f;P,T) - (S(f_1;P',T') + S(f_2;P'',T''))| \\ &\quad + |S(f_1;P,T) - A| + |S(f_2;P,T) - C| \\ &< 4M\delta + 2\epsilon \\ &\leq 3\epsilon. \end{aligned} \tag{6.46}$$

Since (P, T) was an arbitrary pointed partition of width less than δ, the inequality (6.46) shows that $\mathcal{S}_\delta(f) \subset B_{3\epsilon}(A + C)$. Since ϵ was arbitrary, this establishes that f is integrable and that $\int_a^b f = A + C$, as desired.

Conversely, suppose that f is integrable on $[a, b]$. Then f is bounded. Let $\epsilon > 0$. Let P be a partition of $[a, b]$ such that $U(f; P) - L(f; P) < \epsilon$; such P exists by Proposition 6.45. Define indices j', j'' and partitions P', P'' (of $[a, c]$ and $[c, b]$, respectively) just as in (6.41)–(6.42) and (6.43)–(6.44). For each $j \in \{1, \ldots, N\}$, define m_j and M_j as in Proposition 6.36. Let $M' = \sup(f([x_{j'}, c]))$ and $m' = \inf(f([x_{j'}, c])$; since $[x_{j'}, c] \subset [x_{j'}, x_{j'+1}]$ we have $M' \leq M_{j'+1}$ and $m' \geq m_{j'+1}$. Then, applying Proposition 6.36,

$$
\begin{aligned}
U(f_1; P') - L(f_1; P') &= \sum_{j=1}^{j'} (M_j - m_j)\Delta_j(P) + (M' - m')(c - x_{j'}) \\
&\leq \sum_{j=1}^{j'} (M_j - m_j)\Delta_j(P) + (M_{j'+1} - m_{j'+1})(x_{j'+1} - x_{j'}) \\
&= \sum_{j=1}^{j'+1} (M_j - m_j)\Delta_j(P) \\
&\leq \sum_{j=1}^{N} (M_j - m_j)\Delta_j(P) \\
&= U(f; P) - L(f; P) \\
&< \epsilon.
\end{aligned}
$$

Similarly, $U(f_2; P'') - L(f_2, P'') < \epsilon$. Since ϵ was arbitrary, Proposition 6.45 implies that f_1 and f_2 are integrable. ■

Corollary 6.57 *Let $f : [a, b] \to \mathbf{R}$.*

(a) The function f is integrable if and only if its restriction to each closed subinterval of $[a, b]$ is integrable.

(b) Suppose f is integrable, n is a positive integer, and $a < c_1 < c_2 \cdots < c_n < b$. Then

$$\int_a^b f = \int_a^{c_1} f + \int_{c_1}^{c_2} f + \cdots + \int_{c_n}^{b} f.$$

Exercise 6.7 Prove Corollary 6.57.

Definition 6.58 Let $a, b \in \mathbf{R}$, with $a \leq b$, and let f be a real-valued function on $[a, b]$.

(i) We define $\int_a^a f = 0$, and say that this integral exists.

(ii) If $b > a$, we say that $\int_b^a f$ exists if and only if $\int_a^b f$ exists, in which case we define $\int_b^a f = -\int_a^b f$.

▲

Corollary 6.59 *Let $a, b \in \mathbf{R}$ (with the possibilities $a < b$, $a = b$, $a > b$ all allowed).*

(i) Let $c \in \mathbf{R}$. Then $\int_a^b c = c(b-a)$.

(ii) Suppose that f is integrable on the closed interval with endpoints a and b, and that $|f(x)| \leq M$ for every x in this interval. Then

$$\left| \int_a^b f \right| \leq M|b-a|. \tag{6.47}$$

Proof: In view of Definition 6.58, it suffices to establish (i) and (ii) in the case $a < b$, so let us assume $a < b$. Then (i) follows from Example 6.11. For (ii), note that every Riemann sum of f over $[a, b]$ lies in the interval $[-M(b-a), M(b-a)]$. The definition of $\int_a^b f$ then implies that $\int_a^b f$ also lies in this interval. Therefore (6.47) holds. ∎

Note that in Proposition 6.55, if both integrals on the right-hand side of equation (6.40) exist, then so does the integral on the left-hand side, while if the integral on the left-hand side exists, so do both of the integrals on the right-hand side. Hence if any two of the three integrals written in equation (6.40) exist, so does the third, and the equation holds true. Observe also that equation (6.40) can be rewritten as

$$\int_c^b f = \int_a^b f - \int_a^c f,$$

which, using Definition 6.58, can be further rewritten as

$$\int_c^b f = \int_c^a f + \int_a^b f, \tag{6.48}$$

which differs from (6.40) only by a permutation of the letters a, b, c. Furthermore, $\int_c^a f$ exists if and only if $\int_a^c f$ exists, so two of the three integrals in (6.40) exist if and only if two of the three integrals in (6.48) exist. However, if $a < c < b$, then in (6.48) we do not have $c < a < b$; equation (6.48) holds even though the limits of integration do not have the same order-relation as in Proposition 6.55. Pushing these ideas a little further leads to the following:

Corollary 6.60 *Let $a, b, c \in \mathbf{R}$ and let f be a real-valued function defined on an interval that includes a, b, and c. (No ordering or distinctness of a, b, c is assumed.) Then if any two of the three integrals $\int_a^c f$, $\int_a^b f$, $\int_b^c f$ exists, so does the third, and*

$$\int_a^c f = \int_a^b f + \int_b^c f; \tag{6.49}$$

equivalently,

$$\int_a^c f - \int_b^c f = \int_a^b f. \tag{6.50}$$

Exercise 6.8 Prove Corollary 6.60. (Do not forget to handle the cases in which two or three of the numbers a, b, c are equal.)

6.7 The Fundamental Theorem of Calculus

There are essentially two different, but closely-related, theorems that go by the name "The Fundamental Theorem of Calculus"[6] (or, more historically, "The Fundamental Theorem of *Integral* Calculus"; this longer name is more descriptive but is rarely used anymore). One of these involves the integral of a derivative, and the other the derivative of an integral. More precisely, this is the primary distinction between the *conclusions* of these theorems. For each of these types of conclusions, there are actually more than one theorem, differing in their hypotheses. Of the various theorems that go by the name "The Fundamental Theorem of Calculus", we will prove the two that are of the greatest use in calculus, and refer to each of these two as "part of the Fundamental Theorem of Calculus".[7] Later, in optional reading for the student, we discuss some of the other, related, theorems called The Fundamental Theorem of Calculus, and discuss the nomenclature for all these theorems.

The following simple lemma is needed for the statement of the first theorem we will prove.

Lemma 6.61 *Let $U \subset \mathbf{R}$ be an open interval, $f : U \to \mathbf{R}$ a continuous function, and $a, b \in U$. Then $\int_a^b f$ exists.*

Proof: If $a = b$ then, by definition, the integral exists and is 0. If $a \neq b$ then the restriction of f to $[\min\{a, b\}, \max\{a, b\}]$ is continuous, so by Theorem 6.53, the integral of f over this interval exists. If $a < b$ we are done; if $a > b$ the result follows from Definition 6.58. ■

This lemma assures us that the function F in the theorem below is indeed well-defined.

Theorem 6.62 ("part of" the Fundamental Theorem of Calculus) *Let f be a continuous real-valued function on an open interval $U \subset \mathbf{R}$, and let $a \in U$. Define $F : U \to \mathbf{R}$ by*

$$F(x) = \int_a^x f(t)\, dt.$$

[6]Since there is more than one theorem called "The Fundamental Theorem of Calculus", it is tempting to refer to these theorems collectively as "The Fundamental Theorem*s* of Calculus". We choose not do so in these notes, however, since that terminology can give the impression that this group of theorems contains *all* the theorems that are fundamental to calculus, when in fact "The Fundamental Theorem of Calculus" is simply a historical name for one theorem and its relatives.

[7] Needless to say, we could state a single theorem that has each of these two theorems as a part, but the only motivation would be that the two theorems share a name. Combining them into a single theorem, each part of which has its own hypotheses, would be rather artificial, and would be inconvenient for the proofs.

Then F is differentiable and $F' = f$.

Proof: Fix $x_0 \in U$, and let $x \in U$. Corollary 6.59(i) implies that $\int_{x_0}^{x} f(x_0)\,dt = f(x_0)(x - x_0)$ (here we are integrating the *constant* function $t \mapsto f(x_0)$). Using Corollary 6.60 in the form (6.50), we also have $F(x) - F(x_0) = \int_{x_0}^{x} f(t)dt$. Hence

$$F(x) - F(x_0) - f(x_0)(x - x_0) = \int_{x_0}^{x} f(t)\,dt - \int_{x_0}^{x} f(x_0)\,dt = \int_{x_0}^{x} (f(t) - f(x_0))\,dt.$$

Therefore for all $x \in U$ with $x \neq x_0$, we have

$$\left|\frac{F(x) - F(x_0)}{x - x_0} - f(x_0)\right| = \left|\frac{F(x) - F(x_0) - f(x_0)(x - x_0)}{x - x_0}\right| = \frac{\left|\int_{x_0}^{x} (f(t) - f(x_0))\,dt\right|}{|x - x_0|}. \tag{6.51}$$

Now let $\epsilon > 0$ be given, and let $\delta > 0$ be such that for all $x \in U$ with $|x - x_0| < \delta$ we have $|f(x) - f(x_0)| < \epsilon$; such δ exists since f is continuous at x_0. Then for all $x \in B_\delta(x_0) \setminus \{x_0\}$, (6.51) and Corollary 6.59(ii) imply that

$$\left|\frac{F(x) - F(x_0)}{x - x_0} - f(x_0)\right| \leq \frac{\epsilon|x - x_0|}{|x - x_0|} = \epsilon.$$

Hence $\lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0} = f(x_0)$. Thus F is differentiable at x_0, and $F'(x_0) = f(x_0)$. Since x_0 was arbitrary, we are done. ■

An *antiderivative* of a function f on an open set U is a differentiable function F such that $F' = f$. An immediate corollary of Theorem 6.62 is:

Corollary 6.63 *Every continuous function f on an open interval has an antiderivative on that interval.*

Proof: Fix any $a \in U$. Then the function $x \mapsto \int_a^x f(t)\,dt$ is an antiderivative of f on U. ■

Theorem 6.64 (The Fundamental Theorem of Calculus) *Let $U \subset \mathbf{R}$ be an open interval, let $f : U \to \mathbf{R}$ be a continuous function, and let F be an antiderivative of f on U. Then for all $a, b \in U$,*

$$\int_a^b f(t)dt = F(b) - F(a). \tag{6.52}$$

Observe that Theorem 6.64 can be stated equivalently as follows:

Theorem 6.65 (The Fundamental Theorem of Calculus) *Let $U \subset \mathbf{R}$ be an open interval, and let $F : U \to \mathbf{R}$ be a differentiable function whose derivative F' is continuous. Then for all $a, b \in U$,*

$$\int_a^b F'(t)\,dt = F(b) - F(a). \tag{6.53}$$

Proof of Theorem 6.64: Fix $a \in U$, and define $G : U \to \mathbf{R}$ by $G(x) = \int_a^x f(t)\,dt$. By Theorem 6.62, $G' = f$. But by hypothesis, $F' = f$. Recall the following consequence of the Mean Value Theorem: If two differentiable functions H_1, H_2 on an open interval have identical derivatives, then $H_2 - H_1$ is constant (on that interval). Hence $G - F$ is constant. Therefore for all $x \in U$,

$$G(x) - F(x) = G(a) - F(a) = 0 - F(a) = -F(a),$$

so $G(x) = F(x) - F(a)$. Thus for any $b \in U$, $\int_a^b f(t)\,dt = G(b) = F(b) - F(a)$. ■

Remark 6.66 The careful reader will have noticed that Theorem 6.62 is not, in fact, part of Theorem 6.64. For an explanation of the this apparently illogical terminology, see Remark 6.72 (optional reading).

Remark 6.67 Theorem 6.64 (and even a stronger version) can be proven without the use of Theorem 6.62; the earlier theorem simply affords us a proof of Theorem 6.64 than is shorter than other proofs. See Theorem 6.70 and Exercise 6.10 later. ▲

Exercise 6.9 Evaluate $\lim_{n\to\infty} \frac{1}{n}\left[\left(\frac{1}{n}\right)^6 + \left(\frac{2}{n}\right)^6 + \left(\frac{3}{n}\right)^6 + \cdots + \left(\frac{n}{n}\right)^6\right]$.

Problems like the exercise above were common in high-school math-team competitions when the writer of these notes was in high school. Usually, students were given 2 minutes or so to solve such a problem. The trick is to recognize the sequence whose limit is being taken as a sequence of Riemann sums for an appropriate function over an appropriate interval, then use Exercise 6.1 and the Fundamental Theorem of Calculus.

Remark 6.68 "True" integration refers to what we call the "definite integral" in Calculus 1; it's about adding stuff up. This is true whether we are talking about the Riemann integral, a generalization called the Riemann-Stieltjes integral, improper integrals, or the Lebesgue integral. Nothing in the *concept* of integration involves differentiation. Archimedes already had this concept of integration as "adding up stuff" nearly two millennia before derivatives and integrals were defined, when he realized that the area inside a circle could be computed as the limit as $n \to \infty$ of the area of an inscribed regular n-gon. The Fundamental Theorem of Calculus (FTC) relates two completely distinct concepts:

integration and *antidifferentiation*. Because we are able to compute antiderivatives of so many familiar functions, the FTC is a key tool in the computation of (definite) integrals.

It is *because* of the Fundamental Theorem of Calculus that antiderivatives are also called by a name, "indefinite integrals", that involves the word "integral". If you learned indefinite integration before definite integration, you may have received the false impression that "integration" always *means* "antidifferentiation". In this case, when learning about the Riemann-sum definition of the integral (either in Calculus 1 or in Advanced Calculus), you may have wondered, "What does this have to do with integration?" But you should now realize that this is the wrong question. Once you understand what integration actually means, but before you learn the FTC, the right question is "What does *antidifferentiation* have to do with integration?" This question is answered by the FTC. The FTC is the reason that antiderivatives are also called (indefinite) integrals.

If you learned indefinite integration before definite integration, another question you may have asked yourself is, "Where does this symbol '$\int$' " come from?" It comes from definite integration. The history of the symbol is that "$\int$" is an elongated S, the "S" standing for "sum". The reason that the same symbol is used for antiderivatives is, again, the FTC. ▲

Remark 6.69 Observe that the last statement of Theorem 6.62 can be written as

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x), \tag{6.54}$$

a statement about "the derivative of an integral" (more precisely, the derivative of a function defined by an integral in the specific way above), whereas (6.53) is a statement about the integral of a derivative. Although (6.53) and (6.54) look different, they are actually equivalent once we know Corollary 6.63. By simply changing notation, we can rewrite (6.53) as $\int_a^x F'(t)\,dt = F(x) - F(a)$. Since F in this equation is assumed differentiable, and $F(a)$ is just a constant, the right-hand side of this equation is differentiable in x; hence so is the left-hand side. Thus, given (6.53), we deduce that $\frac{d}{dx}\int_a^x F'(t)\,dt = \frac{d}{dx}(F(x) - F(a)) = F'(x)$. Since (from Corollary 6.63) every continuous real-valued function on an open interval has an antiderivative, there is no loss of generality if we replace F' (which was assumed continuous when we wrote (6.53)) in this last equation by an arbitrary continuous function f. But this yields (6.54).

Conversely, our proof of Theorem 6.64 shows that (6.54) implies (6.52), hence also implies (6.53). This equivalence is the reason that both Theorem 6.64 and Theorem 6.62 are often referred to by the same name, "The Fundamental Theorem of Calculus." However, as written, Theorem 6.62 is a stronger theorem than Theorem 6.64, since it implies that every continuous real-valued function on an open interval has an antiderivative, which cannot be deduced from Theorem 6.64. ▲

The remainder of this section is optional reading. (However, if you're wondering why Theorems 6.62 and Theorem 6.64 were given their names in these notes, the answer is contained in Remark 6.72.)

In Theorem 6.65, we assumed that the integrand F' was continuous. This hypothesis can be weakened to the assumption that F' is merely *integrable* over the appropriate interval, thereby obtaining the following stronger theorem (which can also reasonably be called "the Fundamental Theorem of Calculus"):

Theorem 6.70 *Let $U \subset \mathbf{R}$ be an open interval, let $a, b \in U$, and let $F : U \to \mathbf{R}$ be a differentiable function whose derivative F' is integrable over the interval with endpoints a and b. Then equation (6.53) holds.*

Exercise 6.10 (Optional.) Prove Theorem 6.70. *Hint*: It suffices to prove the result in the case $a < b$ (why?). Assume $a < b$. For any partition $P = \{x_0, \ldots, x_N\}$ of $[a, b]$, observe that $F(b) - F(a) = \sum_{j=1}^{N}(F(x_j) - F(x_{j-1}))$. Apply the Mean Value Theorem to F on each interval $[x_{j-1}, x_j]$, deduce that (for any partition P), $L(f; P) \leq F(b) - F(a) \leq U(f; P)$. Now apply an appropriate result that we proved earlier.

Theorem 6.62 can also be strengthened:

Theorem 6.71 *If $f \in \mathcal{R}([b, c])$, and $a \in [b, c]$, then the function $F : [b, c] \to \mathbf{R}$ defined by*

$$F(x) = \int_a^x f(t)\, dt$$

is continuous. If, in addition, f is continuous at $x_0 \in (b, c)$, then F is differentiable at x_0, and $F'(x_0) = f(x_0)$.

Observe that the proof we gave of Theorem 6.62 actually proves the second assertion in Theorem 6.71; we simply replace the *arbitrary* point x_0 in that proof by the *specific* point x_0 in the statement of Theorem 6.71. What is new in Theorem 6.71 is really the *first* assertion: that if we assume only that f is *integrable* (rather than *continuous*) on a closed, bounded interval containing a, we can still deduce something about the function F, namely that it is continuous.

Exercise 6.11 Prove the first assertion in Theorem 6.71.

Most textbooks on introductory or advanced calculus state only Theorems 6.64 and 6.62; usually Theorems 6.70 and 6.71 are stated only in more advanced textbooks on analysis.

Remark 6.72 (Naming the theorems in this section) For purposes of this Remark, minor differences in the statements of the theorems under discussion are ignored.

If you ask different mathematicians (or even the same mathematician at different times), "What is the Fundamental Theorem of Calculus?" you will get different answers. The answer may also depend on the level of who's asking the question. In some textbooks, Theorem 6.64 is called the Fundamental Theorem of Calculus (FTC) or the Fundamental Theorem of Integral Calculus (FTIC), and Theorem 6.62 is stated but not given a name (e.g. [8, 11]). In other textbooks, exactly the opposite is true: Theorem 6.62 is called the FTC, and Theorem 6.64 is stated but not given a name (e.g. [5]).

This is not where the name-discrepancies end. The first edition of Apostol's *analysis* textbook [2] calls Theorem 6.64 the FTIC, and states and proves a generalized version[8] of Theorem 6.71, but does not give a name to the latter theorem. The second edition of Apostol's analysis textbook, [3], calls Theorem 6.64 the *Second* FTIC. In this edition, Apostol stil does not give his version of Theorem 6.71 a name, but says afterwards that part (iii) of this theorem—the only part that gives a relation between integration and differentiation—is "sometimes called the *first* FTIC" in the special case of the pure Riemann integral. Apostol's *calculus* textbook [1] calls Theorem 6.62 the *First* FTC, and Theorem 6.64 the *Second* FTC. Some textbooks implicitly (but never explicitly) combine Theorems 6.62 and 6.64 into one theorem[9], by calling Theorem 6.62 the "FTC, part 1", and call Theorem 6.64 the "FTC, part 2" (e.g. [4, 9]). Some textbooks call Theorem 6.62 the "FTC–*Second* Form" and Theorem 6.64 the "FTC–*First* Form". (Thus, among authors calling Theorems 6.62 and 6.64 parts or forms of the same theorem, there is inconsistency about which part/form is the first, and which is the second.) In [10], no theorem is given a name that includes "FTC"; Theorems 6.64 and 6.62 are stated but not given names. Rudin [6, 7] states only the stronger versions of Theorems 6.64 and 6.62 (Theorems 6.70 and 6.71), calls Theorem 6.70 the FTC, and does not give a name to Theorem 6.71.

What all the theorems whose textbook names include "FTC" (when the theorems are named at all) have in common is that they are theorems about, and only about, an "inverse" relationship of differentiation and integration (their conclusions are purely about the derivative of an integral or the integral of a derivative). Sometimes you may see "FTC" included in the name of Theorem 6.71, but this is less conventional than leaving the theorem un-named, because the first assertion of Theorem 6.71 has nothing to do with a relation between integration and differentiation. (Apostol's treatment in [3], in which he says only that *part* of his un-named version of Theorem 6.71 is called the First FTIC, is more conventional.)

[8]In [2] and [3], Apostol works with a generalized version of the Riemann integral called the Riemann-Stieltjes integral. In some cases, but not all, he specifically states what various theorems reduce to for the Riemann integral. A similar comment applies to Rudin [6, 7].

[9]Presumably, authors' reasons for never doing this explicitly are similar to those mentioned in footnote 7.

The upshot is that students should take none of these name-variants as gospel. It is okay to call any of Theorems 6.62, 6.64, 6.70, and 6.71 the FTC, or part of the FTC, or a form or version of the FTC.

The writer of these notes generally regards Theorem 6.65 (equivalently, Theorem 6.64) as the "true" FTC, but when there is a need to refer to the (extremely important) Theorem 6.62, he uses language that implicitly combines Theorem 6.62 and Theorem 6.64 into one grand FTC; hence the name given here to Theorem 6.62. (This is similar to the approach that calls Theorems 6.64 and 6.62 "parts 1 and 2"—in whatever order—of the FTC.) Theorem 6.65 has a place in mathematics that is rather more special than that of Theorem 6.62, in several respects:

- Theorem 6.65 is the first of several important theorems, covered in a traditional Calculus 3 course, that have a certain formal similarity that is actually very deep. Other theorems in this collection are the "Fundamental Theorem of Line Integrals", Green's Theorem, Stokes's Theorem, and the Divergence Theorem. Each of these theorems pertains to integrating a suitably defined derivative "$d\omega$" of a suitably defined object ω over a "nice" n-dimensional set S with $(n-1)$-dimensional boundary ∂S ($n = 1, 2$, or 3), and makes a statement of the form

  $$n\text{-dimensional integral of } d\omega \text{ over } S = (n-1)\text{-dimensional integral of } \omega \text{ over } \partial S.$$

  (For these purposes, the 0-dimensional integral of a function $F : [a, b] \to \mathbf{R}$ is simply $F(b) - F(a)$.) This collection of theorems generalizes to a single theorem, also called Stokes's Theorem, that holds for all $n \geq 1$ (not just $n = 1, 2, 3$). The FTC (in the form 6.65) is simultaneously a special case of this generalized version of Stokes's Theorem, and a key step in its proof. This more general Stokes's Theorem is extremely important on its own, but is also the inspiration for a large subject in algebraic topology called "homology and cohomology theory".

- As discussed in Remark 6.68, Theorem 6.64 (equivalent to Theorem 6.65) is the reason that we use the symbol "$\int$" for antiderivatives.

▲

6.8 Change of variable

Definition 6.73 Let $U \subset \mathbf{R}$ be open.

(a) A function $g : U \to \mathbf{R}$ is *continuously differentiable* if g is differentiable and g' is continuous.

(b) Let $V \subset \mathbf{R}$, let $g : U \to V$ be a function, and let $\tilde{g} : U \to \mathbf{R}$ be the function defined from g by simply changing the codomain to $\mathbf{R}$. (Equivalently, $\tilde{g} = \iota \circ g$, where $\iota : V \to \mathbf{R}$

is the *inclusion map*; i.e. $\iota(x) = x$ for all $x \in V$.) We say that g is (continuously) differentiable if $\tilde{g}$ is (continuously) differentiable. ▲

Proposition 6.74 (Change-of-variable in one-dimensional integrals) *Let $U, I \subset \mathbf{R}$ be open intervals, $f : U \to \mathbf{R}$ a continuous function, $\varphi : I \to U$ a continuously differentiable function. Then for any $a, b \in I$,*

$$\int_{\varphi(a)}^{\varphi(b)} f = \int_a^b (f \circ \varphi)\, \varphi' \; ; \tag{6.55}$$

equivalently, in "dummy-variable notation",

$$\int_{\varphi(a)}^{\varphi(b)} f(u)\, du = \int_a^b f(\varphi(x))\, \varphi'(x)\, dx. \tag{6.56}$$

Proof: Fix $a \in U$. Define $F : U \to \mathbf{R}$ by $F(y) = \int_{\varphi(a)}^{y} f$. Then, by Theorem 6.62, F is differentiable and $F' = f$. Define $G : I \to \mathbf{R}$ by $G = F \circ \varphi$. Then G is the composition of differentiable functions, so G is differentiable and $G' = (F' \circ \varphi)\varphi' = (f \circ \varphi)\varphi'$. The function $(f \circ \varphi)\varphi'$ is continuous (why?), so by Theorem 6.62, the function $H : I \to \mathbf{R}$ defined by $H(x) = \int_a^x (f \circ \varphi)\varphi'$ is differentiable, and $H' = (f \circ \varphi)\varphi' = G'$. Since I is an interval, "$H' = G'$" implies that $G - H$ is constant (see the proof of Theorem 6.64). Hence for all $x \in I$,

$$G(x) - H(x) = G(a) - H(a) = F(\varphi(a)) - H(a) = \int_{\varphi(a)}^{\varphi(a)} f - \int_a^a (f \circ \varphi)\varphi' = 0 - 0 = 0.$$

Hence $G(x) = H(x)$ for all $x \in I$. In particular, $G(b) = H(b)$, which is exactly equation (6.55). ■

Remark 6.75 Writing the result of Proposition 6.74 in the form (6.56) explains the terminology "change of variable"; we think of the dummy variables in (6.56) as being related by the equation $u = \varphi(x)$. However, observe that in this proposition, φ need not be one-to-one. This is a remarkable feature of the "one-dimensional" change-of-variables formula that is not shared by the change-of-variables formula for multiple integrals (the last topic we will study in this course, if we complete the syllabus).

Remark 6.76 (Helpfulness of Leibniz notation when changing variables) In the Leibniz notation for the derivative of a function f, names are chosen for the independent and dependent variables—say x and y, respectively, related by the equation $y = f(x)$; sometimes we simply write "$y = y(x)$." With this choice of variables, the Leibniz notation for $f'(x)$ is $\frac{dy}{dx}$ (in which we must remember that the right-hand side is not actually a fraction with real numbers in numerator and denominator). In some situations, this notation can lead to problems; in others, it is extremely helpful. The change-of-variables formula(s) in Proposition 6.74 is an instance in which the Leibniz notation is a truly marvelous mnemonic device. In place of introducing a *name* φ for the functional relation between u and x that we're thinking of when we write the formula (6.56), we simply write "$u(x)$" in place of $\varphi(x)$ on the right-hand side, and write $\frac{du}{dx}$ in place of $\varphi'(x)$. In the limits of integration on the left-hand side, instead of writing "$\varphi(a)$" and "$\varphi(b)$", we could write $u(a)$ and $u(b)$, but—since we are thinking of this as a change of variables—we often write "$u = u(a)$" and "$u = u(b)$" instead. For the sake of symmetry, we often use similar notation for the limits of integration on the right-hand side. Equation (6.56) then becomes

$$\int_{u=u(a)}^{u=u(b)} f(u)\,du = \int_{x=a}^{x=b} f(u(x))\frac{du}{dx}\,dx,$$

or, even more familiarly,

$$\int_{x=a}^{x=b} f(u(x))\frac{du}{dx}\,dx = \int_{u=u(a)}^{u=u(b)} f(u)\,du. \tag{6.57}$$

In other words, if we simply pretend that $\frac{du}{dx}$ in (6.57) is a true fraction, whose denominator can be cancelled by the "dx" appearing to its right, then it appears "obvious" that the left-hand side of (6.57) equals the right-hand side. While this logic for equating the left-hand side with the right-hand side is completely bogus, it does allow us to *remember* (6.55) and (6.56)—*which we have rigorously proven*—more easily. This is a tremendous benefit, and both student and seasoned mathematician alike have no reason to be embarrassed by relying on the above "abuse of notation" (pretending that $\frac{du}{dx}$ is a fraction, etc.) to help *remember* (6.55) and (6.56). Just keep in mind that a valid *proof* is needed to deduce that (6.57) is correct; "proof by abuse of notation" (or "proof by misunderstanding notation") is not a valid method of proof. ▲

Remark 6.77 As the student will recall from Calculus 1, Proposition 6.74 is a useful tool for the evaluation of integrals. Reasonable names for this tool are "integration by substitution" and "changing variables in the integral".[10] Calculus 1-2-3 tomes currently on the market usually call this technique, and its analog for indefinite integrals, by the abysmal name "u-substitution". You will not find this terminology in older, "classic" textbooks such as [1–3, 8, 11], or in Rosenlicht [5]. In older books, the technique is named according to the *concept* of substitution, rather than a *letter* that is commonly used in substitutions. Calling this technique "u-substitution" is like calling every function f : (subset of $\mathbf{R}$) $\to \mathbf{R}$ an "x-function". ▲

[10]However, once we learn about changing variables in multiple integrals—a topic at the end of this course, if we complete the syllabus—we will see that "changing variables" is not a great description of (6.56) unless φ restricts to a bijection from the interval with endpoints a, b to the interval with endpoints $\varphi(a), \varphi(b)$.

6.9 Integration of vector-valued functions

This section is an expanded version of Rosenlicht's homework problem VI.6, a problem that illustrates the generality and several strengths of the Riemann-sum approach to the Riemann integral.

Throughout this section, $(V, \| \ \|)$ denotes a complete normed vector space[11], with the associated metric d. Usually we will write simply V rather than $(V, \| \ \|)$, with understanding that V has been given a fixed norm $\| \ \|$ for which the metric space (V, d) is complete. We will write 0_V for the zero element of V. For $c \in \mathbf{R}$ and $v \in V$, we define "vc" to mean cv. Open balls in V will generally be denoted by notation of the form "$B_\epsilon(v)$", but in situations in which both balls in V and balls in $\mathbf{R}$ enter the discussion, we put an appropriate superscript V or $\mathbf{R}$ on the "B".

We will extend the theory of the Riemann integral from the realm of real-valued functions to the realm of vector-valued functions, by which we mean functions from an interval $[a, b]$ to a (complete, normed) vector space V. We do not assume that V is finite-dimensional, except where noted. However, the case $V = \mathbf{R}^n$ (with, say, the Euclidean norm) is an important special case, and it is very helpful to keep this case in mind when trying to grasp what various definitions, propositions, etc., are saying.

Definition 6.78 (Riemann sums) Let $f : [a, b] \to V$ be a function and let $(P, T) = (P, \{t_1, \dots t_N\})$ be a pointed partition of $[a, b]$. The *Riemann sum* for f corresponding to (P, T) is

$$S(f; P, T) = \sum_{j=1}^{N} f(t_j)\Delta_j \,. \tag{6.58}$$

As we did for real-valued functions, we will write

$$\mathcal{S}(f; P) \quad = \quad \{S(f; P, T) : T \text{ is a pointing of } P\},$$

and for each $\delta > 0$, write

$$\mathcal{S}_\delta(f) = \bigcup \{\mathcal{S}(f; Q) : Q \in \mathcal{P}_\delta([a, b])\}.$$

▲

Note that there is *no difference* between the definitions (6.2) and (6.58) of Riemann sums, except that in (6.58) the function f is taking its values in V rather than $\mathbf{R}$. The same definition would work with V replace by *any* vector space (whether or not normed or complete); all that is needed for the definition (6.58) is the vector-space structure on

[11] A complete normed vector space is called a *Banach space*, but to help the student keep in mind the important features we are assuming of our $(V, \| \ \|)$, we will stick to the self-descriptive term "complete normed vector space".

V. For the next definition, we need only a little more: the metric structure on V given by a norm. This definition could be written exactly as Definition 6.6, simply replacing the absolute-value symbols by norm-symbols, but we will use our notation "$\mathcal{S}_\delta(f)$" to state the definition more efficiently (as we did for real-valued functions in Remark 6.14).

Definition 6.79 (Integrability) A function $f : [a, b] \to V$ is *(Riemann) integrable* if there is a vector $A \in V$ such that for each $\epsilon > 0$ there exists $\delta > 0$ such that $\mathcal{S}_\delta(f) \subset B_\epsilon(A)$. More generally, if f is a V-valued function whose domain includes $[a, b]$, we say that f *is integrable on* $[a, b]$ (or *over* $[a, b]$) if $f|_{[a,b]}$ is integrable. ▲

We continue our convention (for these notes) that "integrable" means "Riemann integrable" and that all integrals we discuss are Riemann integrals.

If there exist distinct $A, A' \in V$ both satisfying the condition satisfied by A in Definition 6.79, then for $\epsilon = \|A - A'\|/2$ and $S \in V$ we cannot have both $\|S - A\| < \epsilon$ and $\|S - A'\| < \epsilon$ (the triangle inequality would lead to a contradiction). Therefore if, just as for real-valued functions, if f is integrable on $[a, b]$ then there is a *unique* $A \in V$ satisfying the condition in Definition 6.79. Thus we can define the integral of f exactly as in Definition 6.8, just with $\mathbf{R}$ replaced by V:

Definition 6.80 Let $f : [a, b] \to V$ be integrable. We define the *integral of* f to be the unique $A \in V$ satisfying the condition given in Definition 6.6, and denote this A as $\int_a^b f$ or as $\int_a^b f(x)dx$, etc. for any dummy variable. More generally, if f is a V-valued function on a domain that includes $[a, b]$, and f is integrable on $[a, b]$, we use the notation $\int_a^b f$ (or $\int_a^b f(x)dx$, etc.) for the integral of $f|_{[a,b]}$, and refer to the value of this integral as *the integral of* f *over* $[a, b]$. We define the phrase "$\int_a^b f$ exists" (or "$\int_a^b f(x)\,dx$ exists", etc.) to mean that f is integrable on $[a, b]$. ▲

Notation 6.81 We let $\mathrm{Func}([a, b], V)$ denote the set of *all* functions $[a, b] \to V$, and let $\mathcal{R}([a, b], V) \subset \mathrm{Func}([a, b], V)$ denote the set of integrable functions from $[a, b]$ to V.

The set $\mathrm{Func}([a, b], V)$ is itself a vector space, with zero element the constant function $x \mapsto 0_V$, and with the vector-space operations defined through pointwise operations: for $f, g \in \mathrm{Func}([a, b], V)$ and any $c \in \mathbf{R}$, we define elements $f + g$ and cf of $\mathrm{Func}([a, b], V)$ by $(f + g)(x) := f(x) + g(x)$ and $(cf)(x) := cf(x)$ for all $x \in [a, b]$.

Remark 6.82 If $\dim(V) = 0$, then $V = \{0_V\}$ and $\mathrm{Func}([a, b], V)$ contains only the constant function $x \mapsto 0_V$. All Riemann sums of this function have the value 0_V. Hence this function is integrable, and the value of the integral is 0_V.

Thus, in a discussion of integrating vector-valued functions, the 0-dimensional vector space is not interesting. We have not excluded it from our discussion, though, since a restriction of the form "Assume $\dim(V) \geq 1$" might give the impression that something goes wrong if $\dim(V) = 0$, rather than that this case is simply uninteresting. ▲

Exercise 6.12 Recall that two norms $\| \ \|_1, \| \ \|_2$ on V are called *equivalent* if there exist real numbers $c_1, c_2 > 0$ such that for all $v \in V$ we have $\|v\|_2 \leq c_1\|v\|_1$ and $\|v\|_1 \leq c_2\|v\|_2$. Show that if the given norm $\| \ \|$ on V is replaced by any equivalent norm, neither the set $\mathcal{R}([a,b],V)$ nor the value of any integral changes.

For Exercises 6.13, 6.14, and 6.15 below, you simply need to go through the proofs of the corresponding statements for real-valued functions, and observe that if you replace absolute-value symbols (if they occur at all) by norm-symbols, the same arguments work verbatim.

Exercise 6.13 Show that the statements in Exercise 6.1 for functions $f : [a,b] \to \mathbf{R}$ also hold for functions $f : [a,b] \to V$.

Exercise 6.14 Establish the analog of Example 6.11 for V-valued functions: For any $v \in V$, the constant function $f : [a,b] \to V$ given by $f(x) = v$ is integrable, and

$$\int_a^b v\,dx = (b-a)v.$$

Proposition 6.83 (linearity of the integral) *The set $\mathcal{R}([a,b],V)$ is a vector space (a vector subspace of* $\mathrm{Func}([a,b],V)$*), and the map $\mathcal{R}([a,b],V) \to V$ defined by $f \mapsto \int_a^b f$ is linear.*

Exercise 6.15 Prove Proposition 6.83.

Since a general vector space is not an ordered set (statements such as "$v < w$" for $v, w \in V$ are *meaningless* unless $V = \mathbf{R}$ or $V = \{0_V\}$), there are no analogs of Proposition 6.18 or Corollary 6.19 for V-valued functions (for general V). For the same reason, there are no analogs of upper and lower sums. However, we used upper and lower sums only as a tool to simplify proofs and to aid in visualization of certain facts. Most facts about integrable real-valued functions that do not *explicitly* (i.e. in their statements, not just their proofs) rely on the fact that $\mathbf{R}$ is ordered, *do* generalize to V-valued functions. For some of these facts, we will have to use a different proof-strategy, since we often used the fact that $\mathbf{R}$ is ordered as a crutch to simplify proofs (and often to gain useful insight!). The proofs of results such as "integrable implies bounded" and "continuous implies integrable", given in this section for V-valued functions, would have worked just as well earlier for $\mathbf{R}$-valued functions.

Proposition 6.84 ("Integrable implies bounded") *If $f : [a,b] \to V$ is integrable, then f is bounded.*

Proof: Let $f \in \mathcal{R}([a,b],V)$, and let $A = \int_a^b f$. Let $\delta > 0$ be such that $\mathcal{S}_\delta(f) \subset B_1(A)$. Fix a partition $P = \{x_0, \ldots x_N\}$ of $[a,b]$ of width less than δ.

Assume that f is unbounded. Then f is unbounded on at least one of the intervals $I_j := [x_{j-1}, x_j]$, since there are only finitely many such intervals. Let $j_0 \in \{1, \ldots, N\}$ be such that f is unbounded on I_{j_0}. For each $n \in \mathbf{N}$, choose $z_n \in I_{j_0}$ such that $\|f(z_n)\| > n$; such z_n exist by the unboundedness assumption. For each $j \in \{1, \ldots, N\}$ with $j \neq j_0$, fix any number $t_j \in [x_{j-1}, x_j]$, let $T^{(n)}$ be the pointing $\{t_1^{(n)}, \ldots, t_N^{(n)}\}$ of P for which $t_j^{(n)} = \begin{cases} t_j \text{ if } j \neq j_0, \\ z_n \text{ if } j = j_0\,, \end{cases}$ and let $A' = \sum_{j \neq j_0} f(t_j)\Delta_j$. Then, using the triangle inequality,

$$\begin{aligned} \|S(f;P,T^{(n)}) - A\| = \|f(z_n)\Delta_{j_0} + A' - A\| &\geq \|f(z_n)\Delta_{j_0}\| - \|A - A'\| \\ &= \|f(z_n)\|\Delta_{j_0} - \|A - A'\| \\ &> n\Delta_{j_0} - \|A - A'\|. \end{aligned}$$

For n sufficiently large, $n\Delta_{j_0} - \|A - A'\| > 1$, implying that $S(f;P,T^{(n)}) \notin B_1(A)$, a contradiction.

Hence f is bounded. ■

To get rid of our reliance on upper and lower sums in various proofs, we need to establish Theorem 6.32's "(i) $\iff$ (iv)" implication in a way that does not use the order structure of $\mathbf{R}$ (in particular, a way that does involve statement (ii) or (iii) of that theorem). We do this by copying Rosenlicht's Lemma 1 in [5, Section VI.3], simply replacing "real-valued" replaced by "V-valued", and absolute values by norms:

Proposition 6.85 *Let $f \in \mathrm{Func}([a,b],V)$. Then f is integrable if and only if, for each $\epsilon > 0$, there exists $\delta > 0$ such that for all $S_1, S_2 \in \mathcal{S}_\delta(f)$, we have $\|S_1 - S_2\| < \epsilon$.*

Proof: First assume that f is integrable. Let $A = \int_a^b f$ and let $\epsilon > 0$. Let $\delta > 0$ be such that $\mathcal{S}_\delta(f) \subset B_{\epsilon/2}(A)$. Then for all $S_1, S_2 \in \mathcal{S}_\delta(f)$ we have

$$\|S_1 - S_2\| = d(S_1, S_2) \leq d(S_1, A) + d(A, S_2) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This proves the "only if" assertion of the proposition.

Conversely, assume that for each $\epsilon > 0$, there exists $\delta > 0$ such that for all $S_1, S_2 \in \mathcal{S}_\delta(f)$, we have $\|S_1 - S_2\| < \epsilon$. Let $(P_n)_{n=1}^\infty$ be a sequence of partitions of $[a,b]$, and $(S^{(n)})_{n=1}^\infty$ a sequence of Riemann sums of f, such that for all n we have $\mathrm{wid}(P_n) < \frac{1}{n}$ and $S^{(n)} \in \mathcal{S}(f;P_n)$. Let $\epsilon > 0$, and let δ be such that for all $S_1, S_2 \in \mathcal{S}_\delta(f)$, we have $\|S_1 - S_2\| < \epsilon$. Let $N \in \mathbf{N}$ be any integer greater than $1/\delta$. Then for all $n, m \geq N$ the partitions P_n, P_m both have widths less than δ, so $\|S^{(n)} - S^{(m)}\| < \epsilon$. Therefore the sequence $(S^{(n)})$ in (V, d) is Cauchy. Since (V, d) is complete, this sequence converges; let A denote its limit.

Again let $\epsilon > 0$ be arbitrary, and now let $\delta > 0$ be such that for all $S_1, S_2 \in \mathcal{S}_\delta(f)$, we have $\|S_1 - S_2\| < \frac{\epsilon}{2}$. Let $N \in \mathbf{N}$ be such that $N > \frac{1}{\delta}$ and $\|S^{(N)} - A\| < \frac{\epsilon}{2}$; such N exists since $(S^{(n)})$ converges to A. For every $S \in \mathcal{S}_\delta(f)$ we then have

$$d(S, A) \leq d(S, S^{(N)}) + d(S^{(N)}, A) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Hence f is integrable. ■

Exercise 6.16 Prove that the analog of Proposition 6.55, "Additivity of the integral", holds for V-valued functions. The first half of the proof of Proposition 6.55 can be mimicked fairly easily. For the second half, which made use of upper and lower sums, you will need to figure out how to use Proposition 6.85 in place of Proposition 6.45, the "Step-function lemma".

Definition 6.86 Let V^* denote the set of *continuous* linear transformations from $V \to \mathbf{R}$. (V^* is called the *dual space* or *continuous dual* of V.)

The handout "Some notes on normed vector spaces" (http://dgarchive.com/classes/4212_s19/misc_handouts/normed_vector_spaces.pdf) proves, among other things, several facts we will need concerning linear transformations from V to $\mathbf{R}$. We collect these here into a proposition so that we may refer to them easily:

Proposition 6.87

(a) A linear transformation $\xi : V \to \mathbf{R}$ is continuous if and only if there exists a real number K such that

$$|\xi(v)| \leq K\|v\| \quad \text{for all } v \in V. \tag{6.59}$$

(b) If V is finite-dimensional, *then every linear transformation from V to $\mathbf{R}$ is continuous.*

(c) If V is finite-dimensional, *then any two norms on V are equivalent.*

We mention a few things in passing:

- The "normed vector spaces" handout actually proves (a) and (b) for linear transformations from V to any normed vector space, not just $\mathbf{R}$.

- Facts (b) and (c) are false if is V infinite-dimensional.
- For fact (a), all that we will need is the "only if" part.

One additional fact that we will use, not mentioned in the "normed vector spaces" handout, is that a finite-dimensional vector space, endowed with any norm, is complete. This follows from facts proven (one hopes) in MAA 4211: (i) If d_1, d_2 are equivalent metrics on a set E ("equivalence" being defined the same way as for norms on vector spaces), then (E, d_1) is complete if and only if (E, d_2) is complete. (ii) If two norms on a vector space are equivalent, so are their associated metrics. (iii) If V has finite dimension $n \geq 1$, and $\| \ \|$ is the ℓ^∞ (or the ℓ^2) norm determined by some choice of basis, then V is complete with respect to the associated metric.

Thus, every finite-dimensional normed vector space is a complete normed vector space.

Returning to general V (not necessarily finite-dimensional): given any $f \in \mathrm{Func}([a,b], V)$ and any $\xi \in V^*$, the composition $\xi \circ f$ is a real-valued function on $[a,b]$. The next proposition relates the integrability, and the integrals, of the V-valued function f and the real-valued function $\xi \circ f$. Before we state the proposition, the student should do the following easy exercise relating *Riemann sums* of the V-valued function f and the real-valued function $\xi \circ f$.

Exercise 6.17 Show that for any $f \in \mathrm{Func}([a,b], V)$ and $\xi \in V^*$, and any pointed partition (P, T) of $[a,b]$,

$$\xi\left(S(f; P, T)\right) = S(\xi \circ f; P, T). \tag{6.60}$$

Proposition 6.88 *If $f \in \mathcal{R}([a,b], V)$, then for every $\xi \in V^*$ we have $\xi \circ f \in \mathcal{R}([a,b])$, and*

$$\xi\left(\int_a^b f\right) = \int_a^b \xi \circ f. \tag{6.61}$$

Proof: Let $f \in \mathcal{R}([a,b], V)$, let $\xi \in V^*$, and let $A = \int_a^b f$. Let $K > 0$ be such that (6.59) is satisfied. Let $\epsilon > 0$ be given, let $\epsilon_1 = \frac{\epsilon}{K}$, and let $\delta > 0$ be such that $\mathcal{S}_\delta(f) \subset B^V_{\epsilon_1}(A)$.

Now let (P, T) be a pointed partition of $[a,b]$ of width less than δ. Then $S(f; P, T) \in B^V_{\epsilon_1}(A)$, and, using equation (6.60),

$$\begin{aligned} |S(\xi \circ f; P, T) - \xi(A)| = |\xi\left(S(f; P, T)\right) - \xi(A)| &= |\xi\left(S(f; P, T) - A\right)| \\ &\leq K\|S(f; P, T) - A\| \\ &< K\epsilon_1 = \epsilon, \end{aligned}$$

so $S(\xi \circ f; P, T) \in B^{\mathbf{R}}_\epsilon(\xi(A))$.

Hence $\mathcal{S}_\delta(\xi \circ f) \subset B^{\mathbf{R}}_\epsilon(A)$. Since ϵ was arbitrary, it follows that $\xi \circ f \in \mathcal{R}([a,b])$ and that 6.61 holds. ■

Results much stronger than Proposition 6.88 are true if V is finite-dimensional. We show one of these next, and deduce as a corollary that if V is finite-dimensional, the "if ... then" in Proposition 6.88 can be strengthened to "if and only if". For the next few pages, to make visually clear which objects are elements of V and which are real numbers, we will use boldface for elements of V and for V-valued functions. (However, $\mathbf{R}$ still denotes the reals!)

A function $\mathbf{f} : [a,b] \to \mathbf{R}^n$ is often written in the form $(f_1, \dots, f_n)$, where the f_i are real-valued functions on $\mathbf{R}^n$. We can also write $(f_1, \dots, f_n)$ as $\sum_{i=1}^n f_i \mathbf{e}_i$, where $\{\mathbf{e}_i\}_{i=1}^n$ is the standard basis of $\mathbf{R}^n$ ($\mathbf{e}_i$ is the vector whose i^{th} coordinate is 1, and all of whose other coordinates are 0). The *coordinate functions determined by the basis* $\{\mathbf{e}_i\}_{i=1}^n$ are exactly the usual coordinate functions $\{x_i : \mathbf{R}^n \to \mathbf{R}\}_{i=1}^n$ (the functions defined by $x_i(a_1, a_2, \dots, a_n) = a_i$). Observe that $f_i = x_i \circ \mathbf{f}$. The student should keep this concrete example in mind when reading the next proposition, while remembering that $\mathbf{R}^n$ is just *one example* of an n-dimensional vector space, and that the standard basis of $\mathbf{R}^n$ is just *one example* of a basis of $\mathbf{R}^n$.

Proposition 6.89 *Assume that V has finite dimension $n \geq 1$ and let $\{\mathbf{v}_i\}_{i=1}^n$ be a basis of V. Let $\mathbf{f} \in \mathrm{Func}([a,b], V)$, and let $f_1, \dots, f_n$ be the unique real-valued functions on $[a,b]$ defined by writing $\mathbf{f}$ pointwise in terms of a basis:*

$$\mathbf{f}(x) = \sum_{i=1}^n f_i(x)\, \mathbf{v}_i \quad \textit{for all } x \in [a,b].$$

Then the V-valued function $\mathbf{f}$ is integrable if and only if each of the real-valued functions f_i is integrable. In the integrable case,

$$\int_a^b (f_1\mathbf{v}_1 + \cdots + f_n\mathbf{v}_n) = \left(\int_a^b f_1\right)\mathbf{v}_1 + \cdots + \left(\int_a^b f_n\right)\mathbf{v}_n\ . \tag{6.62}$$

Proof: Let $\{\xi_i : V \to \mathbf{R}\}_{i=1}^n$ be the coordinate functions on V determined by the basis $\{\mathbf{v}_i\}_{i=1}^n$. (Thus $\xi_i(\sum_j a_j \mathbf{v}_j) = a_i$, $\mathbf{w} = \sum_{i=1}^n \xi_i(\mathbf{w})\mathbf{v}_i$ for all $\mathbf{w} \in V$, and $f_i = \xi_i \circ \mathbf{f}$ for $1 \leq i \leq n$.) Then for each $i \in \{1, \dots, n\}$, the function ξ_i is a linear transformation $V \to \mathbf{R}$, so by Proposition 6.87 parts (b) and (a), $\xi_i \in V^*$ and there exists $K_i > 0$ such that $|\xi_i(\mathbf{w})| \leq K_i \|\mathbf{w}\|$ for all $\mathbf{w} \in V$. Select such $K_1, \dots, K_n$ and let $K = \max\{K_i : 1 \leq i \leq n\}$.

First assume that $\mathbf{f}$ is integrable on $[a,b]$, and let $\mathbf{A} = \int_a^b \mathbf{f}(x)\,dx$. For $1 \leq i \leq n$ let $A_i = \xi_i(\mathbf{A})$; thus $\mathbf{A} = \sum_{i=1}^n A_i \mathbf{v}_i$. Let $\epsilon > 0$, let $\epsilon_1 = \epsilon/K$, and let $\delta > 0$ be such that $\mathcal{S}_\delta(\mathbf{f}) \subset B^V_{\epsilon_1}(\mathbf{A})$.

Let (P,T) be a pointed partition of $[a,b]$ of width less than δ. For each $i \in \{1,2,\ldots,n\}$ let $S_i = S(f_i;P,T)$. Define $\mathbf{S} = S(\mathbf{f};P,T)$. Since $\mathrm{wid}(P) < \delta$, we have $\|\mathbf{S}-\mathbf{A}\| < \epsilon_1$.

Fix $i \in \{1,2,\ldots,n\}$. Since $\xi_i : V \to \mathbf{R}$ is linear, we may apply Exercise 6.17 to obtain

$$\xi_i(\mathbf{S}) = \xi_i\left(S(\mathbf{f};P,T)\right) = S(\xi_i \circ \mathbf{f};P,T) = S(f_i;P,T) = S_i.$$

Hence, again using the linearity of ξ_i,

$$|S_i - A_i| = |\xi_i(\mathbf{S}) - \xi_i(\mathbf{A})| = |\xi_i(\mathbf{S}-\mathbf{A})| \le K_i\|\mathbf{S}-\mathbf{A}\| < K\epsilon_1 = \epsilon.$$

Therefore we have produced $\delta > 0$ such that for every arbitrary pointed partition (P,T) of width less than δ, we have $S(f_i;P,T) \in B^{\mathbf{R}}_{\epsilon}(A_i)$. Since ϵ was arbitrary, this proves that f_i is integrable and that $\int_a^b f_i = A_i$. Since $i \in \{1,\ldots,n\}$ was arbitrary, this is true for every i, and

$$\int_a^b \mathbf{f} = \mathbf{A} = \sum_{i=1}^n A_i\mathbf{v}_i = \sum_{i=1}^n \left(\int_a^b f_i\right)\mathbf{v}_i.$$

We have now shown that if $\mathbf{f}$ is integrable on $[a,b]$, then (i) each component function f_i is integrable on $[a,b]$, and (ii) the equality (6.62) holds. For the converse of the integrability implication, assume now that f_i is integrable on $[a,b]$ for $1 \le i \le n$.

Let $A_i = \int_a^b f_i$, $1 \le i \le n$, and let $\mathbf{A} = \sum_{i=1}^n A_i\mathbf{v}_i$. Let $\epsilon > 0$, let $C = \sum_{i=1}^n \|\mathbf{v}_i\|$, and let $\epsilon_1 = \epsilon/C$. For $1 \le i \le n$ let $\delta_i > 0$ be such that $\mathcal{S}_{\delta_i}(f_i) \subset B^{\mathbf{R}}_{\epsilon_1}(A_i)$, and let $\delta = \min\{\delta_i : 1 \le i \le n\}$.

Let (P,T) be a pointed partition of $[a,b]$ of width less than δ, and let $\mathbf{S} = S(\mathbf{f};P,T)$. Then, again using Exercise 6.17,

$$\begin{aligned}
\mathbf{S}-\mathbf{A} &= \sum_{i=1}^n \xi_i(\mathbf{S})\mathbf{v}_i - \sum_{i=1}^n A_i\mathbf{v}_i \\
&= \sum_{i=1}^n \left[\xi_i(S(\mathbf{f};P,T)) - A_i\right]\mathbf{v}_i \\
&= \sum_{i=1}^n \left[S(\xi_i \circ \mathbf{f};P,T) - A_i\right]\mathbf{v}_i \\
&= \sum_{i=1}^n \left(S(f_i;P,T) - A_i\right)\mathbf{v}_i\ .
\end{aligned}$$

Hence

$$\begin{aligned}
\|\mathbf{S}-\mathbf{A}\| &\leq \sum_{i=1}^{n} \|(S(f_i;P,T)-A_i)\,\mathbf{v}_i\| \\
&= \sum_{i=1}^{n} |S(f_i;P,T)-A_i|\;\|\mathbf{v}_i\| \\
&< \sum_{i=1}^{n} \epsilon_1\|\mathbf{v}_i\| \\
&= \epsilon_1 C = \epsilon.
\end{aligned}$$

Therefore we have produced $\delta > 0$ such that $\|\mathbf{S}-\mathbf{A}\| < \epsilon$ for all $\mathbf{S} \in \mathcal{S}_\delta(\mathbf{f})$. Since ϵ was arbitrary, it follows that $\mathbf{f}$ is integrable on $[a,b]$. ■

Corollary 6.90 *Assume that V is finite-dimensional and let $\mathbf{f} \in \mathrm{Func}([a,b],V)$. Then $\mathbf{f} \in \mathcal{R}([a,b],V)$ if and only if for every $\xi \in V^*$ we have $\xi \circ \mathbf{f} \in \mathcal{R}([a,b])$.*

Proof: The "only if"part of the implication follows from Proposition 6.88. For the "if" part, assume that for every $\xi \in V^*$ we have $\xi \circ \mathbf{f} \in \mathcal{R}([a,b])$. If $\dim(V) = 0$ then trivially $\mathbf{f} \in \mathcal{R}([a,b],V)$, so assume that $n := \dim(V) \geq 1$. Let $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ be a basis of V, and, as in the proof of Proposition 6.89, let $\{\xi_i : V \to \mathbf{R}\}_{i=1}^{n}$ be the corresponding coordinate functions on V. Then for each i, we have $\xi_i \in V^*$, so (by our hypothesis) $\xi_i \circ \mathbf{f}$ is integrable. But $\xi_i \circ \mathbf{f}$ is exactly the function f_i in the statement of Proposition 6.89. Hence that Proposition implies that $\mathbf{f}$ is integrable. ■

Remark 6.91 Equation (6.62) *formally* looks very similar to

$$\int_a^b \left(\sum_{i=1}^{m} c_i f_i\right) = \sum_{i=1}^{m} c_i \int_a^b f_i \quad \text{(where } c_1,\ldots c_m \in \mathbf{R}\text{)}, \tag{6.63}$$

just with the real constants c_i in (6.63) replaced by "vector constants" $\mathbf{v}_i$ that happen to form a basis of V. But (6.62) and (6.63) are really very different statements. For real-valued functions $f_1,\ldots,f_m$, the equality (6.63) is one version of the statement that (i) $\mathcal{R}([a,b])$, the set of integrable *real-valued* functions on $[a,b]$, is a vector space and that (ii) "$\int_a^b$" is a linear map $\mathcal{R}([a,b]) \to \mathbf{R}$. The only meaning of "$\int_a^b$" in equation (6.63) is integration of a *real-valued* function on $[a,b]$. The number of functions m is arbitrary; it's not related to the dimension of anything (unlike the n in (6.62)). The corresponding statement for V-valued functions is *not* (6.62); it's that (i) $\mathcal{R}([a,b],V)$ is a vector space and that (ii) the map $\int_a^b : \mathcal{R}([a,b],V) \to V$ is linear:

$$\int_a^b \left(\sum_{j=1}^m c_j \mathbf{f}_j(x) \right) dx = \sum_{j=1}^m c_j \int_a^b \mathbf{f}_j(x)\ dx \tag{6.64}$$
$$\text{for all integers } m > 0, \text{ all } \mathbf{f}_1, \dots, \mathbf{f}_m \in \mathcal{R}([a,b],V), \text{ and all } c_1, \dots, c_m \in \mathbf{R}.$$

In (6.64), like in (6.63), the notation "$\int_a^b$" has only one meaning, but in (6.64) the meaning is integration of a V-*valued* function on $[a,b]$.

Equation (6.62) may be *interpreted, informally*, as saying that the basis vectors $\mathbf{v}_j$ behave as "vector constants" that can be pulled through the integral sign "just like" scalar constants (real numbers). But the "just like" is inaccurate. As noted above, in equation (6.63) the notation "$\int_a^b$" has the same meaning on both sides of the equation; it is a *single* operator (fancy name for function) on *one* vector space, $\mathcal{R}([a,b],V)$. In (6.62), the same notation "$\int_a^b$" is used for *two different operators*, the one on the left-hand side having domain $\mathcal{R}([a,b],V)$, and the one on the right-hand side having domain $\mathcal{R}([a,b])$. The operators are conceptually similar, but they have very different domains. It is important to keep in mind that while "Vector constants can be pulled through the integral sign just like scalar constants" is something that could be conjectured, or even expected, before proving anything, there is no such thing as "proof by analogy".

We will say more about equation (6.62) later in Remark 6.98, after establishing some more results. ▲

Together, the next two propositions generalize Exercise 6.5 from real-valued functions to V-valued functions (with V assumed finite-dimensional).

Proposition 6.92 *Assume that V is finite-dimensional. If $\mathbf{f} : [a,b] \to V$ is integrable, then the real-valued function $x \mapsto \|\mathbf{f}(x)\|$ is integrable.*

Proof: Let $g : [a,b] \to \mathbf{R}$ denote the function $x \mapsto \|\mathbf{f}(x)\|$.

If $\dim(V) = 0$ then $\mathbf{f}$ is the constant function 0_V and g is the constant function 0, which is integrable.

Assume now that $n := \dim(V) \geq 1$ and let $\{\mathbf{v}_i\}_{i=1}^n$ be a basis of V. Since every element of V is a unique linear combination $\sum_i a_i \mathbf{v}_i$, we can define a function $\|\ \|_1 : V \to \mathbf{R}$ by $\|\sum_i a_i \mathbf{v}_i\|_1 = \sum_i |a_i|$. As the student may easily show, $\|\ \|_1$ is a norm on V. (The proof is virtually identical to the proof that the ℓ^1-norm on $\mathbf{R}^n$ is a norm.)

As in Proposition 6.89, let $f_1, \dots f_n$ be the component-functions of $\mathbf{f}$ determined by this basis, i.e. the unique real-valued functions such that $\mathbf{f} = \sum_i f_i \mathbf{v}_i$. By Proposition

6.89, each component-function f_i is integrable, hence also bounded (thus Theorem 6.32 applies to f_i).

By Proposition 6.87(3), the norm $\| \|_1$ is equivalent to the given norm $\| \|$ on V. Let $c > 0$ be such that for all $\mathbf{v} \in V$, $\|\mathbf{v}\| \le c\|\mathbf{v}\|_1$.

Let $\epsilon > 0$. For each $i \in \{1, \dots, n\}$ let $\delta_i > 0$ be such that $U_{\delta_i}(f_i) - L_{\delta_i}(f_i) < \frac{\epsilon}{cn}$; such δ_i exist by the "(i) $\Longrightarrow$ (iii)" implication of Theorem 6.32. Let $\delta = \min\{\delta_1, \dots, \delta_n\}$. Then for each $P \in \mathcal{P}_\delta([a, b])$ and each $i \in \{1, \dots, n\}$ we have

$$U(f_i; P) - L(f_i; P) \le U_\delta(f_i) - L_\delta(f_i) \le U_{\delta_i}(f_i) - L_{\delta_i}(f_i) < \frac{\epsilon}{cn} .$$

Let $P = \{x_0, \dots, x_N\} \in \mathcal{P}_\delta([a, b])$. For $1 \le i \le n$ and $1 \le j \le N$ let $M_{i,j} = \sup\{f_i(x) : x \in [x_{j-1}, x_j]\}$ and $m_{i,j} = \inf\{f_i(x) : x \in [x_{j-1}, x_j]\}$. Observe that for any $s, t \in [x_{j-1}, x_j]$, and any $i \in \{1, \dots, n\}$, we have

$$|f_i(s) - f_i(t)| \le M_{i,j} - m_{i,j}. \tag{6.65}$$

Let $T = \{t_1, \dots, t_N\}$ and $T' = \{t'_1, \dots, t'_N\}$ be arbitrary pointings of P. Then, using the triangle inequality in the form $\|\mathbf{a}\| - \|\mathbf{b}\| \le \|\mathbf{a} - \mathbf{b}\|$, we have

$$\begin{aligned}
S(g; P, T) - S(g; P, T') &= \sum_{j=1}^{N} (\|\mathbf{f}(t_j)\| - \|\mathbf{f}(t'_j)\|)\Delta_j \\
&\le \sum_{j=1}^{N} \|\mathbf{f}(t_j) - \mathbf{f}(t'_j)\| \, \Delta_j \\
&\le \sum_{j=1}^{N} c\|\mathbf{f}(t_j) - \mathbf{f}(t'_j)\|_1 \, \Delta_j \\
&= c\sum_{j=1}^{N} \left(\sum_{i=1}^{n} |f_i(t_j) - f_i(t'_j)| \right) \Delta_j \\
&= c\sum_{i=1}^{n} \left(\sum_{j=1}^{N} |f_i(t_j) - f_i(t'_j)|\Delta_j \right) \\
&\le c\sum_{i=1}^{n} \left(\sum_{j=1}^{N} (M_{i,j} - m_{i,j})\Delta_j \right) \quad \text{(using (6.65))} \\
&= c\sum_{i=1}^{n} (U(f_i; P) - L(f_i; P)) \\
&< c\sum_{i=1}^{n} \frac{\epsilon}{cn} \\
&= \epsilon.
\end{aligned}$$

Thus $S(g; P, T) - S(g; P, T') < \epsilon$ for all pointings T, T' of P. Taking the supremum over T and then the infimum over T', we deduce that $U(g; P) - L(g; P) \leq \epsilon$. Since ϵ was arbitrary, it follows from Proposition 6.45 that g is integrable. ■

We are done restricting attention to finite-dimensional V for now, so we resume using non-boldface letters for elements of V and for V-valued functions.

Proposition 6.93 ("Triangle inequality for integrals") *Let $f \in \mathcal{R}([a, b], V)$, and let $\|f(\cdot)\| : [a, b] \to \mathbf{R}$ denote the function $x \mapsto \|f(x)\|$. Then*

$$\left\| \int_a^b f \right\| \leq \lim_{\delta \to 0} U_\delta(\|f(\cdot)\|). \tag{6.66}$$

Hence if $\|f(\cdot)\|$ is integrable,

$$\left\| \int_a^b f(x)\, dx \right\| \leq \int_a^b \|f(x)\|\, dx. \tag{6.67}$$

Our nickname "triangle inequality for integrals" really refers only to inequality (6.67). The reason for this nickname is discussed later in item 6 of Remark 6.98.

Proof of Proposition 6.93: Let us write $A = \int_a^b f$ and $g = \|f(\cdot)\|$.

Let $\epsilon > 0$, and let $\delta > 0$ be such that $\mathcal{S}_\delta(f) \subset B_\epsilon^V(A)$. Let $P \in \mathcal{P}_\delta([a, b])$ and let $T = \{t_1, \dots, t_N\}$ be a pointing of P. Then, using the triangle inequality,

$$\begin{aligned} \|S(f; P, T)\| = \left\| \sum_{j=1}^N f(t_j)\Delta_j \right\| &\leq \sum_{j=1}^N \|f(t_j\|\, \Delta_j \\ &= S(g; P, T) \\ &\leq U(g; P) \ \ \text{(by the definition of } U(g; P)) \\ &\leq U_\delta(g) \ \ \text{(by the definition of } U_\delta(g)). \end{aligned}$$

But $\|S(f; P, T) - A\| < \epsilon$, so

$$\|A\| \leq \|A - S(f; P, T)\| + \|S(f; P, T)\| < \epsilon + \|S(f; P, T)\| \leq \epsilon + U_\delta(g);$$

i.e. $\|A\| < U_\delta(g) + \epsilon$. Hence, by an order-property of limits of real-valued functions,

$$\|A\| \leq \lim_{\delta \to 0} U_\delta(g) + \epsilon.$$

Since ϵ was arbitrary, we conclude that $\|A\| \leq \lim_{\delta \to 0} U_\delta(g)$, which is (6.66).

If g is integrable, then $\lim_{\delta\to 0} U_\delta(g) = \int_a^b g$ (by Theorem 6.32), so (6.66) reduces to (6.67) in this case. ■

Observe that, by Proposition 6.92, if V is finite-dimensional, then under the hypotheses of Proposition 6.93 the function $\|f(\cdot)\|$ is automatically integrable, so the stronger conclusion (6.67) holds. We record this fact later in Corollary 6.97.

Remark 6.94 (Optional reading, intended for students who have read Section 6.4.) In view of the Darboux theorem mentioned in Section 6.4, we can alternatively write (6.66) as

$$\left\| \int_a^b f(x)\, dx \right\| \leq \overline{\int_a^b} \|f(x)\|\, dx.$$

▲

Proposition 6.95 ("Continuous implies integrable") *If $f : [a,b] \to V$ is continuous, then f is integrable.*

Proof: Let f be a continuous function from $[a,b]$ to V. Since $[a,b]$ is compact, f is uniformly continuous. Let $\epsilon > 0$, and let $\delta > 0$ be such that if $x, y \in [a,b]$ and $|x-y| < \delta$ then $\|f(x) - f(y)\| < \epsilon_1 := \frac{\epsilon}{2(b-a)}$.

Let $P_1 = \{x_0, \ldots, x_{N_1}\}$, $P_2 = \{y_0, \ldots, y_{N_2}\} \in \mathcal{P}_\delta([a,b])$. Let $P = P_1 \cup P_2 = \{z_0, \ldots, z_N\}$; then $P \in \mathcal{P}_\delta([a,b])$ as well. For $1 \leq j \leq N_1$ let $i_j \in \{1, 2, \ldots N\}$ be the index for which $x_j = z_{i_j}$. (It is helpful to draw a diagram of the interval $[a,b]$ to follow the proof from this point on.) Let $T_1 = \{t_1, \ldots, t_{N_1}\}$, $T = \{s_1, \ldots, s_N\}$ be pointings of P_1, P respectively. Then

$$\begin{aligned} S(f;P_1,T_1) - S(f;P,T) &= \sum_{j=1}^{N_1} f(t_j)(x_j - x_{j-1}) - \sum_{i=1}^{N} f(s_i)(z_i - z_{i-1}) \\ &= \sum_{j=1}^{N_1} f(t_j) \left(\sum_{i=i_{j-1}+1}^{i_j} (z_i - z_{i-1}) \right) - \sum_{j=1}^{N_1} \left(\sum_{i=i_{j-1}+1}^{i_j} f(s_i)(z_i - z_{i-1}) \right) \\ &= \sum_{j=1}^{N_1} \left(\sum_{i=i_{j-1}+1}^{i_j} (f(t_j) - f(s_i))(z_i - z_{i-1}) \right). \end{aligned}$$

Note that in the expression "$f(t_j) - f(s_i)$" on the last line, we have $i_{j-1} \leq i - 1 < i \leq i_j$, implying $x_{j-1} \leq z_{i-1} \leq s_i \leq z_i \leq x_j$. Thus s_i lies in $[x_{j-1}, x_j]$, as does t_j. Since $x_j - x_{j-1} < \delta$, we have $|t_j - s_i| < \delta$, implying $\|f(t_j) - f(s_i)\| < \epsilon_1$. Therefore, applying the iterated triangle inequality, we have

$$
\begin{aligned}
\|S(f;P_1,T_1) - S(f;P,T)\| &\leq \sum_{j=1}^{N_1} \left\| \left(\sum_{i=i_{j-1}+1}^{i_j} (f(t_j) - f(s_i))(z_i - z_{i-1}) \right) \right\| \\
&\leq \sum_{j=1}^{N_1} \left(\sum_{i=i_{j-1}+1}^{i_j} \|(f(t_j) - f(s_i))(z_i - z_{i-1})\| \right) \\
&= \sum_{j=1}^{N_1} \left(\sum_{i=i_{j-1}+1}^{i_j} (z_i - z_{i-1}) \|(f(t_j) - f(s_i))\| \right) \\
&< \sum_{j=1}^{N_1} \left(\sum_{i=i_{j-1}+1}^{i_j} (z_i - z_{i-1}) \epsilon_1 \right) && (6.68) \\
&= \epsilon_1 \sum_{i=1}^{N} (z_i - z_{i-1}) && (6.69) \\
&= \epsilon_1 (b-a) && (6.70) \\
&= \frac{\epsilon}{2} \,. && (6.71)
\end{aligned}
$$

Thus, $\|S(f;P_1,T_1) - S(f;P,T)\| < \frac{\epsilon}{2}$. Similarly, $\|S(f;P_2,T_2) - S(f;P,T)\| < \frac{\epsilon}{2}$. Hence

$$
\begin{aligned}
\|S(f;P_1,T_1) - S(f;P_2,T_2)\| &\leq \|S(f;P_1,T_1) - S(f;P,T)\| + \|S(f;P,T) - S(f;P_2,T_2)\| \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}
$$

Therefore for any $S_1, S_2 \in \mathcal{S}_\delta(f)$ we have $\|S_1 - S_2\| < \epsilon$. Since ϵ was arbitrary, it follows from Proposition 6.85 that f is integrable on $[a,b]$. ■

Remark 6.96 The argument above gives a second proof that continuous *real*-valued functions on $[a,b]$ are integrable, without relying on the Proposition 6.45 (the "Step-function lemma").

Corollary 6.97 *Let $f \in \mathcal{R}([a,b],V)$. If either (a) V is finite-dimensional or (b) f is continuous, then the real-valued function $x \mapsto \|f(x)\|$ is integrable, and*

$$
\left\| \int_a^b f(x)\, dx \right\| \leq \int_a^b \|f(x)\|\, dx. \tag{6.72}
$$

Proof: Let g denote the function $x \mapsto \|f(x)\|$. Under hypothesis (a), Proposition 6.92 implies that g is integrable. Under hypothesis (b), g is the composition of the continuous function $f : [a, b] \to V$ with the continuous function $\| \ \| : V \to \mathbf{R}$. (Students: why is the latter function is continuous?) Hence g is continuous, so by Theorem 6.53, g is integrable.

Thus, under either hypothesis (a) or (b), the function g is integrable. Therefore Proposition 6.93 shows that the inequality (6.72) holds. ■

Remark 6.98 It is worthwhile to revisit the Calculus 3 version of (6.62):

$$\int_a^b (f_1(t)\mathbf{i} + f_2(t)\mathbf{j} + f_3(t)\mathbf{k})\, dt = \left(\int_a^b f_1(t)\, dt\right)\mathbf{i} + \left(\int_a^b f_2(t)\, dt\right)\mathbf{j} + \left(\int_a^b f_3(t)\, dt\right)\mathbf{k}. \tag{6.73}$$

In Calc 3 we are taught that equation (6.73) is the *definition* of the integral on the left-hand side. *Now that we have proven Proposition 6.89,* we see that for the case $V = \mathbf{R}^3$, with any norm (since they are all equivalent), the definition given in Calc 3 is equivalent to the loftier Definition 6.80. But the loftier definition, while requiring more sophistication and more work, has several advantages:

1. It shows *from the start* that we can define integrals of V-valued functions for any finite-dimensional vector space, not just $\mathbf{R}^n$.
2. It shows *from the start* that, given a finite-dimensional vector space V, we do not need to introduce a basis of V in order to define integrals of V-valued functions.
3. By defining the integral without reference to a basis, the loftier definition *guarantees* that the value of the integral is independent of the choice of basis (a fact that needs to be *proven*, even for the case $V = \mathbf{R}^3$, if we use the Calc 3 definition).
4. It directly incorporates the principle that *integration is about adding stuff up* (the "stuff density" being the function we're integrating), which equation (6.73), taken as a definition, does not.
5. It is an *elegant* generalization of the definition of integrals of real-valued functions: essentially nothing changed in passing from Definitions 6.6 and 6.8 to Definitions 6.79 and 6.80; all we had to do was to replace absolute-value symbols by norm symbols.
6. It enables us to prove Proposition 6.93 very easily, and to show *why*, for finite-dimensional V, we obtain the inequality in Corollary 6.97 *for any norm whatsoever on V*. We saw that the inequality (6.66) follows simply from applying the (iterated)

triangle inequality to Riemann sums. The stronger inequality (6.72)—simply (6.67) stated a second time—then followed (for finite-dimensional V) as soon as we showed that the pointwise norm of an integrable V-valued function is an integrable real-valued function, which we saw is true for *any* norm on V. Inequality (6.67) or (6.72) can be viewed as *the iterated triangle inequality generalized from finite sums to integrals* (hence the nickname we have given to Proposition 6.93).

With the generalized Calc 3 definition, the result of Corollary 6.97 in the finite-dimensional case can also be proven, with a little cleverness but not much difficulty, for a *Euclidean* norm on V—one that comes from an inner product. A standard argument (starting from the generalized Calc 3 definition), presented in [6,7], makes use of the Cauchy-Schwartz inequality for inner products to obtain (6.72). Unfortunately, this argument obscures the fundamental reason *why* (6.72) *ought* to be true. In addition, not every norm on $\mathbf{R}^n$, or on a general finite-dimensional vector space, is Euclidean (an ℓ^2 norm); it need not even be an ℓ^p-norm for *any* p. Thus, with a general norm on V, the argument based on the Cauchy-Schwartz inequality does not yield even the *integrability* of the pointwise norm of an integrable V-valued function, let alone the inequality (6.72).

7. The loftier definition tells us (after proving Proposition 6.89) *why* equation (6.73) *should* be true; that it's not just a definition introduced for convenient bookkeeping.

8. The loftier definition does not even require V to be finite-dimensional; it requires only that V be a complete normed vector space. (We have seen several examples of infinite-dimensional complete normed vector spaces in MAA 4211–4212: the space $\ell^\infty(\mathbf{R})$, and the space $C(X)$ for any compact metric space X of infinite cardinality.) The Calc 3 definition does not generalize to infinite-dimensional V.

The advantages listed above of the loftier definition are only one side of the coin, however. Even at levels more advanced that of MAA 4211–4212, there are good textbooks (such as [6, 7]) and good teachers who prefer the "generalized Calc 3 definition" of an $\mathbf{R}^n$-valued function (with Calc 3's $\mathbf{R}^3$ replaced by $\mathbf{R}^n$), assume boundedness from the start. This approach defines-away the need to prove Propositions 6.84, 6.88, and 6.89, and thereby enables other results to be written down sooner[12], albeit at the expense of some insight and generality. (Additional time is saved, in this approach, by asserting and proving Proposition 6.92 only for the Euclidean norm, rather than for an *arbitrary* norm.) In this approach, the fact that the integral of an integrable $\mathbf{R}^n$-valued function has a basis-independent characterization is something *proven*, rather than something that drops out of the definition of "integral". This is an instance of something quite common in mathematics: often there are two (or more) approaches to the same topic, with some theorems in approach A being definitions in approach B and vice-versa. ▲

[12]As a practical matter, this can be very important, since semesters have finite length! But several times in his mathematical life, the author of these notes has been grateful that he learned the "invariant" definition, i.e. Definition 6.80, early on, so his preference is to expose students to this approach.

Remark 6.99 The Fundamental Theorem of Calculus (FTC) also generalizes to V-valued functions. However, since in MAA 4212 we have not yet defined derivatives of anything other than real-valued functions, the statement and proof of the FTC in this setting are not included in this chapter.

6.10 Logarithmic and exponential functions

Since the function $t \mapsto \frac{1}{t}$ is continuous on $(0, \infty)$, it is integral over any closed interval with endpoints in this interval exists. Hence we may define a function as follows:

Definition 6.100 The function $\log : (0, \infty) \to \mathbf{R}$ is defined by

$$\log x := \log(x) = \int_1^x \frac{1}{t}\, dt \quad \text{for each } x > 0. \tag{6.74}$$

Remark 6.101 The function "log" defined above is the *natural logarithm* function that you are probably used to denoting "ln". Mathematicians tend to use the notation "log" for the natural log function rather than "ln" (except when teaching lower-level calculus courses and other courses populated largely by engineering majors), and "$\log_{10}$" for the function you may be used to denoting simply as "log", because $\log_{10}$ has no special mathematical significance. Prior to the age of pocket calculators, $\log_{10}$ had much greater *practical* significance than it has now; bankers, scientists, and other people who used to need to do a lot of multiplication by hand, used tables of values of the $\log_{10}$ function in order to reduce multiplication to the addition of logs. Nowadays, these tasks are done by computers and pocket calculators. The $\log_{10}$ function still survives in a few $\log_{10}$-based scales in the sciences, such as the pH scale in chemistry and the decibel scale for sound-intensity. We also still use the phrase "order of magnitude" in a sense coming from the $\log_{10}$ function, since human beings brought up with base-10 arithmetic naturally find it easy to think in terms of how many powers of 10 are involved.

Proposition 6.102 *The function* $\log : (0, \infty) \to \mathbf{R}$ *is differentiable, strictly increasing, bijective, and satisfies the following for all* $x, y \in (0, \infty)$ *and all integers* n*:*

(i) $\log'(x) = \frac{1}{x}$ *(where "*$\log'$*" denotes the derivative of* $\log$*.)*

(ii) $\log(1) = 0$.

(iii) $\log(xy) = \log x + \log y$.

(iv) $\log(\frac{1}{x}) = -\log x$.

(v) $\log \frac{x}{y} = \log x - \log y$.

(vi) $\log(x^n) = n \log x$.

Proof: Since $t \mapsto \frac{1}{t}$ is continuous, the Fundamental Theorem of Calculus implies that log is differentiable and that $\log'(x) = \frac{1}{x}$. Since $\frac{1}{x} > 0$ for all $x > 0$, log is a strictly increasing (hence injective) function; we will show later that its range is all of $\mathbf{R}$.

Property (ii) is immediate from the defining equation (6.74). To establish property (iii), let $x, y > 0$. By Corollary 6.60, we have

$$\log(xy) = \int_1^{xy} \frac{1}{t}\,dt = \int_1^{x} \frac{1}{t}\,dt + \int_x^{xy} \frac{1}{t}\,dt = \log x + \int_x^{xy} \frac{1}{t}\,dt. \tag{6.75}$$

In the last integral in (6.75), we may make the substitution $t = xs$. (More formally, we define the function $\varphi : [\min\{1, y\}, \max\{1, y\}] \to [\min\{x, xy\}, \max\{x, xy\}]$ by $\varphi(s) = xs$. Then $\varphi(1) = x$, $\varphi(y) = xy$, and $\varphi'(s) = x$ for all s. See Remarks 6.75 and 6.76.) Applying the change-of-variable theorem (Proposition 6.74), we have

$$\int_x^{xy} \frac{1}{t}\,dt = \int_1^{y} \frac{1}{xs}\,x\,ds = \int_1^{y} \frac{1}{s}\,ds = \log y.$$

Hence (6.75) implies that $\log(xy) = \log x + \log y$.

Property (iv) now follows from properties (ii) and (iii) (since $x\,\frac{1}{x} = 1$), and property (v) then follows from properties (iii) and (iv) (since $\frac{x}{y} = x\,\frac{1}{y}$).

Property (vi) holds for $n = 0$ by property (ii), and is true trivially for $n = 1$. Using property (iii) and induction, it follows easily that property (vi) holds for all $n \geq 1$ (details are left to the student). From this and property (iv), we then find that property (vi) holds for $n \leq -1$ as well.

We have now proven everything except that the range of log is $\mathbf{R}$. For this, we first note that since $\frac{1}{t} \geq \frac{1}{2}$ for $t \in [1, 2]$, Corollary 6.19 implies

$$\log 2 = \int_1^2 \frac{1}{t}\,dt \geq \int_1^2 \frac{1}{2} dt = \frac{1}{2},$$

so

$$\log 2 \geq \frac{1}{2} > 0. \tag{6.76}$$

Let $y \in [0, \infty)$, and let n be a positive integer such that $n \log 2 \geq y$; such n exists since $\log 2 > 0$. Then, by property (vi), we have $\log(2^n) \geq y$, so $y \in [\log(1), \log(2^n)]$. Since log is differentiable, log is continuous, so the Intermediate Value Theorem implies that there exists $x \in [1, 2^n]$ such that $\log x = y$. Hence log achieves every non-negative real value. Property (iv) then shows that log achieves every non-positive real value as well, hence achieves every real value. Thus the range of log is $\mathbf{R}$, as claimed. ■

Remark 6.103 In deriving (6.76), we used only Corollary 6.19, which yields the "$\geq$" in (6.76), since that sufficed for the proof of Proposition 6.102. But using Exercise 6.6(b), we can actually deduce the strict inequality $\log 2 > \frac{1}{2}$. ▲

Remark 6.104 (A pause to smell the roses) Thanks to our hard work in Sections 6.1–6.7, the proof of part (i) of Proposition 6.102 was very short, so let us take a moment to reflect on something we've achieved with the formula "$\log'(x) = x^{-1}$". When we first learn calculus, the first functions we learn how to differentiate are the power-functions $x \mapsto x^n$, where n is a positive integer or zero. Shortly thereafter, we work out the derivative for negative n as well, discovering the beautiful fact that $\frac{d}{dx}x^n = nx^{n-1}$ for *all* integers n. Later, when we start to study antidifferentiation, our first tool is *recognition*: having seen power-functions arise as derivatives of other power functions, we can easily invert the process. Since $3x^2$ is the derivative of x^3, we know that any multiple of x^2 will have some multiple of x^3 as an antiderivative. More generally, our derivative formula tells us that, for every integer n, x^n is an antiderivative of nx^{n-1}, hence that $\frac{x^n}{n}$ is an antiderivative of x^{n-1}, hence that $\frac{x^{n+1}}{n+1}$ is an antiderivative of x^n—with one exception, the case $n = -1$. We never see x^{-1} arising as the derivative of a multiple of a power function, so at this early stage of our learning, we have no way to find an antiderivative. But the qualitative Corollary 6.63 tells us that there *is* one, and the quantitative Theorem 6.62 gives an outright *formula* for one. The gap is filled![13]

Since $\log : (0, \infty) \to \mathbf{R}$ is bijective, it has an inverse, so we may make the following definition:

Definition 6.105 We define the function $\exp : \mathbf{R} \to (0, \infty)$, to be the inverse of $\log : (0, \infty) \to \mathbf{R}$. ▲

We will shortly establish some important properties of the exponential function "exp". We first record as lemmas a couple of facts we will need. The first of these is an easy consequence of results proven in MAA 4211 (Rosenlicht problem IV.4 and the Intermediate Value Theorem):

Lemma 6.106 *Let $I \subset \mathbf{R}$ be an interval, $f : I \to \mathbf{R}$ a continuous, strictly monotone function. Then* $\mathrm{range}(f)$ *is an interval J, f is a bijection $I \to J$ (hence has an inverse), and the inverse function $f^{-1} : J \to I$ is continuous.*

Exercise 6.18 Prove Lemma 6.106.

[13]Unfortunately, the all-too-popular "early transcendentals" calculus textbooks rob students of an appreciation of how marvelous this is, and make it difficult to fall in love with calculus.

Next is a very useful fact, also easy to prove, that you may have seen proven in MAA 4211 (where you may also have seen the terminology "limit point" rather than the synonymous terminology "cluster point").

Lemma 6.107 ("Substitution lemma for limits") *Let X, Y, Z be metric spaces, p_0 a cluster point of X, and q_0 a cluster point of Y. Let $f : X\setminus\{p_0\} \to Y\setminus\{q_0\}, g : Y\setminus\{q_0\} \to Z$ be functions, and assume that $\lim_{p\to p_0} f(p) = q_0$ and that $\lim_{q\to q_0} g(q)$ exists. Then*

$$\lim_{p\to p_0} g(f(p)) = \lim_{q\to q_0} g(q). \tag{6.77}$$

Remark 6.108 (1) Lemma 6.107 is stated in such a way that f and g do not *need* to be defined at p_0 and q_0 respectively, to allow us to apply the lemma in such cases (one of which arises in the next Proposition). Nothing goes *wrong* if f is defined at p_0, or g is defined at q_0. (2) If, in place of f as above, we have a function $f : X \setminus \{p_0\} \to Y$ whose range may include q_0, but such that $f(B \setminus \{p_0\}) \subset Y \setminus \{p_0\}$ for some ball B containing p_0, then the hypotheses of the lemma are met with $(B, f|_{B\setminus\{p_0\}})$ playing the role of the lemma's (X, f). Hence the equality (6.77) holds under these more general hypotheses.

Exercise 6.19 Prove Lemma 6.107.

Proposition 6.109 *The function* $\exp : \mathbf{R} \to (0,\infty)$ *is differentiable, strictly increasing, bijective, and satisfies the following for all $x, y \in (0,\infty)$ and all integers n:*

(i) $\exp' = \exp$ *(where "$\exp'$" denotes the derivative of* $\exp$*.)*

(ii) $\exp(0) = 1$.

(iii) $\exp(x + y) = \exp(x)\exp(y)$

(iv) $\exp(-x) = \frac{1}{\exp(x)}$.

(v) $\exp(x - y) = \frac{\exp(x)}{\exp(y)}$.

(vi) $\exp(nx) = (\exp(x))^n$.

Proof: The fact that exp is bijective and strictly increasing follow from the fact that it is the inverse of a function with these properties. Since log is continuous, Lemma 6.106 implies that exp is continuous.

We next show that $\exp' = \exp$. Let $x_0 \in \mathbf{R}$ and let $y_0 = \exp(x_0)$ (hence $x_0 = \log y_0$). Define $g : (0,\infty) \setminus \{y_0\} \to \mathbf{R}$ by $g(y) = \frac{y-y_0}{\log y - \log y_0}$. Note that $\lim_{y\to y_0} \frac{1}{g(y)} = \log'(y_0)$ (by definition of $\log'$), hence equals $\frac{1}{y_0}$ (by Proposition 6.102). Therefore

$$\lim_{y\to y_0} g(y) = \lim_{y\to y_0} \frac{1}{\frac{1}{g(y)}} = \frac{1}{\lim_{y\to y_0} \frac{1}{g(y)}} = \frac{1}{\frac{1}{y_0}} = y_0.$$

Since exp is one-to-one and $\exp(x_0) = y_0$, the range of $\exp|_{\mathbf{R}\setminus\{x_0\}}$ does not include y_0. Hence we may define a function $f : \mathbf{R} \setminus \{x_0\} \to (0,\infty) \setminus \{y_0\}$ by $f(x) = \exp(x)$ (implying $\log f(x) = x$). Since exp is continuous, $\lim_{x\to x_0} f(x) = f(x_0) = \exp(x_0) = y_0$. Hence we may apply Lemma 6.107, yielding

$$\lim_{x\to x_0} g(f(x)) = \lim_{y\to y_0} g(y) = y_0 = \exp(x_0). \tag{6.78}$$

But $g(f(x)) = \frac{f(x)-y_0}{\log f(x)-\log y_0} = \frac{\exp(x)-\exp(x_0)}{x-x_0}$. Hence $\lim_{x\to x_0} \frac{\exp(x)-\exp(x_0)}{x-x_0} = \lim_{x\to x_0} g(f(x))$, which, by (6.78), exists and equals $\exp(x_0)$. Thus exp is differentiable at x_0, and $\exp'(x_0) = \exp(x_0)$. Since x_0 was arbitrary, we conclude that the functions $\exp'$ and exp are identical.

Properties, (ii)—(vi) of exp follow from the corresponding properties for log. For example, for (iii), given any $x, y \in \mathbf{R}$, by the bijectivity of log there exist unique $a, b \in (0,\infty)$ such that $x = \log a$ and $y = \log(b)$. We then have

$$x + y = \log a + \log b = \log(ab) = \log(\exp(x)\exp(y)),$$

implying that $\exp(x + y) = \exp(\log(\exp(x)\exp(y))) = \exp(x)\exp(y)$. Derivations of the remaining properties are left to the student. ■

Remark 6.110 Our proof of the differentiability of exp—the inverse of log—and the computation of $\exp'$, can be easily generalized to the following theorem:

> **" 'Baby' Inverse Function Theorem"**: *Let $I, J \subset \mathbf{R}$ be open intervals, let $h : I \to J$ a surjective, continuously differentiable function such that, for all $y \in I$, $h'(y) \neq 0$. Then h is bijective, the function $f = h^{-1} : J \to I$ is differentiable, and for all $x \in J$ we have $f'(x) = \frac{1}{h'(f(x))}$.*

▲

Consider now any positive, real number a. Letting $x = \log a$, property (vi) in Proposition 6.109, read from right to left, says that for any integer n we have $a^n = \exp(n \log a)$. In view of this fact, the following definition does not alter the meaning of a^n for any integer n, but gives meaning to "a^n" for all *real* n:

Definition 6.111 *Let $a, r \in \mathbf{R}$, with $a > 0$. We define the number $a^r \in (0,\infty)$ by*

$$a^r = \exp(r \log a).$$

The next proposition may be summarized as saying that the "usual algebra of exponentiation" for integer exponents holds for real exponents.

Proposition 6.112 *Let $a, b, x, y \in \mathbf{R}$, with $a > 0$ and $b > 0$. Then:*

(i) $a^0 = 1$.

(ii) $a^{x+y} = a^x a^y$.

(iii) $a^{-x} = \frac{1}{a^x}$.

(iv) $(a^x)^y = a^{xy}$.

(v) $(ab)^x = a^x b^x$.

(vi) If $a > 1$ then the map $x \mapsto a^x$ is strictly increasing; if $a < 1$ then this map is strictly decreasing.

Proof: All these properties follow quickly from Proposition 6.109 and Definition 6.111. For example, for (ii) we have

$$a^{x+y} = \exp((x+y)\log a) = \exp(x \log a + y \log a) = \exp(x \log a)\exp(y \log a) = a^x a^y.$$

The remaining parts of the proof are left as an exercise to the student. ∎

Exercise 6.20 Complete the proof of Proposition 6.112.

Observe that Definition 6.111 says nothing about how to define 0^r. Of course, if r is a positive integer, we already have a purely algebraic definition of 0^r, yielding $0^r = 0$. Using the fact that for positive integers n, the unique n^{th} root of 0 is 0, we can naturally extend the definition "$0^r = 0$" to all positive *rational* r. But irrational r cannot be handled by these purely algebraic means. For these, we need a separate definition, which we will write in a way that applies in both the rational-r and irrational-r cases:

Definition 6.113 For every $r > 0$, we define $0^r = 0$.

We do not define the expression "0^r" for $r \leq 0$; in particular, we do not define "0^0". One motivation for not defining "0^0" is part (b) of the following exercise. Part (a) of the exercise provides one motivation for why we *do* define $0^r = 0$ for all $r > 0$.

Exercise 6.21 Using Definitions 6.111 and 6.113, define $\tilde{f} : ([0,\infty) \times [0,\infty)) \setminus \{(0,0)\} \to \mathbf{R}$ by $\tilde{f}(x,y) = x^y$. Let f be the restriction of $\tilde{f}$ to the domain $(0,\infty) \times [0,\infty)$; observe that $\text{domain}(\tilde{f}) = \text{domain}(f) \cup (\{0\} \times (0,\infty))$.

(a) Show that $\tilde{f}$ is the unique continuous extension of f to $([0,\infty) \times [0,\infty)) \setminus \{(0,0)\}$. (Note that you need to show two facts about $\tilde{f}$, not necessarily in the following order:

(i) that $\tilde{f}$ is continuous, and (ii) that any continuous extension of f to $([0,\infty)\times[0,\infty))\setminus\{(0,0)\}$ is the function $\tilde{f}$.)

(b) Show that there does not exist a continuous extension of f to $[0,\infty)\times[0,\infty) = \text{domain}(\tilde{f})\cup\{(0,0)\}$. (*Suggestion*: Use the fact that any such extension would also be an extension of $\tilde{f}$.)

(c) Define $\tilde{g} : ([0,\infty)\times\mathbf{R})\setminus(\{0\}\times(-\infty,0]) \to \mathbf{R}$ by $\tilde{g}(x,y) = x^y$. Let g be the restriction of $\tilde{g}$ to the domain $(0,\infty)\times\mathbf{R}$; observe that $\text{domain}(\tilde{g}) = \text{domain}(g)\cup(\{0\}\times(0,\infty))$. Redo parts (a) and (b) with $\tilde{f}$ and f replaced by $\tilde{g}$ and g, respectively.

Exercise 6.22 Determine how Proposition 6.112 generalizes if we allow $a \geq 0$ and/or $b \geq 0$. Are any restrictions on x and/or y needed (possibly different restrictions for different parts of the proposition)? If so, what?

Remark 6.114 (Rational exponents) Proposition 6.112 shows that Definition 6.111 is consistent with prior definitions (e.g. from MAA 4211) of a^r for *rational* exponents r. For example, for $a > 0$ and n a positive integer, Proposition 6.112(iv) shows that $(a^{1/n})^n = a^{n/n} = a$. Since the function $x \mapsto x^n$ is strictly increasing on $(0,\infty)$, $a^{1/n}$ is therefore the *unique* positive real number c such that $c^n = a$, i.e. the number that in MAA 4211 we defined to be the (positive) n^{th} *root of* a. Similarly, if p, q are integers and $q \neq 0$, Proposition 6.112 shows that, consistently with prior definitions of "$a^{p/q}$", we have

$$(a^{1/q})^p = a^{p/q} = (a^p)^{1/q}. \tag{6.79}$$

However, Definition 6.111 gives a much "cleaner", if less intuitive, definition of a^r for $r \in \mathbf{Q}$ than does taking either the first or second equality in (6.79) to be the definition of $a^{p/q}$, because a rational number does not have a *unique* expression as a quotient of integers; e.g. $\frac{2}{3} = \frac{16}{24} = \frac{-42}{-63}$. When we attempt to use (say) the first equality in (6.79) as the definition of $a^{p/q}$ when p, q are positive integers, we must do one of the following in order to ensure that $a^{p/q}$ is well-defined: (1) require that the exponent be expressed in "lowest terms", i.e. with p and q having no common divisor greater than 1, or (2) show that if $\frac{p}{q} = \frac{p'}{q'}$, where p, q, p', q' are positive integers, then $(a^{1/q'})^{p'} = (a^{1/q})^p$. (Since every rational number *can* be expressed in lowest terms, (2) can be reduced to the case in which $p' = kp$ and $q' = kq$ for some positive integer k.) Approach (1), however, becomes insufficient the moment we try to show that rational exponents obey property (ii) in Proposition 6.112, based only on the algebra of integer exponents and on a definition of $a^{p/q}$ that requires p and q to be relatively prime. For example, $\frac{1}{5} + \frac{3}{10} = \frac{1}{2}$, but you will not likely succeed in showing that $a^{1/5}a^{3/10} = a^{1/2}$ without knowing that $a^{1/5} = (a^{1/10})^2$ and that $(a^{1/10})^5 = a^{1/2}$. Similarly, you will have difficulty showing that $2^{0.6} > 2^{0.5}$ without knowing that $(2^{1/5})^3 = (2^{1/10})^6$ and that $2^{1/2} = (2^{1/10})^5$. Thus, if we attempt to take (6.79) as the definition of a^r for rational non-integer r, then to obtain the results of Proposition 6.112 even just for rational exponents, based on knowing them for integer exponents, we are forced to prove (at least) that $(a^{1/kq})^{kp} = (a^{1/q})^p$ for all positive integers p, q, and k. This is not *difficult* to prove, but the necessity of proving it can't be avoided if we attempt to use one of the equalities in (6.79) as the definition of a^r for non-integer rational r. ▲

Remark 6.115 You may be accustomed to thinking that the algebraic rules in Proposition 6.112 are "obvious", even though you likely have been using one of the equalities in (6.79) as the definition of a^r when r is rational, positive, and expressed in lowest terms, and have likely been taking Proposition 6.112(iii) as the definition of a^r when r is rational and negative. There is nothing incorrect about these prior definitions. However, as Remark 6.114 shows, if we use these prior definitions then the algebraic rules in Proposition 6.112 are not at all obvious once we leave the realm of integer exponents. We're simply so used to these rules that we may forget that there was ever anything non-obvious about them. ▲

Remark 6.116 (Irrational exponents, part 1) Modern students, having grown up with pocket calculators that have an "x^y" button on them, may be not be conscious of the fact that there is nothing obvious about what an expression like "$2^{\sqrt{2}}$" should mean. We can use (6.79) to define what it means to raise a number to a *rational* exponent, but this equation says nothing about irrational exponents.

Let r be an irrational number. We may attempt to define a^r in an *ad hoc* manner, using the decimal expansion of r as a sequence in $\mathbf{Q}$ approaching r from below (e.g. using the fact that $\sqrt{2}$ is the limit of a sequence 1, 1.4, 1.41, 1.414, 1.4142, ...), and tentatively defining a^r to be the limit of this sequence. Why should the limit exist? If we first do the work mentioned in Remark 6.114, we can then show that this sequence is monotone and bounded, hence convergent. But that's not entirely satisfying (nor can it be rigorously justified early in Calculus 1, let alone prior to calculus): should the value of the number a^r depend on the fact that humans have 10 fingers (the reason that we chose the *decimal* expansion of r)? For *rational* exponents there is no such dependence, so we would certainly *hope* that there is none for irrational exponents either. This leads us to want to prove at least the following: if $(r_n)_{n=1}^{\infty}$ is an increasing sequence of rational numbers approaching r, does $\lim_{n\to\infty} a^{r_n}$ is independent of the choice of the sequence (r_n).

Even this is not wholly satisfying. What if we had chosen a sequence (r_n) that *decreases* to r instead of *increasing* to r? (For example, if $r = -\sqrt{2}$, the sequence $-1, -1.4, -1.41, -1.414, -1.4142, \ldots$ is a decreasing sequence.) What if we had chosen a non-monotonic sequence (r_n) with limit r? Do we always get the same limit? Using (6.79) we can, indeed, prove that $\lim_{n\to\infty} a^{r_n}$ exists for every sequence in $\mathbf{Q}$ converging to r, and that the value of this limit is independent of the choice of sequence (r_n).[14] Using this limit to define a^r, we then have a definition of a^r that is valid for all real r and all

[14]But the proofs of these statements are far beyond the level at which today's students are generally taught this approach to defining irrational powers. An "it can be shown" statement is required, and it's for something that really *can't* be shown at the level of Calculus 1.

$a > 0$. Can we then prove Proposition 6.112 based on this approach? Yes, but still more work is involved, and some questions with non-obvious answers have to be addressed. By contrast, when we use Definition 6.111 to define a^r for all real r, we have a definition that works simultaneously for rational and irrational exponents, that looks identical for *all* real exponents, and that renders the proof of Proposition 6.112 essentially trivial. ▲

Definition 6.117 The number $e \in \mathbf{R}$ is defined by $e = \exp(1)$. ▲

From Definitions 6.111 and 6.117, we immediately have

$$e^x = \exp(x)$$

for all $x \in \mathbf{R}$.

Proposition 6.118 *Let $a, r \in \mathbf{R}$, with $a > 0$. The function $\mathbf{R} \to \mathbf{R}$ defined by $x \mapsto a^x$, and the function $(0, \infty) \to \mathbf{R}$ defined by $x \mapsto x^r$, are differentiable, and their derivatives are given by the following formulas:*

(i) $\frac{d}{dx} a^x = a^x \log a$.

(ii) $\frac{d}{dx} x^r = r x^{r-1}$.

Proof: (i) Define $f, g : \mathbf{R} \to \mathbf{R}$ by $f(x) = a^x$ and $g(x) = x \log a$. Then, by definition, $f = \exp \circ g$. Since exp and g are differentiable, the Chain Rule Theorem implies that f is differentiable and that $f'(x) = \exp'(g(x))g'(x) = \exp(g(x)) \log a = a^x \log a$.

(ii) Define $f, g : \mathbf{R} \to \mathbf{R}$ by $f(x) = x^r$ and $g(x) = r \log x$. Then, by definition, $f = \exp \circ g$. Since exp and g are differentiable, the Chain Rule Theorem implies that f is differentiable and that $f'(x) = \exp'(g(x))g'(x) = \exp(g(x)) \frac{r}{x} = x^r \frac{r}{x}$. Using Proposition 6.112(ii)–(iii), this last expression equals $r x^{r-1}$. ■

Remark 6.119 (Exponential and logarithm functions) For each $a > 0$, the function $x \mapsto a^x$ may be called the *exponential function with base a*. All such functions are called "exponential functions". In this terminology, the function exp is the exponential function with base e. This is the function that is meant when the terminology "*the* exponential function" is used (properly).

Note that for any $k \in \mathbf{R}$, $a^{kx} = (a^k)^x$, so every function of the form $a \mapsto a^{kx}$ is also an exponential function.

Since $a^x = \exp(x \log a)$, and $\log a$ is positive for $a > 1$ and negative for $a < 1$, Proposition 6.109 implies that the exponential function with base a is strictly increasing if $a > 1$ and strictly decreasing if $a < 1$ (and is the constant function 1 if $a = 1$), and has range $(0, \infty)$ if $a \neq 1$. Hence for $a \neq 1$ this function has an inverse. This inverse function is called *the logarithm function with base* a, and is denoted $\log_a$ (which is best read "log, base a" [15]; the expression $\log_a x := \log_a(x)$ is best read "log, base a, of x"). Any such function is known as a *logarithm function*, or simply "log function". Observe that $\log_e$ is the same as the *natural logarithm* function that we're denoting simply as log, but which you're probably used to writing as "ln". ▲

Exercise 6.23 Show that for $a, b, x > 0$ and $a \neq 1 \neq b$,

$$\log_b x = \frac{\log_a x}{\log_a b}.$$

Thus, any logarithm function is a constant times any other.

Remark 6.120 (Power functions and their derivatives) For each $r \in \mathbf{R}$, Definition 6.111 defines the expression "x^r" for all $x > 0$, and, as we have seen, this definition agrees with definitions we have previously learned for $r \in \mathbf{Q}$. However, for certain r we do not need $x > 0$ for the expression x^r to be defined. (For example, if n is a positive integer, basic algebra produces a definition of x^n for all $x \in \mathbf{R}$, and we then define $x^{-n} = \frac{1}{x^n}$ for all $x \neq 0$. We also define $x^0 = 1$ for all $x \neq 0$, in order that the property in Proposition 6.112(iii) hold for all real $a \neq 0$ and all integer exponents. For odd integers n, every real number has a unique n^{th} root, so we may define $x^{1/n}$ for all real x.) For each $r \in \mathbf{R}$, and any set $U \subset \mathbf{R}$ such that $x \mapsto x^r$ is defined for all $x \in U$, the function $U \to \mathbf{R}$ given by $x \mapsto x^r$ is called a *power function*, or the r^{th}*-power function*. (Of course, for certain r, we have other names as well, e.g. the *squaring function*, for $r = 2$ and the *cube-root function* for $r = 1/3$.) We still use the name "(r^{th}-)power function" (or these other names) if the codomain $\mathbf{R}$ is replaced by any set containing $\{x^r : x \in U\}$, as in "the squaring function $\mathbf{R} \to [0, \infty)$" or "the cube-root function $(8, \infty) \to (2, \infty)$".

The remainder of this Remark is optional reading; the student may skip to Remark 6.121.

In Calculus 1, one of the first things we learn is that for *positive integers* r,

$$\frac{d}{dx} x^r = r x^{r-1}; \tag{6.80}$$

[15] This is one of two common ways that the notation "$\log_a$" is read. The other, "log to the base a", appears to be idiomatic—it makes no sense grammatically, unlike alternatives such as "log *from* base a" or "log *with* base a"—but *is* the terminology used in the classic textbook [11] by Thomas, from which many current mathematicians learned calculus. Based on the terminology "raising *to* a power", which (at least for integer exponents) is probably much older than any terminology for logs with bases other than 10, if $y = a^x$ there is logical justification to say that "x is the logarithm of y *from* the base a," or that "x is the log, *from* the base a, of y." By contrast, if we apply conventional rules of grammar and usage, "log to the base a" is not consistent with the terminology for powers, or with any other uses of "to" in English.

to derive this fact we use the Binomial Theorem. We also learn that the derivative of a constant function is 0, so that (6.80) holds for $r = 0$ as well (on the domain $\mathbf{R} \setminus \{0\}$), not as a consequence of the Binomial Theorem, but because x^0 has been defined to be 1 for $x \neq 0$. In a *good* Calculus 1 course[16], we learn progressively that (6.80) holds for more and more exponents, until we have shown that it is true for all *rational* exponents:

1. By one of several methods, we learn that

$$\frac{d}{dx}x^{-1} = -x^{-2}, \tag{6.81}$$

which is (6.80) for $r = -1$. Simple methods by which (6.81) can be shown, without using any "laws of exponents" for anything other than integer exponents (the only exponents for which it is *obvious* that rules like Proposition 6.112(ii), (iv), and (v) are valid), are:

   (a) Direct calculation: defining $f : \mathbf{R} \setminus \{0\} \to \mathbf{R}$ by $f(x) = x^{-1}$, we compute

$$f'(x) = \lim_{h\to 0} \frac{\frac{1}{x+h} - \frac{1}{x}}{h} = \lim_{h\to 0} \frac{\frac{x-(x+h)}{x(x+h)}}{h} = \lim_{h\to 0} \frac{-1}{x(x+h)} = \frac{-1}{x^2} .$$

   (b) First learning the quotient rule, then computing

$$\frac{d}{dx}\frac{1}{x} = \frac{x\frac{d}{dx}(1) - 1\frac{d}{dx}(x)}{x} = \frac{x \cdot 0 - 1}{x^2} = -x^{-2}.$$

(Method (a) is really a special case of the proof of the quotient rule, so it is not entirely different from method (b).)

2. By one of several methods, we learn that for *all* positive integers n,

$$\frac{d}{dx}x^{-n} = -nx^{-n-1} \tag{6.82}$$

on $\mathbf{R} \setminus \{0\}$, which is (6.80) for $r =$ [negative integer]. Simple methods by which (6.82) can be shown are, without using any "laws of exponents" for anything other than integer exponents, are:

   (a) First learning the quotient rule, then computing

$$\frac{d}{dx}\frac{1}{x^n} = \frac{x^n\frac{d}{dx}(1) - 1\frac{d}{dx}(x^n)}{(x^n)^2} = \frac{x^n \cdot 0 - nx^{n-1}}{x^{2n}} = -nx^{-n-1}.$$

[16] "Good", as defined by the writer of these notes.

(b) First establishing (6.81), then learning the Chain Rule, and then computing

$$\frac{d}{dx}x^{-n} = \frac{d}{dx}(x^n)^{-1} = -(x^n)^{-2}\frac{d}{dx}(x^n) = -(x^{-2n})nx^{n-1} = -nx^{-n-1}.$$

Combining (6.82) with the previously-established cases of (6.80), we have now learned that (6.80) holds for all *integer* exponents.

3. We show, by one of several methods, that for positive integers n, the function $x \mapsto x^{1/n}$ is differentiable on $(0, \infty)$, and compute that

$$\frac{d}{dx}x^{\frac{1}{n}} = \frac{1}{n}x^{\frac{1}{n}-1}. \tag{6.83}$$

on this interval. Two methods by which (6.82) can be shown, without using any "laws of exponents" for anything other than *rational* exponents, are:

(a) Proving the " 'Baby' Inverse Function Theorem" (see Remark 6.110), and then applying it to $h : x \mapsto x^n$. This shows that the function $f : x \mapsto x^{1/n}$ is differentiable and that

$$\frac{d}{dx}x^{1/n} = f'(x) = \frac{1}{h'(f(x))} = \frac{1}{\frac{d}{dy}y^n\big|_{y=x^{1/n}}} = \frac{1}{n(x^{\frac{1}{n}})^{n-1}} = \frac{1}{n(x^{1-\frac{1}{n}})} = \frac{1}{n}x^{\frac{1}{n}-1}.$$

(b) Restricting attention to $n \geq 2$ (sufficient, since (6.83) has been proven already for $n = 1$), showing first that $f : x \mapsto x^{1/n}$ is continuous on $(0, \infty)$), and then using the algebraic identity $b^n - a^n = (b-a)(b^{n-1} + b^{n-2}a + b^{n-3}a^2 + \cdots + a^{n-1})$ (for $a, b, \in \mathbf{R}$) to compute

$$
\begin{aligned}
f'(x) &= \lim_{h\to 0}\frac{(x+h)^{\frac{1}{n}} - x^{\frac{1}{n}}}{h} \\
&= \lim_{h\to 0}\frac{\left[(x+h)^{\frac{1}{n}} - x^{\frac{1}{n}}\right]\left[\left\{(x+h)^{\frac{1}{n}}\right\}^{n-1} + \left\{(x+h)^{\frac{1}{n}}\right\}^{n-2}x^{\frac{1}{n}} + \cdots + (x^{\frac{1}{n}})^{n-1}\right]}{h\left[\left\{(x+h)^{\frac{1}{n}}\right\}^{n-1} + \left\{(x+h)^{\frac{1}{n}}\right\}^{n-2}x^{\frac{1}{n}} + \cdots + (x^{\frac{1}{n}})^{n-1}\right]} \\
&= \lim_{h\to 0}\frac{\left\{(x+h)^{\frac{1}{n}}\right\}^{n} - (x^{\frac{1}{n}})^{n}}{h\left[\left\{(x+h)^{\frac{1}{n}}\right\}^{n-1} + \left\{(x+h)^{\frac{1}{n}}\right\}^{n-2}x^{\frac{1}{n}} + \cdots + (x^{\frac{1}{n}})^{n-1}\right]} \\
&= \lim_{h\to 0}\frac{(x+h) - x}{h\left[\left\{(x+h)^{\frac{1}{n}}\right\}^{n-1} + \left\{(x+h)^{\frac{1}{n}}\right\}^{n-2}x^{\frac{1}{n}} + \cdots + (x^{\frac{1}{n}})^{n-1}\right]} \\
&= \lim_{h\to 0}\frac{1}{\left\{(x+h)^{\frac{1}{n}}\right\}^{n-1} + \left\{(x+h)^{\frac{1}{n}}\right\}^{n-2}x^{\frac{1}{n}} + \cdots + (x^{\frac{1}{n}})^{n-1}} \\
&= \frac{1}{\underbrace{\left\{x^{\frac{1}{n}}\right\}^{n-1} + \left\{x^{\frac{1}{n}}\right\}^{n-2}x^{\frac{1}{n}} + \cdots + (x^{\frac{1}{n}})^{n-1}}_{n \text{ terms}}} \\
&= \frac{1}{nx^{1-\frac{1}{n}}} \\
&= \frac{1}{n}x^{\frac{1}{n}-1}.
\end{aligned}
$$

If n is odd, so that $x^{1/n}$ is defined for all x, the methods in (a) and (b) extend from the domain $(0,\infty)$ to the domain $\mathbf{R}\setminus\{0\}$.

4. Having established (6.80) for the cases in which n is an integer or the reciprocal of a positive integer, we apply the Chain Rule Theorem to generalize to other rational exponents, as follows[17]: for an arbitrary rational number $r = \frac{m}{n}$, where m, n are

[17] In our good calculus course, we do not investigate other exponents until we have proven the Chain Rule Theorem (CRT). However, once we have established the CRT, an alternative approach to deriving (6.80) for general rational exponents that does not require that we first handle reciprocal-integer exponents separately, is as follows: (1) Introduce implicit differentiation (which depends crucially on the chain rule). (2) *State* the Implicit Function Theorem (for real-valued functions of a single real variable), advising the student that the proof is beyond the scope of a Calculus 1 course. (3) For an arbitrary rational number $r = \frac{m}{n}$, where m, n are integers and $n > 0$, show that the Implicit Function Theorem implies that the equation $y^n = x^m$ (with $(x, y) \in (0,\infty)\times(0,\infty)$) defines y as a differentiable function of x on $(0,\infty)$, hence that the function $x \mapsto x^{m/n}$ is differentiable on $(0,\infty)$. (4) Implicitly differentiate "$y^n = x^m$" with respect to x, and then apply rules of rational exponents, to deduce that $\frac{d}{dx}x^{m/n} = \frac{dy}{dx} = \frac{m}{n}x^{m/n-1}$.

integers and $n > 0$, we have

$$\frac{d}{dx}x^r = \frac{d}{dx}(x^{\frac{1}{n}})^m = m(x^{\frac{1}{n}})^{m-1}\frac{d}{dx}x^{\frac{1}{n}} = m(x^{\frac{1}{n}})^{m-1}\frac{1}{n}x^{\frac{1}{n}-1} = \frac{m}{n}x^{\frac{m-1}{n}+\frac{1}{n}-1} = rx^{r-1}.$$

In our good Calculus 1 course, we have now *shown* that (6.80) is true for all rational exponents, using no laws of exponents that we did not know how to prove in high school. But our method of proof was different for different types of rational exponents. This strongly suggests that there must be some underlying principle, undiscovered as yet, that would give a unified derivation of (6.80) for all rational exponents.

The proof we have given of Proposition 6.118(ii) is exactly this unified derivation; moreover, it works equally well for all real exponents, whether rational or irrational.

Since the rationals are dense in the reals, we could reasonably conjecture now that (6.80) holds for all *real* exponents. It would have been absurdly bold to make such a conjecture based only on knowing that (6.80) holds for nonnegative integer exponents. ▲

Remark 6.121 (Irrational exponents, part 2) Suppose that, instead of using Definition 6.111 to define irrational powers of positive numbers, we have defined them an "elementary" way, using a limit-procedure such as those discussed in Remark 6.116 (assuming we have already defined rational powers by elementary means, the way we would in high school or in MAA 4211, rather than by Definition 6.111). Assume that we have done this in the best possible way, showing that for any $a > 0$ and any rational sequence $(r_n)_{n=1}^{\infty}$ converging to r, (i) $\lim_{n\to\infty} a^{r_n}$ exists and (ii) the value of this limit is independent of the choice of sequence (r_n). Finally, suppose we have shown that the derivative-formula (6.80) holds for rational exponents, just based on these elementary, intuitive definitions. To then show that (6.80) is true (with this value of r) we must do something like the following:

1. Choose a sequence of rational numbers $(r_n)_{n=1}^{\infty}$ converging to r.

2. For fixed $x > 0$, write down the following computation, which we will do in "shoot first and ask questions later" form:

$$\begin{aligned}
\lim_{h\to 0}\frac{(x+h)^r - x^r}{h} &= \lim_{h\to 0}\left(\lim_{n\to\infty}\frac{(x+h)^{r_n} - x^{r_n}}{h}\right) && (6.84)\\
&\stackrel{?}{=} \lim_{n\to\infty}\left(\lim_{h\to 0}\frac{(x+h)^{r_n} - x^{r_n}}{h}\right) && (6.85)\\
&= \lim_{n\to\infty}\left(\frac{d}{dx}x^{r_n}\right)\\
&= \lim_{n\to\infty} r_n x^{r_n-1}\\
&\stackrel{?}{=} rx^{r-1}\ . && (6.86)
\end{aligned}$$

If we can justify equalities (6.85) and (6.86), we will have shown that the r^{th}-power function is differentiable on $(0,\infty)$, and that its derivative is the function $x \mapsto rx^{r-1}$.

Justifying (6.86) is no problem: $(r_n - 1)_{n=1}^{\infty}$ is a rational sequence converging to $r-1$, so, by what we're assuming we've already proven about our elementary definition of irrational powers, x^{r_n-1} converges to x^{r-1}. Since (r_n) converges to r, one of our basic results about real-valued sequences shows that the sequence $(r_n x^{r_n-1})$ converges to rx^{r-1}. We can do all this without ever setting foot in Advanced Calculus; the fact that "the limit of a product is the product of the limits" (for convergent real-valued sequences) can easily be proven even in the most elementary treatment of sequences, such as in Calculus 2.

But justifying the interchange-of-limits equation (6.85) is another matter entirely. The entire notion of "interchange of limits" is far beyond the level of Calculus 1. To this writer's knowledge, justifying (6.85) is impossible at the level of Calculus 1, and would even be difficult in Advanced Calculus without relying on the fact that the elementary definition of a^r for *rational* r gives the same value that Definition 6.111 gives.

The (once standard) approach to defining exponentiation presented earlier in this section is a triumph of calculus, a true gem.[18] It unifies the definitions of a^r for positive integer, negative integer, non-integer rational, and irrational r; Definition 6.111 is the same for all exponents. It leads easily to the derivative formula (6.80) for *all* real exponents. By showing that Definition 6.111 agrees with the "elementary" definition for rational exponents, we see why our elementary derivations of $\frac{d}{dx}x^r$ for rational r (the optional reading in Remark 6.120) *had* to keep giving the same formula for all exponents. With Definition 6.111, our proven continuity of the function exp *guarantees* that for any real $a > 0$, any real number r, and any rational (or even real) sequence (r_n) converging to r, the sequence (a^{r_n}) converges to a limit that is independent of which sequence (r_n) we choose. And at the core of all this were two major theorems from the theory of integration:

[18]This may be difficult for modern students to appreciate, especially if they've been taught out of an "early transcendentals" calculus textbook. Calculus textbooks that define exponential functions early, through a "filling in the holes in the graph" idea—the same idea as the limit-procedures discussed in Remark 6.116—rather than waiting until the groundwork has been laid for Definition 6.111, often say shortly after deriving (6.80) for non-negative integer exponents that "We will show later" that (6.80) holds for all real r. These textbooks slip under the rug the fact that they do *not* derive (6.80) from their first definition of x^r. Rather, they wait until they have *re*defined x^r using Definition 6.111, after which they derive (6.80) exactly as in the proof of Proposition 6.118. The "derivation" of (6.80) in at least one popular early-transcendentals textbook starts by writing "$x^r = (e^{\ln x})^r = e^{r\ln x}$", and then uses the Chain Rule to compute the derivative. But the reasoning is completely reversed from the correct logic! Starting with "$x^r = (e^{\ln x})^r = e^{r\ln x}$" suggests that $x^r = e^{r\ln x}$ *because* $(x^b)^c = x^{bc}$ for all real b, c. The truth is exactly the opposite: "$x^r = e^{r\ln x}$" is the *definition* of x^r, a definition from which we *derive* the fact that $(x^b)^c = x^{bc}$ (a fact that, in "early transcendentals" textbooks, is often relegated to an appendix).

There is nothing wrong in a student's using the familiar formula $(x^b)^c = x^{bc}$ (familiar for *rational exponents only*) as a *mnemonic device* to help remember that $x^r = e^{r\ln x}$, but the above so-called derivation encourages the student to think, wrongly, that this formula for x^r is a *consequence* of "$(x^b)^c = x^{bc}$".

the integrability of continuous functions (Theorem 6.53) and the Fundamental Theorem of Calculus (in the form of Theorem 6.62).

6.11 Index for notation and terminology

References

[1] T. M. Apostol, *Calculus*, 2nd ed., vol. 1, John Wiley & Sons, 1967. 38, 41

[2] T. M. Apostol, *Mathematical Analysis*, Addison-Wesley, 1957. 38, 41

[3] T. M. Apostol, *Mathematical Analysis*, 2nd ed., Addison-Wesley, 1974. 38, 41

[4] J. Hass, M. D. Weir, and G. B Thomas, *University Calculus*, alternate edition, Pearson Addison-Wesley, 2008. 38

[5] M. Rosenlicht, *Introduction to Analysis*, Scott, Foresman and Co., 1968. Reprinted by Dover, 1986. 2, 38, 41, 45

[6] W. Rudin, *Principles of Mathematical Analysis*, 2nd ed., McGraw-Hill, 1964. 27, 38, 57

[7] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill, 1976. 27, 38, 57

[8] A. Schwartz, *Calculus and Analytic Geometry*, 2nd ed., Holt, Rinehart, and Winston, 1967. 38, 41

[9] J. Stewart, *Calculus: Early Transcendentals*, 6th ed., Thomson Brooks/Cole, 2008. 38

[10] A. E. Taylor and W. R. Mann, *Advanced Calculus*, 2nd ed., Xerox College Publishing, 1972. 26, 27, 38

[11] G. B. Thomas, *Calculus and Analytic Geometry*, alternate edition, Addison-Wesley, 1972. 38, 41, 67

[12] D. G. Zill and W. S. Wright, *Calculus: Early Transcendentals*, 4th ed., Jones and Bartlett, 2011.