THEORY OF THE INTEGRAL

Brian S. Thomson
Simon Fraser University

CLASSICALREALANALYSIS.COM

This text is intended as a treatise for a rigorous course introducing the elements of integration theory on the real line. All of the important features of the Riemann integral, the Lebesgue integral, and the Henstock-Kurzweil integral are covered. The text can be considered a sequel to the four chapters of the more elementary text THE CALCULUS INTEGRAL which can be downloaded from our web site. For advanced readers, however, the text is self-contained.

For further information on this title and others in the series visit our website.

www.classicalrealanalysis.com

There are free PDF files of all of our texts available for download as well as instructions on how to order trade paperback copies. We also allow access to the content of our books on GOOGLE BOOKS and on the AMAZON *Search Inside the Book* feature.

COVER IMAGE: This mosaic of M31 merges 330 individual images taken by the Ultraviolet/Optical Telescope aboard NASA's Swift spacecraft. It is the highest-resolution image of the galaxy ever recorded in the ultraviolet. The image shows a region 200,000 light-years wide and 100,000 light-years high (100 arcminutes by 50 arcminutes). Credit: NASA/Swift/Stefan Immler (GSFC) and Erin Grand (UMCP)

—http://www.nasa.gov/mission_pages/swift/bursts/uv_andromeda.html

Citation: *Theory of the Integral*, Brian S. Thomson, ClassicalRealAnalysis.com (2013), [ISBN 1467924393]

Date PDF file compiled: February 9, 2013

ISBN-13: 9781467924399

ISBN-10: 1467924393

CLASSICALREALANALYSIS.COM

Preface

The text is a self-contained account of integration theory on the real line. The usual curricula in real analysis courses do not allow for much time to be spent on the Henstock-Kurzweil integral. Instead extensive accounts of Riemann's integral and the Lebesgue integral are presented. Accordingly the version here would be mostly recommended for supplementary reading.

Even so it would be a reasonable course design to teach this material prior to a course in abstract measure and integration. The student should end up as well-prepared as in more traditional courses. Certainly every professional mathematician should be aware of more than Lebesgue's theory; while nonabsolutely convergent integrals do not play an extensive role in applications, they are part of our history and of our culture.

The reader might want to view first the prequel to this text:

> B. S. Thomson, *The Calculus Integral*, ClassicalRealAnalysis.com (2008).
> ISBN-13: 978-1442180956, ISBN-10: 1442180951

That text is an (experimental) outline of an elementary real analysis course in which the Newton integral plays the key role. Since the presentation in the present textbook also uses the Newton integral (in its various versions) as a motivating tool, the reader may wish also to consult the prequel to see how this would work at an introductory level.

There are innumerable books published on the Lebesgue integral. The reader needing instruction in that theory is faced with too many choices, although many of them are truly excellent. For the more general integral (called here the general Newton integral or simply "the integral") that is best known classically as the Denjoy-Perron integral and, more recently, as the Henstock-Kurzweil integral, there are far fewer choices and not all of them are excellent. I have resisted for many years writing a lengthy account of this integral, partly because the topic is not widely thought of as being of much significance.

As an further experiment, however, I offer this account of the theory of that integral. The challenge as I see it is to present a coherent narrative leading the readers to a deep understanding of the nature of integration on the real line and, moreover, fully preparing them to study abstract measure and integration. But that is just a goal, not necessarily realized here. Most teachers will, doubtless, remain with the usual sequence of instruction: basic calculus (the Riemann and improper Riemann integrals vaguely presented), elementary analysis (the Riemann integral treated in depth), then abstract measure and integration in graduate school.

My guess is that few graduate students, freshly taught this sequence, could survive an oral examination on the statement

$$\int_a^b F'(x)\,dx = F(b) - F(a),$$

giving all conditions on how this might or might not hold. I hope my readers do better.

BST

Contents

Chapter 1

By way of an introduction

The reader has, no doubt, studied the Riemann integral in elementary classes and been apprised of the fact that that integral is entirely inadequate for modern applications. That is a sufficient starting point to embark on this text.

We presume, however, that most readers of this material are interested in the topic of the Henstock-Kurzweil integral for its own sake. Most readers too are familiar with some of the elements of the Lebesgue theory of integration and will want to see how the present theory connects with that. This chapter will give some background to the narrative. The formal theory starts with the next chapter.

1.1 The classical Newton integral

Integration theory on the real line can be motivated in a number of ways. Many presentations start with the ancient Greeks and their study of the method of exhaustion to calculate areas. That then transitions into discussions of Riemann sums and an account of the Riemann integral. Having started that way, then an introduction to Lebesgue's quite different theory of integration demands some new expository skills.

A simpler story to tell is presented in this chapter, starting with the original concept of Newton. He realized that the problem of areas (along with a multitude of problems in physics) could be expressed as a problem in antidifferentation. Specifically (in more modern language) he introduced the initial value problem

$$\frac{dy}{dx} = f(x), \quad y(x_0) = y_0$$

where f is a given function and (x_0, y_0) is a given point. Finding any antiderivative of f solves this problem by finding a curve through the point (x_0, y_0) with tangent slope at each point (x, y) on the curve given by $f(x)$.

Newton's integral is a formal solution of this problem, captured by the following definition. The idea is Newton's although the notation and terminology have different sources (notably Leibnitz and Fourier).

Rather curiously, this definition would be instantly recognizable to mathematicians of earlier centuries as stating exactly what they intended by an integral. Many modern mathematicians, however, might find it odd since they have been schooled in the Riemann or Lebesgue theories where such a statement (with restrictions) is a theorem, not a definition.

Definition 1.1 (Classical Newton integral) *Let $f : [a,b] \to \mathbb{R}$ where $[a,b]$ is a compact interval. Then f is said to be Newton integrable on $[a,b]$ if there is a function $F : [a,b] \to \mathbb{R}$ such that*[a] *$F'(x) = f(x)$ for each x in $[a,b]$. We write*

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

[a]More properly the requirement that $F'(a) = f(a)$ and that $F'(b) = f(b)$ would be understood in terms of left-hand and right-hand derivatives respectively.

Usually such a function F is called an *indefinite integral* or a *primitive* for f. Since there are many possible such primitive functions F for a given f, the definition requires checking that $F(b) - F(a)$ does not depend on which one is chosen. The mean-value theorem of the calculus is the only tool needed for this.

A descriptive definition The definition is purely *descriptive*. It offers no method for finding or constructing an integral—if one happens to know or to find an antiderivative of f one can write the value of the integral. One cannot look at f presumably and determine in every case whether it is or is not integrable. Nor, unless we have found a primitive, does the definition give us any other method for computing the value of the integral.

To the novice this may appear strange. All integrals in a typical calculus class are computed by exactly this method. Why is this not constructive? The reason is that for the full class of all derivatives (not just the continuous ones) there is no constructive procedure for determining a primitive.

The initial value problem "solved" The initial value problem

$$\frac{dy}{dx} = f(x), \quad y(x_0) = y_0$$

is clearly solved by the expression

$$y = y_0 + \int_{x_0}^{x} f(t)\,dt$$

but only in a trivial sense. It just expresses the same problem in different language but with no new insights. In particular, even though we might write down this formula as if it were a solution, we may not have any idea about whether the integral does in fact exist.

Deficiencies Integration theory in the eighteenth century would have been mostly understood in this sense. The inadequacies of the theory were met partly by Cauchy in the 1820s (making the theory more constructive) and by later authors who extended the integral to include larger classes of "integrable" functions. It was not until the early 20th century that a suitable theory of integration emerged that could be used in all applications.

Even so, the Newton integral is an excellent starting point for thinking about the problem of integration and communicating all of the essential ideas to students of the theory. It has one very strong advantage: nearly all calculus students think that this definition (that they actually first encountered as a theorem) describes precisely what they remember of the calculus integral.

1.1.1 Bounded integrable functions and Lipschitz functions

What functions can appear as indefinite integrals? This is an important question in integration theory, although it is entirely transparent for the classical Newton integral. One can write

$$F(x) = \int_a^x f(t)\,dt \qquad (a \le x \le b)$$

in the sense of this integral if and only if F is everywhere differentiable.

If we narrow the question a bit to insist that the function f appearing should be bounded on $[a,b]$ then we must consider the following important and well-known class of functions.

Definition 1.2 (Lipschitz functions) *A function $F : [a,b] \to \mathbb{R}$ is said to be a Lipschitz function if there is a nonnegative number M so that*

$$|F(y) - F(x)| \le M|y - x|$$

for all x, $y \in [a,b]$.

Lemma 1.3 *If the function $f : [a,b] \to \mathbb{R}$ is bounded and integrable, then its indefinite integral must be a Lipschitz function*

Proof. Since we are proving this lemma only for functions integrable in the classical Newton sense, this is an elementary application of the mean-value theorem. Take F as the indefinite integral, assume that $|f(x)| \le M$ for all $x \in [a,b]$ and use the mean-value theorem to select a point ξ between x and y so that

$$\left|\frac{F(y) - F(x)}{y - x}\right| = |f'(\xi)| \le M.$$

For all of our methods of integration in this text the same lemma is true, although a different proof would be needed. All of our more general integrals would allow the following simple argument. Take $a \le x < y \le b$. Integrate on

$[x,y]$ using monotone properties of integrals to obtain

$$-M(y-x) \leq \int_x^y f(t)\,dt \leq M(y-x)$$

and substitute

$$F(y)-F(x) = \int_x^y f(t)\,dt.$$

■

Because Lipschitz functions arise naturally in any investigation of bounded integrable functions we should be alert to the properties of this class of functions and expect them to play a role in integration theory.

1.1.2 Absolutely integrable functions and bounded variation

Continuing the theme of the previous section, we can ask what functions can appear as indefinite integrals of absolutely integrable functions? A function $f : [a,b] \to \mathbb{R}$ is said to be *absolutely integrable* if it is integrable and the function $|f|$ is also integrable. It said to be *nonabsolutely integrable* if it is integrable and the function $|f|$ fails to be integrable.

The following computational lemma quickly illustrates the key property that the indefinite integral must possess. We shall prove the property for the classical Newton integral but this same property is true for all of our integrals.

Lemma 1.4 *If the function $f : [a,b] \to \mathbb{R}$ is absolutely integrable, then its indefinite integral F must have this property: for every subdivision*

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

of the interval $[a,b]$

$$\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| \leq \int_a^b |f(t)\,dt < \infty.$$

Proof. Let G be the indefinite integral of $|f|$ on $[a,b]$. The proof is a simple application of this easy inequality

$$|F(x_i) - F(x_{i-1})| = \left| \int_{x_i}^{x_{i-1}} f(t)\,dt \right| \leq \int_{x_i}^{x_{i-1}} |f(t)|\,dt = G(x_i) - G(x_{i-1}).$$

Then one sums to obtain

$$\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})|$$

$$\leq \sum_{i=1}^{n} [G(x_i) - G(x_{i-1})] = G(b) - G(a) = \int_a^b |f(t)\,dt < \infty.$$

These same steps would prove the lemma for any of our later integration methods; it is not special to the classical Newton integral ■

Functions of bounded variation Any function F that possesses the property stated in the lemma is said to have bounded variation on $[a,b]$. In fact we define $\mathrm{Var}(F,[a,b])$ to be the supremum of all sums of the form

$$\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})|$$

taken for arbitrary subdivisions

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

of the interval $[a,b]$. The value $\mathrm{Var}(F,[a,b])$ is called the *total variation* of F.

Because functions of bounded variation arise naturally in any investigation of absolutely integrable functions we should be alert to the properties of this important class of functions and expect them to play a role in integration theory.

Exercise 1 *Show that a function $F : [a,b] \to \mathbb{R}$ that is Lipschitz must have bounded variation on $[a,b]$.* Answer □

1.1.3 The Newton integral is a nonabsolute integral

There do exist functions that are nonabsolutely integrable. The student who has seen only the Riemann integral and its more important cousin, the Lebesgue integral, would not have encountered such functions. We give the details here.

The following interesting example shows that it is in the nature of all Newton-type integrals to be non-absolute. Define the function $F(x) = x^2 \sin x^{-2}$ on $[0,1]$, interpreting $F(0) = 0$ so that F is continuous. In fact F is differentiable at every point. A direct computation using limits shows that $F'(0) = 0$ while, for $0 < x \leq 1$, standard calculus techniques (product rule, chain rule) supply

$$F'(x) = 2x \sin x^{-2} - \frac{2}{x} \cos x^{-2}.$$

Consequently F' is integrable on $[0,1]$ in the classical Newton sense.

We show that, while F' is integrable, $|F'|$ is not. If it were integrable then, by Lemma 1.4, F would have bounded variation on $[0,1]$. It does not.

For any positive integer k, take points

$$y_k = \frac{1}{\sqrt{k\pi}} \quad \text{and} \quad x_k = \frac{\sqrt{2}}{\sqrt{[2k+1]\pi}}.$$

Observe that $F(y_k) = 0$ while

$$F(x_k) = \pm \frac{2}{[2k+1]\pi}.$$

Consequently, for any large integer N one has

$$\sum_{k=1}^{N} |F(y_k) - F(x_k)| \geq \sum_{k=1}^{N} \frac{2}{[2k+1]\pi}.$$

The value of the sum on the right-hand side grows without bound as N increases. This allows us to check that $\mathrm{Var}(F,[0,1])=\infty$ and so F does not have bounded variation and $|F'|$ cannot be integrable by this (or indeed any) method.

1.2 Continuity and integrability

What functions are integrable in the Newton sense? We can answer this with a host of examples, since every calculus student learns a variety of methods of computing antiderivatives (albeit for a limited set of problems). But the best elementary theory and the best starting point for understanding the nature of the problem is to focus on continuous functions.

We shall show now that, with appropriate continuity assumptions on f, we can be assured that an integral exists without any requirement that we should find it. Our methods will show that we can also describe a procedure that would, in theory, produce the indefinite integral as the limit of a sequence of simpler functions.

We will still have a theory for Newton integrals of discontinuous functions but we will have to be content with the fact that much of the theory is formal, and describes objects which are not necessarily constructible[1].

Historical Note Lebesgue [50], in a short pedagogical note from 1905, gives a direct proof (similar to ours below) that continuous functions have primitives. He points out the importance of proving such a theorem to beginning students of integration theory. Then, as now, most students would first encounter this only within the context of a study of the Riemann integral. One can argue that there are benefits to the student who learns this simple theorem independently and prior to exposure to more elaborate integration theories.[2]

> "Throughout the nineteenth century existence questions played a large role in mathematics. They were solved for the Cauchy and Riemann integrals, regarded as limits of sums. However in reading the works of the mathematicians of the last century one gains the impression that the overwhelming majority of them continued to use the definition of a definite integral due to Newton and Leibniz, in which the definite integral is regarded as the difference of two values of a primitive function. In this connection the question of the existence of a primitive was either not considered at all or was answered by starting from the notion of an integral as a limit of sums, which was logically circular. It was only in 1905 that Lebesgue succeeded in proving the existence of a primitive for every continuous function without resort to the definition of an integral as a limit of sums;

[1] **Note to the instructor:** Just how unconstructible are indefinite integrals in general? See Chris Freiling [30], *How to compute antiderivatives,* Bull. Symbolic Logic 1 (1995), no. 3, 279–316. This is by no means an elementary question.

[2] I am indebted to Russell Gordon for supplying me with these references.

and only then was it justified to regard the integral of a continuous function as the difference of two values of a primitive.” ——Medvedev [61, p. 66]

1.2.1 Upper functions

We will illustrate our method by introducing the notion of an *upper function*. This is a piecewise linear function whose slopes dominate a given function.

Let f be defined at all points of a compact interval $[a,b]$ and suppose that f is bounded. Choose any points

$$a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n = b.$$

Suppose that F is a continuous function on $[a,b]$ that is linear on each interval $[x_{i-1}, x_i]$ and such that

$$\frac{F(x_i) - F(x_{i-1})}{x_i - x_{i-1}} \geq f(\xi)$$

for all points ξ for which $x_{i-1} \leq \xi \leq x_i$ $(i = 1, 2, \ldots, n)$. Then we can call F an *upper function* for f on $[a,b]$.

The method of upper functions is to approximate the indefinite integral that we require by suitable upper functions. Upper functions are piecewise linear functions with the break points (where the corners are) at x_1, x_2, ..., x_{n-1}. The slopes of these line segments exceed the values of the function f in the corresponding intervals. See Figure 1.1 for an illustration of such a function.

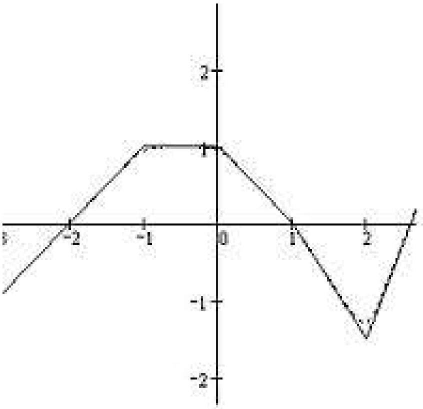

Figure 1.1: A piecewise linear function on $[-3,3]$.

Exercise 2 *Let $f(x) = x^2$ be defined on the interval $[0,1]$. Define an upper function for f using the points $0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$. Sketch the graph of that upper function.*

Answer □

Exercise 3 (step functions) *Let a function f be defined by requiring that, for any integer n (positive, negative, or zero), $f(x) = n$ if $n-1 \leq x < n$. This is a simple example of a step function. Find a formula for an indefinite integral and show that this is an upper function for f.* Answer □

1.2.2 Continuous functions are Newton integrable

For continuous functions we can always determine the existence of an indefinite integral by a limiting process using appropriate upper functions. The lemma is a technical computation that justifies this statement.

Lemma 1.5 *Suppose that $f : [a,b] \to \mathbb{R}$ is a bounded function. Then there exists a Lipschitz function $F : [a,b] \to \mathbb{R}$ so that $F'(x) = f(x)$ for every point $a \leq x \leq b$ at which f is continuous.*

From this lemma there immediately follows our first existence theorem for the Newton integral. The proof of the lemma even contains a method for constructing the indefinite integral as a limit of a sequence of upper functions.

Theorem 1.6 *Suppose that $f : [a,b] \to \mathbb{R}$ is continuous. Then f is Newton integrable on all closed subintervals of $[a,b]$.*

1.2.3 Proof of Lemma 1.5

We use the method of upper functions to prove Lemma 1.5. It will be enough to assume that $f : [0,1] \to \mathbb{R}$ and that f is nonnegative and bounded. The general case is easily argued from this special situation.

Let F_0 denote the function on $[0,1]$ that has $F_0(0) = 0$ and has constant slope equal to

$$c_{01} = \sup\{f(t) : 0 \leq t \leq 1\}.$$

Subdivide $[0,1]$ into $[0,\frac{1}{2}]$ and $[\frac{1}{2},1]$ and let F_1 denote the continuous, piecewise linear function on $[0,1]$ that has $F_0(0) = 0$ and has constant slope equal to

$$c_{11} = \sup\{f(t) : 0 \leq t \leq \frac{1}{2}\}$$

on $[0,\frac{1}{2}]$ and constant slope equal to

$$c_{12} = \sup\{f(t) : \frac{1}{2} \leq t \leq 1\}$$

on $[0,\frac{1}{2}]$. This construction is continued. For example, at the next stage, Subdivide $[0,1]$ further into $[0,\frac{1}{4}]$, $[\frac{1}{4},\frac{1}{2}]$, $[\frac{1}{2},\frac{3}{4}]$, and $[\frac{3}{4},1]$. Let F_2 denote the continuous, piecewise linear function on $[0,1]$ that has $F_0(0) = 0$ and has constant slope equal to

$$c_{11} = \sup\{f(t) : 0 \leq t \leq \frac{1}{4}\}$$

on $[0, \frac{1}{4}]$, constant slope equal to

$$c_{12} = \sup\{f(t) : \frac{1}{4} \le t \le \frac{1}{2}\}$$

on $[\frac{1}{4}, \frac{1}{2}]$, constant slope equal to

$$c_{13} = \sup\{f(t) : \frac{1}{2} \le t \le \frac{3}{4}\}$$

on $[\frac{1}{2}, \frac{3}{4}]$, and constant slope equal to

$$c_{14} = \sup\{f(t) : \frac{3}{4} \le t \le 1\}$$

on $[\frac{3}{4}, 1]$.

In this way we construct a sequence of such functions $\{F_n\}$. Note that each F_n is continuous and nondecreasing. Moreover a look at the geometry reveals that

$$F_n(x) \ge F_{n+1}(x)$$

for all $0 \le x \le 1$ and all $n = 0, 1, 2, \ldots$. In particular $\{F_n(x)\}$ is a nonincreasing sequence of nonnegative numbers and consequently

$$F(x) = \lim_{n\to\infty} F_n(x)$$

exists for all $0 \le x \le 1$. We prove that $F'(x) = f(x)$ at all points x in $[0, 1]$ at which the function f is continuous.

Fix a point x in $[0, 1]$ at which f is assumed to be continuous and let $\varepsilon > 0$. Our argument addresses the case that $0 < x < 1$; if $x = 0$ or $x = 1$ a one-sided argument would have to be introduced, but otherwise the details do not much differ.

Choose $\delta > 0$ so that the oscillation

$$\omega f([x - 2\delta, x + 2\delta])$$

of f on the interval $[x - 2\delta, x + 2\delta]$ does not exceed ε. Let h be fixed so that $0 < h < \delta$. Choose an integer N sufficiently large that

$$|F_N(x) - F(x)| < \varepsilon h \text{ and } |F_N(x+h) - F(x+h)| < \varepsilon h.$$

From the geometry of our construction notice that the inequality

$$|F_N(x+h) - F_N(x) - f(x)h| \le h\omega f([x - 2h, x + 2h]),$$

must hold for large enough N. (Simply observe that the graph of F_N will be composed of line segments, each of whose slopes differ from $f(x)$ by no more than the number $\omega f([x - 2h, x + 2h])$.)

Putting these inequalities together we find that

$$|F(x+h) - F(x) - f(x)h| \le$$

$$|F_N(x+h) - F_N(x) - f(x)h| + |F_N(x) - F(x)| + |F_N(x+h) - F(x+h)| < 3\varepsilon h.$$

This shows that the right-hand derivative of F at x must be exactly $f(x)$. A similar argument will handle the left-hand derivative and we have verified that F is an indefinite integral (or primitive) for f on the interval $[a,b]$.

We should now check that the function F defined here is Lipschitz on $[0,1]$. Let M be an upper bound for the function f. Check, first, that

$$0 \leq F_n(y) - F_n(x) \leq M(y-x)$$

for all $x < y$ in $[0,1]$. From this we can deduce that F is in fact Lipschitz on $[0,1]$.

This argument (thus far) establishes the theorem for the case $[a,b] = [0,1]$ and with f nonnegative. To complete the proof the reader can follow the next steps as outlined here:

1. Suppose that $f : [a,b] \to \mathbb{R}$ and set $g(t) = f(a+t(b-a))$ for all $0 \leq t \leq 1$. If G is an indefinite integral for g on $[0,1]$ show how to find an indefinite integral for f on (a,b). [Hint: If $H(t) = G(a+t(b-a))$ then, by the chain rule,
$$H'(t) = G'(a+t(b-a)) \times (b-a) = f(a+t(b-a)) \times (b-a).$$
Substitute $x = a+t(b-a)$ for each $0 \leq t \leq 1$.]

2. Suppose that $f : [a,b] \to \mathbb{R}$ is a bounded function and that
$$K = \inf\{f(x) : a \leq x \leq b\}.$$
Set $g(t) = f(t) - K$ for all $a < t < b$. Show that g is nonnegative and bounded. Suppose that G is an indefinite integral for g on (a,b); show how to find an indefinite integral for f on (a,b). [Hint: If $G'(t) = g(t)$ then $\frac{d}{dt}(G(t)+Kt) = g(t)+K = f(t)$.]

3. Show how to deduce Theorem 1.6 from the lemma.

1.3 Riemann sums

The expression of a Newton integral by its definition

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

requires finding a function F to serve as an antiderivative. It would be more convenient, both for theory and practice, if we can relate the value of the integral directly to the actual values of the function f.

Approximations of the form

$$\int_a^b f(x)\,dx \approx \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$

have long been used. While Cauchy was the first to use them in a manner that essentially defines an integral, earlier mathematicians considered such sums as approximations to an integral that they would have conceived of as a Newton integral. The early numerical approximations are just refinements of Riemann sum approximations.

Here the points x_i are chosen so as to begin at the left endpoint a and end at the right endpoint b,

$$a = x_0, x_1, x_2, x_3, \ldots, x_{n-1}, x_n = b$$

and the points ξ_i (called the *associated points*) are required to be chosen at or between the corresponding points x_{i-1} and x_i. Most readers would have encountered such sums under the stricter conditions that

$$a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n = b \text{ and } x_{i-1} \leq \xi_i \leq x_i$$

so that the points are arranged in increasing order. This need not always be the case, but it is most frequently so.

Riemann sums and integration theory These sums

$$\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$

will be called *Riemann sums* even though their use predates Riemann's birth by many years. The connection with integration theory also does not originate with Riemann[3] nor are they that late in the history of the subject. Poisson in 1820 proposed such an investigation as "the fundamental proposition of the theory of definite integrals." Euler, by at least 1768, had already used such sums to approximate integrals. Of course, for both Poisson and Euler the integral was understood in our sense as an antiderivative[4].

Thus we use the following language to describe these sums.

Definition 1.7 (Riemann sum) *Suppose that $f : \mathbb{R} \to \mathbb{R}$ and that a collection of points is given*

$$a = x_0, x_1, x_2, x_3, \ldots, x_{n-1}, x_n = b$$

and along with associated points ξ_i at or between x_{i-1} and x_i for $i = 1, 2, \ldots, n$. Then any sum of the form

$$\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$

is called a Riemann sum *for the function f on the interval $[a, b]$.*

Partitions and subpartitions When the collection of points

$$a = x_0, x_1, x_2, x_3, \ldots, x_{n-1}, x_n = b$$

[3] Georg Friedrich Bernhard Riemann (1826–1866). His lecture notes on integration theory date from the 1850s.

[4] See Judith V. Grabiner [36], *Who gave you the epsilon? Cauchy and the origins of rigorous calculus*, American Mathematical Monthly 90 (3), 1983, 185–194.

is chosen to be increasing, i.e., so that

$$a = x_0 < x_1 < x_2 < x_3 < \ldots, x_{n-1} < x_n = b,$$

then we might prefer to write the partition as a interval-point relation:

$$\{([x_{i-1}, x_i], \xi_i) : i = 1, 2, 3 \ldots n\}$$

or perhaps relabeled as

$$\{([a_i, b_i], \xi_i) : i = 1, 2, 3 \ldots n\}.$$

We call such a collection a *partition* of the interval $[a, b]$. Any subset of a partition is called a *subpartition*. The only requirement on a collection

$$\{([a_i, b_i], \xi_i) : i = 1, 2, 3 \ldots m\}$$

to require it to be a subpartition is that the intervals $[a_i, b_i]$ do not overlap and the associated points ξ_i belong to the appropriate interval. For most of the theory of Riemann sums one uses sums over partitions and subpartitions rather than general versions where the points are chosen not to be increasing.

Partitions and subpartitions finer than δ Let δ be a positive number. Then a partition or a subpartition

$$\{([a_i, b_i], \xi_i) : i = 1, 2, 3 \ldots m\}$$

is said to be *finer than* δ if each

$$b_i - a_i < \delta.$$

More generally, if δ is a positive function we use the same phrase to describe the requirement that each

$$b_i - a_i < \delta(\xi_i).$$

In nearly all applications of Riemann sums some such "finer" requirement will appear. Our first observation, however, applies to partitions for which no such "finer" notion is needed.

1.3.1 Mean-value theorem and Riemann sums

The mean-value theorem allows an interpretation in terms of Riemann sums that is a convenient starting point for the theory. If $F : \mathbb{R} \to \mathbb{R}$ is a function that is differentiable at every point of the interval $[a, b]$ then we know that $f = F'$ is Newton integrable and that the mean-value theorem of the calculus can be applied to express the integral in the form

$$\int_a^b f(x)\,dx = F(b) - F(a) = f(\xi)(b - a)$$

for some $\xi \in (a, b)$. This expresses the integral exactly as a very simple kind of Riemann sum with just one term. Here $x_0 = a$ and $x_1 = b$.

Take now the three distinct points

$$a = x_0, x_1, x_2 = b$$

and do the same thing twice. Then

$$\int_a^b f(x)\,dx = F(b) - F(a) = [F(b) - F(x_1)] + [F(x_1) - F(a)]$$

$$= f(\xi_2)(b - x_1) + f(\xi_1)(x_1 - a) = \sum_{i=1}^{2} f(\xi_i)(x_i - x_{i-1})$$

for some points ξ_1 between a and x_1 and ξ_2 between x_1 and b. Again, this expresses the integral exactly as a simple kind of Riemann sum with just two terms.

Exact expression of the integral as a Riemann sum In fact then we can do this for any number of points. Take any collection

$$a = x_0, x_1, x_2, x_3, \ldots, x_{n-1}, x_n = b$$

arranged in any order (not necessarily increasing) and choose the associated points ξ_i between x_{i-1} and x_i for $i = 1, 2, \ldots, n$ in such a way that

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n} [F(x_i) - F(x_{i-1})]$$

$$= \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}).$$

Using our language, we have just proved in this identity that a Newton integral in all situations can be computed *exactly* by some appropriately chosen Riemann sum.

Wonderful or . . . perhaps not This seems both wonderful and, maybe, not so wonderful. In the first place it means that an integral

$$\int_a^b f(x)\,dx$$

can be computed by a simple sum using the values of the function f rather than by using the definition and having, instead, to solve a difficult or impossible indefinite integration problem. On the other hand this only works if we can select the right associated points $\{\xi_i\}$ that make this precise. In theory the mean-value theorem supplies the points, but in practice we would be most often unable to select the correct points.

Riemann sums and area Every Riemann sum

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}). \tag{1.1}$$

that expresses an integral of a nonnegative function can be reinterpreted as an area. Consider that the product

$$f(\xi_i) \times (x_i - x_{i-1})$$

represents the area of a rectangle with height $f(\xi_i)$ and base $(x_i - x_{i-1})$. Then both the integral and the Riemann sum are equal to the total area of a figure composed of n rectangles. Figure 1.2 illustrates.

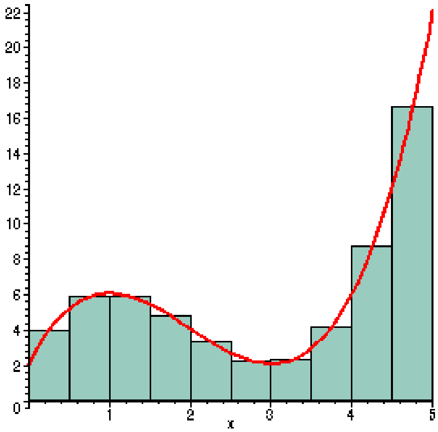

Figure 1.2: A Riemann sum considered as an area.

This then permits us to use the language of areas in all discussions of integrals. This may or may not be of assistance in visualizing the problem being addressed.

Exercise 4 *Show that the integral $\int_a^b x\,dx$ can be computed exactly by any Riemann sum*

$$\int_a^b x\,dx = \sum_{i=1}^{n} \frac{x_i + x_{i-1}}{2}(x_i - x_{i-1}) = \frac{1}{2}\sum_{i=1}^{n}(x_i^2 - x_{i-1}^2).$$

Answer □

Exercise 5 *Subdivide the interval $[0,1]$ at the points $x_0 = 0$, $x_1 = 1/3$, $x_2 = 2/3$ and $x_3 = 1$. Determine the points ξ_i so that*

$$\int_0^1 x^2\,dx = \sum_{i=1}^{3} \xi_i^2 (x_i - x_{i-1}).$$

□

Exercise 6 *Subdivide the interval $[0,1]$ at the points $x_0 = 0$, $x_1 = 1/3$, $x_2 = 2/3$*

and $x_3 = 1$. Determine the points $\xi_i \in [x_{i-1}, x_i]$ so that

$$\sum_{i=1}^{3} \xi_i^2 (x_i - x_{i-1}).$$

is as large as possible. By how much does this sum exceed $\int_0^1 x^2\,dx$? □

Exercise 7 *Subdivide the interval $[0,1]$ at the points $x_0 = 0$, $x_1 = 1/3$, $x_2 = 2/3$ and $x_3 = 1$. Consider various choices of the points $\xi_i \in [x_{i-1}, x_i]$ in the sum*

$$\sum_{i=1}^{3} \xi_i^2 (x_i - x_{i-1}).$$

What are all the possible values of this sum? What is the relation between this set of values and the number $\int_0^1 x^2\,dx$? □

Exercise 8 *Subdivide the interval $[0,1]$ by defining the points $x_0 = 0$, $x_1 = 1/n$, $x_2 = 2/n$, $\dots x_{n-1} = (n-1)/n$, and $x_n = n/n = 1$. Determine the points $\xi_i \in [x_{i-1}, x_i]$ so that*

$$\sum_{i=1}^{n} \xi_i^2 (x_i - x_{i-1}).$$

is as large as possible. By how much does this sum exceed $\int_0^1 x^2\,dx$? □

Exercise 9 *Let $0 < r < 1$. Subdivide the interval $[0,1]$ by defining the points $x_0 = 0$, $x_1 = r^{n-1}$, $x_2 = r^{n-2}$, $\dots$, $x_{n-1} = r^{n-(n-1)} = r$, and $x_n = r^{n-(-n)} = 1$. Determine the points $\xi_i \in [x_{i-1}, x_i]$ so that*

$$\sum_{i=1}^{n} \xi_i^2 (x_i - x_{i-1}).$$

is as large as possible. By how much does this sum exceed $\int_0^1 x^2\,dx$? □

Exercise 10 (error estimate) *Let $f : [a,b] \to \mathbb{R}$ be a Newton integrable function with F as an indefinite integral. Suppose that*

$$\{([x_i, x_{i-1}], \xi_i) : i = 1, 2, \dots n\}$$

is an arbitrary partition of $[a,b]$. Show that

$$\left| \int_{x_{i-1}}^{x_i} f(x)\,dx - f(\xi_i)(x_i - x_{i-1}) \right| \le \omega f([x_i, x_{i-1}])(x_i - x_{i-1})$$

for $i = 1, 2, 3, \dots, n$ and that

$$\left| \int_a^b f(x)\,dx - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| \le \sum_{i=1}^{n} \omega f([x_i, x_{i-1}])(x_i - x_{i-1}). \tag{1.2}$$

Note: that, if the right hand side of the inequality (1.2) is small then the Riemann sum, while not precisely equal to the integral, would be a good estimate. Of course, the right hand side might also be big. Answer □

1.3.2 Uniform Approximation by Riemann sums

We have seen that Newton integrals can be *exactly* computed by Riemann sums. Since we must appeal to the mean-value theorem, this gives no procedure for determining the correct associated points that make the computation exact.

Suppose we relax our goal. Instead of asking for an exact computation, perhaps an approximate computation might be useful:

$$\int_a^b f(x)\,dx \approx \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}) \ ?$$

Here we wish to allow an arbitrary choice of associated points. Thus we will certainly introduce an error, depending on how far $f(\xi_i)$ is from the "correct" choice of associated point. To control the error we need to make the points x_i and x_{i-1} close together. By a *uniform* approximation we mean that we shall specify the smallness by a single small number δ and require that the points be chosen so that, for each $i = 1, 2, 3, \dots, n$,

$$|x_i - x_{i-1}| < \delta.$$

[Later on we will relax this requirement and investigate a pointwise approximation, by requiring that the points be chosen instead so that

$$|x_i - x_{i-1}| < \delta(\xi_i)$$

using a different measure of smallness at each associated point.]

Furthermore, since each term in the sum can add in a small error, we need also to restrict the choice of sequence

$$a = x_0, x_1, x_2, \dots, x_{n-1}, x_n = b$$

so that the total error introduced is not too large. One way to accomplish this is to require that the points are chosen in the natural order:

$$a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n = b.$$

A different way is to limit the size of the variation of the sequence of points by restricting the size of the sum

$$\sum_{i=1}^n |x_i - x_{i-1}|\,.$$

We do the former for Cauchy's theorem and the latter for Robbins's theorem.

1.3.3 Cauchy's theorem

The earliest theorem of this type is due to Cauchy from about 1820. Eighteenth century authors also would certainly have known and recognized this result. It is only attributable to Cauchy because he was the first to articulate what the notion of continuity should mean.

We present it separately here as having largely historical interest. Theorem 1.9 which follows (Robbins's theorem) is the correct technical version for integrability of continuous functions expressed by Riemann sums.

Theorem 1.8 (Cauchy) *Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then, f is Newton integrable on $[a,b]$ and moreover the integral may be uniformly approximated by Riemann sums: for every $\varepsilon > 0$ there is a $\delta > 0$ so that*

$$\sum_{i=1}^{n} \left| \int_{x_{i-1}}^{x_i} f(x)\,dx - f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon \tag{1.3}$$

and

$$\left| \int_a^b f(x)\,dx - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon \tag{1.4}$$

whenever points are given

$$a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n = b$$

for which

$$x_i - x_{i-1} < \delta$$

with associated points $\xi_i \in [x_{i-1}, x_i]$.

Exercise 11 *Prove Theorem 1.8 by using the error estimate in Exercise 10.*

Answer □

Exercise 12 *Show that the integral*

$$\int_0^1 x^2\,dx = \lim_{n\to\infty} \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + \cdots + n^2}{n^3}.$$

Answer □

Exercise 13 *Show that the integral*

$$\int_0^1 x^2\,dx = \lim_{r\to 1-} \left[(1-r) + r(r-r^2) + r^2(r^2-r^3) + r^3(r^3-r^4) + \dots \right].$$

Answer □

Exercise 14 *Show that the integral $\int_0^1 x^5\,dx$ can be exactly computed by the method of Riemann sums provided one has the formula*

$$1^5 + 2^5 + 3^5 + 4^5 + 5^5 + 6^5 + + \cdots + N^5 = \frac{N^6}{6} + \frac{N^5}{2} + \frac{5N^4}{12} - \frac{N^2}{12}.$$

□

1.3.4 Robbins's theorem

There is another version possible for Cauchy's theorem. For Cauchy's theorem the points of the subdivision $a = x_0,\, x_1,\, \dots,\, x_n = b$ were arranged in increasing order. In Robbins's theorem[5] we drop the insistence that the points in the Riemann sum must form an increasing sequence. This allows us to characterize the Newton integral of continuous functions entirely by a statement using Riemann sums.

In place of the familiar requirement in the formulation of Riemann sums that the terms $x_0, x_1, \dots, x_n$ would form an increasing sequence we relax this to place only an upper limit on the variation of the sequence. The reader will recall that the variation of such a sequence would be the sum

$$\sum_{i=1}^{n} |x_i - x_{i-1}|.$$

If the sequence is indeed increasing then certainly

$$\sum_{i=1}^{n} |x_i - x_{i-1}| = x_n - x_0.$$

Theorem 1.9 (Robbins) *A real-valued function f is continuous on an interval $[a,b]$ if and only if it satisfies the following strong uniform integrability criterion: there is a number I so that, for every $\varepsilon > 0$ and $C > 0$, there is a $\delta > 0$ with the property that*

$$\left| I - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon$$

for any choice of points $x_0, x_1, \dots, x_n$ and $\xi_1, \xi_2, \dots, \xi_n$ from $[a,b]$ satisfying

$$\sum_{i=1}^{n} |x_i - x_{i-1}| \leq C$$

where $a = x_0$, $b = x_n$, $0 < |x_i - x_{i-1}| < \delta$ and each ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \dots, n$. In that case, necessarily, f is Newton integrable on $[a,b]$ and

$$I = \int_a^b f(x)\,dx.$$

This theorem gives us some insight into integration theory. Instead of basing the integral on the concept of an antiderivative, it could instead (at least in the case of continuous functions) be obtained directly from a definition of an integral based on the concept of Riemann sums. This gives us two equivalent formulations of the Newton integral of continuous functions: one uses an antiderivative and one uses Riemann sums.

Exercise 15 *Deduce the inequality (1.4) in Cauchy's theorem (Theorem 1.8) from Robbins's theorem.* □

[5]Only one direction in the theorem is due to Robbins and a proof can be found in Herbert E. Robbins, *Note on the Riemann integral*, American Math. Monthly, Vol. 50, No. 10 (Dec., 1943), 617–618.. The other direction is proved in B. S. Thomson, *On Riemann sums*, Real Analysis Exchange 37, no. 1 (2011) 221–242.

1.3.5 Proof of Theorem 1.9

We first prove the easy direction in Robbins's theorem, i.e, we assume that f is continuous and prove that the statement holds with

$$I = \int_a^b f(x)\,dx.$$

Suppose first that f is (uniformly) continuous on $[a,b]$. Then f is integrable on $[a,b]$ in the classical Newton sense. We prove (using a different method than that chosen by Robbins) that f satisfies also this strong integrability condition. Let $\varepsilon > 0$ and $C > 0$ be given. Take δ sufficiently small that

$$|f(x) - f(y)| < \varepsilon/C$$

if x and y are points of $[a,b]$ for which $|x-y| < \delta$. (This uses the uniform continuity of f on the interval $[a,b]$.)

Write $F(x) = \int_a^x f(t)\,dt$. Suppose that $a \le x \le \xi \le y \le b$ and that $0 < y-x < \delta$. Then, by the mean-value theorem, there is a point ξ^* between x and y for which

$$F(y) - F(x) = f(\xi^*)(y-x).$$

Thus we also have

$$|F(y) - F(x) - f(\xi)(y-x)| = |[f(\xi^*) - f(\xi)](y-x)| < \frac{\varepsilon}{C}(y-x).$$

Then, for any choice of points $x_0, x_1, \dots, x_n$ and $\xi_1, \xi_2, \dots, \xi_n$ from $[a,b]$ with the properties in the statement of the theorem,

$$\begin{aligned}
\left|\int_a^b f(x)\,dx - \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1})\right| &= \left|F(b) - F(a) - \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1})\right| \\
&= \left|\sum_{i=1}^n [F(x_i) - F(x_{i-1}) - f(\xi_i)(x_i - x_{i-1})]\right| \\
&\le \sum_{i=1}^n |F(x_i) - F(x_{i-1}) - f(\xi_i)(x_i - x_{i-1})| \\
&< \frac{\varepsilon}{C}\sum_{i=1}^n |x_i - x_{i-1}| \le \varepsilon.
\end{aligned}$$

That completes the proof in this one direction.

In the other direction let us assume that f has the strong integrability property expressed in the statement of the theorem. We shall prove that f is then necessarily continuous on $[a,b]$ with the number I equal to the integral of f on that interval. This is an exercise[6] in elementary integration theory for the experi-

[6] Indeed the experienced reader will recognize that f is necessarily Riemann integrable on $[a,b]$ and so Steps 1–3 can be skipped.

enced reader, and an introduction to some useful methods for the novice.

Step 1 We show that such a function satisfies the following equivalent strong integrability criterion: for every $\varepsilon > 0$ and $C > 0$, there is a $\delta > 0$ with the property that

$$\left|\sum_{j=1}^{m} f(\xi'_j)(x'_j - x'_{j-1}) - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})\right| < \varepsilon$$

for any choice of points

$$x_0, x_1, \ldots, x_n \text{ and } \xi_1, \xi_2, \ldots, \xi_n \text{ and } x'_0, x'_1, \ldots, x'_m \text{ and } \xi'_1, \xi'_2, \ldots, \xi'_m$$

from $[a,b]$ satisfying

$$\sum_{i=1}^{n} |x_i - x_{i-1}| \leq C \text{ and } \sum_{j=1}^{m} |x'_j - x'_{j-1}| \leq C$$

where $a = x_0 = x'_0$, $b = x_n = x'_m$, $0 < |x_i - x_{i-1}| < \delta$, $0 < |x'_j - x'_{j-1}| < \delta$, each ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \ldots, n$, and each ξ'_j belongs to the interval with endpoints x'_j and x'_{j-1} for $j = 1, 2, \ldots, m$,

This is a standard "Cauchy criterion" version of the integrability condition. Such a statement is equivalent to the other version. It is an essential element of general integration theory to prove the equivalence of such statements. (We leave the details to the reader as it is an exercise useful to understanding the nature of integration theory. This is similar to proving that a sequence is convergent if and only if it is a Cauchy sequence. If you review how that proof is done you will find that much of that method works here.)

Step 2 We note now that, if a function satisfies the strong integrability property of the theorem on an interval $[a,b]$ then it satisfies this same strong uniform integrability criterion on every subinterval $[c,d] \subset [a,b]$: there is a number $I(c,d)$ so that, for every $\varepsilon > 0$ and $C > 0$, there is a $\delta > 0$ with the property that

$$\left|I(c,d) - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})\right| < \varepsilon$$

for any choice of points $x_0, x_1, \ldots, x_n$ and $\xi_1, \xi_2, \ldots, \xi_n$ from $[c,d]$ satisfying

$$\sum_{i=1}^{n} |x_i - x_{i-1}| \leq C$$

where $c = x_0$, $d = x_n$, $0 < |x_i - x_{i-1}| < \delta$ and each ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \ldots, n$.

The method of proof is to use the equivalent "Cauchy criterion" as encountered in Step 1. Just check that, if f satisfies a Cauchy criterion on $[a,b]$ then it must also satisfy a Cauchy criterion on each subinterval $[c,d]$.

Step 3 The next step is to verify the identity

$$I(x,z) = I(x,y) + I(y,z)$$

for all $a \leq x < y < z \leq b$ where these expressions have been defined in Step 2. Again this should present no conceptual difficulty, although the steps require some attention to detail and work with inequalities.

Step 4 Finally we are ready to prove the crucial step in Robbins's theorem, i.e, we assume that f satisfies the strong "integrability" criterion and prove that f must be continuous and that

$$I = \int_a^b f(x)\,dx.$$

We know (from Steps 1–3) that such a function f with these properties would have to have the same properties on each subinterval and that that there must be a function $F : [a,b] \to \mathbb{R}$ with $I(x,y) = F(y) - F(x)$ for each $a \leq x < y \leq b$.

Suppose, contrary to what we want to prove, that there is a point z of discontinuity of f in the interval. We will assume that $a < z < b$ and derive a contradiction. (The cases $z = a$ and $z = b$ are similarly handled.) Then there must be a positive number $\eta > 0$ so that, if we choose any points $z_1 < z < z_2$, the interval $[z_1, z_2]$ must contain points c_1 and c_2 for which $|f(c_1) - f(c_2)| > \eta$.

Now we apply the strong integrability hypothesis using

$$I = F(b) - F(a),\ \varepsilon = \eta/4, \text{ and } C = b - a + 4$$

to obtain a choice of δ with $0 < \delta < 1$ that meets the conditions stated in the theorem on $[a,b]$. Choose points $z_1 < z < z_2$ so that $z_2 - z_1 < \delta$ and then select points c_1 and c_2 in the interval $[z_1, z_2]$ for which $f(c_1) - f(c_2) > \eta$.

Construct a sequence

$$a = x_0 < x_1 < \cdots < x_p = z_1$$

along with associated points $\{\xi_i\}$ so that $0 < x_i - x_{i-1} < \delta$ and so that

$$\left| F(z_1) - F(a) - \sum_{i=1}^{p} f(\xi_i)(x_i - x_{i-1}) \right| < \eta/4.$$

This just uses the integrability hypotheses of the function f on the interval $[a, z_1]$.

Choose the least integer r so that

$$r(z_2 - z_1) > 1.$$

Note that

$$1 < r(z_2 - z_1) = (r-1)(z_2 - z_1) + (z_2 - z_1) \leq 1 + (z_2 - z_1) < 1 + \delta < 2.$$

Using r continue the sequence $\{x_i\}$ by defining points

$$x_p = x_{p+2} = x_{p+4} = \cdots = x_{p+2r} = z_1$$

and

$$x_{p+1} = x_{p+3} = x_{p+5} = \cdots = x_{p+2r-1} = z_2.$$

Write $\xi_{p+2j} = c_2$ and $\xi_{p+2j-1} = c_1$ for $j = 1, 2, \ldots, r$.

Finally complete the sequence $\{x_i\}$ by selecting points

$$z_1 = x_{p+2r} < x_{p+2r+1} < \cdots < x_{n-1} < x_n = b$$

along with associated points $\{\xi_i\}$ so that

$$\left| F(b) - F(z_1) - \sum_{i=p+2r+1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \eta/4.$$

This just uses the integrability hypotheses for f on $[z_1, b]$.

Consider now the sum

$$\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$

taken over the entire sequence thus constructed. Observe that

$$\sum_{i=1}^{n} |x_i - x_{i-1}| = \sum_{i=1}^{p} (x_i - x_{i-1}) + \sum_{i=p+1}^{p+2r} |x_i - x_{i-1}| + \sum_{i=p+2r+1}^{n} (x_i - x_{i-1})$$

$$= (z_1 - a) + 2r(z_2 - z_1) + (b - z_1) =$$

$$(b - a) + 2r(z_2 - z_1) \leq (b - a) + 4 = C.$$

Thus the points chosen satisfy the conditions of the definition for the δ selected and we must have

$$\left| F(b) - F(a) - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon < \eta/4.$$

On the other hand

$$\left[F(z_1) - F(a) + F(b) - F(z_1) - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right]$$

$$= \left[F(z_1) - F(a) - \sum_{i=1}^{p} f(x_i)(x_i - x_{i-1}) \right]$$

$$+ \left[F(b) - F(z_1) - \sum_{i=p+2r+1}^{n} f(\xi_i)(x_i - x_{i-1}) \right]$$

$$- \left[\sum_{i=p+1}^{p+2r} f(\xi_i)(x_i - x_{i-1}). \right]$$

From this we deduce that

$$\left| \sum_{i=p+1}^{p+2r} f(\xi_i)(x_i - x_{i-1}) \right| < 3\eta/4.$$

But a direct computation of this sum shows that

$$\sum_{i=p+1}^{p+2r} f(\xi_i)(x_i - x_{i-1}) = [f(c_1) - f(c_2)]r(z_2 - z_1)$$
$$> \eta r(z_2 - z_1) > \eta.$$

This contradiction completes the proof.

1.4 Characterization of Newton's integral

We have characterized the Newton integral of *continuous* functions entirely by a concept expressed in terms of Riemann sums. The Classical Newton integral can be equally well characterized in general by a slight modification, a modification that takes a uniform property and substitutes a more general pointwise property.

This is identical to an old problem of W. H. Young[7]: to determine necessary and sufficient conditions on a function f in order that it should be the derivative of some other function, i.e., in order that it have a Newton integral. We might recall from elementary calculus already one sufficient condition (that f might be continuous) and perhaps one necessary condition (that f should have the intermediate value property).

Theorem 1.10 (Characterization of derivatives) *A function $f : [a,b] \to \mathbb{R}$ is an exact derivative if and only if it has the following strong pointwise integrability property: there is a number I so that, for any choice of $\varepsilon > 0$ and $C > 0$, there must exist a positive function $\delta : [a,b] \to \mathbb{R}^+$ with the property that*

$$\left| I - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon$$

for any choice of points $x_0, x_1, \ldots, x_n$ and $\xi_1, \xi_2, \ldots, \xi_n$ from $[c,d]$ with these four properties:

1. *$a = x_0$ and $b = x_n$.*
2. *$0 < |x_i - x_{i-1}| < \delta(\xi_i)$ for all $i = 1, 2, \ldots, n$.*
3. *ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \ldots, n$.*
4. *$\sum_{i=1}^{n} |x_i - x_{i-1}| \le C$.*

Necessarily then,

$$I = \int_a^b f(x)\,dx.$$

[7] W. H. Young, A note on the property of being a differential coefficient, *Proc. London Math. Soc.* (2) **9** (1911) 360–368.

This theorem too gives us some insight into integration theory. Instead of basing the calculus integral on the concept of an antiderivative it could instead be based on a definition of an integral centered on the concept of Riemann sums. This gives us two equivalent formulations of the Newton integral: one uses an antiderivative and one uses Riemann sums. The latter has some theoretical advantages since it is hard to examine a function and conclude that it is a derivative without actually finding the antiderivative itself. Our main interest in this theorem is that it leads naturally to the correct and natural definition of the integral on the real line.

Exercise 16 *Prove that the condition stated in Theorem 1.10 is necessary. (The sufficiency proof is the trickier part.)* Answer □

1.4.1 Proof of Theorem 1.10

One direction the reader will already have checked in Exercise 16. We prove the converse direction. The proof is structured so as to be similar in many details to the proof of Theorem 1.9. The first three steps are essentially identical and will not need much commentary.

Step 1 We first show that if the function f satisfies the hypotheses of the theorem on an interval $[a,b]$ if and only if it satisfies the following equivalent strong integrability criterion: for every $\varepsilon > 0$ and $C > 0$, there is a positive function δ on $[a,b]$ with the property that

$$\left| \sum_{j=1}^{m} f(\xi'_j)(x'_j - x'_{j-1}) - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon$$

for any choice of points

$$x_0, x_1, \dots, x_n \text{ and } \xi_1, \xi_2, \dots, \xi_n \text{ and } x'_0, x'_1, \dots, x'_m \text{ and } \xi'_1, \xi'_2, \dots, \xi'_m$$

from $[a,b]$ satisfying $a = x_0 = x'_0$, $b = x_n = x'_m$, and

$$\sum_{i=1}^{n} |x_i - x_{i-1}| \leq C \text{ and } \sum_{j=1}^{m} |x'_j - x'_{j-1}| \leq C$$

where $a = x_0 = x'_0$, $b = x_n = x'_n$, $0 < |x_i - x_{i-1}| < \delta(\xi_i)$, $0 < |x'_j - x'_{j-1}| < \delta(\xi'_j)$, each ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \dots, n$, and each ξ'_j belongs to the interval with endpoints x'_j and x'_{j-1} for $j = 1, 2, \dots, m$,

This is a standard “Cauchy” version of the integrability condition. Such a statement is equivalent to the other version. It is an essential element of general integration theory to prove the equivalence of such statements.

Step 2 Now show that, if a function satisfies the hypotheses on an interval $[a,b]$, then it satisfies this same strong “integrability” criterion on every subinterval $[c,d] \subset [a,b]$: there is a number $I(c,d)$ so that, for every $\varepsilon > 0$ and $C > 0$, there is a positive function δ on $[a,b]$ with the property that

$$\left| I(c,d) - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon$$

for any choice of points $x_0, x_1, \dots, x_n$ and $\xi_1, \xi_2, \dots, \xi_n$ from $[c,d]$ satisfying

$$\sum_{i=1}^{n} |x_i - x_{i-1}| \leq C$$

where $c = x_0$, $d = x_n$, $0 < |x_i - x_{i-1}| < \delta(\xi_i)$ and each ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \dots, n$.

Step 3 Next show that

$$I(x,z) = I(x,y) + I(y,z)$$

for all $a \leq x < y, z \leq b$.

Step 4 Let us then suppose that f is a function possessing this strong integrability property on an interval $[a,b]$. We know (by Steps 1–3) that such a function f with these properties would have to have the same properties on each subinterval. Moreover there must be a function $F : [a,b] \to \mathbb{R}$ with $I(x,y) = F(y) - F(x)$ for each $a \leq x < y \leq b$.

We claim now that $F'(x) = f(x)$ at every point x in the interval $[a,b]$. Suppose (contrary to this) that there is a point z in the interval at which it is *not* true that

$$F'(z) = f(z).$$

One possibility is that this is because the upper right-hand (Dini) derivative at z exceeds $f(z)$ by some positive value $\eta > 0$. Another is that the value $f(z)$ exceeds the upper right-hand (Dini) derivative at z by some positive value $\eta > 0$. There are six other possibilities, corresponding to the other three Dini derivatives under which $F'(z) = f(z)$ might fail. It is sufficient for a proof that we show that this first possibility cannot occur. From this we will obtain a contradiction to the statement in the theorem.

Thus we will assume that there must be a positive number $\eta > 0$ so that we can choose an arbitrarily small positive number t so that the interval $[z, z+t]$ has this property:

$$\frac{F(z+t) - F(z)}{t} > f(z) + \eta$$

and hence so that

$$F(z+t) - F(z) > f(z)t + \eta t.$$

We give the details assuming this and that $a < z < b$. Now we apply the theorem using $\varepsilon < \eta/4$, and $C = b - a + 6$ to obtain a choice of positive function

δ that meets the conditions of the theorem. Choose a number $0 < t < 1$ for which $t < \delta(z)$ and $z+t < b$ and with the property that

$$F(z+t) - F(z) > f(z)t + \eta t.$$

Let s be the least integer so that $st > 2$. Note that, consequently,

$$2 < st = (s-1)t + t \leq 2 + t < 3.$$

We first select a sequence of points

$$z = u_0 < u_1 < u_2 < \cdots < u_{k-1} = z+t$$

and points υ_i from $[x_{i-1}, x_i]$ so that $0 < u_i - u_{i-1} < \delta(\upsilon_i)$ and

$$\left| F(z+t) - F(z) - \sum_{i=1}^{k-1} f(\upsilon_i)(u_i - u_{i-1}) \right| < \eta t/2$$

This is possible simply because f possesses the strong integrability property on the interval $[z, z+t]$. Now we add in the point $u_k = z$ and $\upsilon_k = z$.

We compute that

$$\sum_{i=1}^{k} f(\upsilon_i)(u_i - u_{i-1}) = -f(z)t + \sum_{i=1}^{k-1} f(\upsilon_i)(u_i - u_{i-1})$$

$$> -[F(z+t) - F(z) - \eta t] + \sum_{i=1}^{k-1} f(\upsilon_i)(u_i - u_{i-1}) > \eta t/2.$$

while at the same time

$$\sum_{i=1}^{k} |x_i - x_{i-1}| = 2t.$$

Repeat this sequence

$$z = u_0 < u_1 < \cdots < u_{k-1} > u_k = z$$

exactly s times so as to produce a sequence

$$z = u_0, u_1, \ldots u_{r-1}, u_r = z$$

with the property that

$$\sum_{i=1}^{r} f(\upsilon_i)(u_i - u_{i-1}) > \eta st/2 > \eta$$

while at the same time

$$\sum_{i=1}^{r} |u_i - u_{i-1}| = 2st < 6.$$

Now construct a sequence

$$a = z_0 < z_1 < \cdots < z_p = z$$

along with associated points ζ_i so that $0 < z_i - z_{i-1} < \delta(\zeta_i)$ and so that

$$\left| \int_a^z f(x)\,dx - \sum_{i=1}^{p} f(\zeta_i)(z_i - z_{i-1}) \right| < \eta/4.$$

We also need a sequence

$$z = w_0 < w_1 < \dots w_q = b$$

along with associated points ω_i so that $0 < w_i - w_{i-1} < \delta(\omega_i)$ and so that

$$\left| \int_z^b f(x)\,dx - \sum_{i=1}^{q} f(\omega_i)(w_i - w_{i-1}) \right| < \eta/4.$$

Both of these just use the strong integrability property of f on the subintervals $[a,z]$ and $[z,b]$

Now we put these three sequences together in this way

$$a = z_0 < z_1 < \cdots < z_p = z = u_0, u_1, \dots, u_r = z = w_0 < w_1 < \dots w_q = b$$

to form a new sequence $a = x_0, x_1, \dots, x_N = b$ for which $|x_i - x_{i-1}| < \delta(\xi_i)$ and for which

$$\sum_{i=1}^{N} |x_i - x_{i-1}| = (z-a) + 2st + (b-z) = b - a + 2st < b - a + 6 = C.$$

We use ξ_i in each case as the appropriate intermediate point used earlier: thus associated with an interval $[z_{i-1}, z_i]$ we had used ζ_i; associated with an interval $[w_{i-1}, w_i]$ we had used ω_i; while associated with a pair (u_{i-1}, u_i) we use υ_i.

Consider the sum

$$\sum_{i=1}^{N} f(\xi_i)(x_i - x_{i-1})$$

taken over the entire sequence thus constructed. Because the points satisfy the conditions of the theorem for the δ function selected we must have

$$\left| \int_a^b f(x)\,dx - \sum_{i=1}^{N} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon < \eta/4.$$

On the other hand

$$\left[\int_a^z f(x)\,dx + \int_z^b f(x)\,dx - \sum_{i=1}^{N} f(\xi_i)(x_i - x_{i-1}) \right]$$

$$= \left[\int_a^z f(x)\,dx - \sum_{i=1}^{p} f(\zeta_i)(z_i - z_{i-1}) \right]$$

$$+ \left[\int_z^b f(x)\,dx - \sum_{i=1}^{q} f(\omega_i)(w_i - w_{i-1}) \right]$$

$$+ \left[\sum_{i=1}^{r} f(\upsilon_i)(u_i - u_{i-1}) \right].$$

From this we deduce that

$$\sum_{i=1}^{r} f(\xi_i)(u_i - u_{i-1}) < 3\eta/4$$

and yet we recall that

$$\sum_{i=1}^{r} f(\xi_i)(u_i - u_{i-1}) > \eta st/2 > \eta.$$

This contradiction completes the proof.

1.5 How to generalize the integral?

The classical Newton integral is not nearly strong enough for the applications that are demanded of a theory of integration. This was apparent certainly by the early 19th century.[8]

A first suggestion that one might make is to interpret the integral

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

more broadly by no longer insisting that $F'(x) = f(x)$ holds *everywhere*. Thus we introduce an exceptional set N of points x where $F'(x) = f(x)$ might fail. This could be because $f(x)$ is not defined, or $F'(x)$ does not exist, or even where $F'(x)$ and $f(x)$ do exist but have different values.

The simplest such generalization would allow finite sets. Our text *The Calculus Integral* develops the theory of such an integral as a teaching tool that leads the student to the modern versions.

This simple generalization is to say that a function $f : [a,b] \to \mathbb{R}$ is Newton integrable *in the elementary sense* if there is a primitive $F : [a,b] \to \mathbb{R}$ with $F'(x) = f(x)$ for all x in $[a,b]$ with the possible exception of a finite set of points, at each of which the function F is nonetheless continuous. As usual,

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

The theory is not much harder to develop than that for the most severe Newton integral, and is rather more useful and pedagogically more promising. A more ambitious approach leads to the correct theory.

1.5.1 What sets to ignore?

We can ignore finite sets in defining a Newton integral (as we have just seen) provided we require the primitive function to be continuous at each point ignored. How can we determine a larger class of sets that can be ignored?

We have seen in Lemma **??** that the identity

$$F(x) = \int_a^x f(t)\,dt \quad (a \le x \le b)$$

[8] For an account of the integration difficulties of that era see David Bressoud [8], *Radical Approach to Real Analysis*, Washington, D.C.: Mathematical Association of America, 2007.

is closely related to the finding of a positive function δ so that the sum

$$\sum_{i=1}^{n} |F(a_i) - F(b_i) - f(\xi_i)(b_i - a_i)| \tag{1.5}$$

is small for partitions

$$\{([a_i, b_i], \xi_i) : i = 1, 2, 3, \dots\}$$

of $[a, b]$ finer than δ. This is merely because, if $F'(x) = f(x)$ everywhere, then we can choose $\delta(x) > 0$ small enough so that

$$|F(y) - F(x) - f(x)(y - x)| < \varepsilon|y - x|$$

for $|x - y| < \delta(x)$. In that case the sum in (**??**) is easily estimated:

$$\sum_{i=1}^{n} |F(a_i) - F(b_i) - f(\xi_i)(b_i - a_i)| \le \sum_{i=1}^{n} \varepsilon(b_i - a_i) = \varepsilon(b - a).$$

If we introduce a set of points N where $F'(x) = f(x)$ fails we are left trying to control a sum of the form

$$\sum_{i=1}^{n} |F(a_i) - F(b_i) - f(\xi_i)(b_i - a_i)|$$

for a subpartition

$$\{([a_i, b_i], \xi_i) : i = 1, 2, 3, \dots\}$$

of $[a, b]$ and with all associated points ξ_i belonging to the exceptional set N.

The easy and obvious way to control this is to require that the two sums

$$\sum_{i=1}^{n} |F(a_i) - F(b_i)| \quad \text{and} \quad \sum_{i=1}^{n} |f(\xi_i)|\,(b_i - a_i)$$

for such subpartitions are small.

This leads us to the notion of “small Riemann sums” and hence to the idea of sets of measure zero and functions of zero variation. We explore these now. We return to these ideas in Chapter 2 where detailed proofs are available. For our introductory chapter, however, the reader is encouraged to try to establish all statements now. This will help clarify the nature of the integrals to be defined.

1.6 Exceptional sets

We now define the notion of sets of measure zero and functions having zero variation using the theme of “small Riemann sums” and motivated by our discussion in the preceding section.

1.6.1 Sets of measure zero

A set E of real numbers is said to have measure zero[9] if certain Riemann sums of the form

$$\sum_{i=1}^{n} f(\xi_i)(y_i - x_i)$$

defined relative to that set are arbitrarily small. The definition assumes a rather familiar form and is, consequently, closely linked to ideas in integration theory. Indeed we can anticipate from the form of the definition that sets of measure zero are precisely those that play a key role in integration theory: they are the "negligible sets, the sets upon which the values of a function f can have no influence on the value of the integral of f.

Definition 1.11 (Measure zero) *A set E of real numbers is said to have measure zero if for every $\varepsilon > 0$ and all functions $f : E \to \mathbb{R}$ there is a positive function $\delta : E \to \mathbb{R}^+$ such that*

$$\left|\sum_{i=1}^{n} f(\xi_i)(y_i - x_i)\right| < \varepsilon$$

for all subpartitions $\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$ for which $y_i - x_i < \delta(\xi_i)$ and with all associated points $\xi_i \in E$.

Subpartitions anchored in a set Whenever a subpartition subpartitions

$$\pi = \{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$$

is given with all associated points ξ_i chosen from a fixed set E we say that π is *anchored in E*. Evidently this will occur rather frequently so it is useful to have some informal language that expresses the idea quickly.

A characterization of measure zero sets A simple and useful characterization is available in the following lemma. Here we have replaced the requirement to check "all functions" with an easier handled version that requires producing only one function with the small Riemann sums property. We shall find, in the next chapter, two more characterizations including the original definition of Lebesgue.

Lemma 1.12 *A set E of real numbers has measure zero if and only if there is some positive function $g : E \to \mathbb{R}^+$ that, for every $\varepsilon > 0$, there is a positive function $\delta : E \to \mathbb{R}^+$ such that*

$$\sum_{i=1}^{n} g(x)(y_i - x_i) < \varepsilon$$

for all subpartitions $\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$ anchored in E and finer than δ.

[9] Lebesgue's definition was different but equivalent.

As a corollary we have this special condition which could also have been taken as our definition of sets of measure zero.

Corollary 1.13 *A set E of real numbers has measure zero if and only if for every $\varepsilon > 0$, there is a positive function $\delta : E \to \mathbb{R}^+$ such that*

$$\sum_{i=1}^{n} (y_i - x_i) < \varepsilon$$

for all subpartitions $\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$ anchored in E and finer than δ.

Exercise 17 *Show that all subsets of a set of measure zero are also of measure zero.* □

Exercise 18 *Show that a union of a sequence of sets of measure zero is also of measure zero.* □

Exercise 19 *Show that countable sets have measure zero.* □

Exercise 20 (infinite derivatives) *Let $N \subset (a, b)$ be a set of measure zero. Show that there is a nondecreasing function $\phi : [a, b] \to \mathbb{R}$ so that $\phi'(x) = +\infty$ for all $x \in N$.* Answer □

1.6.2 Proof of Lemma 1.12

In one direction this is easy since if this property holds for all such functions f then it certainly holds a particular choice of positive function $g(x)$ on E.

Suppose that $g : E \to \mathbb{R}^+$ is a positive function with the property that, for every $\varepsilon > 0$, there is a positive function $\delta : E \to \mathbb{R}^+$ such that

$$\sum_{i=1}^{n} g(x)(y_i - x_i) < \varepsilon$$

for all subpartitions $\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$ anchored in E and finer than δ. Fix a positive integer k and let E_k denote the set of points $x \in E$ at which $g(x) > 1/k$.

Take any function f on E_k and write

$$E_{mk} = \{x \in E_k : m - 1 \le |f(x)| < m\}$$

for each integer $m = 1, 2, 3, \dots$, noting that the union of the sets E_{mk} is the set E_k itself. Define δ_m on E so that

$$\sum_{i=1}^{n} g(x)(y_i - x_i) < \varepsilon 2^{-m} m^{-1} k^{-1}$$

for all subpartitions $\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$ anchored in E and finer than δ_m.

Now take $\delta(x) = \delta_m(x)$ for $x \in E_{mk}$. Consider a subpartition

$$\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$$

anchored in E_k and finer than δ. Then

$$\left|\sum_{i=1}^{n} f(\xi_i)(y_i - x_i)\right| \leq \sum_{i=1}^{n} |f(\xi_i)|\,(y_i - x_i)$$

$$\leq \sum_{m=1}^{\infty} \sum mkg(\xi_i)(y_i - x_i)$$

$$\leq \sum_{m=1}^{\infty} mk\left[\varepsilon 2^{-m} m^{-1} k^{-1}\right] = \varepsilon.$$

This requires separately summing the terms for which ξ_i belongs to different sets E_{mk}.

This verifies that each set E_k has measure zero by definition. By Exercise 18 a union of a sequence of sets of measure zero is also of measure zero. Hence E is also of measure zero.

1.7 Zero variation

A function F is said to have zero variation on a set E of real numbers if a certain Riemann sum of the form

$$\sum_{i=1}^{n} |F(y_i) - F(x_i)|$$

defined relative to that set is arbitrarily small. The definition assumes again a familiar form and is also closely linked to ideas in integration theory.

Definition 1.14 (Zero variation) *A function $F : \mathbb{R} \to \mathbb{R}$ is said to have zero variation on a set E of real numbers if for every $\varepsilon > 0$ there is a positive function $\delta : E \to \mathbb{R}^+$ such that*

$$\sum_{i=1}^{n} |F(y_i) - F(x_i)| < \varepsilon$$

whenever a subpartition

$$\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$$

is anchored in E and finer than δ.

We should note an obvious connection between these two notions: a set E has measure zero if and only if the identity function $F(x) = x$ has zero variation on the set E. Indeed most of the functions one encounters in the calculus will have this feature: they have zero variation on sets of measure zero. This is a key observation for integration theory as it will turn out.

Exercise 21 *Show that if a function has zero variation on each set in a sequence $E_1, E_2, E_3, \dots$ then that function has zero variation on the union $\bigcup_{n=1}^{\infty} E_n$.* □

Exercise 22 *Show that a continuous function has zero variation on every countable set.*

Answer □

Exercise 23 *Show that a Lipschitz function has zero variation on every set of measure zero.* □

Exercise 24 *Show that a function with a zero derivative at every point of a set E has zero variation on E.* Answer □

Exercise 25 *Show that a function with a finite derivative at every point of a measure zero set E has zero variation on E.* Answer □

Exercise 26 *Show that a function that has zero variation on an open interval (a,b) is necessarily constant on that interval.* Answer □

Exercise 27 (inverse function) *Let $F : [a,b] \to \mathbb{R}$ be a continuous, strictly increasing function and let $G : [F(a), F(b)] \to \mathbb{R}$ be its inverse. Let $N \subset [a,b]$. Show that G has zero variation on the set $F(N)$ if and only if N has measure zero.* Answer □

Exercise 28 (left inverse) *Let $F : [a,b] \to \mathbb{R}$ be a nondecreasing function. Define the left-inverse of F as the function $G : [F(a), F(b)] \to \mathbb{R}$ defined by*

$$G(y) = \inf\{z \in [a,b] : F(z) \geq y\}.$$

Show that

1. *G is nondecreasing and left-continuous.*

2. *G has a jump discontinuity at a point y_0 in $[F(a), F(b))$ if and only if $F(u) = y_0$ for all u in some open interval (s,t).*

3. *$G(F(t)) \leq t$ for every $t \in [a,b]$. Moreover, $G(F(t)) < t$ can occur if and only if F is constant on some interval $[s,t]$.*

4. *$G(y) = u_0$ for all y in some interval $(y_1, y_2) \subset [F(a), F(b))]$ if and only if F has a jump discontinuity at u_0 and $(y_1, y_2) \subset (F(u_0-), F(u_0+)$.*

5. *If F is strictly increasing then G is a left-inverse of F in the sense that $G(F(t) = t$ for all $t \in [a,b]$ and G is continuous.*

6. *If F is strictly increasing, $N \subset [a,b]$, and G has zero variation on the set $F(N)$, then N has measure zero.*

[This exercise is needed in the sequel and placed here for easy reference.]

Answer □

1.8 Generalized Newton integral

We have defined the classical Newton integral (much as Newton himself might have) as requiring for the expression

$$\int_a^b f(t)\,dt = F(b) - F(a)$$

that $F'(x) = f(x)$ at *every* point of the interval $[a,b]$. There are many reasons for relaxing this to allow exceptions, a small set of points at which f may fail to be defined or at which $F'(x) = f(x)$ might fail. To make this work, however, requires some assumptions to be added to the function F.

Here are the variants that can be usefully introduced to beginning students of integration theory, arranged in increasing order of generality (and difficulty). In each case f is said to be integrable in that sense and the value

$$\int_a^b f(t)\,dt = F(b) - F(a)$$

is assigned to the integral:

Classical f is defined on $[a,b]$ and $F'(x) = f(x)$ at every point of $[a,b]$.

Naive f is defined at least on (a,b), F is continuous on $[a,b]$ and $F'(x) = f(x)$ at every point of (a,b).

Elementary f is defined at least on (a,b) except possibly at finitely many points, F is continuous on $[a,b]$, while $F'(x) = f(x)$ at every point of (a,b) with at most finitely many exceptions.

Utility f is defined at least on (a,b) except possibly for a countable set N, F is continuous on $[a,b]$, while $F'(x) = f(x)$ at every point of (a,b) excepting possibly points in the set N.

General f is defined at least on (a,b) except possibly for a set N of measure zero, F is continuous on $[a,b]$ and has zero variation on N, while $F'(x) = f(x)$ at every point of (a,b) excepting possibly points in the set N.

1.8.1 Exercises

Exercise 29 *Show that the function*

$$f(x) = \frac{1}{\sqrt{x}}$$

is not Newton integrable on $[0,1]$ *in the classical sense but is integrable in a naive Newton sense.* □

Exercise 30 *Show that the function*

$$f(x) = \frac{1}{\sqrt{|x|}}$$

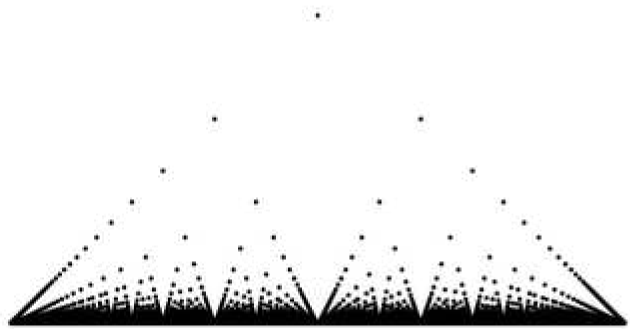

Figure 1.3: Graph of the popcorn function.

is not Newton integrable on $[-1,1]$ in the classical or naive senses but is integrable in the elementary Newton sense. □

Exercise 31 *Show that the popcorn function of Figure 1.3 is Newton integrable in the utility sense, but is not integrable in the classical, naive, or elementary Newton senses. [The popcorn function is the function $P:[0,1]\to\mathbb{R}$ defined by $P(x)=0$ for x irrational and $P(x)=1/q$ if $x=p/q$ expresses the rational number x in its lowest terms.]* Answer □

Exercise 32 *Suppose that $f:[a,b]\to\mathbb{R}$ is integrable in one of the Newton senses with F as an indefinite integrable. Prove that F has this "absolute continuity" property: F has zero variation on every subset $N\subset(a,b)$ of measure zero.* Answer □

Exercise 33 (Justifying the naive integral) *Let F, G, $f:[a,b]\to\mathbb{R}$ be functions on $[a,b]$. Suppose that both F and G are continuous on $[a,b]$ and that $F'(x)=G(x)=f(x)$ for all x in (a,b). Show that $F(b)-F(a)=G(b)-G(a)$.* Answer □

Exercise 34 (Justifying the utility integral) *Let F, G, $f:[a,b]\to\mathbb{R}$ be functions on $[a,b]$. Suppose that*

1. *Both F and G are continuous.*
2. *$F'(x)=f(x)$ for all x in (a,b) with at most countably many exceptions.*
3. *$G'(x)=f(x)$ for all x in (a,b) with at most countably many exceptions.*

Show that $F(b)-F(a)=G(b)-G(a)$. Answer □

Exercise 35 (Justifying the general integral) *Let F, G, $f:[a,b]\to\mathbb{R}$ be functions on $[a,b]$. Suppose that*

1. *Both F and G are continuous.*
2. $F'(x) = f(x)$ *for all x in (a,b) excepting at most a set of points N_1 of measure zero.*
3. $G'(x) = f(x)$ *for all x in (a,b) excepting at most a set of points N_2 of measure zero.*
4. *F has zero variation on N_1.*
5. *G has zero variation on N_2.*

Show that $F(b) - F(a) = G(b) - G(a)$. Answer □

1.8.2 Newton integral: controlled version

There is yet another variant on the Newton integral. In this section let us discuss one of these suggested recently by Hana Bendová and Jan Malý [5][10]. The reader who has seen quite enough of Newton-type integrals may wish to proceed directly to the equivalent Henstock-Kurzweil definition of the integral, since this is the integral that is our main object of study in the text. But it is too tempting to introduce yet one more variant for your consideration.

The controlled version of the Newton integral is, as it turns out, just another way of expressing the general Newton integral without alluding to derivatives or sets of measure zero. This shifts the technical details from derivatives and variation to what the authors call a "control function." The control function is not too difficult to manipulate, but some students might find it rather non-intuitive.

We review the naive Newton integral but in a rather more suggestive language that lends itself to an obvious generalization. In order for $f : (a,b) \to \mathbb{R}$ to be Newton integrable in the naive sense the following must hold:

> There is a continuous function $F : [a,b] \to \mathbb{R}$ and, for each $x \in (a,b)$,
> $$\lim_{y\to x} \frac{F(y) - F(x) - f(x)(y-x)}{y-x} = 0.$$
> In that case the value of the integral may be taken as
> $$\int_a^b f(x)\,dx = F(b) - F(a).$$

Bendová and Malý [5] modify this formulation to produce the following definition, which they offer as a reasonable alternative starting point for teaching the general theory of integration on the real line.

[10] H. Bendová and J. Malý, An elementary way to introduce a Perron-like integral, Annales Academiae Scientiarum Fennicae Mathematica, Volumen 36, 2011, 153–164.

Definition 1.15 (Controlled Newton) *A function $f : (a,b) \to \mathbb{R}$ is said to be Newton integrable in the controlled sense provided there is a continuous function $F : [a,b] \to \mathbb{R}$ and there is a strictly increasing function $\phi : (a,b) \to \mathbb{R}$ called the "control" with the property that, for each $x \in (a,b)$,*

$$\lim_{y\to x} \frac{F(y) - F(x) - f(x)(y-x)}{\phi(y) - \phi(x)} = 0.$$

In that case the value of the integral is taken as

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

The exercise below is needed to justify the definition. It is clear that this integral is more general than the naive Newton integral. It is true, although not immediately apparent, that this integral is equivalent to the general Newton integral. In the article [5] can be found an incomplete sketch of the program that could be followed to develop integration theory from this starting point. They include some of the basic tools for integration like integration by parts and change of variables and go as far as a proof of the monotone convergence theorem. The techniques would be accessible to most students of elementary analysis (particularly so for the excellent mathematics students at Charles University in Prague where the authors of this article teach.)

Theorem 1.16 *A function $f : (a,b) \to \mathbb{R}$ is Newton integrable in the controlled sense if and only if it is integrable in the general Newton sense. Moreover the values of the integral agree.*

Exercise 36 (Justifying the controlled integral) *Show that, if there are two continuous functions F_1, $F_2 : (a,b) \to \mathbb{R}$ that satisfy the hypotheses of Definition 1.15, then $F_1(b) - F_1(a) = F_2(b) - F_2(a)$.*

Answer □

Exercise 37 *Show that the condition that, for each $x \in (a,b)$,*

$$\lim_{y\to x} \frac{F(y) - F(x) - f(x)(y-x)}{\phi(y) - \phi(x)} = 0$$

in Definition 1.15 is sufficient to force F to be continuous at every point of (a,b) (but not necessarily continuous at a and b). □

1.8.3 Proof of Theorem 1.16

This proof appeals to the Lebesgue differentiation theorem in order to prove that a function that is Newton integrable in the controlled senses is also integrable in the general Newton sense. That material appears later in this Chapter (Section 1.14), so this proof might be left until later.

Let $f : (a,b) \to \mathbb{R}$ be Newton integrable in the controlled sense. Then there is an indefinite integral $F : [a,b] \to \mathbb{R}$ and a strictly increasing function $\phi : (a,b) \to \mathbb{R}$ that serves as the control. This means that for each $x \in (a,b)$,

$$\lim_{y\to x} \frac{F(y) - F(x) - f(x)(y-x)}{\phi(y) - \phi(x)} = 0.$$

Let x be a point at which ϕ is differentiable. Then

$$\frac{F(y) - F(x) - f(x)(y-x)}{y-x}$$
$$= \frac{F(y) - F(x) - f(x)(y-x)}{\phi(y) - \phi(x)} \times \left(\frac{\phi(y) - \phi(x)}{y-x}\right) \to 0 \times \phi'(x) = 0$$

as $y \to x$. So $F'(x) = f(x)$ at each such point. By the Lebesgue differentiation theorem[11] it follows that $\phi'(x)$ exists and that $F'(x) = f(x)$ for a.e. point x in (a,b).

It follows that F is an indefinite Newton integral for f provided we can prove one further property of F: if $N \subset (a,b)$ is a set of measure zero then F has zero variation on N. Fix an interval $[c,d] \subset (a,b)$. Let $\varepsilon > 0$ and write

$$\eta = \frac{\varepsilon}{2[\phi(d) - \phi(c)]}.$$

For each $x \in N$ there is a $\delta_1(x) > 0$ so that, for all $0 < |y-x| < \delta_1(x)$,

$$\left|\frac{F(y) - F(x) - f(x)(y-x)}{\phi(y) - \phi(x)}\right| < \eta.$$

Since N is a set of measure zero there is a positive function $\delta_2 : N \to \mathbb{R}^+$ so that

$$\sum_{([u,v],w)\in\pi} |f(w)|\,(v-u) < \varepsilon/2$$

whenever π is a subpartition anchored in N finer than δ_2. Take $\delta = \min\{\delta_1, \delta_2\}$ and use it to verify that F has zero variation on $N \cap [c,d]$. If π is a subpartition of $[c,d]$ anchored in N finer than δ we compute that

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u)| \leq$$
$$\leq \sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)| + \sum_{([u,v],w)\in\pi} |f(w)|\,(v-u)$$
$$< \sum_{([u,v],w)\in\pi} \eta\,|\phi(v) - \phi(u)| + \varepsilon/2$$
$$\leq \eta[\phi(d) - \phi(c)] + \varepsilon/2 = \varepsilon.$$

[11] Not yet proved in the text, but we shall invoke it here nonetheless. See Sections 1.14 and 2.11.

It follows that F has zero variation on $N\cap[c,d]$. Thus N can be expressed as

$$N=\bigcup_{n=1}^{\infty} N\cap[a+n^{-1},b-n^{-1}],$$

i.e., as the union of a sequence of sets on each of which has F has zero variation. Consequently F has zero variation on N.

This completes the proof that every function Newton integrable in the controlled sense must be also integrable in the general Newton sense. The same function F appears as the indefinite integral for both.

Let us now prove the opposite direction. Let $f:(a,b)\to\mathbb{R}$ be Newton integrable in the general sense. Then there is an indefinite integral $F:[a,b]\to\mathbb{R}$ in that sense and we need to construct a strictly increasing function $\phi:(a,b)\to\mathbb{R}$ that will serve as the control in the sense of Definition 1.15.

Let N be the set of points x in (a,b) at which $F'(x)=f(x)$ fails. We know that F has zero variation on N and that N itself is a set of measure zero. Define $\phi_1(x)=x$ for each point x in (a,b) and note that

$$\lim_{y\to x}\frac{F(y)-F(x)-f(x)(y-x)}{\phi_1(y)-\phi_1(x)}=0.$$

for each x in (a,b) that is not in N. Since N is a set of measure zero we can choose (see Exercise 20) a nondecreasing function ϕ_2 for which $\phi_2'(x)=+\infty$ for each $x\in N$. Note that

$$\lim_{y\to x}\frac{f(x)(y-x)}{\phi_2(y)-\phi_2(x)}=0.$$

Finally let us construct the third nondecreasing function ϕ_3.

Since N is a set on which F has zero variation there is, for each integer $n=1,2,3,\dots$ a positive function $\delta_n:N\to\mathbb{R}^+$ so that

$$\sum_{([u,v],w)\in\pi}|F(v)-F(u)|<2^{-n}$$

whenever π is a subpartition anchored in N finer than δ_n.

First, for each integer $n=1,2,3,\dots$ define the function $G_n(x)$ at each point $a<x<b$ by requiring $G_n(x)$ to be the supremum of the values

$$\sum_{([u,v],w)\in\pi}|F(v)-F(u)|$$

taken over all subpartitions π of $[a,x]$ anchored in N and finer than δ_n.

Note that, for any integer n and all $k=1,2,3,\dots,n$, if $x\in N$ and if $0<y-x<\delta_n(x)$ then $([x,y],x)$ is finer than δ_k and so

$$G_k(y)-G_k(x)\ge|F(y)-F(x)|.$$

Similarly if $0<x-y<\delta_n(x)$ then $([y,x],x)$ is finer than δ_k and so

$$G_k(x)-G_k(y)\ge|F(x)-F(y)|.$$

We now ready to define our third control

$$\phi_3(x) = \sum_{k=1}^{\infty} G_k(x).$$

This is a finite-valued function, nondecreasing on (a,b). Note that, if $0 < y-x < \delta_n(x)$ then

$$\phi_3(y) - \phi_3(x) \geq n\,|F(y) - F(x)|\,.$$

and if $0 < x-y < \delta_n(x)$ then

$$\phi_3(x) - \phi_3(y) \geq n\,|F(x) - F(y)|\,.$$

Consequently for each $x \in (a,b)$,

$$\lim_{y\to x} \frac{F(y)-F(x)}{\phi_3(y)-\phi_3(x)} = 0.$$

Putting these three controls together

$$\phi = \phi_1 + \phi_2 + \phi_3$$

we can now check that for each $x \in (a,b)$,

$$\lim_{y\to x} \frac{F(y)-F(x)-f(x)(y-x)}{\phi(y)-\phi(x)} = 0.$$

This supplies the needed control for Definition 1.15 and hence f is integrable in the sense of that definition with the same function F for its indefinite integral.

1.8.4 Continuous linear functionals

The essential features of integration theory are often described in the language of functionals. We view can any integration method as producing a functional

$$f \longrightarrow \int_a^b f(x)\,dx$$

assigning a real number to any function f belonging to a specified class of functions. The structure of that functional can be discussed using this terminology.

(monotone) If $f,\, g : [a,b] \to \mathbb{R}$ are integrable and $f(x) \leq g(x)$ for all $x \in [a,b]$ then

$$\int_a^b f(t)\,dt \leq \int_a^b g(t)\,dt.$$

(linear) If $f_1,\ f_2,\ \dots f_n : [a,b] \to \mathbb{R}$ are integrable and $h(x) = \sum_{i=1}^n c_i f_i(x)$ is a linear combination of these functions, then h is integrable in the same sense and

$$\int_a^b h(x)\,dx = \sum_{i=1}^n c_i \left(\int_a^b f_i(x)\,dx \right).$$

(continuous) If $f_1,\ f_2,\ f_3,\ \dots$ is a uniformly convergent sequence of integrable functions on $[a,b]$ and $f(x) = \lim_{n\to\infty} f_n(x)$. then f is integrable in the

same sense and
$$\int_a^b f(x)\,dx = \lim_{n\to\infty}\int_a^b f_i(x)\,dx.$$

Proofs that each of the Newton variants produces an integration theory with these three properties are requested in the exercises below. (As we shall see, however, the elementary version does not quite allow the continuity property since a uniform limit of functions integrable in that sense might fail to be integrable.)

Notation A slight shift in notation makes these functional properties rather more transparent. Use
$$\Gamma(f) = \int_a^b f(x)\,dx$$
and rewrite the three properties above in rather more of a slogan-like form:

(monotone) If $f \le g$ then $\Gamma(f) \le \Gamma(g)$.

(linear) If $h = \sum_{i=1}^n c_i f_i$ then $\Gamma(h) = \sum_{i=1}^n c_i \Gamma(f_i)$.

(continuous) If $f_n \to f$ uniformly then $\Gamma(f) = \lim_{n\to\infty}\Gamma(f_n)$.

Exercise 38 *Check the monotone property for one or more of the five Newton variants.* Answer □

Exercise 39 *Check the linear property for one or more of the five Newton variants.*

Answer □

Exercise 40 *Show that the continuity property does not hold for the elementary Newton integral but does hold for the other four Newton variants.* Answer □

1.8.5 Which Newton variant should we teach

The classical variant (requiring that an integrable function should be everywhere a derivative) is rather too severe, even for a calculus class. The naive version is useful, particularly because it allows the integration of some unbounded functions that commonly require “improper integration” techniques when discussed in conventional calculus classes.

The elementary version, however, seems the most useful[12] for beginning calculus students, especially as it requires no technical apparatus beyond the

[12] Our textbook *The Calculus Integral* [85] follows this approach and presents a full course of instruction on such an integral.

use of the mean-value theorem. The generality is sufficient for virtually all applications that are intended in calculus courses and the presentation is considerably easier than an introduction to the Riemann integral would require. One can still employ Riemann sums in applications, but the integral itself need not be developed from such sums.

The utility version is useful for more serious classes that wish to have a deeper theory of integration but are not prepared for the Lebesgue integral and beyond. This integrates all regulated functions[13] and is easier in many ways than a serious account of the Riemann integral would require. This is the point of view taken in the elementary analysis text by Elias Zakon [93][14]. Thus, in his text, all integrals concern continuous functions that are differentiable except possibly at the points of some sequence of exceptional points.

The general version (equivalently the controlled version) is, as it will turn out, exactly the generalization that is needed for the vast majority of integration applications on the real line (it is equivalent to the Henstock-Kurzweil integral). This allows a reasonable stab at using the full theory of integration without too much technical preparation. It is not likely to be adapted in undergraduate instruction anytime soon, but one should keep it in mind nonetheless.

1.9 Constructive aspects: the regulated integral

The Newton integral is not constructive. We do not have a method for determining which functions are integrable in any of these senses since we do not have any methodology, in general, for determining if a function f has an antiderivative F. Nor is the value of the integral constructive: if we cannot find F we cannot compute $F(b) - F(a)$ to evaluate the integral. This means that the theory remains largely formal: it offers a way to discuss and develop a theory of integration, but we are somewhat in mystery as to the full nature of the methods until we do much more analysis.

To make the integral constructive we are forced to narrow the scope of the integral. Both the Riemann and Lebesgue methods of integration do this. Some authors suggest avoiding the Riemann methods as they do not lead to a satisfying theory of integration that justifies the effort. Many more authors avoid the Lebesgue methods as too difficult for undergraduate instruction. In this chapter, we too avoid the Lebesgue methods because we are trying not to work quite so hard (at least until later). In Chapter 4 we do present Lebesgue's constructive theory.

The regulated functions are easy to work with and offer a constructive (if lim-

[13]A function is regulated if it has finite one-sided limits at every point. See Section 1.9.1.

[14]E. Zakon, *Mathematical Analysis I*, ISBN 1-931705-02-X, published by The Trillia Group, 2004.

ited) theory of integration on the real line. The Bourbaki[15] suggestion for instruction at a lower level was to focus just on regulated functions. As Berberian [6][16] explained it

> At the outset, I hasten to say that I remain a "Riemann loyalist": pound for pound, the Riemann circle of ideas can't be beat for its instructional value to the beginning student of analysis. Consequently, I wouldn't go so far as to suggest that the theory of regulated functions replace the Riemann integral in the beginning undergraduate analysis course; however, in a graduate course in real variables, the theory of regulated functions can be an entertaining alternative to a routine review of the Riemann integral; and it is in some ways a more instructive prelude to the Lebesgue theory, as I hope to persuade the reader in this brief "comparative anatomy" of integration theories.

A less ambitious teaching integral (intended for first-year students) by Dixmier [25][17] used a theory that is limited to functions that are piecewise continuous and have finite one-sided limits at the discontinuities. This would correspond to our elementary Newton integral, but simplified by refining considerably the class of functions under consideration so as to have available a constructive process.

1.9.1 Step functions and regulated functions

Let $f : [a,b] \to \mathbb{R}$.

1. If $f(x) = \chi_I(x)$ where $I \subset [a,b]$ is an interval $I = [c,d]$, (c,d), $[c,d]$, or $(c,d]$ then f is said to be a single-step function.

2. If f is a finite linear combination of single-step functions then f is said to be a step function.

3. If f is a uniform limit on $[a,b]$ of a sequence of step functions then f is said to be a regulated function.

The following observations are nearly immediate conclusions we can make from our brief study of the Newton integral. In each case we see that we verify integrability and compute the value of the integral.

[15] If you don't happen to know who N. Bourbaki "was" please do some Wiki research. You will find it more entertaining than anything that could be put here in a footnote.

[16] Sterling K. Berberian, Classroom Notes: Regulated Functions: Bourbaki's Alternative to the Riemann Integral. Amer. Math. Monthly 86 (1979), no. 3, 208–211.

[17] Jacques Dixmier, Cours de mathématiques du premier cycle. Cahiers Scientifiques, Fasc. 30. Gauthier-Villars, Paris, 1967.

Lemma 1.17 *If $f(x) = \chi_I(x)$ is a single-step function on $[a,b]$ then f is integrable [elementary sense] and*

$$\int_a^b f(x)\,dx = \int_a^b \chi_I(x)\,dx = \lambda(I).$$

Lemma 1.18 *If f is a step function on $[a,b]$ then f is integrable [elementary sense] and*

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n} c_i \lambda(I_i).$$

where

$$f(x) = \sum_{i=1}^{n} c_i \chi_{I_i}(x)$$

is any representation of f as a linear combination of single-step functions.

Lemma 1.19 *If f is a regulated function on $[a,b]$ then f is integrable [utility sense] and*

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

where $\{f_i\}$ is any representation of f as a uniform limit of step functions.

Exercise 41 *Verify that a single-step function is integrable [elementary sense] and show how to evaluate the integral.* Answer □

Exercise 42 *Show that a function $s : [a,b] \to \mathbb{R}$ is a step function if and only there are points*

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

so that s is constant on each open interval (x_{i-1}, x_i) for $i = 1, 2, \dots, n$ and verify that, in that case,

$$\int_a^b s(x)\,dx = \sum_{i=1}^{n} c_i (x_i - x_{i-1})$$

where c_i is the value of s on (x_{i-1}, x_i). □

Exercise 43 *Show that a regulated function $f : [a,b] \to \mathbb{R}$ has these properties:*

1. *f is bounded.*
2. *f is continuous nearly everywhere on $[a,b]$.*
3. *f has finite one-sided limits at each point of $[a,b]$.*

Answer □

Exercise 44 *Show that a function $f : [a,b] \to \mathbb{R}$ is regulated if and only if it has finite one-sided limits at each point of $[a,b]$.* Answer □

Exercise 45 *Suppose that* $f : [a,b] \to \mathbb{R}$ *is regulated and let* F *be an indefinite integral. Show that* F *has finite one-sided derivatives at each point, in fact that, for all* $a \leq x < b$,

$$D^+F(x) = \lim_{t \to x+} f(t)$$

and, for all $a < x \leq b$,

$$D^-F(x) = \lim_{t \to x-} f(t).$$

□

Exercise 46 *Show that any monotonic function is regulated.* Answer □

Exercise 47 *The regulated functions are not the only class of Newton integrable functions whose integrals can be constructed. Show that if* $F : [a,b] \to \mathbb{R}$ *is a continuous function that has a continuous derivative* F' *on* (a,b) *then the value of the integral of* F' *can be constructed.* Answer □

Exercise 48 *Show that the regulated integral (unlike the Newton integral) is an absolute integration method, i.e. that whenever a function* f *is regulated so too is the function* $|f|$. Answer □

Exercise 49 *Show that Dixmier's teaching integral (unlike the Newton integral) is an absolute integration method, i.e. that whenever a function* f *is integrable by his method so too is the function* $|f|$. Answer □

1.10 Riemann's integral

Integration theory developed in quite a different direction historically. The five versions of Newton's integral might have offered a different history to the subject, but the popularity of Riemann's ideas led to a more stagnant and less insightful development.

By the middle of the nineteenth century Riemann, clearly inspired by Cauchy's clarification of Newton's theory, was lecturing on a general integration theory based on it. This is a standard and time-honored tradition in mathematics. What an earlier mathematician proposes as a theorem, you will propose as a definition. Thus Theorem 1.8 turned into this definition.

Definition 1.20 (Riemann integral) *Let f be a function that is defined at every point of $[a,b]$. Then, f is said to be Riemann integrable on $[a,b]$ if it satisfies the following "uniform integrability" criterion: there is a number I so that, for every $\varepsilon > 0$ there is a $\delta > 0$, with the property that*

$$\left| I - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon$$

whenever points are given

$$a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n = b$$

for which

$$x_i - x_{i-1} < \delta$$

with associated points $\xi_i \in [x_{i-1}, x_i]$.

The number I in the definition would then be written in integral notation as

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

Compatibility The Riemann integral is compatible with the various versions of the Newton integral but the relationship is murky. All continuous functions are integrable in the Riemann sense and in the classical Newton sense with the values of the integrals agreeing (of course). Most bounded functions (but not all) that are integrable in the Newton sense (elementary or utility) are Riemann integrable and, again, the values of the integrals agree. Not all Riemann integrable functions are integrable in the Newton sense (elementary or utility). It is safe, however, to write

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

when we are sure that the function f is integrable in both senses. For unbounded functions the Riemann integral cannot be used.

We will see that all regulated functions are Riemann integrable. So the Riemann offers another (more complicated) constructive approach to integration theory. The class of Riemann integrable functions is larger than the class of regulated functions and one would think that that alone would justify pursuing this theory. The class of Riemann integrable functions, however, does not play a truly significant role in analysis and it is better for many reasons to bypass the theory except for historical curiosity.

The general Newton integral does include the Riemann integral (as well as the Lebesgue integral) and is sufficiently general for all but very specialized integration problems.

1.10.1 Integrability criteria

What functions are Riemann integrable? This seems an important question, and it certainly is an important question if we are to commit ourselves to a study of this integral. In fact, though, we are not much interested in this very limited integral and would not want to make too much of an effort to answer this question. We answer it here only in order to introduce some methods and ideas that will be useful later on.

The Newton integral (along with its several variants) makes a good teaching integral. The Riemann integral is rather overrated as a teaching integral and we wish to make minimal use of it. Even so we pause here to present a few integrability criteria for historical interest and for motivation.

The following list of criteria assists in determining what functions are or are not Riemann integrable. Let $f : [a,b] \to \mathbb{R}$ and let $\lambda(J)$ denote the length of the interval J (allowing also J to be a degenerate interval [i.e., empty or a single point]).

Cauchy Criterion For every $\varepsilon > 0$ there is a $\delta > 0$, with the property that

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{m} [f(\xi_i) - f(\eta_j)]\, \lambda(([a_i,b_i] \cap [c_j,d_j]) \right| < \varepsilon \tag{1.6}$$

whenever

$$\{([a_i,b_i],\xi_i) : i = 1,2,3,\ldots,n\}$$

and

$$\{([c_j,d_j],\eta_j) : j = 1,2,3,\ldots,m\}$$

are partitions of $[a,b]$ that are finer than δ.

Criterion M For every $\varepsilon > 0$ there is a $\delta > 0$, with the property that

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \left| f(\xi_i) - f(\eta_j) \right| \lambda(([a_i,b_i] \cap [c_j,d_j]) < \varepsilon$$

whenever

$$\{([a_i,b_i],\xi_i) : i = 1,2,3,\ldots,n\}$$

and

$$\{([c_j,d_j],\eta_j) : j = 1,2,3,\ldots,m\}$$

are subpartitions of $[a,b]$ that are finer than δ. (The criterion is named in honor of E. J. McShane who used it in a different context.)

Riemann's Criterion For every $\varepsilon > 0$ there is a partition of $[a,b]$

$$\{([a_i,b_i],\xi_i) : i = 1,2,3,\ldots,n\}$$

with the property that

$$\sum_{i=1}^{n} \omega f([a_i,b_i])(b_i-a_i) < \varepsilon.$$

Lebesgue's Criterion f is bounded and almost everywhere continuous[18].

Theorem 1.21 *Each of the four criteria is equivalent to the Riemann integrability of the function f.*

Exercise 50 *Show that the Riemann integral (unlike the Newton integral) is an absolute integration method, i.e. that whenever a function f is Riemann integrable on an interval so too is the function $|f|$.*

Answer □

1.10.2 Proof of Theorem 1.21

We shall prove that the four criteria are equivalent and that any one of them is in turn equivalent to the Riemann integrability of the function f on the interval $[a,b]$. Note first that each of the criteria requires f to be bounded so that we need not address that.

The Cauchy Criterion is equivalent to Riemann integrability Indeed the first criterion is exactly the same as the usual and familiar Cauchy criterion, which in this instance would more likely be written in this form:

For every $\varepsilon > 0$ there is a $\delta > 0$, with the property that

$$\left|\sum_{i=1}^{n} f(\xi_i)\lambda(([a_i,b_i]) - \sum_{j=1}^{m} f(\eta_j)\lambda(([c_j,d_j])\right| < \varepsilon \qquad (1.7)$$

whenever

$$\{([a_i,b_i],\xi_i) : i = 1,2,3,\ldots,n\}$$

and

$$\{([c_j,d_j],\eta_j) : j = 1,2,3,\ldots,m\}$$

are partitions of $[a,b]$ that are finer than δ.

Just check that the inequalities (1.6) and (1.7) are identical.

[18] A function is almost everywhere (a.e.) continuous if the set of points of discontinuity forms a set of measure zero.

Criterion M implies the Cauchy Criterion This is available simply from the triangle inequality. Just notice that

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{m}[f(\xi_i)-f(\eta_j)]\,\lambda(([a_i,b_i]\cap[c_j,d_j])\right|$$
$$\leq \sum_{i=1}^{n}\sum_{j=1}^{m}\left|f(\xi_i)-f(\eta_j)\right|\lambda(([a_i,b_i]\cap[c_j,d_j]).$$

Cauchy Criterion implies Riemann's Criterion This is easy. Take two subpartitions

$$\{([a_i,b_i],\xi_i):i=1,2,3,\dots,n\}$$

and

$$\{([a_i,b_i],\eta_i):i=1,2,3,\dots,n\}$$

and note the relation between

$$\sum_{i=1}^{n}[f(\xi_i)-f(\eta_j)]\,\lambda(([a_i,b_i])$$

and

$$\sum_{i=1}^{n}\omega f([a_i,b_i])(b_i-a_i).$$

In all cases

$$[f(\xi_i)-f(\eta_j)]\leq \omega f([a_i,b_i])$$

and the points ξ_i and η_j can be chosen so that the left-hand side is arbitrarily close to the right-hand side.

Riemann's Criterion implies Criterion M Assuming Riemann's Criterion we select points of $[a,b]$,

$$a=x_0<x_1<x_2<\dots<x_k=b$$

with the property that

$$\sum_{i=1}^{k}\omega f([x_{i-1},x_i])(x_i-x_{i-1})<\varepsilon/3.$$

Let M be an upper bound for $|f|$ and choose δ smaller than $\varepsilon(6Mk)^{-1}$ and also smaller than each of the lengths (x_i-x_{i-1}).

Now consider any collections

$$\{([a_i,b_i],\xi_i):i=1,2,3,\dots,n\}$$

and

$$\{([c_j,d_j],\eta_j):j=1,2,3,\dots,m\}$$

that are subpartitions of $[a,b]$ that are finer than δ.

An interval in one of these partitions is of type 1 (say) if it contains none of the points x_1, x_2, ..., x_{k-1} Note that if either $[a_i, b_i]$ or $[c_j, d_j]$ is not of type 1 then we can use a crude estimate

$$\left|f(\xi_i) - f(\eta_j)\right| \lambda(([a_i, b_i] \cap [c_j, d_j]) \leq 2M\delta.$$

There are fewer than $2k$ of these intervals and so we can control the sum over such intervals.

On the other hand if both of the intervals $[a_i, b_i]$ and $[c_j, d_j]$ are of type 1 then the interval $[a_i, b_i] \cap [c_j, d_j]$ is entirely inside one of the intervals $[x_{i-1}, x_i]$ of the original partition that we have chosen. We can control the contribution from these intervals by relating to the sum

$$\sum_{i=1}^{k} \omega f([x_{i-1}, x_i])(x_i - x_{i-1}).$$

Arguing using these ideas, we find that

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \left|f(\xi_i) - f(\eta_j)\right| \lambda(([a_i, b_i] \cap [c_j, d_j]) \leq 4kM\delta + \varepsilon/3 < \varepsilon.$$

This is the Criterion M.

Riemann's Criterion implies Lebesgue's Criterion We use the oscillation $\omega_f(x)$ of the function f at a points x. The set of points of discontinuity of f is exactly the set

$$E = \{x \in [a, b] : \omega_f(x) > 0\}.$$

let $\varepsilon > 0$. We shall choose a positive function $\delta : [a, b] \to \mathbb{R}^+$ so that

$$\sum_{i=1}^{n} \omega_f(\xi_i)\,[y_i - x_i] < \varepsilon$$

whenever a subpartition of $[a, b]$

$$\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$$

is given that is anchored in E and finer than δ. It will follow then that E is a set of measure zero as required.

Starting with Riemann's Criterion we select a partition of $[a, b]$

$$\{([a_i, b_i], \xi_i) : i = 1, 2, 3, \dots, n\}$$

with the property that

$$\sum_{i=1}^{n} \omega f([a_i, b_i])(b_i - a_i) < \varepsilon/3.$$

Here is how to choose our positive function $\delta : [a, b] \to \mathbb{R}^+$: we require only that the interval $[x - \delta(x), x + \delta(x)]$ intersect one (if possible) or at most two of the intervals in the collection

$$\{[a_i, b_i] : i = 1, 2, 3, \dots, n\}.$$

Note, for example, that if $a_1 < \xi < b_1$ then

$$\omega_f(\xi_1) \leq \omega f([a_1,b_1])$$

while if

$$a_1 < \xi = b_1 = a_2 < b_2$$

then

$$\omega_f(\xi) \leq \omega f([a_1,b_1]) + \omega f([a_2,b_2]).$$

It follows from these computations that

$$\sum_{i=1}^{n} \omega_f(\xi_i)\,[y_i - x_i] \leq 3\sum_{i=1}^{n} \omega f([a_i,b_i])(b_i - a_i) < \varepsilon$$

whenever a subpartition of $[a,b]$

$$\{([x_i,y_i],\xi_i) : i = 1,2,3,\dots\}$$

is anchored in E and finer than δ. Consequently E is a set of measure zero, as Lebesgue's Criterion demands.

Lebesgue's Criterion implies Riemann's Criterion For each x in the interval choose $\delta(x) > 0$ so that

$$\omega f([x-\delta(x), x+\delta(x)]) \leq \omega_f(x) + \varepsilon 2^{-1}(b-a)^{-1}.$$

The set of points of discontinuity of f is exactly the set

$$E = \{x \in [a,b] : \omega_f(x) > 0\}$$

and we are assuming, using Lebesgue's criterion, that the set E is a set of measure zero. Consequently there is a positive function $\delta_1(x)$ on $[a,b]$ smaller than $\delta(x)$ with the property that

$$\sum_{i=1}^{n} \omega_f(\xi_i)\,[y_i - x_i] < \varepsilon/2$$

whenever a subpartition of $[a,b]$

$$\{([x_i,y_i],\xi_i) : i = 1,2,3,\dots\}$$

is anchored in E and finer than δ.

Take any partition

$$\{([a_i,b_i],\xi_i) : i = 1,2,3,\dots,n\}$$

of the interval finer than δ_1. (Such a partition must exist by Cousin's lemma.) Then, remembering that $\omega_f(\xi_i) = 0$ for points ξ_i not in E, we would have

$$\sum_{i=1}^{n} \omega f([a_i,b_i])(b_i - a_i) \leq \sum_{i=1}^{n} \omega_f(\xi_i)\,[b_i - a_i] + \varepsilon/2 < \varepsilon.$$

This is Riemann's criterion for the partition we have just chosen.

Exercise 51 (Gillespie) *Let f be a function that is defined at every point of*

$[a,b]$. Then, f is said to be Cauchy integrable on $[a,b]$ if there is a number I so that, for every $\varepsilon > 0$ there is a $\delta > 0$, with the property that

$$\left| I - \sum_{i=1}^{n} f(x_{i-1})(x_i - x_{i-1}) \right| < \varepsilon$$

whenever points are given

$$a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n = b$$

for which $x_i - x_{i-1} < \delta$. Show that the Cauchy and Riemann integrals are equivalent.

Answer □

1.10.3 Volterra's Example

It is obvious that the Riemann integral does not include the classical Newton integral since there are unbounded derivatives. It was long thought, however, that Riemann's methods did suffice to construct the integral of any bounded derivative. In 1881 Vito Volterra constructed a bounded derivative on $[0,1]$ which is not Riemann integrable.

Since that time, a number of authors have constructed other such examples. Not all of these are easily accessible to the beginning student. A short note of Casper Goffman [32][19] provides a simple example of such a derivative f and uses only elementary techniques to show that f has the desired properties.

Essentially the construction is this: define an open set G that is dense in $[0,1]$ and is the union of a sequence of pairwise disjoint open intervals

$$G = \bigcup_{i=1}^{\infty} (a_i, b_i)$$

for which $\sum_{i=1}^{\infty}(b_i - a_i) = 1/2$. Then $K = [0,1] \setminus G$ is a closed set that cannot be of measure zero (as a simple argument will show). Define a differentiable function $F : [0,1] \to \mathbb{R}$ in such a way that $|F'(x)| \leq 1$ for all x, $F'(x) = 0$ for all $x \in K$ and $F'(x) = 1$ for at least one point x in every interval (a_i, b_i). By Lebesgue's criterion the function F' cannot be Riemann integrable because it is discontinuous at each point of K that is not isolated in K.

Thus the Riemann integral is a rather odd member of the integration family. It offers a simple constructive method for obtaining the integral of a large class of functions, but it does not fit well into the story of the Newton integral which, after all, is our main narrative here. A deeper study of Riemann's integral would reveal yet more flaws. (The next exercise is just one of these.)

[19] C. Goffman, A bounded derivative which is not Riemann integrable. Amer. Math. Monthly 84 (1977), no. 3, 205–206.

Exercise 52 *Let $f, g : [a,b] \to \mathbb{R}$ and suppose that $f(x) = g(x)$ for almost every point x in $[a,b]$. Show that if f and g are both Riemann integrable, then*

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

Show that, however, it is possible for f to be Riemann integrable without g being Riemann integrable. Answer □

1.11 Integral of Henstock and Kurzweil

Definition 1.20, defining the Riemann integral, in retrospect now appears as a serious mistake in the history of integration theory. The correct definition for a general integral on the real line is nearly identical but uses a positive function δ rather than the uniform version (with a constant δ) promoted by Riemann. Theorem 1.9 makes it clear that this better version is what is needed to capture the classical Newton integral and to generalize it. Since Riemann selected to generalize only the uniform version, his integral did not even include the Newton integral of bounded functions (although he would not have noticed this).

Definition 1.22 (Henstock-Kurzweil integral) *Let f be a function that is defined at every point of $[a,b]$. Then, f is said to be Henstock-Kurzweil integrable on $[a,b]$ if it satisfies the following "pointwise integrability" criterion: there is a number I so that, for every $\varepsilon > 0$ there is a positive function $\delta : [a,b] \to \mathbb{R}^+$, with the property that*

$$\left| I - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon$$

whenever points are given

$$a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n = b$$

for which

$$x_i - x_{i-1} < \delta(\xi_i)$$

with associated points $\xi_i \in [x_{i-1}, x_i]$.

It is immediate that the Henstock-Kurzweil integral includes both the Riemann integral and the classical Newton integral. As we see, it is easily motivated and easily presented. It is not clear at this stage how easy it is to develop a practical theory of integration from this definition. Indeed at this stage we should be concerned by the fact that Definition 1.22 (unlike Definition 1.20) is not constructive. Partitions finer than a positive function δ are rather more mysterious than partitions finer than a positive number δ.

Defined everywhere? Note that the function here is defined at *every* point of the interval $[a,b]$. We do not usually insist on this, permitting instead integrable functions to be defined only almost everywhere. The way to make this theorem accessible in general is assign arbitrary values to the function at points where it is undefined. We will prove later that this does not alter integrability nor change the integral in any way. A frequent convention, given a function f defined almost everywhere on an interval (a,b), is to work instead with the function g where we take $g(x) = f(x)$ when that exists and $g(x) = 0$ otherwise.

1.11.1 A Cauchy criterion

Integrability in this sense has a familiar Cauchy-type criterion. Here (and elsewhere) we use $\lambda(I)$ to denote the length of an interval I; if I is a degenerate interval or an empty interval then $\lambda(I)$ is considered to be zero.

Theorem 1.23 (Cauchy criterion) *A necessary and sufficient condition in order for a function $f : [a,b] \to \mathbb{R}$ to be Henstock-Kurzweil integrable on a compact interval $[a,b]$ is that, for all $\varepsilon > 0$, positive function $\delta : [a,b] \to \mathbb{R}^+$ can be found so that*

$$\left| \sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} [f(w) - f(w')]\lambda(I \cap I') \right| < \varepsilon \tag{1.8}$$

for all partitions π, π' of $[a,b]$ finer than δ.

Proof. Start by checking that when π and π' are both partitions of the same interval $[a,b]$ then, for any subinterval I of $[a,b]$

$$\lambda(I) = \sum_{(I',w')\in\pi'} \lambda(I \cap I')$$

from which it is easy to see that

$$\sum_{(I,w)\in\pi} f(w)\lambda(I) = \sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} f(w)\lambda(I \cap I').$$

This allows the difference that would normally appear in a Cauchy type criterion

$$\left| \sum_{(I,w)\in\pi} f(w)\lambda(I) - \sum_{(I',w')\in\pi'} f(w')\lambda(I') \right|$$

to assume the simple form given in (1.8). In particular that statement can be rewritten as

$$\left| \sum_{(I,w)\in\pi} f(w)\lambda(I) - \sum_{(I',w')\in\pi'} f(w)\lambda(I) \right| < \varepsilon. \tag{1.9}$$

∎

1.11.2 The Henstock-Saks Lemma

One of the most important, if elementary, tools of our theory is the following connection between a function and its indefinite integral. Saks was (perhaps) the first to exploit this in a study of integrals as Henstock pointed out later on when some authors referred to this as “Henstock’s Lemma.”

This can be viewed as the generalization of the classical Newton connection between a function f and its indefinite integral F where we insist that $F'(x) = f(x)$ everywhere. This condition is clearly much more general.

Theorem 1.24 (Henstock-Saks) *Suppose that f is a Henstock-Kurzweil integrable function defined at every point of a compact interval $[a,b]$. Then f is Henstock-Kurzweil integrable on every closed subinterval of $[a,b]$. Moreover, for every $\varepsilon > 0$, there exists a positive function $\delta : [a,b] \to \mathbb{R}^+$, so that*

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)| < \varepsilon$$

whenever π is a subpartition of the interval $[a,b]$ finer than δ, where here

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

is an indefinite integral for f.

The exercise asks for the converse direction.

Exercise 53 *Suppose that f, $F : [a,b] \to \mathbb{R}$ and that, for every $\varepsilon > 0$, there exists a positive function $\delta : [a,b] \to \mathbb{R}^+$, so that*

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)| < \varepsilon$$

whenever π is a subpartition of the interval $[a,b]$ finer than δ. Show that f is Henstock-Kurzweil integrable on $[a,b]$ and that

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

Answer □

Exercise 54 *Suppose that f, $F : [a,b] \to \mathbb{R}$ and that, for every $\varepsilon > 0$, there exists a positive number δ, so that*

$$\sum_{[u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)| < \varepsilon$$

whenever π is a subpartition of the interval $[a,b]$ finer than δ. What can you conclude? Answer □

1.11.3 Proof of Theorem 1.24

The first step is to show that if f is Henstock-Kurzweil integrable on $[a,b]$, then f is Henstock-Kurzweil integrable on every closed subinterval of $[a,b]$. Use the techniques we have already seen in the proofs of Theorems 1.9 and 1.10. Here are the details.

Henstock-Kurzweil integrability on subintervals.

> If $f : [a,b] \to \mathbb{R}$ is Henstock-Kurzweil integrable then it is also integrable on any compact subinterval of $[a,b]$.

Let $\varepsilon > 0$. Suppose that f is Henstock-Kurzweil integrable on $[a,b]$ and $[c,d]$ is a compact subinterval. Choose a positive function $\delta : [a,b] \to \mathbb{R}^+$, with the property that

$$\left| \int_a^b f(x)\,dx - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon$$

whenever points are given

$$a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n = b$$

for which

$$x_i - x_{i-1} < \delta(\xi_i)$$

with associated points $\xi_i \in [x_{i-1}, x_i]$.

Observe that for every pair of partitions π_1, and π_2 of the subinterval $[c,d]$ both of which are finer than δ, there is a subpartition π also finer than δ so that $\pi_1 \cup \pi$ and $\pi_1 \cup \pi$ are partitions of the full interval $[a,b]$. In particular then

$$\left| \sum_{(I,w)\in\pi_1} f(w)\lambda(I) - \sum_{(I,w)\in\pi_2} f(w)\lambda(I) \right| =$$

$$\left| \sum_{(I,w)\in\pi\cup\pi_1} f(w)\lambda(I) - \sum_{(I,w)\in\pi\cup\pi_2} f(w)\lambda(I) \right| < \varepsilon$$

The integrability of f on $[c,d]$ follows now from the Cauchy criterion.

The indefinite HK integral

> If $f : [a,b] \to \mathbb{R}$ is Henstock-Kurzweil integrable then there is a function $F : [a,b] \to \mathbb{R}$, called an indefinite integral for f, so that
>
> $$\int_c^d f(x)\,dx = F(d) - F(c)$$
>
> for every compact subinterval $[c,d]$ of $[a,b]$.

We have already supplied the existence of the integral on the subintervals. We simply verify that the function

$$F(t) = \int_a^t f(x)\,dx \quad (a \leq t \leq b)$$

will have this stated property.

To see this first check that if $a < c < d \leq b$ then

$$\int_a^c f(x)\,dx + \int_c^d f(x)\,dx = \int_a^d f(x)\,dx. \tag{1.10}$$

Consequently

$$\int_c^d f(x)\,dx = \int_a^d f(x)\,dx - \int_a^c f(x)\,dx = F(d) - F(c)$$

as we require. Thus the remainder of the proof for this statement requires proving the identity (1.10). We will leave this as an exercise to the reader to attempt this using the Cauchy criterion. [This also follows from Exercise 186 which is a related exercise for upper integrals from Chapter 3.].

The Henstock-Saks property

> If $F(x) = \int_a^x f(t)\,dt$ for all $a \le x \le b$ and $\varepsilon > 0$ then there exists a positive function $\delta : [a,b] \to \mathbb{R}^+$, so that
>
> $$\sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)| < \varepsilon \qquad (1.11)$$
>
> whenever π is a subpartition of the interval $[a,b]$ finer than δ.

Choose a positive function $\delta : [a,b] \to \mathbb{R}^+$, so that

$$\left| F(b) - F(a) - \sum_{([u,v],w)\in\pi} f(w)(v-u) \right| < \varepsilon/4 \qquad (1.12)$$

for every partition π of $[a,b]$ finer than δ. It will be our goal to establish (1.11). whenever π is a subpartition of the interval $[a,b]$ finer than δ.

Fix π and let $\pi' \subset \pi$ be any nonempty subset. We can apply the Cousin lemma on $[a,b]$ to find partitions of any compact subinterval finer than this δ. This gives us a useful way to supplement the subpartition π' so as to form a useful partition of $[a,b]$: we write

$$\pi \setminus \pi' = \{([u_1,v_1],w_1),([u_2,v_2],w_2),\ldots([u_k,v_k],w_k)\}.$$

Our hypothesis requires F to be an indefinite integral for f on each $[u_i,v_i]$ ($i = 1,2,\ldots,k$) and so, for each $i = 1,2,\ldots,k$, we are able to select a partition π_i of the interval $[u_i,v_i]$ that is finer than δ in such a way that

$$\left| F(v_i) - F(u_i) - \sum_{([u,v],w)\in\pi_i} f(w)(v-u) \right| < \varepsilon/(4k). \qquad (1.13)$$

Thus if we augment π' to form

$$\pi'' = \pi \cup \pi_1 \cup \pi_2 \cup \cdots \cup \pi_k$$

we obtain a partition of $[a,b]$ finer than δ and thus also satisfying an inequality of the form (1.12). Computing with these ideas, we see

$$\sum_{([u,v],x)\in\pi'} [F(v) - F(u)] = F(b) - F(a) - \sum_{i=1}^{k} [F(v_i) - F(u_i)]$$

and

$$\sum_{([u,v],w)\in\pi'} f(w)(v-u) = \sum_{([u,v],w)\in\pi''} f(w)(v-u) - \sum_{i=1}^{k}\left(\sum_{([u,v],w)\in\pi_i} f(w)(v-u)\right).$$

Putting these together with the estimates (1.12) and (1.13) we obtain

$$\left|\sum_{([u,v],x)\in\pi'} [[F(v)-F(u)]-f(x)(v-u)]\right| \le \left|F(b)-F(a)-\sum_{([u,v],x)\in\pi''} f(x)(v-u)\right|$$

$$+\sum_{i=1}^{k}\left|[F(v_i)-F(u_i)]-\sum_{([u,v],x)\in\pi_i} f(x)(v-u)\right| < \varepsilon/4+k(\varepsilon/(4k)=\varepsilon/2.$$

Let us emphasize what we now see: if π' is *any* subset of π we have obtained this inequality:

$$\left|\sum_{([u,v],w)\in\pi'} [F(v)-F(u)-f(x)(v-u)]\right| < \varepsilon/2.$$

To complete the proof let

$$\pi^+ = \{([u,v],w)\in\pi : F(v)-F(u)-f(w)(v-u)\ge 0\}$$

and

$$\pi^- = \{([u,v],w)\in\pi : F(v)-F(u)-f(w)(v-u)< 0\}.$$

Then

$$\sum_{([u,v],w)\in\pi^+} |F(v)-F(u)-f(w)(v-u)|$$

$$= \sum_{([u,v],w)\in\pi^+} [F(v)-F(u)-f(w)(v-u)] < \varepsilon/2$$

and

$$\sum_{([u,v],w)\in\pi^-} |F(v)-F(u)-f(w)(v-u)|$$

$$= \sum_{([u,v],w)\in\pi^-} -[F(v)-F(u)-f(w)(v-u)] < \varepsilon/2.$$

Adding the two inequalities proves (1.11).

1.11.4 The Henstock-Kurzweil integral includes all Newton integrals

The Henstock-Kurzweil integral includes all variants of the Newton integrals and in equivalent to the general version. We use this fact in Chapter 3 to justify embarking on a detailed study of the Henstock-Kurzweil integral. Although it is useful to have a wide variety of characterizations of the integral available, the use of Riemann sums often offers the easiest and most transparent route to proving some property that we might need.

Theorem 1.25 *The general Newton integral is equivalent in the Henstock-Kurzweil integral.*

Proof. Since the controlled Newton integral is equivalent to the general Newton integral we could use either in the proof. The former is rather easier and we shall it here.

Assume that the conditions of Definition 1.15 hold for a function $f : (a,b) \to \mathbb{R}$, a continuous function $F : [a,b] \to \mathbb{R}$, and a control function $\phi : (a,b) \to \mathbb{R}$. Fix an interval $[c,d] \subset (a,b)$. Let $\varepsilon > 0$ and write

$$\eta = \frac{\varepsilon}{\phi(d) - \phi(c)}.$$

For each $x \in (a,b)$, there is a $\delta(x) > 0$ so that

$$\left| \frac{F(y) - F(x) - f(x)(y-x)}{\phi(y) - \phi(x)} \right| < \eta$$

if $0 < |y - x| < \delta(x)$.

A simple Cousin argument verifies that f is Henstock-Kurzweil integrable on $[c,d]$ and that

$$\int_c^d f(x)\,dx = F(d) - F(c).$$

If $\pi = \{([u,v],w)\}$ is a partition of $[c,d]$ finer than δ, then

$$\left| F(d) - F(c) - \sum_{([u,v],w)\in\pi} f(w)(v-u) \right| \leq \sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)|$$

$$\leq \sum_{([u,v],w)\in\pi} \eta[\phi(v) - \phi(u)] = \eta[\phi(d) - \phi(c)] = \varepsilon.$$

That proves that f is Henstock-Kurzweil integrable on $[c,d]$ and that F is an indefinite integral. The function F is continuous on $[a,b]$. The extension to the full interval $[a,b]$ now follows with some further work (or else look ahead to the Cauchy property of Section 1.13 to find this among the standard properties of the Henstock-Kurzweil integral).

To prove the converse direction we show that the Henstock-Kurzweil integral is included in the controlled Newton integral. Suppose that $f : [a,b] \to \mathbb{R}$ is Henstock-Kurzweil integrable on $[a,b]$ with F as its indefinite integral. Certainly F is continuous. We need to construct an appropriate control in the sense of Definition 1.15.

Use the Henstock-Saks lemma (Theorem 1.24) to choose a decreasing sequence of positive functions $\delta_n : [a,b] \to \mathbb{R}^+$ so that

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)| < 2^{-n}$$

whenever π is a subpartition of the interval $[a,b]$ finer than δ_n.

First, for each integer $n = 1, 2, 3, \ldots$ define the function $G_n(x)$ at each point

$a < x < b$ by requiring $G_n(x)$ to be the supremum of the values

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)|$$

taken over all partitions π of $[a,x]$ finer than δ_n. Note that $G_n : [a,b] \to \mathbb{R}$ is nondecreasing, $G_n(a) = 0$ and $G_n(b) < 2^{-n}$.

We see that, for any integer n and all $k = 1,2,3,\dots,n$, if $0 < y-x < \delta_n(x)$ then $([x,y],x)$ is finer than δ_k. Thus

$$G_k(y) - G_k(x) \geq |F(y) - F(x) - f(x)(y-x)|.$$

Similarly if $0 < x-y < \delta_n(x)$ then $([y,x],x)$ is finer than δ_k and so

$$G_k(x) - G_k(y) \geq |F(x) - F(y) - f(x)(x-y)|.$$

We now define our control

$$\phi(x) = x + \sum_{k=1}^{\infty} G_k(x).$$

This is a finite-valued function, increasing on (a,b). Note that, if $0 < y-x < \delta_n(x)$ then

$$\phi(y) - \phi(x) \geq n\,|F(y) - F(x) - f(x)(y-x)|.$$

and if $0 < x-y < \delta_n(x)$ then

$$\phi(x) - \phi(y) \geq n\,|F(x) - F(y) - f(x)(x-y)|.$$

Consequently for each $x \in (a,b)$,

$$\lim_{y\to x} \frac{F(y) - F(x) - f(x)(y-x)}{\phi(y) - \phi(x)} = 0.$$

Thus ϕ is the required control verifying that the integral exists and with the same value as the controlled Newton integral. ■

Exercise 55 *Suppose that f, $F : [a,b] \to \mathbb{R}$ and that one of these four statements is true:*

1. *$F'(x) = f(x)$ at every point of $[a,b]$.*
2. *F is continuous on $[a,b]$ and $F'(x) = f(x)$ at every point of (a,b).*
3. *F is continuous on $[a,b]$, while $F'(x) = f(x)$ at every point of (a,b) with at most finitely many exceptions.*
4. *F is continuous on $[a,b]$, while $F'(x) = f(x)$ at every point of (a,b) excepting possibly countably many points.*
5. *F has zero variation on a set $N \subset [a,b]$ of measure zero, while $F'(x) = f(x)$ at every point of $[a,b]$ excepting possibly points in the set N.*

Give a direct proof (without appealing to Theorem 1.25) that f is Henstock-Kurzweil integrable on $[a,b]$ and

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

Answer □

1.12 Integral of Lebesgue

Historically and theoretically the Lebesgue integral on the real line is the most important, more useful and broadly studied certainly than the Henstock-Kurzweil integral. Most textbooks use the Riemann integral as a motivating tool leading to Lebesgue's theory. In fact, though, it was rather Newton's classical integral which motivated Lebesgue himself. The Riemann integral offers a constructive process that yields the Newton integral for some, but not all, derivatives. Lebesgue offered as motivation for his integral the problem of finding a constructive process that would provide the Newton integral of *all* derivatives. He succeeded in finding a process that worked for all *bounded* derivatives. We now know that there is no constructive process[20] possible that will yield the Newton integral for all unbounded derivatives.

The Lebesgue integral can be characterized in many various ways, some of them much simpler than Lebesgue's original thesis would suggest.

Definition 1.26 *A function $f : [a,b] \to \mathbb{R}$ is said to be Lebesgue integrable provided that both f and $|f|$ are Henstock-Kurzweil integrable.*

In the event that both f and $|f|$ are Henstock-Kurzweil integrable it is common (and suggestive) to say that f is absolutely integrable. We use the expression "Lebesgue integrable" instead for cultural reasons.

If $f : [a,b] \to \mathbb{R}$ is Lebesgue integrable then one can write

$$-\infty < -\int_a^b |f(x)|\,dx \le \int_a^b f(x)\,dx \le \int_a^b |f(x)|\,dx < \infty.$$

If $f,g : [a,b] \to \mathbb{R}$ are Lebesgue integrable then one can write

$$\left|\int_a^b f(x)\,dx - \int_a^b g(x)\,dx\right| \le \int_a^b |f(x) - g(x)|\,dx.$$

Both of these ideas are extremely useful and are not available for functions that are Henstock-Kurzweil integrable, but not absolutely (Lebesgue) integrable.

Measure theory? By defining the Lebesgue integral as a special case of the Henstock-Kurzweil integral we are placed in an interesting historical position:

[20] See Randall Dougherty and Alexander S. Kechris [26], The complexity of antidifferentiation. Adv. Math. 88 (1991), no. 2, 145–169. .

whereas Lebesgue had to develop measure theory first (and at great length) before presenting his integration theory, we have the integral already quickly defined, but now must go about (in our later chapters) the task of developing the necessary measure theory that will support the integration theory. The essential point to note here is that the procedure whereby the Lebesgue integral is defined as a special case of the Henstock-Kurzweil integral, is not intended to side-step measure theory. It just changes the order of the topics a bit. Measure theory remains a central aspect of the theory of integration on the real line.

1.13 The Cauchy property

The expression of an integral as

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

is sometimes usefully rewritten as

$$\int_a^b f(x)\,dx = F(b-) - F(a+)$$

by using the one-sided limits

$$F(a+) = \lim_{c\to a+} F(c) \quad \text{and} \quad F(b-) \lim_{d\to b-} F(d).$$

Since an indefinite integral in any of our senses is always continuous this does not introduce any new methods, although it proves to be suggestive of a new method.

Suppose that we are able to check integrability only on all intervals $[c,d] \subset (a,b)$, but not on $[a,b]$ itself. Does it follow that f would have to be integrable on all of $[a,b]$? For some integration methods the answer is yes, provided these one-sided limits $F(a+)$ and $F(b-)$ exist.

The reader will likely recognize this as a method taught in calculus classes under the unfortunate topic known as "improper integrals." For example many students have been taught that the only correct computation of the integral

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx$$

must involve these steps:

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx = \lim_{c\to 0+} \int_c^1 \frac{1}{\sqrt{x}}\,dx = \lim_{c\to 0+} \left[2\sqrt{1} - 2\sqrt{c}\right] = 2.$$

Indeed a student who merely used the function $F(x) = 2\sqrt{x}$ to write

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx = F(1) - F(0) = 2$$

would be severely chastised for her failure to mention that the integrand was unbounded and, hence, not integrable by the Riemann method. Both computations are, however, entirely correct if the integral is interpreted in a Newton sense.

We introduce the following definition, according to which the computation above is explained by the failure of the Riemann integral to possess the Cauchy property.

Definition 1.27 *An integration method is said to possess the* Cauchy property *if every function f defined on an interval $[a,b]$ is integrable provided*

1. *f is integrable by that method on all intervals $[c,d] \subset (a,b)$, and*

2. *the limits*

$$\lim_{c\to a+}\lim_{d\to b-}\int_c^d f(x)\,dx = \lim_{d\to b-}\lim_{c\to a+}\int_c^d f(x)\,dx$$

   both exist and are equal.

The method has the bounded Cauchy property *if this property holds for all bounded functions on any interval $[a,b]$.*

Only the following methods, from among the integration methods that we have studied, possess the Cauchy property:

1. The Newton integral (naive, utility, and general versions).

2. The Henstock-Kurzweil integral.

The following methods *do not* possess the Cauchy property:

1. The Newton integral (classical and elementary).

2. The regulated integral [Section 1.9].

3. The Dixmier teaching integral [Section 1.9].

4. The Riemann integral.

5. The Lebesgue integral.

Both the Riemann and Lebesgue integrals possess the bounded Cauchy property.

Exercise 56 *The example $f(x) = x^{-1/2}$ on $[0,1]$ shows that the Riemann integral does not possess the Cauchy property. Show that it does possess the bounded Cauchy property.* Answer □

Exercise 57 *Show that the elementary version of the Newton integral does not possess the Cauchy property.* Answer □

Exercise 58 *Show that the Lebesgue integral does not possess the Cauchy property. [Hint: use the function $F(x) = x^2 \sin x^{-2}$ on $[0,1]$.* Answer □

1.14 Lebesgue differentiation theorem

Our final application of the methods of this chapter is to prove a famous and useful theorem of Lebesgue asserting that functions of bounded variation are almost everywhere differentiable. We have already commented on the central role that functions of bounded variation must play in the study of absolutely integrable functions. Accordingly we shall very much need this in our study of the Lebesgue integral later on.

In Section 2.11 we shall return to this theorem and present a different (more subtle) proof based on fine covering [i.e., Vitali covering] arguments. We will also remove the hypothesis that F is continuous. The proof in this chapter is based on the Rising Sun Lemma of Riesz which is of interest for other reasons.

Theorem 1.28 (Lebesgue differentiation theorem) *Let $F : [a,b] \to \mathbb{R}$ be a continuous function of bounded variation. Then F is differentiable at almost every point in (a,b).*

Corollary 1.29 *Let $F : [a,b] \to \mathbb{R}$ be a continuous monotonic function. Then F is differentiable at almost every point in (a,b).*

Corollary 1.30 *Let $F : [a,b] \to \mathbb{R}$ be a Lipschitz function. Then F is differentiable at almost every point in (a,b).*

1.14.1 Bounded variation

If a function f is absolutely integrable in any one of our senses on an interval $[a,b]$ then observe the following estimate. Write

$$F(x) = \int_a^x f(t)\,dt \quad (a \leq x \leq b).$$

Take any number of nonoverlapping subintervals of $[a,b]$

$$[a_1,b_1],[a_2,b_2],[a_3,b_3],\dots,[a_N,b_N]$$

and check that

$$\sum_{i=1}^{N} |F(b_i) - F(a_i)| = \sum_{i=1}^{N} \left| \int_{a_i}^{b_i} f(t)\,dt \right| \leq \sum_{i=1}^{N} \int_{a_i}^{b_i} |f(t)|\,dt \leq \int_a^b |f(t)|\,dt.$$

This places an upper bound on sums of the form

$$\sum_{i=1}^{N} |F(b_i) - F(a_i)|.$$

Functions for which such sums always remain bounded are said to have *bounded variation* on $[a,b]$. The least upper bound is written as $\mathrm{Var}(F,[a,b])$ and called the *total variation* of F on $[a,b]$.

Indefinite integrals do not have to be of bounded variation. But if the function being integrated is absolutely integrable then the indefinite integral must be of bounded variation. In fact, as we have just observed,

$$\operatorname{Var}(F,[a,b]) \leq \int_a^b |f(t)|\,dt.$$

The differentiation properties of such functions thus becomes an important concern of integration theory.

Exercise 59 *Show that if F has bounded variation on an interval $[a,b]$ then F is necessarily bounded on $[a,b]$.* □

Exercise 60 *Show that if F has bounded variation on two adjacent intervals $[a,b]$ and $[b,c]$ then F has bounded variation on $[a,c]$. In fact, show that*

$$\operatorname{Var}(F,[a,c]) \leq \operatorname{Var}(F,[a,b]) + \operatorname{Var}(F,[b,c]).$$

□

Exercise 61 *Is the class of functions of bounded variation on an interval $[a,b]$ closed under linear combinations? Under products? Under quotients?* □

1.14.2 The Dini derivatives

The proof of Theorem 1.28 uses the four *Dini derivatives*. To analyze how a derivative $F'(x)$ may fail to exist we split that failure first into a right and left version and then into two further pieces, an upper and a lower.

The two right Dini derivatives are defined as

$$\overline{D}^+ F(u) = \inf_{\delta>0} \sup \left\{ \frac{F(v)-F(u)}{v-u} \,:\, 0 < v-u < \delta \right\}$$

and

$$\underline{D}^+ F(u) = \sup_{\delta>0} \inf \left\{ \frac{F(v)-F(u)}{v-u} \,:\, 0 < v-u < \delta \right\}$$

We will prove that, for almost every point x in (a,b),

$$\overline{D}^+ F(x) > -\infty, \quad \underline{D}^+ F(x) < \infty,$$

and

$$\overline{D}^+ F(x) = \underline{D}^+ F(x).$$

From these three assertions it follows that F has a finite right-hand derivative $D^+ F(x)$ at almost every point x in (a,b). The same arguments would apply to the left-hand Dini derivatives and so we can conclude that F has both a right-hand and a left-hand derivative almost everywhere. Since the points where a right-hand and a left-hand derivative can differ form a set that is countable [see Exercise 62] it would follow that F is differentiable a.e. as stated.

Exercise 62 (Beppo-Levi) *Suppose that a function $F : [a,b] \to \mathbb{R}$ has both a right-hand and a left-hand derivative at every point of a nonempty set $E \subset (a,b)$ and that $D^+F(x) \neq D^-F(x)$ for each $x \in E$. Show that E is countable. [cf. Exercise 344.]*

Answer □

1.14.3 Two easy lemmas

The proof employs the Rising Sun lemma as well as an elementary geometric lemma that Donald Austin[21] used in 1965 to give a simple proof of this theorem. Our proof of the differentiation theorem exploits some of the computations in his proof, but written in different language and employing the Rising Sun lemma for some of the work.

The Rising Sun lemma This lemma is known as the "Rising Sun" or "Setting Sun" Lemma, depending on whether you wish the sun to rise in the east (on the left) or set in the west (on the right). It is due to Frédéric Riesz [71] and, while simple, is justly famous.

Lemma 1.31 (Riesz) *Let $H : [a,b] \to \mathbb{R}$ be a continuous function and define a point in (a,b) to be* shaded *if there is a point $y > x$ in the interval for which $H(x) < H(y)$. Then the set of shaded points is open. If $\{(a_k, b_k)\}$ is the sequence of component intervals of that set then $H(a_k) \leq H(b_k)$ for each k.*

Image the sun placed on the x-axis far to the right of the graph of H considered as a hilly landscape. A point that is shaded (according to the statement in the lemma) is indeed in the shade of some part of the graph to the right. See Figure 1.4 for an illustration from the Wolfram website. The images on the web site are interactive and a good assist to the intuition.

The proof is easy and should be attempted by all novices at least once. Note especially two facts.

1. If $a < x < b$ and $\overline{D}^+H(x) > 0$ then x is a shaded point for H.

2. If $G : [a,b] \to \mathbb{R}$ is continuous then all points x at which $\overline{D}^+G(x) > \alpha$ are contained in some collection of pairwise disjoint open intervals $\{(a_k, b_k)\}$ for which
$$G(b_k) - G(a_k) \leq \alpha(b_k - a_k) \tag{1.14}$$
for each k. [Apply the lemma to $H(x) = F(x) - \alpha x$.]

Exercise 63 *Prove Lemma 1.31.* Answer □

[21] D. Austin, *A geometric proof of the Lebesgue differentiation theorem.* Proc. Amer. Math. Soc. 16 (1965) 220–221.

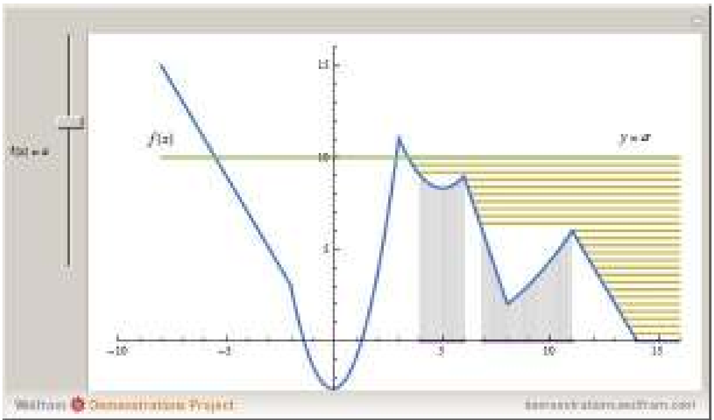

Figure 1.4: Riesz's rising sun lemma.

Austin's lemma To exploit an inequality such as (1.14) that comes from the Rising Sun lemma in our proof we need some easy estimates. The proof is left as an exercise.

Lemma 1.32 (Austin) *Let $G : [a,b] \to \mathbb{R}$, and suppose that $\{(a_k, b_k)\}$ is a finite or infinite sequence of pairwise disjoint open subintervals of $[a,b]$.*

1. *If $G(a) \le G(b)$, then*
$$-\sum_{k\ge1} (G(b_k) - G(a_k)) \le \mathrm{Var}(G,[a,b]) - |G(b) - G(a)|.$$
2. *If $G(b) \le G(a)$, then*
$$\sum_{k\ge1} (G(b_k) - G(a_k)) \le \mathrm{Var}(G,[a,b]) - |G(b) - G(a)|.$$

Exercise 64 *Prove Lemma 1.32.* Answer □

1.14.4 Proof of the Lebesgue differentiation theorem

We now prove the theorem. Recall that we are focusing just on the right-hand derivative, since that is all that we need to establish.

Step 1. The strategy The first step in the proof is to show that at almost every point t in (a,b),
$$\underline{D}^+F(t) = \overline{D}^+F(t).$$
If this is not true then there must exist a pair of rational numbers r and s for which the set
$$E_{rs} = \{t \in (a,b) : \underline{D}^+F(t) < r < s < \overline{D}^+F(t)\}$$

is not a set of measure zero. This is because the union of the countable collection of sets E_{rs} contains all points t for which $\underline{D}^+F(t) \neq \overline{D}^+F(t)$.

Let us show that each such set E_{rs} is a set of measure zero. Write $\alpha = (s-r)/2$, $B = (r+s)/2$, $G(t) = F(t) - Bt$. Note that

$$E_{rs} = \{t \in (a,b) : \underline{D}^+G(t) < -\alpha < 0 < \alpha < \overline{D}^+G(t)\}.$$

This will make our computations easier to visualize.

Step 2. Use the variation Since F is a continuous function of bounded variation on $[a,b]$, so too is the function G. In fact

$$\mathrm{Var}(G,[a,b]) \leq \mathrm{Var}(F[a,b]) + B(b-a).$$

Let $\varepsilon > 0$ and select points

$$a = s_0 < s_1 < \cdots < s_{n-1} < s_n = b$$

so that

$$\sum_{i=1}^{n} |G(s_i) - G(s_{i-1})| > \mathrm{Var}(G,[a,b]) - \alpha\varepsilon.$$

Let $E'_{rs} = E_{rs} \setminus \{s_1, s_2, \ldots, s_{n-1}\}$. Let us call an interval $[s_{i-1}, s_i]$ *black* if $G(s_i) - G(s_{i-1}) \geq 0$ and call it *red* if $G(s_i) - G(s_{i-1}) < 0$.

Step 3. Define the function δ For each $i = 1,2,3,\ldots,n$ we define a function δ_i as follows. Suppose that $[s_{i-1}, s_i]$ is a black interval and x is any point in $(s_{i-1}, s_i) \cap E_{rs}$. Note that x is a shaded point for the function $H(x) = G(x) + \alpha x$ on $[s_{i-1}, s_i]$. We can use the Rising Sun lemma to select a sequence of open intervals $\{(a_{ij}, b_{ij})\}$ of $[s_{i-1}, s_i]$ containing all these shaded points. In particular we have arranged that each

$$G(b_{ij}) - G(a_{ij}) \leq -\alpha(b_{ij} - a_{ij}).$$

Moreover, using Lemma 1.32 we see that we will have the inequality

$$\sum_{j\geq 1} \alpha(b_{ij} - a_{ij}) \leq -\sum_{j\geq 1} (G(b_{ij}) - G(a_{ij})) \leq \mathrm{Var}(G,[s_{i-1}, s_i]) - |G(s_i - G(s_{i-1})|.$$

If, instead, $[s_{i-1}, s_i]$ is a red interval and x is any point in $(s_{i-1}, s_i) \cap E_{rs}$, we note that x is a shaded point for the function $H(x) = -G(x) + \alpha x$ on $[s_{i-1}, s_i]$. We can again apply the Rising Sun lemma to select a sequence of open intervals $\{(a_{ij}, b_{ij})\}$ of $[s_{i-1}, s_i]$ containing all these shaded points. In particular we have arranged that each

$$G(b_{ij}) - G(a_{ij}) \geq \alpha(b_{ij} - a_{ij}).$$

Moreover, again using Lemma 1.32 we see that we will have the inequality

$$\sum_{j\geq 1} \alpha(b_{ij} - a_{ij}) \leq \sum_{j\geq 1} (G(b_{ij}) - G(a_{ij})) \leq \mathrm{Var}(G,[s_{i-1}, s_i]) - |G(s_i - G(s_{i-1})|.$$

We simply choose $\delta_i(x) > 0$ so that $(x - \delta_i(x), x + \delta_i(x))$ is a subinterval of that interval (a_{ij}, b_{ij}) to which x belongs.

Finally define δ to assume the value $\delta_i(x)$ whenever $x \in (s_{i-1}, s_i) \cap E_{rs}$. Then δ is defined and positive on all of E'_{rs}.

Step 4. Estimate the Riemann sum Let π be any subpartition anchored in E'_{rs} and finer than δ. Our estimate will verify that E'_{rs} is a set of measure zero according to Definition 1.11.

Write π_i for the subset of the subpartition π all of whose intervals are subintervals of $[s_{i-1}, s_i]$. Note that if $([u,v],w) \in \pi_i$ then $[u,v]$ is necessarily contained in one of the open intervals (a_{ij}, b_{ij}). By our estimates in Step 2 we see that that

$$\alpha\left(\sum_{([u,v],w)\in\pi_i}(v-u)\right) \leq \mathrm{Var}(G,[s_{i-1},s_i]) - |G(s_i) - G(s_{i-1})|.$$

Consequently

$$\alpha\left(\sum_{([u,v],w)\in\pi}(v-u)\right) = \alpha\left(\sum_{i=1}^{n}\sum_{([u,v],w)\in\pi_i}(v-u)\right)$$
$$\leq \sum_{i=1}^{n}\mathrm{Var}(G,[s_{i-1},s_i]) - \sum_{i=1}^{n}|G(s_i - G(s_{i-1})|$$
$$\leq \mathrm{Var}(G,[a,b]) - \sum_{i=1}^{n}|G(s_i - G(s_{i-1})| < \alpha\varepsilon.$$

We have proved that E'_{rs} is a set of measure zero. So too then is E_{rs} since the two sets differ by only a finite number of points.

Step 5. Show the Dini derivatives are finite a.e. We know now that the function F has a right-hand derivative, finite or infinite, almost everywhere in (a,b). We wish to exclude the possibility of the infinite derivative, except on a set of measure zero.

Let

$$E_\infty = \{t \in (a,b) : \underline{D}^+F(t) = \infty\}.$$

Choose any B so that $F(b) - F(a) \leq B(b-a)$ and set $G(t) = F(t) - Bt$. Note that $G(b) \leq G(a)$ which will allow us to apply the second estimate in Lemma 1.32

Let $\varepsilon > 0$ and choose a positive number α large enough so that

$$\mathrm{Var}(G,[a,b]) - |G(b) - G(a)| < \alpha\varepsilon.$$

Note that every point $x \in E_\infty$ is a shaded point for the function $H(x) = G(x) - \alpha x$ on $[a,b]$.

We use the Rising Sun lemma to select a sequence of open subintervals $\{(a_i, b_i)\}$ of $[a,b]$ containing all these shaded points. In particular we have that each

$$G(b_i) - G(a_i) \geq \alpha(b_i - a_i).$$

We choose $\delta(x) > 0$ so that $(x - \delta(x), x + \delta(x))$ is a subinterval of that interval (a_i, b_i) to which x belongs.

Let π be any subpartition anchored in E_∞ and finer than δ. Our estimate will verify that E_∞ is a set of measure zero. Using Lemma 1.32 we have

$$\alpha \sum_{([u,v],w)\in\pi} (v - u) \leq \mathrm{Var}(G, [a,b]) - |G(b) - G(a)| < \alpha\varepsilon.$$

We have proved that

$$\sum_{([u,v],w)\in\pi} (v - u) < \varepsilon$$

for every such subpartition π. It follows that E_∞ is a set of measure zero. The same arguments will handle the set

$$E_{-\infty} = \{t \in (a,b) : \overline{D}^+ F(t) = -\infty\}.$$

1.14.5 Removing the continuity hypothesis

The earliest proofs of the Lebesgue differentiation theorem (Theorem 1.28) demanded that the function must be continuous. This is not necessary, although any direct use of the Rising Sun Lemma appears to demand continuity. There are a number of ways of proving the full version, valid for all functions of bounded variation, continuous or not. Claude-Alain Faure [28] gives an elementary proof that employs the Rising Sun lemma throughout. Botsko [7] also is quite readable. Our proof later on in Section 2.11 gives a proof based on a Vitali covering argument.

A simple way of extending what we already have is to consider taking an "inverse" as Faure [28] has done. To illustrate the technique we prove this simple version. We give the argument for monotonic functions. (The Jordan decomposition theorem, proved for us later on in Chapter 5), would allow us to deduce the same property of functions of bounded variation.)

Theorem 1.33 *Let $F : [a,b] \to \mathbb{R}$ be a nondecreasing function. Then F is differentiable at almost every point of (a,b).*

Proof. We shall replace F by the function $F_1(t) = F(t) + t$ and argue for F_1. If we can show that F_1 is differentiable a.e. in (a,b) then so too is the function F.

Note that $F_1 : [a,b] \to \mathbb{R}$ is a strictly increasing function. Then the function $G : [F_1((a), F_1((b)] \to \mathbb{R}$ defined by

$$G(y) = \inf\{z \in [a,b] : F_1((z) \geq y\}$$

will serve as a continuous, strictly increasing left-inverse function for F_1, i.e., $G(F_1(t)) = t$ for all t. We use the following feature of this inverse situation: if

$N \subset (a,b)$, then $F_1(N) \subset (F_1((a),F_1((b))$, and if the the function G has zero variation on $F_1(N)$ then necessarily N is a set of measure zero. (All of this is the content of Exercise 28.)

Note that, if $F_1(a) \leq x < y \leq F_1(b)$, with $F_1(u) = x$, $F_1(v) = y$ then

$$F_1(v) - F_1(u) = F(v) - F(u) + (v-u) \geq v-u > 0$$

implies that, for all $F_1(a) \leq x < y \leq F_1(b)$,

$$y - x \geq G(y) - G(x) > 0. \tag{1.15}$$

Any pair $([u,v],w)$ corresponds to the pair $([x,y],z)$ with $F_1(u) = x$, $F_1(v) = y$, and $F_1(w) = z$. This reveals the relation

$$\frac{F_1(v) - F_1(u)}{v-u} = \left(\frac{G(y) - G(x)}{y-x}\right)^{-1},$$

from which it follows that F_1 is differentiable at any point w if G has a finite nonzero derivative at z.

Using these observations we can apply the Lebesgue differentiation theorem to the continuous, increasing function G to deduce that G has a derivative at every point of $[F_1((a),F_1((b)]$ except at the points of a set of measure zero. Let M_1 be the set of of points in $[F(a),F(b)]$ where G does not have a derivative. Let M_2 be the set of points where G has a zero derivative. The set M_1 must be of measure zero. The inequality (1.15) shows that G has zero variation on M_1. But we know that G also has zero variation on the set of points where the derivative vanishes, thus G has zero variation on M_2 as well.

So G has zero variation on $M = M_1 \cup M_2$. Let N be the set of points where F_1 does not have a derivative. Since $F_1(N) \subset M$ then N is a set of measure zero. ∎

1.15 Infinite integrals

For many applications, some familiar to calculus students, one needs an integral for a function $f : \mathbb{R} \to \mathbb{R}$ that would be written in the form

$$\int_{-\infty}^{\infty} f(x)\,dx.$$

The theory so far addresses only the case of integration on a compact interval $[a,b]$ and does not generalize without effort to the case of an unbounded interval.

The Newton integral on $(-\infty,\infty)$. For the various Newton integrals the simplest approach conceptually is merely to replace the usual definition

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

with a statement such as

$$\int_{-\infty}^{\infty} f(x)\,dx = F(\infty) - F(-\infty).$$

Here, we interpret

$$F(\infty) = \lim_{b\to\infty} F(b) \quad \text{and} \quad F(-\infty) = \lim_{a\to-\infty} F(a). \tag{1.16}$$

This pattern would be successful. For example, one could take $F:\mathbb{R}\to\mathbb{R}$ as any continuous function for which $F'(x) = f(x)$ for all x with at most countably many exceptions, provided the two limits in (1.16) can be proved to exist.

This is essentially the method used in elementary calculus. Because there are two further limits imposed on the integral, some properties of finite integrals easily extend to infinite integrals, while some properties do not. Exercise 65 illustrates a property that does not survive the extra limits.

The Lebesgue integral on $(-\infty,\infty)$. For the Lebesgue integral an entirely different approach is taken. Lebesgue integrals invariably split the function into its positive and negative parts

$$f(x) = [f(x)]^+ - [f(x))]^-$$

and require

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} [f(x)]^+\,dx - \int_{-\infty}^{\infty} [f(x)]^-\,dx$$

and

$$\int_{-\infty}^{\infty} |f(x)|\,dx = \int_{-\infty}^{\infty} [f(x)]^+\,dx + \int_{-\infty}^{\infty} [f(x)]^-\,dx.$$

Thus the procedure above taking the two limits in (1.16), while sometimes useful in computations, is not used to define the integral nor is it always valid. (Exercise 66 illustrates.)

Integration theory on $(-\infty,\infty)$? In this text, as it currently stands, we do not have much more to say about the general problem of integration on infinite intervals. A full account of what properties are available for infinite integrals might be useful in a course of instruction, but it offers too much extra technical detail to a course that is already highly detailed.

Some authors take the Henstock-Kurzweil integral and develop it also in a way that allows “infinite” partitions, i.e., partitions of $(-\infty,\infty)$. In that way integrals on compact intervals have a theory that is similar to the theory on infinite intervals and a certain unity of approach results. It seems that developing that here would try the patience of the reader beyond what we are already doing.

Our only advice is to learn first the calculus theory as it is usually taught. After that, study the measure-theoretic integral and the theory of that integral. The techniques are different and the resulting theories differ. By then one is prepared for any application.

Exercise 65 *Give an example of a sequence of functions $f_n : \mathbb{R} \to \mathbb{R}$ that converges uniformly to zero on $(-\infty, \infty)$ and yet*

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\, dx \neq 0.$$

Answer □

Exercise 66 *Comment on the statements*

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}\, dx = \pi \quad \textit{and} \quad \int_{-\infty}^{\infty} \left| \frac{\sin x}{x} \right| dx = \infty.$$

Answer □

1.16 Where are we?

Rather than there being a single theory of integration, we have seen an abundance of integrals. The exact relation among all of these integrals is not so easy to spot. Take any pair of integrals (e.g., the elementary Newton integral and the regulated integral). What is the relationship? Do they have differing properties? A specialist in integration theory has no trouble answering such questions. Even many a graduate student of mathematics might, however, get lost in the details.

There are two themes in this chapter. The first theme is the Newton integral itself. The severely classical version that demands $F'(x) = f(x)$ everywhere is too restrictive. The version that we need eventually for serious applications in analysis allows exceptions forming a set of measure zero. This is the integral we study.

The other theme is the theme of Riemann sums. All of the concepts in integration theory on the real line permit a realization as Riemann sums. We saw that in Cauchy's theorem and in Robbins's theorem. Riemann's integral itself employed that notion but did not take the ideas nearly far enough. Riemann sums have led us to a characterization of the general Newton integral (due to Henstock and Kurzweil) as well as to the notions of measure zero sets and functions having zero variation on a set. We have a significant tool in the use of Riemann sums, provided we exploit that properly. This will lead us eventually to one of the most important of the tools of integration theory—measure theory.

We know what integral we want to study. We want to study the integral that includes all of the others. This leaves the inadequate Riemann integral of our calculus class far behind. It also leaves the Lebesgue integral behind for the moment. We study the Henstock-Kurzweil integral, but taking care to develop the Lebesgue theory along with it. Probably the most important aspect of the theory discussed here is the abundance of new tools that this theory supplies to the study of the Lebesgue integral itself. We do not argue that the Henstock-Kurzweil integral should "replace" the Lebesgue integral—it is merely an interesting and useful supplement to the Lebesgue theory.

Henstock famously (infamously?) announced at the 1962 ICM meeting in Stockholm that “the Lebesgue integral is dead.” One can agree with him, at least, that it is not necessary to develop all of the machinery of measure theory in advance in order to define and develop an adequate theory of integration. The Lebesgue integral, of course, survives as the most important special case of Henstock’s theory. Lebesgue’s methods and the measure theory on which he based his integral remain among the most important aspects of any theory of integration.

1.17 Appendix: Constructive vs. descriptive

We end this chapter with some rather loose and elementary comments on the role of constructive and descriptive definitions in integration theory. This language is used frequently in the literature and is not meant in any highly technical sense. A search for the use of the phrase “descriptive definition” will reveal that it is used freely in integration papers, and rarely in any other context.

Here is an analogy which might help to illustrate the ideas. Suppose that one wishes to introduce from first principles the concept $\sqrt{x}$ (the nonnegative square root). A possible (if clumsy) start might be the following *descriptive* definition:

> A real number x is said to be rootable if there is a real number y for which $y^2 = x$. In that case we write $\sqrt{x} = |y|$.

The definition requires a justification: if $y_1^2 = y_2^2 = x$ then $|y_1| = |y_2|$. Having checked that by elementary algebra, we now have a servicable concept that can be explored. It is completely descriptive (we know nothing as to whether such objects exist, nor would be able to compute them). We can observe a few examples. For example 9 is rootable and $\sqrt{9} = 3$ as can be checked using the definition.

Even so a theory of roots can emerge. Perhaps we would prove the following:

1. If a real number x is rootable then $x \geq 0$.
2. If a real number x is rootable then $(\sqrt{x})^2 = x$.
3. If real numbers x and y are rootable then so too is the product xy and $\sqrt{x}\sqrt{y} = \sqrt{xy}$.
4. etc.
5. A real number x is rootable if and only if $x \geq 0$.
6. If $x \geq 0$ then $\sqrt{x} = \lim_{n\to\infty} x_n$ where $x_1 = 1$ and $x_{n+1} = \frac{1}{2}\left(x_n + \frac{x}{x_n}\right)$.

Items 5 and 6 are the sophisticated steps needed in the exposition of this concept. The latter is known as the Babylonian method or as Heron’s method, named after the first-century Greek mathematician Hero of Alexandria. Calculus students might first study it as an instance of Newton's rule. Thus a descriptive definition can evolve into a full theory of the concept after some work.

How would a constructive definition work? Well start at the end and work backwards.

> For any nonnegative real number x define $\sqrt{x}$ to mean $\sqrt{x} = \lim_{n\to\infty} x_n$ where $x_1 = 1$ and $x_{n+1} = \frac{1}{2}\left(x_n + \frac{x}{x_n}\right)$.

This is completely constructive (although we would be required to prove that the limit exists in order to justify the definition). Now we are obliged to develop the theory with this as a starting point. For example, try to prove statements 2 and 3 directly from the limit definition!

It is a matter of taste, convenience, and pedagogy which of these two routes one might wish to follow. The theory ends up the same.

For integration theory it is somewhat analogous. The definitions of Riemann and Lebesgue are constructive. Descriptive definitions could be given instead and the theory will develop in a different order, but end up in the same place. Neither of the methods of Riemann or Lebesgue are intuitive, however, and one could argue that they belong at the end of the study (just as Heron's method is better appearing at the end in the discussion above, rather than at the start as a definition).

The definitions we have given for variants of the Newton integral are all descriptive. A case can be made that the most natural development of integration theory starts with these variants and develops the constructive theory as a final step. This is especially true for the most general Newton integral since the constructive theory needed requires (as Denjoy illustrated) a countable transfinite process. Even the Lebesgue integral, one can argue, benefits by a descriptive start rather than taking the lengthy process of Lebesgue's development of measure theory as the working definition.

Chapter 2

Covering Theorems

We embark now on a complete theory for the integral on the real line. In Chapter 1 we studied integration theory following mostly the single theme that

$$\int_a^b F'(x)\,dx = F(b) - F(a),$$

requiring that $F : [a,b] \to \mathbb{R}$ is differentiable at all or at most points.

The connection of these integration ideas with Riemann sums has been explored. We now examine this more fully and purse the elements of the theory of the Henstock-Kurzweil integral. We will start by exploring Lebesgue measure zero sets in greater detail. In particular we prove the Mini-Vitali covering theorem that characterizes Lebesgue measure zero sets in terms of full and fine covers.

Here is our goal for both the review and the new material that will be introduced in this chapter:

- Covering relations.
- Riemann sums.
- Lebesgue measure zero sets.
- Full null and fine null sets.
- Mini-Vitali theorem asserting the equivalence of Lebesgue measure zero, full null, and fine null.
- Zero variation and its relation to zero derivative.
- Absolute continuity (several versions).
- A second proof of the equivalence of the general Newton and Henstock-Kurzweil integrals.
- A second proof of the Lebesgue differentiation theorem, this time using the mini-vitali theorem.

2.1 Covering Relations

The language of integration theory and many of our most important techniques, as presented in the next few chapters, depend on an understanding of and facility with partitions and Riemann sums. A partition is a special case of a covering relation. This section defines and reviews all of the terminology and examines all of the techniques needed to carry on to a complete investigation of the integral.

2.1.1 Partitions and subpartitions

Construct a subdivision of a compact interval $[a,b]$ by choosing points

$$a = a_0 < a_1 < a_2 < \cdots < a_{k-1} < a_k = b$$

and then select points $\xi_1, \xi_2, \ldots, \xi_k$ so that each point ξ_i belongs to the corresponding interval $[a_{i-1}, a_i]$. Then the collection

$$\pi = \{([a_{i-1}, a_i], \xi_i) : i = 1, 2, \ldots, k\}$$

is called a *partition* of $[a,b]$. Note that the intervals do not overlap and that their union is the whole of the interval $[a,b]$. The associated points must be selected from their corresponding interval. Any subset of a partition is called a *subpartition*.

We consider this a special kind of *covering relation*.

2.1.2 Covering relations

Families of pairs $([u,v],w)$, where $[u,v]$ is a compact interval and w a point in that interval, are called *covering relations*. Every partition and every subpartition is a covering relation. It is a *relation* because it provides an association of points with intervals.

All covering relations are just subsets of one big covering relation:

$$\hat{\beta} = \{([u,v],w) : u,v,w \in \mathbb{R},\ u < v \text{ and } u \le w \le v\}.$$

We shall most frequently use the Greek symbol β to denote a covering relation. We have already used the Greek symbol π to denote those covering relations which are partitions or subpartitions.

2.1.3 Prunings

Given a number of covering relations arising in a problem we often have to combine them or “prune out” certain subsets of them. We use the following techniques quite frequently:

Definition 2.1 *If β is a covering relation and E a set of real numbers then we write:*

- $\beta[E] = \{([u,v],w) \in \beta : w \in E\}$.
- $\beta(E) = \{([u,v],w) \in \beta : [u,v] \subset E\}$.

to indicate these subsets of the covering relation β from which we have removed inconvenient members.

Note, for example, that if π is a partition of an interval $[a,b]$ and E is a subset of that interval then $\pi[E]$ selects just those elements of the partition whose associated points belong to the set E.

2.1.4 Full covers

A full cover is one that, in very loose language, contains all sufficiently small intervals at a point.

Definition 2.2 *Let E be a set of real numbers. A covering relation β is said to be a* full cover *of E if for each $w \in E$ there is a positive number $\delta(w)$ so that β contains every pair $([u,v],w)$ for which $v-u < \delta(w)$.*

By a *full cover* without reference to any set we mean a full cover of all of $\mathbb{R}$.

Full covers arise naturally as ways to describe continuity, differentiation, integration, and numerous other processes of analysis. The student should attempt many of the exercises in order to gain a facility in covering arguments.

2.1.5 Fine covers

A fine cover[1] is one that, in very loose language, contains arbitrarily small intervals at a point.

Definition 2.3 *Let E be a set of real numbers. A covering relation β is said to be a* fine cover *of E if for each $w \in E$ and any positive number ε the covering relation β contains at least one pair $([u,v],w)$ for which $v-u < \varepsilon$.*

By a *fine cover* without reference to any set we mean a fine cover of all of $\mathbb{R}$.

Fine covers arise in the same way that full covers arise. In a sense the fine cover comes from a negation of a full cover. For example (as you will see in the Exercises) full covers could be used to describe continuity conditions and fine covers would then twist this to describe the situation where continuity fails.

This can be described as a duality. Fine covers are the duals of full covers and full covers are the duals of fine covers. Exercises 82 and 81 make this explicit.

[1] Known also as a *Vitali cover*.

2.1.6 Uniformly full covers

A uniformly full cover is one that, in very loose language, contains all sufficiently small intervals at a point, where the smallness required is considered the same for all points.

Definition 2.4 *Let E be a set of real numbers. A covering relation β is said to be a* uniformly full cover *of E if there is a positive number δ so that β contains every pair $([u,v],w)$ for which $v-u<\delta$.x*

Only occasionally shall we use uniformly full covers. To verify that a covering relation is full just requires us to test what happens at each point. To verify that a covering relation is *uniformly* full requires more: we have to find a positive number δ that works at every point. The exclusive use of uniformly full covers would lead to a restrictive theory: the Riemann integral (which plays only a minor role in this textbook) is based on uniformly full covers. Our integration theory uses full covers and, as a consequence, is much more general and is easier.[2]

Exercises

Exercise 67 *Suppose that G is an open set. Show that*

$$\beta = \{([u,v],w) : u \leq w \leq v, \ \ [u,v] \subset G\}$$

is a full cover of G. □

Exercise 68 *Suppose that β is a full cover of a set E and that G is an open set containing E. Show that $\beta(G)$ is also a full cover of E. [This is described as "pruning the full cover" by the open set G.]* Answer □

Exercise 69 *Suppose that β is a fine cover of a set E and that G is an open set containing E. Show that $\beta(G)$ is also a fine cover of E. [This is described as "pruning the fine cover" by the open set G.]* Answer □

Exercise 70 *Suppose that β is a uniformly full cover of a set E and that G is an open set containing E. Show that $\beta(G)$ is not necessarily a uniformly full cover of E. Would it be a full cover?* □

Exercise 71 *Suppose that β_1 and β_2 are both full covers of a set E. Show that $\beta_1 \cap \beta_2$ is also a full cover of E.* □

Exercise 72 *Suppose that β_1 and β_2 are both fine covers of a set E. Show that $\beta_1 \cap \beta_2$ need not be a fine cover of E.* □

---

[2] It is easier since the requirement in Riemann integration to always check that the covers used are not merely full, but uniformly full, imposes unnecessary burdens on many proofs.

Exercise 73 *Suppose that β_1 is a full cover of a set E and β_2 is a fine cover. Show that $\beta_1 \cap \beta_2$ is also a fine cover of E. Need it be a full cover?* □

Exercise 74 *Suppose that β_1 and β_2 are full covers of sets E_1 and E_2 respectively. Show that $\beta_1 \cup \beta_2$ is a full cover of $E_1 \cup E_2$.* □

Exercise 75 *Suppose that β_1 and β_2 are fine covers of sets E_1 and E_2 respectively. Show that $\beta_1 \cup \beta_2$ is a fine cover of $E_1 \cup E_2$.* □

Exercise 76 *Let $E_1, E_2, E_3, \dots$ be a sequence of sets. Suppose that $\beta_1, \beta_2, \beta_3, \dots$ are full covers of sets $E_1, E_2, E_3, \dots$ respectively. Show that*

$$\beta = \beta_1 \cup \beta_2 \cup \beta_3 \cup \dots$$

is a full cover of $E = \bigcup_{n=1}^{\infty} E_n$. □

Exercise 77 *Let $E_1, E_2, E_3, \dots$ be a sequence of sets. Suppose that $\beta_1, \beta_2, \beta_3, \dots$ are fine covers of sets $E_1, E_2, E_3, \dots$ respectively. Show that*

$$\beta = \beta_1 \cup \beta_2 \cup \beta_3 \cup \dots$$

is a fine cover of $E = \bigcup_{n=1}^{\infty} E_n$. □

Exercise 78 *Let $F : \mathbb{R} \to \mathbb{R}$ and let $\varepsilon : \mathbb{R} \to \mathbb{R}^+$ be an arbitrary positive function. Show that*

$$\beta_1 = \{([u,v],w) : |F(u) - F(v)| \leq \varepsilon(w)\}$$

is a full cover of the set of points at which F is continuous, while

$$\beta_2 = \{([u,v],w) : |F(u) - F(v)| \geq \varepsilon(w)\}$$

is a fine cover of the set of points at which F is not continuous. □

Exercise 79 *Let F, $f : \mathbb{R} \to \mathbb{R}$, and let $\varepsilon : \mathbb{R} \to \mathbb{R}^+$ be an arbitrary positive function. Show that*

$$\beta_1 = \{([u,v],w) : |F(u) - F(v) - f(w)(v-u)| \leq \varepsilon(w)(v-u))\}$$

is a full cover of the set of points x at which $F'(x) = f(x)$ is true, while

$$\beta_2 = \{([u,v],w) : |F(u) - F(v) - f(w)(v-u)| \geq \varepsilon(w)(v-u))\}$$

is a fine cover of the set of points x at which $F'(x) = f(x)$ fails to be true, i.e., either F is not differentiable at x or else F is differentiable at x but $F'(x) \neq f(x)$.- □

Exercise 80 *Let $F : \mathbb{R} \to \mathbb{R}$ and let $\varepsilon > 0$. Show that*

$$\beta_1 = \{([u,v],w) : F(t) > f(w) - \varepsilon \text{ for all } t \in [u,v]\}$$

is a full cover of the set of points at which F is lower semicontinuous. □

Exercise 81 (Duality) *Show that β is fine at a point w if and only if for all β_1 that are full at w there is at least one pair $([u,v],w)$ belonging to both β and β_1.*
Answer □

Exercise 82 (Duality) *Show that β is full at a point w if and only if for all β_1 that are fine at w there is at least one pair $([u,v],w)$ belonging to both β and β_1.*
Answer □

Exercise 83 (Heine-Borel) *Let $\mathcal{G}$ be a family of open sets so that every point in a compact set K is contained in at least one member of the family. Show that the covering relation*

$$\beta = \{(I,x) : x \in I \text{ and } I \subset G \text{ for some } G \in \mathcal{G}\}.$$

is a full cover of K (cf. the Heine-Borel Theorem). □

Exercise 84 (Bolzano-Weierstrass) *Let E be an infinite set that contains no points of accumulation. Show that*

$$\beta = \{(I,x) : x \in I \text{ and } I \cap E \text{ is finite}\}.$$

must be a full cover (cf. the Bolzano-Weierstrass Theorem). □

Exercise 85 *Let $\{x_n\}$ be a sequence of real numbers and let*

$$\beta = \{(I,x) : x \in I \text{ and } I \text{ contains only finitely many of the } x_n\}.$$

If β is a fine cover of a set E what (if anything) can you conclude? Answer □

Exercise 86 *Let $\{x_n\}$ be a sequence of real numbers and let*

$$\beta = \{(I,x) : x \in I \text{ and } I \text{ contains only finitely many of the } x_n\}.$$

If β is not a fine cover of a set E what (if anything) can you conclude?
Answer □

Exercise 87 *Let $\{x_n\}$ be a sequence of real numbers and let*

$$\beta = \{(I,x) : x \in I \text{ and } I \text{ contains only finitely many of the } x_n\}.$$

If β is a full cover of a set E what (if anything) can you conclude? Answer □

Exercise 88 *Let $\{x_n\}$ be a sequence of real numbers and let*

$$\beta = \{(I,x) : x \in I \text{ and } I \text{ contains only finitely many of the } x_n\}.$$

If β is not a full cover of a set E what (if anything) can you conclude?
Answer □

Exercise 89 *Let $\{x_n\}$ be a sequence of real numbers and let*

$$\beta = \{(I,x) : x \in I \text{ and } I \text{ contains infinitely many of the } x_n\}.$$

If β is a fine cover of a set E what (if anything) can you conclude? Answer □

Exercise 90 *Let $\{x_n\}$ be a sequence of real numbers and let*

$$\beta = \{(I,x) : x \in I \text{ and } I \text{ contains infinitely many of the } x_n\}.$$

If β is a not a fine cover of a set E what (if anything) can you conclude?

Answer □

Exercise 91 *Let $\{x_n\}$ be a sequence of real numbers and let*

$$\beta = \{(I,x) : x \in I \text{ and } I \text{ contains infinitely many of the } x_n\}.$$

If β is a full cover of a set E what (if anything) can you conclude? Answer □

Exercise 92 *Let $\{x_n\}$ be a sequence of real numbers and let*

$$\beta = \{(I,x) : x \in I \text{ and } I \text{ contains infinitely many of the } x_n\}.$$

If β is a not a full cover of a set E what (if anything) can you conclude?

Answer □

2.2 Covering arguments

There are a number of standard devices we can use that employ covering relations. The most fundamental for our purposes, perhaps, is the Cousin covering lemma, asserting that full covers always contain partitions. We need this in order justify our definition of the integral in terms of Riemann sums, full covers, and partitions. Advanced readers would be familiar with the Vitali covering argument that appears frequently in later courses of analysis.

2.2.1 Cousin covering lemma

In elementary analysis the Cousin covering lemma can often be used to construct proofs that might normally invoke the Bolzano-Weierstrass theorem or the Heine-Borel theorem. We repeat it here for convenience and to stress the role that it plays in covering arguments in analysis and in integration theory. This also allows us a chance to rewrite the proof in the language of this chapter.

Lemma 2.5 (Cousin covering lemma) *Let β be a full cover. Then β contains a partition of every compact interval.*

Proof. Note, first, that if β fails to contain a partition of some interval $[a,b]$ then it must fail to contain a partition of much smaller subintervals. For example if $a < c < b$, if π_1 is a partition of $[a,c]$ and π_2 is a partition of $[c,b]$, then $\pi_1 \cup \pi_2$ is certainly a partition of $[a,b]$.

We use this feature repeatedly. Suppose that β fails to contain a partition of $[a,b]$. Choose a subinterval $[a_1,b_1]$ with length less than $1/2$ the length of $[a,b]$ so that β fails to contain a partition of $[a_1,b_1]$. Continue inductively, selecting a nested sequence of compact intervals $[a_n, b_n]$ with lengths shrinking to zero so that β fails to contain a partition of each $[a_n, b_n]$.

By the nested interval property there is point z belonging to each of the intervals. As β is a full cover, there must exist a $\delta > 0$ so that β contains (I, z) for any compact subinterval I of $[a, b]$ with length smaller than δ. In particular β contains all $([a_n, b_n], z)$ for n large enough to assure us that $b_n - a_n < \delta$. The set $\pi = \{([a_n, b_n], z)\}\}$ containing a single element is itself a partition of $[a_n, b_n]$ that is contained in β. That contradicts our assumptions. Consequently β must contain a partition of $[a, b]$. Since $[a, b]$ was arbitrary, β must contain a partition of any compact interval. ∎

2.2.2 A simple covering argument

As an illustration of how Cousin's covering lemma can be used in a covering argument we prove a well-known theorem from elementary calculus. Most calculus students learn and use the fact that continuous functions possess the Darboux property (mostly called the "intermediate value" property in elementary calculus classes). The proofs are usually inadequate and given properly only in more advanced courses. The most familiar genuine proofs are based on the Heine-Borel theorem or the Bolzano-Weierstrass theorem. We give a simple covering argument instead.

Theorem 2.6 *Every continuous function $F : \mathbb{R} \to \mathbb{R}$ has the Darboux property.*

Proof. The statement of the theorem is equivalent to this: if $F(x) \neq k$ for all x then either $F(x) > k$ for all x or else $F(x) < k$ for all x. By applying this to the function $G(x) = F(x) - k$ we see that this asserts that if $G(x) \neq 0$ for all x then either $G(x) > 0$ for all x or else $G(x) < 0$ for all x.

Assume, then, that the continuous function G never vanishes. Define

$$\beta = \left\{ ([u, v], w) : \frac{G(v)}{G(u)} > 0 \right\}. \tag{2.1}$$

This can be checked to be a full cover. Accordingly, by the Cousin covering lemma, if $[a, b]$ is any interval then there are points

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

and associated points $\{\xi_i\}$ so that

$$\pi = \{([x_{i-1}, x_i], \xi_i) : i = 1, 2, \ldots, n\} \subset \beta.$$

Hence

$$\frac{G(b)}{G(b)} = \frac{G(x_1)}{G(x_0)} \times \frac{G(x_2)}{G(x_1)} \times \frac{G(x_3)}{G(x_2)} \times \cdots \times \frac{G(x_n)}{G(x_{n-1})} > 0.$$

Consequently, for any interval $[a, b]$, $G(a)$ and $G(b)$ have the same sign, i.e, either $G(x) > 0$ for all x or else $G(x) < 0$ for all x. ∎

Keep it elegant Note that the definition of the covering relation β in (2.1) is focused on the property that we want to prove, not constructed from the hypotheses of the theorem. One might have been tempted instead first to choose $\delta(x) > 0$ for each x so that $|F(y) - F(x)| < |F(x)|/2$ if $|y - x| < \delta(x)$. This is certainly possible because $|F(x)| > 0$ and F is continuous at each point. Then one might define

$$\beta_1 = \{([u, v], w) : 0 < v - u < \delta(w)\}.$$

The same argument would work because $\beta_1 \subset \beta$. But that would be a clumsy use of covering arguments.

Exercise 93 *Prove the Heine-Borel theorem using a covering argument.*

Answer □

Exercise 94 *Prove the Bolzano-Weierstrass theorem using a covering argument.*

Answer □

Exercise 95 *Suppose that $F : \mathbb{R} \to \mathbb{R}$ is a continuous function for which $F'(x) \geq 0$ at every point x with at most countably many exceptions. Construct a covering argument to demonstrate that F is nondecreasing.* Answer □

Exercise 96 *Suppose that $F : \mathbb{R} \to \mathbb{R}$ is a lower semicontinuous function. Construct a covering argument to demonstrate that F is bounded below on every compact interval.* Answer □

Exercise 97 *A function $F : \mathbb{R} \to \mathbb{R}$ is said to have* bounded derived numbers *if there is a number M so that, for each x one can choose $\delta > 0$ so that*

$$\left| \frac{F(x+h) - F(x)}{h} \right| \leq M$$

whenever $0 < |h| < \delta$. Show that F is Lipschitz if and only if F has bounded derived numbers. Answer □

2.2.3 Decomposition of full covers

There is a decomposition of full covers that is often of use in constructing a proof. Here is a good place to put it for easy reference, although it is mostly unmotivated for the moment. (We have used a similar idea in the solution of Exercise 62.) This shows that, while a full cover is a much more general object than a uniformly full cover, it can be broken into pieces that are themselves uniform covers.

Lemma 2.7 (Decomposition Lemma) *Let β be a full cover of a set E. Then there is an increasing sequence of sets $\{E_n\}$ with $E = \bigcup_{n=1}^{\infty} E_n$ and a sequence of nonoverlapping compact intervals $\{I_{kn}\}$ covering E_n so that if x is any point in E_n and I is any subinterval of I_{kn} that contains x then (I,x) belongs to β.*

Proof. Let β be a full cover of a set E. By the nature of the cover there must exist, for each $x \in E$ a positive number $\delta(x)$ on E with the property that (I,x) belongs to β whenever if $x \in E$, $x \in I$ and the length of the interval I is smaller than $\delta(x)$. Define

$$E_n = \{x \in E : \delta(x) > 1/n\}.$$

This is an expanding sequence of subsets of E whose union is E itself. If I is any compact interval that contains a point x in E_n and has length less than $1/n$, then (I,x) must belong to β.

A way of exploiting this property is to introduce the intervals

$$I_{mn} = \left[\frac{m}{n}, \frac{m+1}{n}\right]$$

for integers $m = 0, \pm 1, \pm 2, \ldots$. Then $\beta([E_n \cap I_{mn}])$ has this property: if x is any point in $E_n \cap I_{mn}$ and I is any subinterval of I_{mn} that contains x then (I,x) is a member of $\beta([E_n \cap I_{mn}])$.

Thus the condition of being a full cover, which is a local condition defined in a special way at each point, has been made uniform throughout each piece of the decomposition. If we relabel these sets in a convenient way then we now have our decomposition property. ∎

2.3 Riemann sums

The integral can be characterized as a limit of Riemann sums. The original Riemann integral has such a definition and the Lebesgue integral, although originally defined in a completely different manner, also has such a characterization although not as simple as that for the Riemann integral. The notation for Riemann sums can assume any of the following forms (2.2), (2.3), (2.4), or (2.5), depending on which is convenient:

Take an interval $[a,b]$ and subdivide as follows:

$$a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n = b.$$

Then form a partition of $[a,b]$ by selecting points ξ_i from each of the corresponding intervals:

$$\pi = \{[x_0,x_1],\xi_1), \ ([x_1,x_2],\xi_2), \ \ldots, \ ([x_{n-1},x_n],\xi_n\}.$$

Sums of the following form are called *Riemann sums* with respect to this partition:

$$\sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1}). \tag{2.2}$$

These can also be more conveniently written as

$$\sum_{([u,v],w)\in\pi} f(w)(v-u) \tag{2.3}$$

or

$$\sum_{([u,v],w)\in\pi} f(w)\lambda([u,v]) \tag{2.4}$$

or even as

$$\sum_{(I,w)\in\pi} f(w)\lambda(I). \tag{2.5}$$

Here we are using λ as a *length function*:

$$\lambda([u,v]) = v-u$$

is simply the length of the interval $[u,v]$. We can in this way also conveniently assign a length to the intersection of two compact intervals. For example,

$$\lambda([u_1,v_1]\cap[u_2,v_2])$$

would be the length of the interval $[u_1,v_1]\cap[u_2,v_2]$ (if it is an interval) and would have length zero if the two intervals do not overlap.

General Riemann sums In general, let $h([u,v],w)$ denote any real-valued function which assigns to an interval-point pair $([u,v],w)$ a real value. Let π be any partition or subpartition. Then we will (loosely) call any sum

$$\sum_{([u,v],w)\in\pi} h([u,v],w) \tag{2.6}$$

or

$$\sum_{(I,w)\in\pi} h(I,w) \tag{2.7}$$

a Riemann sum. Such sums will play a role in many diverse investigations.

Exercises

Exercise 98 *Let $F:[a,b]\to\mathbb{R}$ and let π be a partition of $[a,b]$. Verify the computations*

$$\sum_{([u,v],w)\in\pi} (v-u) = b-a$$

and

$$\sum_{([u,v],w)\in\pi} (F(v)-F(u)) = F(b)-F(a).$$

□

Exercise 99 *Let $F:[a,b]\to\mathbb{R}$ and let π be a partition of $[a,b]$. Show that*

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| \geq |F(b)-F(a)|.$$

$\square$

Exercise 100 *Let $F:[a,b]\to\mathbb{R}$ be a Lipschitz function with Lipschitz constant M and let π be a partition of the interval $[a,b]$. Show that*

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| \leq M(b-a)|.$$

$\square$

Exercise 101 *Let F, $f:[a,b]\to\mathbb{R}$ and let π be a partition of $[a,b]$ and suppose that*

$$F(v)-F(u) \geq f(w)(v-u)$$

for all $([u,v],w)\in\pi$. Show that

$$\sum_{([u,v],w)\in\pi} f(w)(v-u)) \leq F(b)-F(a).$$

$\square$

Exercise 102 *Let $F:[a,b]\to\mathbb{R}$ be a function with the property that*

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| = 0.$$

for every partition π of the interval $[a,b]$. What can you conclude? $\square$

Exercise 103 *Let $F:[0,1]\to\mathbb{R}$ be a function with the property that it is monotonic on each of the intervals $[0,\frac{1}{3}]$, $[\frac{1}{3},\frac{2}{3}]$, and $[\frac{2}{3},1]$. What is the largest possible value of*

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)|$$

for arbitrary partitions π of the interval $[a,b]$. $\square$

Exercise 104 *Describe the difference between the two sums*

$$\sum_{([u,v],w)\in\pi} f(w)(v-u)$$

and

$$\sum_{(I,w)\in\pi([c,d])} f(w)(v-u)$$

where $[c,d]$ is an interval. Answer $\square$

Exercise 105 *Describe the difference between the two sums*

$$\sum_{([u,v],w)\in\pi} f(w)(v-u)$$

and

$$\sum_{([u,v],w)\in\pi[E]} f(w)(v-u).$$

where E is a set. Answer □

Exercise 106 *How could you interpret the expression*

$$\sum_{([u,v],w)\in\pi_1\cup\pi_2} f(w)(v-u)?$$

□

Exercise 107 *How could you interpret the expression*

$$\sum_{(([u_1,v_1],w_1)\in\pi_1} \sum_{([u_2,v_2],w_2)\in\pi_2} f(w_1)\lambda([u_1,v_1]\cap[u_2,v_2])?$$

if π_1 and π_2 are both partitions of the same interval $[a,b]$? □

Exercise 108 *Show that*

$$\sum_{(([u_1,v_1],w_1)\in\pi_1} f(w_1)\lambda([u_1,v_1]) - \sum_{([u_2,v_2],w_2)\in\pi_2} f(w_2)\lambda([u_2,v_2]) =$$

$$\sum_{(([u_1,v_1],w_1)\in\pi_1} \sum_{([u_2,v_2],w_2)\in\pi_2} [f(w_1)-f(w_2)]\lambda([u_1,v_1]\cap[u_2,v_2])$$

if π_1 and π_2 are both partitions of the same interval $[a,b]$? □

Exercise 109 *Let $f:[a,b]\to\mathbb{R}$ be a continuous function and let $\varepsilon>0$. What could you require of two partitions π_1 and π_2 of the interval $[a,b]$ in order to conclude that*

$$\left|\sum_{(([u_1,v_1],w_1)\in\pi_1} f(w_1)(v_1-u_1) - \sum_{([u_2,v_2],w_2)\in\pi_2} f(w_2)(v_2-u_2)\right| < \varepsilon?$$

□

2.4 Sets of Lebesgue measure zero

We review the notion of a set of Lebesgue measure zero already studied in Chapter 1. We will present three equivalent versions of this concept. The first is due to Lebesgue and arises from his theory of measure. The second and third use full and fine coverings and estimates using Riemann sums. Before we used the full covering version for our first definition of Lebesgue measure zero. Now we begin with Lebesgue's definition.

2.4.1 Lebesgue measure of open sets

The property that a set E will be a set of Lebesgue measure zero is actually a statement about the family of open sets containing E. A set E is measure zero if there are arbitrarily "small" open sets containing E.

For a precise version of this we require a definition for the Lebesgue measure $\lambda(G)$ of an open set G. Later on we will study Lebesgue's measure in general. The attention here is directed only on that measure for open sets.

Definition 2.8 *Let G be an open set. Then the Lebesgue measure $\lambda(G)$ of an open set G is defined to be the total sum of the lengths of all the component intervals of G.*

According to this definition $\lambda(\emptyset) = 0$ (since there are no component intervals). If G consists of infinitely many bounded component intervals $(\{a_i, b_i)\}$ then the measure is the sum of an infinite series:

$$\lambda(G) = \sum_{i=1}^{\infty}(b_i - a_i).$$

[An unbounded component interval would have length ∞ and so an open set with an unbounded component has infinite measure.]

The only tool we need for working with this concept for the moment is given by the subadditivity property.

Lemma 2.9 (Subadditivity) *Let G_1, G_2, G_3, ... be a sequence of open sets. Then the union*

$$G = \bigcup_{i=1}^{\infty} G_i$$

is also an open set and

$$\lambda(G) \leq \sum_{i=1}^{\infty} \lambda(G_i).$$

Proof. Certainly G is open since any union of open sets is open. Let

$$T = \sum_{i=1}^{\infty} \lambda(G_i).$$

Note that T is simply the sum of the lengths of all the component intervals of all the G_i.

Let $(\{a_j, b_j)\}$ denote the component intervals of G. Take (a_1, b_1) and choose any $[c_1, d_1] \subset (a_1, b_1)$. A compactness argument shows that $[c_1, d_1]$ is contained in finitely many of the component intervals making up the sum T. We conclude that $d_1 - c_1 \leq T$. That would be true for any choice of $[c_1, d_1] \subset (a_1, b_1)$, so that $b_1 - a_1 \leq T$. A similar argument using m components (a_1, b_1), (a_2, b_2), ..., (a_m, b_m) will establish that

$$\sum_{j=1}^{m}(b_j - a_j) \leq T$$

from which

$$\lambda(G) = \sum_{j=1}^{\infty}(b_j - a_j) \leq T$$

evidently follows. ∎

2.4.2 Sets of Lebesgue measure zero

Our first definition of Lebesgue measure zero set expresses this as a property of open sets that contain the set.

Definition 2.10 *Let E be a set of real numbers. Then E is said to have* Lebesgue measure zero *if for every $\varepsilon > 0$ there is an open set G containing E for which $\lambda(G) < \varepsilon$.*

Recall that we have given a completely different definition of measure zero in Chapter 1. Thus we are obliged very quickly to show that these two definitions are equivalent. In the meantime the following exercises should be attempted but now with the new definition. In Section 2.7 we will show that the two definitions (along with a third definition for Lebesgue measure zero) are equivalent.

Exercises

Exercise 110 *The empty set has Lebesgue measure zero.* Answer □

Exercise 111 *Every finite set has Lebesgue measure zero.* Answer □

Exercise 112 *Every infinite, countable set has Lebesgue measure zero.* Answer □

Exercise 113 *The Cantor set has Lebesgue measure zero.* Answer □

2.4.3 Sequences of Lebesgue measure zero sets

One of the most fundamental of the properties of sets having measure zero is how sequences of such sets combine. We recall that the union of any sequence of countable sets is also countable. We now prove that the union of any sequence of Lebesgue measure zero sets is also a Lebesgue measure zero set.

Theorem 2.11 *Let $E_1, E_2, E_3, \ldots$ be a sequence of sets of measure zero. Then the set E formed by taking the union of all the sets in the sequence is also of Lebesgue measure zero.*

Proof. Let $\varepsilon > 0$. Choose open sets $G_n \supset E_n$ so that

$$\lambda(G_n) < 2^{-n}\varepsilon.$$

Then set $G = \bigcup_{n=1}^{\infty} G_n$. Observe, by the subadditivity property (i.e., from Lemma 2.9), that G is an open set containing E for which $\lambda(G) < \varepsilon$. ∎

Exercises

Exercise 114 *Show that E is a set of Lebesgue measure zero if and only if there is a finite or infinite sequence*

$$(a_1,b_1),(a_2,b_2),(a_3,b_3),(a_4,b_4),\dots$$

of open intervals covering the set E so that

$$\sum_{k=1}^{\infty}(b_k-a_k)\le\varepsilon.$$

□

Exercise 115 (compact sets of Lebesgue measure zero) *Let E be a compact set of Lebesgue measure zero. Show that for every $\varepsilon>0$ there is a* finite *collection of open intervals*

$$\{(a_k,b_k):k=1,2,3,\dots,N\}$$

that covers the set E and so that

$$\sum_{k=1}^{N}(b_k-a_k)<\varepsilon.$$

Answer □

Exercise 116 *Show that E is a set of Lebesgue measure zero if and only if there is a finite or infinite sequence*

$$[a_1,b_1],[a_2,b_2],[a_3,b_3],[a_4,b_4],\dots$$

of compact intervals covering the set E so that

$$\sum_{k=1}^{\infty}(b_k-a_k)\le\varepsilon.$$

□

Exercise 117 *Show that every subset of a set of Lebesgue measure zero also has Lebesgue measure zero.* □

Exercise 118 *Suppose that $E\subset[a,b]$ is a set of Lebesgue measure zero. Show that $\int_a^b \chi_E(x)\,dx=0$.* Answer □

Exercise 119 *If E has Lebesgue measure zero, show that the translated set*

$$E+\alpha=\{x+\alpha:x\in E\}$$

also has Lebesgue measure zero. □

Exercise 120 *If E has Lebesgue measure zero, show that the expanded set*

$$cE=\{cx:x\in E\}$$

also has Lebesgue measure zero for any $c>0$. □

Exercise 121 *If E has Lebesgue measure zero, show that the reflected set*

$$-E = \{-x : x \in E\}$$

also has Lebesgue measure zero. □

Exercise 122 *Without referring to Theorem 2.11, show that the union of any two sets of Lebesgue measure zero also has Lebesgue measure zero.* □

Exercise 123 *If $E_1 \subset E_2$ and E_1 has Lebesgue measure zero but E_2 has not, what can you say about the set $E_2 \setminus E_1$?* □

Exercise 124 *Show that any interval (a,b) or $[a,b]$ is not of Lebesgue measure zero.* □

Exercise 125 *Give an example of a set that is not of Lebesgue measure zero and does not contain any interval $[a,b]$.* □

Exercise 126 *A careless student claims that if a set E has measure zero, then it is true that the closure $\overline{E}$ must also have Lebesgue measure zero. After all, if E is contained in $\bigcup_{i=1}^{\infty}(a_i,b_i)$ with small total length then $\overline{E}$ is contained in $\bigcup_{i=1}^{\infty}[a_i,b_i]$, also with small total length. Is this correct?* □

Exercise 127 *If a set E has Lebesgue measure zero what can you say about interior points of that set?* □

Exercise 128 *If a set E has Lebesgue measure zero what can you say about boundary points of that set?* □

Exercise 129 *Suppose that a set E has the property that $E \cap [a,b]$ has Lebesgue measure zero for every compact interval $[a,b]$. Must E also have Lebesgue measure zero?* □

Exercise 130 *Show that the set of real numbers in the interval $[0,1]$ that do not have a 7 in their infinite decimal expansion is of Lebesgue measure zero.* □

Exercise 131 *Describe completely the class of sets E with the following property: For every $\varepsilon > 0$ there is a* finite *collection of open intervals*

$$(a_1,b_1),(a_2,b_2),(a_3,b_3),(a_4,b_4),\dots(a_N,b_N)$$

that covers the set E and so that

$$\sum_{k=1}^{N}(b_k - a_k) < \varepsilon.$$

(These sets are said to have zero content.*)* □

Exercise 132 *Show that a set E has Lebesgue measure zero if and only if there is a sequence of intervals*

$$(a_1,b_1),(a_2,b_2),(a_3,b_3),(a_4,b_4),\dots$$

so that every point in E belongs to infinitely many of the intervals and $\sum_{k=1}^{\infty}(b_k - a_k)$ converges. □

Exercise 133 *Suppose that $\{(a_i,b_i)\}$ is a sequence of open intervals for which*

$$\sum_{i=1}^{\infty}(b_i-a_i)<\infty.$$

Show that the set

$$E=\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}(a_i,b_i)$$

has Lebesgue measure zero. What relation does this exercise have with the preceding exercise? □

Exercise 134 *By altering the construction of the Cantor set, construct a nowhere dense closed subset of $[0,1]$ so that the sum of the lengths of the intervals removed is not equal to 1. Will this set have Lebesgue measure zero?-* □

2.4.4 Almost everywhere a.e. language

Here is some language commonly used in discussions of Lebesgue measure zero sets. Let $P(x)$ be a property that may or not be possessed by a point $x \in \mathbb{R}$. We say that

> $P(x)$ *is true almost everywhere*

or

> $P(x)$ *is true for almost every x*

if the set

$$\{x \in \mathbb{R} : P(x) \text{ is not true}\}$$

is a Lebesgue measure zero set.

There is a convenient extension of this language useful in integration theory:

(mostly everywhere) A statement holds *mostly everywhere* if it holds everywhere with the exception of a finite set of points $c_1, c_2, c_3, \dots, c_n$.

(nearly everywhere) A statement holds *nearly everywhere* if it holds everywhere with the exception of a countable set.

(almost everywhere) A statement holds *almost everywhere* if it holds everywhere with the exception of a set of Lebesgue measure zero.

Nearly everywhere might be abbreviated "n.e." but only in a context where the reader is reminded of such usage. Almost everywhere is very frequently abbreviated "a.e." and most advanced readers are familiar with this usage.

Exercises

Exercise 135 *What would it mean to say that a function is almost everywhere discontinuous?* □

Exercise 136 *What would it mean to say that a function is almost everywhere differentiable? Give an example of function that is almost everywhere differentiable, but not everywhere differentiable.* □

Exercise 137 *What would it mean to say that almost every point in* $\mathbb{R}$ *is irrational? Is this a true statement?* □

Exercise 138 *What would it mean to say that almost everywhere point in a set* A *belongs to a set* B*? Give an example for which this is true and an example for which this is false.* □

Exercise 139 *What would it mean to say that a function is almost everywhere equal to zero?* □

Exercise 140 *What would it mean to say that a function is almost everywhere different from zero?* □

Exercise 141 *Suppose that the function* $f : [a,b] \to \mathbb{R}$ *is integrable and is almost everywhere in* $[a,b]$ *nonnegative. Show that* $\int_a^b f(x)\,dx \geq 0$*.*

Answer □

Exercise 142 *Suppose that the functions* F*,* $G : [a,b] \to \mathbb{R}$ *are continuous almost everywhere in* $[a,b]$*. Is the sum function* $F(x) + G(x)$ *also continuous almost everywhere in* $[a,b]$*.* □

Exercise 143 *Suppose that the functions* F*,* $G : [a,b] \to \mathbb{R}$ *are differentiable almost everywhere in* $[a,b]$*. Is the sum function* $F(x) + G(x)$ *also differentiable almost everywhere in* $[a,b]$*.* □

2.5 Full null sets

Sets of Lebesgue measure zero are defined using open sets that contain them. There is a variant on this using full covers instead. We have already taken advantage of this variant in Chapter 1 because that variant has the closest connection with integration theory as we have presented it. For the moment we refer to this version of measure zero as "full null." Once we have proved the equivalence we can revert to normal usage and just label such sets as "Lebesgue measure zero" or more simply and commonly in discussions of real functions, merely as "measure zero."

This definition has the advantage that, since it is defined using full covers, the definition is more closely related to the differentiation and integration properties of functions. It has the disadvantage that, unlike the Lebesgue measure zero sets, it is not constructive; full covers themselves are not necessarily constructive.

Definition 2.12 *A set E of real numbers is said to be* full null *if for every $\varepsilon > 0$ there is a full cover β of the set E with the property that*

$$\sum_{([u,v],w)\in\pi} (v-u) < \varepsilon \tag{2.8}$$

for every subpartition π chosen from β.

We recall that an equivalent formulation (in Definition 1.11) required what is at first sight much stronger: that

$$\sum_{([u,v],w)\in\pi} |f(w)(v-u)| < \varepsilon \tag{2.9}$$

for arbitrary functions $f : E \to \mathbb{R}$ rather than the narrower condition in (2.8). Thus we can also describe full null sets as those for which all Riemann sums concentrated on them can be arranged to be small. We will further show that the two definitions, full null and Lebesgue measure zero, are equivalent later. For the moment one direction is easy.

Theorem 2.13 *Every set of Lebesgue measure zero is also full null.*

Proof. Assume that a set E Lebesgue measure zero and let $\varepsilon > 0$. Choose an open set G containing E for which $\lambda(G) < \varepsilon$. Let $\{(a_i,b_i)\}$ be the component intervals of G. Define β to be the collection of all pairs $([u,v],w)$ with the property that $w \in [u,v] \subset G$. It is easy to check that β is a full cover of E.

Consider any subpartition π chosen from β. For each $([u,v],w) \in \pi$, $[u,v]$ is a subinterval of some component (a_i,b_i) of G. Holding i fixed, the sum of the lengths of those intervals $[u,v] \subset (a_i,b_i)$ would certainly be smaller than

$(b_i - a_i)$. It follows that

$$\sum_{([u,v],w)\in\pi} (v-u) \leq \sum_{i=1}^{\infty}(b_i - a_i) = \lambda(G) < \varepsilon.$$

This verifies that E is full null. ■

Exercises

Exercise 144 *Show that every subset of a full null set is also a full null set.* □

Exercise 145 *Show that the union of any two full null sets is also a full null set.* □

Exercise 146 *Show that the union of any sequence of full null sets is also a full null set.* □

Exercise 147 *Define a set E to be* uniformly full null *if for every $\varepsilon > 0$ there is a uniformly full cover β of the set E with the property that*

$$\sum_{([u,v],w)\in\pi} (v-u) < \varepsilon \tag{2.10}$$

for every subpartition π chosen from β. Show that uniformly full null sets are the same as sets of zero content. (cf. Exercise 131). □

Exercise 148 (Small Riemann sums) *Show that our definition in this section is equivalent to Definition 1.11, i.e., show that a set E of real numbers is full null if and only if for every $\varepsilon > 0$ and any function $f : E \to \mathbb{R}$ there is a full cover β of E such that*

$$\left|\sum_{i=1}^{n} f(\xi_i)(y_i - x_i)\right| < \varepsilon$$

for all subpartitions $\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \ldots\}$ contained in β. □

2.6 Fine null sets

Sets of Lebesgue measure zero are defined with attention to the open sets that contain them. Full null sets are defined using full covers. There is a third variant on this using fine covers instead. This offers yet a more delicate way of working with Lebesgue measure zero sets, since fine covers can express very subtle properties of derivatives and integrals. We will show in Section 2.7 that all three notions are equivalent.

Definition 2.14 *A set E of real numbers is said to be* fine null *if for every $\varepsilon > 0$ there is a fine cover β of the set E with the property that*

$$\sum_{([u,v],w)\in\pi} (v-u) < \varepsilon \tag{2.11}$$

for every subpartition π chosen from β.

We remark that an equivalent formulation would be to insist on a formally stronger requirement: that

$$\sum_{([u,v],w)\in\pi} |f(w)(v-u)| < \varepsilon \tag{2.12}$$

for arbitrary functions $f : E \to \mathbb{R}$ rather than the narrower condition in (2.11). The reader can take either as a definition of fine null sets.

Exercises

Exercise 149 *Show that every set of Lebesgue measure zero is also fine null.* □

Exercise 150 *Show that every full null set is also fine null.* □

Exercise 151 *Show that every subset of a fine null set is also a fine null set.* □

Exercise 152 *Show that the union of any two fine null sets is also a full null set.* □

Exercise 153 *Show that the union of any sequence of fine null sets is also a fine null set.* □

Exercise 154 (Small Riemann sums) *Show that our definition in this section is equivalent to a fine version of Definition 1.11, i.e., show that a set E of real numbers is fine null if and only if for every $\varepsilon > 0$ and any function $f : E \to \mathbb{R}$ there is a fine cover β of E such that*

$$\left|\sum_{i=1}^{n} f(\xi_i)(y_i - x_i)\right| < \varepsilon$$

for all subpartitions $\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$ contained in β. □

2.7 The Mini-Vitali Covering Theorem

The original Vitali covering theorem asserts that the Lebesgue measure of an arbitrary set can be determined either by open coverings of E, or by full covers of E, or by fine covers of E. Our goals in this chapter are narrower. We want to establish these same facts, but only for sets of Lebesgue measure zero. Later we will return and complete the Vitali covering theorem.

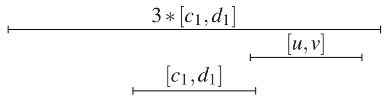

Figure 2.1: Note that $3*[c_1,d_1]$ will then include any shorter interval $[u,v]$ that intersects $[c_1,d_1]$.

Theorem 2.15 (Mini-Vitali covering theorem) *For any set $E \subset \mathbb{R}$ the following are equivalent:*

1. *E is a set of Lebesgue measure zero.*
2. *E is a full null set.*
3. *E is a fine null set.*

As a result of this theorem we can now simply refer to these sets as measure zero sets and use any of the three characterizations that is convenient. The proof requires some simple geometric arguments and an application of the Heine-Borel theorem; it is given in the sections that now follow.

2.7.1 Covering lemmas for families of compact intervals

We begin with some simple covering lemmas for finite and infinite families of compact intervals.

Lemma 2.16 *Let $\mathcal{C}$ be a finite family of compact intervals. Then there is a pairwise disjoint subcollection $[c_i,d_i]$ $(i=1,2,\dots,m)$ of that family with*[a]

$$\bigcup_{[u,v]\in\mathcal{C}} [u,v] \subset \bigcup_{i=1}^{m} 3*[c_i,d_i].$$

[a] By $3*[u,v]$ we mean the interval centered at the same point as $[u,v]$ but with three times the length.

Proof. For $[c_1,d_1]$ simply choose the largest interval. Note that $3*[c_1,d_1]$ will then include any other interval $[u,v]\in\mathcal{C}$ that intersects $[c_1,d_1]$. See Figure 2.1.

For $[c_2,d_2]$ choose the largest interval from among those that do not intersect $[c_1,d_1]$. Note that together $3*[c_1,d_1]$ and $3*[c_2,d_2]$ include any interval of the family that intersects either $[c_1,d_1]$ or $[c_2,d_2]$. Continue inductively, choosing $[c_{k+1},d_{k+1}]$ as the largest interval in $\mathcal{C}$ that does not intersect one the previously chosen intervals $[c_1,d_1]$, $[c_2,d_2]$, ..., $[c_k,d_k]$. Stop when you run out of intervals to select. ∎

The next covering lemma addresses arbitrary families of compact intervals.

Lemma 2.17 *Let $\mathcal{C}$ be any collection of compact intervals. Then the set*

$$G = \bigcup_{[u,v]\in\mathcal{C}} (u,v)$$

is an open set that contains all but countably many points of the set

$$E = \bigcup_{[u,v]\in\mathcal{C}} [u,v]$$

Proof. Let

$$C = \{x : x \notin G \text{ and } x = c \text{ or } x = d \text{ for at least one } [c,d] \in \mathcal{C}\}.$$

We observe that G is open, being a union of a family of open intervals. Clearly G contains all of E except for points that are in the set C. To complete the proof of the lemma, we show that C is countable. Write, for $n = 1,2,3,\dots,$

$$C_n = \{x : x \notin G, \ \ x = c \text{ for at least one } [c,d] \in \mathcal{C} \text{ with } d-c > 1/n\}.$$

$$C_n' = \{x : x \notin G, \ \ x = d \text{ for at least one } [c,d] \in \mathcal{C} \text{ with } d-c > 1/n\}.$$

We easily show that each C_n and C_n' is countable. For example if $c \in C_n$ then the interval $(c, c+1/n)$ can contain no other point of C. This is because there is at least one interval $[c,d]$ from $\mathcal{C}$ with $d-c > 1/n$. Thus $(c,c+1/n) \subset (c,d) \subset G$. Consequently there can be only countably many such points. It follows that the set $C = \bigcup_{n=1}^{\infty}(C_n \cup C_n')$ is a countable subset of E. ∎

2.7.2 Proof of the Mini-Vitali covering theorem

We begin with a simple lemma that is the key to the argument, both for our proof of the mini version as well as the proof of the full Vitali covering theorem.

Lemma 2.18 *Let β be a covering relation and write*

$$G = \bigcup_{([u,v],w)\in\beta} (u,v).$$

Then G is an open set and, if $g = \lambda(G)$, is finite then there must exist a subpartition $\pi \subset \beta$ for which

$$\sum_{([u,v],w)\in\pi} (v-u) \geq g/6. \tag{2.13}$$

In particular

$$G' = G \setminus \bigcup_{([u,v],w)\in\pi} [u,v]$$

is an open subset of G and $\lambda(G') \leq 5g/6$.

Proof. It is clear that the set G of the lemma, expressed as the union of a family of open intervals, must be an open set. Let $\{(a_i,b_i)\}$ be the sequence of component intervals of G. Thus, by definition,

$$g = \lambda(G) = \sum_{i=1}^{\infty}(b_i - a_i).$$

Choose an integer N large enough that

$$\sum_{i=1}^{N}(b_i - a_i) > 3g/4.$$

Inside each open interval (a_i, b_i), for $i = 1, 2, \dots, N$, choose a compact interval $[c_i, d_i]$ so that

$$\sum_{i=1}^{N}(d_i - c_i) > g/2.$$

Write

$$K = \bigcup_{i=1}^{N}[c_i, d_i]$$

and note that it is a compact set covered by the family

$$\{(u,v) : ([u,v],w) \in \beta\}.$$

By the Heine-Borel theorem there must be a finite subset

$$([u_1,v_1],w_1),\ ([u_2,v_2],w_2),\ ([u_3,v_3],w_3),\ \dots,\ ([u_m,v_m],w_m)$$

from β for which

$$K \subset \bigcup_{i=1}^{m}(u_i, v_i).$$

By Lemma 2.16 we can extract a subpartition π from this list so that

$$K \subset \bigcup_{([u,v],w)\in\pi} 3 * [u,v].$$

and so

$$\sum_{([u,v],w\in\pi} 3(v-u) \geq \sum_{i=1}^{N}(d_i - c_i) > g/2.$$

Statement (2.13) then follows. [Not that we need it here, but recall that Lemma 2.16 allows us to claim that the intervals in the subpartition π are disjoint, not merely nonoverlapping.]

The final statement of the lemma requires just checking the length of a finite number of the components of G'. We have removed all the intervals $[u,v]$ from G for which $([u,v],w) \in \pi$. Since the total length removed is greater than $g/6$ what remains cannot be larger than $5g/6$. ∎

Proof of the Mini-Vitali covering theorem: We already know that every set of Lebesgue measure zero is full null, and that every full null set is fine null. To complete the proof we show that every fine null set is a set of Lebesgue measure zero. Let us suppose that E is not a set of Lebesgue measure zero. We show that it is not fine full then. Define

$$\varepsilon_0 = \inf\{\lambda(G) : G \text{ open and } G \supset E\}.$$

Since E is not Lebesgue measure zero, $\varepsilon_0 > 0$.

Let β be an arbitrary fine cover of E. Define

$$G = \bigcup_{([u,v],w)\in\beta} (u,v).$$

This is an open set and, by Lemma 2.17, G covers all of E except for a countable set. It follows that $\lambda(G) \geq \varepsilon_0$, since if $\lambda(G) < \varepsilon_0$ we could add to G a small open set G' that contains the missing countable set of points and for which the combined set $G \cup G'$ is an open set containing E but with measure smaller than ε_0.

By Lemma 2.18 there must exist a subpartition $\pi \subset \beta$ for which

$$\sum_{([u,v],w)\in\pi} (v-u) \geq \varepsilon_0/6.$$

But that means that E is not a fine null set, since this is true for *every* fine cover β.

2.8 Functions having zero variation

A set E is full null (i.e., Lebesgue measure zero) if there is a full cover β of the set E so that

$$\sum_{([u,v],w)\in\pi} (v-u) < \varepsilon$$

whenever π is a subpartition, $\pi \subset \beta$. This generalizes easily by considering instead sums

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)|$$

for some function F. We have used this definition in Chapter 1 but repeat and review it here.

Definition 2.19 *Let F be defined on an open set that contains a set E of real numbers. We say that F* has zero variation *on the set E provided that for every $\varepsilon > 0$ there is a full cover β of the set E so that*

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| < \varepsilon$$

whenever π is a subpartition, $\pi \subset \beta$.

Lemma 2.20 *Let $F : (a,b) \to \mathbb{R}$. Then F has zero variation on the open interval (a,b) if and only if F is constant on (a,b).*

Proof. One direction is obvious; the other direction is an application of the Cousin covering lemma. Suppose that F has zero variation on (a,b). Let $\varepsilon > 0$ and choose a full cover β of the set (a,b) so that

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| < \varepsilon$$

whenever π is a subpartition, $\pi \subset \beta$. If $[c,d] \subset (a,b)$ then there is a partition $\pi \subset \beta$ of the whole interval $[c,d]$. Consequently

$$|F(d)-F(c)| \leq \sum_{([u,v],w)\in\pi} |F(v)-F(u)| < \varepsilon.$$

This holds for every such interval $[c,d]$ and every positive ε. It follows that F must be constant on (a,b). ■

Lemma 2.21 *Let F be defined on an open set that contains each of the sets $E_1, E_2, E_3, \dots$ and suppose that F has zero variation on each E_i ($i = 1,2,3,\dots$). Then F has zero variation on any subset of the union $\bigcup_{i=1}^{\infty} E_i$.*

Proof. Let $\varepsilon > 0$ and, for each integer i, choose a full cover β_i of E_i so that

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| < 2^{-i}\varepsilon \tag{2.14}$$

whenever π is a subpartition, $\pi \subset \beta_i$. Construct β as the union of the sequence $\beta_i[E_i]$. This is a full cover of any subset E of the union $\bigcup_{i=1}^{\infty} E_i$. Now simply check that, if π is a subpartition, $\pi \subset \beta$ then

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| \leq \sum_{i=1}^{\infty} \sum_{([u,v],w)\in\pi[E_i]} |F(v)-F(u)| < \sum_{i=1}^{\infty} 2^{-i}\varepsilon = \varepsilon. \tag{2.15}$$

It follows that F has zero variation on E. ■

Exercises

Exercise 155 *Show that a constant function has zero variation on any set. Is the converse true, i.e., if F has zero variation on a set E must F be constant on E?* Answer □

Exercise 156 *Show that if F has zero variation on a set E then it has zero variation on any subset of E.* □

Exercise 157 *Let E contain a single point x_0. What does it mean for F to have zero variation on E?* Answer □

Exercise 158 *Let E have countably many points. Show that F has zero variation on the set E if and only if F has zero variation on the singleton sets $\{e\}$ for each $e \in E$.* □

Exercise 159 *Show that N is a measure zero set if and only if the function $F(x) = x$ has zero variation on N.* □

Exercise 160 *Suppose that both the functions F and G have zero variation on a set E. Show that so too does every linear combination $rF + sG$.* □

Exercise 161 *Suppose that both the functions F and G have zero variation on a set E. Does it follow that the product FG must have zero variation on E?* □

Exercise 162 *Show that a continuous function has variation zero on every countable set.* □

Exercise 163 *Show that a function that has variation zero on every countable set must be continuous.* □

2.8.1 Zero variation and zero derivatives

There is an intimate connection between the notion of zero variation and the fact of zero derivatives. The following two theorems are central to our theory. Note that zero derivatives imply zero variation and that, conversely, zero variation implies zero derivatives (but only almost everywhere).

Theorem 2.22 *Let $F : \mathbb{R} \to \mathbb{R}$ and suppose that $F'(x) = 0$ at every point of the set E. Then F has zero variation on E.*

Proof. Fix an integer n and write $E_n = (-n, n) \cap E$. Let $\varepsilon > 0$ and consider the collection

$$\beta = \{([u,v],w) : w \in E,\ w \in [u,v] \subset (-n,n),\ |F(v) - F(u)| < \varepsilon(v-u)\}.$$

By our assumption that $F'(x) = 0$ at every point of E we see easily that β is a full cover of E_n. But if $\pi \subset \beta$ is any subpartition we must have

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u)| < \sum_{([u,v],w)\in\pi} \varepsilon(v-u) < 2\varepsilon n.$$

This proves that F has zero variation on each set E_n. It follows from Lemma 2.21 that F has zero variation on the set E which is, evidently, the union of the sequence of sets $\{E_n\}$. ■

Theorem 2.23 *Let $F : \mathbb{R} \to \mathbb{R}$ and suppose that F has zero variation on a set E. Then $F'(x) = 0$ at almost every point of the set E.*

Proof. This theorem is deeper than the preceding and will require, for us, an appeal to our version of the Vitali covering theorem. Let N be the set of points x in E at which $F'(x) = 0$ is false. A fine covering argument allows us to analyze this. There must be *some* positive number $\varepsilon(x)$ for each $x \in N$ so that

$$\beta_1 = \{([u,v],w) : w \in E,\ |F(v) - F(u)| \geq \varepsilon(w)(v-u)\} \tag{2.16}$$

is a fine cover of N. This is how the full/fine arguments work. For, if not, then there would be some point x in E so that, for *every* $\varepsilon > 0$,

$$\beta_2 = \{([u,v],w) : w \in E,\ |F(v) - F(u)| < \varepsilon(v-u)\} \tag{2.17}$$

would have to be full at x. But that says exactly that $F'(x) = 0$. Write

$$N_i = \{w \in N : \varepsilon(w) > 1/i\}$$

for each integer i and note that N is the union of the sequence of sets $\{N_i\}$.

Fix i. Let $\eta > 0$. Since F has zero variation on E we can find a full cover β_3 of E so that

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u)| < \eta \tag{2.18}$$

whenever π is a subpartition, $\pi \subset \beta_3$. The intersection $\beta = \beta_1 \cap \beta_3$ is a fine cover of N.

For the set N_i and any subpartition $\pi \subset \beta[N_i]$ we compute, with some help from (2.16) and (2.18), that

$$\sum_{([u,v],w)\in\pi} (v-u) < \sum_{([u,v],w)\in\pi} \varepsilon(w)|F(v) - F(u)|$$

$$\leq i \sum_{([u,v],w)\in\pi} |F(v) - F(u)| < i\eta.$$

This verifies that each set N_i is a fine null set and so, by the Mini-Vitali covering theorem, also a set of Lebesgue measure zero. Consequently N itself, as the union of a sequence of Lebesgue measure zero sets, is also a set of Lebesgue measure zero. This completes the proof. ∎

2.8.2 Generalization of the zero derivative/variation

We wish to interpret this result in a much more general manner. Let h be any real-valued function that assigns values $h(([u,v],w))$ to pairs $([u,v],w))$. We can define zero variation and zero derivative for h just as easily as we can for a function $F : \mathbb{R} \to \mathbb{R}$.

If $h(I,x)$ is any function which assigns real values to interval-point pairs it will be convenient to have a notation for the following limits:

$$\limsup_{(I,x)\Longrightarrow x} h(I,x) = \inf_{\delta>0} \left(\sup\{h(I,x) : \lambda(I) < \delta,\ x \in I\}\right)$$

and

$$\liminf_{(I,x)\Longrightarrow x} h(I,x) = \sup_{\delta>0} \left(\inf\{h(I,x) : \lambda(I) < \delta,\ x \in I\}\right).$$

These are just expressions for the lower and upper limits of $h(I,x)$ as the interval I (always assumed to contain x) shrinks to the point x.

We say that h has a *zero derivative* at a point w if

$$\limsup_{(I,w)\Longrightarrow w}\left|\frac{h(I,w)}{\lambda(I)}\right| = 0.$$

This is equivalent to requiring that

$$\lim_{\delta\to 0+}\sup\left\{\left|\frac{h(([u,v],w))}{v-u}\right| : u\le w\le v,\ 0<v-u<\delta\right\} = 0.$$

We say too that h has *zero variation* on a set E if for every $\varepsilon>0$ there is a full cover β of E so that

$$\sum_{([u,v],w)\in\pi}|h(([u,v],w))|<\varepsilon$$

whenever π is a subpartition, $\pi\subset\beta$.

A repeat of the proofs just given, with minor changes, allows us to claim that Theorems 2.22 and 2.23 can be extended to these general versions:

Theorem 2.24 *If h has a zero derivative everywhere in a set E then h has zero variation on E.*

Theorem 2.25 *Zero variation for h on a set E implies h has a zero derivative almost everywhere in E.*

Exercise 164 *Show that if*

$$\limsup_{(I,x)\Longrightarrow x} h(I,x) < t$$

at every point x of a set E then

$$\{(I,x) : x\in I,\ h(I,x)<t\}$$

is a full cover of E. □

Exercise 165 *Show that if*

$$\liminf_{(I,x)\Longrightarrow x} h(I,x) < t$$

at every point x of a set E then

$$\{(I,x) : x\in I,\ h(I,x)<t\}$$

is a fine cover of E. □

2.8.3 Zero variation and mapping properties

If a function F has zero variation on a set E then necessarily the image set $F(E)$ has Lebesgue measure zero. This is an important and subtle property of zero variation and we wish to present it early in our study, even though the proof will require accepting some basic facts about Lebesgue measure that we develop only later.

Theorem 2.26 *If a function F has zero variation on a set E then the image set*
$$F(E) = \{y : y = F(x) \text{ for some } x \in E\}$$
has Lebesgue measure zero.

Proof. This is a special case of Theorem 6.7 that we prove much later. It is convenient to place it here because it clarifies the nature of zero variation. We require for the proof some basic facts about the Lebesgue measure not yet proved. For any set A of real numbers the value
$$\lambda(A) = \inf\{\lambda(G) : G \text{ open and } G \supset A\}$$
is called the Lebesgue measure of A. Certainly A is a set of Lebesgue measure zero if and only if $\lambda(A) = 0$.

We need here to know these elementary facts about Lebesgue measure. If
$$A = \bigcup_{n=1}^{\infty} A_n$$
then
$$\lambda(A) \leq \sum_{n=1}^{\infty} \lambda(A_n).$$
Moreover, if $\{A_n\}$ is an increasing sequence of sets and each $\lambda(A_n) \leq t$, then $\lambda(A) \leq t$.

To continue the proof now, let E be a set on which F has zero variation. Let $\varepsilon > 0$. There is then a full cover β of E so that
$$\sum_{([u,v],w)\in\pi} |F(v) - F(u)| < \varepsilon$$
whenever π is a subpartition contained in β.

We can apply the decomposition lemma (Lemma 2.7) to obtain an increasing sequence of set $\{E_n\}$ whose union is E as well as, for each n a sequence of nonoverlapping intervals $\{I_{kn}\}$ with the properties stated in the lemma. The main feature we need is that if x_1 and x_2 are distinct points of E_n inside an interval I_{kn} and $x_1 < x_2$ then $([x_1,x_2],x_1)$ is a member of β.

Let us make a crude estimate of the Lebesgue measure of $F(E_n \cap I_{kn})$. We just check the diameter of the set. If that diameter is not zero, then there must be distinct points x_1 and x_2 of E_n inside an interval I_{kn} with $x_1 < x_2$ such that the number
$$|F(x_1) - F(x_2)| + \varepsilon 2^{-k}$$
exceeds this diameter. We know that $([x_1,x_2],x_1)$ is a member of β. This provides us with an upper estimate for $\lambda(E_n \cap I_{kn})$ from which we deduce, using our first property of Lebesgue measure, that
$$\lambda(F(E_n)) \leq \sum_{k=1}^{\infty} \lambda(E_n \cap I_{kn}) \leq 2\varepsilon.$$

As this holds for all n it follows (again from a property of Lebesgue measure) that

$$\lambda(F(E)) \leq 2\varepsilon.$$

Finally then $\lambda(F(E)) = 0$ and we have proved that $F(E)$ is a set of Lebesgue measure zero. ■

Exercise 166 *Suppose that F is differentiable at every point of a set E and that the image set $F(E)$ is a set of Lebesgue measure zero. Show that $F'(x) = 0$ for almost every point x in E.* Answer □

2.9 Absolutely continuous functions

Our formulation of the notions of zero variation and measure zero are immediately related by the fact that the function $F(x) = x$ has zero variation on a set N precisely when that set N is a set of Lebesgue measure zero. We see, then, that $F(x) = x$ has zero variation on all sets of Lebesgue measure zero. Most functions that we have encountered in the calculus also have this property. We shall see that all differentiable functions have this property. It plays a vital role in the theory; such functions are said to be λ-absolutely continuous[3].

The basic definition is the following, with several variants given subsequently.

Definition 2.27 *A real-valued function defined on an open set containing a set E is said to be* λ-absolutely continuous *on E if F has zero variation on every subset N of E that has Lebesgue measure zero.*

Absolute continuity is stronger than continuity.

Lemma 2.28 *If a function $F : (a,b) \to \mathbb{R}$ is λ-absolutely continuous on the open interval (a,b) then F is continuous at each point of that interval.*

Proof. If F has zero variation on each Lebesgue measure zero subset of (a,b) then F has zero variation on any set $\{x_0\}$ containing a single point x_0 from that interval. If we translate what this would mean into ε, δ language we find that for every $\varepsilon > 0$ there must be a $\delta > 0$ so that

$$|F(v) - F(u)| < \varepsilon$$

if $v - u < \delta$ and $x_0 \in [u,v]$. But this is exactly the statement that F is continuous at the point x_0. ■

[3]**Note to the instructor:** we need to distinguish among several closely related notions of absolute continuity. The phrase is usually used in analysis courses in the sense that Vitali introduced (for functions) and also in the quite different measure sense (for set functions).

The exercises show that most continuous functions we encounter in the calculus will be absolutely continuous. In fact the only continuous function we might have seen so far that is not absolutely continuous is the Cantor function of elementary analysis.

2.9.1 Absolute continuity in the sense of Vitali

Historically the first notion of absolute continuity is due to G. Vitali. It is not expressed in terms of zero variation/zero measure but more in a way that is closely related to uniform continuity.

Definition 2.29 (AC in the sense of Vitali) *A function $F : [a,b] \to \mathbb{R}$ is absolutely continuous in Vitali's sense on $[a,b]$ provided that for every $\varepsilon > 0$ there is a $\delta > 0$ so that*

$$\sum_{i=1}^{n} |F(x_i) - F(y_i)| < \varepsilon$$

whenever $\{[x_i, y_i]\}$ are nonoverlapping subintervals of $[a,b]$ for which

$$\sum_{i=1}^{n} (y_i - x_i) < \delta.$$

This definition does not immediately relate to our fundamental definition of λ-absolute continuity, but the connection is easily established.

Lemma 2.30 *Suppose that $F : [a,b] \to \mathbb{R}$ is absolutely continuous in Vitali's sense on $[a,b]$. Then*

1. *F is uniformly continuous on $[a,b]$.*
2. *F has bounded variation on $[a,b]$.*
3. *F is λ-absolutely continuous on (a,b).*
4. *F maps subsets of (a,b) of Lebesgue measure zero into sets of Lebesgue measure zero.*

One should always ask, on seeing such a list of properties, whether they are sufficient as well as necessary. In this case we will be able to prove (much later) that conditions 1, 2, and 3 are sufficient and (more interesting) that conditions 1, 2, and 4 are also sufficient.

2.9.2 Proof of Lemma 2.30

It is clear just from the definitions that if F is absolutely continuous in Vitali's sense on $[a,b]$ then F is uniformly continuous on $[a,b]$.

Suppose that $F : [a,b] \to \mathbb{R}$ is absolutely continuous in Vitali's sense on $[a,b]$. We show that F has bounded variation on $[a,b]$. Select a $\delta > 0$ so that

$$\sum_{i=1}^{n} |F(x_i) - F(y_i)| < 1$$

whenever $\{[x_i, y_i]\}$ are nonoverlapping subintervals of $[a,b]$ for which

$$\sum_{i=1}^{n} (y_i - x_i) < \delta.$$

Note that if $[c,d]$ is a subinterval of $[a,b]$ with length less than δ then certainly

$$\mathrm{Var}(F,[c,d]) \leq 1.$$

Since $[a,b]$ can be expressed as the union of a finite number of subintervals of $[a,b]$ each with length less than δ, it follows (from Exercise 60) that F has bounded variation on $[a,b]$.

Finally if $F : [a,b] \to \mathbb{R}$ is absolutely continuous in Vitali's sense on $[a,b]$ and $E \subset (a,b)$ is a set of measure zero, let $\varepsilon > 0$ and choose $\delta > 0$ so that

$$\sum_{i=1}^{n} |F(x_i) - F(y_i)| < \varepsilon$$

whenever $\{[x_i, y_i]\}$ are nonoverlapping subintervals of $[a,b]$ for which

$$\sum_{i=1}^{n} (y_i - x_i) < \delta.$$

Let $G \supset E$ be a an open set with $\lambda(G) < \delta$. Define

$$\beta = \{([u,v],w) : w \in E \cap [u,v] \subset G\}.$$

This is a full cover of E. Just check that

$$\sum_{([u,v],w)\in\pi} (v-u) < \delta$$

and hence that

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u)| < \varepsilon$$

whenever π is a subpartition contained in β. It follows that F has zero variation on E. As this applies to any measure zero set it follows that F is λ-absolutely continuous on (a,b).

Finally the fourth condition of Lemma 2.30 now follows from Theorem 2.26.

2.9.3 Absolute continuity in the variational sense

As we said, historically the first notion of absolute continuity is due to G. Vitali. The connection with sets of measure zero soon emerged as the measure theory developed. In fact absolute continuity (in our sense, not in Vitali's sense) is used extensively in the general theory of measure.

Vitali's definition is limited to functions of bounded variation as Lemma 2.30 shows. The need for a notion of absolute continuity that would apply more generally to functions that do not have bounded variation led A. Denjoy to a notion that imitated the Vitali definition. That notion developed later on by S. Saks under the name "generalized absolute continuity in the restricted sense" shortened considerably by the symbols ACG$_\star$. A different (but equivalent) definition is rather more convenient.

Definition 2.31 (AC in the variational sense (ACG$_\star$)) *A function $F : [a,b] \to \mathbb{R}$ is absolutely continuous in the variational sense on $[a,b]$ provided that F is uniformly continuous on $[a,b]$ and λ-absolutely continuous on (a,b).*

We see (because of Lemma 2.30) that a function absolutely continuous in the Vitali sense on an interval $[a,b]$ is necessarily also absolutely continuous in the variational sense (i.e., is ACG$_\star$ as mathematicians of the previous century would have expressed it). To give an example that shows the converse not to be true one can just supply an everywhere differentiable function that does not have bounded variation. The next section explains why differentiable functions would have this property.

2.9.4 Absolute continuity and derivatives

There is an intimate relationship between the differentiability properties of a function and its absolute continuity properties. The first such connection is easy to make. Our lemma shows that all differentiable functions are λ-absolutely continuous.

Lemma 2.32 *Suppose that F is a real-valued function defined on an open set that contains the Lebesgue measure zero set N and that F is differentiable at every point of N. Then F is has zero variation on N.*

Proof. For each natural number n let N_n be the collection of those points x in N at which $|F'(x)| < n$. We show that F has zero variation on each N_n. It follows then that F is has zero variation on $N = \bigcup_{n=1}^{\infty} N_n$.

Let $\varepsilon > 0$. Since There must be a full cover β_1 of N so that

$$\sum_{([u,v],w)\in\pi} (v-u) < \varepsilon/n$$

whenever π is a subpartition, $\pi \subset \beta_1$. Define

$$\beta_2 = \{([u,v],w) : w \in E_n,\ |F(v)-F(u)| < n(v-u)\}.$$

This is evidently a full cover of N_n, because $|F'(w)| < n$ for each $w \in N_n$.

Consequently $\beta = \beta_1 \cap \beta_2$ is a full cover of N_n for which

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| < \sum_{([u,v],w)\in\pi} n(v-u) < \varepsilon$$

whenever π is a subpartition, $\pi \subset \beta$. This proves that F has zero variation on N_n. Since N is the union of the sequence of set N_n this proves our assertion., ∎

Corollary 2.33 *Suppose that $F : [a,b] \to \mathbb{R}$ is a uniformly continuous function that is differentiable at nearly every point of (a,b). Then F is absolutely continuous in the variational sense (i.e., ACG$_\star$) on $[a,b]$.*

Corollary 2.34 *Suppose that $F : [a,b] \to \mathbb{R}$ is a uniformly continuous function that is differentiable at every point of (a,b) with the exception possibly of points in a set N on which F has zero variation. Then F is absolutely continuous in the variational sense (i.e., ACG$_\star$) on $[a,b]$.*

Exercises

Exercise 167 *Show that the function $F(x) = x$ is λ-absolutely continuous on every open interval.* □

Exercise 168 *Show that a linear combination of λ-absolutely continuous functions is λ-absolutely continuous.* □

Exercise 169 *Suppose that $F : (a,b) \to \mathbb{R}$ is is λ-absolutely continuous on the interval (a,b). Show that F must be pointwise continuous at every point of that interval.* □

Exercise 170 *Show that a Lipschitz function defined on an open interval is λ-absolutely continuous there.* □

Exercise 171 *Give an example of an λ-absolutely continuous function that is not Lipschitz.* □

Exercise 172 *Show that the Cantor function is not λ-absolutely continuous on $(0,1)$.* □

Exercise 173 *Suppose that $F : (a,b) \to \mathbb{R}$ is differentiable at each point of the open interval (a,b). Show that F is λ-absolutely continuous on the interval (a,b).*

Answer □

Exercise 174 *Suppose that $F : (a,b) \to \mathbb{R}$ is differentiable at each point of the open interval (a,b) with countably many exceptions but that F is pointwise continuous at those exceptional points. Show that F is λ-absolutely continuous on the interval (a,b).* Answer □

Exercise 175 *Suppose that $F:(a,b)\to\mathbb{R}$ is differentiable at each point of the open interval (a,b) with the exception of a set $N\subset(a,b)$. Suppose further that N is a set of measure zero and that F has zero variation on N. Show that F is λ-absolutely continuous on the interval (a,b).* □

Exercise 176 *Suppose that $F:(a,b)\to\mathbb{R}$ is λ-absolutely continuous on the interval (a,b). Then by definition F has zero variation on every subset of Lebesgue measure zero. Is it possible that F has zero variation on subsets that are not Lebesgue measure zero?* □

Exercise 177 *A function F on an open interval I is said to have* finite derived numbers *on a set $E\subset I$ if, for each $x\in E$, there is a number M_x and one can choose $\delta>0$ so that*

$$\left|\frac{F(x+h)-F(x)}{h}\right|\leq M_x$$

whenever $x+h\in I$ and $|h|<\delta$. Show that F is λ-absolutely continuous on E if F has finite derived numbers there. □

Exercise 178 *Suppose that $F:[a,b]\to\mathbb{R}$ is Lipschitz. Show that F must be absolutely continuous in the Vitali sense on $[a,b]$. Is the converse true, i.e., if F is absolutely continuous in the Vitali sense on $[a,b]$ then is F necessarily Lipschitz.* □

2.10 An application to the Henstock-Kurzweil integral

Our first serious application of the methods of this chapter establishes that the Henstock-Kurzweil integral is equivalent to the general Newton integral. This is our second proof of this fact.

Theorem 2.35 (Properties of the indefinite integral) *Suppose that $f:[a,b]\to\mathbb{R}$ is a Henstock-Kurzweil integrable function on a compact interval $[a,b]$ and that F is an indefinite integral for f. Then*

1. *F is absolutely continuous in the variational sense (ACG$_\star$) on $[a,b]$.*
2. *$F'(x)=f(x)$ at almost every point of (a,b).*

As a corollary, by combining this theorem with Theorem 55, we obtain a second proof of the equivalence of the Henstock-Kurzweil and general Newton integrals. Recall that the first proof of that equivalence in the preceding chapter made use of the Lebesgue differentiation theorem. This one uses the theme of zero variation implies zero derivative (see the proof in the next section) along with the Henstock-Saks lemma.

Corollary 2.36 *The Henstock-Kurzweil and general Newton integrals are equivalent.*

2.10.1 Proof of Theorem 2.35

Let N be a set of measure zero contained in $[a,b]$. We show that F has zero variation on N. Let $\varepsilon > 0$. By Theorem 1.24 (the Henstock-Saks lemma) there exists a full cover β_1 of $[a,b]$ so that

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)-f(w)(v-u)| < \varepsilon/2$$

whenever π is a subpartition of the interval $[a,b]$ contained in β_1. As N has measure zero, there exists a full cover β_2 of N so that

$$\sum_{([u,v],w)\in\pi} |f(w)|\,(v-u) < \varepsilon/2$$

whenever π is a subpartition of the interval $[a,b]$ contained in β_2.

Consequently, setting β as the intersection of β_1 and β_2, we will have a full cover of N for which

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| \leq \sum_{([u,v],w)\in\pi} |f(w)|\,(v-u)$$
$$+\sum_{([u,v],w)\in\pi} |F(v)-F(u)-f(w)(v-u)| < \varepsilon/2+\varepsilon/2 = \varepsilon.$$

whenever π is a subpartition of the interval $[a,b]$ contained in β. It follows that F is absolutely continuous in the variational sense (ACG$_\star$) on $[a,b]$.

The second part of the theorem is a direct application of the material of Section 2.8.2. Define the function

$$h([u,v],w) = F(v)-F(u)-f(w)(v-u).$$

By the Saks-Henstock lemma h has zero variation on the set (a,b). By Theorem 2.25 this means that $h'(x) = 0$ at almost every point x of (a,b). But, observing that

$$\frac{h([u,v],w)}{v-u} = \frac{F(v)-F(u)}{v-u} - f(w),$$

we see that at every point w for which $h'(w) = 0$ it is true that $F'(w) = f(w)$.

2.11 Lebesgue differentiation theorem (2nd proof)

Our next application of the Mini-Vitali theorem is to give our second proof of the famous and useful theorem of Lebesgue asserting that functions of bounded variation are almost everywhere differentiable. We have already studied this theorem in Section 1.14 where a "rising sun lemma" arguement was used. Here we illustrate the use of fine covering (i.e., Vitali covering) arguments. An education in real analysis would be expected to include an exposure to the Rising Sun lemma as well as to the fine [Vitali] covering arguments that can be used to prove Lebesgue's theorem. Historians would also want to study Lebesgue's original proof.

Theorem 2.37 *Let $F : [a,b] \to \mathbb{R}$ be a function of bounded variation. Then F is differentiable at almost every point in (a,b).*

Corollary 2.38 *Let $F : [a,b] \to \mathbb{R}$ be a monotonic function. Then F is differentiable at almost every point in (a,b).*

The proof of the theorem will require an introduction, first, to the upper and lower derivates and then a simple geometric lemma that allows us to use a fine covering argument to show that the set of points where $F'(x)$ does not exist is Lebesgue measure zero.

2.11.1 Upper and lower derivates

The proof uses the *upper and lower derivates*. To analyze how a derivative $F'(x)$ may fail to exist we split that failure into two pieces, an upper and a lower, defined as

$$\overline{D}F(x) = \inf_{\delta>0} \sup \left\{ \frac{F(v)-F(u)}{v-u} : x \in [u,v],\ 0 < v-u < \delta \right\}$$

and

$$\underline{D}F(x) = \sup_{\delta>0} \inf \left\{ \frac{F(v)-F(u)}{v-u} : x \in [u,v],\ 0 < v-u < \delta \right\}$$

We will prove that, for almost every point x in (a,b),

$$\overline{D}F(x) > -\infty, \qquad \underline{D}F(x) < \infty,$$

and

$$\overline{D}F(x) = \underline{D}F(x).$$

From these three assertions it follows that F has a finite derivative $F'(x)$ at almost every point x in (a,b).

The proof will depend on a fine covering argument. For that we need to recognize the following connection between derivates and covers. The proof is trivial; it is only a matter of interpreting the statements.

Lemma 2.39 *Let $F : [a,b] \to \mathbb{R}$, $\alpha \in \mathbb{R}$, and let*

$$\beta = \left\{ ([u,v],w) : \frac{F(v)-F(u)}{v-u} > \alpha,\ w \in [u,v] \subset [a,b] \right\}.$$

Then, β is a full cover of the set

$$E_1 = \{x \in (a,b) : \underline{D}F(x) > \alpha\}$$

and a fine cover of the larger set

$$E_2 = \{x \in (a,b) : \overline{D}F(x) > \alpha\}.$$

2.11.2 Geometrical lemmas

The proof employs an elementary geometric lemma that Donald Austin [2][4] used in 1965 to give a simple proof of this theorem. Our proof of the differentiation theorem is essentially his, but written in different language. See also the version of Michael Botsko [7][5].

Lemma 2.40 (Austin's lemma) *Let* $G : [a,b] \to \mathbb{R}$, $\alpha > 0$ *and suppose that* $G(a) \leq G(b)$. *Let*

$$\beta = \left\{ ([u,v],w) : \frac{G(v)-G(u)}{v-u} < -\alpha,\ w \in [u,v] \subset [a,b] \right\}.$$

Then, for any nonempty subpartition $\pi \subset \beta$,

$$\alpha \left(\sum_{([u,v],w)\in\pi} (v-u) \right) < \mathrm{Var}(G,[a,b]) - |G(b)-G(a)|.$$

Proof. To prove the lemma, let π_1 be a partition of $[a,b]$ that contains the subpartition π. Just write

$$|G(b)-G(a)| = G(b)-G(a) = \sum_{([u,v],w)\in\pi_1} [G(v)-G(u)]$$

$$= \sum_{([u,v],w)\in\pi} [G(v)-G(u)] + \sum_{([u,v],w)\in\pi_1\setminus\pi} [G(v)-G(u)]$$

$$< -\alpha \left(\sum_{([u,v],w)\in\pi} [v-u] \right) + \mathrm{Var}(G,[a,b]).$$

The statement of the lemma follows. ∎

As a corollary we can replace F with $-F$ to obtain a similar statement.

Corollary 2.41 *Let* $G : [a,b] \to \mathbb{R}$, $\alpha > 0$ *and suppose that* $G(b) \leq G(a)$. *Let*

$$\beta = \left\{ ([u,v],w) : \frac{G(v)-G(u)}{v-u} > \alpha,\ w \in [u,v] \subset [a,b] \right\}.$$

Then, for any nonempty subpartition $\pi \subset \beta$,

$$\alpha \left(\sum_{([u,v],w)\in\pi} (v-u) \right) < \mathrm{Var}(G,[a,b]) - |G(b)-G(a)|.$$

---

[4] D. Austin, *A geometric proof of the Lebesgue differentiation theorem.* Proc. Amer. Math. Soc. 16 (1965) 220–221.

[5] M. W. Botsko, *An elementary proof of Lebesgue's differentiation theorem*. Amer. Math. Monthly 110 (2003), no. 9, 834–838.

2.11.3 Proof of the Lebesgue differentiation theorem

We now prove the theorem. The first step in the proof is to show that at almost every point t in (a,b),

$$\underline{D}F(t) = \overline{D}F(t).$$

If this is not true then there must exist a pair of rational numbers r and s for which the set

$$E_{rs} = \{t \in (a,b) : \underline{D}F(t) < r < s < \overline{D}F(t)\}$$

is not a set of Lebesgue measure zero. This is because the union of the countable collection of sets E_{rs} contains all points t for which $\underline{D}F(t) \neq \overline{D}F(t)$.

Let us show that each such set E_{rs} is fine null. By the Mini-Vitali theorem we then know that E_{rs} is a set of Lebesgue measure zero. Write $\alpha = (s-r)/2$, $B = (r+s)/2$, $G(t) = F(t) - Bt$. Note that

$$E_{rs} = \{t \in (a,b) : \underline{D}G(t) < -\alpha < 0 < \alpha < \overline{D}G(t)\}.$$

Since F has bounded variation on $[a,b]$, so too does the function G. In fact

$$\mathrm{Var}(G,[a,b]) \leq \mathrm{Var}(F[a,b]) + B(b-a).$$

Let $\varepsilon > 0$ and select points

$$a = s_0 < s_1 < \cdots < s_{n-1} < s_n = b$$

so that

$$\sum_{i=1}^{n} |G(s_i) - G(s_{i-1})| > \mathrm{Var}(G,[a,b]) - \alpha\varepsilon.$$

Let $E'_{rs} = E_{rs} \setminus \{s_1, s_2, \ldots, s_{n-1}\}$. Let us call an interval $[s_{i-1}, s_i]$ *black* if $G(s_i) - G(s_{i-1}) \geq 0$ and call it *red* if $G(s_i) - G(s_{i-1}) < 0$.

For each $i = 1,2,3,\ldots,n$ we define a covering relation β_i as follows. If $[s_{i-1}, s_i]$ is a black interval then

$$\beta_i = \left\{([u,v],w) : \frac{G(v)-G(u)}{v-u} < -\alpha,\ w \in [u,v] \subset [s_{i-1}, s_i]\right\}.$$

If, instead, $[s_{i-1}, s_i]$ is a red interval then

$$\beta_i = \left\{([u,v],w) : \frac{G(v)-G(u)}{v-u} > \alpha,\ w \in [u,v] \subset [s_{i-1}, s_i]\right\}.$$

Let $\beta = \bigcup_{i=1}^{n} \beta_i$. Because of Lemma 2.39 we see that this collection β is a fine cover of E'_{rs}.

Let π be any nonempty subpartition contained in β. Write

$$\pi_i = \pi \cap \beta_i.$$

By Lemma 2.40 applied to the black intervals and Corollary 2.41 applied to the red intervals we obtain that

$$\alpha\left(\sum_{([u,v],w)\in\pi_i} (v-u)\right) < \mathrm{Var}(G,[s_{i-1},s_i]) - |G(s_i) - G(s_{i-1})|.$$

Consequently

$$\alpha\left(\sum_{([u,v],w)\in\pi}(v-u)\right)=\alpha\left(\sum_{i=1}^{n}\sum_{([u,v],w)\in\pi_i}(v-u)\right)$$
$$\leq\sum_{i=1}^{n}\operatorname{Var}(G,[s_{i-1},s_i])-\sum_{i=1}^{n}|G(s_i-G(s_{i-1})|$$
$$\leq\operatorname{Var}(G,[a,b])-[\operatorname{Var}(G,[a,b])-\alpha\varepsilon]=\alpha\varepsilon.$$

We have proved that β is a fine cover of E'_{rs} with the property that

$$\sum_{([u,v],w)\in\pi}(v-u)<\varepsilon$$

for every subpartition $\pi\subset\beta$. It follows that E'_{rs} is fine null, and hence a set of Lebesgue measure zero. So too then is E_{rs} since the two sets differ by only a finite number of points.

We know now that the function F has a derivative, finite or infinite, almost everywhere in (a,b). We wish to exclude the possibility of the infinite derivative, except on a set of Lebesgue measure zero.

Let

$$E_\infty=\{t\in(a,b):\underline{D}F(t)=\infty\}.$$

Choose any B so that $F(b)-F(a)\leq B(b-a)$ and set $G(t)=F(t)-Bt$. Note that $G(b)\leq G(a)$ which will allow us to apply Corollary 2.41.

Let $\varepsilon>0$ and choose a positive number α large enough so that

$$\operatorname{Var}(G,[a,b])-|G(b)-G(a)|<\alpha\varepsilon.$$

Define

$$\beta=\left\{([u,v],w):\frac{G(v)-G(u)}{v-u}>\alpha,[u,v]\subset[a,b]\right\}.$$

This is a fine cover of E_∞. Let π be any subpartition $\pi\subset\beta$. By our corollary then

$$\alpha\sum_{([u,v],w)\in\pi}(v-u)<\operatorname{Var}(G,[a,b])-|G(b)-G(a)|<\alpha\varepsilon.$$

We have proved that β is a fine cover of E_∞ with the property that

$$\sum_{([u,v],w)\in\pi_i}(v-u)<\varepsilon$$

for every subpartition $\pi\subset\beta$. It follows that E_∞ is fine null, and hence a set of Lebesgue measure zero. The same arguments will handle the set

$$E_{-\infty}=\{t\in(a,b):\overline{D}F(t)=-\infty\}.$$

2.11.4 Fubini differentiation theorem

The formula
$$\frac{d}{dx}\sum_{n=1}^{\infty}F_n(x)=\sum_{n=1}^{\infty}\frac{d}{dx}F_n(x)$$
is not generally valid without assumptions about uniform convergence, Fubini's differentiation theorem says that, with some assumptions on the nature of the functions F_n, we can have this differentiation formula, not everywhere, but almost everywhere. Prove this occasionally useful results as an application of the Lebesgue differentiation theorem:

Theorem 2.42 (Fubini) *Let $\{F_n\}$ be a sequence of monotonic, nondecreasing functions on the interval $[a,b]$ and suppose that $F(x)=\sum_{n=1}^{\infty}F_n(x)$ is absolutely convergent for all $a\le x\le b$. Then, for almost every x in (a,b),*
$$F'(x)=\sum_{n=1}^{\infty}F_n'(x).$$

Proof. Our main tool, apart from ordinary computations, is the fact that monotonic functions are differentiable almost everywhere. This is proved in Theorem 2.37.

Let us simplify the proof by deciding that $F_n(a)=0$ for all n, so that F and all functions F_n are nonnegative. We know from the Lebesgue differentiation theorem applied to all of these monotonic functions that, except for x in a set of Lebesgue measure zero, all of the derivatives, $F'(x)$ and $F_n'(x)$ exist. Thus it is only the identity for these values of x that we need to establish.

Note that
$$F(x)\ge\sum_{n=1}^{m}F_n(x)$$
for every integer m so that for almost every x,
$$F'(x)\ge\sum_{n=1}^{m}F_n'(x)$$
and, consequently,
$$F'(x)\ge\sum_{n=1}^{\infty}F_n'(x). \tag{2.19}$$

To simplify we can assume that
$$F(b)-\sum_{n=1}^{m}F_n(b)\le 2^{-m}.$$
If this were not the case then we could put parentheses in the series, group terms together, and relabel so that this would be the case. Consider the series
$$G(x)=\sum_{n=1}^{\infty}\left(F(x)-\sum_{k=1}^{n}F_k(x)\right).$$

Note that

$$0 \le G(x) - \sum_{n=1}^{m} \left(F(x) - \sum_{k=1}^{n} F_k(x) \right) \le \sum_{n=m+1}^{\infty} 2^{-n} = 2^{-m}.$$

Thus we see that G is also the sum of a series of functions.

A repeat of the argument we just gave to establish (2.19) will provide the analogous statement for this series:

$$0 \le \sum_{n=1}^{\infty} \left(F'(x) - \sum_{k=1}^{n} F_k'(x) \right) \le G'(x) \tag{2.20}$$

The function G has a finite derivative at almost every point. So in order for the inequality in (2.20) to hold for this series at a particular value of x the terms must tend to zero. Writing that out we now know that, for almost every x,

$$\lim_{n\to\infty} \left(F'(x) - \sum_{k=1}^{n} F_k'(x) \right) = 0.$$

This is exactly the conclusion of the theorem. ∎

2.12 An application to the Riemann integral

What are necessary and sufficient conditions in order that a function may be an indefinite integral in the Riemann sense? The problem was explicitly posed in a short note [80] published by Erik Talvila in 2008. A solution was given in [84]. As an application of the material in this chapter we shall present a solution here.

While we are not so much interested in the Riemann integral for its own sake, we are interested in the methods of integration theory. This proof illustrates those methods and is a good illustration of how powerful the techniques of this chapter can be.

Theorem 2.43 *A necessary and sufficient condition in order for a function $F : [a,b] \to \mathbb{R}$ to be representable in the form*

$$F(x) = C + \int_a^x f(t)\,dt \quad (a \le x \le b)$$

for some constant C and for some Riemann integrable function f on $[a,b]$ is that, for all $\varepsilon > 0$, a positive function $\delta : [a,b] \to \mathbb{R}^+$ can be found so that

$$\sum_{i=1}^{n} \left| \frac{F(\xi_i) - F(x_{i-1})}{\xi_i - x_{i-1}} - \frac{F(x_i) - F(\xi_i')}{x_i - \xi_i'} \right| (x_i - x_{i-1}) < \varepsilon \tag{2.21}$$

for every choice of partition

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

that is finer than δ and every choice of associated points $x_{i-1} < \xi_i \le \xi_i' < x_i$.

Proof. For the proof that the condition is necessary let us suppose that F is the indefinite integral of a Riemann integrable function f. Let $\varepsilon > 0$ and choose $\delta > 0$ so that

$$\sum_{i=1}^{n} \omega_f([x_{i-1}, x_i])\,(x_i - x_{i-1}) < \varepsilon$$

for every subdivision $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ that is finer than δ. Here, as always,

$$\omega_f([c,d]) = \sup\{|f(x) - f(y)| : x, y \in [c,d]\}$$

is used to denote the oscillation of the function f on a closed interval $[c,d]$. Since f is Riemann integrable this is possible (indeed it is one of Riemann's own characterizations of integrability, equivalent to that given in Section 1.10.1 as "Riemann's criterion").

Observe that, if $s \le f(x) \le t$ on an interval $[c,d]$, then

$$s - t \le \frac{F(\xi) - F(c)}{\xi - c} - \frac{F(d) - F(\xi')}{d - \xi'} \le t - s$$

for every $c < \xi \le \xi' < d$. It follows that

$$\left| \frac{F(\xi) - F(c)}{\xi - c} - \frac{F(d) - F(\xi')}{d - \xi'} \right| \le \omega_f([c,d]).$$

Consequently, using a subdivision $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ that is finer than δ,

$$\sum_{i=1}^{n} \left| \frac{F(\xi_i) - F(x_i)}{\xi_i - x_i} - \frac{F(x_i) - F(\xi_i')}{x_i - \xi_i'} \right| (x_i - x_{i-1})$$

$$\le \sum_{i=1}^{n} \omega_f([x_{i-1}, x_i])\,(x_i - x_{i-1}) < \varepsilon$$

proving (2.21) for any choice of associated points $x_{i-1} < \xi_i \le \xi_i' < x_i$.

In the opposite direction we suppose $\varepsilon > 0$ and that $\delta > 0$ has been chosen so that the condition (2.21) is satisfied for such subdivisions.

First we claim that F is Lipschitz. We note that F must be bounded, even continuous, otherwise the condition (2.21) is easily violated. Suppose then that $|F(x)| < K$ for all $x \in [a,b]$.

Fix a number $0 < t < \delta$. We work in the interval $[a, b-t]$. For any $x \in [a, b-t]$ we use the interval $[x, x+t]$ and observe, for any $0 < h < t/2$, that

$$\left| \frac{F(x+h) - F(x)}{h} - \frac{F(x+t) - F(x+t/2)}{t/2} \right| (x + t - x) < \varepsilon$$

because of the condition (2.21). Consequently

$$\left| \frac{F(x+h) - F(x)}{h} \right| < \frac{4K + \varepsilon}{t}.$$

This imposes a bound on all the right-hand derived numbers of the continuous function F in the interval $[a, b-t]$. It follows that this bound also serves as a

Lipschitz constant for F in $[a,b-t]$. By identical arguments, working on the left side, we can show that this same bound is a Lipschitz constant for F on the interval $[a+t,b]$. It follows that F is Lipschitz on $[a,b]$.

Since F is Lipschitz the derivative $F'(x)$ is a bounded function that exists at all points x in a set D for which $[a,b]\setminus D$ has Lebesgue measure zero.

We define $f(x)=F(x)$ for $x\in D$ and, at points x not in D, we write

$$f(x)=\inf_{t>0}\sup\{F'(y):y\in D,\ |x-y|<t\}.$$

We shall now prove that f is continuous at almost every point of D and hence at almost every point of $[a,b]$. It is certainly bounded since F' is bounded by the Lipschitz constant for F. Then, by Lebesgue's criterion (see Section 1.10.1), we will know that f is Riemann integrable on $[a,b]$.

Let $\omega_f(x)$ denote the oscillation of the function f at a point x, i.e.,

$$\omega_f(x)=\inf_{t>0}\sup\{|f(x+h)-f(x)|:x+h\in[a,b],\ |h|<t\}.$$

The function f is continuous at a point x if and only if $\omega_f(x)=0$. Thus the collection of discontinuity points of f can be expressed as the union of an increasing sequence of sets $\{E_m\}$ where

$$E_m=\{x\in[a,b]:\omega_f(x)>1/m\}\quad(m=1,2,3,\dots).$$

We show that each $|E_m|=0$, i.e., that each is a set of Lebesgue measure zero.

For each $x\in D\cap E_m$ we may choose a sequence of nonzero numbers $h_n\to 0$ so that

$$|f(x+h_n)-f(x)|\geq 1/(2m).$$

By the way in which f was defined we may select these points so that $x+h_n$ are in D.

Thus for each point x that is in $D\cap E_m$ we may define β to be the collection of all the interval-point pairs of the form $([x,y],x)$ or $([y,x],x)$ with $|x-y|$ smaller than δ and for which $y\in D$ and

$$|f(y)-f(x)|\geq 1/(2m).$$

This β must form a fine cover of $D\cap E_m$.

We intend to apply the Mini-Vitali theorem. Take any subpartition from β; this corresponds to a nonoverlapping collection of intervals $[x_1,y_1]$, $[x_2,y_2]$, …, $[x_p,y_p]$ chosen from this fine cover. This will have the property that $0<y_k-x_k<\delta$ and

$$|f(y_k)-f(x_k)|\geq 1/(2m)\quad(k=1,2,\dots,p).$$

For each $k=1,2,\dots,p$ select points ξ_k, ξ_k' with $x_k<\xi_k\leq\xi_k'<y_k$ in such a way that

$$\left|\frac{F(\xi_k)-F(x_k)}{\xi_k-x_k}-F'(x_k)\right|<\varepsilon$$

and
$$\left|\frac{F(y_k)-F(\xi_k')}{y_k-\xi_k'}-F'(y_k)\right|<\varepsilon.$$

Now observe that
$$\frac{1}{2m}(y_k-x_k)\le|f(y_k)-f(x_k)|(y_k-x_k)\le$$
$$\left|\frac{F(\xi_k)-F(x_k)}{\xi_k-x_k}-F'(x_k)\right|(y_k-x_k)+\left|\frac{F(y_k)-F(\xi_k)}{y_k-\xi_k'}-F'(y_k)\right|(y_k-x_k)$$
$$+\left|\frac{F(\xi_k)-F(x_k)}{\xi_k-x_k}-\frac{F(y_k)-F(\xi_k)}{y_k-\xi_k'}\right|(y_k-x_k).$$

But
$$\sum_{k=1}^{p}\left|\frac{F(\xi_k)-F(x_k)}{\xi_k-x_k}-\frac{F(y_k)-F(\xi_k)}{y_k-\xi_k'}\right|(y_k-x_k)<\varepsilon$$
by the assumed condition (2.21). (This isn't a full subdivision of $[a,b]$ but the sum remains smaller than ε.)

The other inequalities we have imposed then show that
$$\sum_{k=1}^{p}(y_k-x_k)\le(2m)\varepsilon[1+2(b-a)].$$
This argument works for any such subpartition chosen from this fine cover β. Since $\varepsilon>0$ is arbitrary we can claim, by the Mini-Vitali theorem, that
$$D\cap E_m$$
is a set of Lebesgue measure zero for each m. Thus the set of discontinuities of f in D has been expressed as the union of a sequence of sets of measure zero.

Thus we have proved that f is bounded and a.e. continuous on $[a,b]$. By Lebesgue's criterion (see Section 1.10.1 again) the function f is Riemann integrable. The indefinite integral of f can differ from F only by a constant. It follows that
$$F(x)=C+\int_a^x f(t)\,dt \quad (a\le x\le b) \tag{2.22}$$
for some constant C, f f is Riemann integrable and the representation in (2.22) can be interpreted in the Riemann sense. ■

2.12.1 Riesz's problem

Theorem 2.43 just proved is closely related to a similar problem solved by Riesz [70] for functions of bounded variation and can be deduced from it with some extra work (left to the reader).

Theorem 2.44 (Riesz) *In order for a function $F : [a,b] \to \mathbb{R}$ to be represented in the form*

$$F(x) = C + \int_a^x f(t)\,dt \quad (a \le x \le b)$$

for some constant C and for some function f that has bounded variation on $[a,b]$ it is necessary and sufficient that there is a constant K so that

$$\sum_{i=1}^{n} \left| \frac{F(\xi_i) - F(x_{i-1})}{\xi_i - x_{i-1}} - \frac{F(x_i) - F(\xi_i')}{x_i - \xi_i'} \right| \le K \tag{2.23}$$

for every subdivision $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ and every choice of points $x_{i-1} < \xi_i \le \xi_i' < x_i$.

This property has been labeled *bounded slope variation* and has received some attention by later authors. This is more often expressed by placing a bound on the sums

$$\sum_{i=1}^{n-1} \left| \frac{F(x_{i+1}) - F(x_i)}{x_{i+1} - x_i} - \frac{F(x_i) - F(x_{i-1})}{x_i - x_{i-1}} \right| \tag{2.24}$$

but the equivalent formulation here makes many computations more transparent. For details connecting the two expressions (2.23) and (2.24), see Ene [27, p. 719].

Note that our condition (2.21) in Theorem 2.43 is easily implied by the stronger condition expressed in (2.23) here. Thus, in particular, we already know that a function F satisfying this stronger condition is the indefinite Riemann integral of a function f constructed in the proof there. Just prove now that f has bounded variation on $[a,b]$.

2.12.2 Other variants

Theorem 2.43 and Theorem 2.44 belong to a collection of problems, each of which asks for necessary and sufficient conditions for a given function $F : [a,b] \to \mathbb{R}$ to be expressed as

$$F(x) = C + \int_a^x f(t)\,dt \quad (a \le x \le b)$$

for some constant C and for some function f belonging to a given class of functions. The most famous would be for the class of Lebesgue integrable functions: the answer is that F should be absolutely continuous in the sense of Vitali, a result which is (as one might expect) due to Vitali. We will return to this kind of problem later on in Section 4.15.

Let us complete our discussion with two more examples, left as exercises.

Exercise 179 *Show that a necessary and sufficient condition in order for a function $F : [a,b] \to \mathbb{R}$ to be representable in the form*

$$F(x) = C + \int_a^x f(t)\,dt \quad (a \le x \le b)$$

for some constant C and for some continuous function f on $[a,b]$ is that, for all $\varepsilon > 0$ and $\kappa > 0$, a positive δ can be found so that

$$\sum_{i=1}^{n} \left| \frac{F(\xi_i) - F(x_{i-1})}{\xi_i - x_{i-1}} - \frac{F(x_i) - F(\xi_i')}{x_i - \xi_i'} \right| |x_i - x_{i-1}| < \varepsilon \tag{2.25}$$

for every collection of points

$$a = x_0, x_1, x_2, \dots, x_n = b$$

from $[a,b]$ for which $\sum_{i=1}^{n} |x_i - x_{i-1}| < \kappa$ and $0 < |x_i - x_{i-1}| < \delta$ and for every choice of associated points ξ_i and ξ_i' for which

$$x_{i-1} < \xi_i \le \xi_i' < x_i \text{ or } x_{i-1} > \xi_i \ge \xi_i' > x_i.$$

□

Exercise 180 (F. Riesz) *Let $1 < p < \infty$. Show that a necessary and sufficient condition in order for a function $F : [a,b] \to \mathbb{R}$ to be representable in the form*

$$F(x) = C + \int_a^x f(t)\,dt \quad (a \le x \le b)$$

for some constant C and for some function f on $[a,b]$ with the property that $|f|^p$ is integrable is that, for some constant K and for every collection of points

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b,$$

the inequality

$$\sum_{i=1}^{n} \frac{|F(x_i) - F(x_{i-1})|^p}{(x_i - x_{i-1})^{p-1}} \le K \tag{2.26}$$

must hold. Answer □

Chapter 3

The Integral

We have already introduced a number of variants for an integration theory on the real line. Of these the broadest theory available is that of the general Newton integral or (equivalently) the Henstock-Kurzweil integral. The simpler versions of Newton's integral, as well as the Riemann integral, and the Lebesgue integral are all contained in these equivalent theories.

We shall take as our definition of the integral a simple version of the Henstock-Kurzweil theory. This integral has historically been known under different names depending on the choice of definition: Denjoy's restricted integral, Perron's integral, the Denjoy-Perron integral, the Riemann-complete integral, the Kurzweil integral, the Henstock-Kurzweil integral, and the gage integral. Mostly we call it simply *the integral*.

The theory is formal, not constructive. The Riemann integral is constructive, but too restricted for an adequate theory of integration. The Lebesgue theory (as we shall see in Chapter 4) is constructive too and is quite adequate even if not quite as general. Our viewpoint on the Lebesgue integral shall be that it is contained as a special case of the integral here presented. As part of that viewpoint we accept the responsibility to develop the constructive aspects of the theory. Thus we will return to Lebesgue's original presentation, but backwards.

Our development should prove a little easier than the usual introductions to that integral. Lebesgue's definition of an integral requires an extensive development of the underlying measure theory first. Then it is necessary to prove all properties of the integral using that definition. Since we already have an integration theory developed, we need only check that Lebesgue's methods can be used to construct the value of the integral.

3.1 Upper and lower integrals

The integral is studied by means of an upper and a lower integral. This is a useful way to develop the theory and so we can leave the Henstock-Kurzweil definition behind us for a moment and start the theory of this integral in this way. This notion of using upper and lower integrals goes back at least to 1875 and is due to Jean-Gaston Darboux (1842–1917).

Definition 3.1 *For a function $f:[a,b]\to\mathbb{R}$ we define an upper integral by*

$$\overline{\int_a^b} f(x)\,dx = \inf_{\beta}\sup_{\pi\subset\beta}\left\{\sum_{([u,v],w)\in\pi} f(w)(v-u)\right\}$$

where the supremum is taken over all partitions π of $[a,b]$ contained in β, and the infimum over all full covers β of the interval $[a,b]$.

Note that the first step is to estimate the largest possible value for the Riemann sums for partitions π of $[a,b]$ contained in β, and the second step is to refine this by shrinking to smaller and smaller full covers β.

Similarly we define a lower integral as

$$\underline{\int_a^b} f(x)\,dx = \sup_{\beta}\inf_{\pi\subset\beta}\left\{\sum_{([u,v],w)\in\pi} f(w)(v-u)\right\}$$

where, again, π is a partition of $[a,b]$ and β is a full cover.

Exercises

Exercise 181 *Check that*

$$\underline{\int_a^b} f(x)\,dx = -\overline{\int_a^b}[-f(x)]\,dx.$$

□

Exercise 182 *Let $f:[a,b]\to\mathbb{R}$. Show that*

$$\underline{\int_a^b} f(x)\,dx \le \overline{\int_a^b} f(x)\,dx.$$

Answer □

Exercise 183 *Show that a function f can be altered at a finite number of points without changing the values of the upper and lower integrals.* □

Exercise 184 *Show that a function f can be altered at a countable number of points without changing the values of the upper and lower integrals.* □

Exercise 185 *Show that a function f can be altered at the points of a set of measure zero without changing the values of the upper and lower integrals.*

Answer □

Exercise 186 *Let $f:[a,c]\to\mathbb{R}$ and suppose that $a<b<c$. Show that*

$$\overline{\int_a^c} f(x)\,dx = \overline{\int_a^b} f(x)\,dx + \overline{\int_b^c} f(x)\,dx,$$

assuming the sum makes sense. Answer □

Exercise 187 *Define a partition π to be* endpointed *if only elements of the form $([u,w],w)$ or $([w,v],w)$ appear and there is no element $([u,v],w)\in\pi$ for which $u<w<v$. Show that a restriction in the definition of integrals to use endpointed partitions only would not change the theory at all.* Answer □

3.1.1 The integral and integrable functions

If the upper and lower integrals are identical we write the common value as

$$\int_a^b f(x)\,dx = \underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx$$

allowing finite or infinite values. We say in this case that the integral is *determined*. When the integral is not determined then it will be the case that

$$\underline{\int_a^b} f(x)\,dx < \overline{\int_a^b} f(x)\,dx$$

and there is no integral (although one can work with the lower and upper integrals separately).

If the integral is determined and this value is also finite then f is said to be *integrable* and

$$\int_a^b f(x)\,dx$$

is called the *integral*, now assuming a finite value. Our first goal will be to check that this account is equivalent to the usual presentation of the Henstock-Kurzweil integral that we offered in Chapter 1.

Exercises

Exercise 188 *Let $f:[a,b]\to\mathbb{R}$ show that a sufficient condition for f to be integrable on $[a,b]$ with $c=\int_a^b f(x)\,dx$ is that for every $\varepsilon>0$ there is a full cover so that*

$$\left| c - \sum_{([u,v],w)\in\pi} f(w)(v-u)\right| < \varepsilon$$

for every partition π of $[a,b]$ contained in β. Answer □

Exercise 189 *Let $f : [a,b] \to \mathbb{R}$ be an integrable function and let π be any partition of $[a,b]$. Show that*

$$\left| \int_a^b f(x)\,dx - \sum_{([u,v],w)\in\pi} f(w)(v-u) \right| \leq \sum_{([u,v],w)\in\pi} \omega f([u,v])\lambda([u,v]).$$

Here $\omega f(I)$ denotes the oscillation *of the function f on the interval I, defined as*

$$\sup_{s,t\in I} |f(s)-f(t)|.$$

□

3.1.2 HK criterion

Our first criterion for integrability returns us to Definition 1.22 and shows that the upper/lower integral approach is equivalent to the original Henstock-Kurzweil definition.

Theorem 3.2 *Let $f : [a,b] \to \mathbb{R}$. A necessary and sufficient condition in order for*

$$-\infty < \underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx < \infty$$

is that there is a number I so that for all $\varepsilon > 0$ a full cover β of $[a,b]$ can be found so that

$$\left| \sum_{([u,v],w)\in\pi} f(w)(v-u) - I \right| < \varepsilon$$

for all partitions π of $[a,b]$ contained in β. In that case then, necessarily,

$$I = \underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx.$$

Proof. In Exercise 188 we checked that this condition is sufficient. On the other hand, if we know that f is integrable with $I = \int_a^b f(x)\,dx$ then, using the definition of the upper integral, for any $\varepsilon > 0$ we choose a full cover β_1 so that

$$\sum_{([u,v],w)\in\pi} f(w)(v-u) < I + \varepsilon$$

for all partitions π of $[a,b]$ contained in β_1. Similarly, using the definition of the lower integral, we choose a full cover β_2 so that

$$\sum_{([u,v],w)\in\pi} f(w)(v-u) > I - \varepsilon$$

for all partitions π of $[a,b]$ contained in β_2. Take $\beta = \beta_1 \cap \beta_2$. This is a full cover with the property stated. ■

At this stage we now know that a function is integrable (in the sense of this chapter) if and only if it is integrable in the general Newton sense or in the sense of the Henstock-Kurzweil definition. This means that many of the properties of the integral are immediate for us in view of our studies in Chapters 1 and 2. For example the fundamental Henstock-Saks criterion (see Section 1.11.2) can be used as well as the fact that the relation

$$F(t) = \int_a^t f(x)\,dx \quad (a \leq t \leq b)$$

holds precisely when $F(a) = 0$ and $F'(t) = f(t)$ for every t in $[a,b]$ excepting only for t in a set $N \subset [a,b]$ that has measure zero and on which F has zero variation. Had we started instead with this chapter we would have to return to our first two chapters in order to complete the theory of the integral with these details.

3.1.3 Cauchy criterion

As in nearly all theories concerning limits there is a "Cauchy" criterion available. This should be compared to the closely related one for the Riemann integral discussed in Section 1.10.1.

Theorem 3.3 (Cauchy criterion) *A necessary and sufficient condition in order for a function $f : [a,b] \to \mathbb{R}$ to be integrable on a compact interval $[a,b]$ is that, for all $\varepsilon > 0$, a full cover β of $[a,b]$ can be found so that*

$$\left| \sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} [f(w) - f(w')]\lambda(I \cap I') \right| < \varepsilon \tag{3.1}$$

for all partitions π, π' of $[a,b]$ contained in β.

Proof. Start by checking that when π and π' are both partitions of the same interval $[a,b]$ then, for any subinterval I of $[a,b]$

$$\lambda(I) = \sum_{(I',w')\in\pi'} \lambda(I \cap I')$$

from which it is easy to see that

$$\sum_{(I,w)\in\pi} f(w)\lambda(I) = \sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} f(w)\lambda(I \cap I').$$

This allows the difference that would normally appear in a Cauchy type criterion

$$\left| \sum_{(I,w)\in\pi} f(w)\lambda(I) - \sum_{(I',w')\in\pi'} f(w')\lambda(I') \right|$$

to assume the simple form given in (3.1). In particular that statement can be rewritten as

$$\left| \sum_{(I,w)\in\pi} f(w)\lambda(I) - \sum_{(I',w')\in\pi'} f(w)\lambda(I) \right| < \varepsilon. \tag{3.2}$$

The condition is necessary. For if f is integrable then the first Cauchy criterion supplies a full cover β so that

$$\left|\sum_{(I,w)\in\pi} f(w)\lambda(I) - c\right| < \varepsilon/2$$

for all partitions π of $[a,b]$ contained in β. Any two Riemann sums would both be this close to c and hence within ε of each other.

Suppose the condition holds. We can see from (3.2) that the upper and lower integrals must be finite. We wish to show that they are equal.

Using the definition of the upper integral, there is at least one partition π of $[a,b]$ contained in β with

$$\sum_{(I,w)\in\pi} f(w)\lambda(I) > \overline{\int_a^b} f(x)\,dx - \varepsilon$$

Using the definition of the lower integral, there is at least one partition π' of $[a,b]$ contained in β with

$$\sum_{(I,w)\in\pi'} f(w)\lambda(I) < \underline{\int_a^b} f(x)\,dx + \varepsilon.$$

Together with (3.2) these show that

$$\overline{\int_a^b} f(x)\,dx - \underline{\int_a^b} f(x)\,dx < 2\varepsilon.$$

Since ε is an arbitrary positive number the upper and lower integrals are equal. ■

3.1.4 McShane's criterion

A nearly imperceptible change in the statement of the Cauchy criterion for integrability strengthens the property in such a way as to characterize the Lebesgue integral.

Definition 3.4 (McShane's criterion) *A function $f : [a,b] \to \mathbb{R}$ is said to* satisfy McShane's criterion *on $[a,b]$ provided that for all $\varepsilon > 0$ a full cover β can be found so that*

$$\sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} \left|f(w) - f(w')\right| \lambda(I\cap I') < \varepsilon$$

for all partitions π, π' of $[a,b]$ contained in β.

Theorem 3.5 *If a function $f : [a,b] \to \mathbb{R}$ satisfies McShane's criterion then f is absolutely integrable on $[a,b]$ (i.e., both f and $|f|$ are integrable on $[a,b]$).*

Note: the converse is proved later on.

Proof. Just check that the expression in Theorem 3.3

$$\left| \sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} [f(w) - f(w')]\lambda(I\cap I') \right| < \varepsilon \tag{3.3}$$

is smaller than the expression

$$\sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} \left|f(w) - f(w')\right| \lambda(I\cap I') < \varepsilon.$$

That would prove, using Theorem 3.3 that the integrability of f follows from the McShane criterion.

But it is also true that this expression

$$\left| \sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} [|f(w)| - |f(w')|]\lambda(I\cap I') \right| < \varepsilon \tag{3.4}$$

is smaller than the expression

$$\sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} \left|f(w) - f(w')\right| \lambda(I\cap I') < \varepsilon.$$

That would prove, again using Theorem 3.3 that the integrability of $|f|$ follows from the same McShane criterion. ∎

3.2 Elementary properties of the integral

All of our elementary properties of the integral are anticipated by the classical Newton integral which shares all the same properties in somewhat weaker forms. For the Newton-type integrals the proofs are obtained by direct appeal to corresponding properties of derivatives. Here we can prove properties either by appealing to the equivalence of our integral with the general Newton integral, or by deriving the properties directly from the expression of the integral as a limit of Riemann sums. In many cases (for example the monotone convergence property) the use of the upper or lower integrals simplifies technical details.

3.2.1 Integration and order

Theorem 3.6 *Suppose that f, g are both integrable on a compact interval $[a,b]$ and that $f(x) \leq g(x)$ for almost every x in that interval. Then*

$$\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx.$$

3.2.2 Integration of linear combinations

Theorem 3.7 *Suppose that f, g are both integrable on a compact interval $[a,b]$. Then so too is any linear combination $rf+sg$ and*

$$\int_a^b [rf(x)+sg(x)]\,dx = r\left(\int_a^b f(x)\,dx\right)+s\left(\int_a^b g(x)\,dx\right).$$

3.2.3 Integrability on subintervals

Theorem 3.8 *Suppose that f is integrable on a compact interval $[a,b]$. Then f is integrable on any compact subinterval of $[a,b]$.*

3.2.4 Additivity

Theorem 3.9 *If f is integrable on both of the intervals $[a,b]$ and $[b,c]$, then f is integrable on $[a,c]$ and*

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx.$$

3.2.5 A simple change of variables

Let $\phi : [a,b] \to \mathbb{R}$ be a strictly increasing differentiable function. We would expect from elementary formulas of the calculus that

$$\int_{\phi(a)}^{\phi(b)} f(x)\,dx = \int_a^b f(\phi(t))\phi'(t)\,dt.$$

If f is itself everywhere a derivative then this could be justified. If f is assumed only to be integrable then a different proof, using ϕ to map full covers and partitions, is needed.

Later on in Section 5.12 we will return to this problem and find more general conditions for a change of variables. The convenient tool there is the Stieltjes integral which helps clarify these ideas and offers some technical simplifications.

Theorem 3.10 (Change of variable) *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a strictly increasing, differentiable function. If $f : \mathbb{R} \to \mathbb{R}$ is integrable on $[\phi(a),\phi(b)]$ then*

$$\int_{\phi(a)}^{\phi(b)} f(x)\,dx = \int_a^b f(\phi(t))\phi'(t)\,dt.$$

Proof. Let $\varepsilon > 0$ and define β to be the collection of all pairs $([x,y],z)$ subject only to the conditions that

$$\left|\frac{\phi(y)-\phi(x)}{y-x} - \phi'(z)\right| < \frac{\varepsilon}{2(b-a)|(1+|f(\phi(z)|)}.$$

Since ϕ is everywhere differentiable this is a full cover. Note that we can write $\phi(y)-\phi(x)$ also as $\lambda(J)$ where $J=\phi([x,y])$ is just the compact interval that ϕ maps $[x,y]$ onto.

Write

$$\beta_1' = \{(\phi([x,y]),\phi(x)) : ([x,y],z)\in\beta_1\}$$

and check that β_1' is also a full cover. Observe that elements $(J,x) = (\phi([x,y]),\phi(z))$ of β_1' must satisfy

$$|f(\phi(x))\lambda(\phi([x,y])) - f(\phi(x))\phi'(x)\lambda([x,y])| < \varepsilon\lambda([x,y])/2(b-a).$$

The expression $f(\phi(t))\lambda(\phi([x,y]))$ here is better viewed as $f(x)\lambda(J)$.

Choose a full cover β_2' so that

$$\left|\int_{\phi(a)}^{\phi(b)} f(x)\,dx - \sum_{(J,x)\in\pi'} f(x)\lambda(J)\right| < \varepsilon/2$$

for all partitions $\pi'\subset\beta_2'$ of the interval $[\phi(a),\phi(b)]$. Write β_2 for the collection of all (I,x) for which $(I,x)=(\phi(J),\phi(t))$ for some $(J,t)\in\beta_2'$. This is a full cover of $[a,b]$.

Write $\beta=\beta_1\cap\beta_2$. Check that β is a full cover of $[a,b]$ and check that

$$\left|\int_{\phi(a)}^{\phi(b)} f(x)\,dx - \sum_{(I,x)\in\pi} f(\phi(x))\phi'(x)\lambda(I)\right| < \varepsilon$$

for all partitions $\pi\subset\beta$ of the interval $[a,b]$. An appeal to the HK criterion then completes the proof. ∎

3.2.6 Integration by parts

Integration by parts formula:

$$\int_a^b F(x)G'(x)\,dx = F(b)G(b) - F(a)G(b) - \int_a^b F'(x)G(x)\,dx. \tag{3.5}$$

The formula can be derived from the product rule for derivatives:

$$\frac{d}{dx}(F(x)G(x)) = F(x)G'(x) + F'(x)G(x)$$

which holds at any point where both functions are differentiable. One must then give strong enough hypotheses that the function $F(x)G(x)$ is an indefinite integral for the function

$$F(x)G'(x)+F'(x)G(x)$$

in the sense needed for our integral.

The most general statement is the following: if f and g are both integrable on $[a,b]$ and F and G are their indefinite integrals on that interval then $Fg+fG$

is integrable on $[a,b]$ and

$$\int_a^b (F(x)g(x)+f(x)G(x))\,dx = F(b)G(b)-F(a)G(b).$$

In particular the usual formula (3.5) holds if and only if one of the two integrals in that statement exists. The proof is easiest to deduce from the Stieltjes version

$$\int_a^b F(x)\,dG(x)+G(x)dF(x) = F(b)G(b)-F(a)G(b) \tag{3.6}$$

that we will study in a later chapter. The reader may wish to try, however, to prove it directly.

Remark: For the Lebesgue integral the integration by parts formula is available but not quite as straightforward. It is possible that $Fg+fG$ is integrable on $[a,b]$ but that only one of Fg and fG is Lebesgue integrable (i.e., absolutely integrable) on $[a,b]$. For example take $F(x)=x$ and $G(x)=x\cos x^{-2}$ on $[0,1]$. It is also possible neither is Lebesgue integrable: take $F(x)=x^{1/2}\sin x^{-1}$ and $G(x)=x^{1/2}\cos x^{-1}$.

3.2.7 Derivative of the integral

If f is integrable on an interval $[a,b]$ then the formula

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

holds at almost every point in (a,b). This is merely by definition had we started with the Newton version of the integral. Since the Henstock-Kurzweil integral is equivalent to the general Newton integral we know this fact already. To make a claim, however, at some particular point the following simple observation is sometimes useful.

Theorem 3.11 *Let $f:[a,b]\to\mathbb{R}$ be an integrable function on the interval $[a,b]$. Let*

$$F(t)=\int_a^t f(x)\,dx \quad (a\le t\le b).$$

Assume that $x_0\in[a,b]$ is a point of continuity of f. Then

1. *If $a<x_0<b$ then $F'(x_0)=f(x_0)$.*
2. *If $a=x_0$ then the right hand derivative $F'_+(a)=f(a)$.*
3. *If $x_0=b$ then the left hand derivative $F'_-(b)=f(b)$.*

Proof. Let x_0 be a point of continuity of f and let $\varepsilon>0$. Then there is a $\delta>0$ so that $|f(x)-f(x_0)|<\varepsilon$ if $|x-x_0|<\delta$ and $x\in[a,b]$. Let $[u,v]\subset[a,b]$ be any interval that contains x_0 and has length less than δ. Simply compute

$$\left|\int_u^v f(x)\,dx - f(x_0)(v-u)\right| = \left|\int_u^v f(x)\,dx - \int_u^v f(x_0)\,dx\right|$$

$$\leq \int_u^v |f(x) - f(x_0)|\,dx \leq \varepsilon(v-u).$$

From this the conclusions of the theorem are easy to check. ∎

3.2.8 Null functions

A function is a null function if it is equal to zero at every point with only a small set of exceptions (a set of measure zero). The primary properties of null functions are easily proved from the equivalence of our integral with the general Newton integral.

Theorem 3.12 *Let $f : [a,b] \to \mathbb{R}$ be a null function. Then f is integrable on $[a,b]$ and*

$$\int_a^b f(x)\,dx = 0.$$

Theorem 3.13 *Let $f : [a,b] \to \mathbb{R}$ be an integrable function with the property that*

$$\int_c^d f(x)\,dx = 0 \text{ for all } [c,d] \subset [a,b].$$

Then f is a null function.

Corollary 3.14 *Let $f : [a,b] \to \mathbb{R}$ be a nonnegative integrable function with the property that*

$$\int_a^b f(x)\,dx = 0.$$

Then f is a null function.

3.2.9 Monotone convergence theorem

The two formulas

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b \left(\lim_{n\to\infty} f_n(x) \right) dx$$

and

$$\int_a^b \left(\sum_{k=1}^{\infty} g_k(x) \right) dx = \sum_{k=1}^{\infty} \left(\int_a^b g_k(x)\,dx \right)$$

are extremely useful and have multiple applications in analysis. They are not generally valid. If the sequence of integrable functions $\{f_n\}$ is monotone then the former identity does hold. If the sequence of integrable functions $\{g_k\}$ is nonnegative then the latter does hold. We prove both of these assertions. (One follows easily from the other.)

Theorem 3.15 (Monotone convergence theorem) *Let $f_n : [a,b] \to \mathbb{R}$ ($n = 1,2,3,\dots$) be a nondecreasing sequence of integrable functions and suppose that*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for almost every x in $[a,b]$. Then

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b f_n(x)\,dx. \tag{3.7}$$

In particular, if the limit exists and is finite the function f is integrable on $[a,b]$ and the identity (3.7) holds. If the limit is infinite then the function f is not integrable but the integral is determined and

$$\int_a^b f(x)\,dx = \infty.$$

Here we are using the ideas from Section 3.1 that allow us to express an integral as infinite. This was not available to us in our study of the Newton integral but the Henstock-Kurzweil theory of upper and lower integrals allowed this. The proof of Theorem 3.15 is given in Section 3.2.11 below.

3.2.10 Summing inside the integral

We establish here that the summation formula

$$\int_a^b \left(\sum_{n=1}^\infty f_n(x) \right) dx = \sum_{n=1}^\infty \left(\int_a^b f_n(x)\,dx \right)$$

is possible for nonnegative integrable functions.

Theorem 3.16 (summing inside the integral) *Let $f_n : [a,b] \to \mathbb{R}$ ($n = 1,2,3,\dots$) be a sequence of nonnegative integrable functions and suppose that*

$$f(x) = \sum_{n=1}^\infty f_n(x)$$

for almost every x. Then

$$\int_a^b f(x)\,dx = \sum_{n=1}^\infty \left(\int_a^b f_n(x)\,dx \right).$$

In particular, if the series converges the function f is integrable on $[a,b]$ and the identity (3.10) holds. If the series diverges then the function f is not integrable but the integral is determined and

$$\int_a^b f(x)\,dx = \infty.$$

The proof is obtained from the two lemmas given in Section 3.2.11 below.

3.2.11 Two convergence lemmas

The monotone convergence theorem and the formula for summing inside the integral are directly related by the following observation. If

$$f_1(x) \le f_2(x) \le f_3(x) \le \dots$$

and

$$\lim_{n\to\infty} f_n(x) = f(x)$$

then

$$f(x) = f_1(x) + \sum_{n=1}^{\infty} (f_n(x) - f_{n-1}(x))$$

expresses f as the sum of a series. Thus it is enough to prove Theorem‘3.16. This is obtained from the following two lemmas.

Lemma 3.17 *Suppose that $f, f_1, f_2, \dots$ is a sequence of nonnegative functions defined on a compact interval $[a,b]$. If, for almost every x*

$$f(x) \ge \sum_{n=1}^{\infty} f_n(x),$$

then

$$\underline{\int_a^b} f(x)\,dx \ge \sum_{n=1}^{\infty} \left(\underline{\int_a^b} f_n(x)\,dx \right). \tag{3.8}$$

Proof. We can assume that the inequality assumed is valid for *every* x; simply redefine $f_n(x) = 0$ for those points in the null set where the inequality doesn't work. The resulting functions will have the same lower integrals as f_n.

Let $\varepsilon > 0$. Take any integer N and choose full covers β_n ($n = 1, 2, \dots, N$) so that all the Riemann sums[1]

$$\sum_{\pi} f_n(w)(v-u) \ge \underline{\int_a^b} f_n(x)\,dx - \varepsilon 2^{-n}$$

whenever $\pi \subset \beta_n$ is a partition of $[a,b]$. (If the integrals here are not finite then there is nothing to prove, since both sides of the inequality (3.8) will be infinite.)

Let

$$\beta = \bigcap_{n=1}^{N} \beta_n.$$

This too is a full cover, one that is contained in all of the others.

[1] We simplify our notation for Riemann sums a bit by replacing

$$\sum_{([u,v],w)\in\pi} f(w)(v-u) \text{ by } \sum_{\pi} f(w)(v-u).$$

Take any partition of $[a,b]$ with $\pi \subset \beta$, and compute

$$\sum_{\pi} f(w)(v-u) \geq \sum_{\pi} \left(\sum_{n=1}^{N} f_n(w)(v-u) \right) = \sum_{n=1}^{N} \left(\sum_{\pi} f_n(w)(v-u) \right) \geq$$

$$\sum_{n=1}^{N} \left(\underline{\int_a^b} f_n(x)\,dx - \varepsilon 2^{-n} \right).$$

This gives a lower bound for all Cauchy sums and hence, since ε is arbitrary, shows that

$$\underline{\int_a^b} f(x)\,dx \geq \sum_{n=1}^{N} \left(\underline{\int_a^b} f_n(x)\,dx \right).$$

As this is true for all N the inequality (3.8) must follow. ■

Lemma 3.18 *Suppose that $f, f_1, f_2, \ldots$ is a sequence of nonnegative functions defined on a compact interval $[a,b]$. If, for almost every x*

$$f(x) \leq \sum_{n=1}^{\infty} f_n(x),$$

then

$$\overline{\int_a^b} f(x)\,dx \leq \sum_{n=1}^{\infty} \left(\overline{\int_a^b} f_n(x)\,dx \right). \tag{3.9}$$

Proof. As before, we can assume that the inequality assumed is valid for *every* x; simply redefine $f(x) = 0$ for those points in the null set where the inequality doesn't work. The resulting function will have the same integral and same upper integral as f.

This lemma is similar to the preceding one, but requires a bit of bookkeeping and a new technique with the covers. Let $t < 1$ and choose for each $x \in [a,b]$ the first integer $N(x)$ so that

$$t f(x) \leq \sum_{n=1}^{N(x)} f_n(x).$$

Choose, again and using the same ideas as before, full covers β_n ($n = 1, 2, \ldots$) so that $\beta_1 \supset \beta_2 \supset \beta_3 \ldots$ and all Riemann sums[2]

$$\sum_{\pi} f_n(w)(v-u) \leq \overline{\int_a^b} f_n(x)\,dx + \varepsilon 2^{-n}$$

whenever $\pi \subset \beta_n$ is a partition of $[a,b]$. (Again, if the integrals here are not finite then there is nothing to prove, since the larger side of the inequality (3.9) will be

[2] As before, we simplify our notation for Riemann sums by replacing

$$\sum_{([u,v],w)\in\pi} f(w)(v-u) \text{ by } \sum_{\pi} f(w)(v-u).$$

infinite.)

Let

$$E_n = \{x \in [a,b] : N(x) = n\}.$$

We use these sets to carve up the covering relations. Write

$$\beta_n[E_n] = \{([u,v],w) \in \beta_n : w \in E_n\}.$$

There must be a full cover β so that

$$\beta[E_n] \subset \beta_n[E_n]$$

for all $n = 1,2,3,\dots$.

Take any partition of $[a,b]$ with $\pi \subset \beta$. Let N be the largest value of $N(x)$ for the finite collection of pairs $(I,x) \in \pi$. We need to carve the partition π into a finite number of disjoint subsets by writing, for $j = 1,2,3,\dots,N$,

$$\pi_j = \{([u,v],w) \in \pi : w \in E_j\}$$

and

$$\sigma_j = \pi_j \cup \pi_{j+1} \cup \dots \cup \pi_N.$$

for integers $j = 1,2,3,\dots,N$. Note that

$$\sigma_j \subset \beta_j$$

and that

$$\pi = \pi_1 \cup \pi_2 \cup \dots \cup \pi_N.$$

Check the following computations, making sure to use the fact that for $x \in E_i$,

$$tf(x) \leq f_1(x) + f_2(x) + \dots + f_i(x).$$

$$\sum_{\pi} tf(w)(v-u) = \sum_{i=1}^{N} \sum_{\pi_i} tf(w)(v-u)$$

$$\leq \sum_{i=1}^{N} \sum_{\pi_i} (f_1(w) + f_2(w) + \dots + f_i(w))(v-u)$$

$$= \sum_{j=1}^{N} \left(\sum_{\sigma_j} f_j(w)(v-u) \right) \leq$$

$$\sum_{j=1}^{N} \left(\overline{\int_a^b} f_j(x)\,dx + \varepsilon 2^{-j} \right) \leq \sum_{j=1}^{\infty} \left(\overline{\int_a^b} f_j(x)\,dx \right) + \varepsilon.$$

This gives an upper bound for all Cauchy sums and hence, since ε is arbitrary, shows that

$$\overline{\int_a^b} tf(x)\,dx \leq \sum_{n=1}^{\infty} \left(\overline{\int_a^b} f_n(x)\,dx \right).$$

As this is true for all $t < 1$ the inequality (3.9) must follow too. ∎

Exercises

Exercise 190 *Give an example to show that it is possible that* $\int_a^b f(x)\,dx = \infty$ *in Theorem 3.16.* □

Exercise 191 *Give an example to show that it is possible for the Theorem 3.16 to fail if we drop the assumption that the functions are nonnegative in the theorem.* □

Exercise 192 *Let* $f_n : [a,b] \to \mathbb{R}$ *(*$n = 1,2,3,\ldots$*) be a sequence of absolutely integrable functions and suppose that*

$$\sum_{n=1}^{\infty} |f_n(x)| < \infty$$

for almost every x *and that*

$$\sum_{n=1}^{\infty} \left(\int_a^b |f_n(x)|\,dx \right) < \infty.$$

Then show that

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

is finite for almost every x *in* $[a,b]$*, is absolutely integrable, and that*

$$\int_a^b f(x)\,dx = \sum_{n=1}^{\infty} \left(\int_a^b f_n(x)\,dx \right).$$

□

3.3 Equi-integrability

We describe a uniform version of integrability that is useful in discussions of the convergence of sequences of integrable functions.

Definition 3.19 (equi-integrability) *Suppose that* $\{f_n\}$ *is a sequence of integrable functions defined at every point of a compact interval* $[a,b]$*. Then* $\{f_n\}$ *is said to be equi-integrable on* $[a,b]$ *if, for every* $\varepsilon > 0$*, there is a full cover* β *of* $[a,b]$ *so that*

$$\left| \int_a^b f_n(x)\,dx - \sum_{([u,v],w)\in\pi} f_n(w)(v-u) \right| < \varepsilon$$

whenever π *is a partition of the interval* $[a,b]$ *chosen from* β*.*

Uniform convergence is a sufficient condition for equi-integrability, but the condition itself is much more general.

Lemma 3.20 *Suppose that $\{f_n\}$ is a sequence of integrable functions defined at every point of a compact interval $[a,b]$ and that $\{f_n\}$ is uniformly convergent on $[a,b]$. Then $\{f_n\}$ is equi-integrable on $[a,b]$.*

Equi-integrability along with pointwise convergence gives a simply stated criterion for taking the limit inside the integral.

Theorem 3.21 *Suppose that $\{f_n\}$ is a sequence of equi-integrable functions defined at every point of a compact interval $[a,b]$ and that $\{f_n\}$ is pointwise convergent almost everywhere on $[a,b]$ to a function f. Then f is integrable on $[a,b]$ and*

$$\int_a^b f(x)\,dx = \lim_{n\to\infty}\int_a^b f_n(x)\,dx. \tag{3.10}$$

We will return to the subject of equi-integrability in Section 4.13. Here we are concerned only with integrability conditions expressible in the simple language of the Henstock-Kurzweil integral. For the (narrower) Lebesgue integral some stronger conditions are needed that appear in the next chapter.

Exercise 193 *Prove Lemma 3.20.* □

Exercise 194 *Prove Theorem 3.21.* Answer □

Exercise 195 *Does equi-integrability of a sequence $\{f_n\}$ and pointwise convergence of $\{f_n\}$ to a function f imply uniform convergence of the sequence $\{f_n\}$?* Answer □

Exercise 196 *Definition 3.19 expresses what might be called equi-integrability in the Henstock-Kurzweil sense. Give a definition for equi-integrability in the Riemann sense. Can you prove any theorems using this concept?* Answer □

Chapter 4

Lebesgue's Integral

Lebesgue's program is the construction of the value of the integral

$$\int_a^b f(x)\,dx$$

directly from the measure and the values of the function f in the integral. Our formal definition of the integral *appears* to do this. Since full covers are not themselves, in general, constructible from the function being integrated we cannot claim that our integral is constructed in the sense Lebesgue intends.

For his program he invented the integral as a heuristic device, imagined what properties it should possess, and then went about discovering how to construct it based on this fiction. At the end he had to take his construction as the definition itself. For us to follow the same program is much easier: we have an integral, we know many of its properties, and we can use this information to construct the value of the integral for any absolutely integrable function.

This chapter presents an introduction to Lebesgue's methods, but backwards in a sense from conventional presentations. We already have a formal definition of the integral, so we do not need to define an integral by Lebesgue's method. Nor do we need to develop its elementary properties since we have already done so by other means. We need only show how to construct the value of an object $\int_a^b f(x)\,dx$ that we have already defined. In the course of that construction we discover the many methods of measure theory which can be used to study integration theory at a deeper level.

4.1 The Lebesgue integral

The Lebesgue integral is a special case of the general integral. It is not merely a special case, but certainly the most important special case.

Definition 4.1 (The Lebesgue integral) *A function $f : [a,b] \to \mathbb{R}$ is said to be* Lebesgue integrable *if f is absolutely integrable, i.e., if both f and $|f|$ are integrable on $[a,b]$.*

By definition all Lebesgue integrable functions are integrable in our sense. There are functions that are integrable but not Lebesgue integrable; they are said to be *nonabsolutely integrable*. The theory of such functions is less powerful and more delicate[1] than the theory of the Lebesgue integrable functions. There are also fewer applications. We return to nonabsolutely integrable functions in Chapter 6. In this chapter we study properties of absolutely (i.e., Lebesgue) integrable functions. In particular we shall find that a good constructive theory is available using Lebesgue's theory of measure.

4.2 Lebesgue measure

We define the following three versions of Lebesgue measure (similar to the three versions of a measure zero set) for a set $E \subset \mathbb{R}$:

- $\lambda(E) = \inf\{\lambda(G) : \ G$ open and $G \supset E\}$.

- $\lambda^*(E) = \inf_{\beta}\left(\sup_{\pi\subset\beta} \sum_{([u,v],w)\in\pi} (v-u) \right)$

  where the infimum is taken over all full covers β of the set E and $\pi \subset \beta$ is an arbitrary subpartition.

- $\lambda_*(E) = \inf_{\beta}\left(\sup_{\pi\subset\beta} \sum_{([u,v],w)\in\pi} (v-u) \right)$

  where the infimum is taken over all fine covers β of the set E and $\pi \subset \beta$ is an arbitrary subpartition.

The first of these λ is Lebesgue's original version of his measure. We have already (in Section 2.4.1) defined the Lebesgue measure of open sets. This definition extends that, by a simple infimum, to all sets. The definition of the full measure λ^* is closely related to the integral.

Lemma 4.2 *Let E be a set of real numbers contained in an interval $[a,b]$. Then*

$$\lambda^*(E) = \overline{\int_a^b} \chi_E(x)\,dx.$$

The three definitions are equivalent, a fact which is proved as the Vitali covering theorem in Section 4.3 below.

[1] The best analogy that captures the difference is in the theory of convergent series: absolutely convergent series permit a stronger and more useful theory than do the nonabsolutely [conditionally] convergent series.

4.2.1 Basic property of Lebesgue measure

Theorem 4.3 *Lebesgue measure* λ *is a nonnegative real-valued set function defined for all sets of reals numbers that is a* measure[a] *on* $\mathbb{R}$*, i.e., it has the following properties:*

1. $\lambda(\emptyset) = 0.$
2. *For any sequence of sets* $E, E_1, E_2, E_3, \ldots$ *for which*
$$E \subset \bigcup_{n=1}^{\infty} E_n$$
the inequality
$$\lambda(E) \leq \sum_{n=1}^{\infty} \lambda(E_n)$$
must hold.

[a]Most authors would call this an *outer measure*.

This result is often described by the following language that splits the property (2) in two parts:

Subadditivity: $\lambda\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \lambda(E_n).$

Monotonicity: $\lambda(A) \leq \lambda(B)$ if $A \subset B$.

Since we have three representations of the Lebesgue measure, as λ, λ^*, or as λ_* we can prove this using any one of the three. The exercises ask for all three; any one would suffice in view of the Vitali covering theorem proved in the next section.

Exercises

Exercise 197 *Prove Lemma 4.2, i.e., establish the identity*
$$\lambda^*(E) = \overline{\int_a^b} \chi_E(x)\,dx$$
for any set $E \subset [a,b]$*. Note that this exercise uses the full version* λ^**, not the original Lebesgue version* λ*.* Answer □

Exercise 198 *Prove that* λ *is a measure in the sense of Theorem 4.3.* Answer □

Exercise 199 *Prove that* λ^* *is a measure in the sense of Theorem 4.3.* □

Exercise 200 *Prove that λ_* is a measure in the sense of Theorem 4.3.*

□

Exercise 201 *Prove that $\lambda(A) < t$ if and only if there is an open set G that contains all but countably many points of A and for which $\lambda(G) < t$.* □

4.3 Vitali covering theorem

These three measures are identical and we can use any version. The identity $\lambda = \lambda^* = \lambda_*$ is Vitali's theorem, although his theorem is normally expressed in different language.[2]

The proof is just a bit more difficult than the proof of the narrower version, the mini-Vitali theorem given in Section 2.7, where we showed that sets of measure zero were equivalent to both full null and fine null sets.

Theorem 4.4 (Vitali Covering Theorem) $\lambda = \lambda^* = \lambda_*$.

4.3.1 Classical version of Vitali's theorem

Vitali's covering theorem asserts that the measure of an arbitrary set can be determined from full and fine covers of that set. The basic computation about fine covers is the following lemma, known as the classical version of Vitali's theorem.

Lemma 4.5 (Vitali covering theorem) *Let β be a fine cover of a bounded set E and suppose that $\varepsilon > 0$. Then there must exist a subpartition $\pi \subset \beta$ for which*

$$\lambda\left(E \setminus \bigcup_{([u,v],w)\in\pi} [u,v]\right) < \varepsilon. \tag{4.1}$$

Proof. For the proof of this theorem we need only one simple fact (Exercise 201) about the Lebesgue measure $\lambda(E)$ of a real set A:

> $\star$ $\lambda(A) < \varepsilon$ *if and only if there is an open set G containing all but countably many points of A and for which $\lambda(G) < \varepsilon$.*

Thus the proof is really about open sets. Indeed in our proof we use only the Lebesgue measure of open sets and several covering lemmas.

The proof is just a repeated application of Lemma 2.18. Since E is bounded there is an open set U_1 containing E for which $\lambda(U_1) < \infty$. If $\lambda(U_1) < \varepsilon$ then,

[2] The language here, will no doubt, shock some traditionalists for whom it appears to suggest Lebesgue inner and outer measure. But this has nothing to do with inner/outer measure. The measures λ^* and λ_* are those derived from full and fine covers.

since $E \subset U_1$, $\lambda(E) < \varepsilon$ and there is nothing more to prove: take $\pi = \emptyset$ and the statement (4.1) is satisfied. If $\lambda(U_1) \geq \varepsilon$ we start our process.

We prune β by the open set U_1: define $\beta_1 = \beta(U_1)$. Note that this, too, is a fine cover of E. Set

$$G_1 = \bigcup_{([u,v],w)\in\beta_1} (u,v).$$

Then G_1 is an open set and $g_1 = \lambda(G_1) < \lambda(U_1)$ is finite. We know from Lemma 2.17, that G_1 covers all of E except for a countable set. [We shall ignore countable sets in this proof, to keep the bookkeeping simple]. By Lemma 2.18 there must exist a subpartition $\pi_1 \subset \beta_1$ for which

$$U_2 = G_1 \setminus \bigcup_{([u,v],w)\in\pi} [u,v]$$

is an open subset of G_1 and

$$\lambda(U_2) \leq 5g_1/6 \leq 5\lambda(U_1)/6.$$

Define

$$E_1 = E \setminus \bigcup_{([u,v],w)\in\pi_1} [u,v].$$

If $\lambda(U_2) < \varepsilon$ then $\lambda(E_1) < \varepsilon$. This is because U_2 is an open set containing all of E_1 except possibly some countable set; thus $\star$ stated above implies that $\lambda(E_1) < \varepsilon$. But if $\lambda(E_1) < \varepsilon$ the process can stop: take $\pi = \pi_1$ and the statement (4.1) is satisfied.

If $\lambda(U_2) \geq \varepsilon$ we continue our process. Define $\beta_2 = \beta(U_2)$ and note that this is a fine cover of E_1 (i.e., the points in E not already handled by the subpartition π_1 or the countably many points of E discarded in the first stage of our proof).

Set

$$G_2 = \bigcup_{([u,v],w)\in\beta_2} (u,v).$$

Then G_2 is an open set and

$$g_2 = \lambda(G_2) \leq \lambda(U_2).$$

As before, we know from Lemma 2.17, that G_2 covers all of E_1 except for a countable set. [We are ignoring countable sets in this proof, throw these points away].

Again applying Lemma 2.18, we find a subpartition $\pi_2 \subset \beta_2$ for which

$$U_3 = G_2 \setminus \bigcup_{([u,v],w)\in\pi_2} [u,v]$$

is an open subset of G_2 and $\lambda(U_3) \leq 5g_2/6$. Define

$$E_2 = E_1 \setminus \bigcup_{([u,v],w)\in\pi_2} [u,v]$$

$$= E \setminus \bigcup_{([u,v],w)\in\pi_1\cup\pi_2} [u,v].$$

If $\lambda(U_3) < \varepsilon$ then $\lambda(E_2) < \varepsilon$. This because U_3 is an open set containing all of E_2 except possibly some countable set; thus $\star$ stated above implies that $\lambda(E_2) < \varepsilon$. But if $\lambda(E_2) < \varepsilon$ the process can stop: take $\pi = \pi_1 \cup \pi_2$ and the statement (4.1) is satisfied. [Be sure to check that the intervals from π_1 have been arranged to be disjoint from the intervals in π_2.]

This process is continued, inductively, until it stops. It certainly *must* stop since

$$\lambda(U_{k+1}) < \frac{5}{6}\lambda(U_k) \leq \cdots \leq \left(\frac{5}{6}\right)^k \lambda(U_1)$$

so that eventually $\lambda(U_{k+1}) < \varepsilon$ and $\lambda(E_k) < \varepsilon$. Take

$$\pi = \pi_1 \cup \pi_2 \cup \ldots \pi_k$$

and the statement (4.1) is satisfied. ∎

4.3.2 Proof that $\lambda = \lambda^* = \lambda_*$.

The inequality

$$\lambda_* \leq \lambda^* \leq \lambda$$

is trivial. First of all, any full cover is also a fine cover so that $\lambda_* \leq \lambda^*$ must be true. Second, if $\lambda(E) < t$ there is an open set G containing E for which it is also true that $\lambda(G) < t$. But then we can define a covering relation β to consist of all pairs $([u,v],w)$ provided $w \in [u,v] \subset G$. This is a full cover of E. Note that

$$\sum_{([u,v],w)\in\pi} (v-u) \leq \lambda(G) < t$$

whenever $\pi \subset \beta$ is an arbitrary subpartition. It follows that $\lambda^*(E) < t$. As this is true for all t,

$$\lambda^*(E) \leq \lambda(E).$$

Finally, then, Lemma 4.5 completes the proof. Let β be any fine cover of a bounded set E and suppose that $\varepsilon > 0$. Then there must exist a subpartition $\pi \subset \beta$ for which

$$\lambda\left(E \setminus \bigcup_{([u,v],w)\in\pi} [u,v]\right) < \varepsilon. \tag{4.2}$$

In particular, using subadditivity measure property of λ,

$$\lambda(E) \leq \lambda\left(E \setminus \bigcup_{([u,v],w)\in\pi} [u,v]\right) + \sum_{([u,v],w)\in\pi} \lambda([u,v])$$

$$< \sum_{([u,v],w)\in\pi} (v-u) + \varepsilon.$$

So, since this is true for any fine cover of E,

$$\lambda(E) \leq \lambda_*(E) + \varepsilon.$$

It follows that $\lambda(E) \leq \lambda_*(E)$ for all bounded sets E.

That establishes the identity $\lambda = \lambda^* = \lambda_*$ for all bounded sets. The extension to unbounded sets can be accomplished with the standard measure properties.

4.4 Density theorem

As an application of the Vitali covering theorem we prove the density theorem. This asserts that for an arbitrary set E almost every point is a point of density, a point x where

$$\frac{\lambda(E \cap [u,v])}{\lambda([u,v])} \to 1$$

as $[u,v]$ shrinks to x.

Theorem 4.6 *Almost every point of an arbitrary set E is a point of density.*

Proof. To define this with a bit more precision write

$$\underline{d}(E,x) = \sup_{\delta > 0} \inf \left\{ \frac{\lambda(E \cap [u,v])}{\lambda([u,v])} : u \leq x \leq v,\ 0 < v - u < \delta \right\}.$$

This is called the *lower density* of E at x. The theorem asserts that

$$\underline{d}(E,x) = 1$$

at almost every point x of E.

We may assume that E is bounded. Take any $\alpha < 1$ and define

$$E_\alpha = \{x \in E : \underline{d}(E,x) < \alpha\}$$

and

$$E' = \{x \in E : \underline{d}(E,x) < 1\}.$$

We show that E_α is necessarily a set of measure zero. It follows that E' is then a set of measure zero since evidently

$$E' = \bigcup_{n=1}^{\infty} E_{\frac{n}{n+1}}.$$

Fix $\alpha < 1$ and any open set G containing E_α, and define

$$\beta = \{([u,v],w) : u \leq x \leq v,\ \lambda(E \cap [u,v]) < \alpha\lambda([u,v])\}.$$

This is a fine cover of E_α, and since G is an open set containing E_α, the pruned relation $\beta(G)$ is also a fine cover of E_α. Let $\varepsilon > 0$. By the Vitali covering theorem (Lemma 4.5) there must exist a subpartition $\pi \subset \beta(G)$ for which

$$\lambda\left(E_\alpha \setminus \bigcup_{([u,v],w) \in \pi} [u,v]\right) < \varepsilon. \tag{4.3}$$

Now we simply compute, using subadditivity, that

$$\lambda(E_\alpha) \leq \lambda\left(E_\alpha \setminus \bigcup_{([u,v],w)\in\pi} [u,v]\right) + \sum_{([u,v],w)\in\pi} \lambda(E_\alpha \cap [u,v])$$
$$\leq \varepsilon + \sum_{([u,v],w)\in\pi} \lambda(E \cap [u,v])$$
$$\leq \varepsilon + \alpha \sum_{([u,v],w)\in\pi} \lambda([u,v]) \leq \varepsilon + \alpha\lambda(G).$$

We deduce that $\lambda(E_\alpha) \leq \lambda(G)$ for all such open sets G and hence that $\lambda(E_\alpha) \leq \alpha\lambda(E_\alpha)$. This is possible only if $\lambda(E_\alpha) = 0$. ∎

4.4.1 Approximate Cousin lemma

Density arguments play an important role in many advanced studies in analysis. As an illustration we generalize the notion of a full cover by introducing a density computation into the concept.

Definition 4.7 *A covering relation β is said to be approximately full at a point x_0 provided there is a set Δ_{x_0} containing x_0 and for which $\underline{d}(\Delta_{x_0}, x_0) = 1$ so that $([y,z],x)$ belongs to β for all $y \leq x \leq z$ with y, $z \in \Delta_{x_0}$.*

A covering relation β is said to be approximately full if it is approximately full at every point. Note that all full covers are necessarily also approximately full. There is a covering lemma entirely analogous to the Cousin lemma available for these more general covers.

Lemma 4.8 (Approximate Cousin) *A covering relation that is approximately full contains partitions of every compact interval.*

Proof. Suppose that β is approximately full. Let us say that an interval (a,b) is *regular* if β contains a partition of every closed subinterval of $[a,b]$. Let R be the union of all regular intervals. (If there are no regular intervals then $R = \emptyset$.) Clearly R is open. Let (c,d) be an open interval contained in R. We claim that then (c,d) itself is regular. Suppose $[a,b] \subset [c,d]$.

If $c < a < b < d$ then a simple compactness argument (Heine-Borel) shows that β contains a partition of $[a,b]$. Alternatively an ordinary covering argument works too. Let

$$\alpha = \{([u,v],w) : \ \beta \text{ contains a partition of } [u,v]\ \}.$$

Then α is a full cover of R and so, by the ordinary Cousin lemma, α contains a partition of $[a,b]$. That can be used to deduce that β contains a partition of $[a,b]$.

If $c = a < b = d$ then again we can argue that β contains a partition of $[a,b]$. Indeed select $a < a' < b' < b$ so that $([a,a'],a)$ and $([b',b],b)$ are in β and select a partition $\pi \subset \beta$ of the interval $[a',b']$. Then

$$\pi' = \pi \cup \{([a,a'],a),\, ([b',b],b)\}$$

is a partition of $[a,b]$ contained in β.

Thus the lemma is proved if we are able to show that $R = \mathbb{R}$. Suppose not. Then $E = \mathbb{R} \setminus R$ is a nonempty closed set. For each $x \in E$ choose a set Δ_x having density 1 at x so that $([x,y],x)$ and $([z,x],x)$ belong to β if $z < x < y$ and z, $y \in \Delta_x$. Choose $\delta(x) > 0$ so that

$$\frac{\lambda(\Delta_x \cap (x,x+h))}{h} > 1/2$$

and

$$\frac{\lambda(\Delta_x \cap (x-h,x))}{h} > 1/2$$

if $0 < h < \delta(x)$.

Write, for any integer n,

$$E_n = \{x \in E : \delta(x) > 1/n\},$$

observing that the union of the sequence of sets $\{E_n\}$ is all of E.

By the Baire category theorem[3] there is an interval (c,d) and an integer n so that E_n is dense in $E \cap (c,d)$ and $d - c < 1/n$.

We complete the proof by showing that (c,d) must then be a regular interval. This would be impossible since no regular interval can contain any points of E.

First observe that if $[u,v] \subset [c,d]$ and u, v are points of E_n then the set

$$\Delta_u \cap \Delta_v \cap [u,v] \neq \emptyset.$$

If that set were empty it would violate the measure conditions

$$\lambda(\Delta_u \cap (u,v)) > \frac{v-u}{2} \text{ and } \lambda(\Delta_v \cap (u,v)) > \frac{v-u}{2}.$$

Simply select a point w in that set and notice that

$$\pi = \{([u,w],u),\, ([w,v],w)\}$$

is a partition of $[u,v]$ that is contained in β.

Now, using similar ideas, observe that if $[u,v] \subset [c,d]$ and $u \in E$, $v \in E_n \cap (u, u+\delta(u))$, then there is a partition of $[u,v]$ that is contained in β. Again, if $[u,v] \subset [c,d]$ and $v \in E$, $u \in E_n \cap (v, v-\delta(v))$, then there is a partition of $[u,v]$ that is contained in β.

These observations are enough for us to determine that β contains a partition of every closed subinterval of $[c,d]$. For example to find a partition of $[c,d]$ itself one possibility is that we can select points $c < c_1 < d_1 < d$ so that c_1,

[3] See Section 6.17 for a full account. Many textbooks, including our textbooks [88] and [13], also have extensive instructional materials on Baire category arguments.

$d_1 \in E$ and $(c,c_1) \cap E = (d_1,d) \cap E = \emptyset$. (Note that (c,c_1) and (d_1,d) would necessarily be regular intervals.) If c_1 and d_1 are in E_n we would be done. We would have a partition from β of each of the intervals $[c,c_1]$, $[c_1,d_1]$, and $[d_1,d]$.

If c_1 and d_1 are not in E_n then we would continue, using the fact that E_n is dense in E here, and select $d_2 > c_2$ so that $c_2 \in E_n \cap (c_1, c_1 + \delta(c_1))$ and $d_2 \in E_n \cap (d_1 - \delta(d_1), d_1)$. We know that β contains a partition of each interval $[c,c_1]$, $[c_1,c_2]$, $[c_2,d_2]$, $[d_2,d_1]$, and $[d_1,d]$. Consequently β contains a partition of $[c,d]$. In this way we find that β contains a partition of each subinterval of $[c,d]$ which is impossible unless $E = \emptyset$ and so $R = \mathbb{R}$. ∎

4.5 Additivity

Lebesgue measure is *subadditive* in general on the union of two sets E_1 and E_2. The subadditivity formula is

$$\lambda(A \cap (E_1 \cup E_2)) \leq \lambda(A \cap E_1) + \lambda(A \cap E_2)$$

We know that this same subadditivity formula holds for a sequence of sets $\{E_i\}$:

$$\lambda\left(A \cap \left(\bigcup_{i=1}^{\infty} E_i\right)\right) \leq \sum_{i=1}^{\infty} \lambda(A \cap E_i).$$

We now ask for conditions under which we can claim equality (not inequality). The additivity formula we wish to investigate is

$$\lambda\left(A \cap \left(\bigcup_{i=1}^{\infty} E_i\right)\right) = \sum_{i=1}^{\infty} \lambda(A \cap E_i)?$$

Our first observation is that this is possible if the sets $\{E_i\}$ are separated by open sets. This means merely that there exist open sets G_i and G_j that have no point in common, with $E_i \subset G_i$ and $E_j \subset G_j$. This is stronger than the requirement that E_i and E_j have no point in common. But note that two disjoint closed sets can always be separated in this fashion.

Lemma 4.9 *Let E_1 and E_2 be sets that are separated by open sets. Then, for any set A*

$$\lambda(A \cap (E_1 \cup E_2)) = \lambda(A \cap E_1) + \lambda(A \cap E_2).$$

Proof. Let us use the full version λ^*. We know that

$$\lambda^*(A \cap (E_1 \cup E_2)) \leq \lambda^*(A \cap E_1) + \lambda^*(A \cap E_2).$$

Let us prove the opposite direction. Let β be any full cover of $A \cap (E_1 \cup E_2)$. Select G_1 and G_2, disjoint open sets containing E_1 and E_2 (respectively). Then $\beta(G_1 \cup G_2)$ is necessarily a full cover of $A \cap (E_1 \cup E_2)$. Note that $\beta(G_1)$ is a full cover of $A \cap E_1$ and that $\beta(G_2)$ is a full cover of $A \cap E_2$. If $t_1 < \lambda^*(A \cap E_1)$ and

$t_2 < \lambda^*(A \cap E_2)$ then there must be subpartitions $\pi_1 \subset \beta(G_1)$ and $\pi_2 \subset \beta(G_2)$ with

$$\sum_{([u,v],w)\in\pi_1} (v-u) > t_1$$

and

$$\sum_{([u,v],w)\in\pi_2} (v-u) > t_2.$$

It follows that β contains a subpartition $\pi = \pi_1 \cup \pi_2$ for which

$$\sum_{([u,v],w)\in\pi} (v-u) > t_1 + t_2.$$

From this we deduce that

$$\lambda^*(A \cap (E_1 \cup E_2)) > t_1 + t_2.$$

Then

$$\lambda^*(A \cap (E_1 \cup E_2)) \geq \lambda^*(A \cap E_1) + \lambda^*(A \cap E_2)$$

follows. ■

Corollary 4.10 *Let $E_1, E_2, E_3, \dots$ be a sequence of pairwise disjoint subsets of $\mathbb{R}$ and write*

$$E = \bigcup_{i=1}^{\infty} E_i.$$

Suppose that each pair of sets in the sequence are separated by open sets. Then, for any set A,

$$\lambda(A \cap E) = \sum_{i=1}^{\infty} \lambda(A \cap E_i).$$

Proof. We know from the usual measure properties that

$$\lambda(A \cap E) \leq \sum_{i=1}^{\infty} \lambda(A \cap E_i).$$

We also know that

$$\lambda(A \cap (E_1 \cup E_2)) = \lambda(A \cap E_1) + \lambda(A \cap E_2).$$

An inductive argument would show, too, that for any $n > 1$,

$$\lambda(A \cap (E_1 \cup E_2 \cdots \cup E_n)) = \lambda(A \cap E_1) + \lambda(A \cap E_2) + \cdots + \lambda(A \cap E_n).$$

Thus, from the monotonicity property of measures,

$$\sum_{i=1}^{n} \lambda(A \cap E_i) \leq \lambda(A \cap E) \leq \sum_{i=1}^{\infty} \lambda(A \cap E_i).$$

From this the corollary evidently follows. ■

Corollary 4.11 *Let E_1, E_2, E_3, ... be a sequence of pairwise disjoint closed subsets of $\mathbb{R}$ and write $E = \bigcup_{i=1}^{\infty} E_i$. Then, for any set A,*

$$\lambda(A \cap E) = \sum_{i=1}^{\infty} \lambda(A \cap E_i).$$

To push the countable additivity one step further we use the previous corollary in a natural way. This looks like a highly technical lemma, but it is the basis and motivation for our definition of measurable sets and the theory is more natural than it might appear. The proof is left as an exercise; working through a proof should make it clear how and why the measurability definition in the next section works.

Lemma 4.12 *Let $E_1, E_2, E_3, \ldots$ be a sequence of pairwise disjoint subsets of $\mathbb{R}$ and write $E = \bigcup_{i=1}^{\infty} E_i$. Suppose that for every $\varepsilon > 0$ and for every n there is an open set G_n so that $E_n \setminus G_n$ is closed and so that $\lambda(G_n) < \varepsilon$. Then, for any set A,*

$$\lambda(A \cap E) = \sum_{i=1}^{\infty} \lambda(A \cap E_i).$$

4.6 Measurable sets

4.6.1 Definition of measurable sets

Definition 4.13 *An arbitrary subset E of $\mathbb{R}$ is* measurable[a] *if for every $\varepsilon > 0$ there is an open set G with $\lambda(G) < \varepsilon$ and so that $E \setminus G$ is closed.*

[a] Most advanced courses will start with a different definition of measurable and later on show that this property used here is equivalent in certain settings. See Section 4.8.2 for the connections.

Thus a set is measurable if it is "almost closed." Immediately from this definition we see that all closed sets are measurable and that all null sets are measurable. The definition is exactly designed to produce the following Theorem.

Theorem 4.14 *Let $E_1, E_2, E_3, \ldots$ be a sequence of pairwise disjoint measurable subsets of $\mathbb{R}$ and write $E = \bigcup_{i=1}^{\infty} E_i$. Then, for any set A,*

$$\lambda(A \cap E) = \sum_{i=1}^{\infty} \lambda(A \cap E_i).$$

Proof. This follows immediately from Lemma 4.12. ∎

4.6.2 Properties of measurable sets

The union and intersection of a sequence of measurable sets is also a measurable set. The complement of a measurable set is also a measurable set. This is described often by asserting that the class of measurable sets forms a Borel family.

Theorem 4.15 *The class of all measurable subsets of $\mathbb{R}$ forms a Borel family[a] that contains all closed sets and all null sets.*

[a]The definition of a Borel family is outlined in the proof.

Proof. The class of all measurable subsets of $\mathbb{R}$ forms a Borel family: it a collection of sets that is closed under the formation of unions and intersections of sequences of its members, and contains the complement of each of its members. Here are the details of the proof. Items (3), (4), and (5) are specifically the requirements that the class of measurable sets forms a Borel family.

We prove that the family of all measurable sets has the following properties:

1. Every null set is measurable.
2. Every closed set is measurable.
3. If $E_1, E_2, E_3,$ is a sequence of measurable sets then the union $\bigcup_{n=1}^{\infty} E_n$ is also measurable.
4. If $E_1,\ E_2,\ E_3,$ is a sequence of measurable sets then the intersection $\bigcap_{n=1}^{\infty} E_n$ is also measurable.
5. If E is measurable then the complement $\mathbb{R}\setminus E$ is also measurable.

Items (1) and (2) are easy. Let us prove (5) first. Let E be measurable and E' is its complement. Let $\varepsilon > 0$ and choose an open set G_1 so that $E\setminus G_1$ is closed and $\lambda(G_1) < \varepsilon/2$. Let O be the complement of $E\setminus G_1$; evidently O is open.

First find an open set G_2 with $\lambda(G_2) < \varepsilon/2$ so that $O\setminus G_2$ is closed. [Simply display the component intervals of O, handle the infinite components first, and then a finite number of the bounded components.] Now observe that

$$E'\setminus(G_1\cup G_2) = O\setminus G_2$$

is a closed set while $G_1\cup G_2$ is an open set with measure smaller than ε. This verifies that E' is measurable.

Now check (e): let $\varepsilon > 0$ and choose open sets G_n so that $\lambda(G_n) < \varepsilon 2^{-n}$ and each $E_n\setminus G_n$ is closed. Observe that the set $G = \bigcup_{n=1}^{\infty} G_n$ is an open set for which

$$\lambda(G) \le \sum_{n=1}^{\infty} \lambda(G_n) \le \sum_{n=1}^{\infty} \varepsilon 2^{-n} = \varepsilon.$$

Finally

$$E' = E\setminus G = \bigcap_{n=1}^{\infty}(E_n\setminus G_n)$$

is closed.

For (4), write E'_n for the complementary set to E_n. Then the complement of the set $A = \bigcup_{n=1}^{\infty} E_n$ is the set $B = \bigcap_{n=1}^{\infty} E'_n$. Each E'_n is measurable by (5) and hence B is measurable by (d). The complement of B, namely the set A, is measurable by (5) again. ∎

4.6.3 Increasing sequences of sets

If

$$E_1 \subset E_2 \subset E_3 \subset \dots$$

is an increasing sequence of sets then we would expect that

$$\lambda\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \lambda(E_n).$$

This is particularly easy to prove if the sets are measurable. We show that this identity holds in general.

Theorem 4.16 *Suppose that $\{E_n\}$ is an increasing sequence of sets. Then*

$$\lambda\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \lambda(E_n).$$

Proof. Suppose first that the sets are measurable. Then simply write $E_0 = \emptyset$ and $A_n = E_n \setminus E_{n-1}$ for each $n = 1, 2, 3, \dots$. Then these sets are also measurable and Lemma 4.14 shows us that

$$\lambda\left(\bigcup_{n=1}^{\infty} E_n\right) = \lambda\left(\bigcup_{n=1}^{\infty} A_n\right)$$

$$= \sum_{n=1}^{\infty} \lambda(A_n) = \sum_{n=1}^{\infty} (\lambda(E_n) - \lambda(E_{n-1})) = \lim_{n\to\infty} \lambda(E_n).$$

Now we drop the assumption that the sets $\{E_n\}$ are measurable. Observe first that

$$\lambda\left(\bigcup_{n=1}^{\infty} E_n\right) \geq \lim_{m\to\infty} \lambda(E_m)$$

merely because each set E_m is contained in this union.

To prove the opposite inequality, begin by choosing measurable sets $H_n \supset E_n$ with the same measures, i.e., so that $\lambda(E_n) = \lambda(H_n)$. (For example, start with a sequence of open sets G_{nm} containing E_n with $\lambda(E_n) \leq \lambda(G_{nm}) \leq \lambda(E_n) + 1/n$ and take $H_n = \bigcap_{m=1}^{\infty} G_{nm}$.)

Write $V_m = \bigcap_{k=m}^{\infty} H_k$ and $V = \bigcup_{m=1}^{\infty} V_m$. These sets are all measurable because we choose the $\{H_k\}$ to be measurable. We obtain

$$\lambda(V) = \lim_{m\to\infty} \lambda(V_m).$$

But $E_m \subset V_m \subset H_m$ so that $V \supset E$ and $\lambda(E_m) = \lambda(V_m) = \lambda(H_m)$. Consequently

$$\lambda\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \lambda(V) = \lim_{m\to\infty} \lambda(V_m) = \lim_{m\to\infty} \lambda(E_m).$$

This completes the proof. ∎

Exercise 202 (Borel–Cantelli) *Let $\{A_n\}$ be a sequence of measurable sets with $\sum_{n=1}^{\infty} \lambda(A_n) < \infty$. Then*

$$\lambda\left(\limsup_{n\to\infty} A_n\right) = 0.$$

Here by the limsup of a sequence of sets we mean the collection of points that belong to infinitely many of the sets in the sequence. This can be defined also as

$$\limsup_{n\to\infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n.$$

□

4.6.4 Existence of nonmeasurable sets

We turn now to a search for Lebesgue nonmeasurable sets. The first proof that nonmeasurable sets must exist is due to G. Vitali (1875–1932). It uses the axiom of choice which has to this point not been needed in the text.

Theorem 4.17 *There exist subsets of $\mathbb{R}$ that are not Lebesgue measurable.*

Proof. Let $I = [-\frac{1}{2}, \frac{1}{2}]$. We define an equivalence relation on this interval by relating points to rational numbers; we use Q to denote the set of all rationals. For $x, y \in I$ write $x \sim y$ if $x - y \in Q$. For all $x \in I$, let

$$K(x) = \{y \in I : x - y \in Q\} = \{x + r \in I : r \in Q\}.$$

We show that $\sim$ is an equivalence relation. It is clear that $x \sim x$ for all $x \in I$ and that if $x \sim y$ then $y \sim x$. To show transitivity of $\sim$, suppose that $x, y, z \in I$ and $x - y = r_1$ and $y - z = r_2$ for $r_1, r_2 \in Q$. Then $x - z = (x - y) + (y - z) = r_1 + r_2$, so $x \sim z$. Thus the set of all equivalence classes $K(x)$ forms a partition of I: $\bigcup_{x\in I} K(x) = I$, and if $K(x) \neq K(y)$, then $K(x) \cap K(y) = \emptyset$.

Let A be a set containing exactly one member of each equivalence class. (The existence of such a set A follows from the axiom of choice.) We show that A is nonmeasurable. Let $0 = r_0, r_1, r_2, \ldots$ be an enumeration of $Q \cap [-1, 1]$, and define

$$A_k = \{x + r_k : x \in A\}$$

so that A_k is obtained from A by the translation $x \to x + r_k$.

Then

$$[-\frac{1}{2},\frac{1}{2}] \subset \bigcup_{k=0}^{\infty} A_k \subset [-\tfrac{3}{2},\tfrac{3}{2}]. \tag{4.4}$$

To verify the first inclusion, let $x \in [-\frac{1}{2},\frac{1}{2}]$ and let x_0 be the representative of $K(x)$ in A. We have $\{x_0\} = A \cap K(x)$. Then $x - x_0 \in Q \cap [-1,1]$, so there exists k such that $x - x_0 = r_k$. Thus $x \in A_k$. The second inclusion is immediate: the set A_k is the translation of $A \subset [-\frac{1}{2},\frac{1}{2}]$ by the rational number $r_k \in [-1,1]$.

Suppose now that A is measurable. It is easy to see that then each of the translated sets A_k is also measurable and that $\lambda(A_k) = \lambda(A)$ for every k. But the sets $\{A_i\}$ are pairwise disjoint. If $z \in A_i \cap A_j$ for $i \neq j$, then $x_i = z - r_i$ and $x_j = z - r_j$ are in different equivalence classes. This is impossible, since $x_i - x_j \in Q$. It now follows from (4.4) and the countable additivity of λ for measurable set that

$$1 = \lambda([-\frac{1}{2},\frac{1}{2}]) \leq \lambda(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \lambda(A_k) \leq \lambda([-\tfrac{3}{2},\tfrac{3}{2}]) = 3. \tag{4.5}$$

Let $\alpha = \lambda(A) = \lambda(A_k)$. From (4.5), we infer that

$$1 \leq \alpha + \alpha + \cdots \leq 3. \tag{4.6}$$

But it is clear that no number α can satisfy both inequalities in (4.6). The first inequality implies that $\alpha > 0$, but the second implies that $\alpha = 0$. Thus A is nonmeasurable. ■

The proof has invoked the axiom of choice in order to construct the nonmeasurable set. One might ask whether it is possible to give a more constructive proof, one that does not use this principle. This question belongs to the subject of logic rather than analysis, and the logicians have answered it. In 1964, R. M. Solovay showed that, in Zermelo–Fraenkel set theory with a weaker assumption than the axiom of choice, it is consistent that all sets are Lebesgue measurable. On the other hand, the existence of nonmeasurable sets does not imply the axiom of choice. Thus it is no accident that our proof had to rely on the axiom of choice: it would have to appeal to some further logical principle in any case.[4]

4.7 Measurable functions

The most important class of functions defined on the real line are the measurable functions. All of the functions that one normally encounters are measurable.

[4]See also K. Ciesielski, "How good is Lebesgue measure?" *Math. Intelligencer* 11(2), 1989, pp. 54–58, for a discussion of material related to this section and for references to the literature. That same author's text, *Set Theory for the Working Mathematician*, Cambridge University Press, London (1997) is an excellent source for students wishing to go deeper into these ideas.

Definition 4.18 *An arbitrary function $f : \mathbb{R} \to \mathbb{R}$ is* measurable *if for any real number r*

$$A_r = \{x \in \mathbb{R} : f(x) < r\}$$

is a measurable set.

A function $f : [a,b] \to \mathbb{R}$ would be measurable if there is a measurable function $g : \mathbb{R} \to \mathbb{R}$ and $f(x) = g(x)$ for all $x \in [a,b]$.

Exercises

Exercise 203 *Show that the function $f(x) = \chi_A(x)$ is measurable if and only if the set A is a measurable set.* □

Exercise 204 *Show that a function $f : \mathbb{R} \to \mathbb{R}$ is measurable if and only if $f^{-1}(G)$ is a measurable set for each open set $G \subset \mathbb{R}$.* □

Exercise 205 *Suppose that, for $n = 1,2,3,\dots,n$, each function $f_n : \mathbb{R} \to \mathbb{R}$ is measurable. Show that the function*

$$g(x) = f_1(x) + f_2(x) + f_3(x) + \cdots + f_n(x)$$

is measurable. □

Exercise 206 *Suppose that, for $n = 1,2,3,\dots,n$, each function $f_n : \mathbb{R} \to \mathbb{R}$ is measurable. Show that the functions*

$$g(x) = \max\{f_1(x), f_2(x), f_3(x), \dots, f_n(x)\}$$

and

$$h(x) = \min\{f_1(x), f_2(x), f_3(x), \dots, f_n(x)\}$$

are measurable. □

Exercise 207 *Let $f : \mathbb{R} \to \mathbb{R}$ be measurable. Show that the functions*

$$|f|(x) = |f(x)| = \max\{f(x), -f(x)\},$$
$$[f]^+(x) = \max\{f(x), 0\},$$

and

$$[f]^-(x) = \max\{0, -f(x)\},$$

are measurable. Answer □

4.7.1 Measurable functions are almost bounded

Measurable functions defined on a compact interval $[a,b]$ can be considered to be almost bounded, i.e., bounded once a small set of points is removed. This is a prelude to a deeper theorem of Lusin which we prove later on. In Lusin's theorem it will be shown that, not merely are all measurable functions on a compact interval almost bounded—they are almost continuous.

Theorem 4.19 *Suppose that $f : [a,b] \to \mathbb{R}$ is measurable. Then f is "almost bounded" on $[a,b]$ in the sense that, for every $\varepsilon > 0$, there is an open set G with $\lambda(G) < \varepsilon$ such that f is bounded on the closed set $[a,b] \setminus G$.*

Proof. For each integer n, let

$$E_n = \{x \in [a,b] : |f(x)| > n\}.$$

These sets all have finite Lebesgue measure, they form a decreasing sequence of sets and

$$\bigcap_{n=1}^{\infty} E_n = \emptyset.$$

Consequently $\lambda(E_n) \to 0$. Now just take an integer N so that $\lambda(E_N) < \varepsilon$ and use $E = E_N$. Observe that $|f(x)| \leq N$ for all $x \in [a,b] \setminus E$. Since $\lambda(E) < \varepsilon$ we can also find an open set $G \supset E$ for which $\lambda(G) < \varepsilon$ and for which the statement of the theorem must hold. ∎

Note This argument fails for measurable functions defined on a set of infinite measure. We have explicitly used in the proof the fact that $\lambda([a,b]) < \infty$. In fact the theorem is untrue: take $f(x) = x$ defined for all real numbers x. There is no small open set G, not even any open set G of finite measure, for which f is bounded on $\mathbb{R} \setminus G$.

4.7.2 Continuous functions are measurable

Lemma 4.20 *A function $f : \mathbb{R} \to \mathbb{R}$ that is continuous everywhere is measurable.*

Proof. To prove that f is measurable we need to verify that, for any real number r,

$$A_r = \{x \in \mathbb{R} : f(x) < r\}$$

is a measurable set. But we already know that, for continuous functions, such sets are open. As all open sets (by Theorem 4.15) are measurable we are done. ∎

We know too that a continuous function $f : [a,b] \to \mathbb{R}$ is also measurable by our definition since f agrees on $[a,b]$ with the continuous function g defined by $g(t) = f(t)$ for $a \leq t \leq b$, $g(t) = g(b)$ for $t > b$, and $g(t) = g(a)$ for $t < a$.

Exercise 208 (composition) *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function. Show that the composition $f \circ g$ must be measurable but that the composition $g \circ f$ may not be.* Answer □

4.7.3 Derivatives are measurable

Suppose that $f:\mathbb{R}\to\mathbb{R}$ is almost everywhere the derivative of some function. Then f is measurable[5]. If we combine that fact with the definition of the calculus integral we see that all integrable functions must be measurable.

Lemma 4.21 *A function $f:\mathbb{R}\to\mathbb{R}$ that is almost everywhere the derivative of some function is measurable.*

Proof. We suppose that $F:\mathbb{R}\to\mathbb{R}$ and $F'(x)=f(x)$ almost everywhere, say everywhere in $\mathbb{R}\setminus N$ where N is a set of measure zero. Consider the set $E=\{x:\overline{D}F(x)>r\}$ for any r. Let m, n be positive integers and define β_{mn} to be the covering relation consisting of all pairs $([u,v],w)$ for which $u\le w\le v$, and for which $0<v-u<1/m$ and

$$\frac{F(v)-F(u)}{v-u}\ge r+1/n.$$

Write

$$E_{mn}=\bigcup\{[u,v]:([u,v],w)\in\beta_{mn}\}.$$

Each set E_{mn} is thus a fairly simple object: it is a union of a family of compact intervals. In Lemma 2.17 we have seen that this means it has a simple structure: it differs from an open set by a countable set. In particular each E_{mn} is an measurable set. We check that

$$E=\bigcup_{n=1}^{\infty}\bigcap_{m=1}^{\infty}E_{mn}. \tag{4.7}$$

To begin suppose that $x\in E$. Then $\overline{D}F(x)>r$. There must be at least one integer n with $\overline{D}F(x)>r+1/n$. Moreover, for every integer m there would have to be at least one compact interval $[u,v]$ containing x with length less than $1/m$ so that

$$\frac{F(v)-F(u)}{v-u}\ge r+1/n.$$

Hence x is a point in the set on the right-hand side of the proposed identity. Conversely, should x belong to that set, then there is at least one n so that for all m, x belongs to E_{mn}. It would follow that $\overline{D}F(x)>r$ and so $x\in E$.

The identity (4.7) now exhibits E as a combination of sequences of measurable sets and so E too is an measurable set because the measurable sets form a Borel family (Theorem 4.15). Finally then

$$\{x:f(x)>r\}=\big(\{x:\overline{D}F(x)>r\}\cap[\mathbb{R}\setminus N]\big)\cup N'$$

where N' is an appropriate subset of N. This exhibits the set $\{x:f(x)>r\}$ as

[5] A theorem of Lusin states the converse: if f is measurable then there is a continuous function F for which $F'(x)=f(x)$ almost everywhere. This should not be confused with the fundamental theorem of the calculus.

the union of a measurable set and a set of measure zero. Consequently that set is measurable. This is true for all r and verifies that f is a measurable function. ■

Exercise 209 *Let $f : \mathbb{R} \to \mathbb{R}$. Show that the set of points where an arbitrary function is differentiable is a measurable set.* Answer □

4.7.4 Integrable functions are measurable

If we combine Lemma 4.21 with the descriptive definition of the integral we see that all integrable functions must be measurable.

Lemma 4.22 *If $f : [a,b] \to \mathbb{R}$ is integrable then f is measurable.*

4.7.5 Simple functions

A function $f : \mathbb{R} \to \mathbb{R}$ is *simple* if there is a finite collection of measurable sets $E_1, E_2, E_3, \dots, E_n$ and real numbers $r_1, r_2, r_3, \dots, r_n$ so that

$$f(x) = \sum_{k=1}^{n} r_k \chi_{E_k}(x)$$

for all real x.

Lemma 4.23 *Any simple function is measurable.*

Proof. Suppose that

$$f(x) = \sum_{k=1}^{n} r_k \chi_{E_k}(x)$$

and s is any real number. It is easy to sort out, for any value of s, exactly what the set

$$A_s = \{x : f(x) < s\}$$

must be in terms of the sets $\{E_k\}$. In each case we see that A_s is some simple combination of measurable sets and so is itself measurable. ■

4.7.6 Measurable functions are almost simple

Simple functions are easily seen to be measurable. In a sense all measurable functions are very close to being simple themselves. We show that bounded measurable functions can be written as the uniform limit of simple functions. All nonnegative measurable functions can be written as the limit of a monotonic sequence of simple functions, or equivalently as the sum of a series of simple functions.

Theorem 4.24 *If $f : \mathbb{R} \to \mathbb{R}$ is a bounded, measurable function then f is the uniform limit of a sequence of simple functions $\{f_k\}$.*

Proof. To simplify the arithmetic, start with the situation of a nonnegative, measurable function $f : [0,\infty) \to \mathbb{R}$. One simply checks that the following procedure expresses any such function $f : [0,\infty) \to \mathbb{R}$ as a nondecreasing limit of a sequence $\{f_k\}$ of simple functions: Fix an integer k. Subdivide $[0,k]$ into subintervals

$$[(j-1)2^{-k}, j2^{-k}] \quad (j = 1,2,3,\dots,k2^k)$$

and, for all $x \in [a,b]$, define $f_k(x)$ to be $(j-1)2^{-k}$ if

$$(j-1)2^{-k} \le f(x) < j2^{-k}$$

and to be k if $f(x) \ge k$. If f is bounded then this limit is in fact uniform. The general case of a bounded measurable function $f : \mathbb{R} \to \mathbb{R}$ is easily deduced from this. ∎

For unbounded functions we must certainly drop the uniform convergence since all simple functions are bounded. We do have the following, however, whose proof is contained in the paragraph above.

Theorem 4.25 *Every nonnegative, measurable function $f : \mathbb{R} \to \mathbb{R}$ can be written as a nondecreasing limit of a sequence of simple functions $\{f_k\}$.*

Here too is an equivalent formulation with a convenient algorithm for determine the sequence of simple functions.

Theorem 4.26 *Every nonnegative, measurable function $f : \mathbb{R} \to \mathbb{R}$ can be written as the sum of a series of nonnegative simple functions by the following inductive procedure: Take $\{r_k\}$ to be any sequence of positive numbers for which $r_k \to 0$ and $\sum_{k=1}^{\infty} r_k = +\infty$. Define the sets*

$$A_k = \left\{ x : f(x) \ge r_k + \sum_{j<k} r_j \chi_{A_j}(x) \right\}$$

inductively, starting with $A_0 = \emptyset$. Then

$$f(x) = \sum_{k=1}^{\infty} r_k \chi_{A_k}(x)$$

at every x.

The proof is just a matter of deciding whether and why this works.

4.7.7 Limits of measurable functions

In many instances all one might know of a function is that it is a limit of a sequence of known functions or the sum of an infinite series of such functions. The next theorem asserts that we can deduce the measurability of the function from the separate functions in the limit or the sum.

Theorem 4.27 *Let $f_n : \mathbb{R} \to \mathbb{R}$ be a sequence of measurable functions. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a function for which*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for almost every x. Then f is measurable.

Proof. We fix a real number r and verify that

$$\{x \in \mathbb{R} : f(x) < r\}$$

is a measurable set. We use the fact that sets of the form

$$\{x \in \mathbb{R} : f_n(x) < s\}$$

are measurable. This follows from the measurability of each function f_n.

Let N be the null set consisting of points x where we do not have

$$f(x) = \lim_{n\to\infty} f_n(x)$$

and let $E = \mathbb{R} \setminus N$. Then both E and N are measurable.

We claim the following set identity:

$$\{x \in E : f(x) < r\} = \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{x \in E : f_n(x) < r - 1/k\}.$$

This is a matter of close interpretation. If x_0 belongs to the simple set on the left of the proposed identity, then $x_0 \in E$ and $f(x_0) < r$. There must exist a k so that $f(x_0) < r - 1/k$. Then there must exist an integer m so that

$$f_n(x) < r - 1/k$$

for all $n \geq m$. That places x_0 in the set on the right.

In the other direction if x_0 belongs to the complicated set on the right of the proposed identity, then for some k and m, $f_n(x_0) < r - 1/k$ for all $n \geq m$. It follows that

$$f(x_0) \leq r - 1/k < r.$$

That places x_0 in the set on the left.

Each set

$$\{x \in E : f_n(x) < r - 1/k\} = E \cap \{x \in \mathbb{R} : f_n(x) < r - 1/k\}$$

thus is measurable since it is the intersection of a measurable set and an open set. As measurable sets form a Borel family the intersections and unions of these sets remain measurable.

Finally then

$$\{x \in \mathbb{R} : f(x) < r\}$$

is seen to be the union of the measurable set

$$\{x \in E : f(x) < r\}$$

and some subset of N. This checks the measurability of the function f. ∎

Corollary 4.28 *Let $g_n : \mathbb{R} \to \mathbb{R}$ be a sequence of measurable functions. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a function for which*

$$f(x) = \sum_{n=1}^{\infty} g_n(x)$$

for almost every x. Then f is measurable.

4.8 Construction of the integral

We now give Lebesgue's construction of the integral in a series of steps, starting with characteristic functions, then simple functions, then nonnegative measurable functions, and finally all absolutely integrable functions.

4.8.1 Characteristic functions of measurable sets

Lemma 4.29 *Let E be a subset of an interval $[a,b]$. Then the function χ_E is integrable on $[a,b]$ if and only if E is a measurable set, and in that case*

$$\lambda(E) = \int_a^b \chi_E(x)\,dx.$$

Proof. For any set $E \subset [a,b]$, measurable or not, we can easily establish the (Exercise 197) identity

$$\lambda^*(E) = \overline{\int_a^b} \chi_E(x)\,dx.$$

The two concepts in this identity are defined by the same process. Thus the proof of the lemma depends only on showing that integrability of $\chi_E(x)$ is equivalent to the measurability of E.

We already know that if $\chi_E(x)$ is integrable then it is a measurable function. But this can happen only if E is a measurable set. Conversely let us suppose that E is measurable and verify that χ_E is integrable on $[a,b]$. In fact we show that this function satisfies the McShane criterion on this interval (see Theorem 3.5).

Since E is measurable we know that

$$\lambda(E) + \lambda([a,b] \setminus E) = b - a.$$

Let $\varepsilon > 0$. Select open sets $E \subset G_1$ and $[a,b] \setminus G_2$ so that

$$\lambda(G_1) < \lambda(E) + \varepsilon/2$$

and

$$\lambda(G_2) < \lambda([a,b] \setminus E) + \varepsilon/2.$$

Then, use the identity

$$\lambda(G_1 \cup G_2) = \lambda(G_1) + \lambda(G_2) - \lambda(G_1 \cap G_2)$$

to get

$$\lambda(G_1 \cap G_2)$$

$$= \lambda(G_1) + \lambda(G_2) - \lambda(G_1 \cup G_2)$$
$$< [\lambda(E) + \varepsilon/2] + [\lambda([a,b] \setminus E) + \varepsilon/2] - (b-a) = \varepsilon.$$

This will enable us to apply the McShane criterion to establish that χ_E is integrable on $[a,b]$. Define β as the collection of all pairs $([u,v],w)$ for which either

$$w \in E \text{ and } [u,v] \subset G_1$$

or else

$$w \in [a,b] \setminus E \text{ and } [u,v] \subset G_2.$$

This is a full cover of $[a,b]$.

Choose any two partitions π, π' of $[a,b]$ contained in β. We compute

$$\sum_{([u,v],w)\in\pi} \sum_{([u',v'],w')\in\pi'} \left|\chi_E(w) - \chi_E(w')\right| \lambda([u,v] \cap [u',v']). \tag{4.8}$$

Note, in this sum, that terms for which both w and w' are in E or for which neither is in E vanish. Terms for which $w \in E$ and $w' \in [a,b] \setminus E$ must have

$$\left|\chi_E(w) - \chi_E(w')\right| = 1,$$

$[u,v] \subset G_1$ and $[u',v'] \subset G_2$. In particular

$$[u,v] \cap [u',v'] \subset (G_1 \cap G_2).$$

The same is true if $w' \in E$ and $w \in [a,b] \setminus E$. Remembering that

$$\lambda(G_1 \cap G_2) < \varepsilon,$$

we see that the sum in (4.8) is smaller than ε. By the McShane criterion χ_E is absolutely integrable on $[a,b]$. ∎

4.8.2 Characterizations of measurable sets

From our studies so far we can obtain a number of characterizations of measurable sets. Our definition of choice was to describe measurable sets as those that are almost closed. The original Lebesgue definition is equivalent to assertion (3) below, but expressed in the language of inner and outer measures. For Lebesgue, assertion (2) would have been interpreted as a definition of integrability, rather than a "property" of measurable sets. Assertion (4) is known as Carathéodory's criterion and is particularly useful in the study of abstract measure theory on spaces more general than the real line. The final assertion (5) is closely related to the McShane condition for Lebesgue integrability.

Theorem 4.30 *Let E be a set of real numbers. Then the following assertions are equivalent:*

(almost closed) *E is measurable.*

(integrability) *χ_E is integrable on every compact interval $[a,b]$.*

(Lebesgue's definition) *For every compact interval $[a,b]$,*

$$\lambda([a,b]\cap E)+\lambda([a,b]\setminus E)=b-a. \tag{4.9}$$

(Carathéodory's criterion) *For every set $T\subset\mathbb{R}$,*

$$\lambda(T)\geq\lambda(T\cap E)+\lambda(T\setminus E). \tag{4.10}$$

(McShane's criterion) *For every $\varepsilon>0$ and every compact interval $[a,b]$, there is a full cover β of $[a,b]$ so that, whenever π, π' are subpartitions of $[a,b]$ with $\pi\subset\beta[E]$ and $\pi'\subset\beta[[a,b]\setminus E]]$,*

$$\sum_{([u,v],w)\in\pi}\ \sum_{([u',v'],w')\in\pi'}\lambda([u,v]\cap[u',v'])<\varepsilon.$$

.

Proof. First note that a set E is measurable if and only if $E\cap[a,b]$ is measurable for every compact interval $[a,b]$. In one direction this is because $[a,b]$ is a measurable set (it is closed) and the intersection of measurable sets is also measurable. In the other direction, if $E\cap[a,b]$ is measurable for every compact interval $[a,b]$, then $E=\bigcup_{n=1}^{\infty}E\cap[-n,n]$ expresses E as a measurable set.

The first three conditions (a), (b), and (c) we have explicitly shown to be equivalent in the proof of the lemma. Let us check that (d) implies (c). Observe that the inequality,

$$\lambda(T)\leq\lambda(T\cap E)+\lambda(T\setminus E)$$

holds in general, so that the condition (4.10) is equivalent to the assertion of equality:

$$\lambda(T)=\lambda(T\cap E)+\lambda(T\setminus E).$$

Thus (c) is a special case of (d) with $T=[a,b]$. On the other hand, (a) implies (d). Measurability of E implies that E and $\mathbb{R}\setminus E$ are disjoint measurable sets for which

$$\lambda(T)=\lambda(T\cap E)+\lambda(T\setminus E)$$

must hold for any set $T\subset\mathbb{R}$. Finally the fifth condition (e) is just a rewriting of the McShane criterion for integrability of the function χ_E on $[a,b]$. We have seen in the proof of the lemma that measurability of $E\cap[a,b]$ is equivalent to that criterion applied to χ_E on $[a,b]$. ∎

4.8.3 Integral of simple functions

Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is *simple* if there is a finite collection of measurable sets $E_1, E_2, E_3, \dots, E_n$ and real numbers $r_1, r_2, r_3, \dots, r_n$ so that

$$f(x) = \sum_{k=1}^{n} r_k \chi_{E_k}(x)$$

for all real x. Since this is a finite linear combination it follows from the integration theory and the integration of characteristic functions (Lemma 4.29) that such a function is necessarily integrable on any compact interval $[a,b]$ and that

$$\int_a^b f(x)\,dx = \sum_{k=1}^{n} \left(\int_a^b r_k \chi_{E_k}(x)\,dx \right) = \sum_{k=1}^{n} r_k \lambda(E_k \cap [a,b]).$$

Thus the integral of simple functions can be constructed from the values of the function in a finite number of steps using the Lebesgue measure.

4.8.4 Integral of bounded measurable functions

The value of the integral on a compact interval $[a,b]$ for simple functions has been seen to be constructible directly from the values of the Lebesgue measure itself. This extends, by Theorem 4.24, to all bounded measurable functions since they are uniform limits of simple functions.

The reader might recall that we used a similar constructive argument for the regulated integral of Section 1.9. There, starting with step functions (much less general than simple functions) we were able to extend the integral constructively to all regulated functions. Regulated functions are uniform limits of step functions; bounded measurable functions are uniform limits of simple functions. The theories are completely analogous.

Theorem 4.31 *Let f be a bounded, measurable function on an interval $[a,b]$. Then f is Lebesgue integrable and, for any representation of f as the uniform limit of a sequence of simple functions,*

$$f(x) = \lim_{n\to\infty} f_n(x) \quad (a \le x \le b)$$

the identity

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

must hold.

Proof. This requires only an appeal to the uniform convergence theorem for integrals (see Section 3.3). ∎

4.8.5 Integral of nonnegative measurable functions

We have seen (Theorem 4.26) that every nonnegative measurable function can be represented by simple functions. Consequently the integral of such a function can be constructed.

Theorem 4.32 *Let f be a nonnegative, measurable function on an interval $[a,b]$. Then, for any representation of f as the sum of a series of nonnegative, simple functions*

$$f(x) = \sum_{k=1}^{\infty} f_n(x) \quad (a \leq x \leq b)$$

the identity

$$\int_a^b f(x)\,dx = \sum_{k=1}^{\infty} \left(\int_a^b f_n(x)\,dx \right)$$

must hold (finite or infinite). Moreover f is Lebesgue integrable on $[a,b]$ if and only if this series of integrals converges to a finite value.

Proof. This requires only an appeal to the monotone convergence theorem (see Section 3.2.9). ∎

Corollary 4.33 *Let f be a nonnegative, measurable function on an interval $[a,b]$. Then*

$$\int_a^b f(x)\,dx$$

exists (finitely or infinitely). Moreover f is integrable on $[a,b]$ if and only if this value is finite.

Proof. This follows from the theorem. ∎

4.8.6 Fatou's Lemma

Theorem 4.34 (Fatou's lemma) *Let f_n be a sequence of nonnegative, measurable functions defined at every point of an interval $[a,b]$. Then, assuming that*

$$f(x) = \liminf_{n\to\infty} f_n(x)$$

is finite almost everywhere,

$$\int_a^b \liminf_{n\to\infty} f_n(x)\,dx \leq \liminf_{n\to\infty} \int_a^b f_n(x)\,dx..$$

Proof. Fatou's lemma is proved using the monotone convergence theorem, Theorem 3.15. Let f denote the limit inferior of the f_n. For every natural number k define the function

$$g_k(x) = \inf_{n\geq k} f_n(x).$$

Then the sequence $g_1, g_2, \ldots$ is a nondecreasing sequence of measurable functions and converges pointwise to f. For $k \le n$, we have $g_k(x) \le f_n(x)$, so that

$$\int_a^b g_k(x)\,dx \le \int_a^b f_n(x)\,dx,$$

hence

$$\int_a^b g_k(x)\,dx \le \inf_{n\ge k}\int_a^b f_n(x)\,dx.$$

Using the monotone convergence theorem, the last inequality, and the definition of the limit inferior, it follows that

$$\int_a^b \liminf_{n\to\infty} f_n(x)\,dx = \lim_{k\to\infty}\int_a^b g_k(x)\,dx \le$$

$$\lim_{k\to\infty}\inf_{n\ge k}\int_a^b f_n(x)\,dx = \liminf_{n\to\infty}\int_a^b f_n(x)\,dx.$$

■

Exercises

Exercise 210 *On the interval* $[0,1]$ *for every natural number* n *define*

$$f_n(x) = \begin{cases} n & \text{for } x\in(0,1/n), \\ 0 & \text{otherwise.}\end{cases}$$

Show that

$$\int_0^1 \liminf_{n\to\infty} f_n(x)\,dx < \liminf_{n\to\infty}\int_0^1 f_n(x)\,dx.$$

□

Exercise 211 *On the interval* $[0,\infty)$ *for every natural number* n *define*

$$f_n(x) = \begin{cases} \frac{1}{n} & \text{for } x\in[0,n], \\ 0 & \text{otherwise.}\end{cases}$$

Show that $\{f_n\}$ *is uniformly convergent and that*

$$\int_0^\infty \liminf_{n\to\infty} f_n(x)\,dx < \liminf_{n\to\infty}\int_0^\infty f_n(x)\,dx.$$

□

Exercise 212 *On the interval* $[0,\infty)$ *for every natural number* n *define*

$$f_n(x) = \begin{cases} -\frac{1}{n} & \text{for } x\in[n,2n], \\ 0 & \text{otherwise.}\end{cases}$$

Show that $\{f_n\}$ *is uniformly convergent and that the inequality in Fatou's lemma*

$$\int_0^\infty \liminf_{n\to\infty} f_n(x)\,dx \le \liminf_{n\to\infty}\int_0^\infty f_n(x)\,dx.$$

fails. □

Exercise 213 (reverse Fatou lemma) *Let $\{f_n\}$ be a sequence of measurable functions defined on an interval $[a,b]$. Suppose that there exists a Lebesgue integrable function g on $[a,b]$ such that $f_n \leq g$ for all n. Show that*

$$\int_a^b \limsup_{n\to\infty} f_n(x)\,dx \geq \limsup_{n\to\infty} \int_a^b f_n(x)\,dx.$$

Answer □

4.8.7 Dominated convergence theorem

Fatou's lemma provides a simple proof of a well-known and highly useful convergence theorem. We cannot expect to take limits inside the integral sign merely in the presence of pointwise convergence. This theorem asserts that this is indeed possible if the sequence of functions is controlled by a dominating function,

Theorem 4.35 (dominated convergence) *Let $\{f_n\}$ be a sequence of measurable functions defined on an interval $[a,b]$. Assume that the sequence converges pointwise and is dominated[a] by some nonnegative, Lebesgue integrable function g. Then the pointwise limit is an integrable function and*

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

[a]To say that the sequence is "dominated" by g means that $|f_n(x)| \leq g(x)$ for all natural numbers n and all points x in $[a,b]$.

Proof. If f denotes the pointwise limit of the sequence, then f is also measurable and dominated by g, hence integrable. Furthermore,

$$|f(x) - f_n(x)| \leq 2g(x)$$

for all n and

$$\limsup_{n\to\infty} |f(x) - f_n(x)| = 0.$$

By the reverse Fatou lemma,

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

Using linearity and monotonicity of the integral,

$$\left|\int_a^b f(x)\,dx - \int_a^b f_n(x)\,dx\right| = \left|\int_a^b f(x) - f_n(x))\,dx\right| \leq \int_a^b |f(x) - f_n(x)|\,dx,$$

and the statement is proved. ∎

4.9 Derivatives

4.9.1 Derivative of the integral

It is the case for all of our integrals (since they are all contained in the general Newton integral) that

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x) \tag{4.11}$$

for a.e. point x in an interval $[a,b]$ in which f is assumed to be integrable. For the Lebesgue integral we can add

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

for a.e. point x in an interval in which f is assumed to be Lebesgue integrable. In most presentations of the Lebesgue theory of integration this would now have to be proved. For us, since we started with the Newton integral (i.e., the equivalent Henstock-Kurzweil integral), this is already given.

Lebesgue points A stronger condition than merely the derivative statement in equation (4.11) is described in the following definition.

Definition 4.36 *Suppose that f is Lebesgue integrable on an open interval that includes the point x_0. Then that point is said to be a Lebesgue point for f provided*

$$\lim_{h\to 0}\frac{1}{h}\int_{x_0}^{x_0+h} |f(x)-f(x_0)|\,dx = 0.$$

It is easy to check that every point of continuity of f is necessarily also a Lebesgue point for f. Indeed, a Lebesgue point is a special kind of point of "approximate" continuity. This would be made more precise in advanced courses. It is stronger than merely a point where the derivative exists (see Exercise 215).

Theorem 4.37 *Let f be Lebesgue integrable on $[a,b]$. Then almost every point of $[a,b]$ is a Lebesgue point of f.*

Proof. Let r be any rational number. Then the function $g_r(x) = f(x) - r$ is Lebesgue integrable and thus

$$\lim_{h\to 0}\frac{1}{h}\int_{x}^{x+h} |f(t)-r|\,dt = |f(x)-r| \tag{4.12}$$

for a.e. point x in (a,b). Let

$$E(r) = \{x\in(a,b) : \text{(4.12) fails}\}.$$

Then $\lambda(E(r)) = 0$. Let E be the union of the countable collection of sets $E(r)$ taken over all possible rational numbers r. Then $\lambda(E) = 0$.

We shall how that every point x_0 in $(a,b)\setminus E$ is a Lebesgue point for f. Let x_0 be such a point and let $\varepsilon>0$. Choose a rational number r_n such that

$$|f(x_0)-r_n|<\frac{\varepsilon}{3}. \tag{4.13}$$

We then have

$$||f(x)-r_n|-|f(x)-f(x_0)||<\tfrac{1}{3}\varepsilon.$$

on $[a,b]$ so that

$$\left|\frac{1}{h}\int_{x_0}^{x_0+h}|f(x)-r_n|\,dx-\frac{1}{h}\int_{x_0}^{x_0+h}|f(x)-f(x_0)|\,dx\right|\le\frac{\varepsilon}{3} \tag{4.14}$$

whenever $x_0+h\in[a,b]$. Since $x_0\notin E$, (4.12) applies, so there exists $\delta>0$ such that

$$\left|\frac{1}{h}\int_{x_0}^{x_0+h}|f(x)-r_n|\,dx-|f(x_0)-r_n|\right|<\frac{\varepsilon}{3}$$

if $|h|<\delta$. From (4.13), we infer that, for $|h|<\delta$,

$$\frac{1}{h}\int_{x_0}^{x_0+h}|f(x)-r_n|\,dx<\frac{2\varepsilon}{3}$$

so

$$\frac{1}{h}\int_{x_0}^{x_0+h}|f(x)-f(x_0)|\,dx<\varepsilon \tag{4.15}$$

by (4.14).

We have shown that for all $x_0\notin E$ and every $\varepsilon>0$ there exists $\delta>0$ such that (4.15) holds whenever $|h|<\delta$. Since $\lambda(E)=0$, we conclude that almost every $x\in[a,b]$ is a Lebesgue point of f. ∎

Exercise 214 *Suppose that $f:[a,b]\to\mathbb{R}$ is Lebesgue integrable and that x_0 is a point of continuity of f. Show that x_0 is a Lebesgue point of f. Is the converse true?* □

Exercise 215 *Suppose that $f:[a,b]\to\mathbb{R}$ is Lebesgue integrable and that x_0 is a Lebesgue point of f. Show that $F'(x_0)=f(x_0)$.* □

4.9.2 Lebesgue points and points of approximate continuity

Let $f:\mathbb{R}\to\mathbb{R}$ be a function that is locally Lebesgue integrable, i.e. integrable on each compact interval. Let us compare points of approximate continuity with Lebesgue points and points where the fundamental theorem of the calculus holds.

Approximate continuity will be studied in a bit greater detail below in Section 4.11.2. Let us anticipate the def inion here. Recall that $\underline{d}(E,x_0)$ denotes for us the lower density of the set E at a point x_0 (as defined in Section 4.4).

Definition 4.38 (approximate continuity) *Let f be a function defined in an open set containing a point x_0. If there exists a set E such that*

$$\underline{d}(E,x_0) = 1 \text{ and } \lim_{x\to x_0, x\in E} f(x) = f(x_0),$$

we say that f is approximately continuous *at x_0. If f is approximately continuous at all points of its domain, we simply say that f is* approximately continuous.

We use this language for our theorem:

1. A point x_0 is in $\mathcal{A}_f$ if f is approximately continuous at x_0.
2. A point x_0 is in $\mathcal{L}_f$ if x_0 is a Lebesgue point for f.
3. A point x_0 is in $\mathcal{D}_f$ if $F'(x_0) = f(x_0)$ where F is any indefinite integral for f in a neighborhood of x_0.

The connections are given in the following statement.

Theorem 4.39 *Let $f : \mathbb{R} \to \mathbb{R}$ be a locally Lebesgue integrable function. Then*

1. *$\mathcal{L}_f \subset \mathcal{A}_f$.*
2. *$\mathcal{L}_f \subset \mathcal{D}_f$.*
3. *If f is locally bounded[a] then $\mathcal{A}_f = \mathcal{L}_f \subset \mathcal{D}_f$.*
4. *In general $\mathcal{A}_f \setminus \mathcal{D}_f$ and $\mathcal{D}_f \setminus \mathcal{A}_f$ are both first category and measure zero.*

[a] i.e., bounded on each compact interval.

Proof. Let us begin by noting that the first two statements are easy and can be left to the reader to sort out. We now give a detailed proof for the third statement: we show that $\mathcal{A}_f \subset \mathcal{L}_f$ for bounded functions. We suppose f is a bounded measurable function on $[a,b]$. If f is approximately continuous at $x_0 \in (a,b)$ we verify that x_0 is a Lebesgue point for f. Choose a set E such that $\underline{d}(E,x_0) = 1$ and f is continuous at x_0 restricted to the set E. Write $A = \mathbb{R} \setminus E$. Let M be an upper bound for $|f|$, and let $h > 0$. Then

$$\frac{1}{h}\int_{x_0}^{x_0+h} |f(x) - f(x_0)|\,dx =$$

$$\frac{1}{h}\int_{x_0}^{x_0+h} \chi_E(x)|f(x) - f(x_0)|\,dx + \frac{1}{h}\int_{x_0}^{x_0+h} \chi_A(x)|f(x) - f(x_0)|\,dx.$$

Let $\varepsilon > 0$ and choose $\delta > 0$ such that (i) if $t \in E$ and $|t - x_0| < \delta$ then $|f(t) - f(x_0)| < \varepsilon/2$, and (ii) if $h < \delta$, then

$$\frac{\lambda([x_0, x_0+h] \setminus E)}{h} < \frac{\varepsilon}{4M}.$$

For $h < \delta$, we calculate

$$\frac{1}{h}\int_{x_0}^{x_0+h} |f(x) - f(x_0)|\,dx$$
$$\leq \frac{\varepsilon}{2h}\lambda([x_0, x_0+h]\cap E) + \frac{2M}{h}\lambda([x_0, x_0+h]\setminus E)$$
$$\leq \varepsilon\frac{h}{2h} + 2M\frac{\varepsilon}{4M} = \varepsilon.$$

A similar calculation holds if $h < 0$. Since ε is arbitrary, we conclude that that x_0 is a Lebesgue point for f.

Finally let us address the fourth statement in the theorem. It is clear that $\mathcal{A}_f \setminus \mathcal{D}_f$ and $\mathcal{D}_f \setminus \mathcal{A}_f$ are both sets of measure zero. Indeed we know that $\mathbb{R}\setminus\mathcal{D}_f$ and $\mathbb{R}\setminus\mathcal{A}_f$ are both sets of measure zero. We will rely now on the interested reader to review the necessary notions of Baire category to complete the proof and establish that these sets are also sets of the first category. ■

Exercise 216 *Give an example of an unbounded, locally Lebesgue integrable function $f : \mathbb{R} \to \mathbb{R}$ that illustrates that $\mathcal{A} \subset \mathcal{L}$ can be false in general, i.e., find such a function that is approximately continuous at a point that is not itself also a Lebesgue point.* □

4.9.3 Derivatives of functions of bounded variation

As a consequence of Lebesgue's program to this point we can prove also some facts about derivatives of monotonic functions and derivatives of functions of bounded variation. These are due to Lebesgue, but our proofs are rather easier since we do not need much of the measure theory to obtain them.

Theorem 4.40 *Let $F : [a,b] \to \mathbb{R}$ be a function of bounded variation. Then $F'(x)$ exists almost everywhere in $[a,b]$ and*

$$\int_a^b |F'(x)|\,dx \leq \mathrm{Var}(F, [a,b]).$$

Proof. We know from the Lebesgue differentiation theorem that F is a.e. differentiable. Let $f(x) = |F'(x)|$ at every point at which $F'(x)$ exists and as zero elsewhere. Then f is a nonnegative function. At every point w in $[a,b]$ there is a $\delta > 0$ so that, whenever $u \leq w \leq v$ and $0 < v - u < \delta$,

$$f(w) - \varepsilon \leq \frac{|F(v) - F(u)|}{v - u}.$$

At points w where $f(w) = 0$ this is obvious, while at points w where $F'(w)$ exists this follows from the definition of the derivative.

Take β as the collection of all pairs $([u,v], w)$ subject to the requirement only that

$$|F(v) - F(u)| > [f(w) - \varepsilon](v - u)$$

if $w \in [a,b]$ and $[u,v] \subset [a,b]$. This collection β is a full cover.

Every partition $\pi \subset \beta$ of the interval $[a,b]$ satisfies

$$\sum_{([u,v],w)\in\pi} [f(w)-\varepsilon](v-u) < \sum_{([u,v],w)\in\pi} |F(v)-F(u)| \leq \operatorname{Var}(F,[a,b]).$$

It follows that

$$-\varepsilon(b-a) + \overline{\int_a^b} f(x)\,dx \leq \operatorname{Var}(F,[a,b]).$$

Since ε is an arbitrary positive number,

$$\overline{\int_a^b} f(x)\,dx \leq \operatorname{Var}(F,[a,b]).$$

Since f is almost everywhere a derivative it is necessarily measurable. Thus we may use the integral in place of the upper integral. ■

Corollary 4.41 *Let $F : [a,b] \to \mathbb{R}$ be a nondecreasing function. Then $F'(x)$ exists almost everywhere in $[a,b]$ and*

$$\int_a^b F'(x)\,dx \leq F(b)-F(a).$$

Corollary 4.42 (Lebesgue decomposition) *Let $F : [a,b] \to \mathbb{R}$ be a continuous, nondecreasing function. Then $F'(x)$ exists almost everywhere in $[a,b]$ and*

$$F(t) = \int_a^t F'(x)\,dx + S(t) \quad (a \leq t \leq b)$$

expresses F as the sum of an integral and a continuous, nondecreasing singular[a] function.

[a] In this context *singular* simply indicates a function with a zero derivative almost everywhere.

Proof. Simply define

$$S(t) = F(t) - \int_a^t F'(x)\,dx \quad (a \leq t \leq b).$$

Check that $S'(t) = 0$ almost everywhere (trivial) and so S is singular. That S is continuous is evident since it is the difference of two continuous functions. That S is nondecreasing follows from the theorem, since

$$S(d)-S(c) = F(d)-F(c) - \int_c^d F'(x)\,dx \geq 0$$

for any $[c,d] \subset [a,b]$. ■

4.9.4 Characterization of the Lebesgue integral

Recall that a function f is Lebesgue integrable on an interval $[a,b]$, according to our definition, if both f and $|f|$ are integrable on that interval. Thus we define the integral (in the Henstock-Kurzweil or general Newton sense) first and obtain the Lebesgue integral as a special case. For Lebesgue the next theorem would be appear as a definition.

Theorem 4.43 *Let $f : [a,b] \to \mathbb{R}$. Then f is Lebesgue integrable if and only if f is measurable and*

$$\int_a^b |f(x)|\, dx < \infty.$$

Proof. We know, from Exercise 207, that the functions $|f|$, $[f]^+$, and $[f]^-$ are also measurable. The finiteness of this integral implies (by Corollary 4.33) that each of these functions are integrable. In particular both functions $f = [f]^+ - [f]^-$ and $|f|$ are integrable. Thus f must be absolutely integrable. Conversely if f is absolutely integrable, this means that $|f|$ is integrable and consequently, by definition, it has a finite integral. ∎

Our final theorem for Lebesgue's program shows that the integral is constructible by his methods for all Lebesgue integrable functions. We see in the Section 4.9.6 that this is as far as one can go.

Theorem 4.44 *If f is Lebesgue integrable on a compact interval $[a,b]$ then f, $|f|$, $[f]^+$, and $[f]^-$ are measurable and*

$$\int_a^b |f(x)|\, dx = \int_a^b [f(x)]^+\, dx + \int_a^b [f(x)]^-\, dx$$

and

$$\int_a^b f(x)\, dx = \int_a^b [f(x)]^+\, dx - \int_a^b [f(x)]^-\, dx$$

Proof. If f is Lebesgue integrable then we know that f and $|f|$ are integrable. It follows that $[f]^+ = (f + |f|)/2$ and $[f]^- = (|f| - f)/2$ are both integrable. All functions are measurable since all are integrable. Since

$$|f(x)| = [f(x)]^+ + [f(x)]^-$$

and

$$f(x) = [f(x)]^+ - [f(x)]^-$$

the integration formulas are immediately available. ∎

4.9.5 McShane's Criterion

Lebesgue's integral can also be characterized by the McShane criterion. Using normal inequality techniques we easily observe that the expression

$$\left| \sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} [f(w) - f(w')]\lambda(I \cap I') \right| < \varepsilon \tag{4.16}$$

that we use for the Cauchy criterion must be smaller than a quite similar expression:

$$\left|\sum_{(I,w)\in\pi}\sum_{(I',w')\in\pi'}[f(w)-f(w')]\lambda(I\cap I')\right| \leq$$
$$\sum_{(I,w)\in\pi}\sum_{(I',w')\in\pi'}\left|f(w)-f(w')\right|\lambda(I\cap I').$$

It takes a sharp (and young) eye to spot the difference, but the larger side of this inequality may be strictly larger. This leads to a stronger integrability criterion than that in the Cauchy criterion. This is the motivation for the criterion, named after E. J. McShane. We prove that McShane's criterion is a necessary and sufficient condition for Lebesgue integrability.

Definition 4.45 (McShane's criterion) *A function $f:[a,b]\to\mathbb{R}$ is said to* satisfy McShane's criterion *on $[a,b]$ provided that for all $\varepsilon>0$ a full cover β can be found so that*

$$\sum_{(I,w)\in\pi}\sum_{(I',w')\in\pi'}\left|f(w)-f(w')\right|\lambda(I\cap I')<\varepsilon$$

for all partitions π, π' of $[a,b]$ contained in β.

Theorem 4.46 *Let $f:[a,b]\to\mathbb{R}$. Then f is Lebesgue integrable on an interval if and only if it satisfies McShane's criterion on that interval.*

Proof. It is immediate that if f satisfies McShane's criterion it also satisfies the Cauchy criterion. Thus the function f is integrable. We then observe that, since

$$\left||f(x)|-|f(x')|\right|\leq\left|f(x)-f(x')\right|,$$

it is clear that whenever f satisfies McShane's criterion so too does $|f|$. Thus $|f|$ too is integrable on $[a,b]$. The inequalities of the theorem simply follow from the inequalities $-|f(x)|\leq f(x)\leq|f(x)|$ which hold for all x.

Here is the proof in the other direction. To simplify the notation let us write

$$S(f,\pi,\pi')=\sum_{([u,v],w)\in\pi}\sum_{([u',v'],w')\in\pi'}\left|f(w)-f(w')\right|\lambda([u,v]\cap[u',v']) \tag{4.17}$$

for any two partitions π, π' of $[a,b]$. Some preliminary computations will help. If g_1, g_2, ..., g_n are functions on $[a,b]$ then,

$$S\left(\sum_{i=1}^{n}g_i,\pi,\pi'\right)\leq\sum_{i=1}^{n}S(g_i,\pi,\pi'). \tag{4.18}$$

If

$$\overline{\int_a^b}|f(x)|\,dx<t$$

then there must exist a full cover β with the property that for any two partitions π, π' of $[a,b]$ from β,

$$S(f,\pi,\pi')<2t. \tag{4.19}$$

Finally

$$S(f,\pi,\pi') \leq \sup\{|f(t)| : a \leq t \leq b\} \cdot 2(b-a). \tag{4.20}$$

Each of the statements (4.18), (4.19), and (4.20) require only simple computations that we leave to the reader.

Now for our argument. We assume that f is absolutely integrable and verify the criterion. But f can be written as a difference of two nonnegative integrable functions. If both of these satisfy the criterion then, using (4.18) we deduce that so too does f. Consequently for the remainder of the proof we assume that f is nonnegative and integrable.

The first step is to observe that every characteristic function of a measurable set satisfies the McShane criterion. This is proved in Lemma 4.29. Using (4.18) we easily deduce, as our second step, that every nonnegative simple function also satisfies the McShane criterion.

The third step is to show that every nonnegative, bounded measurable function also satisfies this criterion. But such a function is the uniform limit of a sequence of nonnegative simple functions. It follows then, from (4.20), that such functions satisfy the McShane criterion. For if f is a bounded measurable function, $\varepsilon > 0$, choose a simple function g so that

$$|f(t) - g(t)| < \varepsilon/(4[b-a])$$

for all $a \leq t \leq b$. Now using McShane's criterion on g we can select a full cover β for which $S(g,\pi,\pi') < \varepsilon/2$ for all partitions π, π' of $[a,b]$ from β. Then

$$S(f,\pi,\pi') \leq S(f-g,\pi,\pi') + S(g,\pi,\pi') \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

The final step requires an appeal to the monotone convergence theorem. Set $f_N(t) = \min\{N, f(t)\}$ and use the monotone convergence theorem to find an integer N large enough so that

$$\int_a^b [f(x) - f_N(x)]\,dx < \varepsilon/4.$$

Using (4.19) select a full cover β_1 for which $S(f - f_N,\pi,\pi') < \varepsilon/2$ for all partitions π, π' of $[a,b]$ from β_1. Select a full cover β_2 for which $S(f_N,\pi,\pi') < \varepsilon/2$ for all partitions π, π' of $[a,b]$ from β_2. Then set $\beta = \beta_1 \cap \beta_2$. This is a full cover and we can check that

$$S(f,\pi,\pi') \leq S(f - f_N,\pi,\pi') + S(f_N,\pi,\pi') \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

for all partitions π, π' of $[a,b]$ from β. This verifies the McShane criterion for an arbitrary nonnegative integrable function f. ∎

Exercises

Exercise 217 *Suppose that each of the functions $f_1, f_2, \ldots, f_n : [a,b] \to \mathbb{R}$ satisfies McShane's criterion on a compact interval $[a,b]$ and that a function*

$L:\mathbb{R}^n \to \mathbb{R}$ is given satisfying

$$|L(x_1,x_2,\dots,x_n) - L(y_1,y_2,\dots,y_n)| \leq M\sum_{i=1}^{n}|x_i - y_i|$$

for some number M and all $(x_1,x_2,\dots,x_n)$ and $(y_1,y_2,\dots,y_n)$ in $\mathbb{R}^n$. Show that the function $g(x) = L(f_1(x), f_2(x),\dots,f_n(x))$ satisfies McShane's criterion on $[a,b]$. □

Exercise 218 *Let F, $f:\mathbb{R}\to\mathbb{R}$. A necessary and sufficient condition in order that f be the derivative of F at each point is that for every $\varepsilon > 0$ there is a full cover β of the real line with the property that for every compact interval $[a,b]$ and every partition $\pi \subset \beta$ of $[a,b]$,*

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)| < \varepsilon(b-a). \tag{4.21}$$

Answer □

Exercise 219 (Freiling's criterion) *Let $f:\mathbb{R}\to\mathbb{R}$. Show that necessary and sufficient condition*[6] *in order that f be the derivative of some function F at each point is that for every $\varepsilon > 0$ there is a full cover β of the real line with the property that for every compact interval $[a,b]$ and every pair of partitions π_1, $\pi_2 \subset \beta$ of $[a,b]$,*

$$\left|\sum_{(I,z)\in\pi}\ \sum_{(I',z')\in\pi'} [f(z) - f(z')]\lambda(I\cap I')\right| < \varepsilon\lambda([a,b]). \tag{4.22}$$

Answer □

Exercise 220 *Let $f:\mathbb{R}\to\mathbb{R}$. Characterize the following property: for every $\varepsilon > 0$ there is a full cover β of the real line with the property that for every compact interval $[a,b]$ and every pair of partitions π_1, $\pi_2 \subset \beta$ of $[a,b]$,*

$$\sum_{(I,z)\in\pi}\ \sum_{(I',z')\in\pi'} |f(z) - f(z')|\lambda(I\cap I') < \varepsilon\lambda([a,b]).$$

□

4.9.6 Nonabsolutely integrable functions

A function f is *nonabsolutely integrable* on an interval $[a,b]$ if it is integrable, but not absolutely integrable there, i.e., f is integrable $[a,b]$ but $|f|$ is not integrable. Lebesgue's program will not construct the integral of a nonabsolutely integrable

[6] This is from Chris Freiling, *On the problem of characterizing derivatives*. Real Anal. Exchange 23 (1997/98), no. 2, 805–812. For a different account of the problem of characterizing derivatives return to Section **??**.

function. The only method that his program offers is the hope that

$$\int_a^b f(x)\,dx = \int_a^b [f(x)]^+\,dx - \int_a^b [f(x)]^-\,dx?$$

Theorem 4.47 *If f is nonabsolutely integrable on a compact interval $[a,b]$ then*

$$\int_a^b |f(x)|\,dx = \int_a^b [f(x)]^+\,dx = \int_a^b [f(x)]^-\,dx = \infty.$$

Proof. If f is nonabsolutely integrable then it is measurable. It follows from Exercise 207 that the functions $|f|$, $[f]^+$, and $[f]^-$ are also measurable. If, for example,

$$\int_a^b [f(x)]^+\,dx < \infty,$$

contrary to what we wish to prove, then we must conclude (from Theorem 4.43) that $[f]^+$ is integrable. But if $[f]^+$ is integrable then from the identity

$$[f(x)]^- = [f(x)]^+ - f(x)$$

we could conclude that $[f]^-$ must also be integrable and consequently each of the functions f, $|f|$, $[f]^+$, and $[f]^-$ must be integrable, contradicting the hypothesis of the theorem. ∎

4.10 Convergence of sequences of functions

Let f, f_1, f_2, ... be a sequence of functions each defined on some compact interval $[a,b] \subset \mathbb{R}$. There are various ways in which one can interpret the phrase "the sequence of functions $\{f_n\}$ converges to f." From the point of view of integration theory we would want to know the relations among these various ways as well as whether one can write

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

4.10.1 Review of elementary theory

In an elementary course one is likely to have learned only a few modes of convergence. We might, likely, be interested only in continuous functions or integrable functions (either Riemann or Newton integrable).

pointwise $\{f_n\}$ converges pointwise to f on $[a,b]$ if $\lim_{n\to\infty} f_n(x) = f(x)$ for each $x \in [a,b]$.

uniformly (unif) $\{f_n\}$ converges uniformly to f on $[a,b]$ if for all $\varepsilon > 0$ there is an integer N so that

$$|f_n(x) - f(x)| < \varepsilon$$

for all $n \geq N$ and all $x \in [a,b]$.

in mean $\{f_n\}$ converges in mean to f on $[a,b]$ if

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

where all functions are assumed to be integrable (in an appropriate sense).

The only comparisons between these three modes on a compact interval $[a,b]$ can be expressed this way:

[unif] $\Rightarrow$ [pointwise] and [unif] $\Rightarrow$ [mean].

The implications do not reverse, nor is there any general connection between [mean] and [pointwise]. Note that mean convergence does imply convergence of the integrals since, if $\{f_n\}$ converges in mean to f on $[a,b]$, then

$$\left| \int_a^b f_n(x)\, dx - \int_a^b f(x)\, dx \right| \leq \int_a^b |f_n(x) - f(x)|\, dx \to 0$$

as $n \to \infty$.

The exercises are standard in elementary courses and should be attempted or reviewed before proceeding to the more advanced modes of convergence.

Exercise 221 *Give an example to illustrate that*

[pointwise] $\not\Rightarrow$ *[unif] and [pointwise]* $\not\Rightarrow$ *[mean].*

□

Exercise 222 *Give an example to illustrate that*

[mean] $\not\Rightarrow$ *[unif] and [mean]* $\not\Rightarrow$ *[pointwise].*

□

Exercise 223 *Show that, if* $\{f_n\}$ *is a sequence of continuous functions converging uniformly to* f *on* $[a,b]$*, then* f *is also continuous. [Show that this is not true for pointwise convergence.]* □

Exercise 224 *Show that, if* $\{f_n\}$ *is a sequence of Riemann integrable functions converging uniformly to* f *on* $[a,b]$*, then* f *is also Riemann integrable and* $f_n \to f$ *[mean] on* $[a,b]$*. [Show that this is not true for pointwise convergence.]* □

Exercise 225 *Is the preceding exercise still true if "Riemann integrable" is replaced by "Newton integrable" in one of the five senses of Section 1.8?*

Answer □

‘

4.10.2 Modes of convergence

In an advanced course (with the measure theory developed in this chapter) there are number of important new modes of convergence. Usually, now, one assumes that the functions $f, f_1, f_2, \ldots$ in the sequence are each measurable and defined on some compact interval $[a,b] \subset \mathbb{R}$. Each mode of convergence represents some subtle way in which the functions f_n in the sequence can be considered to get close to the function f.

almost everywhere (a.e.) $\{f_n\}$ converges almost everywhere to f on $[a,b]$ if $\lim_{n\to\infty} f_n(x) = f(x)$ for each $x \in E$ except possibly for x in a set of Lebesgue measure zero.

almost uniformly (a.u.) $\{f_n\}$ converges almost uniformly to f on $[a,b]$ if for all $\varepsilon > 0$ there is a measurable set $A \subset [a,b]$ so that $\lambda([a,b] \setminus A) < \varepsilon$ and $\{f_n\}$ converges uniformly to f on A.

in measure (meas) $\{f_n\}$ converges in measure to f on $[a,b]$ if for all $\varepsilon > 0$

$$\lim_{n\to\infty} \lambda(\{x \in [a,b] : |f_n(x) - f(x)| \geq \varepsilon\}) = 0.$$

in mean $\{f_n\}$ converges in mean to f on $[a,b]$ if

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

Note For many applications one would want to express convergence on a given measurable set E, rather than on an interval $[a,b]$. If E is bounded then simply take an interval $[a,b]$ containing E and replace the sequence $\{f_n\}$ by the sequence $\{f_n\chi_E\}$. The theory is unchanged. If E is unbounded there are different considerations and a somewhat different theory. We leave that to a later advanced course in measure and integration.

4.10.3 Comparison of modes of convergence on $[a,b]$

The implications in Figure 4.1 display[7] all the connections among the various modes of convergence on a given interval $[a,b]$. These are all explored in the Exercises. The hardest one of these to prove is known as Egorov's theorem and is presented separately in Section 4.10.6.

The exercises also supply the necessary counterexamples to show that the missing arrows in the figure are correct. The most interesting of these counterexamples is presented separately in Section 4.10.4.

[7]These figures have been popular for many years, since appearing in M. E. Munroe, *Introduction to Measure and Integration*, Addison-Wesley (1953).

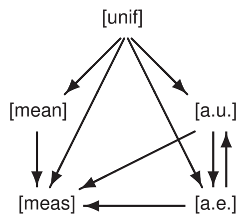

Figure 4.1: Comparison of modes of convergence on $[a,b]$

Exercise 226 *Which are the trivial or easy implications in Figure 4.1. Supply the arguments (except in the completely trivial cases).* Answer □

Exercise 227 *By examining Figure 4.1 state which of the possible implications requires a counterexample justifying the missing arrow.* Answer □

Exercise 228 *Let $\{f_n\}$ be a sequence of Lebesgue integrable functions on $[a,b]$ such that $f_n \to f$ [mean]. Show that that $f_n \to f$ [meas] on $[a,b]$.* Answer □

4.10.4 A sliding sequence of functions

Most of the required counterexamples are easy to construct. The one that might cause some difficulty is the one that will illustrate that

$$\text{[meas]} \not\Rightarrow \text{[a.e.]}$$

on an interval $[a,b]$. This is a little tricky and so we present one in this section.

For nonnegative integers n, k, with $0 \leq k < 2^n$ and $m = 2^n + k$, let

$$E_m = \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right].$$

Let $f_1 = \chi_{[0,1]}$ and, for $n > 1$, $f_m = \chi_{E_m}$. We show that this sequence supplies an example of a sequence of measurable functions on the interval $[0,1]$ that converges in measure to the zero function, and yet which is not convergent at even a single point of $[0,1]$.

We see that

$$\begin{aligned}
f_2 &= \chi_{[0,\frac{1}{2}]}, \ f_3 = \chi_{[\frac{1}{2},1]}, \\
f_4 &= \chi_{[0,\frac{1}{4}]}, \ f_5 = \chi_{[\frac{1}{4},\frac{1}{2}]}, \ f_6 = \chi_{[\frac{1}{2},\frac{3}{4}]}, \ f_7 = \chi_{[\frac{3}{4},1]}, \\
f_8 &= \chi_{[0,\frac{1}{8}]}, \ \dots
\end{aligned}$$

Every point $x \in [0,1]$ belongs to infinitely many of the sets E_m, and so

$$\limsup_m f_m(x) = 1,$$

while

$$\liminf_m f_m(x) = 0.$$

Thus $\{f_m\}$ converges at no point in $[0,1]$, yet $\lambda(E_m) = 2^{-n}$ for $m = 2^n + k$. As $m \to \infty$, $n \to \infty$ also. For every $\eta > 0$,

$$\lambda(\{x : f_m(x) \geq \eta) \leq \frac{1}{2^n}.$$

It follows that $f_m \to 0$ [meas] on the interval $[0,1]$.

Remarkably this sequence of functions does not converge at even a single point. This illustrates convergence in measure as a particularly weak form of convergence, useful in the theory precisely because it is so weak.

4.10.5 Riesz's Theorem

We have just witnessed a sequence of measurable functions converging in measure on an interval and not convergent at a single point. If we are prepared to consider subsequences, however, we can claim that convergence in measure does imply pointwise a.e. convergence.

Theorem 4.48 (Riesz) *Suppose that $\{f_n\}$ is a sequence of measurable functions on $[a,b]$ such that $f_n \to f$ [meas]. While it need not be true that $f_n \to f$ [a.e.] on $[a,b]$, there must be some subsequence $\{f_{n_k}\}$ of $\{f_n\}$ that does converge a.e. to f.*

Proof. For each integer k, choose n_k such that

$$\lambda\left(\left\{x \in [a,b] : |f_n(x) - f(x)| \geq \frac{1}{2^k}\right\}\right) < \frac{1}{2^k}$$

for every $n \geq n_k$. We choose the sequence $\{n_k\}$ to be increasing. Let

$$A_k = \left\{x \in [a,b] : |f_{n_k}(x) - f(x)| \geq \frac{1}{2^k}\right\},$$

and let $A = \limsup_k A_k$, i.e. A is the set of points that belong to infinitely many of the sets in the sequence of sets. Since

$$\sum_{k=1}^{\infty} \lambda(A_k) < 1 < \infty,$$

it follows that $\lambda(A) = 0$ by the Borel–Cantelli lemma (Exercise 202). Let $x \notin A$. Then x is a member of only finitely many of the sets A_k. Thus there exists K such that, if $k \geq K$,

$$|f_{n_k}(x) - f(x)| < \frac{1}{2^k}.$$

It follows that $f_{n_k} \to f$ [a.e.]. ∎

Exercise 229 (Fatou's lemma for convergence in measure) *Let f_n be a sequence of nonnegative, measurable functions defined at every point of an interval $[a,b]$. Suppose that $\{f_n\}$ converges in measure to a function f on $[a,b]$. Show that*

$$\int_a^b f(x)\,dx \leq \liminf_{n\to\infty} \int_a^b f_n(x)\,dx.$$

[cf. Fatou's lemma in Section 4.8.6.] Answer □

Exercise 230 *Let f_n be a sequence of measurable functions defined at every point of an interval $[a,b]$. Show that $\{f_n\}$ converges in measure to zero on $[a,b]$ if and only if*

$$\lim_{n\to\infty} \int_a^b \frac{|f_n(t)|}{1+|f_n(t)|}\,dt = 0.$$

□

4.10.6 Egorov's Theorem

We will show that

[a.e.] ⇒ [a.u.] on a compact interval $[a,b]$.

This is known as Egorov's theorem and it supplies a crucial link in our connections of the modes [a.e.] and [a.u.]. Since the proof is not completely transparent and is historically of great significance we supply all details.

Theorem 4.49 (Egorov) *Suppose that the functions f_1, f_2, ... are each measurable and defined on a compact interval $[a,b]$. Suppose that $f_n \to f$ [a.e.] on $[a,b]$. Then f is measurable and $f_n \to f$ [a.u.] on $[a,b]$.*

Proof. We already know that the a.e. limit of a sequence of measurable functions must be measurable. For every pair of integers n, k let

$$A_{nk} = \bigcap_{m=n}^{\infty} \{x \in [a,b] : |f_m(x) - f(x)| < \frac{1}{k}\}.$$

Since all functions are measurable, it follows that each of the sets A_{nk} is measurable. Let

$$E = \{x \in [a,b] : \lim_n |f_n(x) - f(x)| = 0\}.$$

Since $f_n \to f$ [a.e.], E is measurable, $\lambda(E) = \lambda([a,b])$, and for each integer k,

$$E \subset \bigcup_{n=1}^{\infty} A_{nk}.$$

For fixed k, the sequence $\{A_{nk}\}_{n=1}^{\infty}$ is expanding, so that

$$\lim_n \lambda(A_{nk}) = \lambda\left(\bigcup_{n=1}^{\infty} A_{nk}\right) \geq \lambda(E) = \lambda([a,b]).$$

Since $\lambda(E) < \infty$,

$$\lim_n \lambda(E \setminus A_{nk}) = 0. \tag{4.23}$$

Now let $\varepsilon > 0$. It follows from (4.23) that there exists n_k such that

$$\lambda(E \setminus A_{n_k k}) < \varepsilon 2^{-k}. \tag{4.24}$$

We have shown that for each $\varepsilon > 0$ there exists n_k such that inequality (4.24) holds. Let

$$A = \bigcap_{k=1}^{\infty} A_{n_k k}.$$

We now show that

$$\lambda([a,b] \setminus A) = \lambda(E \setminus A) < \varepsilon$$

and that $f_n \to f$ [unif] on A. It is clear that A is measurable. Furthermore,

$$\lambda(E \setminus A) = \lambda\left(\bigcup_{k=1}^{\infty} (E \setminus A_{n_k k})\right) \le \sum_{k=1}^{\infty} \lambda(E \setminus A_{n_k k}) < \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \varepsilon.$$

We see from the definition of the sets A_{nk} that, for $m \ge n_k$,

$$|f_m(x) - f(x)| < \frac{1}{k}$$

for every $x \in A_{n_k k}$ and therefore for every $x \in A$. Thus $f_n \to f$ [unif] on A as we wished to show. ■

The exercise shows that a version of Egorov's theorem on sets of infinite measure would not be available.

Exercise 231 *Consider the sequence of functions $f(x) = x/n$ on $(-\infty,\infty)$. Show that $\{f_n\}$ converges pointwise on $(-\infty,\infty)$ but that there can be no "small" set N so that $\{f_n\}$ converges uniformly on $(-\infty,\infty) \setminus N$.* Answer □

4.10.7 Dominated convergence on an interval

Let us return to the notion of dominated convergence that we introduced in Section 4.8.7. This allows us to complete the picture illustrated in Figure 4.1 connecting the various modes of convergence. We assume that the functions f, f_1, f_2, ... in the sequence are each measurable and defined on an interval $[a,b]$ with the very special property that there is a nonnegative Lebesgue integrable function $g : [a,b] \to \mathbb{R}$ that *dominates* the entire sequence. That is we assume that

$$\int_a^b g(x)\,dx < \infty$$

and that

$$|f_n(x)| \le g(x) \text{ for all } x \in [a,b].$$

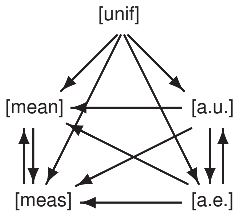

Figure 4.2: Comparison of modes of dominated convergence.

Figure 4.2 shows that with this extra dominated assumption there are some further connections among the various modes of convergence. In the exercises we are asked to supply the extra proofs now needed. Since we have already addressed such concerns in Theorem 4.35 the reader can supply some of these proofs without further difficulty.

Bounded convergence Dominated convergence is often replaced by a simpler assumption—bounded convergence. If

$$|f_n(x)| \leq M \text{ for all } x \in [a,b].$$

for some positive number M then the function $g(x) = M$ dominates the functions in the sequence f_1, f_2, . . . as we require.

Exercise 232 *What are the implications that must be proved to justify the connections in Figure 4.2 that did not appear in Figure 4.1.*

Answer □

Exercise 233 *Show that*

$$\text{[meas]} \Rightarrow \text{ [mean]}$$

on an interval $[a,b]$ assuming that the sequence is dominated. Answer □

4.11 Lusin's theorem

We have defined measurable functions by first defining measurable sets and then requiring a function f to be measurable provided that all the associated sets of the form $\{x : f(x) > r\}$ are measurable.

There is another useful way to view measurable functions and can equally well be used as a starting point: they are almost continuous. This important theorem was discovered independently by Guiseppe Vitali (1875–1932) and Nikolai Lusin (1883–1950). It is almost universally called Lusin's theorem. (Lusin, often transliterated as Luzin, was a student of Egorov, who is known mainly for the theorem on almost uniform convergence that we have just proved in Section 4.10.6.)

Theorem 4.50 *Let $f : [a,b] \to \mathbb{R}$ be a measurable function. Then to each pair (ε, η) of positive numbers corresponds a continuous function $g : [a,b] \to \mathbb{R}$ such that*

$$\lambda(\{x \in [a,b] : |f(x) - g(x)| \geq \eta\}) < \varepsilon.$$

Proof. Let us suppose first that f is bounded. By Exercise 4.24 there exists a simple function s, such that

$$|s(x) - f(x)| < \eta \quad (x \in [a,b]).$$

Let $c_1, \dots, c_m$ be the values that s assumes on $[a,b]$ and, for each $i = 1, \dots, m$, let

$$E_i = \{x \in [a,b] : s(x) = c_i\}.$$

The sets E_i are measurable, pairwise disjoint, and cover $[a,b]$. Choose closed sets $F_1, \dots, F_m$ such that, for each $i = 1, \dots, m$, $F_i \subset E_i$ and

$$\lambda(E_i \setminus F_i) < \frac{\varepsilon}{m}.$$

Let

$$F = F_1 \cup \cdots \cup F_m.$$

Then F is a closed subset of $[a,b]$ for which $\lambda([a,b] \setminus F) < \varepsilon$.

We may then construct a continuous function g on $[a,b]$ that agrees with s just at points of the closed set F. The simplest way to do this is just to take any component (c,d) of the open set $G = (a,b) \setminus F$ and define g to be linear on $[c,d]$. Since

$$\lambda([a,b] \setminus F) < \varepsilon,$$

g is the desired function.

Finally, to remove the assumption that f is bounded, simply replace f by a bounded measurable function that differs from it on a set of sufficiently small measure. (See Exercise 4.19.) ∎

From this theorem we easily deduce the following theorem showing that a measurable function must be the [a.u.] limit of a sequence of continuous functions. Use Theorem 4.50 to select a sequence of continuous $g_k \to f$ [meas] on $[a,b]$. Use Exercise 4.48 to pass to a subsequence $g_{k_m} \to f$ [a.e.] on $[a,b]$. Then, by Egorov's theorem, this subsequence converges [a.u.] to f on $[a,b]$. This proves the theorem:

Theorem 4.51 *Let $f : [a,b] \to \mathbb{R}$ be a measurable function. Then there exists a sequence of continuous functions $\{g_k\}$ on $[a,b]$ such that $g_k \to f$ [a.u.] on $[a,b]$.*

Finally we can obtain the usual form of Lusin's theorem.

Theorem 4.52 (Lusin) *Let $f : [a,b] \to \mathbb{R}$ be a measurable function and let $\varepsilon > 0$. Then there exists a continuous function $g : [a,b] \to \mathbb{R}$ such that $f(x) = g(x)$ for all x in a closed set $F \subset [a,b]$ for which $\lambda([a,b] \setminus F) < \varepsilon$.*

Proof. As we have observed in Theorem 4.51, there exists a sequence of continuous functions $\{g_k\}$ on $[a,b]$ such that $g_k \to f$ [a.u.] on $[a,b]$. This means that there is a measurable set E with $\lambda([a,b]\setminus E) < \varepsilon/2$ so that $g_k \to f$ [unif] on E. Choose C closed so that $C \subset E$ with $\lambda(E\setminus C) < \varepsilon/2$. Then $\lambda([a,b]\setminus C) < \varepsilon$.

Because of the uniform convergence, f is continuous on the closed set C (i.e. continuous restricted to that set, not necessarily continuous at each point of that set). Simply now extend f from C to a continuous function $g : [a,b] \to \mathbb{R}$ that agrees with f on C. ∎

4.11.1 Littlewood's three principles

Lusin's theorem, along with Egorov's theorem, is considered part of the fundamentals of real analysis. Littlewood famously expressed it this way:

> "The extent of knowledge required is nothing like so great as is sometimes supposed. There are three principles, roughly expressible in the following terms:
>
> - Every set is nearly a finite union of intervals;
> - Every function is nearly continuous;
> - Every convergent sequence of functions is nearly uniformly convergent.
>
> Most of the results . . . are fairly intuitive applications of these ideas, and the student armed with them should be equal to most occasions when real variable theory is called for. If one of the principles would be the obvious means to settle the problem it it were 'quite' true, it is natural to ask if the 'nearly' is near enough, and for a problem that is actually solvable it generally is."
>
> ——from J. E. Littlewood, *Lectures on the Theory of Functions*, Oxford University Press (1944), p. 26.

One finds usually that any set one encounters can be proved to be measurable; if so, then measurable sets can be approximated by unions of intervals. Similarly one finds that any function one encounters can be proved to be measurable; if so, then measurable functions can be approximated by continuous functions. Finally, most modes of convergence of a sequence of measurable functions can be approximated by uniform convergence.

4.11.2 Denjoy-Stepanoff theorem

Lusin's theorem asserts a close connection between the measurability of a function and a kind of weak continuity condition. It is closely related to a similar theorem wherein measurable functions are characterized as those that are approximately continuous almost everywhere. The definition of approximate continuity uses the notion of density introduced in Section 4.4.

Definition 4.53 (approximate continuity) *Let f be a function defined in an open set containing a point x_0. If there exists a set E such that*

$$\underline{d}(E,x_0)=1 \text{ and } \lim_{x\to x_0, x\in E} f(x)=f(x_0),$$

we say that f is approximately continuous *at x_0. If f is approximately continuous at all points of its domain, we simply say that f is* approximately continuous.

If a function is defined on a closed interval $[a,b]$, then approximate continuity at the end points is defined in the obvious way, invoking one-sided densities. Note that f is approximately continuous at x_0 if there exists a set E having x_0 as a density point, such that the function f restricted to E is continuous at x_0. For example, if $A\subset\mathbb{R}$ is measurable, then the characteristic function χ_A is approximately continuous at every point that is either a point of density of A or else a point of density of $\mathbb{R}\setminus A$. In particular almost every point is a point of approximate continuity of any function $f(x)=\chi_A(x)$ provided A is a measurable set. This suggests that a similar thing might be true for all measurable functions.

Indeed, we show that measurability of functions can be characterized using the concept of approximate continuity. A. Denjoy [23] introduced the concept and showed that measurable functions have this property. Stepanoff [78] proved the more difficult converse.

Theorem 4.54 (Denjoy-Stepanoff) *A function is measurable if and only if it is approximately continuous at almost every point.*

Proof. We prove this only in the one direction (the easy direction). We suppose that f is measurable. Let $\varepsilon>0$. By Lusin's theorem there exists a continuous function g such that

$$\lambda(\{x: g(x)\neq f(x)\})<\varepsilon. \tag{4.25}$$

Let $E=\{g(x)=f(x)\}$. By Theorem 4.6, almost every point of E is a density point of E. If $x_0\in E$ and x_0 is a density point of E, we have

$$\lim_{x\to x_0,\ x\in E} f(x)=\lim_{x\to x_0} g(x)=g(x_0)=f(x_0).$$

Thus f is approximately continuous at x_0. Since x_0 is an arbitrary density point of E, f is approximately continuous at each density point of E. From (4.25), we infer that f is approximately continuous except perhaps on a set of measure less than ε. Since ε is arbitrary, f is approximately continuous a.e.

For the opposite direction we supply some interesting references. In addition to the original paper of Stepanoff cited above, the more adventurous reader might want to track down the simpler proof of Kamke [42][8]. An advanced reader

[8]E. Kamke, Zur Definition der approximate stetigen Funktionen, Fund. Math. I0 (1927), 431–433.

should certainly consult the article of Lukeš [55][9] that demonstrates the theorem as an application of purely topological methods. ■

Exercise 234 *Show that a function f is approximately continuous at a point x_0 if and only if*

$$\lim_{h\to 0+} \frac{\lambda(\{x \in (x_0-h, x_0+h) : |f(x)-f(x_0)| \geq \varepsilon\})}{h} = 0$$

for every $\varepsilon > 0$. □

Exercise 235 *Show that a function that is approximately continuous at every point must be in the first Baire class, i.e., it is the pointwise limit of a sequence of continuous functions.* Answer □

Exercise 236 *Show that a function that is approximately continuous at every point must have the Darboux property, i.e., it has the intermediate value property (familiar to students of the calculus since all continuous functions have this property).* Answer □

4.12 Absolute continuity of the integral

We shall need a type of absolute continuity property of integrals. This is closely related to the various other concepts that we have previously introduced using this phrase.

Theorem 4.55 *Suppose that $f : [a,b] \to \mathbb{R}$ is a Lebesgue integrable function on $[a,b]$. Then for every $\varepsilon > 0$ there is a $\delta > 0$ so that if G is an open set for which $\lambda(G) < \delta$ then*

$$\int_a^b \chi_G(x)|f(x)|\,dx < \varepsilon.$$

Proof. For illustrative purposes only we begin the proof with the bounded case. Suppose that $|f(x)| < N$ for all $x \in [a,b]$. Choose $\delta = \varepsilon/N$ and observe that, if $\lambda(G) < \delta$ then simple inequalities provide

$$\int_a^b \chi_G(x)|f(x)|\,dx \leq N \int_a^b \chi_G(x)\,dx \leq N\lambda(G) < \varepsilon.$$

The argument in the bounded case suggests how to proceed. For each positive integer n let

$$A_n = \{x \in [a,b] : n-1 \leq |f(x)| < n\}.$$

From the fact that f is measurable we can deduce that each A_n is measurable. Note that the sets in the sequence are pairwise disjoint and that their union is

[9] Jaroslav Lukeš, A topological proof of Denjoy-Stepanoff theorem. Časopis Pěst. Mat. 103 (1978), no. 1, 95–96, 98.

all of $[a,b]$. We can select an open set G_n for which $B_n = A_n \setminus G_n$ is closed and $\lambda(G_n) < \varepsilon 2^{-n} n^{-1}$. Using all this we compute

$$\int_a^b \chi_{A_n}(x)|f(x)|\,dx = \int_a^b \chi_{B_n}(x) f(x)\,dx + \int_a^b \chi_{A_n \cap G_n}(x) f(x)\,dx$$
$$\leq \int_a^b \chi_{B_n}(x)|f(x)|\,dx + n[\varepsilon 2^{-n} n^{-1}].$$

Note that $\{B_n\}$ is a disjointed sequence of closed subsets of $[a,b]$. Define B as the union of the sequence. Then

$$\sum_{n=1}^{\infty} \int_a^b \chi_{A_n}(x)|f(x)|\,dx \leq \sum_{n=1}^{\infty} \int_a^b \chi_{B_n}(x)|f(x)|\,dx + \varepsilon$$
$$= \int_a^b \chi_B(x)|f(x)|\,dx + \varepsilon \leq \int_a^b |f(x)|\,dx + \varepsilon < \infty.$$

In particular there must be an integer N sufficiently large that

$$\sum_{n=N+1}^{\infty} \int_a^b \chi_{A_n}(x)|f(x)|\,dx < \varepsilon/2.$$

Choose $\delta = \varepsilon/(2N)$ and let G be any open set for which $\lambda(G) < \delta$. Since

$$G = \{x \in G : |f(x)| < N\} \cup \bigcup_{n=N+1}^{\infty} (G \cap A_n)$$

we have

$$\int_a^b \chi_G(x)|f(x)|\,dx \leq N\lambda(G) + \sum_{n=N+1}^{\infty} \int_a^b \chi_{A_n}(x)|f(x)|\,dx < \varepsilon.$$

■

Exercise 237 *Theorem 4.55 requires the assumption that the function f is Lebesgue integrable. Suppose that $f : [a,b] \to \mathbb{R}$ is a nonabsolutely integrable function on $[a,b]$ (i.e., f is integrable, but not Lebesgue integrable). Show that for every $\delta > 0$ there must exist a disjoint collection*

$$[a_1,b_1], [a_2,b_2], [a_3,b_3], \ldots, [a_n,b_n]$$

of subintervals of $[a,b]$ for which $\sum_{i=1}^n (b_i - a_i) < \delta$ and yet

$$\sum_{i=1}^n \left| \int_{a_i}^{b_i} f(x)\,dx \right| > 1.$$

Answer □

4.13 Convergence and equi-integrability

One of our central concerns, and indeed one of the central concerns of the early history of integration theory, has been the convergence of integrals. Suppose that $f_1, f_2, f_3, \ldots$ is a sequence of Lebesgue integrable functions on an interval $[a,b]$ and that

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

Then under what extra conditions can we conclude that f is also Lebesgue integrable on $[a,b]$ and that

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx? \tag{4.26}$$

Some of the answers to this question are given by the following conditions.

(uniform convergence) $\lim_{n\to\infty} f_n(x) = f(x)$ uniformly on $[a,b]$.

(bounded convergence) for some positive number M, $|f_n(x)| \le M$ for all integers n and a.e. points $x \in [a,b]$.

(dominated convergence) for some Lebesgue integrable function g and for all integers n, the inequality $|f_n(x)| \le g(x)$ holds for a.e. point $x \in [a,b]$.

These conditions, as we know, are sufficient but by no means necessary in order to conclude the validity of the limit statement (4.26). We have seen this material before in Section 4.10.7.

4.13.1 Equi-integrability

In this section and in Section 4.13.3 we shall add two new conditions that also suffice and are closer to being necessary. We have already, in Section 3.3, investigated a kind of equi-integrability arising naturally in the setting of the Henstock-Kurzweil integral. For the Lebesgue integral stronger conditions are needed.

The Vitali condition below is well-known and often cited[10]. Some authors refer to the property as "uniform absolute continuity" which, in view of the language of Section 4.12, better captures the property than the term "equi-integrability." The second condition we study is a generalization of McShane's characterization of Lebesgue integrability.

(Vitali equi-integrability) for all $\varepsilon > 0$ there is a $\delta > 0$ so that if G is an open set for which $\lambda(G) < \delta$ then

$$\int_a^b |f_n(x)|\chi_G(x)\,dx < \varepsilon$$

for all integers n.

[10]The original is in G. Vitali, Sul integrazione per serie, Rend. Circ. Mat. Palermo 23 (1907), 137–155.

Exercises

Exercise 238 *Show that [uniform convergence] does not imply [bounded convergence], but does imply [dominated convergence].*

Answer □

Exercise 239 *Show that [bounded convergence] implies [dominated convergence].*

Answer □

Exercise 240 *Show that [dominated convergence] implies the [Vitali equi-integrability] condition.*

Answer □

Exercise 241 *Give an example of a sequence of nonnegative, continuous functions $f_n : [0,1] \to \mathbb{R}$ converging pointwise to zero for which*

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = 0$$

and yet not satisfying the [dominated convergence] condition. Answer □

Exercise 242 *Suppose that $g_1,\ g_2,\ g_3,\ \dots$ is a sequence of nonnegative Lebesgue integrable functions on an interval $[a,b]$ and suppose that $\lim_{n\to\infty} f_n(x) = 0$ for a.e. $x \in [a,b]$. Show that*

$$\lim_{n\to\infty} \int_a^b g_n(x)\,dx = 0$$

if and only if the sequence $\{g_n\}$ satisfies the Vitali equi-integrability condition.

Answer □

4.13.2 A stronger convergence theorem

Since this additional condition is weaker than the dominated convergence condition, we can obtain a stronger convergence theorem by using it.

Theorem 4.56 (Vitali) *Let $f_1,\ f_2,\ f_3,\ \dots$ be a sequence of Lebesgue integrable functions on an interval $[a,b]$ and suppose that*

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

If the sequence $\{f_n\}$ satisfies the Vitali equi-integrability condition, then f is Lebesgue integrable on $[a,b]$ and

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx. \tag{4.27}$$

Proof. Let us assume that $f_1, f_2, f_3, \dots$ is a sequence of Lebesgue integrable functions on $[a,b]$ and that

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

The function f is certainly measurable. We show that it is also Lebesgue integrable.

If the sequence $\{f_n\}$ satisfies the Vitali equi-integrability condition, then we can select a positive number $\eta > 0$ so that if G is an open set for which $\lambda(G) < \eta$ then, for each integer i,

$$\int_a^b |f_i(x)|\chi_G(x)\,dx < 1.$$

Cover (a,b) with a finite number of open intervals, each of length less than η, say

$$(a,b) \subset (a_1,b_1) \cup (a_2,b_2) \cup \cdots \cup (a_m,b_m).$$

Then

$$\int_a^b |f_i(x)|\,dx \le \sum_{i=1}^m \int_{a_i}^{b_i} |f_i(x)|\,dx < m.$$

This uniform bound and Fatou's lemma shows that

$$\int_a^b |f(x)|\,dx \le m$$

and, hence, that f is Lebesgue integrable on $[a,b]$.

Let $\varepsilon > 0$. Select a positive number $\delta > 0$ so that if G is an open set for which $\lambda(G) < \delta$ then, for each integer i,

$$\int_a^b |f_i(x)|\chi_G(x)\,dx < \varepsilon/3.$$

Fatou's lemma (yet again) shows that

$$\int_a^b |f(x)|\chi_G(x)\,dx \le \varepsilon/3$$

would also be true. Now apply Egorov's theorem to select an open set G for which $\lambda(G) < \delta$ and so that $f_n \to f$ uniformly on the closed set $K = [a,b] \setminus G$. We compute

$$\int_a^b |f_n(x) - f(x)|\,dx \le \int_a^b \chi_K(x)|f_n(x) - f(x)|\,dx$$

$$+ \int_a^b \chi_G(x)|f_n(x)|\,dx + \int_a^b \chi_G(x)|f(x)|\,dx.$$

$$\le \int_a^b \chi_K(x)|f_n(x) - f(x)|\,dx + 2\varepsilon/3.$$

The uniform convergence of the sequence $\{\chi_K(x)|f_n(x) - f(x)|\}$ on $[a,b]$

shows that, for large enough n,

$$\int_a^b |f_n(x) - f(x)|\, dx < \varepsilon.$$

Finally then, the limit statement (4.27) now follows. ∎

4.13.3 McShane equi-integrability condition

Definition 4.45 expresses a necessary and sufficient condition for a function to be Lebesgue integrable on an interval. This is easily and naturally converted into an equi-integrability condition (although McShane himself did not). Kurzweil and Schwabik[11] have shown that this condition is essentially equivalent to the Vitali equi-integrability condition. The latter is expressed exclusively in measure-theoretic terms, while the McShane condition is stated directly in terms of Riemann sums.

As before we suppose that $f_1,\ f_2,\ f_3,\ \dots$ is a sequence of Lebesgue integrable functions on an interval $[a,b]$ and that

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

(McShane equi-integrability) for all $\varepsilon > 0$ a full cover β can be found so that

$$\sum_{(I,w)\in\pi}\ \sum_{(I',w')\in\pi'} \left|f_n(w) - f_n(w')\right| \lambda(I \cap I') < \varepsilon$$

for all integers n and all subpartitions π, π' of $[a,b]$ contained in β.

Our second convergence theorem is similar to the Vitali theorem just proved.

Theorem 4.57 *Let $f_1,\ f_2,\ f_3,\ \dots$ be a sequence of Lebesgue integrable functions on an interval $[a,b]$ and suppose that*

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

If the sequence $\{f_n\}$ satisfies the McShane equi-integrability condition, then f is Lebesgue integrable on $[a,b]$ and

$$\lim_{n\to\infty} \int_a^b f_n(x)\, dx = \int_a^b f(x)\, dx. \tag{4.28}$$

Proof. We can suppose that

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

(Changing "a.e." to "every" is easily done.)

[11] In J. Kurzweil and S. Schwabik, McShane equi-integrability and Vitali's convergence theorem. Math. Bohem. 129 (2004), no. 2, 141–157.

Let $\varepsilon > 0$. If the sequence $\{f_n\}$ satisfies the uniform McShane condition, then a full cover β can be found so that

$$\sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} |f_n(w) - f_n(w')| \lambda(I\cap I') < \varepsilon/2 \tag{4.29}$$

for all integers n and all subpartitions π, π' of $[a,b]$ contained in β. We would also necessarily have from this inequality that

$$\left| \int_a^b f_n(x)\,dx - \sum_{(I,w)\in\pi} f_n(w)\lambda(I) \right| < \varepsilon \tag{4.30}$$

for all integers n and every partition π of $[a,b]$ contained in β.

Let $n \to \infty$ in (4.29) to obtain that

$$\sum_{(I,w)\in\pi} \sum_{(I',w')\in\pi'} |f(w) - f(w')| \lambda(I\cap I') \leq \varepsilon/2 \tag{4.31}$$

for all subpartitions π, π' of $[a,b]$ contained in β. This means, the McShane criterion, that f is Lebesgue integrable on $[a,b]$. Finally then, the limit statement (4.27) now follows.

Finally letting $n \to \infty$ in (4.30) we can deduce the limit statement (4.28). ■

4.14 Young-Daniell-Riesz Program

Our main interest in this chapter is the constructive method of Lebesgue that provides a way to construct the value of the integral for all absolutely integrable functions. This requires developing the measure theory that is the primary tool of the subject. This is the method used by Lebesgue in his famous presentation of his integral.

At the same time, the British mathematician W. H. Young[12] [91] initiated a theory of the same integral by employing the method of monotone sequences. These ideas were elaborated by Daniell[13] [20] who produced a general method of monotone sequences that has had a considerable impact on later generations of analysts. F. Riesz showed that this method could be used to present an elegant constructive account of the Lebesgue integral without first invoking measure theory. We know, of course, that the Newton and Henstock-Kurzweil methods do this as well, but in a non-constructive manner.

Let us begin by returning to the theory of the regulated integral. We recall that the starting point is the integral of step functions, extended by employing uniform limits. The new theory starts also with the integral of step functions, but employs monotone limits.

[12] William Henry Young, Proc. London Math. Soc. (2) 2 (1904), 52–66.

[13] Percy John Daniell, Ann. of Math. (2) 19 (1917/18), 279–294.

4.14.1 Recall the program for the regulated integral

One starts the program with the integral for step functions. (See Section 1.9.1.)

Definition 4.58 (step function) *A function $f : [a,b] \to \mathbb{R}$ is said to be a step function if there are points*

$$a = x_0 < x_1 < x_2 < \ldots, x_n = b$$

such that f is constant on each open interval (x_{i-1}, x_i).

Step functions are easily shown to be integrable (even integrable in the elementary Newton sense) and the value of the integral constructed as a simple sum

$$\int_a^b f(x)\,dx = \sum_{i=1}^n c_i(x_{i-1} - x_i)$$

where f assumes the value c_i on the interval (x_{i-1}, x_i).

The extension step that we studied in Chapter 1 is to define a function $f : [a,b] \to \mathbb{R}$ as regulated provided it is a uniform limit on $[a,b]$ of a sequence of step functions. Consequently any such regulated function is integrable and the value for its integral is constructed by using

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

where $\{f_n\}$ is any sequence of step functions converging uniformly to f.

Our point of view [since we start with a full integration theory] is that this analysis offers a constructive method for determining the value of an integral whose value has been defined nonconstructively [either by an antiderivative or as a rather mysterious limit of Riemann sums]. A different point of view could be taken that this program offers an alternative definition of an integral whose properties would have to be determined from that definition. The regulated program, indeed, is used in some textbooks and courses of instruction, but seems to be less popular than other methods.

Exercise 243 *Verify the identity*

$$\int_a^b f(x)\,dx = \sum_{i=1}^n c_i(x_{i-1} - x_i)$$

where f assumes the value c_i on each interval (x_{i-1}, x_i), by giving a construction for a continuous primitive function $F : [a,b] \to \mathbb{R}$ for which $F' = f$ except at a finite number of points. Answer □

4.14.2 Riesz sequences

The Young-Daniell-Riesz program we present is nearly identical in basic outline, but takes a technically more challenging approach in the limit step: in place of uniform limits, which are mostly easy to handle, it uses monotone limits, requiring rather more technical and subtle arguments. The payoff, however, is large. This program gives a method for constructing the value of the integral for all absolutely integrable [i.e., Lebesgue integrable] functions.

Semicontinuous functions

Definition 4.59 (l.s.c. function) *A function $f : [a,b] \to \mathbb{R} \cup \infty$ is said to be lower semicontinuous [l.s.c.] if for every point $x \in [a,b]$ and every $t < f(x)$ there is a $\delta > 0$ so that*

$$f(y) > t \text{ for all } y \in [a,b] \cap (x - \delta, x + \delta).$$

Similarly $f : [a,b] \to \mathbb{R} \cup -\infty$ is upper semicontinuous [u.s.c.] if for every point $x \in [a,b]$ and every $t > f(x)$ there is a $\delta > 0$ so that

$$t > f(y) \text{ for all } y \in [a,b] \cap (x - \delta, x + \delta).$$

A function is continuous if and only if it is both u.s.c. and l.s.c. The characteristic function of an open interval is l.s.c. while the characteristic function of a closed interval is u.s.c.

Riesz sequences of step functions

Definition 4.60 (Riesz sequence) *A sequence of functions $f_n : [a,b] \to \mathbb{R}$ for $n = 1, 2, 3, \ldots$ is said to be a Riesz sequence if*

1. *each f_n is a lower semicontinuous step function.*
2. *the sequence is monotone, i.e., for each x and each n, $f_n(x) \le f_{n+1}(x)$.*
3. *the integrals are bounded, i.e.,*

$$\sup_n \int_a^b f_n(x)\, dx < \infty.$$

At the same time we will consider also monotone decreasing Riesz sequences: *decreasing* monotone limits and upper semicontinuous step functions, i.e., for each x and each n, $f_n(x) \ge f_{n+1}(x)$. and we would assume that the integrals are bounded below,

$$\inf_n \int_a^b f_n(x)\, dx > -\infty.$$

4.14.3 Convergence of Riesz sequences of step functions

Our first theorem shows that every Riesz sequence converges to a function that is integrable. Indeed, the value of the integral of such a limit function is directly constructible from the integrals of the members in the sequence.

Theorem 4.61 *Suppose that $f_n : [a,b] \to \mathbb{R}$ for $n = 1,2,3,\dots$ is a Riesz sequence. Then*

$$\lim_{n\to\infty} f_n(x) = f(x) \tag{4.32}$$

exists (finitely or infinitely) for all x and is finite for a.e. x. The function f is lower semicontinuous, bounded below, finite almost everywhere, and absolutely integrable. Moreover

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b f_n(x)\,dx. \tag{4.33}$$

Proof. Since the sequence is monotone it is clear that

$$\lim_{n\to\infty} f_n(x) = f(x)$$

exists as a finite number or as $+\infty$ for all x. A monotone sequence of l.s.c. functions converges to another l.s.c. function. All l.c.s. functions are bounded below (although it is also obvious that $f \geq f_1$ gives a lower bound too). The statement (4.33) follows immediately from the monotone convergence theorem. Because of the bounds on the integrals of the functions f_n in a Riesz sequence, the inequality

$$-\infty < \int_a^b f_1(x)\,dx \leq \int_a^b f(x)\,dx < \infty$$

shows that f must be a.e. finite. ∎

4.14.4 Representing semicontinuous functions by Riesz sequences

The converse direction is essential to us. We wish to know how far this program takes us. We have a constructive method using Riesz sequences for determine the value of the integral for a large class of functions, certainly larger now that the class of regulated functions. How large?

Theorem 4.62 *Suppose that the function $f : [a,b] \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous and absolutely integrable. Then there exists a Riesz sequence. $g_n : [a,b] \to \mathbb{R}$ for $n = 1,2,3,\dots$ so that*

$$\lim_{n\to\infty} g_n(x) = f(x)$$

for a.e. x in $[a,b]$. Moreover, for any such representation of f as a limit of a Riesz sequence,

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

Proof. The set of points $N = \{x \in [a,b] : f(x) = \infty\}$ is necessarily of measure zero since we are assuming that f is integrable. Define $f_1(x) = f(x)$ for $x \in$

$[a,b]\setminus N$ and $f_1(x)=0$ for $x\in N$. Since f is integrable, so too is f_1 and for any integer $n=1,2,3,\dots$, and there must exist a full cover β_1 of $[a,b]$ for which

$$\left|\int_a^b f_1(x)\,dx - \sum_{([u,v],w)\in\pi} f_1(w)(v-u)\right| < 1/n.$$

for any partition $\pi\subset\beta_1$ of $[a,b]$.

Define

$$\beta_2 = \{([u,v],w) : w\in N \text{ and } f(t)>n \text{ for all } t\in[u,v]\ \}.$$

This is a full cover of N since we are assuming that f is l.s.c. on $[a,b]$ and $f(w)=\infty$ for each $w\in N$.

There is a full cover β_3 of the measure zero set N so that

$$\sum_{([u,v],w)\in\pi} (v-u) < 1/n^2.$$

for any subpartition $\pi\subset\beta_2$ of $[a,b]$.

The collection

$$\beta_3 = \{([u,v],w) : w\in[a,b]\setminus N \text{ and } f(t)>f(w)-1/n \text{ for all } t\in[u,v]\ \}$$

is a full cover of $[a,b]\setminus N$ since we are assuming that f is l.s.c. on $[a,b]$.

Finally then

$$\beta = \beta_1\cap([\beta_2\cap\beta_3]\cup\beta_3)$$

is evidently a full cover of $[a,b]$. Take any partition $\pi\subset\beta$ of $[a,b]$.

We know that

$$\left|\int_a^b f_1(x)\,dx - \sum_{([u,v],w)\in\pi} f_1(w)(v-u)\right| < 1/n.$$

where $f_1(w)=0$ and $f(w)=\infty$ if $w\in N$.

Define a step function f_n on $[a,b]$ this way: if $([u,v],w)\in\pi$ and $w\in N$ then set $f_n(t)=n$ for all $t\in(u,v)$. If $([u,v],w)\in\pi$ and $w\in[a,b]\setminus N$ then set $f_n(t)=f(w)-1/n$ for all $t\in(u,v)$. Note that f_n is continuous at each point t at which we have defined it and that $f_n(t)\le f(t)$ at all such points. There are still the finitely many points at the endpoints of the intervals; at these points we simply arrange that f_n is lower semi-continuous at each such point t and that $f_n(t)\le f(t)$.

We claim now to have defined a l.s.c. step function f_n with $f_n\le f$ and such that

$$\int_a^b f_n(t)\,dt \ge \int_a^b f(t)\,dt - (b-a+2)/n. \tag{4.34}$$

To check (4.34), take any element $([u,v],w)\in\pi$. Now $f_1(w)(v-u)=0$ for any $w\in N$ and

$$\int_u^v f_n(t)\,dt = n(v-u).$$

If instead $w \notin N$ then

$$\int_u^v f_n(t)\,dt = [f(w) - 1/n](v-u).$$

Putting these together give

$$\begin{aligned}\int_a^b f_n(t)\,dt &= \sum_{([u,v],w)\in\pi} [f_1(w) - 1/n](v-u) + \sum_{([u,v],w)\in\pi[N]} n(v-u) \\ &\geq \sum_{([u,v],w)\in\pi} f_1(w)v-u) - (b-a)/n - n/n^2 \\ &\geq \int_a^b f(t)\,dt - 1/n - (b-a)/n - n/n^2\end{aligned}$$

proving (4.34).

This is nearly the sequence of l.s.c. step functions that we want, however it is not monotone. Take

$$g_n = \max\{f_1, f_2, f_3, \dots, f_n\}.$$

This is also a sequence of l.s.c. step functions, but now it is a monotone sequence. Then it remains just to check the details so that we can be sure that

$$\lim_{n\to\infty} g_n(x) = f(x)$$

for a.e. x in $[a,b]$ and

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

■

4.14.5 Characterization of the Lebesgue integral

Some notation will help us express our ideas a bit better. Let $\mathcal{S}[a,b]$ denote some set of integrable functions on the interval $[a,b]$. We then write $\mathcal{S}^{\uparrow}[a,b]$ for the set of all functions f integrable on $[a,b]$ for which there is a sequence of functions $\{f_n\}$ for which

1. each f_n is in $\mathcal{S}[a,b]$.
2. the sequence is monotone nondecreasing, i.e., for each x and each n, $f_n(x) \leq f_{n+1}(x)$.
3. the integrals are bounded, i.e.,
$$\sup_n \int_a^b f_n(x)\,dx < \infty.$$
4. $f(x) = \lim_{n\to\infty} f_n(x)$ a.e. in $[a,b]$.

A similar definition would apply to define $\mathcal{S}^{\downarrow}[a,b]$ but using monotone decreasing sequences of functions from $\mathcal{S}[a,b]$. We write $\mathcal{L}[a,b]$ for the class of all Lebesgue integrable functions on $[a,b]$. Because of the monotone convergence theorem we know already that

$$\mathcal{L}[a,b] = \mathcal{L}^{\uparrow}[a,b] = \mathcal{L}^{\downarrow}[a,b].$$

A sum and difference notation is useful as well. If $\mathcal{S}_1[a,b]$ and $\mathcal{S}_2[a,b]$ are sets of integrable functions on the interval $[a,b]$, then

$$\mathcal{S}_1[a,b] + \mathcal{S}_2[a,b]$$

would denote the set of all functions f such that $f(x) = f_1(x) + f_2(x)$ for a.e. $x \in [a,b]$ for some choice of $f_1 \in \mathcal{S}_1[a,b]$ and $f_2 \in \mathcal{S}_2[a,b]$. Similarly

$$\mathcal{S}_1[a,b] - \mathcal{S}_2[a,b]$$

would denote the set of all functions f such that $f(x) = f_1(x) - f_2(x)$ for a.e. $x \in [a,b]$.

Theorem 4.63 *Let $\mathcal{S}_\ell[a,b]$ denote the class of all l.s.c. step functions on the interval $[a,b]$ and let $\mathcal{S}_u[a,b]$ denote the class of all u.s.c. step functions on $[a,b]$. Then*

$$\mathcal{L}[a,b] = \mathcal{S}_\ell^{\uparrow\downarrow}[a,b] = \mathcal{S}_u^{\downarrow\uparrow}[a,b]$$
$$= \mathcal{S}_\ell^{\uparrow}[a,b] + \mathcal{S}_u^{\downarrow}[a,b] = \mathcal{S}_\ell^{\uparrow}[a,b] - \mathcal{S}_\ell^{\uparrow}[a,b] = \mathcal{S}_u^{\downarrow}[a,b] - \mathcal{S}_u^{\downarrow}[a,b].$$

Proof. We know, because of Theorems 4.61 and 4.62, exactly what functions belong to $\mathcal{S}_\ell^{\uparrow}[a,b]$ and $\mathcal{S}_u^{\downarrow}[a,b]$. In particular if g is equal a.e. to an integrable l.s.c. function on $[a,b]$ then g belongs to $\mathcal{S}_\ell^{\uparrow}[a,b]$.

It is evident that $\mathcal{L}[a,b] \supset \mathcal{S}_\ell^{\uparrow}[a,b]$. We know from the monotone convergence theorem that $\mathcal{L}[a,b] = \mathcal{L}^{\downarrow}[a,b]$ and hence that $\mathcal{L}[a,b] \supset \mathcal{S}_\ell^{\uparrow\downarrow}[a,b]$. The same inclusion is true for $\mathcal{S}_u^{\downarrow\uparrow}[a,b]$. We prove the opposite direction.

Suppose first that f is a nonnegative function in $\mathcal{L}[a,b]$. Let $\varepsilon > 0$. By Theorem 4.26, we can express f as the pointwise sum of a series of simple functions

$$f(x) = \sum_{k=1}^{\infty} c_k \chi_{E_k}(x)$$

where $\{E_k\}$ is a sequence of measurable subsets of $[a,b]$ (not necessarily disjoint) and c_k are positive numbers. Note that

$$\int_a^b f(x)\,dx = \sum_{k=1}^{\infty} c_k \left(\int_a^b \chi_{E_k}(x)\,dx\right) = \sum_{k=1}^{\infty} c_k \lambda(E_k).$$

Choose compact sets K_k and open sets G_k so that

$$K_k \subset E_k \subset G_k$$

and such that

$$\lambda(G_k - K_k) < \frac{\varepsilon}{c_k 2^{k+2}}.$$

Define the function

$$L(x) = \sum_{k=1}^{\infty} c_k \chi_{G_k}(x).$$

This is a sum of l.s.c. functions and so $L(x)$ is defined at each point (most likely equal to ∞ at some points) and is itself a l.s.c. function. Certainly $L(x) \geq f(x)$ at each point.

Similarly (but not quite similarly), first choose an integer N so that

$$\sum_{k=N+1}^{\infty} c_k \lambda(E_k) < \varepsilon/4$$

and write the function

$$U(x) = \sum_{k=1}^{N} c_k \chi_{K_k}(x).$$

This is a finite sum of u.s.c. functions and so $U(x)$ is defined at each point and is itself a u.s.c. function. Certainly $U(x) \leq f(x)$ at each point.

We have

$$0 \leq L(x) - U(x) \leq \sum_{k=1}^{N} c_k b \chi_{G_k \setminus K_k}(x) + \sum_{k=N+1}^{\infty} c_k b \chi_{E_k}(x).$$

From this we deduce that

$$\int_a^b [L(x) - U(x)]\, dx < \varepsilon.$$

It follows, in particular, that L is in $\mathcal{S}_\ell^{\uparrow}[a,b]$, $f \leq L$ and

$$\int_a^b f(x)\, dx \leq \int_a^b L(x)\, dx \leq \int_a^b f(x)\, dx + \varepsilon.$$

We also know a similar fact for U, namely that that U is in $\mathcal{S}_\ell^{\downarrow}[a,b]$, $f \geq U$ and

$$\int_a^b f(x)\, dx \geq \int_a^b U(x)\, dx \geq \int_a^b f(x)\, dx - \varepsilon.$$

Thus we easily construct a sequence of functions $\{L_n\}$ from $\mathcal{S}_\ell^{\uparrow}[a,b]$ for which $f \leq L_n$, $L_n \to f$ pointwise a.e. and

$$\int_a^b f(t)\, dt = \lim_{n\to\infty} \int_a^b L_n(t)\, dt.$$

But we need a monotone nonincreasing sequence. Simply write

$$L_n' = \min\{L_1, L_2, L_3, \dots, L_n\}$$

to produce a monotone sequence $\{L_n'\}$ with the correct properties. In the same way we can also construct a monotone nondecreasing sequence of functions $\{U_n'\}$ from $\mathcal{S}_\ell^{\downarrow}[a,b]$ for which $f \geq U_n$, $U_n' \to f$ pointwise a.e. and

$$\int_a^b f(t)\, dt = \lim_{n\to\infty} \int_a^b U_n'(t)\, dt.$$

The proof so far assumes that f is nonnegative. In general take $f =$

$f^+ - f^-$, expressing f as a difference of integrable functions that are nonnegative. Produce the corresponding monotone sequences $\{L_n^+\}$, $\{U_n^+\}$, $\{L_n^-\}$, and $\{U_n^-\}$ for these two functions f^+ and f^-. Then the sequences

$$L_n = L_n^+ - U_n^- \text{ and } U_n = U_n^+ - L_n^-$$

will do the duty for the function f.

It follows that every integrable function f belongs to both $\mathcal{S}_\ell^{\uparrow\downarrow}[a,b]$ and $\mathcal{S}_u^{\downarrow\uparrow}[a,b]$. This completes the proof for the identities

$$\mathcal{L}[a,b] = \mathcal{S}_\ell^{\uparrow\downarrow}[a,b] = \mathcal{S}_u^{\downarrow\uparrow}[a,b].$$

We have seen that every integrable function can be approximated by semicontinuous functions and we know that semicontinuous functions can be approximated by functions in $\mathcal{S}_\ell$. This is the basis for establishing that $\mathcal{L} = \mathcal{S}_\ell^\uparrow - \mathcal{S}_\ell^\uparrow$. We already know that $\mathcal{L} \supset \mathcal{S}_\ell^\uparrow - \mathcal{S}_\ell^\uparrow$.

Let f be an arbitrary element of $\mathcal{L}$. Choose, for each $n = 1, 2, 3, \dots$ a step function g_n so that

$$\int_a^b |f(x) - g_n(x)|\, dx < 2^{-n}. \tag{4.35}$$

Note that

$$\sum_{n=1}^\infty \int_a^b |g_n(x) - g_{n-1}(x)|\, dx$$

$$\leq \sum_{n=1}^\infty \left\{ \int_a^b |g_n(x) - f(x)|\, dx + \int_a^b |f(x) - g_{n-1}(x)|\, dx \right\}$$

$$< \sum_{n=1}^\infty \left(2^{-n} + 2^{-n+1}\right) = 3.$$

This implies that the series

$$\sum_{n=1}^\infty |g_n(x) - g_{n-1}(x)|$$

converges for a.e. point $x \in [a,b]$. From that it follows that

$$\lim_{n\to\infty} g_n(x)$$

exists as a finite value for a.e. point $x \in [a,b]$. Applying Fatou's lemma to the inequality (4.35) shows us that, in fact,

$$\lim_{n\to\infty} g_n(x) = f(x) \text{ a.e.}$$

Define

$$s_n = [g_1]^+ + \sum_{k=1}^n [g_k - g_{k-1}]^+, \quad f_1 = \lim_{n\to\infty} s_n,$$

$$t_n = [g_1]^- + \sum_{k=1}^n [g_k - g_{k-1}]^-, \quad \text{and } f_2 = \lim_{n\to\infty} t_n.$$

We observe that

$$\lim_{n\to\infty} s_n(x) - \lim_{n\to\infty} t_n(x) = f_1(x) - f_2(x) = \lim_{n\to\infty} g_n(x) = f(x) \text{ a.e.}$$

and that the sequences $\{s_n\}$ and $\{t_n\}$ are monotone nondecreasing sequences of step functions for which

$$\sup_n \int_a^b s_n(x)\,dx < \infty \text{ and } \sup_n \int_a^b t_n(x)\,dx < \infty.$$

This proves the identity $\mathcal{L} = \mathcal{S}^{\uparrow} - \mathcal{S}^{\uparrow}$, where $\mathcal{S}$ is the collection of all step functions. If we adjust each of the functions in the two sequences at a finite number of points in an appropriate manner we can produce sequences of l.s.c. step functions with exactly the same properties. Thus we can conclude also that $\mathcal{L} = \mathcal{S}_{\ell}^{\uparrow} - \mathcal{S}_{\ell}^{\uparrow}$. The remaining identities in the statement of the theorem are left to the reader. ∎

For an alternative account of these methods we refer the reader to the literature on the Riesz method of presenting the Lebesgue integral. For example

`http://www.math.canterbury.ac.nz/~d.bridges/`

includes a link to Professor D. S. Bridges [10] lecture notes on a Riesz-method development of the Lebesgue integral. In that theory one *defines* $\mathcal{L}[a,b]$ to be

$$\mathcal{S}_{\ell}^{\uparrow}[a,b] - \mathcal{S}_{\ell}^{\uparrow}[a,b].$$

The obligation then is to develop all of the theory of the Lebesgue integral from that definition, including the necessary measure theory. An advantage of such a theory is that essentially the same steps can be used in abstract settings to develop an integration theory. One gains a deeper understanding of the structure of Lebesgue integration theory than the measure-theoretic approach offers. Many students, though, may not find the details all that palatable.

4.14.6 Vitali-Carathéodory property

It is useful to extract from the preceding arguments a general statement about the approximation of measurable functions by semicontinuous functions. The following theorem has been attributed by Saks [73, p.76] to Vitali [14] and Carathéodory[15].

[14] G. Vitali [90] Una proprietà delle funzioni misurabili, Istit. Lombardo Rend. (2), 38 (1905) 599–603.

[15] C. Carathéodory [16, p. 406], *Vorlesungen über reelle Funktionen*, Leipzig-Berlin (1918).

Theorem 4.64 (Vitali-Carathéodory) *Let $f : [a,b] \to \mathbb{R}$ be a measurable function. Then there exist two monotone sequences of functions $\{L_n\}$ and $\{U_n\}$ (possibly infinite-valued) such that*

1. L_n *are l.s.c. and* U_n *are u.s.c. on* $[a,b]$.
2. L_n *are bounded below and* U_n *are bounded above.*
3. *the sequence* $\{L_n\}$ *is nonincreasing and* $\{U_n\}$ *is nondecreasing.*
4. $L_n(x) \geq f(x) \geq U_n(x)$ *for every* x *in* $[a,b]$.
5. $\lim_{n\to\infty} L_n(x) = \lim_{n\to\infty} U_n(x) = f(x)$ *for a.e.* x *in* $[a,b]$.
6. *If* E *is a measurable subset of* $[a,b]$ *and* $f\chi_E$ *is integrable on* $[a,b]$ *then*
$$\lim_{n\to\infty} \int_a^b L_n\chi_E(x)\,dx = \lim_{n\to\infty} \int_a^b U_n\chi_E(x)\,dx = \int_a^b f(x)\chi_E(x)\,dx.$$

Note on infinite values: We can allow f in the statement of the theorem to be infinite-valued provided, as usual, the set of points where f assumes infinite values is of measure zero. We cannot, however, insist that the semicontinuous functions U_n and L_n are finite-valued, even if we require f to assume finite values (see Exercise 252). If f happens to be bounded, then there is no trouble making U_n and L_n also bounded.

4.14.7 Exercises

Exercise 244 *Show that $\mathcal{S}_\ell[a,b]$ has these "lattice" properties:*

1. *For all f, $g \in \mathcal{S}_\ell[a,b]$ and all nonnegative r and s, $rf + sg \in \mathcal{S}_\ell[a,b]$.*
2. *For all f, $g \in \mathcal{S}_\ell[a,b]$ both functions*
$$f \wedge g = \min\{f,g\} \quad \text{and} \quad f \vee g = \max\{f,g\}$$
belong to $\mathcal{S}_\ell[a,b]$

□

Exercise 245 *Show that $\mathcal{S}_\ell^\uparrow[a,b]$ has these "lattice" properties:*

1. *For all f, $g \in \mathcal{S}_\ell^\uparrow[a,b]$ and all nonnegative r and s, $rf + sg \in \mathcal{S}_\ell^\uparrow[a,b]$.*
2. *For all f, $g \in \mathcal{S}_\ell^\uparrow[a,b]$ both functions*
$$f \wedge g = \min\{f,g\} \quad \text{and} \quad f \vee g = \max\{f,g\}$$
belong to $\mathcal{S}_\ell^\uparrow[a,b]$

□

Exercise 246 *Show that $\mathcal{S}_\ell^\uparrow[a,b]$ and $\mathcal{S}_u^\downarrow[a,b]$ include all regulated functions.* □

Exercise 247 *Show that $\mathcal{S}_\ell^{\uparrow\uparrow}[a,b] = \mathcal{S}_\ell^\uparrow[a,b]$.* □

Exercise 248 *Define $\mathcal{C}[a,b]$ to be the set of all continuous functions on $[a,b]$. Characterize $\mathcal{C}^\uparrow[a,b]$, $\mathcal{C}^\downarrow[a,b]$, $\mathcal{C}^{\uparrow\downarrow}[a,b]$, and $\mathcal{C}^{\downarrow\uparrow}[a,b]$.* □

Exercise 249 *Show that $E \subset [a,b]$ is a set of measure zero if and only if there is a Riesz sequence $\{f_n\}$ of step functions so that $\lim_{n\to\infty} f_n(x) = \infty$ for all $x \in E$.* □

Exercise 250 *Suppose that $E \subset [a,b]$ is a set of measure zero and $\varepsilon > 0$. Show that there must exist a Riesz sequence $\{f_n\}$ of nonnegative step functions so that $\lim_{n\to\infty} f_n(x) = \infty$ for all $x \in E$ and*

$$\int_a^b f_n(x)\,dx < \varepsilon$$

for all $n = 1,2,3,\dots$. □

Exercise 251 *It is easy to misread the statement $\mathcal{L} = \mathcal{S}_\ell^\uparrow - \mathcal{S}_\ell^\uparrow$ as asserting that if f is Lebesgue integrable on an interval then f^+ and f^- must belong to $\mathcal{S}_\ell^\uparrow$. Show that this is not necessarily so.* Answer □

Exercise 252 *Show that there is a Lebesgue integrable function $f : [0,1] \to \mathbb{R}$ such that for no finite-valued l.s.c. function L is $f(x) \le L(x)$ for every x in $[0,1]$.* Answer □

4.15 Characterizations of the indefinite integral

Under what conditions can we be sure that a function $F : [a,b] \to \mathbb{R}$ can be written as

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

for a constant C and an integrable function f belonging to some specified class. We have already solved this problem for the class of Riemann integrable functions and several smaller classes as well. See Section 2.12.

The property and the characterization itself for absolutely integrable (i.e., Lebesgue integrable) functions were given by Giuseppe Vitali in 1905, only shortly after the publication by Lebesgue of his integration theory. We repeat the definition here for convenience, although it has played a role earlier in our discussions.

Definition 4.65 *Suppose that $F : [a,b] \to \mathbb{R}$ is a function. Then F is* absolutely continuous in the Vitali sense[a] *on $[a,b]$ if for all $\varepsilon > 0$ there is a $\delta > 0$ so that*

$$\sum_i |F(v_i) - F(u_i)| < \varepsilon$$

whenever $\{[u_i, v_i]\}$ are nonoverlapping subintervals of $[a,b]$ for which $\sum_i [v_i - u_i] < \delta$.

[a]Most texts call this (as did Vitali himself) simply "absolute continuity."

There are several simple consequences of this definition that we will require in order to better understand this concept. Let us recall from Lemma 2.30 that, if $F : [a,b] \to \mathbb{R}$ is absolutely continuous in Vitali's sense on $[a,b]$, then

1. F is uniformly continuous on $[a,b]$.
2. F has bounded variation on $[a,b]$ and, consequently, is a.e. differentiable in $[a,b]$.
3. F is λ-absolutely continuous on (a,b), i.e. F has zero variation on every measure zero subset of (a,b).
4. F maps subsets of (a,b) of Lebesgue measure zero into sets of Lebesgue measure zero.

4.15.1 Indefinite integral of nonnegative, integrable functions

We now present some answers to our general problem. Proofs follow below in Section 4.15.4. The first statement is for nonnegative integrable functions.

Theorem 4.66 *Let $F : [a,b] \to \mathbb{R}$. A necessary and sufficient condition in order that F can be written as*

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

for a constant C and a nonnegative *integrable function f is that F is absolutely continuous in the Vitali sense and monotonic nondecreasing.*

4.15.2 Indefinite integral of Lebesgue integrable functions

The second statement is for arbitrary absolutely integrable (i.e., Lebesgue) functions.

Theorem 4.67 *Let $F : [a,b] \to \mathbb{R}$. A necessary and sufficient condition in order that F can be written as*

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

for a constant C and an absolutely integrable *function f is that F is absolutely continuous in the Vitali sense.*

As a corollary we can rewrite this assertion in different language.

Corollary 4.68 *Let $F : [a,b] \to \mathbb{R}$. A necessary and sufficient condition in order that F can be written as*

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

for a constant C and an absolutely integrable *function f is that*

1. *F is absolutely continuous in the variational sense (ACG$_*$) on $[a,b]$.*
2. $\mathrm{Var}(F,[a,b]) < \infty$.

4.15.3 Indefinite integral of nonabsolutely integrable functions

The final statement is for arbitrary nonabsolutely integrable functions.

Theorem 4.69 *Let $F : [a,b] \to \mathbb{R}$. A necessary and sufficient condition in order that F can be written as*

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

for a constant C and a nonabsolutely integrable *function f are that*

1. *F is absolutely continuous in the variational sense (ACG$_*$) on $[a,b]$.*
2. $\mathrm{Var}(F,[a,b]) = \infty$.
3. *F is differentiable[a] almost everywhere in (a,b).*

[a] It is possible but not easy to show that when F is absolutely continuous in the variational sense, F *must* be almost everywhere differentiable. Thus (3) follows from (1).

4.15.4 Proofs

The necessity of the conditions in the three theorems can be addressed first. Suppose that

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

for a constant C and an integrable function f.

If f is nonnegative then F is certainly nondecreasing We check that it is also absolutely continuous in the Vitali sense.

Let $f_n(x) = \min\{f(x), n\}$ and note that f_n is measurable and nonnegative, and that $\lim_{n\to\infty} f_n(x) = f(x)$ everywhere. Then, by the monotone convergence theorem, on every subinterval $[c,d] \subset [a,b]$,

$$0 < \int_c^d f(x)\,dx - \int_c^d f_n(x)\,dx < \int_c^d [f(x) - f_n(x)]\,dx \to 0.$$

Choose N so large that

$$\int_a^b f(x)\,dx < \int_a^b f_N(x)\,dx + \varepsilon/2.$$

Choose $\delta = \varepsilon/(2N)$. Then check that, if $[c_i,d_i]$ are nonoverlapping subintervals of $[a,b]$ with $\sum_i (d_i - c_i) < \delta$, then

$$\begin{aligned}0 \le \sum_i [F(d_i) - F(c_i)] &= \sum_i \int_{c_i}^{d_i} f(x)\,dx \\ &\le \sum_i \int_{c_i}^{d_i} f_N(x)\,dx + \varepsilon/2 \\ &\le \sum_i N((d_i - c_i) + \varepsilon/2 < N\delta + \varepsilon/2 < \varepsilon.\end{aligned}$$

This verifies that F is absolutely continuous in the Vitali sense.

If we assume instead that f is absolutely integrable we can again obtain the fact that F is absolutely continuous in the Vitali sense merely by splitting f into its positive and negative parts.

Finally, if f is merely integrable, then we already know that the relation

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

requires that F is continuous everywhere, and that F is absolutely continuous. The fundamental theorem of the calculus requires $F'(x) = f(x)$ almost everywhere in $[a,b]$. Thus each of the necessity parts of the three theorems is proved.

Conversely the stated conditions in the theorems are sufficient to verify that

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

for some function f as stated and constant C. For the third theorem we already know this from the fundamental theorem of the calculus.

That same theorem shows that the proof of the first theorem is also complete provided we know that F is differentiable almost everywhere and that $F'(x) \ge 0$ almost everywhere. But we already know that nondecreasing functions are almost everywhere differentiable. Take $f(x) = F'(x)$ at points where the derivative exists and $f(x) = 0$ elsewhere and the first theorem is proved.

We complete the proof of the second theorem in the same way. The assumption that F is absolutely continuous in the Vitali sense assures us that F is continuous and has bounded variation. So again F is almost everywhere differentiable and again the same argument supplies the representation.

Exercises

Exercise 253 *Show that a function that is Lipschitz on $[a,b]$ is also absolutely continuous in the Vitali sense on $[a,b]$.* □

Exercise 254 *Give an example of a uniformly continuous on an interval $[a,b]$ that is not absolutely continuous in the Vitali sense there.* □

Exercise 255 *Given an example of a function that is not Lipschitz on $[a,b]$ but is absolutely continuous in the Vitali sense on $[a,b]$.* □

Exercise 256 *Show that any continuously differentiable function on an interval $[a,b]$ is absolutely continuous in the Vitali sense on $[a,b]$.* □

Exercise 257 *Show that a differentiable function on an interval $[a,b]$ need not be absolutely continuous in the Vitali sense on $[a,b]$ but that it must be absolutely continuous in the variational sense.* □

Exercise 258 *Show that a function may be absolutely continuous in the variational sense but not absolutely continuous in the Vitali sense.*

Answer □

Exercise 259 (Fichtenholz) *Suppose that $F:[a,b]\to\mathbb{R}$ satisfies the following condition: for every $\varepsilon>0$ there is a $\delta>0$ so that whenever $\{[c_i,d_i]\}$ is* any *sequence of subintervals of $[a,b]$ satisfying $\sum_i(d_i-c_i)<\delta$ then necessarily $\sum_i|F(d_i)-F(c_i)|<\varepsilon$. Show that this condition is strictly stronger than absolutely continuity in the Vitali sense.*

Answer □

4.16 Extending Lebesgue's integral

The Lebesgue integral integrates all bounded, measurable functions on an interval $[a,b]$. Is there a "better" integral, one that integrates *all* bounded functions?

More precisely we would like to have an integral which extends Lebesgue's integral in such a way that it integrates all bounded functions and agrees with the Lebesgue integral if the function integrated happens to be measurable. In this section we consider this problem.

A simple scheme is to search for a more general way to take limits of Riemann sums. We know that a uniform limit of Riemann sums produces only the very limited Riemann integral; a pointwise limit of Riemann sums handles all Lebesgue integrable functions. Here is a general framework.

Fix a compact interval $[a,b]$ and let Π denote the collection of all partitions π of this interval. By an *integration scheme* on $[a,b]$ we shall mean a filter $\mathcal{F}$ on Π, i.e., $\mathcal{F}$ is a nonempty collection of nonempty subsets of Π with the two properties

1. If $\alpha\in\mathcal{F}$ and if $\alpha\subset\alpha'\subset\Pi$ then $\alpha'\in\mathcal{F}$.

2. If $\alpha_1, \alpha_2 \in \mathcal{F}$ then $\alpha_1 \cap \alpha_2 \in \mathcal{F}$.

Let now $f : [a,b] \to \mathbb{R}$ and simply write upper and lower integrals relative to the filter $\mathcal{F}$ in the usual way (as in Section 3.1).

Definition 4.70 *For a function $f : [a,b] \to \mathbb{R}$ and a filter $\mathcal{F}$ on Π we define*

$$(\mathcal{F})\overline{\int_a^b} f(x)\,dx = \inf_{\alpha \in \mathcal{F}} \sup_{\pi \in \alpha} \left\{ \sum_{([u,v],w) \in \pi} f(w)(v-u) \right\}$$

and

$$(\mathcal{F})\underline{\int_a^b} f(x)\,dx = \sup_{\alpha \in \mathcal{F}} \inf_{\pi \in \alpha} \left\{ \sum_{([u,v],w) \in \pi} f(w)(v-u) \right\}$$

A function f would be said to be $\mathcal{F}$*-integrable* if the upper and lower integrals agree and are finite.

There are three filters that we should consider:

- $\mathcal{R}$ denotes the collection of all nonempty subsets $\alpha \subset \Pi$ with the property that there is a $\delta > 0$ so that every partition $\pi \in \Pi$ with $v - u < \delta$ for all $([u,v],w) \in \pi$ must belong to α.

- $\mathcal{H}$ denotes the collection of all nonempty subsets $\alpha \subset \Pi$ with the property that for each $x \in [a,b]$ there is a $\delta(x) > 0$ so that every partition $\pi \in \Pi$ with $v - u < \delta(w)$ for all $([u,v],w) \in \pi$ must belong to α.

- $\mathcal{U}$ is an ultrafilter on Π, i.e., $\mathcal{U}$ is maximal in the sense that there is no larger filter, no other filter that contains $\mathcal{U}$.

The integral based on $\mathcal{R}$ is the Riemann integral on $[a,b]$. The integral based on $\mathcal{H}$ is precisely the Henstock-Kurzweil integral. The integral based on an ultrafilter has the following property.

Theorem 4.71 *Let $\mathcal{U}$ be an ultrafilter*[a] *that includes $\mathcal{H}$. Then for every function $f : [a,b] \to \mathbb{R}$,*

$$(\mathcal{H})\underline{\int_a^b} f(x)\,dx \le (\mathcal{U})\underline{\int_a^b} f(x)\,dx = (\mathcal{U})\overline{\int_a^b} f(x)\,dx \le (\mathcal{H})\overline{\int_a^b} f(x)\,dx.$$

In particular every bounded function is $\mathcal{U}$-integrable and the $\mathcal{U}$-integral extends both the Lebesgue integral and the Henstock-Kurzweil integral.

[a] It follows from the axiom of choice that such an ultrafilter must exist.

There is no practical importance, however, in such a theorem. An ultrafilter cannot be constructed; we can claim that an ultrafilter exists only by appealing to a logical principle such as the axiom of choice. Nonetheless we have an integral, based like our usual integrals on Riemann sums, that extends the Lebesgue integral. We will not have available all of the properties of the Lebesgue integral; in particular we cannot claim translation invariance nor that a monotone convergence theorem would hold for the extended integral.

For a different approach to the problem of extending the Lebesgue integral see Chapter 12, Section 12.6 of our text Bruckner, Bruckner, and Thomson, *Real Analysis*, 2nd Ed., ClassicalRealAnalysis.com (2008). There the extension is accomplished as an application of the Hahn-Banach theorem. That theorem produces a linear functional that behaves very much like an integral but there is no actual connection with integration theory since the integral is not constructed by any recognizable method.

Exercise 260 *What property of a filter $\mathcal{F}$ would guarantee that*

$$(\mathcal{F})\underline{\int_a^b} f(x)\,dx \leq (\mathcal{F})\overline{\int_a^b} f(x)\,dx?$$

□

Exercise 261 *Prove Theorem 4.71.* Answer □

4.17 The Lebesgue integral as a set function

In many presentations of the Lebesgue integral (although not in Lebesgue's original thesis) the integral is defined over arbitrary measurable sets E and denoted as

$$\int_E f(x)\,dx.$$

Then the integral over a compact interval $[a,b]$ would be written as

$$\int_{[a,b]} f(x)\,dx$$

and all of the theory is stated, as far as is possible, for the more general set-valued integral (rather than the interval-valued integral of this chapter). We can define this set-valued integral in somewhat greater generality by using estimates arising from full and fine covers.

Definition 4.72 *Let $f:\mathbb{R}\to\mathbb{R}$ be a function and β a covering relation. We write*

$$\operatorname{Var}(f\lambda,\beta) = \sup_{\pi\subset\beta}\left\{\sum_{([u,v],w)\in\pi} |f(w)|\lambda(([u,v])\right\}$$

where the supremum is taken over all π, arbitrary subpartitions contained in β.

Definition 4.73 (Full and Fine Variations) *Let $f : \mathbb{R} \to \mathbb{R}$ and let E be any set of real numbers. Then we define the full and fine variational measures associated with $f\lambda$ by the expressions:*

$$V^*(f\lambda E) = \inf\{\mathrm{Var}(f\lambda, \beta) : \ \beta \text{ a full cover of } E\}$$

and

$$V_*(f\lambda, E) = \inf\{\mathrm{Var}(f\lambda, \beta) : \ \beta \text{ a fine cover of } E\}.$$

In the special case where f is a nonnegative function and E an arbitrary set we write

$$\int_E f(x)\,dx = V^*(f\lambda, E)$$

and we will check later to see if fine variation can be used as well. We have already sufficient techniques to study this set-valued integral and so we shall develop the theory in the exercises.

Exercises

Exercise 262 (measure estimates for Lebesgue's integral) *Suppose that $f : \mathbb{R} \to \mathbb{R}$ is an arbitrary nonnegative function and that $0 \le r < f(x) < s$ for all x in a set E. Then*

$$r\lambda(E) \le \int_E f(x)\,dx \le s\lambda(E).$$

Answer □

Exercise 263 (comparison with upper integral) *Show that if f is a nonnegative function and E is an arbitrary set contained in an interval $[a,b]$ then*

$$\int_E f(x)\,dx = \overline{\int_a^b} \chi_E(x) f(x)\,dx.$$

□

Exercise 264 (comparison with Lebesgue integral) *Show that if f is a nonnegative measurable function and E is a measurable set contained in an interval $[a,b]$ then*

$$\int_E f(x)\,dx = \int_a^b \chi_E(x) f(x)\,dx$$

where the integral may be interpreted as a Lebesgue integral. (In particular the value of the integral $\int_E f(x)\,dx$ can be constructed by Lebesgue's methods.) □

Exercise 265 (measure properties) *Show that if f is a nonnegative function and $E, E_1, E_2, E_3, \dots$ is a sequence of sets for which $E \subset \bigcup_{n=1}^{\infty} E_n$ then*

$$\int_E f(x)\,dx \le \sum_{n=1}^{\infty} \int_{E_n} f(x)\,dx$$

i.e., the set function integral is a measure in the sense of Theorem 4.3. □

Exercise 266 (absolute continuity (zero/zero)) *Show that if f is a nonnegative function and E is a set of Lebesgue measure zero then*

$$\int_E f(x)\,dx = 0.$$

Answer □

Exercise 267 *Show that if f is a nonnegative function and*

$$\int_E f(x)\,dx = 0$$

then $f(x) = 0$ for almost every point $x \in E$. □

Exercise 268 *Show that if f is a nonnegative function and $E_1, E_2, E_3, \dots$ is a sequence of pairwise disjoint closed sets for which $E = \bigcup_{n=1}^{\infty} E_n$ then*

$$\int_E f(x)\,dx = \sum_{n=1}^{\infty} \int_{E_n} f(x)\,dx$$

i.e., the set function integral is additive over disjoint closed sets as in Corollary 4.11. □

Exercise 269 *Suppose that f is a nonnegative, bounded function and that E is a measurable set. Show that for every $\varepsilon > 0$ there is an open set G so that $E \setminus G$ is closed and*

$$\int_{E\setminus G} f(x)\,dx < \varepsilon.$$

[This is a warm-up to the next exercise where bounded is dropped.] □

Exercise 270 *Suppose that f is a nonnegative, measurable function and that E is a measurable set. Show that for every $\varepsilon > 0$ there is an open set G so that $E \setminus G$ is closed and*

$$\int_{E\setminus G} f(x)\,dx < \varepsilon.$$

[This is an improvement on the preceding exercise where it was assumed that the function is bounded.] □

Exercise 271 *Show that if f is a nonnegative measurable function and $E_1, E_2, E_3, \dots$ is a sequence of pairwise disjoint measurable sets for which $E = \bigcup_{n=1}^{\infty} E_n$ then, for any set A,*

$$\int_{A\cap E} f(x)\,dx = \sum_{n=1}^{\infty} \int_{A\cap E_n} f(x)\,dx$$

i.e., the set function integral is additive over disjoint sets as in Lemma 4.14 provided we assume that the sets and the function are measurable. □

Exercise 272 *Show that if f is a nonnegative measurable function and $E_1 \subset E_2 \subset E_3 \subset \dots$, is an increasing sequence of measurable sets for which $E =$*

$\bigcup_{n=1}^{\infty} E_n$ then
$$\int_E f(x)\,dx = \lim_{n\to\infty} \int_{E_n} f(x)\,dx.$$
□

Exercise 273 *Suppose that $f:\mathbb{R}\to\mathbb{R}$ and that f is nonnegative and bounded. Then for every $\varepsilon>0$ there is a $\delta>0$ so that if G is an open set with $\lambda(G)<\delta$ then*
$$\int_G f(x)\,dx < \varepsilon.$$
[This is a warm-up to the next exercise where bounded is dropped.] Answer □

Exercise 274 (absolute continuity (ε, δ)) *Suppose that $f:\mathbb{R}\to\mathbb{R}$, that f is nonnegative and measurable, and that*
$$\int_E f(x)\,dx < \infty.$$
Then for every $\varepsilon>0$ there is a $\delta>0$ so that if G is an open set with $\lambda(G)<\delta$ then
$$\int_{E\cap G} f(x)\,dx < \varepsilon.$$
Answer □

Exercise 275 (construction of the Lebesgue integral) *Suppose that $f:\mathbb{R}\to\mathbb{R}$ and that f is a nonnegative, measurable function. Let $r>1$ and write*
$$A_{kr} = \{x : r^{k-1} < f(x) \le r^k\}.$$
Then, for any set E,
$$\int_E f(x)\,dx \le \sum_{k=-\infty}^{\infty} r^k \lambda(E\cap A_{kr}) \le r\int_E f(x)\,dx.$$
[In particular as $r \nearrow 1$ the sum approaches the value of the integral.] Answer □

Exercise 276 (full and fine characterization) *Suppose that $f:\mathbb{R}\to\mathbb{R}$ and that f is a nonnegative, measurable function. Show that*
$$\int_E f(x)\,dx = V^*(f\lambda, E) = V_*(f\lambda, E).$$
Answer □

4.18 The abstract Lebesgue integral

Having mastered the classical Lebesgue theory of integration, we are now in a position to discuss (albeit briefly) the abstract version of integration theory. We remove from consideration all features that are special to integration theory on the real line, and attempt to capture a theory that can be applied in many general situations.

4.18.1 Assumptions of the general theory

We wish to define a Lebesgue-type integral

$$\int_X f(x)\,d\mu(x)$$

for real-valued functions f defined on an abstract set X (called the "space"). The space X is assumed to have some structure that permits a measure theory similar to the theory of Lebesgue measure on the real line. We assume a special class of subsets exists (to play the role of the Lebesgue measurable sets) and we assume a measure μ is supplied to play the same role as the Lebesgue measure λ.

1. X is a nonempty set called the space of integration.
2. There is a class of subsets of X denoted as $\mathcal{M}$ and called the measurable subsets of X.
3. There is a nonnegative set function μ defined on the class $\mathcal{M}$.

In order to use the elements of the Lebesgue theory we assume that $(X,\mathcal{M},\mu)$ is a *measure space*. This requires that $\mathcal{M}$ is a Borel family (as defined in Theorem 4.15) and that the nonnegative set function μ defined on the class $\mathcal{M}$ is a measure (i.e., it is a countably additive set function having the property expressed in Theorem 4.14).

4.18.2 Defining the integral

In outline the theory of the integral defined relative to a measure space $(X,\mathcal{M},\mu)$ can follow the same general steps as we used in the construction of the integral according to Lebesgue's original ideas:

1. A function $f : X \to \mathbb{R} \cup \{+\infty\}$ is said to be an *nonnegative elementary function* if
$$f(x) = \sum_{i=1}^{\infty} c_i \chi(E_i)$$
for numbers $\{c_1, c_2, c_3, \dots\}$ in $\mathbb{R}^+ \cup \{+\infty\}$ and sets $\{E_1, E_2, E_3, \dots\}$ belonging to $\mathcal{M}$.
2. For any set $E \in \mathcal{M}$ and any such nonnegative elementary function the integral is defined as
$$\int_E f(x)\,d\mu(x) = \sum_{i=1}^{\infty} c_i \mu(E \cap E_i).$$

3. A function
$$f : X \to \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$$
is said to be *measurable* if f^+ and f^- are both elementary.

4. For any set $E \in \mathcal{M}$ and any such measurable function the integral is defined as
$$\int_E f(x)\,d\mu(x) = \int_E f^+(x)\,d\mu(x) - \int_E f^-(x)\,d\mu(x),$$
provided this is defined (i.e., provided one at least of these two values in the difference is finite).

This definition is remarkably compact and displays the basics of the theory of abstract integration most efficiently. Nonetheless it conceals some hard technical work, work that we can indicate only briefly.

Exercise 277 *If this outline were to be followed to define an integral, what justifications would be needed?*

Answer □

4.18.3 Properties of the integral

Having defined the abstract Lebesgue integral one then proceeds to develop its properties. This is very much a copying exercise. The same properties that hold for the classical Lebesgue integral will hold for the abstract Lebesgue integral (with reasonable exceptions). To some extent the methods used to establish the classical properties extend to this situation too. Thus the additive properties of integrals, the monotone convergence theorem, the dominated convergence theorem, etc. can all be stated and proved in this more general setting.

Some differences arise. First we have defined the Lebesgue integral on a finite interval $[a,b]$ and such intervals have finite Lebesgue measure. Here the integral
$$\int_E f(x)\,d\mu(x)$$
might be taken on a set E for which $\mu(E) = \infty$. That could require some attention to details requiring more careful statement of theorems or altered arguments.

More importantly much of the classical theory takes place in a setting (the real line) where much more structure is apparent. The abstract theory will have no direct counterparts to assertions that are special to the real line. For example, Lusin's theorem (Theorem 4.50) cannot be carried over to the abstract theory without also assuming that the space X carries some extra structure that allows functions to be "continuous."

Our advanced textbook [13] can be consulted for a complete introduction to abstract integration theory. Free downloads of PDf files for viewing on laptops and tablets make this a particularly accessible choice. The reader who has survived thus far in the present text should be adequately equipped for this more advanced study.

Exercise 278 *Assume that the definition of the integral as given can be justified. If so, then verify this additive property: If f and g are both nonnegative elementary functions, then*

$$\int_E [f(x)+g(x)]\,d\mu(x) = \int_E f(x)\,d\mu(x) + \int_E g(x)\,d\mu(x).$$

Answer □

Exercise 279 *Assume that the definition of the integral as given can be justified. If so, then verify this additive property: If f is a nonnegative elementary function and $\{E_i\}$ is a pairwise disjoint sequence of measurable sets and*

$$E = \bigcup_{i=1}^{\infty} E_i,$$

then

$$\int_E f(x)\,d\mu(x) = \sum_{i=1}^{\infty} \int_{E_i} f(x)\,d\mu(x)$$

□

Exercise 280 *Assume that the definition of the integral as given can be justified. If so, then verify this monotone convergence property: If $\{g_i\}$ is a sequence of nonnegative elementary functions and*

$$f(x) = \sum_{i=1}^{\infty} g_i(x),$$

then f is an elementary function and

$$\int_E f(x)\,d\mu(x) = \sum_{i=1}^{\infty} \int_E g_i(x)\,d\mu(x).$$

Answer □

Chapter 5

Stieltjes Integrals

Recall that the *total variation* of a function F on a compact interval is the supremum of sums of the form

$$\operatorname{Var}(F,[a,b]) = \sum_{([u,v],w)\in\pi} |F(v)-F(u)|$$

taken over all possible partitions π of $[a,b]$. This is a measure of the variability of the function F on this interval. Functions of bounded variation play a significant role in real analysis generally, especially so in the theory of integration. The earliest application was to the study of arc length of curves, a subject we will review in this chapter as well.

Our main tool in the study of this important class of functions is a slight generalization of the integral, called the Stieltjes integral. Our definitions for this integral will now be of the Henstock-Kurzweil type. Ideas related to the usual integral will certainly return.

5.1 Stieltjes integrals

The definition of the total variation $\operatorname{Var}(F,[a,b])$ contains what looks very much like one of our Riemann sums, but in place of the usual sum

$$\sum_{([u,v],w)\in\pi} f(w)(v-u)$$

we are here checking values of the sum

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)|.$$

This might suggest to us that integration methods would prove a useful tool in the study of functions of bounded variation.

Let us, accordingly, enlarge the scope of our integration theory by considering limits of Riemann sums that are more general than we have used so far. Let

$f, G : [a,b] \to \mathbb{R}$ and by analogy with

$$\int_a^b f(x)\,dx \sim \sum_{([u,v],w)\in\pi} f(w)(v-u)|$$

we introduce new integrals by making only the obvious changes suggested by the following slogans:

$$\int_a^b f(x)\,dG(x) \sim \sum_{([u,v],w)\in\pi} f(w)(G(v)-G(u))$$

$$\int_a^b f(x)\,|dG(x)| \sim \sum_{([u,v],w)\in\pi} f(w)|G(v)-G(u)|$$

$$\int_a^b f(x)\,[dG(x)]^+ \sim \sum_{([u,v],w)\in\pi} f(w)[G(v)-G(u)]^+$$

$$\int_a^b f(x)\,[dG(x)]^- \sim \sum_{([u,v],w)\in\pi} f(w)[G(v)-G(u)]^-$$

as well as a few other variants we consider in later sections:

$$\int_a^b \sqrt{|dG(x)|\,dx} \sim \sum_{([u,v],w)\in\pi} \sqrt{|G(v)-G(u)|(v-u)}$$

and

$$\int_a^b \sqrt{[dG(x)]^2+[dx]^2} \sim \sum_{([u,v],w)\in\pi} \sqrt{|G(v)-G(u)|^2+(v-u)^2}.$$

Notation Here we are using the expression $[r]^+$ for any real number to denote $(|r|+r)/2$ (or equivalently $\max\{r,0\}$). Similarly $[r]^-$ for any real number is $(|r|-r)/2$ (or equivalently $\max\{-r,0\}$). Note that, with these definitions,

$$|r| = [r]^+ + [r]^- \text{ and } r = [r]^+ - [r]^-.$$

We will refer to all of these as *Stieltjes integrals*, although it is only the first variant of these,

$$\int_a^b f(x)\,dG(x),$$

that the Dutch mathematician Thomas Stieltjes (1856–1894) himself used and the one that most people would mean by the terminology.

5.1.1 Definition of the Stieltjes integral

The slogans in the preceding section should be enough to lead the reader to the correct definition of the various Stieltjes integral. Even so, let us give a precise definition for the simplest case. This is just a copying exercise: take the usual definition and repeat it with the Riemann sums adjusted in the manner required.

Definition 5.1 *For functions G, $f : [a,b] \to \mathbb{R}$ we define an* upper integral *by*

$$\overline{\int_a^b} f(x)\, dG(x) = \inf_{\beta} \sup_{\pi \subset \beta} \sum_{([u,v],w) \in \pi} f(w)(G(v) - G(u))$$

where the supremum is taken over all partitions π of $[a,b]$ contained in β, and the infimum over all full covers β.

Similarly we define a *lower integral*, as

$$\underline{\int_a^b} f(x)\, dG(x) = \sup_{\beta} \inf_{\pi \subset \beta} \sum_{([u,v],w) \in \pi} f(w)(G(v) - G(u))$$

where, again, π is a partition of $[a,b]$ and β is a full cover.

If the upper and lower integrals are identical, i.e., if

$$\underline{\int_a^b} f(x)\, dG(x) = \overline{\int_a^b} f(x)\, dG(x),$$

we say the integral is *determined* and we write the common value as

$$\int_a^b f(x)\, dG(x).$$

We are interested, mostly, in the case in which the integral is determined and finite.

Exercise 281 *Show that, in general,*

$$\underline{\int_a^b} f(x)\, dG(x) \leq \overline{\int_a^b} f(x)\, dG(x).$$

Answer □

5.1.2 Definition of the Riemann-Stieltjes integral

The most familiar version of the Stieltjes integral is the Riemann-Stieltjes integral. The reader is likely to encounter the latter integral in the literature, rather than the more general one defined here. The difference is only in the use of uniform full covers rather than full covers in general. Just as the Riemann integral is a weak cousin compared to the integral we have studied, so too is the Riemann version of a Stieltjes integral a weak special case.

Even so the reader should be aware of the distinction. We occasionally allude to the Riemann-Stieltjes integral but it is not our main preoccupation. The definition is just a direct copy of Definition 1.20 adjusted to this more general integrand.

Definition 5.2 (Riemann-Stieltjes integral) *Let f, G be a function that are defined at every point of $[a,b]$. Then, f is said to be Riemann-Stieltjes integrable on $[a,b]$ with respect to G if it satisfies the following "uniform integrability" criterion: there is a number I so that, for every $\varepsilon > 0$ there is a $\delta > 0$, with the property that*

$$\left| I - \sum_{i=1}^{n} f(\xi_i)(G(x_i) - G(x_{i-1})) \right| < \varepsilon$$

whenever points are given

$$a = x_0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n = b$$

for which

$$x_i - x_{i-1} < \delta$$

with associated points $\xi_i \in [x_{i-1}, x_i]$.

The number I in the definition might then be written in integral notation as

$$I = (RS) \int_a^b f(x)\, dG(x).$$

We recognize this weaker integral as just a special case of the Stieljtes integral of this chapter. On occasion it is useful to know if the integral exists in this restricted sense (just as it might be useful to know if an integrable function happens to be also Riemann integrable).

Exercises

Exercise 282 *Let $G : [a,b] \to \mathbb{R}$. Show that*

$$\int_a^b dG(x) = G(b) - G(a).$$

$\square$

Exercise 283 *Let $G : \mathbb{R} \to \mathbb{R}$ defined so that $G(x) = 0$ for all $x \neq 0$ and $G(1) = 1$. Compute*

$$\overline{\int_0^2} |dG(x)| \quad \text{and} \quad \underline{\int_0^2} |dG(x)|.$$

$\square$

Exercise 284 *Let $G : [0,1] \to \mathbb{R}$ and let $f(x) = 0$ for all $x \neq 1/2$ with $f(1/2) = 1$. What are*

$$\overline{\int_0^1} f(x)\, dG(x) \text{ and } \underline{\int_0^1} f(x)\, dG(x)?$$

$\square$

Exercise 285 *Let G, $f : [0,1] \to \mathbb{R}$ and let $G(x) = 0$ for all $x \leq 1/2$ and with*

$G(x) = 1$ for all $x > 1/2$. What are

$$\overline{\int_0^1} f(x)\,dG(x) \text{ and } \underline{\int_0^1} f(x)\,dG(x)?$$

Answer □

Exercise 286 *Let G, $f : [a,b] \to \mathbb{R}$ and let f be continuous and let G be a step function, i.e. there are points*

$$a < \xi_1 < \xi_2 < \cdots < \xi_m < b$$

so that G is constant on each interval (ξ_{i-1}, ξ_i). What are possible values for

$$\overline{\int_a^b} f(x)\,dG(x) \text{ and } \underline{\int_a^b} f(x)\,dG(x)?$$

Answer □

Exercise 287 *Let G, $F : [-1,1] \to \mathbb{R}$ be defined by $F(x) = 0$ for $-1 \leq x < 0$, $F(x) = 1$ for $0 \leq x \leq 1$, $G(x) = 0$ for $-1 \leq x \leq$, and $G(x) = 1$ for $0 < x \leq 1$. Discuss $\int_{-1}^1 F(x)\,dG(x)$ and $\int_{-1}^1 G(x)\,dF(x)$.* Answer □

Exercise 288 *If $a < b < c$ is the formula*

$$\int_a^b f(x)\,dG(x) + \int_b^c f(x)\,dG(x) = \int_a^c f(x)\,dG(x)$$

valid? Answer □

Exercise 289 *Show that a function f can be altered at a finite number of points where G is continuous without altering the values of the upper and lower integrals. Give an example to show that continuity may not be dropped here.* □

Exercise 290 *Show that a function f can be altered at a countable number of points where G is continuous without altering the values of the upper and lower integrals.* □

Exercise 291 *For integrals of the form $\displaystyle\int_a^b f(x)\,|dG(x)|$ what changes have to be made in the various criteria?* Answer □

Exercise 292 *For integrals of the form $\displaystyle\int_a^b f(x)\,[dG(x)]^+$ what changes have to be made in the various criteria?* □

Exercise 293 *Let $F : [0,2] \to \mathbb{R}$ with $F(t) = 0$ for all $t \neq 1$ and $F(1) = 1$. Show that*

$$\underline{\int_0^2} |dF(x)| < \overline{\int_0^2} |dF(x)| = \mathrm{Var}(F, [0,2]).$$

□

Exercise 294 *Let $F : [a,b] \to \mathbb{R}$. Show that the total variation of F can be expressed as an upper integral:*

$$\mathrm{Var}(F,[a,b]) = \overline{\int_a^b} |dF(x)|.$$

□

Exercise 295 *Let $F : [a,b] \to \mathbb{R}$ and suppose that one at least of the integrals*

$$\overline{\int_a^b} |dF(x)|, \ \overline{\int_a^b} [dF(x)]^+ \ \text{ or } \ \overline{\int_a^b} [dF(x)]^-$$

is finite. Show that F is a function of bounded variation on $[a,b]$ and that, for all $a < t \leq b$,

$$F(t) - F(a) = \overline{\int_a^t} [dF(x)]^+ - \overline{\int_a^t} [dF(x)]^-. \tag{5.1}$$

The identity (5.1) is a representation of F as a difference of two nondecreasing functions. □

Exercise 296 *Let $F : [a,b] \to \mathbb{R}$ be a continuous function. Show that F has bounded variation on $[a,b]$ if and only if there is a continuous, strictly increasing function $G : [a,b] \to \mathbb{R}$ for which $F(d) - F(c) < G(d) - G(c)$ for all $a \leq c < d \leq b$.* □

Exercise 297 *What basic properties of the ordinary integral $\int_a^b f(x)\,dx$ can you prove for Stieltjes integrals without any but the most obvious of changes in the proofs?* □

5.1.3 Henstock's zero variation criterion

Since the Stieltjes integral is defined in exactly the same way as the ordinary integral one expects almost the same properties. Indeed this integral has the same linear, additive, and monotone properties (suitably expressed). There also must be an indefinite integral. Finally, the most important of these properties that carries over, is the Henstock criterion. We give that now.

Theorem 5.3 *Let $F, G, f : [a,b] \to \mathbb{R}$. Then a necessary and sufficient condition for the existence of the Stieltjes integral and the formula*

$$\int_c^d f(x)\,dG(x) = F(d) - F(c) \quad [c,d] \subset [a,b]$$

is that

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0.$$

The proof would merely be a copying exercise of material from earlier. Note that we are taking advantage of our general Stieltjes notation here to allow us to interpret the integral

$$\int_a^b |dF(x) - f(x)\,dG(x)|$$

as a limit of the Riemann sums

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(x)[G(v) - G(u)]|\,.$$

5.2 Regulated functions

We have already introduced the important class of regulated functions in Section 1.9.1. We review that definition here and apply the ideas to our Stieljtes integrals. Recall that the one-sided limit $F(c+)$ exists if, for all sequences of positive numbers t_n tending to zero,

$$\lim_{n\to\infty} F(c + t_n) = F(c+).$$

Similarly, we say $F(c-)$ exists if, for all sequences of positive numbers t_n tending to zero,

$$\lim_{n\to\infty} F(c - t_n) = F(c-).$$

Definition 5.4 *Let $F : [a,b] \to \mathbb{R}$. Then*

- *F is said to be* regulated *if the one-sided limit $F(c+)$ exists and is finite for all $a \le c < b$ and the limit on the other side $F(c-)$ exists and is finite for all $a < c \le b$.*

- *F is said to be* naturally regulated *if F is regulated and, for all $a < c < b$, either*

  $$F(c+) \le F(c) \le F(c-)$$

  or else

  $$F(c-) \le F(c) \le F(c+).$$

Theorem 5.5 *Let $F : [a,b] \to \mathbb{R}$ be monotonic. Then F is naturally regulated.*

Proof. Simply notice that

$$F(c-) = \sup\{F(t) : a \le t < c\} \le F(c)$$
$$\le \inf\{F(t) : c < t \le b\} = F(c+).$$

for all $a < c < b$. ∎

Theorem 5.6 *Let $F : [a,b] \to \mathbb{R}$ be a function of bounded variation. Then F is regulated and has at most countably many discontinuities*[a].

[a]In fact it can be proved that *all* regulated functions have at most countably many discontinuities.

Proof. Suppose that $a < c \leq b$ and $F(c-)$ does not exist. Then there is a positive number ε and a sequence of numbers c_n increasing to c so that, for all n,

$$F(c_n) - F(c_{n+1}) < -\varepsilon < \varepsilon < F(c_{n+2}) - F(c_{n+1}).$$

But then, for all m,

$$\infty > \mathrm{Var}(F,[a,b]) \geq \sum_{n=1}^{m} |F(c_n) - F(c_{n+1})| > m\varepsilon.$$

This is impossible. Similarly $F(c+)$ must exist for all $a \leq c < b$.

Let us show that there are only countably many points $c \in [a,b)$ for which $F(c) \neq F(c+)$. Let c_1, c_2, $\ldots c_m$ denote some set of m points from (a,b) for which

$$|F(c_m+) - F(c)| > 1/n.$$

Then there is a disjointed collection of intervals $[c_i,t_i]$ for which

$$|F(t_i) - F(c_i)| > 1/(2n).$$

In particular

$$\infty > \mathrm{Var}(F,[a,b]) \geq \sum_{i=1}^{m} |F(t_i) - F(c_i)| > m/(2n).$$

Thus there are only finitely many such choices of points c_1, c_2, $\ldots c_m$ for which

$$|F(c_m+) - F(c_m)| > 1/n.$$

It follows that there are only countably many choices of points c_i for which

$$|F(c_i+) - F(c_i)| > 0.$$

A similar argument handles the points $c \in (a,b)]$ for which $F(c) \neq F(c-)$. It follows that the set of points of discontinuity must be countable. ∎

5.2.1 Approximate additivity of naturally regulated functions

Our study of the integration properties of naturally regulated functions will require a simple lemma.

Lemma 5.7 (Approximate additivity) *Suppose that $F : [a,b] \to \mathbb{R}$ is a function that is naturally regulated. Then at any point $a < c < b$, and for any $\varepsilon > 0$ there is $\delta > 0$ so that, for all $c - \delta < u < c < v < c + \delta$,*

$$|F(v) - F(c)| + |F(c) - F(u)| \geq |F(v) - F(u)|$$

and

$$|F(v) - F(u)| \geq |F(v) - F(c)| + |F(c) - F(u)| - \varepsilon. \tag{5.2}$$

Proof. Since F is naturally regulated we know that

$$|F(c+) - F(c-)| = |F(c+) - F(c)| + |F(c-) - F(c)|$$

for each $a < c < b$. At such points there is a $\delta > 0$ so that

$$|F(u) - F(c-)| < \varepsilon/4 \text{ and } |F(v) - F(c+)| < \varepsilon/4$$

for all $c - \delta < u < c < v < c + \delta$. In particular

$$|F(c+) - F(c-)| \leq |F(c+) - F(v)| + |F(v) - F(u)| + |F(u) - F(c-)|$$
$$\leq |F(v) - F(u)| + \varepsilon/2$$

and so

$$|F(v) - F(c)| + |F(c) - F(u)| \leq$$
$$|F(v) - F(c+)| + |F(c+) - F(c)| + |F(c-) - F(c)| + |F(c-) - F(u)|$$
$$\leq |F(c+) - F(c-)| + \varepsilon/2 \leq |F(v) - F(u)| + \varepsilon.$$

Thus

$$|F(v) - F(u)| \geq |F(v) - F(c)| + |F(c) - F(u)| - \varepsilon.$$

The other inequality

$$|F(v) - F(c)| + |F(c) - F(u)| \geq |F(v) - F(u)|$$

is obviously true. ∎

5.3 Variation expressed as an integral

We begin by pointing out the obvious relation between the Jordan variation and a certain Stieljtes integral.

Lemma 5.8 *Suppose that $F : [a,b] \to \mathbb{R}$. Then*

$$\mathrm{Var}(F, [a,b]) = \overline{\int_a^b} |dF(x)|.$$

Our interest is in the special case where this integral exists and we are not forced to use the upper integral.

Lemma 5.9 *Suppose that $F : [a,b] \to \mathbb{R}$ is a function of bounded variation that is naturally regulated. Then*

$$\mathrm{Var}(F,[a,b]) = \int_a^b |dF(x)|.$$

Proof. It is clear that

$$\mathrm{Var}(F,[a,b]) \geq \overline{\int_a^b} |dF(x)|.$$

In fact these are equal for all functions, but we do not need that. Let $\varepsilon > 0$ and select points

$$a = s_0 < s_1 < \cdots < s_{n-1} < s_n = b$$

so that

$$\sum_{i=1}^{n} |F(s_i) - F(s_{i-1})| > \mathrm{Var}(F,[a,b]) - \varepsilon.$$

Define a covering relation β to include only those pairs $([u,v],w)$ for which either $w \neq s_1, s_2, \dots, s_{n-1}$ and $[u,v]$ contains no point $s_1, s_2, \dots, s_{n-1}$, or else $w = s_i$ for some $i = 1,2,\dots,n-1$ and

$$|F(v) - F(u)| \geq |F(v) - F(s_i)| + |F(s_i) - F(u)| - \varepsilon/n. \tag{5.3}$$

It is clear that β is full at every point w. For points $w \neq s_1, s_2, \dots, s_{n-1}$ this is transparent, while for points $w = s_i$ for some $i = 1,2,\dots,n-1$, Lemma 5.7 may be applied.

We use a standard endpointed argument. Take any partition π of $[a,b]$ chosen from β. Scan through π looking for any elements of the form $([u,v],s_i)$ for $u < s_i < w$ and $i = 1,2,\dots,n-1$. Replace each one by the new elements $([u,s_i],s_i)$ and $([s_i,v],s_i)$. Call the new partition π'. Because of (5.3) we see that

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u)| \geq \sum_{([u,v],w)\in\pi'} |F(v) - F(u)| - \varepsilon.$$

Write $\pi_i = \pi'([s_{i-1},s_i])$ and note that, by the way we have arranged π', each π_i is a partition of the interval $[s_{i-1},s_i]$. Consequently

$$\begin{aligned}\sum_{([u,v],w)\in\pi} |F(v) - F(u)| &\geq \sum_{([u,v],w)\in\pi'} |F(v) - F(u)| - \varepsilon \\ &\geq \sum_{i=1}^{n} \sum_{([u,v],w)\in\pi_i} |F(v) - F(u)| - \varepsilon \\ &\geq \sum_{i=1}^{n} |F(s_i) - F(s_{i-1})| - \varepsilon > \mathrm{Var}(F,[a,b]) - 2\varepsilon.\end{aligned}$$

We have shown that for *every* partition π of $[a,b]$ contained in β this sum is larger than $\mathrm{Var}(F,[a,b]) - 2\varepsilon$. It follows that

$$\underline{\int_a^b} |dF(x)| \geq \mathrm{Var}(F,[a,b]) - 2\varepsilon.$$

Since ε is arbitrary the inequality

$$\mathrm{Var}(F,[a,b]) \leq \underline{\int_a^b} |dF(x)| \leq \overline{\int_a^b} |dF(x)| \leq \mathrm{Var}(F,[a,b])$$

must hold and the theorem is proved. ∎

Corollary 5.10 *Suppose that $F : [a,b] \to \mathbb{R}$ is a function of bounded variation that is naturally regulated. Then*

$$\mathrm{Var}(F,[a,b]) = \int_a^b |dF(x)| = \int_a^t [dF(x)]^+ + \int_a^t [dF(x)]^-.$$

Proof. The proof of the lemma can easily be adjusted to prove that all three of these integrals must exist. The identity is trivial: the expression

$$|dF(x)| = [dF(x)]^+ + [dF(x)]^-$$

integrated over $[a,b]$ produces the required identity. ∎

The role of the naturally regulated assumption is exhibited in Exercise 293. It can be checked that if a function is not naturally regulated then the integral is not determined and the variation must be displayed using the upper integrals.

5.4 Representation theorems

5.4.1 Jordan decomposition

The structure of functions of bounded variation is particularly simplified by a theorem of Jordan: every function of bounded variation is merely a linear combination of monotonic functions. We prove this for functions that are naturally regulated, by interpreting the statement as an integration assertion about certain Stieltjes integrals. The statement is true in general for all functions of bounded variation, but then the upper integrals would be needed (cf. Exercise 295).

Theorem 5.11 *Let $F : [a,b] \to \mathbb{R}$ be a function of bounded variation and suppose that F is naturally regulated. Then, for all $a < t \leq b$,*

$$F(t) - F(a) = \int_a^t [dF(x)]^+ - \int_a^t [dF(x)]^-. \tag{5.4}$$

The identity (5.4) is a representation of F as a difference of two functions, both nondecreasing, both naturally regulated.

Proof. The existence of the integrals is given in Corollary 5.10. The identity is trivial: the expression

$$dF(x) = [dF(x)]^+ - [dF(x)]^-$$

integrated over $[a,b]$ produces the required identity. (If you have forgotten the meaning of $[r]^+$ and $[r]^-$ see page 226.) ∎

Corollary 5.12 *Let $F : [a,b] \to \mathbb{R}$ be a function of bounded variation and suppose that F is continuous. Then, for all $a < t \leq b$,*

$$F(t) - F(a) = \int_a^t [dF(x)]^+ - \int_a^t [dF(x)]^-. \tag{5.5}$$

The identity (5.5) is a representation of F as a difference of two functions, both continuous and nondecreasing.

5.4.2 Jordan decomposition theorem: differentiation

We know that all functions of bounded variation and all monotonic functions are almost everywhere differentiable. This and the integral representation given in Theorem 5.11 allows the following corollary.

Corollary 5.13 *Let $F : [a,b] \to \mathbb{R}$ be a function of bounded variation and suppose that F is naturally regulated. Write*

$$F_1(t) = \int_a^t [dF(x)]^+ \quad (a \leq t \leq b), \tag{5.6}$$

and

$$F_2(t) = \int_a^t [dF(x)]^- \quad (a \leq t \leq b), \tag{5.7}$$

Then

$$F(t) - F(a) = F_1(t) - F_2(t) \;\; \text{and} \;\; T(t) = \mathrm{Var}(F,[a,t]) = F_1(t) + F_2(t).$$

Moreover, at almost every t in $[a,b]$,

$$F'(t) = F_1'(t) - F_2'(t), \;\; F_1'(t) = \max\{F'(t),0\}, \;\; F_2'(t) = \max\{-F'(t),0\},$$

$$T'(t) = F_1'(t) + F_2'(t) = |F'(t)| \;\; \text{and} \;\; F_1'(t)F_2'(t) = 0.$$

Proof. There are three tools needed for the differentiation statements: the Lebesgue differentiation theorem (that monotonic functions have derivatives a.e.), the Henstock zero variation criterion for integrals, and the zero variation implies zero derivative a.e. rule.

We illustrate with a proof for one of the statements in the corollary. Define

$$h([u,v],w) = F_1(v) - F_1(u) - [F(v) - F(u)]^+.$$

The identity $F_1(t) = \int_a^t [dF(x)]^+$ requires that h have zero variation on (a,b). This, in term, requires that

$$\lim_{h\to 0+} \frac{F_1(t+h) - F_1(t) - \max\{F(t+h) - F(t),0\}}{h}$$

$$= \lim_{h\to 0+} \frac{F_1(t) - F_1(t-h) - \max\{F(t) - F(t-h), 0\}}{h} = 0$$

for almost every t in (a,b). From that we deduce that $F_1'(t) = \max\{F'(t), 0\}$ must be true for almost every t in (a,b). Proofs for the other statements are similar. ■

5.4.3 Representation by saltus functions

A nondecreasing function is said to be a saltus function if it is a "pure jump function." We can describe such functions simply this way. A function $f:\mathbb{R}\to\mathbb{R}$ is said to be a function with a single simple jump if there are numbers x, u, and v so that $f(t) = 0$ for $t < x$, $f(x) = u$ and $f(t) = u + v$ for $t > x$. A graph of such a function is easy to sketch and to visualize. By a saltus function we intend a convergent sum of a series of single simple jump functions.

We assume that sequences $\{x_n\}$, $\{u_n\}$, and $\{v_n\}$ are given with corresponding jump functions $f_n:\mathbb{R}\to\mathbb{R}$ so that $f_n(t) = 0$ for $t < x_n$, $f_n(x_n) = u_n$ and $f_n(t) = u_n + v_n$ for $t > x_n$.

Then, provided the series $\sum_{n=1}^{\infty} u_n$ and $\sum_{n=1}^{\infty} v_n$ are assumed to be absolutely convergent, the function

$$s(t) = \sum_{n=1}^{\infty} f_n(t) = \sum_{x_n \le t} u_n + \sum_{x_n < t} v_n$$

is well-defined and is considered as a *saltus function*. All of the change in the values of the function s (its growth) occurs on a discrete set of points, the points of the sequence $\{x_n\}$. We sketch out the theory of such functions, leaving most details to the reader to fill in.

Theorem 5.14 *Let $F:[a,b]\to\mathbb{R}$ be a monotonic nondecreasing function and let C be the set of points of continuity of F in $[a,b]$. Then, for all $a < t \le b$,*

$$F(t) - F(a) = \int_a^t \chi_C(x)\,dF(x) + \int_a^t [1 - \chi_C(x)]\,dF(x). \tag{5.8}$$

and

$$\int_a^t [1 - \chi_C(x)]\,dF(x) = [F(t) - F(t-)] + \sum_{s\in[a,t)\setminus C} [F(s+) - F(s-)]$$

The identity (5.8) is a representation of F as a sum of two functions, the first continuous and nondecreasing, the second a saltus function.

Exercise 298 *Suppose that*

$$s(t) = \sum_{x_n \le t} u_n + \sum_{x_n < t} v_n$$

is a representation of a saltus function s for some sequence of points $\{x_n\}$ in an interval (a,b). Show that s has bounded variation in $[a,b]$ and that

$$V(s,[a,b]) = \sum_{n=1}^{\infty}(|u_n|+|v_n|).$$

□

Exercise 299 *Suppose that*

$$s(t) = \sum_{x_n \le t} u_n + \sum_{x_n < t} v_n$$

is a representation of a saltus function s for some sequence of points $\{x_n\}$ in an interval (a,b). Show that s is continuous at each point x in $[a,b]$ that does not appear in the sequence $\{x_n\}$. □

Exercise 300 *Show that any saltus function s is singular, i.e., that $s'(t) = 0$ for almost every point t.* Answer □

Exercise 301 *Prove Theorem 5.14.* □

Exercise 302 *Assume that sequences $\{x_n\}$, $\{u_n\}$, and $\{v_n\}$ are given with the series $\sum_{n=1}^{\infty} u_n$ and $\sum_{n=1}^{\infty} v_n$ both assumed to be absolutely convergent. Let*

$$s(t) = \sum_{x_n \le t} u_n + \sum_{x_n < t} v_n$$

be a representation of a saltus function s. Show that the total variation of s is also a saltus function and that

$$T(t) = \sum_{x_n \le t} |u_n| + \sum_{x_n < t} |v_n|$$

gives this total variation function. □

5.4.4 Representation by singular functions

The decomposition of a function of bounded variation into its continuous and saltus parts can be extended further by splitting the continuous part into an absolutely continuous part plus a singular function.

Theorem 5.15 *Let $F : [a,b] \to \mathbb{R}$ be a continuous monotonic function. Let D be the set of points of differentiability of F in $[a,b]$. Then*

$$F(t) - F(a) = \int_a^t \chi_D(x)\, dF(x) + \int_a^t [1 - \chi_D(x)]\, dF(x) \tag{5.9}$$

and

$$\int_a^t \chi_D(x)\, dF(x) = \int_a^t F'(x)\, dx.$$

The identity (5.9) is a representation of F as a sum of two monotonic functions, the first absolutely continuous in the sense of Vitali and the second a continuous singular[a] function.

[a]A monotonic function $S : [a,b] \to \mathbb{R}$ would be said to be *singular* provided $S'(x) = 0$ for almost every point x in (a,b).

The proof reduces to checking that the integrals exist and then checking that the two expressions have the stated properties. Note that the latter integral exists as a Lebesgue integral.

5.5 Reducing a Stieltjes integral to an ordinary integral

The Stieltjes integral reduces to an ordinary integral in a number of interpretations. When the integrating function G is an indefinite integral the whole theory reduces to ordinary integration. The formula is compelling since, as calculus students often learn,

$$dG(x) = G'(x)\, dx$$

can be assigned a meaning. That meaning is convenient here too and suggests that

$$\int_a^b f(x)\, dG(x) = \int_a^b f(x) G'(x)\, dx.$$

Theorem 5.16 *Suppose that G, f, $g : \mathbb{R} \to \mathbb{R}$ and that g is integrable on a compact interval $[a,b]$ with an indefinite integral*

$$G(d) - G(c) = \int_c^d g(x)\, dx \quad (a \le c < d \le b).$$

Then the Stieltjes integral

$$\int_a^b f(x)\, dG(x)$$

exists if and only if fg is integrable on $[a,b]$, in which case

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

Proof. The proof depends simply on the Henstock criterion. The existence of the ordinary integral

$$\int_a^b g(x)\,dx$$

with an indefinite integral G is equivalent to the zero criterion:

$$\int_a^b |dG(x) - g(x)\,dx| = 0$$

Whenever this identity holds, then one checks that, for any function f,

$$\int_a^b |f(x)dG(x) - f(x)g(x)\,dx| = 0$$

would also be true. For example, if we have a bounded f this is trivial; for unbounded one only has to split $[a,b]$ into the sequence of sets

$$\{x \in [a,b] : n-1 \leq |f(x)| < n\}$$

and argue on each of these (cf. Exercise 304).

The existence of the Stieltjes integral

$$\int_a^b f(x)\,dG(x)$$

with an indefinite integral F is equivalent to the zero criterion:

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0.$$

Together these give

$$\int_a^b |dF(x) - f(x)g(x)\,dx| \leq$$

$$\int_a^b |dF(x) - f(x)\,dG(x)| + \int_a^b |f(x)dG(x) - f(x)g(x)\,dx| = 0.$$

From this it is easy to read off the required identity. ∎

5.6 Properties of the indefinite integral

Theorem 5.17 *Suppose that*

$$F(t) = \int_a^t f(x)\,dG(x) \quad (a \le t \le b).$$

Then

1. F *is continuous at every point at which* G *is continuous.*
2. F *is absolutely continuous in any set* $E \subset (a,b)$ *in which* G *is absolutely continuous.*
3. F *has zero variation on any set* $E \subset (a,b)$ *on which* G *has zero variation.*
4. F *has bounded variation on* $[a,b]$ *if* f *is bounded and if* G *has bounded variation.*
5. *If* G *is Vitali absolutely continuous on* $[a,b]$ *and if* f *is bounded then* F *is also Vitali absolutely continuous on* $[a,b]$.
6. *If* G *is a saltus function on* $[a,b]$ *and* f *is nonnegative then so too is the indefinite integral* F. *Moreover the jumps of* F *occur precisely at points that are jumps of* G *for which* f *does not vanish.*

Theorem 5.18 (Differentiation properties) *Suppose that*

$$F(t) = \int_a^t f(x)\,dG(x) \quad (a \le t \le b).$$

Then

1. *For almost every point* x *in* $[a,b]$
$$\lim_{y \to x} \frac{F(y) - F(x) - f(x)(G(y) - G(x))}{y - x} = 0.$$
2. *For almost every point* x *in* $[a,b]$,
$$\overline{D}F(x) = f(x)\overline{D}G(x) \text{ and } \underline{D}F(x) = f(x)\underline{D}G(x)$$
or else
$$\overline{D}F(x) = f(x)\underline{D}G(x) \text{ and } \underline{D}F(x) = f(x)\overline{D}G(x)$$
depending on whether $f(x) \ge 0$ *or* $f(x) \le 0$.
3. *In particular,* $F'(x) = f(x)G'(x)$ *at almost every point* x *at which either* F *or* G *is differentiable.*
4. *Finally,* $F'(x) = 0$ *at almost every point* x *where* $f(x) = 0$.

The proof for each of these statements depends simply on the Henstock criterion. The existence of the Stieltjes integral

$$\int_a^b fx)\,dG(x)$$

with an indefinite integral F is equivalent to the zero criterion:

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0$$

From the latter will flow each of the statements of the theorem. The individual proofs are left in the Exercises to the reader.

Exercises

Exercise 303 *Suppose that*

$$\int_a^b |dF(x) - f(x)\,dx| = 0.$$

Show that if g is any bounded function on $[a,b]$ then

$$\int_a^b |g(x)dF(x) - f(x)g(x)\,dx| = 0.$$

□

Exercise 304 *Suppose that*

$$\int_a^b |dF(x) - f(x)\,dx| = 0.$$

Show that if g is any real-valued function on $[a,b]$ then

$$\int_a^b |g(x)dF(x) - f(x)g(x)\,dx| = 0.$$

□

Exercise 305 *Suppose that*

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0.$$

Show that F is continuous at any point at which G is continuous. Is the converse necessarily true?

□

Exercise 306 *Suppose that*

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0.$$

Show that F has zero variation on any set on which G has zero variation. Is the converse necessarily true?

□

Exercise 307 *Suppose that*

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0$$

and suppose that G *has bounded variation on* $[a,b]$ *and that* f *is bounded. Show that* F *has bounded variation on* $[a,b]$.

□

Exercise 308 *Suppose that*

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0.$$

Show that

$$\lim_{y\to x} \frac{F(y) - F(x) - f(x)(G(y) - G(x))}{y - x} = 0$$

almost everywhere by using the zero variation implies zero derivative criterion.

□

Exercise 309 *Complete the remaining arguments needed to establish the parts of the theorem.*

□

Exercise 310 *Suppose that*

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0.$$

Show that, for every point x *in* $[a,b]$

$$\lim_{y\to x} \frac{F(y) - F(x)}{G(y) - G(x)} = f(x)$$

except perhaps for points x *in a set* N *in which* G *has* fine *variation zero.*

□

Exercise 311 *Suppose that at every point* x *of a compact interval* $[a,b]$

$$\lim_{y\to x} \frac{F(y) - F(x) - f(x)[G(y) - G(x)]}{y - x} = 0.$$

Show that

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0.$$

□

Exercise 312 *Suppose that at every point* x *of a compact interval* $[a,b]$

$$\lim_{y\to x} \frac{F(y) - F(x) - f(x)[G(y) - G(x)]}{y - x} = 0$$

except for points x *in a set* N *for which both* F *and* G *have zero variation. Show that*

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0.$$

□

Exercise 313 *Suppose that*

$$\int_a^b |dF(x) - f(x)\,dG(x)| = 0.$$

Show that, at almost every point x,

$$\overline{D}F(x) = f(x)\overline{D}G(x) \text{ and } \underline{D}F(x) = f(x)\underline{D}G(x)$$

if $f(x) \geq 0$ while

$$\overline{D}F(x) = f(x)\underline{D}G(x) \text{ and } \underline{D}F(x) = f(x)\overline{D}G(x)$$

if $f(x) \leq 0$. In particular $F'(x) = 0$ at almost every point x where $f(x) = 0$. □

5.6.1 Existence of the integral from derivative statements

The existence of the integral

$$\int_a^b f(x)\,dG(x)$$

can be deduced from a variety of differentiation statements. For example, using Exercise 312, we can prove the following simple version:

Theorem 5.19 *Suppose that at every point x of a compact interval $[a,b]$*

$$\lim_{y\to x} \frac{F(y) - F(x) - f(x)[G(y) - G(x)]}{y - x} = 0$$

except for points x in a set N for which both F and G have zero variation. Then the Stieltjes integral exists and

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

5.6.2 Existence of the integral for continuous functions

Theorem 5.20 *Let $f, G : \mathbb{R} \to \mathbb{R}$ and suppose that f is continuous on a compact interval $[a,b]$ and that G is monotonic nondecreasing throughout that interval. Then the Stieltjes integral exists*[a] *and*

$$\left| \int_a^b f(x)\,dG(x) \right| \leq \|f\|_\infty [G(b) - G(a)].$$

where $\|f\|_\infty = \max_{t\in[a,b]} |f(t)|$.

[a] This integral exists also in the Riemann-Stieltjes sense.

Proof. The inequality is easy since, for any pair $([u,v],w)$ with $[u,v] \subset [a,b]$,

$$|f(w)(G(v) - G(u)| \leq \|f\|_\infty [G(v) - G(u)]. \tag{5.10}$$

To prove that the integral exists we can invoke a version of the McShane criterion here. The details are left as an exercise. ∎

The next theorem is similar.

Theorem 5.21 *Let $f, G : \mathbb{R} \to \mathbb{R}$ and suppose that f is continuous on a compact interval $[a,b]$ and that G has bounded variation throughout that interval. Then the Stieltjes integral exists*[a] *and*

$$\left| \int_a^b f(x)\, dG(x) \right| \leq \|f\|_\infty \operatorname{Var}(G,[a,b]).$$

where $\|f\|_\infty = \max_{t \in [a,b]} |f(t)|$.

[a]This integral exists also in the Riemann-Stieltjes sense.

5.7 Integration by parts

Integration by parts for the Stieltjes integral assumes the following form[1]:

Theorem 5.22 *Let $F, G : \mathbb{R} \to \mathbb{R}$. Then*

$$\int_a^b [F(x)\, dG(x) + G(x)\, dF(x)] = F(b)G(b) - F(a)G(a) - \int_a^b dF(x)\, dG(x)$$

in the sense that if one of the integrals exists, so too does the other with the stated identity.

Proof. First check a simple identity: that, for any u and v,

$$F(u)[G(v) - G(u)] + G(u)[F(v) - F(u)]$$
$$= F(v)G(v) - G(u)G(u) - [F(v) - F(u)][G(v) - G(u).$$

This suggests that

$$\int_a^b |F(x)\, dG(x) + G(x)\, dF(x) - dF(x)\, dG(x) - dF(x)\, dG(x)| = 0 \qquad (5.11)$$

is simply true because of an identity. If indeed this is true then the statement in the theorem is obvious because

$$\int_a^b dF(x) dG(x) = F(b)G(b) - F(a)G(a).$$

To complete the proof we have to address just one concern here. If a partition π of the interval $[a,b]$ contains only pairs $([u,v],u)$ or $([u,v],v)$ [i.e., $([u,w],w)$ with w only at an endpoint] then our simple identity would indeed

[1]For the *Riemann*-Stieltjes integral the extra term $\int_a^b dF(x)\, dG(x)$ does not appear, since this would be zero whenever the integral exists in that sense. (See Corollary 5.24, which should look familiar to fans of the Riemann-Stieltjes integral.)

supply

$$\sum_{([u,v],w)\in\pi} [F(w)[G(v)-G(u)]+G(w)[F(v)-F(u)]-F(v)G(v)-G(u)G(u)]$$

$$= \sum_{([u,v],w)\in\pi} [F(v)-F(u)][G(v)-G(u)].$$

That surely proves (5.11) if we are allowed to use only such partitions. But what happens if we permit (as we must) partitions π containing a pair $([u,v],w)\in\pi$ for which $u<w<v$?

To clear this up note that we can always adjust full covers and partitions π by replacing any pair $([u,v],w)\in\pi$ for which $u<w<v$ by the two items $([u,w],w)$ and $([w,v],w)$. That does not change the sums here because, for example,

$$F(w)[G(v)-G(u)] = F(w)[G(w)-G(u)]+F(w)[G(v)-G(w)].$$

This "endpointed" argument (which we have seen before in Exercise 187) means that in these simple Stieltjes integrals the partitions used can all be restricted to ones where only elements of the form $([u,v],u)$ or $([u,v],v)$ can appear. ■

Corollary 5.23 *Let $F, G : \mathbb{R}\to\mathbb{R}$ and suppose that*

$$\int_a^b |dF(x)\,dG(x)| = 0.$$

Then

$$\int_a^b [F(x)\,dG(x)+G(x)\,dF(x)] = F(b)G(b)-F(a)G(a).$$

If, in addition one of the following two integrals exists then so too does the other and

$$\int_a^b F(x)\,dG(x)+\int_a^b G(x)\,dF(x) = F(b)G(b)-F(a)G(a).$$

Corollary 5.24 *Let $F, G : \mathbb{R}\to\mathbb{R}$ and suppose that F is continuous and G has bounded variation. Then*

$$\int_a^b F(x)\,dG(x)+\int_a^b G(x)\,dF(x) = F(b)G(b)-F(a)G(a).$$

Proof. The assumption that F is continuous and G has bounded variation requires that

$$\int_a^b |dF(x)\,dG(x)| = 0.$$

Thus Theorem 5.22 can be applied. But we know, from Theorem 5.20, that the integral $\int_a^b F(x)dG(x)$ must exist. It follows, from Corollary 5.23, that $\int_a^b G(x)dF(x)$ must also exist and that the integration by parts formula is valid. ■

5.8 Lebesgue-Stieltjes measure

The variation of a function F on an interval $[a,b]$ can be described by the identity

$$\mathrm{Var}(F,[a,b]) = \sup \sum_{([u,v],w)\in\pi} |F(v)-F(u)|$$

where the supremum is taken over all possible partitions π of the interval $[a,b]$. We recall that a somewhat similar expression describes the Lebesgue measure $\lambda(E)$ of a set E:

$$\lambda(E) = \inf_{\beta} \sup_{\pi\subset\beta} \sum_{([u,v],w)\in\pi} (v-u).$$

Here π denotes an arbitrary subpartition contained in β and the infimum is taken over all full covers β of the set E. There is an obvious generalization of Lebesgue measure available by replacing $(v-u)$ by $|F(v)-F(u)|$.

Definition 5.25 *Let F be a function defined at least on an open set G and we suppose that $E \subset G$. Then we write*

$$\lambda_F(E) = \inf_{\beta} \sup_{\pi\subset\beta} \sum_{([u,v],w)\in\pi} |F(v)-F(u)|.$$

Here π denotes an arbitrary subpartition contained in β. The set function λ_F defined for all subsets of G is called the Lebesgue-Stieltjes measure *associated with F or, often, the* variational measure *associated with F.*

In the literature often the Lebesgue-Stieltjes measure is studied only for monotonic functions that are continuous on the left-hand side at every point. It is convenient for us to usurp this language for the completely general case. The definition of the Lebesgue-Stieltjes measure is closely related to the Stieltjes integral, just as the definition of Lebesgue measure in Lemma 4.2 was expressible as an upper integral.

Lemma 5.26 *If F is defined on a compact interval $[a,b]$ and $E \subset (a,b)$ then*

$$\lambda_F(E) = \overline{\int_a^b} \chi_E(x)\,|dF(x)|.$$

By comparing this definition with some earlier notions that are almost identical we will be able to deduce the following properties of this measure:

Properties of the Lebesgue-Stieltjes measures

1. λ_F is a measure, i.e., if F is defined on an open set G and E, E_1, E_2, E_3, ... are subsets of G for which $E \subset \bigcup_{n=1}^{\infty} E_n$ then this inequality must hold:

$$\lambda_F(E) \le \sum_{n=1}^{\infty} \lambda_F(E_n).$$

2. If F is monotonic then
$$\lambda_F([a,b]) = |F(b+) - F(a-)|,$$
$$\lambda_F((a,b)) = |F(b-) - F(a+)|,$$
and
$$\lambda_F(\{x_0\}) = |F(x_0+) - F(x_0-)|.$$

3. F has zero variation on a set E if and only if $\lambda_F(E) = 0$.

4. F is continuous at a point x_0 if and only if $\lambda_F(\{x_0\}) = 0$.

5. F is continuous at every point of an open interval (a,b) if and only if $\lambda_F(C) = 0$ for every countable subset of (a,b).

6. F is absolutely continuous on an interval (a,b) if and only if $\lambda_F(N) = 0$ for every subset N of (a,b) that has measure zero.

7. $\lambda_F((a,b)) = 0$ if and only if F is constant on (a,b).

8. F is locally bounded at a point x_0 if and only if $\lambda_F(\{x_0\}) < \infty$.

9. If F is defined on a compact interval $[a,b]$ then F has bounded variation on $[a,b]$ if and only if $\lambda_F((a,b)) < \infty$.

10. If F is defined on an open set G and has a bounded derivative at each point of a bounded subset E of G then $\lambda_F(E) < \infty$.

11. If F is defined on an open set G and $\lambda_F(E) < \infty$ then F is differentiable at almost every point of E.

It is clear from the definitions that F has zero variation on a set E if and only if $\lambda_F(E) = 0$. Thus the assertions (4)–(8) are immediate from our early study of zero variation. The other assertions are proved in the exercises.

Exercises

Exercise 314 *Prove that λ_F is a measure.* Answer □

Exercise 315 *Show that if F is monotonic then F is monotonic then*
$$\lambda_F([a,b]) = |F(b+) - F(a-)|,$$
$$\lambda_F((a,b)) = |F(b-) - F(a+)|,$$
and
$$\lambda_F(\{x_0\}) = |F(x_0+) - F(x_0-)|.$$

□

Exercise 316 *Show that, if the one-sided limits $F(x_0+)$ and $F(x_0-)$ exist then*
$$\lambda_F(\{x_0\}) = |F(x_0+) - F(x_0)| + |F(x_0-) - F(x_0)|.$$
□

Exercise 317 *Suppose that F is defined on an open set G. Show that F is locally bounded at a point $x_0 \in G$ if and only if $\lambda_F(\{x_0\}) < \infty$.* □

Exercise 318 *Suppose that F is defined on a compact interval $[a,b]$. Show that F has bounded variation on $[a,b]$ if and only if $\lambda_F((a,b)) < \infty$. Show that $\lambda_F((a,b)) \leq \mathrm{Var}(F,[a,b])$ but that the inequality may be strict unless F is continuous.*

Answer □

Exercise 319 *Suppose that F is defined on an open set G and has a bounded derivative at each point of a bounded subset E of G. Show that $\lambda_F(E) < \infty$.*

Answer □

5We recall that every function of bounded variation is

5.9 Mutually singular functions

Definition 5.27 *Let F, $G : [a,b] \to \mathbb{R}$ be functions of bounded variation. Then F and G are said to be* mutually singular *provided that*
$$\int_a^b \sqrt{|dF(x)\,dG(x)|} = 0.$$

Lemma 5.28 *Let F, $G : [a,b] \to \mathbb{R}$ be functions of bounded variation. If F and G are mutually singular, then $F'(x)G'(x) = 0$ almost everywhere in $[a,b]$.*

Proof. This follows easily (as usual) from the zero variation implies zero derivative a.e. rule together with the fact that both $F'(x)$ and $G'(x)$ must exist a.e.. ■

Our main theorem shows that mutually singular functions grow on separate parts of the interval $[a,b]$ in a sense made precise here.

Theorem 5.29 *Let F, $G : [a,b] \to \mathbb{R}$ be functions of bounded variation. Then F and G are mutually singular on $[a,b]$ if and only for every $\varepsilon > 0$ there is a full cover β with the property that every partition π of $[a,b]$ contained in β can be split into two disjoint subpartitions $\pi = \pi' \cup \pi''$ so that*
$$\sum_{([u,v],w)\in\pi'} |F(v) - F(u)| < \varepsilon$$
and
$$\sum_{([u,v],w)\in\pi''} |G(v) - G(u)| < \varepsilon.$$

Proof. Suppose that
$$\int_a^b \sqrt{|dF(x)dG(x)|} = 0.$$
Let $\varepsilon > 0$ and select a full cover β so that
$$\sum_{([u,v],w)\in\pi} \sqrt{|[F(v)-F(u)][G(v)-G(u)]|} < \varepsilon$$
for all partitions π of $[a,b]$ contained in β. Split such a π as follows:
$$\pi' = \{([u,v],w) : |[F(v)-F(u)]| \leq |[G(v)-G(u)]|\}$$
and
$$\pi'' = \{([u,v],w) : |[F(v)-F(u)]| > |[G(v)-G(u)]|\}.$$
Verify that $\pi = \pi' \cup \pi''$ and that
$$\sum_{([u,v],w)\in\pi'} |[F(v)-F(u)]| \leq \sum_{([u,v],w)\in\pi'} \sqrt{|[F(v)-F(u)][G(v)-G(u)]|} < \varepsilon$$
and that
$$\sum_{([u,v],w)\in\pi''} |[G(v)-G(u)]| \leq \sum_{([u,v],w)\in\pi''} \sqrt{|[F(v)-F(u)][G(v)-G(u)]|} < \varepsilon.$$
This proves one direction in the theorem.

For the converse select a number $M > 0$ and a full cover β_1 so that
$$\sum_{([u,v],w)\in\pi} [|[F(v)-F(u)]| + |[G(v)-G(u)]|] < M$$
for all partitions π of $[a,b]$ from β_1. This is possible merely because the functions F and G have bounded variation. Select a full cover β_2 with the property presented in the statement of the theorem (for ε). Let $\beta = \beta_1 \cap \beta_2$. This is a full cover. Consider any partition π of $[a,b]$ contained in β. There must be, by hypothesis, a split $\pi = \pi' \cup \pi''$ so that
$$\sum_{([u,v],w)\in\pi'} |[F(v)-F(u)]| < \varepsilon$$
and
$$\sum_{([u,v],w)\in\pi''} |[G(v)-G(u)]| < \varepsilon.$$
We now compute
$$\sum_{([u,v],w)\in\pi} \sqrt{|[F(v)-F(u)][G(v)-G(u)]|} =$$
$$\sum_{([u,v],w)\in\pi'} \sqrt{|[F(v)-F(u)][G(v)-G(u)]|}$$
$$+ \sum_{([u,v],w)\in\pi''} \sqrt{|[F(v)-F(u)][G(v)-G(u)]|}$$

$$\leq \sqrt{\sum_{([u,v],w)\in\pi'} |[F(v)-F(u)]|}\sqrt{\sum_{([u,v],w)\in\pi'} |[G(v)-G(u)]|}$$
$$+\sqrt{\sum_{([u,v],w)\in\pi''} |[F(v)-F(u)]|}\sqrt{\sum_{([u,v],w)\in\pi''} |[G(v)-G(u)]|}$$
$$\leq 2\sqrt{M\varepsilon}.$$

Here we have used the Cauchy-Schwartz inequality. Since ε is an arbitrary positive number it follows that

$$\int_a^b \sqrt{|dF(x)\,dG(x)|} = 0.$$

Consequently F and G must be mutually singular. ■

5.10 Singular functions

The reader will doubtless have encountered elsewhere the notion of a singular function and been given the usual remarkable example of such a function, the Cantor function (Devil's staircase). The next theorem states that there are further characterizations of this notion, in particular one given exactly by a Stieltjes-type integral.

Theorem 5.30 *Let $F : [a,b] \to \mathbb{R}$ be a function of bounded variation. Then the following are equivalent:*

1. *For every $\varepsilon > 0$ there is a full cover β with the property that every partition π of $[a,b]$ contained in β can be split into two disjoint subpartitions $\pi = \pi' \cup \pi''$ so that*
$$\sum_{([u,v],w)\in\pi'} |F(v)-F(u)| + \sum_{([u,v],w)\in\pi''} (v-u) < \varepsilon.$$
2. *$F'(x) = 0$ almost everywhere in $[a,b]$.*
3. $\displaystyle\int_a^b \sqrt{|dF(x)|\,dx} = 0.$

Proof. In view of the preceding section, it is enough here to show that the second statement implies the third. Suppose that $F'(x) = 0$ almost everywhere. Let $\varepsilon > 0$ and choose a sequence of open intervals $\{(c_i,d_i)\}$ with total length smaller than ε so that $F'(x) = 0$ for all $x \in [a,b]$ not in one of the intervals. Define two covering relations. The first β_1 consists of all pairs $([u,v],w)$ subject only to the condition that if w is in $[a,b]$ and not covered by an open interval $\{(c_i,d_i)\}$ then

$$|F(v)-F(u)| < \varepsilon(v-u)/(b-a).$$

The second β_2 consists of all pairs $([u,v],w)$ subject only to the condition that if w is contained in one of the open intervals $\{(c_i,d_i)\}$ then so too is $[u,v]$. Then β_1, β_2, and $\beta = \beta_1 \cap \beta_2$ are all full covers.

Note that if π is a subpartition contained in β_1 consisting of pairs $([u,v],w)$ not covered by an open interval from $\{(c_i,d_i)\}$ then

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| \leq \sum_{([u,v],w)\in\pi} \varepsilon(v-u)/(b-a) \leq \varepsilon.$$

Note that if π is a subpartition contained in β_2 consisting of pairs $([u,v],w)$ that are covered by an open interval from $\{(c_i,d_i)\}$ then

$$\sum_{(I,x)\in\pi} (v-u) \leq \sum_{i=1}^{\infty}(d_i-c_i) < \varepsilon.$$

Thus any partition of $[a,b]$ chosen from β can be split into two subpartitions with these inequalities. This verifies the conditions asserted in Theorem 5.29 for F and the function $G(x) = x$. But that is exactly our third condition in the statement of the theorem. ∎

5.11 Length of curves

A curve is a pair of continuous functions $F, G : [a,b] \to \mathbb{R}$. We consider that the curve is the pair of functions itself, rather than that the curve is the geometric set of points

$$\{(F(t),G(t)) : t \in [a,b]\}$$

that is the object we might likely think about when contemplating a curve.

Definition 5.31 *Suppose that $F, G : [a,b] \to \mathbb{R}$ is a pair of continuous functions. By the* length *of the curve given by the pair F and G we shall mean*

$$\int_a^b \sqrt{[dF(x)]^2 + [dG(x)]^2}.$$

That this integral is determined (but may be infinite) is pointed out in the proof of the next theorem.

Theorem 5.32 *A curve given by a pair of continuous functions $F, G : [a,b] \to \mathbb{R}$ has finite length if and only if both functions F and G have bounded variation.*

Proof. Note that as F and G are continuous, then so too is the interval function

$$h([u,v]) = \sqrt{[F(v)-F(u)]^2 + [G(v)-G(u)]^2}.$$

A simple application of the Pythagorean theorem will verify that the function h here is a continuous, subadditive interval function. The existence of the integral can be established by a repetition of the argument of Lemma 5.9.

Thus the integral

$$\int_a^b \sqrt{[dF(x)]^2+[dG(x)]^2}$$

in the definition must necessarily be determined, although it might have an infinite value. It will have a finite value if and only if h has bounded variation. That follows from a simple computation:

$$\max\left\{\int_a^b |dF(x)|, \int_a^b |dG(x)|\right\} \le \int_a^b \sqrt{[dF(x)]^2+[dG(x)]^2}$$

and

$$\int_a^b \sqrt{[dF(x)]^2+[dG(x)]^2} \le \int_a^b |dF(x)| + \int_a^b |dG(x)|.$$

■

5.11.1 Formula for the length of curves

In the elementary (computational) calculus one usually assumes that a curve is given by a pair of continuously differentiable functions (i.e., a pair F, G of continuous functions for which F' and G' are also continuous). In that case the familiar formula for length used in elementary applications is

$$\int_a^b \sqrt{[F'(x)]^2+[G'(x)]^2}\,dx.$$

We study this now. Note that the formula is rather compelling if we think that $dF(x) = F'(x)\,dx$ and $dG(x) = G'(x)\,dx$ would be possible here.

Lemma 5.33 *For any pair of continuous functions F, $G : [a,b] \to \mathbb{R}$ of bounded variation on $[a,b]$ define the following function*

$$L(t) = \int_a^t \sqrt{[dF(x)]^2+[dG(x)]^2} \qquad (a < t \le b).$$

Then

$$L'(t) = \sqrt{[F'(t)]^2+[G'(t)]^2}$$

almost everywhere in $[a,b]$.

Proof. We are now quite familiar with the zero variation implies zero derivative a.e. rule. This is all that is needed here to establish this fact, since the statement in the Lemma can be expressed, by the Henstock zero variation criterion, as

$$\int_a^b \left| dL(x) - \sqrt{[dF(x)]^2+[dG(x)]^2} \right| = 0.$$

. ■

Lemma 5.34 *The function L in the lemma is Vitali absolutely continuous if and only if both F and G are Vitali absolutely continuous.*

Proof. This follows easily from the inequalities of Lemma 5.32. ∎

The length of the curve is now available as a familiar formula precisely in the case where the two functions defining the curve are absolutely continuous.

Lemma 5.35 *For any pair of continuous functions $F, G : [a,b] \to \mathbb{R}$ of bounded variation on $[a,b]$,*

$$\int_a^b \sqrt{[dF(x)]^2 + [dG(x)]^2} \geq \int_a^b \sqrt{[F'(x)]^2 + [G'(x)]^2}\, dx.$$

The two expressions are equal if and only if both F and G are Vitali absolutely continuous on $[a,b]$.

Proof. Using the function L introduced above we see that this assertion is easily deduced from the fact that

$$L(t) \geq \int_a^t L'(x)\, dx$$

with equality precisely when L is Vitali absolutely continuous. ∎

Exercises

Exercise 320 *For any continuous function $F : [a,b] \to \mathbb{R}$ define the length of the graph of F to mean*

$$\int_a^b \sqrt{[dx]^2 + [dF(x)]^2}.$$

Show that the graph has finite length if and only if F has bounded variation. Discuss the availability of the familiar formula for length used in elementary applications:

$$\int_a^b \sqrt{1 + [F'(x)]^2}\, dx.$$

□

Exercise 321 *Let $F, G : [a,b] \to \mathbb{R}$ where $[a,b]$ is a compact interval. Suppose that the* Hellinger integral[2]

$$H(t) = \int_a^t \frac{dF(x)\, dG(x)}{dx} \quad (a < t \leq b)$$

exists. Show that $H'(t) = F'(t)G'(t)$ at almost every point t in $[a,b]$ at which both F and G are differentiable. Answer □

Exercise 322 (Reduction theorem) *Let $F, G : [a,b] \to \mathbb{R}$ where $[a,b]$ is a compact interval. Suppose that F is Vitali absolutely continuous on $[a,b]$ and that G is a Lipschitz function. Show that*

$$\int_a^t \frac{dF(x)\, dG(x)}{dx} = \int_a^b F'(x) dG(x) = \int_a^b F'(x)G'(x)\, dx.$$

[2] Named after Ernst Hellinger (1883–1950).

□

Exercise 323 *Let* $F, G : [a,b] \to \mathbb{R}$ *where* $[a,b]$ *is a compact interval. Suppose that* F *is Vitali absolutely continuous on* $[a,b]$ *and that* G *is the indefinite integral of a function of bounded variation. Show that*

$$\int_a^t \frac{dF(x)\,dG(x)}{dx} = \int_a^b F'(x)dG(x) = \int_a^b F'(x)G'(x)\,dx.$$

□

5.12 Change of variables

We include now an application of the Stieltjes integral to a discussion of change of variables formulas for integrals on the real line, giving some particularly easy and transparent versions as well as a deeper one. The traditional change of variables formula is

$$F(G(b)) - F(G(a)) = \int_{G(a)}^{G(b)} f(s)\,ds = \int_a^b f(G(t))dG(t) = \int_a^b f(G(t))g(t)\,dt$$

where

$$F(x) = \int_{G(a)}^{x} f(s)\,ds \text{ and } G(t) = \int_a^t g(u)\,du.$$

Certainly some assumptions are needed in order for the integrals to exist and also for the identities to hold. We address this now. We recall that the part of this identity that asserts that

$$\int_a^b f(G(t))dG(t) = \int_a^b f(G(t))g(t)\,dt$$

is available trivially. We state it explicitly here.

Lemma 5.36 *Let* g *be an integrable function with an indefinite integral* G *on an interval* $[a,b]$ *and suppose that* f *is a real-valued function on* $G([a,b])$. *Then*

$$\int_a^b f(G(x))dG(x) = \int_a^b f(G(t))g(t)\,dt$$

where, if one of the integrals exists, so too does the other and the stated identity is valid.

5.12.1 Easy change of variables

We now apply Robbins's theorem (Theorem 1.9) to give what may be the simplest nontrivial version of a change of variables formula. Robbins gave no applications of his adjusted Riemann sums result in [72], contenting himself with a brief statement and proof amounting to little more than a single page. His paper alludes, however, to this idea occurring during an investigation of change of variables formulas. Thus, no doubt, the theorem which follows uses the simple method he had in mind.

Lemma 5.37 *Let G be a continuous function of bounded variation on an interval $[a,b]$ and suppose that f is continuous on $G([a,b])$. Then*

$$\int_{G(a)}^{G(b)} f(x)\,dx = \int_a^b f(G(t))dG(t)$$

where the integrals exist[a]*.*

[a]In fact the two integrals must exist in the stronger Riemann and Riemann-Stieltjes senses respectively.

Proof. Let $\varepsilon > 0$ and define $C = \mathrm{Var}(G,[a,b])$. Since f is continuous we may apply Robbins's theorem (Theorem 1.9) to find a $\delta > 0$ so that

$$\left| \int_{G(a)}^{G(b)} f(x)\,dx - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon$$

for any choice of points $x_0, x_1, \ldots, x_n$ and $\xi_1, \xi_2, \ldots, \xi_n$ from the interval $G([a,b])$ satisfying

$$\sum_{i=1}^{n} |x_i - x_{i-1}| \le C$$

where $G(a) = x_0$, $G(b) = x_n$, $0 < |x_i - x_{i-1}| < \delta$ and each ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \ldots, n$.

Choose $\delta_1 > 0$ so that

$$|G(s) - G(t)| < \delta$$

if s and t are points in $[a,b]$ with $|s-t| < \delta_1$. Choose any points $a = t_0 < t_1 < \cdots < t_n = b$ and $t_{i-1} \le \tau_i \le t_i$ for which $0 < t_i - t_{i-1} < \delta_1$ and consider the Stieltjes sum

$$\sum_{i=1}^{n} f(G(\tau_i))[G(t_i) - G(t_{i-1})].$$

Consider the points $x_i = G(t_i)$, $\xi_i = G(\tau_i)$. Note that $x_0 = G(a)$, $x_n = G(b)$, $|x_i - x_{i-1}| < \delta$ and

$$\sum_{i=1}^{n} |x_i - x_{i-1}| = \sum_{i=1}^{n} |G(t_i) - G(t_{i-1})| \le \mathrm{Var}(G,[a,b]) = C.$$

In order to use what we now have easily we would need to know that ξ_i is between the points x_{i-1} and x_i, i.e., that $G(\tau_i)$ is between $G(t_{i-1})$ and $G(t_i)$. This may not be the case. Should one of these fail we return to our original Stieltjes sum and replace the offending term by using

$$f(G(\tau_i))[G(t_i) - G(t_{i-1})] = f(G(\tau_i))[G(\tau_i) - G(t_{i-1})] + f(G(\tau_i))[G(t_i) - G(\tau_i)].$$

Having prepared our sum in this way we can then proceed as described and claim, that in each case, ξ_i is between the points x_{i-1} and x_i.

Consequently, using the steps above, we have

$$\left|\int_{G(a)}^{G(b)} f(t)\,dt - \sum_{i=1}^{n} f(G(\tau_i)[G(t_i) - G(t_{i-1})]\right|$$
$$= \left|\int_{G(a)}^{G(b)} f(t)\,dt - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})\right| < \varepsilon.$$

This proves the existence of the Stieltjes integral and establishes the formula. ∎

5.12.2 Another easy change of variables

Exactly the same method used in Lemma 5.37 gives another elementary version of a change of variables formula; replace an appeal to Robbins's theorem (Theorem 1.9) by an appeal to Theorem **??**. (This time the integral cannot be interpreted in the stronger Riemann sense.)

Theorem 5.38 *Let G be a continuous function of bounded variation on an interval $[a,b]$ and suppose that F is differentiable on $G([a,b])$. Then*

$$F(G(b)) - F(G(a)) = \int_{G(a)}^{G(b)} F'(x)\,dx = \int_a^b F'(G(t))dG(t).$$

5.12.3 A general change of variables

Finally we give a more formal version that continues the theme and uses a recognizably similar argument.

Theorem 5.39 *Let g and G be functions defined on an interval $[a,b]$ for which $G'(t) = g(t)$ for a.e. point t in $[a,b]$ and suppose that f and F are functions defined on an interval $[c,d]$ that includes $G([a,b])$ for which $F'(x) = f(x)$ for a.e. point x in $[c,d]$. Then*

$$F(G(b)) - F(G(a)) = \int_a^b f(G(t))g(t)\,dt \tag{5.12}$$

where

1. *the identity (5.12) holds if and only if the composition $F \circ G$ is absolutely continuous in the variational sense*[a] *on $[a,b]$.*

2. *the identity (5.12) holds in the sense of the Lebesgue integral if and only if the composition $F \circ G$ is absolutely continuous in Vitali's sense on $[a,b]$.*

[a] i.e., is ACG$_*$ in classical terminology from the 1930s.

Proof. Recall that a continuous function $H : [a,b] \to \mathbb{R}$ is absolutely continuous in the variational sense if the function has zero variation on all sets of measure zero; this means that, for all $\varepsilon > 0$ and all sets $N \subset (a,b)$ of measure zero, there is a positive function δ on N so that

$$\sum_{i=1}^{k} |H(q_i) - H(p_i)| < \varepsilon$$

if $[p_1,q_1]$, $\dots [p_k,q_k]$ are nonoverlapping subintervals of $[a,b]$ satisfying, for some choice of $\tau_i \in N \cap [p_i,q_i]$, the inequalities

$$0 < q_i - p_i < \delta(\tau_i) \qquad (i = 1,2,3,\dots,k).$$

The proof uses only one idea that is not a near trivial manipulation of Riemann sums. We need to know that a function H that has a derivative $H'(t)$ at each point of a set E for which $H(E)$ is of measure zero must have $H'(t) = 0$ at a.e. point of E. See, for example, [77] who also use this fact to prove their version of this theorem. This also follows from the variational material of our Chapter 2. (See Exercise 166.).

The condition that the composition $F \circ G$ should be absolutely continuous in the variational sense is clearly necessary since all indefinite [Henstock-Kurzweil] integrals have this property. We shall show that it is also sufficient. Thus let us assume that $F \circ G$ is absolutely continuous in the variational sense on $[a,b]$.

Let $N_1 \subset [a,b]$ be the measure zero set of points $t \in [a,b]$ at which $G'(t)$ does not exist. Let M be the measure zero set of points $x \in [c,d]$ at which $F'(x) = f(x)$ fails. Let N_2 be the set of points t in $[a,b]$ at which $G'(t)$ exists, is not equal to zero and for which $G(t)$ is in M. Since M has measure zero it follows (from our remark above) that N_2 also has measure zero.

Let $g_1(t) = 0$ if t is in either of the sets N_1 or N_2 and let $g_1(t) = g(t)$ at all other values of t. Since g and g_1 agree almost everywhere it is enough to prove the theorem using g_1 instead of g.

Let $\varepsilon > 0$ and for each point t in $[a,b]$ but not in $N_1 \cup N_2$ choose $\delta(t) > 0$ so that

$$|F(G(p)) - F(G(q)) - f(G(t))g(t)(q-p)| < \frac{\varepsilon}{2(b-a)}(q-p)$$

if $t \in [p,q]$ and $0 < q-p < \delta(t)$. This just uses the fact that we can compute the derivative of $F \circ G$ at each such point.

For all points t in $N_1 \cup N_2$ choose $\delta(t) > 0$ so that

$$\sum_{i=1}^{k} |F(G(q_i)) - F(G(p_i))| < \frac{\varepsilon}{2}$$

if the intervals $[p_1,q_1], \dots [p_k,q_k]$ are nonoverlapping and satisfy

$$0 < q_i - p_i < \delta(\tau_i)$$

for some choice of $\tau_i \in (N_1 \cup N_2) \cap [p_i,q_i]$. This just uses the fact that $F \circ G$ is absolutely continuous in the variational sense on $[a,b]$ which we have assumed.

Choose any partition of $[a,b]$ that is finer than δ, i.e., take any points $a =$

$t_0 < t_1 < \cdots < t_n = b$ and $t_{i-1} \le \tau_i \le t_i$ for which $0 < t_i - t_{i-1} < \delta(\tau_i)$. We must have

$$\left| F(G(b)) - F(G(a)) - \sum_{i=1}^{n} f(G(\tau_i))g_1(\tau_i)(t_i - t_{i-1}) \right|$$

$$\le \sum_{i=1}^{n} |F(G(t_i)) - F(G(t_{i-1})) - f(G(\tau_i))g_1(\tau_i)(t_i - t_{i-1})|$$

$$\le \sum_{\tau_i \in N_1 \cup N_2} |F(G(t_i)) - F(G(t_{i-1}))|$$

$$+ \sum_{\tau_i \notin N_1 \cup N_2} |F(G(t_i)) - F(G(t_{i-1})) - f(G(\tau_i))g_1(\tau_i)(t_i - t_{i-1})|$$

$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2(b-a)} \sum_{i=1}^{n} (t_i - t_{i-1}) \le \varepsilon.$$

By definition then the identity

$$F(G(b)) - F(G(a)) = \int_a^b f(G(t))g_1(t)\,dt$$

and hence also the identity (5.12) holds [in our general sense, i.e., in the sense of the Henstock-Kurzweil integral].

For the second part of the theorem we need, also, to remember that a function is absolutely continuous in the Vitali sense if and only if it is absolutely continuous in the variational sense and has bounded variation. Since indefinite Lebesgue integrals are absolutely continuous in the Vitali sense, part two follows from part one. ∎

5.12.4 Change of variables for Lipschitz functions

Theorem 5.39 is a more general version of a theorem on change of variables given by Serrin and Varberg [77]. In a sense it appears definitive but, on inspecting the proof, it is clear that it is not deep and merely gives a formal condition for the formula. The formula itself is then essentially always true provided one can establish integrability. But, in any application, it might not be so easy or straightforward to determine properties of the composition $F \circ G$.

In general, it is possible for two function F and G to be absolutely continuous and yet the composition $F \circ G$ is not (see [73, p. 286]). When F is Lipschitz (as Corollaries 5.40 and 5.41 now illustrate) this is not a difficulty. It is easy to establish that the composition $F \circ G$ is absolutely continuous (in either sense) when F is Lipschitz and G is absolutely continuous.

Corollary 5.40 *Let g be Lebesgue integrable on $[a,b]$, let G be its indefinite integral, and suppose that F is a Lipschitz function defined on the interval $G([a,b])$. Then*

$$F(G(b)) - F(G(a)) = \int_a^b F'(G(t))g(t)\,dt$$

where the integral exists as a Lebesgue integral.

Corollary 5.41 *Let g be integrable on $[a,b]$, let G be its indefinite integral, and suppose that F is a Lipschitz function defined on the interval $G([a,b])$. Then*

$$F(G(b)) - F(G(a)) = \int_a^b F'(G(t))g(t)\,dt.$$

5.12.5 Theorem of Kestelman, Preiss, and Uher

Theorem 5.39 can also be used to clarify the situation for the Riemann integral. The complete picture is available in Kestelman [43] and Preiss and Uher [69]. We reproduce this result here.

Theorem 5.42 (Kestelman-Preiss-Uher) *Suppose that g is Riemann integrable on an interval $[a,b]$ with an indefinite integral*

$$G(t) = \int_a^t g(u)\,du \quad (a \le t \le b)$$

and that f is a bounded function on $G([a,b])$. Then the identity

$$\int_{G(a)}^{G(b)} f(s)\,ds = \int_a^b f(G(t))dG(t) = \int_a^b f(G(t))g(t)\,dt$$

holds with all integrals interpreted in the Riemann sense provided either f is Riemann integrable on $G([a,b])$, or the second integral exists as a Riemann-Stieltjes integral, or finally the function $(f\circ G)g$ is Riemann integrable on $[a,b]$.

Proof. Suppose first that f is Riemann integrable on $G([a,b])$ and that F is its indefinite integral. By Corollary 5.40 we have immediately that

$$F(G(b)) - F(G(a)) = \int_{G(a)}^{G(b)} f(s)\,ds = \int_a^b f(G(t))g(t)\,dt$$

where the function $(f\circ G)g$ must be Lebesgue integrable on $[a,b]$.

Thus it is sufficient that we prove that this function is also Riemann integrable as well. Clearly the function is bounded so it is enough to prove that it is continuous a.e. on $[a,b]$ (i.e., to use Lebesgue's criterion for Riemann integrability).

Our analysis[3] is similar to the methods in Kestelman [43]. Roy Davies [21] gave an alternative proof that directly uses Riemann's criterion for integrability.

[3]This is harder than one might think. If $f(G(t))$ is itself Riemann integrable on $[a,b]$ then certainly so too is $f(G(t))g(t)$. Kestelman [43] includes an example to show that even if $f(G(t))g(t)$

Let S_1 be the set of points in $[a,b]$ at which g is not continuous. Let N be the set of points in $G([a,b])$ at which f is not continuous. Let S_2 be the set of points t in $[a,b]\setminus S_1$ at which $G(t)\in N$ and $g(t)\neq 0$.

The function $(f\circ G)g$ is continuous at any point t that is not in $S_1\cup S_2$. The set S_1 is a set of measure zero because g is Riemann integrable. The set S_2 maps by G into the zero measure set N and G is differentiable with a nonzero derivative at each point of S_2.

Recall that we previously used (in the proof of Theorem 5.39) the fact that a function H that has a derivative $H'(t)$ at each point of a set E for which $H(E)$ is of measure zero must have $H'(t)=0$ at a.e. point of E. This implies here that S_2 must be a measure zero set. Consequently $(f\circ G)g$ is continuous a.e. in $[a,b]$ as we require.

Let us now suppose that the function $(f\circ G)g$ is Riemann integrable on $[a,b]$ and prove that f must also be Riemann integrable. Then it must follow that

$$\int_{G(a)}^{G(b)} f(s)\,ds = \int_a^b f(G(t))g(t)\,dt$$

by what we just proved. It is sufficient, then, simply to show that f is continuous a.e. in $G([a,b])$.

Our proof is similar to the analysis given in the first part of the theorem. Preiss and Uher [69] directly use the Riemann criterion for integrability. There is also a proof of this fact in Navrátil [64] where he uses the Darboux integral instead. (Both of these papers are in Czech which presents difficulties to some of us.)

Let A be the set of points in $[a,b]$ at which either g is not continuous or $(f\circ G)g$ is not continuous. This is a set of measure zero since both of these functions are Riemann integrable. It is also true that $G(A)$ is a set of measure zero in $G([a,b])$ since G is Lipschitz.

Let B be the set of points t in $[a,b]$ at which g is continuous and $g(t)=0$. This need not be a set of measure zero but $G(B)$ is a set of measure zero in $G([a,b])$ since G' vanishes on B.

Finally let C be the set of points t in $[a,b]$ at which g is continuous and $(f\circ G)g$ is continuous and $g(t)\neq 0$. We show that f is continuous at every point of $G(C)$. Since $[a,b]=A\cup B\cup C$ and since both $G(A)$ and $G(B)$ have measure zero we will have proved that f is a.e. continuous in $G([a,b])$.

Suppose $x=G(t)$ for some $t\in C$ and that $G'(t)=g(t)>0$. Then G is strictly increasing on an interval containing the point t. If f is discontinuous at x then $(f\circ G)g$ would have to be discontinuous at t which is not the case. The same argument works if $G'(t)=g(t)<0$.

This completes the proof except for mention of the Riemann-Stieltjes integral

---

is integrable on $[a,b]$ and f is integrable on $G([a,b])$ it may well happen that $f(G(t))$ is not Riemann integrable on $[a,b]$. This, he remarks, is the source of the difficulty for the problem.

in the statement of the theorem. But the methods of Lemma 5.36 show that

$$\int_a^b f(G(t))dG(t) = \int_a^b f(G(t))g(t)\,dt$$

where the existence of one integral in the Riemann sense implies the existence of the other. ∎

Chapter 6

Nonabsolutely Integrable Functions

The study of the Lebesgue integral usually marks the culmination of the study of integration theory on the real line for most mathematics students. They are prepared now for the more abstract theories of integration on measure spaces and studies of the important function spaces.

But the story is still not complete; part of the narrative remains. What about those functions that are integrable, but not absolutely integrable? If f is integrable on an interval $[a,b]$ but

$$\int_a^b |f(x)|\,dx = \infty$$

then f is not Lebesgue integrable. Its indefinite integral

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

has infinite variation on the interval $[a,b]$ since it is always true that

$$\mathrm{Var}(F,[a,b]) = \int_a^b |f(x)|\,dx.$$

To complete the story of the integral on the real line we must persist[1] to study the nonabsolute case and to the study of indefinite integrals that do not have bounded variation. Most of the theory was developed in the decades shortly after Lebesgue's thesis. The standard account is given in

> Stanisław Saks, *Theory of the Integral*. 2nd revised edition. English translation by L. C. Young. Monografje Matematyczne, vol. 7. Warsaw, 1937.

[1] **Note to the instructor**: Well *you* may not want to persist. These topics, while well-known to all specialists in real analysis, are not necessary to the backgrounds of all students, who should be encouraged now to study general measure theory and return to this subject later. The level of this chapter is, accordingly, somewhat raised above the expository level of the preceding chapters.

and much of what we shall do can be found there but expressed in different language. Many mathematicians know none of this theory since the usual courses of instruction move directly to the measure-theoretic treatment of integration theory that does not address such questions.

Since we have committed our text to a relatively full account of the integral on the real line we must forge ahead. The Lebesgue integral does not encompass even the simple calculus integral of Section 1.1 with which we began our studies: there are derivatives that are nonabsolutely integrable. All bounded derivatives are, of course, Lebesgue integrable so that it is in the realm of the unbounded derivatives and some rather delicate considerations that this chapter will lead.

6.1 Variational Measures

The Jordan variation is restricted to the study of functions of bounded variation on a compact interval $[a,b]$. When $\mathrm{Var}(f,[a,b]) = \infty$ there is not much more to be said. For a large part of real analysis this is a sufficiently useful tool. But there are differentiable functions which do not have bounded variation and all nonabsolutely integrable functions have indefinite integrals that are not of bounded variation.

Jordan's theory was extended in the early 20th century to handle functions of finite variation on arbitrary compact sets by A. Denjoy, N. Lusin, and S. Saks. This theory was clarified later by the introduction, by R. Henstock, of measures carrying the variational information of a function. This theory includes the Jordan version and the Denjoy-Lusin-Saks versions and is the appropriate technical tool for the full range of problems arising in our study.

We have already introduced the Lebesgue-Stieltjes measures λ_f (in Section 5.8). We return to that study now with an additional variational measure that is dual to the measure λ_f called the fine variation.

6.1.1 Full and fine variational measures

The variation of a function f on an interval $[a,b]$ is described by the identity

$$\mathrm{Var}(f,[a,b]) = \sup_{\pi}\left(\sum_{([u,v],w)\in\pi} |f(v)-f(u)|\right) \tag{6.1}$$

where the supremum is taken over all possible partitions π of the interval $[a,b]$. We recall that a similar expression describes the Lebesgue-Stieltjes measure

$$\lambda_f(E) = \inf_{\beta}\left\{\sup_{\pi\subset\beta}\left(\sum_{([u,v],w)\in\pi} |f(v)-f(u)|\right)\right\} \tag{6.2}$$

where the supremum is taken over all possible subpartitions π contained in β and the infimum is taken over all full covers β of the set E. The two expressions (6.1) and (6.2) are clearly closely related but the exact relationship needs some thinking (see Exercise 335).

The generalization of Lebesgue measure to the Lebesgue-Stieltjes measure arises by replacing $(v-u)$ by $|f(v)-f(u)|$. It is more convenient for our purposes to write

$$\Delta f([u,v]) = f(v) - f(u)$$

so that $\Delta f(I)$ is an interval function that computes the *increment* of the function f on the interval I. This is often useful in conjunction with the notation $\omega f(I)$ denoting the *oscillation* of the function f on the interval I, defined, we recall, as

$$\omega f(I) = \sup_{u,v \in I} |f(v) - f(u)|.$$

We review the Lebesgue-Stieltjes measure construction and add to it a new variational measure based on fine covers instead of full covers.

Definition 6.1 *Let $f : \mathbb{R} \to \mathbb{R}$ be a function and β a covering relation. We write*

$$\operatorname{Var}(\Delta f, \beta) = \sup_{\pi \subset \beta} \left\{ \sum_{([u,v],w) \in \pi} |\Delta f([u,v])| \right\}$$

where the supremum is taken over all subpartitions π contained in β.

Definition 6.2 (Full and Fine Variations) *Let $f : \mathbb{R} \to \mathbb{R}$ and let E be any set of real numbers. Then we define the full and fine variational measures associated with f by the expressions:*

$$\lambda_f(E) = V^*(\Delta f, E) = \inf\{\operatorname{Var}(\Delta f, \beta) : \ \beta \text{ a full cover of } E\}$$

and

$$\lambda_f^{\star}(E) = V_*(\Delta f, E) = \inf\{\operatorname{Var}(\Delta f, \beta) : \ \beta \text{ a fine cover of } E\}.$$

Note that the star $\star$ (not an asterisk $*$) indicates the fine variation. In general the inequality $\lambda_f^{\star}(E) \le \lambda_f(E)$ holds and identity holds only for a certain (important) class of functions. These set functions share the same properties as the measure λ. Specifically they are countably subadditive for sequences of sets and they are countably additive for disjoint sequences of closed sets.

6.1.2 Finite variation and σ-finite variation

Definition 6.2 allows us to extend the notion of bounded variation to describe the situation on arbitrary sets.

1. f has *bounded variation* on an interval $[a,b]$ if $\operatorname{Var}(f,[a,b]) < \infty$.

2. f has *finite variation* on a set E if $\lambda_f(E) < \infty$.

3. f has *σ-finite variation* on a set E if there is a sequence of sets $\{E_n\}$ covering E and $\lambda_f(E_n) < \infty$ for each $n = 1, 2, 3, \ldots$.

We shall state now and prove (eventually in the material below) that the Lebesgue differentiation theorem of Chapter 2 can be extended to a larger class of functions. Recall that our original statement of Lebesgue's theorem required that the function have bounded variation on the whole of some interval. Here we require only that it have σ-finite variation on a set (not necessarily an interval).

Theorem 6.3 (Lebesgue differentiation theorem) *Let f be a continuous function defined on some open set that contains a set E on which f has σ-finite variation. Then f is differentiable λ-almost everywhere in E and has a finite or infinite derivative λ_f-almost everywhere in E.*

The proof follows from Theorem 6.20 that we shall prove much later.

6.1.3 The Vitali property

The two measures λ_f and $\lambda_f^\star$ together express the variation of the function f. We recall that they are analogous to the full and fine versions of Lebesgue measure, λ^* and λ_*.

Those two measures are identical (i.e., $\lambda^* = \lambda_*$) because of the Vitali covering theorem. Accordingly the identity

$$\lambda_f = \lambda_f^\star$$

(when it holds) would be considered a generalization of the Vitali covering theorem. It is not the case that $\lambda_f = \lambda_f^\star$ in general, but for a most important class of functions this will be true. When the Vitali theorem holds for these measures we say that the function f has the Vitali property.

Definition 6.4 *Let $f : \mathbb{R} \to \mathbb{R}$ and let E be any set of real numbers. Then we say that the function f has the* Vitali property *on E provided that the two measures λ_f and $\lambda_f^\star$ agree on all subsets of E.*

6.1.4 Kolmogorov equivalence

The variation describes a convenient equivalence relation between functions. The notion originated with the Russian mathematician Kolmogorov, and was exploited in this context by Henstock who used the terminology "variational equivalence."

Definition 6.5 (Kolmogorov equivalent) *Two functions f and g are said to be* Kolmogorov equivalent *on E if*

$$V^*(\Delta f - \Delta g, E) = 0.$$

By means of this equivalence relation we can lift a number of properties that we already know for functions of bounded variation to a more general class of functions. When two functions are equivalent in this sense then they must share many properties in common. Here is a list of such properties. Proofs are left for the exercises.

Implications of Kolmogorov equivalence. If the functions f and g are Kolmogorov equivalent on E then:

1. $f'(x) = g'(x)$ at almost every point in E at which g is differentiable. [A partial converse is given in Exercise 329.]

2. f is continuous at every point in E at which g is continuous.

3. f is locally bounded at every point in E at which g is locally bounded.

4. f has the Vitali property on E if and only if g has the Vitali property on E.

5. f has finite variation on E if and only if g has finite variation on E.

6. f has zero variation on E if and only if g has zero variation on E.

7. $\lambda_f(E) = \lambda_g(E)$ and $\lambda_f^\star(E) = \lambda_g^\star(E)$.

6.1.5 Variation of continuous, increasing functions

In special cases it is easy to estimate the full and fine variations. Note that as a result of this first computation we see that continuous, increasing functions possess the Vitali property.

Theorem 6.6 *Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and strictly increasing. Then, for any set E,*

$$\lambda_f^\star(E) = \lambda_f(E) = \lambda(f(E))$$

and f has the Vitali property on every set.

Proof. If β is a full [fine] cover of E then check that

$$\beta' = \{(f(I), f(x)) : (I, x) \in \beta\}$$

is a full [fine] cover of $f(E)$. Note too that $\Delta f(I) = \lambda(f(I))$ for such a function. From this we deduce that

$$\lambda^*(f(E)) = \lambda_f(E)$$

and

$$\lambda_*(f(E)) = \lambda_f^\star(E).$$

By the Vitali covering theorem $\lambda^* = \lambda_*$ so that the identity in the theorem now follows. ∎

6.1.6 Variation and image measure

In general the full variation is larger than the image measure.

Theorem 6.7 *For an arbitrary function $f:\mathbb{R}\to\mathbb{R}$ and any real set E,*

$$\lambda(f(E)) \leq \lambda_f(E).$$

Proof. Let $\lambda_f(E) < t$ and select a full cover β of E so that $\mathrm{Var}(\Delta f, \beta) < t$. We apply the decomposition lemma, Lemma 2.7, for β. There is an increasing sequence of sets $\{E_n\}$ with $E = \bigcup_{n=1}^{\infty} E_n$ and a sequence of nonoverlapping compact intervals $\{I_{kn}\}$ covering E so that if x is any point in E_n and I is any subinterval of I_k that contains x then (I, x) belongs to $\beta([E_n \cap I_{kn}])$.

Thus let us estimate the λ-measure of the set $f(E_n \cap I_{kn})$. Our estimate need only be crude: if $f(x_1)$, $f(x_2)$ with $x_1 < x_2$ are any two points in this set then certainly $([x_1, x_2], x_1) \in \beta(I_k)$. Thus

$$|f(x_1) - f(x_2)| = |\Delta f([x_1, x_2])| \leq \mathrm{Var}(\Delta f, \beta(I_{kn}))$$

so it follows that

$$\lambda(f(E_n \cap I_{kn}) \leq \mathrm{Var}(\Delta f, \beta(I_{kn})).$$

Hence, using Exercise 324 and usual properties of Lebesgue measure,, we have that

$$\lambda(f(E_n)) \leq \sum_k \lambda(f(E_n \cap I_{kn}) \leq \sum_k \mathrm{Var}(\Delta f, \beta(I_{kn}) \leq \mathrm{Var}(\Delta f, \beta) < t.$$

Note that the sequence $\{E_n\}$ is expanding and that its union is the whole set E; it follows that $\{f(E_n)\}$ is expanding and that its union is the whole set $f(E)$. Accordingly then, by Theorem 4.16,

$$\lim_{n\to\infty} \lambda(f(E_n)) = \lambda(f(E)).$$

It follows that

$$\lambda(f(E)) \leq t.$$

Since t was merely chosen so that $\lambda_f(E) < t$ it follows that $\lambda(f(E)) \leq \lambda_f(E)$ as required. ∎

6.1.7 Variational classifications of real functions

Let us review and enlarge some of our terminology for the behavior of functions. All of the following ideas are expressible in the language of the variation. Let $f:\mathbb{R}\to\mathbb{R}$ and let E be any set of reals.

(zero variation) f has *zero variation* on E if $\lambda_f(E) = 0$.

(finite variation) f has *finite variation* on E if $\lambda_f(E) < \infty$.

(σ-finite variation) f has *σ-finite variation* on E if $E \subset \bigcup_{k=1}^{\infty} E_k$ so that $\lambda_f(E_k) < \infty$ for each $k = 1, 2, 3, \dots$.

(Kolmogorov equivalent) f and g are *Kolmogorov equivalent* on E if $V^*(\Delta f - \Delta g, E) = 0$.

(Vitali property on a set) f has the *Vitali property* on E provided that, for all subsets A of E, $\lambda_f(A) = \lambda_f^{\star}(A)$.

(continuous at a point) f is *continuous* at a point x_0 provided that $\lambda_f(\{x_0\}) = 0$.

(weakly continuous at a point) f is *weakly continuous* at a point x_0 provided that $\lambda_f^{\star}(\{x_0\}) = 0$.

(λ-absolutely continuous on a set) f is *λ-absolutely continuous* on E provided that, for every set $N \subset E$ that has Lebesgue measure zero, $\lambda_f(N) = 0$.

(λ-singular on E) f is *λ-singular* on E provided $\lambda_f(E \setminus N) = 0$ for some set $N \subset E$ that has Lebesgue measure zero.

(mutually singular) Two functions f and g are said to be *mutually singular* on a set E if $E = E_1 \cup E_2$ and $\lambda_f(E_2) = \lambda_g(E_1) = 0$.

(saltus function) f is a *saltus function* on an open interval (a, b) if there is a countable set C so that $\lambda_f((a, b) \setminus C) = 0$ and $\lambda_f((a, b) \cap C) < \infty$.

Since each of these terms is definable or describable directly in terms of the variational measures it should be expected that there are many interrelationships. Some of these are explored in the exercises.

Exercises

Exercise 324 *Let β be a covering relation and $f : \mathbb{R} \to \mathbb{R}$. If $\{I_k\}$ is a sequence of nonoverlapping subintervals of an interval I (open or closed) then show that*

$$\sum_{k=1}^{\infty} \mathrm{Var}(\Delta f, \beta(I_k)) \leq \mathrm{Var}(\Delta f, \beta(I)).$$

$\square$

Exercise 325 (Subadditivity property) *Let h_1 and h_2 be real-valued functions defined on interval-point pairs. Then, for any set E, show that*

$$V_*(h_1 + h_2, E) \leq V_*(h_1, E) + V^*(h_2, E)$$

and

$$V^*(h_1 + h_2, E) \leq V^*(h_1, E) + V^*(h_2, E).$$

Answer □

Exercise 326 *Let $f, g : \mathbb{R} \to \mathbb{R}$. Write $f \sim g$ on E if f and g are Kolmogorov equivalent on E. Show that this is an equivalence relation.* □

Exercise 327 *Let $f, g : \mathbb{R} \to \mathbb{R}$. Show that, if f and g are Kolmogorov equivalent on a set E, then $\lambda_f(E) = \lambda_g(E)$ and $\lambda^\star_f(E) = \lambda^\star_g(E)$.* □

Exercise 328 *Let $f, g : \mathbb{R} \to \mathbb{R}$. Show that, if f and g are Kolmogorov equivalent on each of the sets E_1, E_2, E_3, ... then f and g are Kolmogorov equivalent on the union of these sets.* □

Exercise 329 *Let $f, g : \mathbb{R} \to \mathbb{R}$. Show that, if $f'(x) = g'(x)$ at every point of a set E then f and g are Kolmogorov equivalent on E.* □

Exercise 330 *Let $f : \mathbb{R} \to \mathbb{R}$. Show that f is λ-singular on a set E if $f'(x) = 0$ at almost every point x of E.* Answer □

Exercise 331 *Let $f : \mathbb{R} \to \mathbb{R}$. Show, conversely, that if f is λ-singular on a set E then $f'(x) = 0$ at almost every point x of E.* □

Exercise 332 *Show that if $f : \mathbb{R} \to \mathbb{R}$ has finite variation or σ-finite variation on a set E then f is continuous at each point of E with countably many exceptions.* Answer □

Exercise 333 *Show that a function $f : \mathbb{R} \to \mathbb{R}$ is weakly continuous at a point x_0 if and only if there are sequences $c_n \nearrow x_0$ and $d_n \searrow x_0$ so that $d_n - c_n > 0$ and*

$$f(d_n) - f(c_n) \to 0.$$

Answer □

Exercise 334 *Let $f : \mathbb{R} \to \mathbb{R}$. Show that f must be weakly continuous at every point with at most countably many exceptions.* Answer □

Exercise 335 *Let $f : \mathbb{R} \to \mathbb{R}$. Establish the following relation between the Jordan variation and the variational measures:*

$$\lambda_f((a,b)) \leq \mathrm{Var}(f, [a,b]) \leq \lambda_f([a,b]) = \lambda_f((a,b)) + \lambda_f(\{a\}) + \lambda_f(\{b\}).$$

In particular show that

$$\lambda_f((a,b)) = \lambda_f([a,b]) = \mathrm{Var}(f, [a,b])$$

if f is continuous at a and b. □

Exercise 336 *Let $f : \mathbb{R} \to \mathbb{R}$. Show that f has bounded variation on $[a,b]$ if and only if f has finite variation on (a,b). Give an example to show that, even so, $\mathrm{Var}(f, [a,b])$ may be different from $\lambda_f((a,b))$.* Answer □

Exercise 337 *Let $E \subset (a,b)$ be a compact set and let $\{(a_i,b_i)\}$ be the component intervals of $(a,b)\setminus E$. Suppose that f is a continuous function satisfying $f(x)=0$ for all $x \in E$ and that*

$$\sum_i \omega f([a_i,b_i]) < \infty.$$

Show that $\lambda_f(E)=0$. □

Exercise 338 (local recurrence) *A function $f:\mathbb{R}\to\mathbb{R}$ is* locally recurrent *at a point x if there is a sequence of points x_n with $x_n \neq x$ and $\lim_{n\to\infty} x_n = x$ so that $f(x)=f(x_n)$ for all n. Let $f:\mathbb{R}\to\mathbb{R}$ and suppose that f is locally recurrent at every point of a set E. Show that $\lambda_f^\star(E)=0$.* Answer □

Exercise 339 (local monotonicity) *A function $f:\mathbb{R}\to\mathbb{R}$ is* locally nondecreasing *at a point x if there is a $\delta>0$ so that $\Delta f(I)\geq 0$ for every compact interval I containing x for which $\lambda(I)<\delta$. Let $f:\mathbb{R}\to\mathbb{R}$ and suppose that f is locally nondecreasing at every point of a set E and that $\lambda_f(\{x\})<\infty$ for each x in E. Show that f has σ-finite variation on E.* Answer □

Exercise 340 (continuous functions have σ-finite fine variation) *Let $f:\mathbb{R}\to\mathbb{R}$ be a continuous function. Show that $\lambda_f^\star$ must be σ-finite.* Answer □

Exercise 341 (Lebesgue differentiation theorem) *Prove Theorem 6.3:*

> *Let f be a continuous function defined on some open set that contains a set E on which f has σ-finite variation. Then f is differentiable at almost every point of E.*

Hint: *You may assume here the conclusion of Theorem 6.20 that there is a sequence of compact sets covering E on each of which f is Kolmogorov equivalent to some continuous function of bounded variation.* Answer □

6.2 Derivates and variation

If the derivates of a function $f:\mathbb{R}\to\mathbb{R}$ are finite on a set E this has implications for the variation λ_f on E.

6.2.1 Ordinary derivates and variation

Theorem 6.8 *Let $f:\mathbb{R}\to\mathbb{R}$ and suppose that f is differentiable at every point x of a set E. Then*

$$\lambda_f(E)=\lambda_f^\star(E)=\int_E |f'(x)|\,dx.$$

In particular f has σ-finite variation, is λ-absolutely continuous, and has the Vitali property on that set.

Proof. The fact that $f'(x)$ exists on E leads immediately to the variational identity

$$V^*(\Delta f - f'\lambda, E) = 0.$$

From this, using Exercise 325, we can deduce that

$$V^*(\Delta f, E) \leq V^*(\Delta f - f'\lambda, E) + V^*(f'\lambda, E)$$

and hence that

$$\lambda_f(E) = V^*(\Delta f, E) \leq V^*(f'\lambda, E) = \int_E |f'(x)|\, dx.$$

The opposite inequality is proved the same way.

Again, using the other inequality in Exercise subaddprop, we can deduce that

$$V_*(f'\lambda, E) \leq V^*(\Delta f - f'\lambda, E) + V_*(\Delta f, E) = \lambda_f^\star(E)$$

Since f' is measurable the identity

$$\int_E |f'(x)|\, dx = V_*(f'\lambda, E) = V^*(f'\lambda, E)$$

can be used to complete the proof. ∎

Theorem 6.9 *Let $f : \mathbb{R} \to \mathbb{R}$ and suppose at every point x of a set E that $\lambda_f(\{x\}) < \infty$ and that either $\overline{D}f(x) < \infty$ or $\underline{D}f(x) > -\infty$. Then f has σ-finite variation in E.*

Proof. For example let us consider that the set E consists of all points at which $\underline{D}f(x) > -\infty$. Write

$$E_n = \{x : \underline{D}f(x) > -n\}.$$

Note that E is the union of the sequence of sets $\{E_n\}$.

Observe that the function $f_n(x) = f(x) + nx$ is locally nondecreasing at each $x \in E_n$. It follows (from Exercise 339) that f_n has σ-finite variation on E_n. But

$$\lambda_f \leq \lambda_{f_n} + n\lambda.$$

Thus f too has σ-finite variation on E_n. In consequence, f has σ-finite variation on E. ∎

6.2.2 Dini derivatives and variation

For many functions a closer analysis is needed than would be available using the upper and lower derivates: we require one-sided versions.

Definition 6.10 (Dini derivatives) *Let $f:\mathbb{R}\to\mathbb{R}$ and suppose that $x\in\mathbb{R}$. Then the four values*

$$D^+f(x)=\inf_{\delta>0}\sup\left\{\frac{f(x+h)-f(x)}{h}:0<h<\delta\right\}$$

$$D_+f(x)=\sup_{\delta>0}\inf\left\{\frac{f(x+h)-f(x))}{h}:0<h<\delta\right\}$$

$$D^-f(x)=\inf_{\delta>0}\sup\left\{\frac{f(x)-f(x-h)}{h}:0<h<\delta\right\}$$

$$D_-f(x)=\sup_{\delta>0}\inf\left\{\frac{f(x)-f(x-h)}{h}:0<h<\delta\right\}$$

are called the Dini derivatives *of f at x.*

We do not need much more information than this for our main theorem. The reader interested in pursuing the Dini derivatives further should try Exercises 342–351. We will return in Section 6.16 to the Dini derivatives and show how a continuous function can be recovered by integrating one of its Dini derivatives.

Theorem 6.11 *Let $f:\mathbb{R}\to\mathbb{R}$ be a continuous function and suppose that at every point x of a set E either*

$$-\infty<D_+f(x)\leq D^+f(x)<\infty$$

or

$$-\infty<D_-f(x)\leq D^-f(x)<\infty.$$

Then f has σ-finite variation in E and is λ-absolutely continuous there.

Proof. We first show that, for any positive integer c, f has σ-finite variation and is λ-absolutely continuous on the set of points

$$A=\{x:-c<D_+f(x)\leq D^+f(x)<c\}.$$

The geometry of this situation is expressed by the covering relation

$$\beta=\{[x,x+h],x):|\Delta f([x,x+h])|<c\lambda([x,x+h])\}.$$

This relation has none of the properties we have so far encountered, but a modification of our methods will handle.

First apply the ideas of the decomposition from Section 2.7 for β. There is an increasing sequence of sets $\{A_n\}$ with $A=\bigcup_{n=1}^{\infty}A_n$ and a sequence of compact intervals $\{I_{kn}\}$ covering A so that if x is any point in A_n and $[x,x+h]$ is any subinterval of I_{kn} then $([x,x+h],x)$ belongs to β.

In particular if $\{[c_i,d_i]\}$ is a sequence of subintervals of I_{kn} with endpoints in the set A_n, then a brief computation shows that

$$\sum_{i=1}^{\infty}\omega f([c_i,d_i])\leq\sum_{i=1}^{\infty}2c\lambda([c_i,d_i])\leq 2c\lambda(I_{kn}).$$

Let C_{nk} denote the closure of the set $A_n \cap I_{kn}$. Since f is continuous this same inequality extends to points in that closure. Thus if $\{[c_i, d_i]\}$ is a sequence of intervals with endpoints in the compact set C_{nk}, then

$$\sum_{i=1}^{\infty} \omega f([c_i, d_i]) \leq \sum_{i=1}^{\infty} 2c\lambda([c_i, d_i]) \leq 2c\lambda(I_{kn}) < \infty.$$

Define a function g_n so that $g_n(x) = f(x)$ for all $x \in C_{nk}$ and extend to all of the real line so as to be continuous and linear on all of the complementary intervals to C_{nk}. Such a function g_n is evidently continuous and has bounded variation. The same inequality shows that g_n is absolutely continuous in the sense of Vitali and so also λ-absolutely continuous.

The computations of Exercise 337 can be used here to check that

$$V^*(\Delta f - \Delta g_n, C_{kn}) = 0.$$

This shows that f is Kolmogorov equivalent on each set C_{nk} to a continuous function of bounded variation. In particular λ_f is finite on each set C_{nk}. It follows that λ_f is σ-finite on A. The function f also inherits from g_n the property of being λ-absolutely continuous on C_{nk}.

Finally the set E of the theorem can be expressed as a union of a sequence of sets of the same type as A, so that λ_f is σ-finite and vanishes on null subsets of each member of the sequence. The theorem follows. ■

6.2.3 Lipschitz numbers

A Lipschitz condition on a function is a global upper estimate of the ratio

$$\left| \frac{F(y) - F(x)}{y - x} \right| = \left| \frac{\Delta F([x,y])}{\lambda([x,y])} \right|$$

We can make this same estimate locally in which case the estimates are called Lipschitz numbers and they serve as a local estimate of the growth of a function. We refine this a bit by introducing a lower estimate as well. In Section 6.2.4 we show how these numbers relate to the variations.

If $h(I,x)$ is any function which assigns real values to interval-point pairs we recall that in Section 2.8.2 we introduced the following notation for the limits:

$$\limsup_{(I,x) \Longrightarrow x} h(I,x) = \inf_{\delta > 0} \left(\sup\{h(I,x) : \lambda(I) < \delta,\ x \in I\} \right)$$

and

$$\liminf_{(I,x) \Longrightarrow x} h(I,x) = \sup_{\delta > 0} \left(\inf\{h(I,x) : \lambda(I) < \delta,\ x \in I\} \right).$$

These are just convenient expressions for the lower and upper limits of $h(I,x)$ as the interval I (always assumed to contain x) shrinks to the point x. As usual if the $\limsup$ and $\liminf$ are same then the common value (including ∞ and $-\infty$)

would be written as

$$\lim_{(I,x) \Longrightarrow x} h(I,x).$$

When working with such limits Exercises 358 and 359 offer useful estimates of some associated variations.

Definition 6.12 *Let $f : \mathbb{R} \to \mathbb{R}$. Then*

$$\overline{\mathrm{lip}}_f(x) = \limsup_{(I,x) \Longrightarrow x} \left| \frac{\Delta f(I)}{\lambda(I)} \right|$$

$$\underline{\mathrm{lip}}_f(x) = \liminf_{(I,x) \Longrightarrow x} \left| \frac{\Delta f(I)}{\lambda(I)} \right|$$

are called the upper and lower Lipschitz number *of f at a point x.*

Lemma 6.13 *Let $f : \mathbb{R} \to \mathbb{R}$. For any real number r the sets*

$$\{x : \overline{\mathrm{lip}}_f(x) < r\} \text{ and } \{x : \underline{\mathrm{lip}}_f(x) < r\}$$

are measurable.

This is nearly identical to Lemma 4.21.

6.2.4 Six growth lemmas

The growth lemmas we present all follow easily from the general limit lemmas of Exercises 358 and 359. Proofs are left to the student. They can be considered as generalizations of these simple facts, well-known to calculus students:

1. If $f'(x) \leq r$ for all x then
$$f(b) - f(a) \leq r(b-a).$$
2. If $f'(x) \geq r$ for all x then
$$f(b) - f(a) \geq r(b-a).$$

Now, however, the derivative is replaced by upper and lower Lipschitz estimates, the interval $[a,b]$ is replaced by an arbitrary set and the increments are replaced by variational measures.

Lemma 6.14 *Let $f : \mathbb{R} \to \mathbb{R}$. If $\underline{\mathrm{lip}}_f(z) < r$ for every $z \in E$ then*

$$\lambda_f^\star(E) \leq r\lambda(E).$$

Lemma 6.15 *Let $f : \mathbb{R} \to \mathbb{R}$. If $\overline{\mathrm{lip}}_f(z) > r > 0$ for every $z \in E$ then*

$$r\lambda(E) \leq \lambda_f(E).$$

Lemma 6.16 *Let $f : \mathbb{R} \to \mathbb{R}$. If $\overline{\text{lip}}_f(z) < r$ for every $z \in E$ then*
$$\lambda_f(E) \leq r\lambda(E).$$

Lemma 6.17 *Let $f : \mathbb{R} \to \mathbb{R}$. If $\underline{\text{lip}}_f(z) > r > 0$ for every $z \in E$ then*
$$r\lambda(E) \leq \lambda^{\star}_f(E).$$

Lemma 6.18 *Let $f : \mathbb{R} \to \mathbb{R}$. If $\lambda_f(E) < \infty$, then $\overline{\text{lip}}_f(x) < \infty$ for almost every point x in E.*

Lemma 6.19 *Let $f : \mathbb{R} \to \mathbb{R}$. If $\lambda^{\star}_f(E) < \infty$ then $\underline{\text{lip}}_f(z) < \infty$ for almost every point x in E.*

Exercises

Exercise 342 *Show that*
$$\underline{D}f(x) \leq D_+f(x) \leq D^+f(x) \leq \overline{D}f(x) \text{ and } \overline{D}f(x) = \max\{D^-f(x), D^+f(x)\}.$$
□

Exercise 343 (Beppo Levi) *Let $f : \mathbb{R} \to \mathbb{R}$ and suppose that f has one-sided derivatives $f'_+(x)$ and $f'_-(x)$ at each point of a set E. Show that the set of points x in E at which*
$$f'_+(x) \neq f'_-(x)$$
is countable. [See Exercise 62.] □

Exercise 344 (Grace Chisolm Young) *Let $f : \mathbb{R} \to \mathbb{R}$. Show that the sets of points*
$$\{x : D^-f(x) < D_+f(x)\}$$
and
$$\{x : D^+f(x) < D_-f(x)\}$$
are both countable. Answer □

Exercise 345 *It is easy to misinterpret the theorem of Beppo Levi (Exercise 343). To avoid this construct a continuous function $f : \mathbb{R} \to \mathbb{R}$ so that for some uncountable set E the right-hand derivative $f'_+(x)$ exists at each point of E and the left-hand derivative $f'_-(x)$ fails to exist at each point of E.* □

Exercise 346 (William Henry Young) *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Show that the sets of points*

$$\{x : D^- f(x) = D^+ f(x)\}$$

and

$$\{x : D_- f(x) = D_+ f(x)\}$$

are both residual subsets of $\mathbb{R}$. Answer □

Exercise 347 *Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Show that the set of points at which f has a right-hand derivative but no left-hand derivative is a meager subset of $[a,b]$.* □

Exercise 348 *Let $f : [a,b] \to \mathbb{R}$ be a continuous function with $f([a,b]) = [c,d]$. Write*

$$D = \{x \in [a,b] : D^+ f(x) \leq 0\}.$$

Show that either f is nondecreasing on $[a,b]$ or else $f(D)$ contains a compact subinterval of $[c,d]$. Answer □

Exercise 349 (Anthony P. Morse) *Let $f : [a,b] \to \mathbb{R}$ be a continuous function with $f([a,b]) = [c,d]$. Write*

$$A = \{x \in [a,b] : D^+ f(x) \geq 0\},$$
$$B = \{x \in [a,b] : D^+ f(x) < 0\},$$

and

$$C = \{x \in [a,b] : D^+ f(x) = 0\}.$$

Suppose that A is dense in $[a,b]$. Show that B is a meager subset of $[a,b]$ and $f(B)$ is a meager subset of $[c,d]$. Moreover, show that either f is nondecreasing on $[a,b]$ or else $f(C)$ contains a residual subset of some compact subinterval of $[c,d]$. Answer □

Exercise 350 (Darboux property of Dini derivatives) *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and suppose that the Dini derivative $D^+ f(x)$ is unbounded both above and below on each interval. Show, for every real number r and compact interval $[a,b]$, that f maps the set*

$$E_r = \{x \in [a,b] : D^+ f(x) = r\}$$

onto a residual subset of some compact interval. (In particular $D^+ f(x)$ assumes every real number at many points in any subinterval.) □

Exercise 351 *For any continuous function $f : \mathbb{R} \to \mathbb{R}$ and any real number r show that the sets*

$$\{x : D^+ f(x) \leq r\} \text{ and } \{x : D_+ f(x) \leq r\}$$

are Lebesgue measurable. Answer □

Exercise 352 *Let $f : \mathbb{R} \to \mathbb{R}$. Verify that*

$$\overline{\text{lip}}_f(x) = \max\{|\overline{D}f(x)|, |\underline{D}f(x)|\}$$

and also

$$\overline{\text{lip}}_f(x) = \max\{|\overline{D}^+ f(x)|,\ |\underline{D}^+ f(x)|,\ |\overline{D}^- f(x)|,\ |\underline{D}^- f(x)|\}.$$

□

Exercise 353 *Let $f : \mathbb{R} \to \mathbb{R}$. Suppose that f has a derivative at x (finite or infinite). Show that $\overline{\text{lip}}_f(x) = \underline{\text{lip}}_f(x) = |f'(x)|$.* □

Exercise 354 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function, and suppose that $\overline{\text{lip}}_f(x) = \underline{\text{lip}}_f(x) < \infty$. Show that f has a finite derivative at x and that*

$$\overline{\text{lip}}_f(x) = \underline{\text{lip}}_f(x) = |f'(x)|.$$

Answer □

Exercise 355 *If $f : \mathbb{R} \to \mathbb{R}$ is continuous and $\underline{\text{lip}}_f(x) = \infty$ show that either $f'(x) = \infty$ or $f'(x) = -\infty$. Give an example to show that continuity cannot be dropped.* □

Exercise 356 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Show that $\underline{\text{lip}}_f(z) < \infty$ for almost every point x.* Answer □

Exercise 357 *For this exercise and the next three exercises we shall use the following generalized variations. Let h be any real-valued function defined on interval-point pairs and define*

$$\text{Var}(h,\beta) = \sup_{\pi\subset\beta} \left\{ \sum_{([u,v],w)\in\pi} |h([u,v],w)| \right\}$$

where the supremum is taken over all π, arbitrary subpartitions contained in β;

$$h^*(E) = \inf\{\text{Var}(h,\beta) :\ \beta \textit{ a full cover of } E\}$$

and

$$h_*(E) = \inf\{h,\beta) :\ \beta \textit{ a fine cover of } E\}.$$

Show that h^* and h_* are measures and that $h_* \le h^*$. □

Exercise 358 (limsup comparison lemma) *Suppose that, for every x in a set E,*

$$s < \limsup_{(I,x)\Longrightarrow x} \left| \frac{h(I,x)}{\lambda(I)} \right| < r$$

Show that

$$s\lambda(E) \leq V^*(h,E) \leq r\lambda(E)$$

and

$$V_*(h,E) \leq r\lambda(E).$$

Answer □

Exercise 359 (liminf comparison lemma) *Suppose that, for every x in a set E,*

$$s < \liminf_{(I,x) \Longrightarrow x} \left| \frac{h(I,x)}{k(I,x)} \right| < r$$

Show that

$$s\lambda(E) \leq V_*(h,E) \leq r\lambda(E)$$

and

$$s\lambda(E) \leq V^*(h,E).$$

Answer □

Exercise 360 *Deduce all of the growth lemmas in Section 6.2.4 from the liminf comparison and limsup comparison lemmas (i.e., Exercises 358 and 359).* □

Exercise 361 *Let $f : \mathbb{R} \to \mathbb{R}$. If $\overline{\text{lip}}_f(z) < \infty$ for every $z \in E$ then show that f has σ-finite variation in E and is λ-absolutely continuous there.* Answer □

6.3 Continuous functions with σ-finite variation

We begin now a deeper analysis of those continuous functions that have σ-finite full variation on a set. Because of part (3) of this theorem we now can deduce the Lebesgue differentiation theorem (Theorem 6.3) asserting that these functions are almost everywhere differentiable.

Theorem 6.20 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and E a real set. Then the following are equivalent:*

1. *f has σ-finite variation on E,*
2. *there is a sequence $\{E_n\}$ of compact sets covering E so that f has finite variation on each E,*
3. *there is a sequence $\{E_n\}$ of compact sets covering E so that on each E_n, f is Kolmogorov equivalent to some continuous function of bounded variation.*

Proof. The implication (2) $\Longrightarrow$ (1) is trivial. The implication (3) $\Longrightarrow$ (2) is easy: if (3) holds then, for some continuous function of bounded variation $g_n : \mathbb{R} \to \mathbb{R}$, the equivalence relation

$$V^*(\Delta f - \Delta g_n, E_n) = 0$$

implies that $\lambda_f(E_n) = \lambda^*_{g_n}(E_n) < \infty$.

Thus the proof is completed by showing that (1) $\Longrightarrow$ (3). It is enough to consider the situation for which E is a bounded set for which $\lambda_f(E) < \infty$. Choose a full cover β of E and a real number t so that

$$\mathrm{Var}(\Delta f, \beta) < t < \infty.$$

Apply the decomposition in Lemma 2.7 to β. Accordingly there is an increasing sequence of sets $\{B_n\}$ with $E = \bigcup_{n=1}^{\infty} B_n$ and a sequence of nonoverlapping compact intervals $\{I_{kn}\}$ covering E so that if x is any point in B_n and I is any subinterval of I_{kn} that contains x then (I, x) belongs to β.

Let $A_{kn} = B_n \cap I_{kn}$. We check some facts about the variation of f on A_{kn}. Suppose that $\{[a_i, b_i]\}$ is any disjointed sequence of compact subintervals of I_{kn} each of which contains at least one point, say x_i, of B_n. Then $\{([a_i, b_i], x_i)\}$ must form a subpartition contained in β. Consequently

$$\sum_i |f(b_i) - f(a_i)| \leq \mathrm{Var}(\Delta f, \beta) < t.$$

Now let C_{kn} denote the closure of A_{kn}, i.e., C_{kn} is the smallest compact set that contains A_{kn}. We extend these considerations to estimating the variation of f on the larger set C_{kn}. Suppose now that $\{[a_i, b_i]\}$ is any disjointed sequence of compact subintervals of I_{kn} each of which contains at least one point of C_{nk}. We enlarge each interval slightly as needed to ensure that the intervals remain disjointed but contain also a point, now, of the dense subset A_{kn}. As f is continuous we can do this without much of an increase in the sums, and so we can certainly guarantee that for the given sequence $\{[a_i, b_i]\}$ that

$$\sum_i |f(b_i) - f(a_i)| < 2t < \infty.$$

Let us define a function g_{nk} so as to be equal to $f(x)$ on the compact set C_{kn} and extended to the real line so as to be linear and continuous on the intervals complementary to C_{kn}. Such a function g_{nk} is continuous and has bounded variation.

The computations of Exercise 337 can be used here to check that

$$V^*(\Delta f - \Delta g_{nk}, C_{kn}) = 0.$$

As every compact set from the sequence $\{C_{kn}\}$ can be treated the same way, we have verified the implication (1) $\Longrightarrow$ (3) provided we merely relabel the full collection $\{C_{kn}\}$ as a single sequence $\{E_n\}$. ∎

6.3.1 Variation on compact sets

We can refine our analysis of σ-finite variation with a few further steps.

Theorem 6.21 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and E a compact set. Then the following are equivalent:*

1. *f has σ-finite variation on E.*
2. *Every nonempty compact subset S of E has a portion $S \cap (a,b)$ on which f has finite variation.*
3. *f has σ-finite variation on every null set $Z \subset E$ that is a $\mathcal{G}_\delta$ set.*

Proof. By a $\mathcal{G}_\delta$ set we mean a set Z of the form $Z = \bigcap_{n=1}^{\infty} G_n$ for some sequence $\{G_n\}$ of open sets. Every closed set can be written in this form.

We begin with (a) $\Longrightarrow$ (b). As we have seen in Theorem 6.20, if f has σ-finite variation on E, then there is a sequence of compact sets $\{E_n\}$ covering the compact set S so that $\lambda_f(E_n) < \infty$ for each n. By the Baire category theorem there must be a portion $S \cap (a,b)$ of E contained in one at least from the sequence $\{E_n\}$. In particular, for some n, $\lambda_f(S \cap (a,b) \leq \lambda_f(E_n) < \infty$ as required to prove (b).

Let us now prove that (b) $\Longrightarrow$ (a) Suppose that every nonempty closed subset S of E has a portion $S \cap (a,b)$ on which f has finite variation. Let G denote the real set consisting of all real x with the property that there is a $\delta(x) > 0$ so that f has σ-finite variation on the set $E \cap (x - \delta(x), x + \delta(x))$. Note that

$$G = \bigcup_{x \in G} (x - \delta(x), x + \delta(x))$$

so G is open.

Consider the set $G \cap E$. Any point in this set would be contained in an open interval (c,d) with rational endpoints so that f has σ-finite variation on $G \cap (c,d)$. It follows that f has σ-finite variation on $G \cap E$. If $G \supset E$ then, we deduce that f has σ-finite variation on E as we wished to prove to verify (a).

Suppose, in order to obtain a contradiction that G does not contain E. Let $E' = E \setminus G$. This would be a nonempty closed subset of E and so, by hypothesis, there would have to be a portion $E' \cap (a,b)$ on which f has finite variation. But if f has finite variation on $E' \cap (a,b)$ and also, evidently, has σ-finite variation on $E \setminus E'$ then f must have σ-finite variation on $E \cap (a,b)$. Every point of this set should belong to G which is impossible in view of the assumption that $E' \cap (a,b)$ is a portion. This contradiction completes our proof that (b) $\Longrightarrow$ (a).

The implication (a) $\Longrightarrow$ (c) is trivial. To complete the proof, then, it will suffice to verify that (c) $\Longrightarrow$ (b). Suppose that f has σ-finite variation on every set $Z \subset E$ that is a $\mathcal{G}_\delta$ set of λ-measure zero.

Let S be a nonempty closed subset of E. To verify (b) we need to find a portion of S on which f has finite variation. If S is a null set then we are almost there. A closed set is also of type $\mathcal{G}_\delta$. Thus, λ_f is σ-finite on S by hypothesis. As we have already argued above, in this situation we are assured that S has a portion $S \cap (a,b)$ on which f has finite variation.

Suppose instead that S is a closed set having positive measure. Exercise 362, which follows the proof, shows exactly how to choose a null subset Z of S that is a $\mathcal{G}_\delta$-set that is dense in S. By our assumption (c), there must be a portion $Z \cap (a,b)$ on which f has σ-finite variation. We apply Theorem 6.20 to obtain a sequence of compact sets $\{K_n\}$ whose union includes $Z \cap (a,b)$ so that each $\lambda_f(K_n) < \infty$.

Apply the Osgood-Baire theorem, now to the sequence of compact sets $\{K_n\}$ that covers the $\mathcal{G}_\delta$-set $Z \cap (a,b)$. Recall that the Osgood-Baire theorem, stated in Section 6.17 for closed sets, applies equally well to $\mathcal{G}_\delta$-sets. Thus we can conclude that there is a portion $Z \cap (c,d)$ and an integer k so that $Z \cap (c,d) \subset K_k$. Since Z is dense in the compact set S we also have $S \cap (c,d) \subset K_k$. In particular

$$\lambda_f(S \cap (c,d)) \leq \lambda_f(K_n) < \infty.$$

We have obtained again (but this time without the additional assumption that S has measure zero) exactly property (b). ■

Exercise 362 *Let S be a compact set. Show that there is a subset Z of S that is of type $\mathcal{G}_\delta$, is a null set, and is dense in S.* Answer □

6.3.2 λ-absolutely continuous functions

As a corollary to Theorem 6.21 immediately we have a special observation, since an λ-absolutely continuous function must have finite variation (indeed zero variation) on every set of λ-measure zero.

Corollary 6.22 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function that is λ-absolutely continuous on a compact set E. Then f has σ-finite variation and is differentiable a.e. on E.*

Corollary 6.23 *Let $f : [a,b] \to \mathbb{R}$ be absolutely continuous in the variational sense on $[a,b]$. Then f has σ-finite variation and is differentiable a.e. on $[a,b]$.*

6.4 Vitali property and differentiability

In this section we show that differentiability on a set implies the Vitali property on that set and, conversely, that the Vitali property on a set implies almost everywhere differentiability.

Theorem 6.24 *Let $f : \mathbb{R} \to \mathbb{R}$ have a finite derivative at every point of a set E. Then f has the Vitali property on E and, moreover,*

$$\lambda_f(E) = \lambda_f^\star(E) = \int_E |f'(x)|\,dx.$$

Proof. This is already proved in Theorem 6.8. ∎

Theorem 6.25 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function that has the Vitali property on a set E. Then f has a finite derivative at almost every point of E and, except at the points of a set N for which $\lambda_f(N) = 0$, f has a finite or infinite derivative $f'(z)$.*

Proof. We need work only with the Lipschitz numbers here. Recall that if $\underline{\mathrm{lip}}_f(z) = \infty$ then necessarily f has an infinite derivative, $f'(z) = \infty$ or $f(z) = -\infty$ (see Exercise 355). Also if

$$\underline{\mathrm{lip}}_f(z) = \overline{\mathrm{lip}}_f(z) < \infty$$

then f has a finite derivative at z (see Exercise 354).

It is enough to prove the theorem under the assumption that E is a bounded set. We examine

$$A = \{x \in E : \underline{\mathrm{lip}}_f(x) < \overline{\mathrm{lip}}_f(x)\}.$$

As is usual in arguments of this type, introduce rational numbers $0 < r < s$ and the subsets

$$A_{rs} = \{x \in A : \underline{\mathrm{lip}}_f(x) < r < s < \overline{\mathrm{lip}}_f(x)\}.$$

Note that A is the countable union of this collection of sets taken over all rationals r and s with $r < s$.

By the growth lemmas of Section 6.2.4 we obtain

$$\lambda_f^\star(A_{rs}) \le r\lambda(A_{rs}) \le s\lambda(A_{rs}) \le \lambda_f(A_{rs}).$$

Our assumption that f has the Vitali property on E gives the identity $\lambda_f = \lambda_f^\star$ on each of these subsets of E. None of these numbers are infinite, $r < s$, and so the inequality makes sense only in the case that $\lambda_f(A_{rs}) = \lambda(A_{rs}) = 0$. Consequently $\lambda_f(A) = \lambda(A) = 0$.

At every point x in $E \setminus A$ we know that either

$$\underline{\mathrm{lip}}_f(x) = \overline{\mathrm{lip}}_f(x) < \infty$$

or else

$$\underline{\mathrm{lip}}_f(x) = \overline{\mathrm{lip}}_f(x) = +\infty.$$

In the former case, as we have already noted, f has a finite derivative and in the latter case f has an infinite derivative. This latter case can occur only on a set of Lebesgue measure zero (as a consequence of Lemma 6.19). ∎

Corollary 6.26 *Let $f:\mathbb{R}\to\mathbb{R}$ be a continuous function that has the Vitali property on a set E and let us specify the following subsets of E at which the derivative exists finitely or infinitely:*

1. $E_d = \{x \in E : f$ *is differentiable at* $x\}$.
2. $E_\infty = \{x \in E : f'(x) = \pm\infty\}$.

Then

$$\lambda_f^\star(E) = \lambda_f(E) = \int_{E_d} |F'(x)|\,dx + \lambda_f(E_\infty).$$

6.5 The Vitali property and variation

The Vitali property is closely related to the finiteness of the variation. Indeed, since the fine variation $\lambda_f^\star$ of a continuous function f is always σ-finite, we know that the identity $\lambda_f^\star(E) = \lambda_f(E)$ can only hold if f has σ-finite variation on E.

6.5.1 Monotonic functions

Theorem 6.27 *Let $f:\mathbb{R}\to\mathbb{R}$ be a continuous, monotonic nondecreasing function. Then f has the Vitali property.*

Proof. Suppose first that f is continuous and is *strictly* increasing. Then certainly f has the Vitali property. Theorem 6.6 immediately supplies the identity

$$\lambda_f^\star(E) = \lambda_f(E) = \lambda(f(E)).$$

Now suppose that f is continous and increasing, but not necessarily stricly so. Let $\varepsilon > 0$ and define a new function $g(x) = f(x) + \varepsilon x$. The function g is continuous and strictly increasing so, by our first observation, $\lambda_{g*} = \lambda_g{}^*$. From Exercise 325 we deduce the inequalities

$$\lambda_f \le \lambda_g{}^* \le \lambda_f + \varepsilon\lambda^*$$

and

$$\lambda_f^\star \le \lambda_{g*} \le \lambda_f^\star + \varepsilon\lambda^*.$$

From these two inequalities and the identity $\lambda_{g*} = \lambda_g^*$ we can deduce $\lambda_f = \lambda_f^\star$. ■

Exercise 363 *Let $f:\mathbb{R}\to\mathbb{R}$ be a monotonic, nondecreasing function. Show that if $\lambda_f^\star(\{x\}) = \lambda_f(\{x\})$ for a point x then f must be continuous at x.* □

6.5.2 Functions of bounded variation

Theorem 6.28 *Let $f:\mathbb{R}\to\mathbb{R}$ be a continuous function that is locally of bounded variation. Then f has the Vitali property on the real line.*

Proof. Fix a compact interval $[a,b]$ and let g be the total variation function of f on $[a,b]$. We know that this relation between a function and its total variation function requires the identity

$$V^*(\Delta g - |\Delta f|,(a,b)) = 0.$$

In particular $\lambda_f(E) = \lambda_g(E)$ and $\lambda_f^\star(E) = \lambda_g^\star(E)$ for all subsets E of (a,b). By the previous theorem $\lambda_g(E) = \lambda_g^\star(E)$ and so $\lambda_f(E) = \lambda_f^\star(E)$ follows. This argument produces the identity we require on all bounded sets, and the extension to arbitrary sets follows from measure properties. ∎

6.5.3 Functions of σ-finite variation

Theorem 6.29 *Let $f:\mathbb{R}\to\mathbb{R}$ be a continuous function. Then f has σ-finite variation on a set E if and only if f has the Vitali property on E.*

Proof. We already know that the Vitali property for a continuous function will imply σ-finite variation. Let us prove the converse.

Suppose that f is continuous function that has σ-finite variation on E. By Theorem 6.20 there is a sequence of compact sets $\{E_n\}$ covering E and a sequence of functions g_n each continuous and locally of bounded variation so that

$$V^*(\Delta f - \Delta g_n, E_n) = 0 \tag{6.3}$$

We know then, from the previous theorem, that $\lambda_{g_n *} = \lambda_{g_n}^*$. We also know that the equivalence (6.3) requires that $\lambda_{g_n}{}^* = \lambda_f$ and $\lambda_{g_n *} = \lambda_f^\star$ on all subsets of E_n.

Introduce the notation

$$A_n = E_n \setminus \bigcup_{k<n} E_k$$

so that $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty E_n$ and the sets $\{A_n\}$ are pairwise disjoint, measurable sets. The student should justify that the following computations are permitted:

$$\lambda_f(E) = \sum_{n=1}^\infty \lambda_f(E\cap A_n) = \sum_{n=1}^\infty \lambda_{g_n}{}^*(E\cap A_n) =$$

$$\sum_{n=1}^\infty \lambda_{g_n *}(E\cap A_n) = \sum_{n=1}^\infty \lambda_f^\star(E\cap A_n) = \lambda_f^\star(E).$$

As this applies as well to any subset of E we see that f must have the Vitali property on E as required. ∎

Corollary 6.30 *If $f:\mathbb{R}\to\mathbb{R}$ is a continuous function that is λ-absolutely continuous on a compact set E, then f has the Vitali property on E.*

Proof. Use Corollary 6.22. ∎

6.6 Characterization of the Vitali property

The class of functions satisfying the Vitali property on a set is fundamental to an understanding of any program seeking to study all aspects of the relation among the concepts of derivative, integral and variation. We have already found a number of characterizations in Theorem 6.20 and Theorem 6.21. Here are some more. Some are easy consequences of what we have proved [e.g., (1) and Theorem 340 immediately imply (2)]. Others are left as entertainments for the student.

Theorem 6.31 *Let $f:\mathbb{R}\to\mathbb{R}$ be a continuous real function and let E be a compact set. The following are equivalent:*

1. *f has the Vitali property on E.*
2. *f has σ-finite variation on E.*
3. *there is a sequence of compact sets $\{E_n\}$ with $E=\bigcup_{n=1}^{\infty}E_n$ so that for each n there is a continuous function g_n that is locally of bounded variation so that f and g_n are Kolmogorov equivalent on E_n.*
4. *f has a derivative (finite or infinite) at λ_f-almost every point of E.*
5. *There is a continuous, increasing function g so that*
$$\limsup_{(I,x)\Longrightarrow x}\left|\frac{\Delta f(I)}{\Delta g(I)}\right|<\infty$$
at every point $x\in E$.
6. *There is a continuous, increasing function g and a real function f_1 so that*
$$V^*(\Delta f-f_1\Delta g,E)=0.$$
7. *There is a continuous, increasing function g so that the composed function $f\circ g$ has a finite derivative everywhere in the compact set $g^{-1}(E)$.*

6.7 Characterization of λ-absolute continuity

The Vitali property expresses the most important property arising in studies of the derivative. The special subclass of λ-absolutely continuous functions plays its most significant role in the integration theory. Here are some similar characterizations for this class, most easily proved from previously proved statements or techniques.

Theorem 6.32 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and let E be a compact set. The following are equivalent:*

1. *f is λ-absolutely continuous on E.*
2. *f has σ-finite variation on E and is λ-absolutely continuous there.*
3. *there is a sequence of compact sets $\{E_n\}$ with $E = \bigcup_{n=1}^{\infty} E_n$ so that for each n there is a continuous function g_n that is of locally of bounded variation and absolutely continuous in the sense of Vitali so that f and g_n are Kolmogorov equivalent on E_n.*
4. *f has a finite derivative at λ_f-almost every point of E.*
5. *There is an increasing, λ-absolutely continuous function g so that*
$$\limsup_{(I,x)\to x} \left| \frac{\Delta f(I)}{\Delta g(I)} \right| < \infty$$
at every point $x \in E$.
6. *There is an increasing, λ-absolutely continuous function g and a real function f_1 so that*
$$V^*(\Delta f - f_1 \Delta g, E) = 0.$$

6.8 Mapping properties

For any set E and any function $f : \mathbb{R} \to \mathbb{R}$ the image of E under the mapping f is written as
$$f(E) = \{f(x) : x \in E\}.$$
We already know some properties of the image set for continuous functions. We recall from elementary studies that the image of any compact interval $[a,b]$ under f is again a compact interval. It is easy to check that that the image of any compact set E under f is again a compact set $f(E)$. A natural question is whether the image of an measurable set must also be measurable .

Theorem 6.33 *Let $f : \mathbb{R} \to \mathbb{R}$ be an measurable function and P an measurable set. The following are equivalent:*

(M) *$f(E)$ is measurable for every measurable subset E of P,*

(N) *$\lambda(f(N)) = 0$ for every subset N of P for which $\lambda(N) = 0$.*

Proof. Suppose that E is measurable and that the second statement of the theorem holds. We need consider only the case where E is bounded. Since f is measurable , then by definition, we can find open sets G_n so that $\lambda(G_n) < 1/n$,

$E \setminus G_n$ is compact and f is equal to a continuous function $g_n : \mathbb{R} \to \mathbb{R}$ on the compact set $E \setminus G_n$.

In particular

$$E = Z \cup \bigcup_{n=1}^{\infty} (E \setminus G_n)$$

where

$$Z = E \cap \bigcap_{n=1}^{\infty} G_n$$

has λ-measure zero. By hypothesis $f(Z)$ must be a set of λ-measure zero and hence is measurable . Also each

$$f(E \setminus G_n) = g_n(E \setminus G_n)$$

is a compact set (since the continuous function g_n maps compact sets to compact sets). In particular each set here is also measurable . Thus

$$f(E) = f(Z) \cup \bigcup_{n=1}^{\infty} f(E \setminus G_n)$$

displays $f(E)$ as the union of a sequence of measurable sets. Thus $f(E)$ is also measurable .

Conversely suppose that the first statement of the theorem does not hold, yet the second does. Then there is a set $Z \subset P$ for which $\lambda(Z) = 0$ and yet $f(Z)$ does not have λ-measure zero. For (b) to be true, however, $f(Z)$ should be an measurable set of positive measure. Such a set must have a subset A that is not measurable .

We shall not pause to prove this assertion but leave it as a project for the student to find elsewhere (or prove). A proof will require use of a logical principle that is beyond our current course of study.

Then there is a set $Z_1 \subset Z$ with $f(Z_1) = A$. The set Z_1 must be measurable merely because $\lambda(Z_1) \leq \lambda(Z) = 0$. But then f maps an measurable set Z_1 to a set $f(Z_1) = A$ that is not measurable. We have contradicted the second statement thus completing the proof. ∎

6.9 Lusin's conditions

Definition 6.34 *A function $f : \mathbb{R} \to \mathbb{R}$ is said to satisfy* Lusin's conditions *on a set P when these equivalent conditions hold:*

(M) *$f(E)$ is measurable for every measurable subset E of P,*

(N) *$\lambda(f(N)) = 0$ for every subset N of P for which $\lambda(N) = 0$.*

Theorem 6.35 *If $f : \mathbb{R} \to \mathbb{R}$ is λ-absolutely continuous on an measurable set P then f satisfies Lusin's conditions on P.*

Proof. This follows immediately from Theorem 6.7 that asserts that $\lambda(f(N))$ is smaller than the full variation of f on N. Thus for every null set $N \subset P$,

$$\lambda(f(N)) \leq \lambda_f(N) = 0.$$

■

6.10 Banach-Zarecki Theorem

In the converse direction we should expect that Lusin's conditions play a role in characterizing the important property of absolute continuity.

Theorem 6.36 (Banach-Zarecki) *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function and E a compact set. Then the following are necessary and sufficient conditions in order that f is λ-absolutely continuous on E:*

1. *f has σ-finite variation on E, and*

2. *f satisfies Lusin's condition on E.*

Proof. Certainly if f is λ-absolutely continuous then we already know that (a) holds because of Theorem 6.21 and that (b) holds because of Theorem 6.35.

Conversely let us suppose that (a) and (b) now hold. We know from Theorem 6.20 that when f has σ-finite variation on a compact set E, there is a sequence $\{E_n\}$ of compact sets covering E and a sequence of continuous functions of bounded variation g_n so that f and g_n are Kolmogorov equivalent on E_n. Recall in the proof that the construction there required $f = g_n$ on the set E_n. We can insist on that here. Moreover the functions g_n in the proof that extended f were also chosen to be merely linear or constant in the intervals complementary to E_n. We can insist also on that here.

We note that the condition (b) of the theorem asserting that f satisfies Lusin's condition on E means that g_n satisfies this same condition on E_n. Moreover by the nature of the construction the function g_n satisfies Lusin's condition on all sets. The proof is completed now by addressing the special case of proving that g_n is λ-absolutely continuous.

Note that each g_n constructed in our proof above satisfies the hypotheses of Exercise 364 below. Indeed, since g_n has bounded variation on every interval it is differentiable outside of a set N of λ-measure zero. The assumption of Lusin's condition on g_n then provides $\lambda(g_n(N)) = 0$. The finiteness of

$$\lambda_{g_n}(\mathbb{R} \setminus N) = \int_{\mathbb{R} \setminus N} |g_n'(x)|\, dx$$

follows from the fact that g_n, as constructed have finite variation.

Now let Z be any set for which $\lambda(Z) = 0$. Let $\varepsilon > 0$ and choose $\delta > 0$ by applying the Exercise 364 to this function g_n. Choose an open set $G \subset Z$ with

$\lambda(G) < \delta$. Choose any full cover β of Z; then $\beta(G)$ is also a full cover of Z and the exercise provides

$$V^*(\Delta g_n, Z) \leq \mathrm{Var}(\Delta g_n, \beta(G)) < \varepsilon.$$

From this we deduce that $\lambda_{g_n}(Z) = 0$. In consequence g_n is λ-absolutely continuous.

From this we can prove that f is λ-absolutely continuous on the set E in question. For if Z is a set of λ-measure zero then $\lambda^{g_n}(Z) = 0$ will imply that

$$\lambda_f(E_n \cap Z) = \lambda^{g_n}(E_n \cap Z) = 0$$

and hence that

$$\lambda_f(E \cap Z) \leq \sum_{n=1}^{\infty} \lambda_f(E_n \cap Z) = 0.$$

This will then show that f is λ-absolutely continuous on E. ■

Corollary 6.37 *Let $f : [a,b] \to \mathbb{R}$. The following are necessary and sufficient conditions in order that f is absolutely continuous in the variational sense on $[a,b]$:*

1. *f is continuous,*
2. *f has σ-finite variation on $[a,b]$, and*
3. *f satisfies Lusin's conditions on $[a,b]$.*

Corollary 6.38 *Let $f : [a,b] \to \mathbb{R}$. The following are necessary and sufficient conditions in order that f is absolutely continuous in the sense of Vitali on $[a,b]$:*

1. *f is continuous,*
2. *f has bounded variation on $[a,b]$, and*
3. *f satisfies Lusin's conditions on $[a,b]$.*

A crucial step in the proof of the theorem uses the following classical problem:

Exercise 364 *Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function. Suppose that g is differentiable at each point with the exception of points in a set N for which $\lambda(g(N)) = 0$ and suppose that $\int_{\mathbb{R}\setminus N} |g'(x)|\,dx < \infty$. Show that, for every $\varepsilon > 0$, there is a $\delta > 0$ so that any sequence of nonoverlapping intervals $\{[c_n, d_n]\}$ for which $\sum_n \lambda([c_n, d_n]) < \delta$ it follows that*

$$\sum_n |\Delta g([c_n, d_n])| < \varepsilon.$$

Answer □

6.11 Local Lebesgue integrability conditions

A measurable function f is Lebesgue integrable on an interval $[a,b]$ provided that the integral $\overline{\int_a^b}|f(x)|\,dx$ is finite. If the integral is not finite then f cannot be Lebesgue integrable on $[a,b]$. But need it be Lebesgue integrable on some subinterval? The theorem we now prove gives a sufficient condition in order for an measurable functions to have a local integrability property. In the theorem we use the following notation for a function f and a closed set E: the function f_E is defined as $f_E(x) = f(x)$ whenever $x \in E$ and $f_E(x) = 0$ otherwise.

Theorem 6.39 *Let E be a nonempty closed subset of $[a,b]$ and f an measurable function. Suppose that*

$$-\infty < \underline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f(x)\,dx < \infty.$$

Then E contains a portion $E \cap (c,d)$ so that f_E is Lebesgue integrable on $[c,d]$.

Proof. We make a simplifying assumption that allows a small technical detail later. We remove from the set E all points that are isolated on either the right side or the left side or both sides. There are only countably many such points and that does not influence either measure or integration statements. While the resulting set is not closed, it is a set of type $\mathcal{G}_\delta$ so that we may still apply the Baire-Osgood theorem to it.

Choose t so that

$$-t < \underline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f(x)\,dx < t$$

and a full cover β of $[a,b]$ so that[2]

$$\left|\sum_\pi f\lambda\right| < t$$

for all partitions $\pi \subset \beta$ of $[a,b]$. Let $[c,d]$ be any subinterval and let $\pi \subset \beta$ be a partition of $[c,d]$. Choose $\pi' \subset \beta$ so that it consists of a partition of $[a,c]$ and $[d,b]$. Then

$$\left|\sum_{\pi\cup\pi'} f\lambda\right| < t$$

so that

$$\left|\sum_\pi f\lambda\right| \leq t + \left|\sum_{\pi'} f\lambda\right|$$

In particular we can write

[2] We are using $\sum_\pi f\lambda$ to denote the sum $\sum_{(I,x)\in\pi} f(x)\lambda(I)$ in this proof as many such sums will be considered.

$$T(c,d) = \sup\left\{ \left|\sum_{\pi} f\lambda\right| : \pi \subset \beta \text{ is a partition of } [c,d] \right\} < \infty.$$

We need a decomposition argument for β similar to that in Section 2.2.3. Choose $\delta(x) > 0$ so that $x \in I \subset [a,b]$ and $\lambda(I) < 2\delta(x)$ requires $(I,x) \in \beta$. Define

$$E_n^+ = \{x \in E : \delta(x) > 1/n,\ 0 \le f(x) \le n\}$$

and

$$E_n^- = \{x \in E : \delta(x) > 1/n,\ 0 \ge f(x) \ge -n\}.$$

This sequence of sets exhausts the set E so that, by the Baire-Osgood theorem, there must be a portion of E so that one of the sets is dense there. Thus we are able to choose an integer m and a subinterval $[c,d]$ so that $d-c < 1/m$ and so that E_m^+ (say) is dense in the nonempty portion $E \cap (c,d)$.

We shall investigate the Lebesgue integrability of f_E on $[c,d]$. For that, let π be an arbitrary partition of $[c,d]$ chosen from β. We shall estimate

$$\sum_{\pi} f_E^+ \lambda \quad \text{and} \quad \sum_{\pi} f_E^- \lambda$$

(where, as usual, f_E^+ and f_E^- denote the positive and negative parts of f_E).

Define $\pi_1 = \pi[E]$ and $\pi_2 = \pi \setminus \pi_1$. We alter π_1 in two different ways. The first alteration denoted as π_1' will replace each $(I,x) \in \pi_1$ by (I,x') where $x' \in E_m^+$. Since $x \in E$ and is not isolated on either side in E, and since E_m^+ is dense in this portion of E, such points are available. For any such point x' we see that the pair $(I,x') \in \beta$ because $\lambda(I) < 1/m < \delta(x')$. The second alteration denoted as π_1'' will replace each $(I,x) \in \pi_1$ for which $f(x) < 0$ by (I,x'') where $x' \in E_m^+$. For the same reasons as before, the pair $(I,x'') \in \beta$. We will make use of the fact that, for the adjusted points x' and x'', we have the inequalities $0 \le f(x') \le m$ and $f(x) < 0 \le f(x'')$.

Now we do our computations:

$$\left|\sum_{\pi_1 \cup \pi_2} f\lambda\right| \le T(c,d) \tag{6.4}$$

$$\left|\sum_{\pi_1' \cup \pi_2} f\lambda\right| \le T(c,d) \tag{6.5}$$

$$\left|\sum_{\pi_1'} f\lambda\right| \le m(d-c) \le 1. \tag{6.6}$$

Combining (6.5) and (6.6) we see that

$$\left|\sum_{\pi_2} f\lambda\right| \le T(c,d) + 1 \tag{6.7}$$

Thus we can estimate

$$\sum_{\pi} f_E^+\lambda = \sum_{\pi_1} f_E^+\lambda = \sum_{\pi_1''} f_E^+\lambda \le \sum_{\pi_1''} f\lambda$$

$$\le \sum_{\pi_1''} f\lambda + \left[\sum_{\pi_2} f\lambda + T(c,d) + 1\right] = \sum_{\pi_1''\cup\pi_2} f\lambda + T(c,d) + 1 \le 2T(c,d) + 1.$$

As such sums have this upper bound we can conclude that

$$\overline{\int_c^d} f_E^+(x)\,dx$$

is finite and hence that the measurable function f_E^+ is Lebesgue integrable on $[c,d]$.

Now we show that f_E^- is also Lebesgue integrable on $[c,d]$. Since

$$f_E^-(x) = f_E^+(x) - f(x)$$

for every $x \in E$, we find that

$$\sum_{\pi} f_E^-\lambda = \sum_{\pi_1} f_E^-\lambda = \sum_{\pi_1} f_E^+\lambda - \sum_{\pi_1} f\lambda$$

$$= \left[\sum_{\pi_1} f_E^+\lambda + \sum_{\pi_2} f_E^+\lambda\right] - \sum_{\pi_1} f\lambda$$

$$\le [2T(c,d)+1] - \sum_{\pi_1} f\lambda - \left[\sum_{\pi_2} f\lambda - T(c,d) - 1\right]$$

$$= [3T(c,d)+2] - \sum_{\pi} f\lambda \le 4T(c,d) + 2.$$

Once again such sums have this upper bound we can conclude that the measurable function f_E^- is Lebesgue integrable on $[c,d]$. Finally then $f_E = f_E^+ + f_E^-$ too must be Lebesgue integrable on $[c,d]$. This gives us our portion $E \cap (c,d)$ and completes the proof. ∎

6.12 Continuity of upper and lower integrals

The indefinite integral of an integrable function is continuous. We can express this by saying that, if f is integrable on a compact interval $[a,b]$, then for every $\varepsilon > 0$ there is a $\delta > 0$ so that

$$-\varepsilon < \int_c^d f(x)\,dx < \varepsilon$$

for every subinterval $[c,d] \subset [a,b]$ for which $\lambda([c,d]) < \delta$. We wish a version of this that does not assume integrability and that can be used for a characterization.

Definition 6.40 *A function f is said to have* continuous upper and lower integrals *on a compact interval $[a,b]$ if for every $\varepsilon > 0$ there is a $\delta > 0$ so that*

$$-\varepsilon < \underline{\int_c^d} f(x)\,dx \leq \overline{\int_c^d} f(x)\,dx < \varepsilon$$

for every subinterval $[c,d] \subset [a,b]$ for which $\lambda([c,d]) < \delta$.

Lemma 6.41 *Suppose that $f : [a,b] \to \mathbb{R}$ has continuous upper and lower integrals on a compact interval $[a,b]$. Then*

$$-\infty < \underline{\int_c^d} f(x)\,dx \leq \overline{\int_c^d} f(x)\,dx < \infty$$

for every subinterval $[c,d] \subset [a,b]$.

Proof. There must be a $\delta > 0$ so that

$$-1 < \underline{\int_c^d} f(x)\,dx \leq \overline{\int_c^d} f(x)\,dx < 1$$

for every subinterval $[c,d] \subset [a,b]$ for which $\lambda([c,d]) < \delta$. Subdivide

$$a = a_0 < a_1 < \cdots < a_{n-1} < a_n = b$$

in such a way that each $a_i - a_{i-1} < \delta$. Then compute, using Exercise 186, that

$$\overline{\int_a^b} f(x)\,dx = \sum_{i=1}^{n} \overline{\int_{a_{i-1}}^{a_i}} f(x)\,dx \leq n < \infty.$$

A similar argument handles the lower integral. ■

Exercises

Exercise 365 (Cauchy extension property) *Let f be integrable on every subinterval $[c,d] \subset (a,b)$. Show that f is integrable on $[a,b]$ if and only if if f has continuous upper and lower integrals on $[a,b]$.* Answer □

Exercise 366 (Harnack extension property) *Let $F : \mathbb{R} \to \mathbb{R}$, let E be a closed subset of $[a,b]$, and let $\{(a_i,b_i)\}$ be the sequence of intervals complementary to E in (a,b). Suppose that*

1. *$f(x) = 0$ for all $x \in E$,*
2. *f is integrable on all intervals $[a_i,b_i]$, and*
3. $\displaystyle\sum_{i=1}^{\infty} \sup_{a_i \leq c_i < d_i \leq b_i} \left| \int_{c_i}^{d_i} f(x)\,dx \right| < \infty.$

Show that f is integrable on $[a,b]$ and
$$\int_a^b f(x)\,dx = \sum_{i=1}^{\infty}\int_{a_i}^{b_i} f(x)\,dx.$$

Answer □

6.13 A characterization of the integral

The class of Lebesgue integrable functions on an interval $[a,b]$ can be characterized as those measurable functions f for which
$$\int_a^b |f(x)|\,dx < \infty.$$
We now show that the full class of integrable functions (absolutely or nonabsolutely) on an interval $[a,b]$ can be characterized as those measurable functions that have continuous upper and lower integrals.

Theorem 6.42 *A function f is integrable on $[a,b]$ if and only if f is measurable and f has continuous upper and lower integrals on $[a,b]$.*

Proof. We already know that an integrable function has these properties. Conversely suppose that f is measurable and that f has continuous upper and lower integrals on $[a,b]$. An open interval $(s,t) \subset (a,b)$ will be called "accepted" if f is integrable on every $[c,d] \subset (s,t)$. Let G be the union of all accepted intervals. This is an open subset of (a,b). Note that, if $[c,d] \subset G$, then by the Heine-Borel property $[c,d]$ can be written as the union of a finite collection of intervals $\{[c_i,d_i]\}$ each of which is inside an accepted interval. It follows that f is integrable on $[c,d]$ too.

Let
$$G = \bigcup_{i=1}^{\infty}(a_i,b_i),$$
displaying G as a union of its component intervals. We claim first that f must be integrable on each of the compact intervals $[a_i,b_i]$. This follows directly from the Cauchy extension property (Exercise 365) using the hypothesis that f has continuous upper and lower integrals. We shall use a single function F to represent the indefinite integral of f on each of these intervals, but we are cautioned not to use F outside of the intervals.

In particular if $G = (a,b)$ then the proof is completed since then f must be integrable on $[a,b]$ as required. Suppose not, i.e., that the theorem fails and $G \neq (a,b)$. Then $E = [a,b] \setminus G$ is a nonempty closed set. Note that E can have no isolated points. Indeed if $c \in E$ is isolated then $(c-t,c) \subset G$ and $(c,c+t) \subset G$ for some $t > 0$ and another application of the Cauchy extension property would show that $(c-t,c+t)$ is accepted so that $(c-t,c+t) \subset G$ which is not possible.

The goal of the proof now will be to obtain a portion $E\cap(c',d')$ of E with the property that (c',d') is accepted, which would be impossible. Portions cannot be empty and no point of E would be allowed to belong to an accepted interval. The local integrability Theorem 6.39 and the Harnack extension property (Exercise 366) will play key roles.

The assumption that f satisfies the continuity condition in Definition 6.40 together with Lemma 6.41 shows that the upper and lower integrals of f are finite. Thus, we can apply Theorem 6.39 to find a portion $E\cap[c,d]$ so that f_E is Lebesgue integrable on $[c,d]$.

Since f has continuous upper and lower integrals on $[c,d]$ it follows from Lemma 6.41 that

$$-\infty < \underline{\int_c^d} f(x)\,dx \le \overline{\int_c^d} f(x)\,dx < \infty.$$

Since f_E is Lebesgue integrable on $[c,d]$ it follows that

$$\int_c^d |f_E(x)|\,dx < \infty.$$

Thus we can select a real number $M>0$ and a full cover β of $[c,d]$ so that for any partition π of $[c,d]$ from β both

$$\left|\sum_\pi f\lambda\right| < M$$

and

$$\sum_\pi |f_E|\lambda < M.$$

We need a decomposition argument for β similar to that in Section 2.2.3. Choose $\delta(x)>0$ so that $x\in I\subset[a,b]$ and $\lambda(I)<2\delta(x)$ requires $(I,x)\in\beta$. Define

$$E_n = \{x\in E\cap[c,d] : \delta(x) > 1/n\}.$$

This sequence of sets exhausts the set $E\cap[c,d]$ so that, by the Baire-Osgood theorem, there must be a portion of so that one of the sets is dense there. Let us agree that E_m is dense in $E\cap(c',d')$ and that $[c',d']$ is smaller in length than $1/m$. Let $\{(c_i,d_i)\}$ denote the component intervals of $(c',d')\setminus E$. There must be infinitely many such component intervals since otherwise it would follow that f is integrable on $[c',d']$. We claim that

$$\sum_{i=1}^{\infty} \omega F([c_i,d_i] = \infty. \tag{6.8}$$

For, if not, then the Harnack extension property (Exercise 366) shows that $f-f_E$ must be integrable on $[c',d']$ and hence f is integrable there. But that contradicts the fact that $[c',d']$ must contain points of E.

From the continuity of F we know that

$$\omega F([c_i,d_i] = |F(s) - F(t)| \tag{6.9}$$

for some subinterval $[s,t] \subset [c_i,d_i]$. Consequently we may choose a sequence of intervals $\{[s_k,t_k]\}$, chosen from different component intervals $[c_i,d_i]$ in such a way that either

$$0 \leq \sum_{k=1}^{\infty} F(t_k) - F(s_k) = \infty \tag{6.10}$$

or

$$0 \geq \sum_{k=1}^{\infty} F(t_k) - F(s_k) = -\infty. \tag{6.11}$$

Let us assume the former. If (6.11) holds instead the same argument with a slight adjustment in the inequalities will work.

Now we fix an integer p and carefully construct a partition π of the interval $[c,d]$ from β. The first step is to choose π' from β to form a partition of $[c,c']$, then π'' from β to form a partition of $[d',d]$. For each of the intervals $\{[s_k,t_k]\}$ for $k = 1,2,3\dots,p$ we select a partition π_k of $[s_k,t_k]$ in such a way that

$$|F(t_k) - F(s_k) - \sum_{\pi_k} f\lambda| < 2^{-k}. \tag{6.12}$$

This is possible since f is integrable on each such interval and F is an indefinite integral. To complete the partition we take the remaining intervals, not yet covered by

$$\pi' \cup \pi'' \cup \bigcup_{k=1}^{p} \pi_k.$$

There are only finitely many of these intervals, say I_1, I_2, $\dots$, I_q. Each is a subinterval of $[c',d']$ and each one contains many points of E; thus each one also contains a point of E_m. Select a point x_i from $E_m \cap I_i$ ($i = 1,2,\dots,q$) and note that (I_i,x_i) belongs to β. Thus if we set

$$\pi''' = \{(I_i,x_i) : i = 1,2,\dots,q\}$$

then we have obtained a partition

$$\pi = \pi' \cup \pi'' \cup \pi''' \cup \bigcup_{k=1}^{p} \pi_k$$

of the interval $[c,d]$ that is contained in β.

Consequently, by the way in which we chose M and β,

$$\left|\sum_{\pi} f\lambda\right| \leq M.$$

We know too that

$$\left|\sum_{\pi'''} f\lambda\right| \leq \sum_{\pi'''} |f_E|\lambda \leq M.$$

We combine these inequalities with (6.12) and the simple inequality

$$\sum_{k=1}^{p} 2^{-k} \leq 1$$

to obtain

$$\sum_{k=1}^{p} F(t_k) - F(s_k) \leq \left| \sum_{\pi' \cup \pi''} f\lambda \right| + 2M + 1.$$

This is true for any p and conflicts with our assumption that the inequality (6.10) holds.

Since neither inequality (6.10) nor (6.11) can hold it follows that inequality (6.9) also fails, thus f is integrable on $[c', d']$. In other words (c', d') is accepted, which would be impossible. This completes the proof. ■

6.14 Denjoy's program

For nonabsolutely integrable functions the integral is not constructive by any of the methods of Lebesgue. Even the classical Newton integral is nonconstructive in a serious way.

If we know in advance that $F'(x) = f(x)$ everywhere, then certainly we can "construct" the value of the integral by using the formula

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

But even if we are assured that f is a derivative of some function, but we are not provided that function itself, then there may be no constructive method of determining either the value of the integral or the antiderivative function itself. This may surprise some calculus students since much of an elementary course is devoted to various methods of *finding* antiderivatives.

> "The ... solution to the primitive problem might go something like this: First, get the right function. Then, show it has the derivative you were searching for.
>
> Now try to ignore the fact that this ... is probably the most powerful method of integration known and, in practice, has successfully computed more antiderivatives than all other solutions combined. Consider instead how hard it is to "guess" the antiderivative. For example, try to find the integral of $\sin x/x$. Well, ok you might be able to "guess" some sort of infinite series solution. But things get worse than this. Dougherty and Kechris [26] have shown (using Y. Matiyasevich's work on diophantine representation of recursively enumerable sets ...) that there are derivatives which are analytically expressible (in terms of an explicit formula using the basic elementary functions sin, cos, exponents, absolute values, etc., and the elementary operations of multiplication, division, composition, and infinite

sums) but whose primitive is immensely complicated, so that for example, there is no way to analytically express the primitive."

—Chris Freiling, "HOW TO COMPUTE ANTIDERIVATIVES," BULLETIN of SYMBOLIC LOGIC, Vol. 1, No. 3 (1995).

After Lebesgue's constructive integral was presented there still remained this problem. All *bounded* derivatives can be handled by his methods, but there exist unbounded derivatives that are nonabsolutely integrable. What procedure (outside of our formal integration theory) would handle these?

Starting with the class of absolutely integrable functions, Arnaud Denjoy discovered in 1912 that a series of extensions of this class could be constructed that would eventually encompass all derivatives and, indeed, all nonabsolutely integrable functions. He called his process totalization. Added to Lebesgue's methods, totalization reveals exactly how constructive our integral is. His process completely catalogs the class of nonabsolutely integrable functions. In effect the integral that is discussed in this text could be (and has been) called *the Denjoy integral*.

6.14.1 Integration method

If we wish to construct an integral, as distinct from describing one in a nonconstructive manner [as we did here for the Newton integrals and the Henstock-Kurzweil integral] we need a language that will help outline the procedure. We have already accumulated a few constructive ideas in this text which we should review first.

We might use the symbol $\mathcal{J}_s$ to denote a starting point in integration theory that uses just the step functions. The integral of any step function is easy to compute as a finite sum. By taking uniform limits of step functions we can compute the integral of all regulated functions. We might write $\mathcal{J}_{su}$ to indicate the uniform limit step. Then the integral of any regulated function is available as a limit of some sequence of integrals, each of which we can compute as a finite sum.

That still does not go far enough. So we might consider monotone nondecreasing limits (as we did in Section 4.14). The symbol $\mathcal{J}_{su\uparrow}$ would then indicate this new procedure. Naturally we can then consider monotone nonincreasing limits. The symbol $\mathcal{J}_{su\uparrow\downarrow}$ could then indicate this new procedure. In fact, we know that we have by now arrived at a constructive procedure in four steps that encompasses all Lebesgue integrable functions.

Can we go further? Additional uniform limits or additional monotone limits do not profess us any further, but there are other constructive extension procedures that do. To describe them we adopt the following notation and language.

Definition 6.43 *By an* integration method *we shall mean a class $\mathcal{J}$ of locally integrable*[a] *functions $f:\mathbb{R}\to\mathbb{R}$ with these properties.*

1. *If $f:\mathbb{R}\to\mathbb{R}$ is zero everywhere then f belongs to $\mathcal{J}$.*
2. *If $f\in\mathcal{J}$ and I is a compact interval then $f\chi_I$ also belongs to $\mathcal{J}$.*
3. *If $f:\mathbb{R}\to\mathbb{R}$ is a function for which $f\chi_{[a,b]}$ and $f\chi_{[b,c]}$ both belong to $\mathcal{J}$, then so too does $f\chi_{[a,c]}$.*

[a] i.e., functions that are integrable on every compact interval $[a,b]$.

Although it is not part of the definition, our interest is in integration methods (i.e., classes of locally integrable functions) for which the integral is somehow constructible. For example, the class $\mathcal{R}$ of all locally Riemann integrable functions and the class $\mathcal{L}$ of all locally Lebesgue integrable functions are of this type and can be considered constructive, as too is the regulated integral of Chapter 1. Each of the varieties of Newton integral would be an integration method according to this definition, but could not be described as constructive.

Definition 6.44 *Let $f:\mathbb{R}\to\mathbb{R}$ and let $\mathcal{J}$ be an integration method. Then a point x_0 is said to be an $\mathcal{J}$-singular point of f if there exist arbitrarily small compact intervals I containing the point x_0 such that $f\chi_I$ does not belong to $\mathcal{J}$.*

The set of all $\mathcal{J}$-singular points of a function f is clearly closed. Moreover, if I is a compact interval that contains no singular points of f then necessarily $f\chi_I$ must belong to $\mathcal{J}$. (This is Exercise 367.)

Exercise 367 *Let $f:\mathbb{R}\to\mathbb{R}$ and let $\mathcal{J}$ be an integration method. Suppose that I is a compact interval that contains no $\mathcal{J}$-singular points of f. Show that $f\chi_I$ must belong to $\mathcal{J}$.* Answer □

6.14.2 Cauchy extension

Suppose that we are given an integration method $\mathcal{J}$. We construct a larger class $\mathcal{J}^C$ of integrable functions from $\mathcal{J}$ by the following extension process usually attributed to Cauchy. His extension process is normally applied to the class $\mathcal{R}$ of all locally Riemann integrable functions. One writes $\mathcal{R}^C$ for the extended class of all functions that have "improper" Riemann integrals on each compact interval. Calculus students will certainly remember the definition. For readers not schooled in such matters $\mathcal{R}^C$ is defined by the following general definition.

Definition 6.45 *Let $\mathcal{J}$ be an integration method. Then $\mathcal{J}^C$ denotes the class of all locally integrable functions $f:\mathbb{R}\to\mathbb{R}$ for which each $\mathcal{J}$-singular point of f is isolated in the set of singular points.*

One needs to be sure that the class of functions $\mathcal{J}^C$ so defined has all three properties of Definition 6.43. This is easy to show.

The definition could be formulated (and should be formulated) without reference to the entire class of locally integrable functions. We prefer to express it as a lemma.

Lemma 6.46 *Let $\mathcal{J}$ be an integration method. Then a function $f:\mathbb{R}\to\mathbb{R}$ belongs to the class $\mathcal{J}^C$ if and only if*

1. *The set of $\mathcal{J}$-singular points of f is finite in any compact interval $[a,b]$.*
2. *For each compact interval $[a,b]$ if a or b is a singular point but (a,b) contains no singular points then the limit*
$$\lim_{s\to 0+,t\to 0+}\int_{a+s}^{b-t} f(x)\,dx$$
exists.

The Cauchy extension property from Exercise 365 shows us that any function satisfying these two properties would have to be locally integrable and that the value of the integral of any f in $\mathcal{J}^C$ can, in all cases, be constructed from the values of the integrals of functions in $\mathcal{J}$. For example if $[a,b]$ contains no singular points then we already know the value of the integral $\int_a^b f(x)\,dx$ because $f\chi_{[a,b]}$ is in $\mathcal{J}$. On the other hand if either endpoint of $[a,b]$ is a singular point and (a,b) contains no singular points then $f\chi_{[a,b]}$ is not in $\mathcal{J}$, but the value of the integral is determined from
$$\int_a^b f(x)\,dx = \lim_{s\to 0+,t\to 0+}\int_{a+s}^{b-t} f(x)\,dx.$$
Each of these integrals is determined by $\mathcal{J}$ because the interval $[a+s,b-t]$ contains no singular points. Similar arguments show that the integral of any f in $\mathcal{J}^C$ on any compact interval is computable from the integrals of functions in $\mathcal{J}$.

Thus not only is the class $\mathcal{J}^C$ constructible directly from the class $\mathcal{J}$ but the value of the integral of any f in $\mathcal{J}^C$ can be constructed from the values of the integrals of function in $\mathcal{J}$.

6.14.3 Harnack extension

There is one more extension procedure that we shall apply to an integration method $\mathcal{J}$. We construct a larger class $\mathcal{J}^{CH}$ of integrable functions from $\mathcal{J}$ by the following extension process usually attributed to Harnack. This method (as we express it) includes the Cauchy extension as a special case. The Cauchy extension used finite sets of singular points; we extend this to certain closed sets of singular points and add some more assumptions.

If $f:\mathbb{R}\to\mathbb{R}$ and (a,b) is an open interval that contains no $\mathcal{J}$-singular points of f, then we recall that $f\chi_I$ belongs to $\mathcal{J}$ for any compact interval $I\subset(a,b)$.

Thus we can define, for any such interval (a,b),

$$\|f\|_{(a,b)} = \sup\left\{\left|\int_c^d f(x)\,dx\right| : [c,d] \subset (a,b)\right\}.$$

Definition 6.47 *Let $\mathcal{J}$ be an integration method. Then $\mathcal{J}^{CH}$ denotes the class of all locally integrable functions $f:\mathbb{R}\to\mathbb{R}$ for which*

1. *$f\chi_E$ is locally Lebesgue integrable, where E is the set of all $\mathcal{J}$-singular points.*

2. *For each compact interval $[a,b]$ the series*

$$\sum_{k=1}^{\infty} \|f\|_{(a,b)\cap(a_k,b_k)} < \infty$$

   where $\{(a_k,b_k)\}$ is the sequence of component intervals of the open set $\mathbb{R}\setminus E$.

Again one needs to be sure that the class of functions $\mathcal{J}^{CH}$ so defined has all three properties of Definition 6.43. This, too, is easy to show.

Note that the way we have defined this

$$\mathcal{J} \subset \mathcal{J}^{C} \subset \mathcal{J}^{CH}.$$

This is so because the conditions of the definition are trivially satisfied if the set E of all $\mathcal{J}$-singular points has no accumulation points. In that case $f\chi_E$ is certainly locally Lebesgue integrable (it is zero a.e.) and the series in the definition converges (there are only finitely many terms).

Again the definition can be formulated without reference to the entire class of locally integrable functions. We can express it as a lemma.

Lemma 6.48 *Let $\mathcal{J}$ be an integration method. Then a function $f : \mathbb{R} \to \mathbb{R}$ belongs to the class $\mathcal{J}^{CH}$ if and only if*

1. *$f\chi_E$ is locally Lebesgue integrable, where E is the set of all $\mathcal{J}$-singular points.*

2. *For each compact interval $[a,b]$ if a or b is in E but $(a,b) \cap E = \emptyset$ then the limit*
$$\lim_{s \to 0+, t \to 0+} \int_{a+s}^{b-t} f(x)\,dx$$
exists.

3. *For each compact interval $[a,b]$ the series*
$$\sum_{k=1}^{\infty} \|f\|_{(a,b)\cap(a_k,b_k)} < \infty$$
where $\{(a_k,b_k)\}$ is the sequence of component intervals of the open set $\mathbb{R} \setminus E$.

The Cauchy extension property from Exercise 365 and the Harnack extension property of Exercise 366 together show us that any function satisfying these four properties would have to be locally integrable. The value of the integral of any f in $\mathcal{J}^{CH}$ on any compact interval $[a,b]$ can be constructed from the values of the integrals of functions in $\mathcal{J}$.

For example (as before) if $[a,b]$ contains no singular points then we already know the value of the integral $\int_a^b f(x)\,dx$ because $f\chi_{[a,b]}$ is in $\mathcal{J}$. On the other hand if either endpoint of $[a,b]$ is a singular point and (a,b) contains no singular points then (again as before) $f\chi_{[a,b]}$ is not in $\mathcal{J}$, but the value of the integral is determined from
$$\int_a^b f(x)\,dx = \lim_{s \to 0+, t \to 0+} \int_{a+s}^{b-t} f(x)\,dx.$$
Each of these integrals is determined by $\mathcal{J}$ because the interval $[a+s, b-t]$ contains no singular points.

Let $[a,b]$ be an interval that contains points of E. Let $\{(a_k,b_k)\}$ be the sequence of component intervals of $(a,b) \setminus E$. We know that $f\chi_{[a_k,b_k]}$ is in $\mathcal{J}^C$ so we can compute the value of the integral
$$\int_{a_k}^{b_k} f(x)\,dx$$
from the values we know from $\mathcal{J}$. But we also know that
$$\sum_{k=1}^{\infty} \|f\|_{[a_k,b_k]} < \infty.$$

Consequently, by the Harnack extension property (Exercise 366),

$$\int_a^b f(x)\,dx = \int_a^b f(x)\chi_K(x)\,dx + \sum_{k=1}^{\infty}\int_{a_k}^{b_k} f(x)\,dx.$$

Thus the class $\mathcal{J}^{CH}$ is constructible directly from the class $\mathcal{J}$ and the value of the integral of any f in $\mathcal{J}^{CH}$ can be constructed from the values of the integrals of functions in $\mathcal{J}$.

6.14.4 Transfinite sequence of extensions of the integral

We can now describe the procedure used by Denjoy to provide a constructive analysis of the nonabsolutely integrable functions. We start with $\mathcal{L}$, the class of all locally Lebesgue integrable functions. We know that the integral for all functions in $\mathcal{L}$ can be constructed (either using measure theory or using monotone sequences of step functions). Consequently if we set

$$\mathcal{L}_0 = \mathcal{L} \text{ and } \mathcal{L}_1 = \mathcal{L}_0^{CH}$$

we obtain a larger class of locally integrable functions, a class whose integrals can be constructed from $\mathcal{L}$ using the Cauchy-Harnack procedure. Note that all functions in $\mathcal{L}_1 \setminus \mathcal{L}_0$ are nonabsolutely integrable on some compact interval. The procedure continues inductively setting

$$\mathcal{L}_{n+1} = \mathcal{L}_n^{CH} \quad (n = 0, 1, 2, 3, \dots).$$

The families

$$\mathcal{L} = \mathcal{L}_0 \subset \mathcal{L}_1 \subset \mathcal{L}_2 \subset \mathcal{L}_3 \subset \mathcal{L}_4 \dots$$

form larger and larger classes of locally integrable functions, each a proper extension of the one before. We can also define

$$\mathcal{L}_\omega = \bigcup_{n=1}^{\infty} \mathcal{L}_n$$

which, it can be shown, is larger than each of the members in the sequence. The Cauchy-Harnack extension procedure continues to produce proper extensions so that one can also define

$$\mathcal{L}_{\omega+1} = \mathcal{L}_\omega^{CH},\ \mathcal{L}_{\omega+2} = \mathcal{L}_{\omega+1}^{CH},\ \mathcal{L}_{\omega+3} = \mathcal{L}_{\omega+2}^{CH},\ \dots$$

in the same way. This process is made more formally correct by invoking the transfinite ordinal numbers.

6.14.5 The totalization process

A full account of the totalization process of Denjoy requires use of the countable ordinal numbers, briefly defined as follows. A well-ordered set is a totally ordered[3] set all of whose nonempty subsets have a minimal element. The existence of well-ordered uncountable sets is guaranteed only by appealing to some logical principle (usually Zorn's Lemma or the Axiom of Choice).

Definition 6.49 *The countable ordinals are defined to be an uncountable well-ordered set Ω such that the set of predecessors*

$$\{\eta \in \Omega : \eta \prec \xi\}$$

of any element $\xi \in \Omega$ is countable and every countable subset of Ω has an upper bound. The first element of Ω is labeled as 0.

For a proof that such a well-ordered set exists and for the properties of ordinals the interested reader is encouraged to study further. An excellent resource on this and other set-theoretic topics needed by analysts is K. Ciesielski [19]. But this brief description is enough for us to proceed.

Theorem 6.50 (Denjoy) *Let $\mathcal{L}$ denote the collection of all locally Lebesgue integrable functions and let $\mathcal{HK}$ denote the collection of all locally Henstock-Kurzweil integrable functions. Write, for countable ordinals ξ,*

$$\mathcal{L}_0 = \mathcal{L},\ \mathcal{L}_\xi = \left(\bigcup_{\eta \prec \xi} \mathcal{L}_\eta\right)^{CH},\ \text{and}\ \mathcal{L}_\Omega = \bigcup_{\xi \in \Omega} \mathcal{L}_\xi .$$

Then

$$\mathcal{HK} = \mathcal{L}_\Omega .$$

Proof. The way that we have defined the extension procedure CH guarantees that $\mathcal{HK} \supset \mathcal{L}_\Omega$. Thus we need to show that if f is locally integrable then f belongs to $\mathcal{L}_\xi$ for some countable ordinal ξ. Fix such a function f.

For each countable ordinal ξ we can write E_ξ for the closed set of $\mathcal{L}_\xi$-singular points of f. This gives us a decreasing transfinite sequence of closed sets $\{E_\xi\}$. There are three possibilities that occur to us:

1. $E_\xi = \emptyset$ for some ordinal ξ,
2. The sequence stabilizes, i.e., all $E_\xi = E \neq \emptyset$ for large enough ξ, or
3. Every member of the sequence is distinct.

The first possibility is the one that we want, since if $E_\xi = \emptyset$ for some ordinal ξ, then f must belong to $\mathcal{L}_\xi$. That would complete the proof.

Let consider possibility (3). If every member is distinct then we have produced a strictly descending uncountable and transfinite sequence of closed sets.

[3]A set X equipped with a partial order relation $\prec$ is totally ordered if for all distinct $x,\ y \in X$ either $x \prec y$ or $y \prec x$.

This is impossible. If there were such a sequence we could find a sequence of points $\{x_\xi\}$ so that $x_\xi \in E_\xi$ but $x_\xi \notin E_\eta$ for any ordinal η preceding ξ. Corresponding then to each such point we can find an open interval (a_ξ, b_ξ) with rational endpoints containing that point and not containing any points from E_η for any ordinal η preceding ξ. That would produce an uncountable collection of rational numbers, which is not possible.

Finally let us dispense as well with possibility (2). If the sequence stabilizes, i.e., all $E_\xi = E$ for large enough ξ, then we shall apply Theorem 6.39 to the nonempty closed set E. Since f is locally integrable, there is a nonempty portion $E\cap(a,b)$ satisfying the conclusion of that theorem. Thus $f\chi_{[a,b]}$ would necessarily belong to $\mathcal{L}_\xi{}^{CH}$ and hence to $\mathcal{L}_\eta$ for all large enough η. The interval (a,b) can't therefore contain any $\mathcal{L}_\eta$-singular points. But it should contain every point of $E\cap(a,b)$. Thus again we have an impossibility. ∎

Do we need all the countable ordinals? One might ask whether all of this transfinite induction is really necessary to capture the integral. Denjoy showed that even the classical Newton integral cannot be constructively so described by less than the totality of all the countable ordinals. More precisely, if $\mathcal{N}$ denotes the class of all functions locally integrable in the classical Newton sense (i.e., all derivatives) then while $\mathcal{N} \subset \mathcal{L}_\Omega$ there is no countable ordinal ξ for which $\mathcal{N} \subset \mathcal{L}_\xi$. For modern treatments and extensions of Denjoy's ideas see [26] and [30].

6.15 The Perron-Bauer program

The transfinite procedure of Denjoy solves the problem of inverting all derivatives, but in a way that was not at the time considered completely satisfactory. Transfinite methods were often used at the time, but it was invariably asked whenever such methods appeared whether a proof could also be constructed without an appeal to transfinite induction.

Thus it was natural that some authors took up the task of defining and studying an integral that would accomplish the same as the Denjoy integral, but in a much simpler way. It was this that led Perron [65][4] in 1914 and Bauer [4] [5] shortly after to propose an integral that would avoid most of the complexities of Denjoy's program.

[4]O. Perron, Über den Integralbelgriff, Sitzber', Heidelberg Akad. Wiss. Abt. A 16 (1914), 1–16.

[5]H. Bauer, Der Perronschen Integralbegriff und seine Beziehung zum Lebesgueschen, Monatsch. Math. Phys. 26(1915), 153–198.

6.15.1 Major and minor functions

The Perron-Bauer idea for capturing the classical Newton integral was to employ the method of major and minor functions. This method appears as well in other parts of mathematics (notably differential equations and potential theory). Applied to the primitive problem it gives a purely formal solution.

We should, perhaps, not be too quick to dismiss this, as the Henstock-Kurzweil method is also largely formal. We can neither construct major and minor functions, nor construct full covers that solve the primitive problem in general. On the other hand the Henstock-Kurzweil method has proven itself to offer a flexible and coherent account of integration theory. The Perron method is rather more awkward.

Definition 6.51 *Let $f : [a,b] \to \mathbb{R}$. A function $U : [a,b] \to \mathbb{R}$ is said to be a Perron-major function for f if*

$$\underline{D}U(x) \geq f(x)$$

at every point of the interval. Similarly $L : [a,b] \to \mathbb{R}$ is said to be a Perron-minor function for f if

$$\overline{D}F(x) \leq f(x)$$

at every point of the interval.

Exercise 369 shows that the major functions dominate the minor functions. For the purposes of developing properties of the Perron integral it is useful to allow exceptional sets. This does not change the integral but it does make some of the proofs easier to construct. Since we intend merely to show that the Perron integral is equivalent to our integral (i.e., the Henstock-Kurzweil integral) this is not a concern to us. Exercises 370–372 give some of the other variants that have appeared in the literature.

Exercise 368 *Suppose that $F : [a,b] \to \mathbb{R}$ is everywhere differentiable. Show that F is both a Perron-major function a Perron-minor function for F'.* □

Exercise 369 *Let $f : [a,b] \to \mathbb{R}$, let $U : [a,b] \to \mathbb{R}$ be a Perron-major function for f and let $L : [a,b] \to \mathbb{R}$ be a Perron-minor function for f. Show that*

$$[L(b) - L(a)] \leq [U(b) - U(a)].$$

Answer □

Exercise 370 *Let f be defined at every point of the interval $[a,b]$ with infinite values allowed. Suppose that $U, L : [a,b] \to \mathbb{R}$ satisfy*

$$-\infty \neq \underline{D}U(x) \geq f(x) \ \text{ and } \ +\infty \neq \overline{D}F(x) \leq f(x)$$

at every point of the interval. Show that

$$[L(b) - L(a)] \leq [U(b) - U(a)].$$

Answer □

Exercise 371 *Let f be defined at every point of the interval $[a,b]$ with infinite values allowed. Suppose that U, $L : [a,b] \to \mathbb{R}$ satisfy*

$$\underline{D}U(x) \geq f(x), \overline{D}F(x) \leq f(x) \text{ for a.e. } x \text{ and } -\infty < \underline{D}U(x) \leq \overline{D}F(x) < +\infty$$

for all but countably many points x in that interval. Show that

$$[L(b) - L(a)] \leq [U(b) - U(a)].$$

Answer □

Exercise 372 *Let f be defined at every point of the interval $[a,b]$ with infinite values allowed. Suppose that U, $L : [a,b] \to \mathbb{R}$ are continuous functions having σ-finite variation on $[a,b]$ and that they satisfy*

$$\underline{D}U(x) \geq f(x), \overline{D}F(x) \leq f(x)$$

for a.e. x. Suppose further that the set of values assumed by U at the points where $U'(x) = -\infty$ has measure zero, and that the set of values assumed by L at the points where $L'(x) = +\infty$ has measure zero. Show that

$$[L(b) - L(a)] \leq [U(b) - U(a)].$$

Answer □

6.15.2 Major and minor functions applied to other integrals

With this definition we can quickly establish the following two suggestive theorems, relating the existence of major and minor functions to the Lebesgue and to the Henstock-Kurzweil integrals.

There is an immediate connection between major and minor functions and upper and lower integrals.

Theorem 6.52 *Let $f : [a,b] \to \mathbb{R}$ and suppose that U is a Perron-major function for f and L is a Perron-minor function for f. Then*

$$[L(b) - L(a)] \leq \underline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f(x)\,dx \leq [U(b) - U(a)].$$

Proof. Let $\varepsilon > 0$. If U is a major function for f then the collection

$$\beta = \left\{([u,v],w) : \frac{U(v) - U(u)}{v - u} > f(w) - \varepsilon\right\}$$

is a full cover for the interval $[a,b]$. Take any partition $\pi \subset \beta$ of the interval $[a,b]$ and observe that

$$[U(b) - U(a)] = \sum_{[u,v],w)\in\pi} [U(v) - U(u)] > \left(\sum_{[u,v],w)\in\pi} f(w)(v-u)\right) - \varepsilon(b-a).$$

Since this is true for all such partitions π
$$\overline{\int_a^b} f(x)\,dx \leq [U(b) - U(a)] + \varepsilon(b-a).$$
The right-hand inequality in the theorem follows. The other inequality is similar. ∎

We can characterize the Lebesgue integral by the existence of absolutely continuous major and minor functions. This theorem follows easily from the Vitali-Carathéodory theorem (see Section 4.14.6).

Theorem 6.53 *Let $f : [a,b] \to \mathbb{R}$. Then f is Lebesgue integrable on $[a,b]$ if and only if, for every $\varepsilon > 0$, there are functions L, $U : [a,b] \to \mathbb{R}$ with these properties:*

1. *U is a Perron-major function for f.*
2. *L is a Perron-minor function for f.*
3. *U and L are absolutely continuous in the Vitali sense on $[a,b]$.*
4. *$[U(b) - U(a)] - [L(b) - L(a)] < \varepsilon$.*

If these conditions hold then
$$[U(b) - U(a)] - \varepsilon < \int_a^b f(x)\,dx < [L(b) - L(a)] + \varepsilon.$$

6.15.3 The Perron "integral"

We are now in a position to define the Perron integral and show that it is equivalent to the Henstock-Kurzweil integral. We have placed quotation marks about the word "integral" out of respect for Denjoy who did not consider that the Perron method deserved the credit of being called a method of integration and used every occasion to attack Perron's creation with enthusiasm.

Theorems 6.52 and 6.53 suggest that the following definition would be a plausible attempt to base an integration theory on the notion of major and minor functions. It ignores, however, the difficulty of producing even a single pair of major and minor functions except in special situations. Note that we have chosen to use *continuous* major and minor functions in this presentation.

Definition 6.54 *Let $f : [a,b] \to \mathbb{R}$. Then f is said to be Perron integrable on $[a,b]$ if for every $\varepsilon > 0$, there are functions $L, U : [a,b] \to \mathbb{R}$ with these properties:*

1. *U is a continuous Perron-major function for f.*
2. *L is a continuous Perron-minor function for f.*
3. *$[U(b) - U(a)] - [L(b) - L(a)] < \varepsilon$.*

If these conditions hold then the value of the integral is defined to be

$$(\mathcal{P}) \int_a^b f(x)\,dx = \inf\{[U(b) - U(a)] : U \text{ a continuous major function for } f\}$$
$$= \sup\{[L(b) - L(a)] : L \text{ a continuous minor function for } f\}.$$

Because of Exercise 368 and Theorems 6.52 and 6.53 we have immediately the following situation:

Theorem 6.55 *The Perron integral includes both the Lebesgue integral and the classical Newton integral and is itself included in the Henstock-Kurzweil integral.*

6.15.4 Hake-Alexandroff-Looman theorem

Hake in 1921 [1] used the constructive definition of Denjoy's integral to show that every Denjoy integrable function is also Perron integrable. Our proof below is identical. The opposite direction (showing that every Perron integrable function is Denjoy integrable) was obtained by P. Alexandroff [1;2] and (independently). Looman [4] For us this latter direction is more transparent since, as we already have seen, the equivalent Henstock-Kurzweil integral is easily shown to include the Perron integral.

Lemma 6.56 *Let $\mathcal{P}$ denote the collection of all locally Perron integrable functions. Then*

$$\mathcal{P}^{CH} = \mathcal{P}.$$

Proof. For the moment the reader is referred to Saks [73, pp. 247–250] for a proof. Lemma 3.1 on page 247 there handles the Cauchy extension and Lemma 3.4 on page 249 handles the Harnack extension. [A later version of the present manuscript may include some or all of these details; for now the reader is being sent back to Saks.] ∎

Theorem 6.57 (Hake-Alexandroff-Looman) *The Perron integral is equivalent to the Henstock-Kurzweil integral (and hence also to the Denjoy integral and the general Newton integral).*

Proof. Because of Theorem 6.55 we can write

$$\mathcal{L} \subset \mathcal{P} \subset \mathcal{HK}$$

and hence by Lemma 6.56, inductively, for all countable ordinals ξ,

$$\mathcal{L}_\xi \subset \mathcal{P}^{CH} = \mathcal{P}.$$

It follows from Theorem 6.50 that $\mathcal{HK} = \mathcal{L}_\Omega \subset \mathcal{P}$. Consequently $\mathcal{HK} = \mathcal{L}_\Omega = \mathcal{P}$ as stated. ∎

6.15.5 Marcinkiewicz theorem

Determining the integrability of a function by the Perron definition requires producing a multitude of major and minor functions. A remarkable observation of Marcinkiewicz [56] reveals that the finding of a single pair of continuous major and minor functions for a function f is sufficient to deduce the integrability of f. This observation supports Denjoy's attack on the Perron integral by revealing the essential difficulty of producing any major and minor functions in general.

Theorem 6.58 (Marcinkiewicz) *Let $f : [a,b] \to \mathbb{R}$ be a measurable function that has at least one continuous major function and one continuous minor function. The f is integrable on $[a,b]$.*

Proof. For the moment the reader is referred to Saks [73, p. 253] for a proof. The proof there shows that f would have to be Perron integrable on $[a,b]$ if such a pair of major and minor functions exist. We have already shown that "Perron integrable" and "integrable" are identical. See also Sarkhel [74] and Tolstov [81]. ∎

6.16 Integral of Dini derivatives

If F is a continuous function on an interval $[a,b]$ and has a finite Dini derivative, say $D_+F(x)$, at each point then f is determined up to an additive constant by that Dini derivative. One suspects that

$$F(x) - F(a) = \int_a^x D_+F(t)\,dt$$

but this is not necessarily true and even when it is true we need some further methods to handle.

6.16.1 Motivation

We require a variant, due to John Hagood [37], of the Cousin covering lemma that is more appropriate for handling the Dini derivatives of continuous functions.

This was exploited in Hagood and Thomson [38] to yield the integral discussed in this section.

Full covers are particularly suited to describing properties of the ordinary derivative. For example if $\underline{D}F(x) > r$ then the covering relation

$$\beta = \{(I,x) : \Delta F(I) > r\lambda(I)$$

has the property that for some $\delta > 0$, if $x \in I$ and $\lambda(I) < \delta$ then necessarily $(I,x) \in \beta$. Indeed $\underline{D}F(x) > r$ if and only if β has this property.

We conclude this chapter by determining how to recover a function from one of its Dini derivatives and so will require a one-sided analogue. The simplest version could come from the observation that $D_+F(x) > r$ if and only if the covering relation

$$\beta = \{(I,x) : \Delta F(I) > r\lambda(I)$$

has the property that for some $\delta > 0$, if $0 < h < \delta$ then necessarily $([x,x+h],x) \in \beta$.

But in fact our covering relation needs to be designed to handle the upper Dini derivative, not the lower. For that the description is more delicate: $D^+F(x) > r$ if and only if the covering relation

$$\beta = \{(I,x) : \Delta F(I) > r\lambda(I)\}$$

has the property that for any $\varepsilon > 0$, there is at least one value of h with $0 < h < \varepsilon$ for which $([x,x+h],x) \in \beta$. We strengthen this by insisting that F is continuous. In that case, if we found h so that

$$\frac{F(x+h) - F(x)}{(x+h) - x} > r,$$

notice that there must be a $\delta > 0$ so that

$$\frac{F(x'+h) - F(x)}{(x'+h) - x} > r$$

for every value of x' in the interval $[x-\delta, x]$.

Definition 6.59 *Let K be a compact set with endpoints $a = \inf K$ and $b = \sup K$. A covering relation β is said to be a* quasi-Cousin cover *of K provided that*

1. *There is at least one pair $([a,d],a) \in \beta$ with $a < d \le b$.*
2. *For every $a < x < b$, $x \in K$ there is a $\delta > 0$ so that there is at least one $x < d \le b$ for which all pairs $([c,d],x) \in \mathcal{S}$ whenever $x - \delta < c \le x$.*
3. *There is a $\delta > 0$ so that all pairs $([c,b],b) \in \beta$ whenever $b - \delta < c < b$.*

6.16.2 Quasi-Cousin covering lemma

Even though the notion of a quasi-Cousin cover is much weaker than that of a full cover the covering lemma generalizes.

Lemma 6.60 (Quasi-Cousin covering lemma) *Let β be a quasi-Cousin cover of a compact set K with endpoints $a = \inf K$ and $b = \sup K$. Then β contains a subpartition π so that*

$$K \subset \bigcup_{(I,x)\in\pi} I \subset [a,b].$$

Proof. Let us assume first that $K = [a,b]$. Let E be the set of all points z, with $b \geq z > a$ and with the property that β contains a partition π of $[a,z]$.

Argue that (i) $E \neq \emptyset$, (ii) if $\sup E = t$ then t cannot be less than b, (iii) if $\sup E = b$ then $b \in E$.

We know that (i) is true since there is at least one pair $([a,d],a) \in \beta$ with $a < d \leq b$ and so $d \in E$. Thus we may set $t = \sup E$ and be assured that $a < d \leq t \leq b$. To see (ii) note that it is not possible for $t < b$ for if so then there is a $\delta > 0$ and $d' > t$ for which all pairs $(t,[c,d']) \in \beta$ with $t - \delta < c \leq t$. But that supplies a point $t' \in (c-\delta,c] \cap E$ and the partition of $[a,t']$ can be enlarged by including $(t,[t',d'])$ to form a partition of $[a,d']$; thus $d' \in E$. But this violates $t = \sup E$.

Finally for (iii) if $t = b$ and yet $b \notin E$ then, repeating much the same argument, there is a $\delta > 0$ for which all pairs $(b,[c,b]) \in \beta$ with $b - \delta < c < b$. But that supplies a point $t' \in (b-\delta,b) \cap E$ and the partition for $[a,t']$ can be enlarged by including $(b,[t',b])$ to form a partition π for $[a,b]$. This shows that $b \in E$ after all.

Now let us handle the general case for an arbitrary compact set $K \subset [a,b]$. Let $G = (a,b) \setminus K$ and

$$\beta_1 = \{(I,x) : \ x \in I \text{ and } I \subset G\}.$$

Since β is a quasi-Cousin cover of K we can check that $\beta \cup \beta_1$ is a quasi-Cousin cover of $[a,b]$. By the first part of the proof there is a partition $\pi \subset \beta \cup \beta_1$ of $[a,b]$. Remove those elements of π that do not belong to β to form a subpartition with exactly the required properties. ∎

The proof contains explicitly the statement of the corollary:

Corollary 6.61 *Let β be a quasi-Cousin cover of a compact interval $[a,b]$. Then β contains a partition of $[a,b]$ (although not necessarily of other subintervals of $[a,b]$).*

Exercises

Exercise 373 (variant on the quasi-Cousin covering) *Let K be a compact set and β a covering relation. Suppose that, for each $x \in K$, there are s, $t > 0$ so that all pairs*

$$([x', x+s], x) \in \beta$$

whenever $x-t \le x' \le x$. *Show that* β *contains a subpartition* π *for which*

$$K \subset \bigcup_{(I,x)\in\pi} I.$$

□

Exercise 374 *Let* $f:\mathbb{R}\to\mathbb{R}$ *be continuous at each point of an open interval* (a,b) *and suppose that* $D^+f(x) > m$ *for each* $x\in(a,b)$. *Then* $f(d)-f(c) > m(d-c)$ *for each* $[c,d]\subset(a,b)$. □

6.16.3 Estimates of integrals from derivates

As a warm-up to our theorem about Dini derivatives let us show that the ordinary derivates are easily handled.

Lemma 6.62 *Let* $F, f : [a,b]\to\mathbb{R}$. *If* F *is continuous at* a *and* b *and*

$$\underline{D}F(x) \ge f(x)$$

at every point of (a,b), *then*

$$\overline{\int_a^b} f(x)\,dx \le F(b)-F(a).$$

Proof. Let $\varepsilon > 0$. Take the covering relation

$$\beta_1 = \{(I,x) : \Delta F(I) \ge (f(x)-\varepsilon)\lambda(I)\}$$

and

$$\beta_2 = \{(I,x) : x = a \text{ or } b,\ x\in I \text{ and } |\Delta F(I)| + |f(x)|\lambda(I) < \varepsilon\}.$$

Check that $\beta = \beta_1\cup\beta_2$ is a full cover of $[a,b]$. At the endpoints a or b the continuity of F needs to be used in the verification, while at the points in (a,b) the inequality $\underline{D}F(x) \ge f(x)$ is used.

Any partition $\pi\subset\beta$ of the interval $[a,b]$ will satisfy

$$\sum_{(I,x)\in\pi} f(x)\lambda(I) \le \sum_{(I,x)\in\pi} [\Delta F(I)+\varepsilon\lambda(I)] + 2\varepsilon = F(b)-F(a)+\varepsilon(2+b-a).$$

This is true for all partitions π from this β and all $\varepsilon > 0$ and so the conclusion that

$$\underline{\int_a^b} f(x)\,dx \le F(b)-F(a)$$

now follows. ∎

Lemma 6.63 *Let* $F, f : [a,b]\to\mathbb{R}$. *If* F *is continuous at* a *and* b *and*

$$\overline{D}F(x) \le f(x)$$

at every point of (a,b), *then*

$$\underline{\int_a^b} f(x)\,dx \ge F(b)-F(a).$$

Proof. Apply Lemma 6.62 to the functions $-F$ and $-f$. ∎

6.16.4 Estimates of integrals from Dini derivatives

For Dini derivatives there is a weaker version of Theorem 6.62 available using similar arguments (but employing quasi-Cousin covers as well as full covers). Note that this weaker version uses lower and upper rather than upper and lower integrals; in particular no corollary can be derived asserting the integrability of the Dini derivative (indeed it may not be integrable).

Theorem 6.64 *Suppose that $F : [a,b] \to \mathbb{R}$ is continuous and that g is a finite-valued function. If $D^+F(x) \geq g(x)$ at every point $a < x < b$, then,*

$$F(b) - F(a) \geq \underline{\int_a^b} g(x)\,dx. \tag{6.13}$$

If $D_+F(x) \leq g(x)$ at every point $a < x < b$, then

$$F(b) - F(a) \leq \overline{\int_a^b} g(x)\,dx. \tag{6.14}$$

Proof. Let $\varepsilon > 0$. Take the covering relation β_1 of all pairs $([x,y],z)$ with

$$\Delta F([x,y]) \geq (f(z) - \varepsilon)\lambda([x,y])$$

and β_2 of all pairs $([a,y],a)$ and $([x,b],b)$ for which

$$\Delta F([a,y]) - f(a)\lambda([a,y]) > -\varepsilon$$

and

$$\Delta F([x,b]) \geq f(b)\lambda([x,b]) - \varepsilon.$$

It is easy to verify that $\beta = \beta_1 \cup \beta_2$ is a quasi-Cousin cover of $[a,b]$. At the endpoints a or b the continuity of F needs to be used in the verification, while at the points in (a,b) the inequality $D^+F(x) \geq g(x)$ is used.

This may not seem too much of a help since the integral is defined by full covers, not by quasi-Cousin covers. But let β_3 be *any* full cover of $[a,b]$. Check that, as defined, $\beta_3 \cap \beta$ must be a quasi-Cousin cover of $[a,b]$. Thus there is at least one partition π from β_3 that is also in β. For that partition a familiar argument gives us

$$\sum_{(I,x)\in\pi} f(x)\lambda(I) \leq \sum_{(I,x)\in\pi} [\Delta F(I) + \varepsilon\lambda([x,y])] + 2\varepsilon = F(b) - F(a) + \varepsilon(2 + b - a).$$

Note that this means any full cover of $[a,b]$ contains at least one partition π with this property. Thus, while we can say nothing about the *upper* integral, we certainly can assert that the *lower* integral must always be lesser than $F(b) - F(a) + \varepsilon(2 + b - a)$ and from this the theorem follows. ∎

As a consequence of this theorem we observe that if an everywhere finite function g is *assumed* to be integrable on $[a,b]$ and lies between the two derivates then an integral identity holds. The assumption that g is integrable cannot be dropped here.

Corollary 6.65 *Let $F:[a,b]\to\mathbb{R}$ be continuous and $g:[a,b]\to\mathbb{R}$ be integrable on $[a,b]$ and suppose that*

$$D_+F(x) \le g(x) \le D^+F(x)$$

at every point x on $[a,b]$. Then

$$F(b)-F(a) = \int_a^b g(x)\,dx.$$

Exercises

Exercise 375 *Suppose that $F:\mathbb{R}\to\mathbb{R}$ is a continuous function and that $D^+F(x) > r$ at every point of an interval $[a,b]$. Verify that the covering relation*

$$\beta = \{(I,x) : \Delta F(I) > r\lambda(I)\}$$

satisfies the first two conditions (but not necessarily the third) in Definition 6.59.- □

Exercise 376 *Continuing the previous exercise, let $\varepsilon > 0$ and let*

$$\beta' = \{([x-t,x)],x) : |F(x-t)-F(x)| < \varepsilon\}.$$

Show that $\beta\cup\beta'$ is a quasi-Cousin cover of $[a,b]$. □

Exercise 377 *Show that every full cover of an interval $[a,b]$ is also a quasi-Cousin cover for any compact subset of $[a,b]$.* □

Exercise 378 *Let $f:\mathbb{R}\to\mathbb{R}$ and suppose that the function $D^+f(x)$ is finite-valued and continuous at a point x_0. Show that f is differentiable at x_0.* □

6.17 Appendix: Baire category theorem

Students working on the proof of Theorem 4.8 and some of the material of Chapter 6 will need to understand the Baire category theorem. Here is a full exposition suitable for most courses of instruction or review.

6.17.1 Meager sets

A set of real numbers is countable if it can be expressed as a countable union of a sequence of finite sets. If I is an interval and E is a countable set then $I\setminus E$ is dense in I.

This generalizes to meager sets. A of real numbers is *meager* if it can be expressed as a countable union of a sequence of nowhere dense sets. If I is an interval and E is a meager set then $I\setminus E$ is dense in I. The proof for meager sets and for countable sets is exactly the same, using the nested interval argument. For example: if E_n is a sequence of nowhere dense sets [finite sets] inside an interval I, then take any subinterval $(c,d)\subset I$. There must be a nested sequence of intervals with $[c_n,d_n]\subset I$ and $[c_n,d_n]\cap E_n=\emptyset$. There is a point that belongs to all of the intervals and that point fails to belong to $E=\bigcup_{n=1}^{\infty}E_n$. This shows that $I\setminus E$ is dense in I.

The complement of the meager set E is said to be *residual* in I. Residual sets are dense as we have just seen. This is usually described as the Baire category theorem.

6.17.2 Portions

If E is a closed set and (a,b) an open interval then

$$E\cap(a,b)$$

is called a *portion* of E provided only that $E\cap(a,b)\neq\emptyset$. It is possible that a portion could be trivial in that $E\cap(a,b)$ might contain only a single point of E; such a point is said to be an *isolated point* of E and we should be alert to the possibility that a portion might merely contain such a point.

6.17.3 Baire-Osgood Theorem

Our interest is in situations where $E, E_1, E_2, E_3, \dots$ is a sequence of closed sets and we wish to be assured that one of the sets E_n contains a portion of E. This requires a compactness argument; the nested interval property is particularly suited to this problem.

Exercise 379 *Suppose that E and E_1 are nonempty closed sets and that E_1 contains no portion of E. Then there must exist a portion*

$$E\cap(a,b)$$

so that $E_1\cap(a,b)=\emptyset$. Answer □

Exercise 380 *Suppose that E, E_1, E_2, $\dots$, E_n are nonempty closed sets and that*

$$E\subset\bigcup_{k=1}^{n}E_k.$$

Show that at least one of the sets E_k must contain a portion of E. Answer □

The Baire-Osgood theorem, one of the basic tools in advanced analysis, takes this exercise and extends the result to infinite sequences of closed sets.

Exercise 381 (Baire-Osgood Theorem) *Suppose that E, E_1, E_2, ..., E_n, ... are nonempty closed sets and that*

$$E \subset \bigcup_{k=1}^{\infty} E_k.$$

Then at least one of the sets E_k must contain a portion of E. Answer □

Exercise 382 *On occasions one will need this theorem without having to assume that E is closed. Show that theorem remains true if $E = \bigcap_{j=1}^{\infty} G_j$ where $\{G_j\}$ is some sequence of open sets.* Answer □

Exercise 383 *If the closed set E is contained in a sequence of sets $\{E_n\}$ but we cannot be assured that they are closed sets then a simple device is to replace them by their closures. [The closure of a set E is the set $\overline{E}$ defined as the smallest closed set containing E.] If we do this show that the conclusion of the theorem would have to be, not that some set E_n contains a portion of E, but that some set E_n is dense in a portion of E.* □

6.17.4 Language of meager/residual subsets

The exploitation of the Osgood-Baire theorem can often be clarified by using the language of meager and residual subsets. If E is a closed set[6] of real numbers then a meager subset is one that represents a "small," insubstantial part of E; what remains after a meager subset is removed would be called a residual subset. It would be considered a "large" subset since only an insubstantial part has been removed. Residual sets are dense, but more than dense. A countable intersection of residual sets would still be dense.

Definition 6.66 *Let E be a closed set. A subset A of E is said to be* a meager subset of E *provided that there exists a sequence of closed sets $\{E_n\}$ none of which contains a portion of E so that*

$$A \subset \bigcup_{n=1}^{\infty} E_n.$$

Definition 6.67 *Let E be a closed set. A subset A of E is said to be* a residual subset of E *provided that the complementary subset $E \setminus A$ is a meager subset of E.*

[6] In this section the language is restricted to subsets of closed sets. In view of Exercise 382 all of this would apply equally well to subsets of $\mathcal{G}_\delta$ sets, that is sets that are intersections of some sequence of open sets.

Chapter 7

Integration in $\mathbb{R}^n$

In this chapter we shall sketch a theory of integration for functions of several variables. This is just a sketch to illustrate that the methods developed in the text extend without too much trouble to higher dimensions. The reader is, by now, ready for a full treatment using any of the standard presentations but may find it convenient to see a rapid account extending some of our techniques here.

The exercises do the technical work and, for the most part we have been content to give references to where the techniques needed can be found. We consider this final chapter more of a guide to thinking about this subject and the exercises and discussions in the Answers section are more a dialogue than a course of study.

7.1 Some background

We must assume the reader is familiar with the rudiments of analysis in the space $\mathbb{R}^n$. In particular these facts will be used.

- $\mathbb{R}^n$ is the collection of all n-tuples of real numbers $x = (x_1, x_2, \dots, x_n)$.
- Addition in $\mathbb{R}^n$ is defined by
$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n).$$
- Scalar multiplication in $\mathbb{R}^n$ is defined by
$$r(x_1, x_2, \dots, x_n) = (rx_1, rx_2, \dots, rx_n).$$
- Distances in $\mathbb{R}^n$ are defined by
$$\|(x_1, x_2, \dots, x_n) - (y_1, y_2, \dots, y_n)\|$$
$$= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots (x_n - y_n)^2}.$$

- The open ball with center $x = (x_1, x_2, \dots, x_n)$ and radius r in $\mathbb{R}^n$ is
$$B(x;r) = \{y : \|x-y\| < r\}.$$

7.1.1 Intervals and covering relations

By a closed interval in $\mathbb{R}$ we mean, of course, the set
$$I = [a,b] = \{x : a \le x \le b\}.$$
That set has two endpoints and the interior is the open interval (a,b) between them. The symbol $|I|$ denotes the length of I, i.e., $|I| = b-a$.

By an interval in $\mathbb{R}^2$ we mean a product of two intervals in $\mathbb{R}$. Thus the closed rectangle
$$I = [a,b]\times[c,d] = \{(x,y) : a \le x \le b,\ c \le y \le d\}.$$
That set has four vertices, (a,c), (b,c), (b,c), and (b,d). The symbol $|I|$ denotes the area of I, i.e., $|I| = (b-a)(d-c)$ which is the product of the length and width of the rectangle.

These ideas and notation extend without difficulty to any dimension greater than two. By an *interval* in $\mathbb{R}^n$ we shall mean a cartesian product of one-dimensional intervals. It will be a closed interval if it is a product of closed intervals. Thus
$$I = [a_1,b_1]\times[a_2,b_2]\times\cdots\times[a_n,b_n]$$
is the set of points in $\mathbb{R}^n$ described by these inequalities:
$$\{(x_1,x_2,\dots,x_n) : a_1 \le x_1 \le b_1,\ a_2 \le x_2 \le b_2, \dots,\ a_n \le x_n \le b_n\}.$$
This interval has 2^n vertices. The symbol $|I|$ denotes the n-dimensional volume of I, i.e.,
$$|I| = (b_1-a_1)(b_2-a_2)(b_3-a_3)\dots(b_n-a_n)$$
which is the product of the length of all the edges in the interval.

Two intervals are *nonoverlapping* if their intersection has no interior points. Thus nonoverlapping intervals are either disjoint or else they meet only at some boundary points. A *packing* is a finite covering relation
$$\{(I_1,x_1),\ (I_2,x_2),\ (I_2,x_2),\dots,\ (I_k,x_k)\}$$
where each I_i is an interval and x_i is a point in the corresponding interval I_i, and distinct pairs of intervals I_i and I_j do not overlap.

By a *full interval cover* of a set $E \subset \mathbb{R}^n$ we mean a covering relation β that consists of pairs (I,x) again for which each I is an interval and x is a point in the corresponding interval I, and which is full in the following (by now familiar) sense: for each $x \in E$ there is a positive $\delta(x)$ so that β contains every pair (I,x) for which I is an interval containing x and contained in the open ball $B(x;\delta(x))$.

Exercise 384 (additivity of the volume) *Show that the n-dimensional volume is an additive interval function, i.e., show that if J is a closed interval in $\mathbb{R}^n$ and*

π a packing for which

$$J = \bigcup_{(I,x)\in\pi} I$$

then

$$|J| = \sum_{(I,x)\in\pi} |I|.$$

Answer □

Exercise 385 (Cousin's lemma) *Show that if β is a full interval cover of a closed interval J in $\mathbb{R}^n$ then there is a packing $\pi \subset \beta$ for which*

$$J = \bigcup_{(I,x)\in\pi} I.$$

Answer □

7.2 Measure and integral

The measure theory and the integration are defined by means of full interval covers and packings. This is the analogue of the Riemann sums expression that was available in dimension one for all of our integrals in the early chapters.

Definition 7.1 *Let $E \subset \mathbb{R}^n$ and let f be a nonnegative real-valued function defined on E. Then we define the upper integral*

$$\overline{\int_E} f(x)\,dx = \inf_{\beta} \sup_{\pi\subset\beta} \sum_{(I,x)\in\pi} f(x)|I|$$

where the supremum is with regard to all packings $\pi \subset \beta$ where β is an arbitrary full interval cover of E. We use also the notation

$$\mathcal{L}^n(E) = \overline{\int_E} dx$$

and refer to the set function $\mathcal{L}^n$ as Lebesgue measure *in $\mathbb{R}^n$.*

The reader might well have expected a higher dimensional integral to look more like the one-dimensional version. For example if $f : \mathbb{R}^2 \to \mathbb{R}$ perhaps we would expect an indefinite integral $F : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$F(x,y) = \int_a^x \int_b^y f(s,t)\,dsdt.$$

But the theory is far better expressed by the set function

$$E \to \int_E f(x)\,dx$$

and it is this idea and notation that we pursue.

Note that if E is a bounded set then the upper integral could have been simply stated as an interval function by noticing that

$$\overline{\int_I} f(x)\chi_E(x)\,dx = \overline{\int_E} f(x)\,dx$$

for every interval I that contains E. Thus the theory could have been developed by Riemann sums over partitions of intervals. We prefer to pass immediately to the set version $E \to \int_E f(x)\,dx$ which is closer to the mainstream of integration theory.

We shall not introduce a lower integral (as might be expected) but we will instead define what is meant by a $\mathcal{L}^n$-measurable set and a $\mathcal{L}^n$-measurable function. When E is a $\mathcal{L}^n$-measurable set and a f is a $\mathcal{L}^n$-measurable function then the Lebesgue integral

$$\int_E f(x)\,dx$$

will be defined to be the value

$$\overline{\int_E} [f(x)]^+\,dx - \overline{\int_E} [f(x)]^-\,dx$$

provided this has a meaning (i.e., is not $\infty - \infty$). Thus the upper integral will serve us only as a tool that leads quickly to a formal expression for the value of the Lebesgue integral and the Lebesgue measure.

7.2.1 Lebesgue measure in $\mathbb{R}^n$

We use the special notion

$$\mathcal{L}^n(E) = \overline{\int_E} dx$$

and refer to this as n-dimensional *Lebesgue [outer] measure* on $\mathbb{R}^n$. This is defined for all subsets E of $\mathbb{R}^n$ as is the upper integral

$$\overline{\int_E} f(x)\,dx$$

which is defined for all functions f that assign a nonnegative number at every point of the set E.

We shall discover that for intervals $\mathcal{L}^n(I) = |I|$ so that Lebesgue measure is an extension of the volume function from the class of closed intervals to the class of all subsets of $\mathbb{R}^n$. Some authors prefer to keep the same notation in which case $|E|$ is defined for all subsets of $\mathbb{R}^n$ as

$$|E| = \overline{\int_E} dx.$$

7.2.2 The fundamental lemma

The fundamental lemma that we need that describes the key property of the upper integral and the measure is the following, seen already in its one-dimensional version in Lemma 3.18. The same proof works here to give essentially the same conclusion.

Lemma 7.2 *Let $E \subset \mathbb{R}^n$ and let $f, f_1, f_2, f_3, \dots$ be a sequence of nonnegative real-valued functions defined on E. Suppose that*

$$f(x) \leq \sum_{k=1}^{\infty} f_k(x)$$

for every $x \in E$. Then

$$\overline{\int_E} f(x)\,dx \leq \sum_{k=1}^{\infty} \overline{\int_E} f_k(x)\,dx.$$

The two corollaries follow immediately and show that the set functions

$$E \to \overline{\int_E} dx$$

and

$$E \to \overline{\int_E} f(x)\,dx$$

are measures on $\mathbb{R}^n$ in the sense we make precise in Section 7.4 below.

Corollary 7.3 *Let $E, E_1, E_2, E_3, \dots$ be a sequence of subsets of $\mathbb{R}^n$. Suppose that*

$$E \subset \bigcup_{k=1}^{\infty} E_k.$$

Then

$$\mathcal{L}^n(E) \leq \sum_{k=1}^{\infty} \mathcal{L}^n(E_k).$$

Corollary 7.4 *Let $E, E_1, E_2, E_3, \dots$ be a sequence of subsets of $\mathbb{R}^n$. Suppose that*

$$E \subset \bigcup_{k=1}^{\infty} E_k$$

and that f is a nonnegative function defined at least on the set $\bigcup_{k=1}^{\infty} E_k$. Then

$$\overline{\int_E} f(x)\,dx \leq \sum_{k=1}^{\infty} \overline{\int_{E_k}} f(x)\,dx.$$

Exercise 386 *Show, for all intervals I in $\mathbb{R}^n$, that $\mathcal{L}^n(I) = |I|$.* □

Exercise 387 *Let f and g be nonnegative functions on a set $E \subset \mathbb{R}^n$ and such that $f(x) \leq g(x)$ for all $x \in E$. Show that*

$$\overline{\int_E} f(x)\,dx \leq \overline{\int_E} g(x)\,dx.$$

□

Exercise 388 *Let f be a nonnegative function on a set $E \subset \mathbb{R}^n$ and such that $r \leq f(x) \leq s$ for all $x \in E$ for some real numbers r and s. Show that*

$$r\mathcal{L}^n(E) \leq \overline{\int_E} f(x)\,dx \leq s\mathcal{L}^n(E)$$

□

Exercise 389 *Suppose that $E_1, E_2 \subset \mathbb{R}^n$ are separated by open sets, i.e., there is a disjoint pair of open sets G_1 and G_2 in $\mathbb{R}^n$ so that $E_1 \subset G_1$ and $E_2 \subset G_2$. Show that*

$$\overline{\int_{E_1 \cup E_2}} f(x)\,dx = \overline{\int_{E_1}} f(x)\,dx + \overline{\int_{E_2}} f(x)\,dx.$$

□

Exercise 390 *Suppose that $E_1, E_2 \subset \mathbb{R}^n$ are separated, i.e.,*

$$\inf\{\|e_1 - e_2\| : e_1 \in E_1,\ e_2 \in E_2\} > 0.$$

Show that

$$\overline{\int_{E_1 \cup E_2}} f(x)\,dx = \overline{\int_{E_1}} f(x)\,dx + \overline{\int_{E_2}} f(x)\,dx.$$

□

Exercise 391 *Suppose that $E_1, E_2 \subset \mathbb{R}^n$ are separated by open sets, i.e., there is a disjoint pair of open sets G_1 and G_2 in $\mathbb{R}^n$ so that $E_1 \subset G_1$ and $E_2 \subset G_2$. Show that*

$$\mathcal{L}^n(E_1 \cup E_2) = \mathcal{L}^n(E_1) + \mathcal{L}^n(E_2).$$

□

Exercise 392 *Show that*

$$\overline{\int_E} f(x)\,dx = 0$$

if and only if $f(x)$ is equal to zero for $\mathcal{L}^n$-almost every x in E. □

Exercise 393 *Show that*

$$\overline{\int_{E \cup N}} f(x)\,dx = 0$$

for any set N for which $\mathcal{L}^n(N) = 0$. □

7.3 Measurable sets and measurable functions

For the definition of measurability we can repeat our theory from earlier. We could choose to generalize to higher dimensions by taking any one of the characterizations of Corollary 4.30 and apply it in this setting. We choose here to take the simplest definition.

Later on in Section 7.4 we take another of the six characterizations of measurability in dimension one proved in that corollary.

Definition 7.5 *A subset E of $\mathbb{R}^n$ is said to be $\mathcal{L}^n$-measurable if for every $\varepsilon > 0$ there is an open set G with $\mathcal{L}^n(G) < \varepsilon$ and so that $E \setminus G$ is closed.*

With only minor changes in wording we can prove, using the methods we have already developed, that the usual properties of one-dimensional Lebesgue measure are enjoyed also by $\mathcal{L}^n$. Here is a fast summary.

- Let $E_1, E_2, E_3, \ldots$ be a sequence of pairwise disjoint $\mathcal{L}^n$-measurable subsets of $\mathbb{R}^n$ and write $E = \bigcup_{i=1}^{\infty} E_i$. Then, for any set $A \subset \mathbb{R}^n$,
$$\mathcal{L}^n(A \cap E) = \sum_{i=1}^{\infty} \mathcal{L}^n(A \cap E_i).$$
- The class of all $\mathcal{L}^n$-measurable subsets of $\mathbb{R}^n$ forms a Borel family that contains all closed sets and all $\mathcal{L}^n$-measure zero sets.
- If $E_1 \subset E_2 \subset E_3 \subset \ldots$ is an increasing sequence of subsets of $\mathbb{R}^n$ then
$$\mathcal{L}^n\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mathcal{L}^n(E_n).$$

7.3.1 Measurable functions

Definition 7.6 *Let E be a $\mathcal{L}^n$-measurable subset of $\mathbb{R}^n$ and f a real-valued function defined on E. Then f is said to be $\mathcal{L}^n$-measurable if*
$$\{x \in E : f(x) > r\}$$
is a $\mathcal{L}^n$-measurable subset of $\mathbb{R}^n$ for every real number r.

Definition 7.7 *Let E be a $\mathcal{L}^n$-measurable subset of $\mathbb{R}^n$ and f a $\mathcal{L}^n$-measurable function defined on E. Then the Lebesgue integral*
$$\int_E f(x)\,dx$$
is be defined to be the value
$$\overline{\int_E} [f(x)]^+\,dx - \overline{\int_E} [f(x)]^-\,dx$$
provided that both of these are not infinite. If both of these are finite then f is said to be Lebesgue integrable *on E and the integral $\int_E f(x)\,dx$ has a finite value.*

The key reason for this definition and for the restriction of the integration theory to measurable functions is the following fundamental additive property.

Theorem 7.8 *Let E be a $\mathcal{L}^n$-measurable subset of $\mathbb{R}^n$ and f, g be $\mathcal{L}^n$-measurable functions defined on E. Then*

$$\int_E (f(x)+g(x))\,dx = \int_E f(x)\,dx + \int_E g(x)\,dx$$

provided that these are defined. (In particular this identity is valid if both f and g are Lebesgue integrable on E.)

Combining this additive theorem with the property of Lemma 7.2 we have immediately one of our most useful tools in the integration theory.

Theorem 7.9 *Let be a $\mathcal{L}^n$-measurable subset of $\mathbb{R}^n$ and let f_1, f_2, f_3, ... be a sequence of nonnegative real-valued functions defined and Lebesgue integrable on E. Suppose that the series*

$$f(x) = \sum_{k=1}^{\infty} f_k(x)$$

converges for every $x \in E$. Then

$$\int_E f(x)\,dx = \sum_{k=1}^{\infty} \int_E f_k(x)\,dx.$$

In particular, f is Lebesgue integrable on E if and only if the series of integrals converges.

Exercise 394 *Show that, for any simple function*

$$f(x) = \sum_{k=1}^{n} c_k \chi_{E_i}(x)$$

where $E_1, E_2, E_3, \dots, E_n$ are $\mathcal{L}^n$-measurable, that

$$\int_E f(x)\,dx = \sum_{k=1}^{n} c_k \mathcal{L}^n(E \cap E_k).$$

Answer □

Exercise 395 *Show that any nonnegative $\mathcal{L}^n$-measurable function $f : \mathbb{R}^n \to \mathbb{R}$ can be written in the form*

$$f(x) = \sum_{k=1}^{\infty} c_k \chi_{E_k}(x)$$

for appropriate $\mathcal{L}^n$-measurable sets $E_1, E_2, E_3, \dots$, and that

$$\int_E f(x)\,dx = \sum_{k=1}^{\infty} c_k \mathcal{L}^n(E \cap E_k).$$

Answer □

Exercise 396 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a $\mathcal{L}^n$-measurable function that is integrable on an interval I. Show that, for every $\varepsilon > 0$ there is a full interval cover β of I so that if $\pi \subset \beta$ is a packing with $J \subset I$ for each $(J,x) \in \pi$ then*

$$\sum_{(J,x)\in\pi} \left| \int_J f(t)\,dt - f(x)|J| \right| < \varepsilon.$$

□

7.3.2 Notation

We have preserved the notation from the elementary calculus in the expression

$$\int_E f(x)\,dx$$

interpreting now x as a dummy variable representing an arbitrary point in $\mathbb{R}^n$. There are other suggestive notations that assist in some situations. For example if $f : \mathbb{R}^2 \to \mathbb{R}$ and E is a subset of $\mathbb{R}^2$ then the integral may appear instead as

$$\int\int_E f(x_1,x_2)\,dx_1 dx_2$$

or

$$\int\int_E f(x,y)\,dxdy.$$

The "double" integral $\int\int$ represents the fact that the dimension is two and contains a hint that an iterated integral may be useful in its computation (see Section 7.5 below).

7.4 General measure theory

The set function

$$E \to \overline{\int_E} f(x)\,dx$$

is defined for every subset E of $\mathbb{R}^n$. Such set functions play a role in many investigations and the students should be made acquainted with the usual general theory and its techniques.

Definition 7.10 *A set function $\mathcal{M}$ defined for all subsets E of $\mathbb{R}^n$ is said to be a* measure *on $\mathbb{R}^n$ provided that*

1. $\mathcal{M}(\emptyset) = 0$.
2. $0 \leq \mathcal{M}(E) \leq \infty$ *for all subsets E of $\mathbb{R}^n$.*
3. $\mathcal{M}(E_1) \leq \mathcal{M}(E_2)$ *if $E_1 \subset E_2 \subset \mathbb{R}^n$.*
4. $\mathcal{M}\left(\bigcup_{k=1}^{\infty} E_k\right) \leq \sum_{k=1}^{\infty} \mathcal{M}(E_k)$ *for any sequence $\{E_k\}$ of subsets of $\mathbb{R}^n$.*

If, moreover,

$$\mathcal{M}(E_1 \cup E_2) = \mathcal{M}(E_1) + \mathcal{M}(E_2).$$

whenever

$$\inf\{\|e_1 - e_2\| : e_1 \in E_1,\ e_2 \in E_2\} > 0$$

then $\mathcal{M}$ is said to be a metric measure *on $\mathbb{R}^n$.*

Note that Lebesgue measure $\mathcal{L}^n$ and the set function

$$\mathcal{M}(E) = \overline{\int_E} f(x)\,dx$$

for any nonnegative function $f : \mathbb{R}^n \to \mathbb{R}$ are metric measures according to this definition. Many authors reserve the term "measure" for set functions defined only on special classes of sets and with stronger additive properties; they would then prefer the term "outer measure" for the concept introduced in this definition. In your readings this should not be hard to keep track of.

For the definition of measurability we take another one of the six characterizations of measurability in dimension one that we presented in Corollary 4.30.

Definition 7.11 *A subset E of $\mathbb{R}^n$ is said to be $\mathcal{M}$-measurable if for every set $A \subset \mathbb{R}^n$*

$$\mathcal{M}(A) = \mathcal{M}(A \cap E) + \mathcal{M}(A \setminus E).$$

We can prove that this definition of measurability, applied to the Lebesgue measure is equivalent to that we are currently using in Definition 7.5. Using this new definition a more general theory emerges that applies to any measure on $\mathbb{R}^n$ (or indeed on any suitable space equipped with a measure). Here is a fast summary.

- Let $E_1, E_2, E_3, \ldots$ be a sequence of pairwise disjoint $\mathcal{M}$-measurable subsets of $\mathbb{R}^n$ and write $E = \bigcup_{i=1}^{\infty} E_i$. Then, for any set $A \subset \mathbb{R}^n$,
$$\mathcal{M}(A \cap E) = \sum_{i=1}^{\infty} \mathcal{M}(A \cap E_i).$$
- The class of all $\mathcal{M}$-measurable subsets of $\mathbb{R}^n$ forms a Borel family that contains all $\mathcal{M}$-measure zero sets.

- If $\mathcal{M}$ is a metric measure then the class of all $\mathcal{M}$-measurable subsets contains all closed sets.

This material is standard and should be part of the background for any advanced student. Almost all texts that discuss outer measures will provide detailed proofs of these facts. You may wish to consult Chapters 2 and 3 of our text Bruckner, Bruckner, and Thomson, *Real Analysis*, 2nd Ed., ClassicalRealAnalysis.com (2008). Those chapters are available for free download.

7.5 Iterated integrals

In many cases the computation of a integral in a higher dimensional space can be accomplished only through a series of one-dimensional integrations. We do not have anything that is as convenient and useful as the calculus computation

$$\int_a^b F'(x)\,dx = F(b) - F(a)$$

that did most of the work in our first calculus course. But if we can reduce an integral in $\mathbb{R}^n$ to several ordinary integrals then the computations can be carried out.

The reader has likely seen in some elementary calculus classes the computation

$$\int\int_{[a,b]\times[c,d]} f(x,y)\,dx\,dy = \int_c^d \left(\int_a^b f(x,y)\,dx\right)\,dy = \int_a^b \left(\int_c^d f(x,y)\,dy\right)\,dx.$$

Another similar, and no doubt also familiar, kind of computation appears in the form

$$\int\int_E f(x,y)\,dx\,dy = \int_a^b \left(\int_{L(u)}^{U(u)} f(u,v)\,dv\right)\,du$$

when E is the set

$$E = \{(u,v) \in \mathbb{R}^2 : a \le u \le b,\ L(u) \le v \le U(u)\}.$$

To formulate the problem correctly we need to consider how best to state it. For example, what would we wish to state for a three dimensional Lebesgue integral

$$\int\int\int_{[a,b]\times[c,d]\times[e,f]} F(x,y,z)\,dxdydz?$$

We might wish to have three iterations

$$\int_a^b \left(\int_c^d \left(\int_e^f F(x,y,z)\,dz\right)\,dy\right)\,dx$$

performed in the order here as 3–2–1. But there are six possible orders in which we could iterate. We also might wish to iterate this as

$$\int_e^f \left(\int\int_{[a,b]\times[c,d]} F(x,y,z)\,dxdy\right)\,dz$$

in the order (1×2)–3. There are three possible such orders in which this might be performed. To capture all of these it is best to keep to a level of abstraction. This is more convenient inside the general theory of measure and integration by using product measures. We will be a little less ambitious.

7.5.1 Formulation of the iterated integral property

Let m and n be positive integers and consider the Lebesgue integral

$$\int_I f(x)\,dx$$

for a function $f : I \to \mathbb{R}^{m+n} \to \mathbb{R}$ where I is an interval in $\mathbb{R}^{m+n}$. Every point x in $\mathbb{R}^{m+n}$ can be written as

$$x = (u,v) \quad (u \in \mathbb{R}^m,\ v \in \mathbb{R}^n)$$

and the interval $I = A(1) \times A(2)$ where $A(1)$ is an interval in $\mathbb{R}^m$ and $A(2)$ is an interval in $\mathbb{R}^n$.

We shall ask for conditions on a function $f : \mathbb{R}^{m+n} \to \mathbb{R}$ that is integrable on the interval $I = A(1) \times A(2)$ so that

- For every[1] $u \in A(1)$ the function
$$v \to f(u,v)$$
is integrable over $A(2)$ and
- the function
$$u \to \int_{A(2)} f(u,v)dv$$
is integrable on $A(1)$, and
- the identity
$$\int_{A(1)\times A(2)} f(x)\,dx =$$
$$\int\int_{A(1)\times A(2)} f(u,v)\,dudv = \int_{A(1)} \left(\int_{A(2)} f(u,v)\,dv\right) du \tag{7.1}$$
is valid.

Exercise 397 *Check that formula 7.1 holds if $f(x) = \chi_I(x)$ where $I = A(1) \times A(2)$ is an interval in $\mathbb{R}^{m+n}$.* □

Exercise 398 *Check that formula 7.1 holds if $f(x)$ is a step function on $I = A(1) \times A(2)$ assuming values $c_1, c_2, \dots c_k$ on subintervals $I_1, I_2, \dots, I_k$ of I.* □

[1] Here we insist on *every* u, but as we know we could and should sometimes ignore a set of measure zero where this fails. That will be covered in Section 7.5.2.

Exercise 399 *Check that formula 7.1 holds if $f(x)$ is a bounded function for which there exists a sequence of step functions $S_1, S_2, S_3, \ldots$ on $I = A(1) \times A(2)$ such that $f(x) = \lim_{k\to\infty} S_k(x)$ for every $x \in I$.* Answer □

Exercise 400 *Show that if f is a continuous function on the closed interval I then Exercise 399 can be applied to verify the formula 7.1.* □

Exercise 401 *Let f_1, f_2, and G be continuous functions on the closed interval I and define a function*

$$f(x) = \begin{cases} f_1(x) & \text{if } x \in I \text{ and } g(x) > 0, \\ f_2(x) & \text{if } x \in I \text{ and } g(x) \leq 0. \end{cases}$$

Show that Exercise 399 can be applied to verify the formula 7.1. Answer □

Exercise 402 (counterexample #1) *There are a number of standard counterexamples that show some caution is needed in applying the iterated technique to multiple integrals of unbounded functions. On the interval $[-1,1] \times [-1,1]$ in $\mathbb{R}^2$ define the function*

$$f(x,y) = xy(x^2+y^2)^{-2} \quad f(0,0) = 0.$$

Examine the integrals

$$\int\int_{[-1,1]\times[-1,1]} f(x,y)\,dxdy,$$

$$\int_{-1}^{1}\left(\int_{-1}^{1} f(x,y)\,dx\right)dy,$$

and

$$\int_{-1}^{1}\left(\int_{-1}^{1} f(x,y)\,dy\right)dx.$$

Answer □

Exercise 403 (counterexample #2) *On the interval $[0,1] \times [-1,1]$ in $\mathbb{R}^2$ define the function*

$$f(x,y) = yx^{-3} \quad \text{if } x > 0 \text{ and } -x < y < x$$

with $f(x,y) = 0$ elsewhere. Examine the integrals

$$\int\int_{[0,1]\times[-1,1]} f(x,y)\,dxdy,$$

$$\int_{-1}^{1}\left(\int_{0}^{1} f(x,y)\,dx\right)dy,$$

and

$$\int_{0}^{1}\left(\int_{-1}^{1} f(x,y)\,dy\right)dx.$$

Answer □

Exercise 404 (counterexample #3) *On the interval* $[0,1]\times[0,1]$ *in* $\mathbb{R}^2$ *define the function*

$$f(x,y) = 2(x-y)(x+y)^{-3} \quad (x>0,\ y>0)$$

with $f(x,y)=0$ *elsewhere. Examine the integrals*

$$\int\int_{[0,1]\times[0,1]} f(x,y)\,dxdy,$$

$$\int_0^1\left(\int_0^1 f(x,y)\,dx\right)dy,$$

and

$$\int_0^1\left(\int_0^1 f(x,y)\,dy\right)dx.$$

Answer □

Exercise 405 *A "clever" student points out that all this trouble over integrals in* $\mathbb{R}^2$ *(or indeed in any dimension) can easily be avoided by simply* defining *double integrals as being two iterated integrals. Thus instead of proving that*

$$\int\int_{[a,b]\times[c,d]} f(x,y)\,dx\,dy = \int_a^b\left(\int_c^d f(x,y)\,dy\right)dx$$

we just take that as a definition. Any comments? Answer □

7.5.2 Fubini's theorem

In the preceding section we obtained a limited version of the iterated integral property, one that applied only to bounded functions and which required in the iteration (7.1) that the inside integral

$$\int_{A(2)} f(u,v)\,dv$$

exist for every value of u. The most general theorem, usually described as Fubini's theorem, asserts that this iteration is available for all integrable functions provided that we accept a set of measure zero where the inside integral might not exist.

Here are the ingredients of that theorem.

Let m and n be positive integers and we suppose that $f:\mathbb{R}^{m+n}\to\mathbb{R}$ is a function Lebesgue integrable on an interval $I=A(1)\times A(2)$ where $A(1)$ is an interval in $\mathbb{R}^m$ and $A(2)$ is an interval in $\mathbb{R}^n$. As before every point x in $\mathbb{R}^{m+n}$ is to be written as

$$x=(u,v) \quad (u\in\mathbb{R}^m,\ v\in\mathbb{R}^n).$$

Then

- There is a set $N(1)\subset A(1)$ with m-dimensional Lebesgue measure equal to zero.

- For every $u \in A(1) \setminus N(1)$ the function
$$v \to f(u,v)$$
is integrable over $A(2)$ and
- the function
$$u \to \int_{A(2)} f(u,v)dv$$
is integrable on $A(1)$, and
- the identity
$$\int_{A(1)\times A(2)} f(x)\,dx = \int\int_{A(1)\times A(2)} f(u,v)\,dudv$$
$$= \int_{A(1)\setminus N(1)} \left(\int_{A(2)} f(u,v)\,dv \right) du \tag{7.2}$$
is valid.

This theorem is proved as Theorem 7–1, pp. 300–303 in E. J. McShane, *Unified Integration*, Academic Press (1983). There is a version in Chapter 6 of R. Henstock, *Lectures on the Theory of Integration*, World Scientific (1988). His version is more general (and less accessible) but uses the same defining structure essentially. The reader is, however, encouraged now to learn this theorem in the setting of general measure theory where the arguments are simpler and more straightforward. For that there are an abundance of excellent texts. We cannot resist recommending, from among them, Bruckner, Bruckner, and Thomson, *Real Analysis*, 2nd Ed., ClassicalRealAnalysis.com (2008).

7.6 Expression as a Stieltjes integral

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a $\mathcal{L}^n$-measurable function that is integrable on a measurable set E. We shall show that the Lebesgue integral
$$\int_E f(x)\,dx$$
can be realized as a one-dimensional Stieltjes integral. Let us fix f and E for our discussion in this section and suppose that $\mathcal{L}^n(E) < \infty$. We define for each real number s the function
$$w(s) = \mathcal{L}^n(\{x \in E : f(x) > s\})$$
called the *distribution function* of the function f on the set E.

Then the following properties of the distribution function are easily established:

- The function $w : \mathbb{R} \to [0,\infty)$ is nonincreasing with
$$\lim_{s\to\infty} w(s) = 0 \text{ and } \lim_{s\to-\infty} w(s) = \mathcal{L}^n(E).$$

- $\mathcal{L}^n(\{x \in E : a < f(x) \leq b\}) = w(b) - w(a)$.
- $w(s+) = w(s)$ (i.e., w is continuous on the right at each point).
- $w(s-) = \mathcal{L}^n(\{x \in E : f(x) \geq s\})$.

The representation theorem expresses the Lebesgue integral of f in terms of the Stieltjes integral

$$\int_a^b s\,dw(s).$$

We know from our study of the Stieltjes integral that this must exist since w is a nonincreasing function.

Theorem 7.12 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a $\mathcal{L}^n$-measurable function that is integrable on a measurable set E and that $\mathcal{L}^n(E) < \infty$. Then*

$$\int_{\{x \in E : a < f(x) \leq b\}} f(x)\,dx = \int_a^b s\,dw(s) \tag{7.3}$$

and

$$\int_E f(x)\,dx = \int_{-\infty}^{\infty} s\,dw(s). \tag{7.4}$$

There are numerous textbooks where the details of this development can be found. A most readable account appears in pp. 76–79 of Wheeden and Zygmund, *Measure and Integral*, Marcel Dekker (1977).

Exercise 406 *Prove the identity (7.3) in Theorem 7.12:*

$$\int_{\{x \in E : a < f(x) \leq b\}} f(x)\,dx = \int_a^b s\,dw(s).$$

Answer □

Exercise 407 *Deduce the identity (7.4) from the identity (7.3) in Theorem 7.12.*

Answer □

Chapter 8

ANSWERS

8.1 Answers to problems

Exercise 1, page 5

Assume that there is a nonnegative number M so that

$$|F(y) - F(x)| \leq M|y - x|$$

for all x, $y \in [a,b]$. Just check directly that

$$\mathrm{Var}(F,[a,b]) \leq M(b-a) < \infty.$$

Exercise 2, page 7

The maximum value of f in each of the intervals $[0,\frac{1}{4}]$, $[\frac{1}{4},\frac{1}{2}]$, $[\frac{1}{2},\frac{3}{4}]$, and $[\frac{3}{4},1]$ is $1/8$, $1/4$, $9/16$, and 1 respectively. Thus define F to be $x/8$ in the first interval, $1/32 + 1/4(x - 1/1/4)$ in the second interval, $1/32 + 1/16 + 9/16(x - 1/2)$ in the third interval, and to be $1/32 + 1/16 + 9/64 + (x - 3/4)$ in the final interval. This should be (if the arithmetic was correct) a continuous, piecewise linear function whose slope in each segment exceeds the value of the function f.

Exercise 3, page 7

Start at 0 and first of all work to the right. On the interval $(0,1)$ the function f has the constant value 1. So define $F(x) = x$ on $[0,1]$. Then on the the interval $(1,2)$ the function f has the constant value 2. So define $F(x) = 1 + 2(x-1)$ on $[1,2]$. Continue until you see how to describe F in general. This is the same construction we used for upper functions.

Exercise 4, page 14

The choice of midpoint

$$\frac{x_i + x_{i-1}}{2} = \xi_i$$

for the Riemann sum gives a sum

$$= \frac{1}{2}\sum_{i=1}^{n}(x_i^2 - x_{i-1}^2) = \frac{1}{2}\left[b^2 - x_{n-1}^2 + x_{n-1}^2 - x_{n-2}^2 + \cdots - a^2\right] = (b^2 - a^2)/2.$$

To explain why this works you might take the indefinite integral $F(x) = x^2/2$ and check that

$$\frac{F(d) - F(c)}{d - c} = \frac{c + d}{2}$$

so that the mean-value always picks out the midpoint of the interval $[c,d]$ for this very simple function.

Exercise 10, page 15

Just take, first, the points ξ_i^* at which we have the exact identity

$$\int_{x_{i-1}}^{x_i} f(x)\,dx - f(\xi_i^*)(x_i - x_{i-1}) = 0$$

Then, for any other point ξ_i,

$$\left|\int_{x_{i-1}}^{x_i} f(x)\,dx - f(\xi_i)(x_i - x_{i-1})\right|$$

$$= |f(\xi_i) - f(\xi^*)|(x_i - x_{i-1}) \leq \omega f([x_i, x_{i-1}])(x_i - x_{i-1}).$$

The final comparison with

$$\sum_{i=1}^{n} \omega f([x_i, x_{i-1}])(x_i - x_{i-1})$$

follows from this.

To get a good approximation of the integral by Riemann sums it seems that we might need

$$\sum_{i=1}^{n} \omega f([x_i, x_{i-1}])(x_i - x_{i-1})$$

to be small. Observe that the pieces in the sum here can be made small if (a) the function is continuous so that the oscillations are small, or (b) points where the function is not continuous occur in intervals $[x_i, x_{i-1}]$ that are small. Loosely then we can make these sums small if the function is mostly continuous, i.e., where it is not continuous can be covered by some small intervals that don't add up to much. The modern statement of this is "the function needs to be continuous almost everywhere."

Exercise 11, page 17

Let $\varepsilon > 0$ and choose $\delta > 0$ so that

$$\omega f([c,d]) < \frac{\varepsilon}{(b-a)}$$

whenever $[c,d]$ is a subinterval of $[a,b]$ for which $d-c<\delta$. Note then that if

$$\{([x_i,x_{i-1}],\xi_i) : i = 1,2,\dots n\}$$

is a partition of $[a,b]$ with intervals shorter than δ then

$$\sum_{i=1}^{n} \omega f([x_i,x_{i-1}])(x_i - x_{i-1}) < \sum_{i=1}^{n}[\varepsilon/(b-a)](x_i - x_{i-1}) = \varepsilon.$$

Consequently, by Exercise 10,

$$\left| \int_a^b f(x)\,dx - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right|$$

$$\leq \sum_{i=1}^{n} \left| \int_{x_{i-1}}^{x_i} f(x)\,dx - f(\xi_i)(x_i - x_{i-1}) \right| < \varepsilon.$$

Exercise 12, page 17

Of course we can more easily use the definition of the integral and compute that $\int_0^1 x^2\,dx = 1/3 - 0$. This exercise shows that, under certain simple conditions, not merely can we approximate the value of the integral by Riemann sums, we can produce a sequence of numbers which converges to the value of the integral. Simply divide the interval at the points 0, $1/n$, $2/n$, $\dots$, $n-1)/n$, and 1. Take $\xi = i/n$ [the right hand endpoint of the interval]. Then the Riemann sum for this partition is

$$\sum_{i=1}^{n} \left(\frac{i}{n}\right)^2 \frac{1}{n} = \frac{1^2+2^2+3^2+4^2+5^2+6^2+\cdots+n^2}{n^3}.$$

As $n \to \infty$ this must converge to the value of the integral by Theorem 1.8. The student is advised to find the needed formula for

$$1^2+2^2+3^2+4^2+5^2+6^2+\cdots+N^2.$$

and determine whether the limit is indeed the correct value $1/3$.

Exercise 13, page 17

Determine the value of the integral

$$\int_0^1 x^2\,dx$$

in the following way. Let $0<r<1$ be fixed. Subdivide the interval $[0,1]$ by defining the points $x_0=0$, $x_1=r^{n-1}$, $x_2=r^{n-2}$, $\dots$, $x_{n-1}=r^{n-(n-1)}=r$, and $x_n=r^{n-(-n)}=1$. Choose the points $\xi_i\in[x_{i-1},x_i]$ as the right-hand endpoint of the interval. Then

$$\sum_{i=1}^{n}\xi_i^2(x_i-x_{i-1})=\sum_{i=1}^{n}\left(r^{n-i}\right)^2\left(r^{n-i}-r^{n-i+1}\right).$$

Note that for every value of n this is a Riemann sum over subintervals whose length is smaller than $1-r$.

As $r\to 1-$ this must converge to the value of the integral by Theorem 1.8. The student is advised to carry out the evaluation of this limit to determine whether the limit is indeed the correct value $1/3$.

Exercise 16, page 24

If f is everywhere the derivative of a function $F;[a,b]\to\mathbb{R}$ then

$$\int_a^b f(x)\,dx=F(b)-F(a)$$

exists and we can take I as this value. For each point ξ in $[c,d]$ choose $\delta(\xi)$ sufficiently small that

$$\left|\frac{F(y)-F(x)}{y-x}-F'(\xi)\right|<\frac{\varepsilon}{C}$$

whenever x and y are points in $[c,d]$ for which $x\le\xi\le y$ and $0<y-x<\delta(\xi)$. This gives us

$$\left|F(y)-F(x)-F'(\xi)(y-x)\right|<\frac{\varepsilon}{C}(y-x).$$

Then, for any choice of points $x_0,x_1,\dots,x_n$ and $\xi_1,\xi_2,\dots,\xi_n$ from $[c,d]$ with the four properties of the statement of the lemma,

$$\left|\int_a^b F'(x)\,dx-\sum_{i=1}^{n}F'(\xi_i)(x_i-x_{i-1})\right|$$

$$=\left|\sum_{i=1}^{n}\left[F(x_i)-F(x_{i-1})-F'(\xi_i)(x_i-x_{i-1})\right]\right|$$

$$\le\sum_{i=1}^{n}\left|F(x_i)-F(x_{i-1})-F'(\xi_i)(x_i-x_{i-1})\right|<\frac{\varepsilon}{C}\sum_{i=1}^{n}|x_i-x_{i-1}|\le\varepsilon.$$

Exercise 20, page 31

Since $N \subset (a,b)$ is a set of measure zero, for each integer $n = 1,2,3,\dots$, we can choose a positive function $\delta_n : N \to \mathbb{R}^+$ so that

$$\sum_{([u,v],w)\in\pi} n\,|v-u| < 2^{-n}$$

whenever π is a subpartition anchored in N finer than δ_n.

First, for each integer $n = 1,2,3,\dots$ define the function $G_n(x)$ at each point $a < x < b$ by requiring $G_n(x)$ to be the supremum of the values

$$\sum_{([u,v],w)\in\pi} n\,|v-u|$$

taken over all subpartitions π of $[a,x]$ anchored in N and finer than δ_n. Note that $G_n : [a,b] \to \mathbb{R}$ is nondecreasing, $G_n(a) = 0$ and $G_n(b) < 2^{-n}$.

We see that, for any integer n and all $k = 1,2,3,\dots,n$, if $x \in N$ and if $0 < y-x < \delta_n(x)$ then $([x,y],x)$ is finer than δ_k and so

$$G_k(y) - G_k(x) \geq k\,(y-x)\,.$$

Similarly if $0 < x-y < \delta_n(x)$ then $([y,x],x)$ is finer than δ_k and so

$$G_k(x) - G_k(y) \geq k\,(x-y)\,.$$

We now ready to define

$$\phi(x) = \sum_{k=1}^{\infty} G_k(x).$$

This is a finite-valued function, nondecreasing on (a,b). Note that, if $0 < y-x < \delta_n(x)$ then

$$\frac{\phi(y)-\phi(x)}{y-x} \geq n$$

and if $0 < x-y < \delta_n(x)$ then

$$\frac{\phi(x)-\phi(y)}{x-y} \geq n.$$

Consequently for each $x \in N$,

$$\lim_{y\to x} \frac{\phi(y)-\phi(x)}{y-x} = +\infty.$$

Exercise 22, page 33

Suppose that $H : \mathbb{R} \to \mathbb{R}$ is continuous, and that Z is a countable set. We show that H has zero variation on Z. Let $\varepsilon > 0$. List the points in Z in a sequence $\{z_n\}$. For each integer n we use the fact that H is continuous at z_n to choose

$\delta(z_n) > 0$ so that

$$|H(z_n+h) - H(z_n)| < \frac{\varepsilon}{2^{n+1}}$$

whenever $|h| < \delta(z_n)$.

Consider a subpartition

$$\pi = \{([a_i, b_i], \xi_i) : i = 1, 2, 3, \ldots, m\}$$

finer than δ and anchored in Z. For that partition we simply note that if $\xi_i = h_n$ for some n then

$$|H(b_i) - H(a_i)| \leq |H(b_i) - H(z_n)| + |H(z_n) - H(a_i)| < \frac{2\varepsilon}{2^{n+1}}.$$

Using this observation we easily compute that

$$\sum_{i=1}^{n} |H(b_i) - H(a_i)| < \varepsilon.$$

Exercise 24, page 33

Hint: It is enough to prove this for bounded sets E.

Exercise 25, page 33

It is enough to suppose that E is contained in some interval (a,b). If F has a finite derivative at every point of a set E of measure zero, then, for each $x \in E$ select $\delta_1(x) > 0$ so that

$$\left| \frac{F(y) - F(x)}{y - x} - F'(x) \right| < \varepsilon$$

if $0 < |y - x| < \delta_1(x)$. Also. for each $x \in E$ select $\delta_2(x) > 0$ so that

$$\sum_{i=1}^{n} [|F'(x)| + \varepsilon](y_i - x_i) < \varepsilon$$

for any subpartition

$$\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \ldots\}$$

anchored in E and finer than δ_2. Take δ as the minimum of δ_1 and δ_2 and consider any subpartition

$$\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \ldots\}$$

anchored in E and finer than δ. Note simply that

$$\sum_{i=1}^{n} |F(y_i) - F(x_i)| \leq \sum_{i=1}^{n} \varepsilon(y_i - x_i) + \sum_{i=1}^{n} |F'(x)|(y_i - x_i) < \varepsilon.$$

Exercise 26, page 33

This is easy to prove—make sure to use Cousin's lemma.

Exercise 27, page 33

Let $M \subset [F(a),F(b)]$ and $N \subset [a,b]$. Suppose that G has zero variation on M. Let $\varepsilon > 0$. There must be a positive function $\delta_1 : M \to \mathbb{R}^+$ so that

$$\sum_{([x,y],z)\in\pi} |G(y)-G(x)| < \varepsilon$$

whenever π is a subpartition anchored in M and finer than δ_1. Suppose that $F(N) \subset M$. For each $w \in N$ use the continuity of F to determine $\delta_2(w)$ so that

$$|F(t)-F(w)| < \delta_1(F(w))/2$$

provided $|t-w| < \delta_2(w)$. Let π_2 be a subpartition anchored in N and finer than δ_2. There is a corresponding subpartition π_1 obtained by replacing each $([u,v],w) \in \pi_2$ by the pair

$$([x,y],z) = ([F(u),F(v)],F(w)).$$

Note that each interval-point pair $([u,v],w)$ corresponds to the pair $([x,y],z)$ with $F(u)=x$, $F(v)=y$, and $F(w)=z$, $w \in N$, $z \in F(N) \subset M$, and $x \le z \le y$ with

$$|y-x| = |F(v)-F(u)| \le |F(v)-F(w)| + |F(w)-F(u)| < \delta_1(F(w)) = \delta_1(z).$$

The subpartition π_1 is thus anchored in M and is finer than δ_1. Consequently

$$\sum_{([u,v],w)\in\pi_2} (v-u) = \sum_{([x,y],z)\in\pi_1} |G(y)-G(x)| < \varepsilon.$$

This verifies that N is a set of measure zero. The reverse direction is clear since the roles of F and G are completely reversible.

Exercise 28, page 33

Exercise 27 is just the special case of this exercise for which F is assumed to be both continuous and strictly increasing. Here we extend that exercise by dropping the continuity of F. An inverse does not necessarily exist, but a left-inverse does for strictly increasing functions. We shall need this variant in the text for one of our proofs and so, for convenience, we place it here.

We give a detailed solution as some of the arguments are a bit tedious to realize. The best approach here is to sketch a good picture illustrating the situation. Then the details are easier to construct or to follow. (Use a simple function F with one interval of constancy and two jumps; your picture for G then will have two intervals of constancy and one jump.)

Item 1 If y_1 and y_2 belong to $[F(a),F(b)]$ with $y_1 \leq y_2$, then

$$\{u \in [a,b] : F(u) \geq y_2\} \subset \{u \in [a,b] : F(u) \geq y_1\}$$

and so

$$G(y_1) = \inf\{u \in [a,b] : F(u) \geq y_1\} \leq \inf\{u \in [a,b] : F(u) \geq y_2\} = G(y_2).$$

This shows that G is nondecreasing as stated in Item 1.

Let us check that G is left-continuous. Take a point $y_0 \in (F(a),F(b)]$. By the definition of $G(y_0)$, $F(t) < y_0$ for all $t < G(y_0)$. Let $\varepsilon > 0$ and fix a point u_0 chosen so that $G(y_0) - \varepsilon < u_0 < G(y_0)$. Then for every point y with $F(u_0) < y < y_0$ we have

$$G(y) = \inf\inf\{u \in [a,b] : F(u) \geq y\} \geq u_0 \geq G(y_0) - \varepsilon.$$

This inequality verifies that G is left-continuous at the point y_0.

Item 2 Suppose that $G(y_0) < G(y_0+)$ for some point $y_0 \in [F(a),F(b))$. i.e, that there is a jump at the point y_0. (We know that G is left-continuous, so the jump must occur on the right.)

By the definition of $G(y_0)$ and the fact that F is nondecreasing, we know that $F(t) \geq y_0$ for all $t > G(y_0)$. We also have, for any u with

$$G(y_0) < u < G(y_0+),$$

that $G(y) > u$ for all $y > y_0$. Thus $F(u) < y$ whenever $y > y_0$. Thus $F(u) < y_0$. That proves that F assumes the value y_0 for all points in the interval between $G(y_0)$ and $G(y_0+)$.

In the converse direction, suppose that F assumes the value y_0 for all points in the interval (u_1,u_2). Then, by definition, $G(y_0) \leq u_1$. For values $y > y_0$, and every $u \in (u_1,u_2)$,

$$F(u) = y_0 < y$$

and so $G(y) \geq u$ by the way $G(y)$ is defined. Let $u \to u_2$ from above to see that $G(y) \geq u_2$ for these values of y. That means $G(y_0+) \geq x_2$. This exhibits a jump for G at the point y_0 because we see that

$$G(y_0) \leq u_1 < u_2 \leq G(y_0+).$$

Item 3 First observe that if we use $y = F(t)$ in the definition of $G(y)$ we obtain that $G(F(t)) \leq t$. Moreover, suppose that $G(F(t)) < t$ does occur for some value of t. Then there must be a point $u < t$ for which $F(u) \geq F(t)$. That must mean that F is constant on the entire interval $[u,t]$ as stated. It is also true that if we assume F is constant on an entire interval $[u,t]$ then $G(F(t)) \leq u < t$.

Item 4 Suppose the G assumes some constant value x_0 in an interval (y_1,y_2). If $u > u_0$ then $F(u) \geq y$ for all y in the interval (y_1,y_2) because of the way $G(y)$ is defined. Let $y \to y_2$ from above and we see that $F(u) \geq y_2$ and so $F(u_0+) \geq y_2$.

Also if $u < u_0$ then $F(u) < y$ for all y in the interval (y_1, y_2) again because of the way $G(y)$ is defined. Let $y \to y_1$ from below and we see that $F(u) \le y_1$ and so $F(u_0-) \le y_1$.

In the converse direction let y be a value between $F(u_0-)$ and $F(u_0+)$. If $u < u_0$ then $F(u) < y$ and so $G(y) \ge u_0$. Applying Item 3 here and recalling that both functions F and G are nondecreasing, we wee that

$$u_0 \le G(y) \le G(F(u_0+) \le G(F(u)) \le u.$$

Let $u \to u_0$ from above and we deduce that $G(y) = u_0$ for all values of y between $F(u_0-)$ and $F(u_0+)$.

Item 5 This follows directly from item 1 and item 2. If F is strictly increasing then it cannot be constant on any interval, thus we see that $G(F(t)) = T$ for all t and that G has no jump discontinuities.

Item 6 Now we just repeat the argument from Exercise 27, with the necessary modifications to handle the fact that F may not be continuous. Note that we do assume F is strictly increasing as before, so it is only the discontinuity points that might interfere with the argument. Also, in Exercise 27 we did use the fact that G was an inverse, but we needed really only a left-inverse.

Let C be the set of points at which F fails to be continuous. Then C is countable because F is nondecreasing.

Let $M \subset [F(a), F(b)]$ and $N \subset [a,b]$. Suppose that G has zero variation on M. Let $\varepsilon > 0$. There must be a positive function $\delta_1 : M \to \mathbb{R}^+$ so that

$$\sum_{([x,y],z)\in\pi} |G(y) - G(x)| < \varepsilon$$

whenever π is a subpartition anchored in M and finer than δ_1.

Suppose that $F(N) \subset M$. Write $N_1 = N \setminus C$. For each $w \in N_1$ use the continuity of F to determine $\delta_2(w)$ so that

$$|F(t) - F(w)| < \delta_1(F(w))/2$$

provided $|t - w| < \delta_2(w)$. Let π_2 be a subpartition anchored in N and finer than δ_2. There is a corresponding subpartition π_1 obtained by replacing each $([u,v],w) \in \pi_2$ by the pair

$$([x,y],z) = ([F(u), F(v)], F(w)).$$

Note that each interval-point pair $([u,v],w)$ corresponds to the pair $([x,y],z)$ with $F(u) = x$, $F(v) = y$, and $F(w) = z$, $w \in N_1$, $z \in F(N_1) \subset M$, and $x \le z \le y$ with $|y-x| = |F(v) - F(u)| \le |F(v) - F(w)| + |F(w) - F(u)| < \delta_1(F(w)) = \delta_1(z)$.

The subpartition π_1 is thus anchored in M and is finer than δ_1. Consequently

$$\sum_{([u,v],w)\in\pi_2}(v-u)=\sum_{([x,y],z)\in\pi_1}|G(y)-G(x)|<\varepsilon.$$

This verifies that N_1 is a set of measure zero. Since C is also a set of measure zero (in fact C is countable) we see that $N=N_1\cup C$ must also be a set of measure zero.

Exercise 31, page 35

Take $F:[0,1]\to\mathbb{R}$ as any constant function. Note that $F'(x)=0=P(x)$ for every irrational number x and $F'(x)=0\neq P(x)$ for every rational number x. This is a countable set of exceptions which is allowed for the utility version, but not for the classical, naive, or elementary Newton versions.

A similar closely related function, the Dirichelet function, is also Newton integrable in the utility sense, but is not integrable in the classical, naive, or elementary senses. The Dirichelet function is the function $D:[0,1]\to\mathbb{R}$ defined by $D(x)=0$ for x irrational and $D(x)=1$ if $x=p/q$ is a rational number.

The function P is continuous at every rational number. It is regulated (see Section 1.9.1). It is also Riemann integrable. The function D is continuous nowhere on $[0,1]$. it is not regulated nor Riemann integrable. The continuity properties of a function certainly play a role in some theories. Note, here, that the utility version of the integral simply ignores the values of the function off some countable set and so both functions P and D are integrable in this sense.

To find a function that is integrable in the general sense but not in the utility sense, just take a set $N\subset(0,1)$ that is uncountable and has measure zero. Define $h(x)=0$ if $x\in[0,1]\setminus N$ and $h(x)=1$ for $x\in N$. Then the general integral simply ignores the set N while the utility integral cannot.

Exercise 32, page 35

Let N be any set of measure zero. Let N_1 the set of points x in N at which $F'(x)$ does not exist. According to any of the versions of the integral, F has zero variation on N_1. But by Exercise 25 the function F has zero variation on $N_2=N\setminus N_1$. Consequently F has zero variation on $N=N_1\cup N_2$.

Exercise 33, page 35

Each of the definitions of the various Newton integrals requires just one serious justification. If there does exist such a function F then there would exist many such functions. Are we sure for all such functions that $F(b)-F(a)$ is always the same number? If so that would permit us to assign that value to the integral $\int_a^b f(x)\,dx$. The lemmas supply the answers and justify the definitions.

We give these explicitly and in some detail since they would be an essential feature of any course electing to present a rigorous account of integration theory along these lines. This first exercise justifies the naive integral.

Define $H(x) = F(x) - G(x)$ and observe that H is continuous on $[a,b]$ and has a zero derivative at each point in (a,b). By Exercise 24 it follows that H has zero variation on (a,b). By Exercise 26 then H is a constant on (a,b). Since H is continuous on $[a,b]$, $H(b) = H(a)$ and, consequently $F(b) - F(a) = G(b) - G(a)$ as required.

Exercise 34, page 35

Write $H = F - G$. Let Z_1 be the countable set of points z in (a,b) at which $F'(x) = f(x)$ fails. Let Z_2 be the countable set of points z in (a,b) at which $G'(x) = f(x)$ fails. The set $Z = Z_1 \cup Z_2$ is countable. Since H is continuous, H has zero variation on Z by Exercise 22. But (exactly as in the preceding lemma) H has a zero derivative and hence has zero variation on $(a,b) \setminus Z$. Consequently H has zero variation on all of (a,b). Again we conclude that H is a constant on (a,b), that $H(b) = H(a)$ and, consequently, $F(b) - F(a) = G(b) - G(a)$ as required.

Exercise 35, page 36

Define $H(x) = F(x) - G(x)$ and observe that H is continuous on $[a,b]$ and has a zero derivative at each point in $(a,b) \setminus (N_1 \cup N_2)$. By Exercise 24 it follows that H has zero variation on $(a,b) \setminus (N_1 \cup N_2)$.

By Exercise 25 the function F has zero variation on $N_2 \setminus N_1$. By the same exercise the function G has zero variation on $N_1 \setminus N_2$. But since, F has zero variation on N_1 and G has zero variation on N_2, it follows that $H = F - G$ has zero variation on $N_1 \cup N_2$. Consequently H has zero variation on (a,b).

By Exercise 26 then H is a constant on (a,b). Since H is continuous on $[a,b]$, $H(b) = H(a)$ and, consequently $F(b) - F(a) = G(b) - G(a)$ as required.

Exercise 36, page 37

For each function there would be a control so that there would be two strictly increasing functions ϕ_1, $\phi : (a,b) \to \mathbb{R}$ to use. Take $H = F_1 - F_2$ and $\phi = \phi_1 + \phi_2$ and check that

$$\lim_{y \to x} \frac{H(y) - H(x)}{\phi(y) - \phi(x)} = 0.$$

A simple Cousin argument applied to any compact interval $[c,d] \subset (a,b)$ will reveal that H is constant on $[c,d]$. That shows that

$$F_1(d) - F_1(c) = F_2(d) - F_2(c)$$

and uniform continuity extends that to

$$F_1(b)-F_1(a)=F_2(b)-F_2(a).$$

Exercise 38, page 41

Here is the basic simple argument. Let F and G be indefinite integrals for f and g on $[a,b]$ and $f \leq g$ and set $H=F-G$. Then H is continuous and $H'(x)=G'(x)-F'(x)\geq 0$ for points x in (a,b) at which both derivatives exist. Argue that (in all cases) the function H is nondecreasing. Thus $H(b)\geq a$ and so $G(b)-G(a)\geq F(b)-F(a)$. Consequently

$$\int_a^b g(x)\,dx-\int_a^b f(x)\,dx=[G(b)-G(a)]-[F(b)-F(a)]\geq 0.$$

For the utility and general versions it will take an argument to show that the function H is nondecreasing. The methods of Section 1.7 will work.

Exercise 39, page 41

This begins with a simple calculus exercise. Let $h=\sum_{i=1}^n c_i f_i$. Take F_i as the indefinite integrals for the f_i and just check that the function

$$H=\sum_{i=1}^n c_i F_i$$

is a continuous function for which the relation $H'(x)=h(x)$ must hold except for the collection of points where one at least of the identities $F_i'(x)=f_i(x)$ fails.

For the classical, naive, and elementary versions of the integral it is immediate that H is an indefinite integral of h. In that case

$$\int_a^b h(t)\,dt=H(b)-H(a)=\sum_{i=1}^n c_i[F_i(b)-F_i(a)]$$
$$=\sum_{i=1}^n c_i\left(\int_a^b f_i(x)\,dx\right)$$

as required.

The utility and general versions use the same steps but require a bit more argument. For example if Z_i is the set of points at which the relation $F_i'(x)=f_i(x)$ fails, then let $Z=\bigcup_{i=1}^n Z_i$. This is a countable set (for the utility version) or a measure zero set (for the general version). In the utility case it is clear than H is an indefinite integral for h. In the general case, just note that, by Exercise 32, each F_i has zero variation on Z. From that it is easy to see that H must also have zero variation on Z. So also, in the general case it is clear than H is an indefinite integral for h.

Exercise 40, page 41

The reason the property fails for the elementary version of the Newton integral is that the finite exceptional set does not survive a limit operation. For example take P as the popcorn function of Exercise 31. The popcorn function is the function $P : [0,1] \to \mathbb{R}$ defined by $P(x) = 0$ for x irrational and $P(x) = 1/q$ if $x = p/q$ expresses the rational number x in its lowest terms. Now, for each integer n write $P_n(x) = 0$ if for x irrational and $P(x) = 1/q$ if $x = p/q$ expresses the rational number x in its lowest terms and $q \leq n$; otherwise $P_n(x) = 0$. One checks that P_n is integrable in the elementary sense since it fails to be zero only on a finite set. But $P_n \to P$ uniformly on $[0,1]$, and yet P fails to be integrable in the elementary sense (although it is integrable in the utility sense).

For the remaining Newton integrals the following arguments work.

Checking the derivative Suppose that F_n is the indefinite integral of f_n on $[a,b]$ and that C_n is the set of points x where $F_n'(x) = f_n(x)$ fails. Suppose that $f_n \to f$ uniformly. We check that F_n converges to a continuous function F for which $F'(x) = f(x)$ except possibly at points x belonging to the set $C = \bigcup_{n=1}^{\infty} C_n$.

Let $\varepsilon > 0$. Choose an integer N so that

$$|f_n(x) - f_m(x)| < \varepsilon$$

for all $x \in [a,b]$ and all n, $m \geq N$. Using the monotone property compute that, for any $a \leq s \neq t \leq b$, and all n, $m \geq N$

$$|[F_n(t) - F_m(t)] - [F_n(s) - F_m(s)]| \leq \varepsilon(t-s).$$

From this we deduce that $\{F_n\}$ is uniformly convergent to a continuous function which we will denote as F.

We wish now to show that $F'(x) = f(x)$ at every point x of (a,b) that is not in C. Rewrite the inequality above as

$$|[F_n(t) - F_n(s)] - [F_m(t) - F_m(s)]| \leq \varepsilon(t-s).$$

Deduce that, for $n \geq N$

$$\left| \frac{F_n(t) - F_n(s)}{t-s} - \frac{F(t) - F(s)}{t-s} \right| \leq \varepsilon.$$

This inequality provides $F'(x) = f(x)$ for points x not in C.

Handling zero variation To complete the proof we need to be sure that F is an indefinite integral for f in the sense of one of the Newton integrals (other than the elementary sense). For all but the general sense this is evident. [Eg., for the utility sense each set C_n is countable and, hence, so too is the union C.]

Let us suppose that each set C_n has measure zero. Then C too must have measure zero. We show that F has zero variation on C.

Let $\varepsilon > 0$. Choose an integer N so that

$$|f_n(x) - f_m(x)| < \frac{\varepsilon}{2(b-a)}$$

for all $x \in [a,b]$ and all n, $m \geq N$. As before this means too that

$$|[F_N(t) - F_N(s)] - [F(t) - F(s)]| \leq \frac{\varepsilon(t-s))}{2(b-a)}.$$

By Exercise 32 the function F_N has zero variation on C. Choose a positive function $\delta : C \to \mathbb{R}^+$ so that that

$$\sum_{i=1}^{n} |F_N(y_i) - F_N(x_i)| < \varepsilon/2$$

whenever a subpartition

$$\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$$

is anchored in C and finer than δ.

But then, for such subpartitions, we also have

$$\sum_{i=1}^{n} |F(y_i) - F(x_i)| < \sum_{i=1}^{n} |F_N(y_i) - F_N(x_i)| + \frac{\varepsilon}{2(b-a)} \sum_{i=1}^{n} (y_i - x_i)$$
$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

This verifies that F has zero variation on C. The rest of the proof is now apparent.

Exercise 41, page 44

Suppose that $f(x) = \chi_I(x)$ where $I \subset [a,b]$ is an interval $I = [c,d]$, (c,d), $[c,d]$, or $(c,d]$. The case $a < c < d < b$ is enough to consider. Define $F(x) = 0$ for $a \leq x \leq c$, Define $F(x) = x - c$ for $c \leq x \leq d$. Define $F(x) = d - c$ for $d \leq x \leq a$. Then $F'(x) = f(x)$ for all x except at the two points c and d. F is continuous on $[a,b]$. Consequently

$$\int_a^b f(x)\,dx = \int_a^b \chi_I(x)\,dx = F(b) - F(a) = d - c = \lambda(I).$$

Exercise 43, page 44

Every step function $s : [a,b] \to \mathbb{R}$ is bounded there, is continuous on $[a,b]$ with at most countably many [actually finitely many] exceptions, and has finite one-sided limits at each point of $[a,b]$. Each of these properties is preserved under uniform limits.

Exercise 44, page 44

We know from Exercise 43 that every regulated function has finite one-sided limits at each point of $[a,b]$. We prove the other direction and assume that f has finite one-sided limits at each point of $[a,b]$.

Let n be an integer. At each point $x \in [a,b]$ select $\delta(x) > 0$ so that the oscillation of f in the intervals $[a,b] \cap (x-\delta(x),x)$ and $[a,b] \cap (x,x+\delta(x)$ is smaller than $1/n$. Since f has finite one-sided limits at each x this can be done.

Use the Cousin lemma to select a partition

$$\{([a_i,b_i],\xi_i) : i = 1,2,3,\ldots,n\}$$

of the interval $[a,b]$ finer than δ. We claim that we can select a step function $g_n : [0,1] \to \mathbb{R}$ so that

$$|f(x) - g_n(x)| < 1/n$$

for all $x \in [0,1]$.

Take any pair $([a_i,b_i],\xi_i)$ from the partition. If $a_i < \xi_i < b_i$ then we just arrange for g to assume a constant value (chosen from the values of f itself) on each interval (a_i,ξ_i) and (ξ_i,b_i). If $a_i = \xi_i < b_i$ or $a_i < \xi_i = b_i$ then we do much the same on (a_i,b_i). There are a finite number of points remaining and we simply make f and g_n agree at those points. For example, at the points ξ_i we would set $g_n(\xi_i) = f(\xi_i)$.

By Exercise 42 we can see that g_n is a step function. Since $\{g_n\}$ is a sequence of step functions converging uniformly to f, the function f must be regulated.

Exercise 46, page 45

Every monotonic function clearly has one-sided limits at each point. Thus, Exercise 44, such functions are regulated.

Exercise 47, page 45

Set $f = F'$. The function f is continuous on any interval $[c,d] \subset (a,b)$. In particular the integral

$$\int_c^d f(x)\,dx$$

can be constructed by the methods of this chapter (as, say a limit of a sequence of simpler integrals of step functions). This means that one can compute

$$F(b) - F(a) = \lim_{m\to\infty} (F(b-m^{-1}) - F(a+m^{-1})$$

and so

$$\int_a^b f(x)\,dx = \lim_{m\to\infty} \int_{a+m^{-1}}^{b-m^{-1}} f(x)\,dx.$$

The point of the exercise is just the observation that "constructibility" of the integral does not necessarily mean as a uniform limit of integrals of simpler functions.

Exercise 48, page 45

There are a number of ways to do this. One is just to show that if f has finite one-sided limits at each point then the function $|f|$ has the same property.

Exercise 49, page 45

Just show that if f is piecewise continuous and has finite one-sided limits at the discontinuities then the function $|f|$ has the same property.

Exercise 50, page 48

Use one or more of these criteria to deduce that whenever a function f is Riemann integrable on an interval $[a,b]$ so too is the function $|f|$. The simplest, perhaps, is the Lebesgue criterion. If f is bounded and a.e. continuous then $|f|$ is clearly bounded and easily shown to be a.e. continuous.

Exercise 51, page 52

The earliest proof of this statement that I can find is Gillespie [35] from 1915, although one can imagine that by that time it was common knowledge that restricting the Riemann sums to use only left-end associated points would not alter the Riemann integral.

Gillespie's proof uses the fact that, if f is bounded and not Riemann integrable, then there must exist a closed set of positive measure such that the oscillation of the function at each point in the set exceeds some positive number $c > 0$. He uses that to show that f cannot be Cauchy integrable. Bressoud [9, p. 300] also proposes this as an exercise in his text, but it seems the hint there is not quite correct.

Exercise 52, page 53

Let N be the measure zero set where $f(x) \neq g(x)$. Let $\varepsilon > 0$ and choose a positive function $\delta_1 : N \to \mathbb{R}^+$ so that

$$\sum_{i=1}^{n} [|f(\xi_i)| + |g(\xi_i)|]\,(y_i - x_i) < \varepsilon/3$$

whenever a subpartition of $[a,b]$

$$\{([x_i, y_i], \xi_i) : i = 1, 2, 3, \dots\}$$

is anchored in N and finer than δ_1. This just uses the definition (i.e., the small Riemann sums property of measure zero sets).

Choose a positive number δ_2 so that both

$$\left|\int_a^b f(x)\,dx - \sum_{i=1}^n f(\xi_i)(y_i - x_i)\right| < \varepsilon/3$$

and

$$\left|\int_a^b g(x)\,dx - \sum_{i=1}^n g(\xi_i)(y_i - x_i)\right| < \varepsilon/3$$

whenever a partition of $[a,b]$

$$\{([x_i,y_i],\xi_i) : i = 1,2,3,\dots\}$$

is given that is finer than δ_2.

Let $\delta : [a,b] \to \mathbb{R}^+$ be chosen smaller than δ_2 and smaller than δ_1 on N. Using the Cousin lemma, choose one partition of $[a,b]$

$$\{([x_i,y_i],\xi_i) : i = 1,2,3,\dots\}$$

that is finer than δ. In the obvious way you should now be able to use that partition to deduce that

$$\left|\int_a^b f(x)\,dx - \int_a^b g(x)\,dx\right| < \varepsilon.$$

From that the identity is evident.

As to finding an example for which f is Riemann integrable and g is not this is should present no difficulties. You can even arrange to have g unbounded, or bounded but continuous nowhere.

Exercise 53, page 55

This is the converse of Theorem 1.24 and is very easy to prove. Suppose that

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)| < \varepsilon$$

for every partition of $[a,b]$ finer than some δ. Then immediately this shows that

$$\left|F(b) - F(a) - \sum_{([u,v],w)\in\pi} f(w)(v-u)\right|$$

$$= \left|\sum_{([u,v],w)\in\pi} [F(v) - F(u) - f(w)(v-u)\right|$$

$$\leq \sum_{([u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)| < \varepsilon.$$

Use this to deduce that $F(b) - F(a) = \int_a^b f(x)\,dx$

Exercise 54, page 55

Compare with the definition of the Riemann integral.

Exercise 55, page 61

The reader should remark upon the fact that this exercise is quite easy to handle; it follows with little trouble from the definitions of measure zero and zero variation.

An inspection of the five statements shows that any one of the first four implies statement 5. So it is enough to assume that the final statement does hold. Thus we assume that F has zero variation on a set N of measure zero, while $F'(x) = f(x)$ at every point of (a,b) except for $x \in N$. We shall apply the integrability criterion in Exercise 54 to provide a proof.

Choose a positive function $\delta_1 : N \to \mathbb{R}^+$, so that

$$\sum_{([u,v],w)\in\pi} |f(w)|\,(v-u) < \varepsilon/4$$

whenever π is a subpartition finer than δ_1 with associated points in N. Choose a positive function $\delta_2 : N \to \mathbb{R}^+$, so that

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u)| < \varepsilon/4$$

whenever π is a subpartition finer than δ_2 with associated points in N.

Since $F'(w) = f(w)$ for all $w \notin N$, choose $\delta_3(w)$ so that

$$\left|\frac{F(v)-F(u)}{v-u} - f(w)\right| < \frac{\varepsilon}{2(b-a)}$$

whenever $0 < v-u < \delta(w)$ with $w \in [u,v]$.

Now we can construct our function $\delta(x)$ on $[a,b]$ so that $\delta(x) = \delta_3(x)$ if $x \in N$ and $\delta(x) = \min\{\delta_1(x), \delta_2(x)\}$ otherwise. Let π be a subpartition of the interval $[a,b]$ finer than δ. We estimate

$$\sum_{[u,v],w)\in\pi} |F(v) - F(u) - f(w)(v-u)|\,.$$

If w is a point where $F'(w) = f(w)$ then

$$|F(v) - F(u) - f(w)(v-u)| \leq \varepsilon\frac{v-u}{2(b-a)}$$

and the contribution to the sum of such terms is evidently smaller than $\varepsilon/2$.

If w is not a point where $F'(w) = f(w)$ then we can treat the sum of such terms by estimating using the larger value

$$|F(v) - F(u) - f(w)(v-u)| \leq |F(v) - F(u)| + |f(w)|(v-u).$$

The sum of the terms

$$|F(v) - F(u)|$$

where $w \in N$ cannot exceed $\varepsilon/4$. The sum of the terms

$$|f(w)|(v-u)$$

where $w \in N_j$ cannot exceed $\varepsilon/4$. Putting these together shows that

$$\sum_{[u,v],w)\in\pi} |F(v)-F(u)-f(w)(v-u)| < \varepsilon.$$

as required. An application of Exercise 54 now shows that f is Henstock-Kurzweil integrable on $[a,b]$ and that

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

Exercise 56, page 63

If $f : [a,b] \to \mathbb{R}$ is bounded and Riemann integrable on all subintervals $[c,d] \subset (a,b)$ then, by the Lebesgue criterion we can deduce that f is a.e. continuous on $[a,b]$. Consequently f is in fact Riemann integrable on all of $[a,b]$.

Exercise 57, page 63

Take the function $f : [0,1] \to \mathbb{R}$ defined by $f(1/n) = 1$ for each $n = 1,2,3,\dots$ and $f(x) = 0$ elsewhere. Check that f is integrable in the elementary Newton sense in any interval $[c,1]$ for $0 < c < 1$ and that

$$\int_c^d f(x)\,dx = 0$$

for all $0 < c < d \leq 1$. Check that it is not, however, integrable in that sense on $[0,1]$.

Exercise 58, page 63

We considered this function in Section 1.1.3. We learned there that F' was Newton integrable but that $|F'|$ was not integrable, because the function F does not have bounded variation on $[0,1]$.

In fact F is differentiable at every point and $F'(0) = 0$ while, for $0 < x \leq 1$,

$$F'(x) = 2x\sin x^{-2} - \frac{2}{x}\cos x^{-2}.$$

Note that F' is continuous on $(0,1]$ so that it is integrable by all of our methods on each interval $[c,1]$ for $0 < c < 1$.

In order to show that F' is not integrable on $[0,1]$ in the Riemann sense it would be enough to observe that F' is unbounded. To show that F' is not integrable on $[0,1]$ in the Lebesgue sense we simply recall that the Lebesgue integral (by definition) is an absolute integral. We know that $|F'|$ is not integrable.

Thus while F' is integrable, it is not Lebesgue integrable. [It is nonabsolutely integrable in other words.] This verifies that the Lebesgue integral does not possess the Cauchy property.

Exercise 62, page 65

The result in the exercise was published by Beppo Levi in 1906 (Ricerche sulle funzioni derivate, Rnd. dei Lincei, (5), Vol. XV, 1906 p. 437.).

The proof is well-worth attempting and also well-worth studying. It contains a number of ideas that will prove useful in similar situations.

We suppose that $F : [a,b] \to \mathbb{R}$ and

$$A = \{x \in (a,b) : D^+F(x) < D^-F(x)\}$$

and

$$B = \{x \in (a,b) : D^-F(x) < D^+F(x)\}.$$

(Some early authors [eg., A. Rosenthal and W. Sierpiński] called such points on the graph of a function “angular points” or “cusps.”)

We show that A is countable. A similar argument would show that B is countable.

Inserting rational numbers r and s Note that if

$$D^+F(x) < D^-F(x)$$

at some point x we can select two rational numbers r and s so that

$$D^+F(x) < r < s < D^-F(x).$$

This means that A is contained in the union of the sets

$$A_{rs} = \{x \in (a,b) : D^+F(x) < r < s < D^-F(x)\}$$

taken over the countable collection of rational numbers r and s. Thus one needs only to show that each set A_{rs} is countable [since a countable union of countable sets is also countable].

Setting up the function δ For each $x \in A_{rs}$ we know that

$$D^+F(x) < r.$$

Thus we may select $\delta_1(x) > 0$ so that

$$\frac{F(y) - F(x)}{y - x} < r$$

if $0 < y - x < \delta_1(x)$. Since it is also true that

$$D^-F(x) > s$$

we can select also $\delta_2(x) > 0$ so that
$$\frac{F(y) - F(x)}{y - x} > s$$
if $0 > y - x > -\delta_1(x)$. Set $\delta(x)$ as the minimum of $\delta_1(x)$ and $\delta_1(x)$.

Decomposing a set using δ Fix an integer $n \geq 1$ and consider the set
$$A_{rsn} = \{x \in A_{rs} : \delta(x) > 1/n\}.$$
This decomposes the set A_{rs} into a sequence of sets so that
$$A_{rs} = \bigcup_{n=1}^{\infty} A_{rsn}.$$

If x_1 and x_2 with $x_1 < x_2$ are two points in A_{rsn} closer together than $1/n$ it follows that $0 > x_1 - x_2 > -\delta(x_2) \geq -\delta_1(x_2)$ and $0 < x_2 - x_1 < \delta(x_1) \leq \delta_2(x_1)$. Consequently
$$\frac{F(x_1) - F(x_2)}{x_1 - x_2} > s$$
$$\frac{F(x_2) - F(x_1)}{x_2 - x_1} < r$$
must both be true, which is evidently impossible since $r < s$. This can only mean that two such points in A_{rsn} do not in fact exist. Consequently A_{rsn} must be a finite set. It follows that A_{rs} is the union of a sequence of finite sets and so is countable.

Note: The steps here can be used in a variety of situations. The decomposition of the set using δ will be formalized in Section 2.2.3 to make it rather easier to use.

Exercise 63, page 66

Hint: The set of shaded points is open. Just check that every shaded point is contained in an open interval of points that are also shaded. Now take $\{(a_k, b_k)\}$ as the sequence of component intervals of that open set. Check that if $H(a_k) > H(b_k)$ for some k, then that would contradict the fact that a_k and b_k are not shaded points.

Exercise 64, page 67

Suppose that $G(a) \leq G(b)$, Take a finite number of the intervals as given
$$[a_1, b_1], [a_2, b_2], [a_3, b_3], \ldots, [a_N, b_N]$$
and then supplement them with further nonoverlapping intervals
$$[c_1, d_1], [c_2, d_2], [c_3, d_3], \ldots, [c_M, d_M]$$

to form a full partition of $[a,b]$. Just write

$$|G(b)-G(a)| = G(b)-G(a) = \sum_{i=1}^{N}[G(b_i)-G(a_i)] + \sum_{j=1}^{M}[G(d_i)-G(c_i)]$$
$$\leq \sum_{i=1}^{N}(G(b_k)-G(a_k)) + \mathrm{Var}(G,[a,b]).$$

The first statement of the lemma follows.

For the second statement, replace G with $-G$. Or else just repeat the same steps with a minor change. Thus suppose now that $G(b) \leq G(a)$, Again take a finite number of the intervals

$$[a_1,b_1],[a_2,b_2],[a_3,b_3],\dots,[a_N,b_N]$$

and then supplement them with further nonoverlapping intervals

$$[c_1,d_1],[c_2,d_2],[c_3,d_3],\dots,[c_M,d_M]$$

to form a full partition of $[a,b]$. Write

$$|G(b)-G(a)| = G(a)-G(b) = -\sum_{i=1}^{N}[G(b_i)-G(a_i)] - \sum_{j=1}^{M}[G(d_i)-G(c_i)]$$
$$\leq -\sum_{i=1}^{N}(G(b_k)-G(a_k)) + \mathrm{Var}(G,[a,b]).$$

The second statement of the lemma follows.

Exercise 65, page 73

Consider these two computations:

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\frac{1}{n}\chi_{[0,n]}(x)\,dx = 1$$

and

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\frac{1}{n}\chi_{[0,n^2]}(x)\,dx = \infty.$$

Exercise 66, page 73

First of all, to avoid leaping to conclusions, note the fact that the integrand is undefined at $x=0$ is entirely unimportant unless you have been schooled only in the Riemann theory of integration. That should be ignored. The integrand is continuous at every point provided one assigns the value 1 at $x=0$.

The curious thing about the example is that, while calculus students are entirely prepared to accept a situation such as

$$\int_{-\infty}^{\infty} f(x)\,dx = C \quad \text{and} \quad \int_{-\infty}^{\infty}|f(x)|\,dx = \infty,$$

students of the Lebesgue integral are not. All Lebesgue integrable functions are absolutely integrable. Thus

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}\, dx = \pi$$

cannot be accepted in the Lebesgue theory. One could, however, compute

$$\lim_{b\to\infty} \int_0^b \frac{\sin x}{x}\, dx = \pi/2$$

by conventional means and then claim to interpret the infinite integral in some "improper" Lebesgue sense. The computations are of some intrinsic interest but the example, from the point of view of integration theory, merely points out that an infinite integral need not and should not be interpreted only as an absolute integral.

Exercise 68, page 80

Take any particular point x in E and check that $\beta(G)$ is full at that point x. Remember that, since G is open, there is a positive number δ_1 so that $(x-\delta_1, x+\delta_1) \subset G$. There is also a positive number δ_2 so that all pairs $([u,v],x)$ with $x \in [u,v]$ and $0 < v-u < \delta_2$ must belong to β.

Exercise 69, page 80

This is nearly identical to the preceding exercise, Exercise 68.

Exercise 81, page 81

This is a dual of the next exercise, Exercise 82.

Exercise 82, page 81

This is a dual of the preceding exercise, Exercise 81.

Exercise 85, page 82

For each x in E there would have to be at least one interval $(x, x+c)$ or $(x-c, x)$ that does not contain any points of the sequence.

Exercise 86, page 82

There would have to be at least one point x_0 in E at which β is not fine. That would mean that all intervals $(x, x+c)$ and $(x-c, x)$ contain infinitely many points of the sequence.

Exercise 87, page 82

For each x in E there would have to be at least one interval $(x-c, x+c)$ that does not contain any points of the sequence other than possibly x itself.

Exercise 88, page 82

There would have to be at least one point x_0 in E at which β is not full. That would mean that all intervals (x_0, x_0+c) or else all intervals (x_0-c, x_0) contain infinitely many points of the sequence.

Exercise 89, page 82

For each x in E every interval $(x, x+c)$ or else every interval $(x-c, x)$ contains infinitely many points of the sequence.

Exercise 90, page 82

There would have to be at least one point x_0 in E at which β is not fine. Thus some interval (x_0, x_0+c) or else some interval (x_0-c, x_0) contains no points of the sequence.

Exercise 91, page 83

For each x in E every interval $(x, x+c)$ and also every interval $(x-c, x)$ contains infinitely many points of the sequence.

Exercise 92, page 83

There would have to be at least one point x_0 in E at which β is not full. Thus some interval (x_0, x_0+c) or else some interval (x_0-c, x_0) contains no points of the sequence.

Exercise 93, page 85

Recall from Exercise 83 the following fact.

Let $\mathcal{G}$ be a family of open sets so that every point in a nonempty, compact set K is contained in at least one member of the family. Then the covering relation

$$\beta_1 = \{(I,x) : x \in I \text{ and } I \subset G \text{ for some } G \in \mathcal{G}\}.$$

is a full cover of K. Let

$$\beta_2 = \{(I,x) : x \notin K \text{ and } I \cap K = \emptyset\ \}$$

and check that this is a full cover of $\mathbb{R} \setminus K$.

Consequently $\beta = \beta_1 \cup \beta_2$ is a full cover. Write $a = \min K$ and $b = \max K$. Choose a partition $\pi \subset \beta$ of the interval $[a,b]$. Then corresponding to each $(I,x) \in \pi[K]$ is an open set from $\mathcal{G}$ covering I. This gives finitely many such open sets covering the compact set K. We have proved:

> **[Heine-Borel]** Let $\mathcal{G}$ be a family of open sets covering a compact set K. Then there are finitely many open sets G_1, G_2, ..., G_n from $\mathcal{G}$ that also cover K.

Exercise 94, page 85

Recall from Exercise 84 the following fact.

Let E be an infinite set that contains no points of accumulation. Then

$$\beta = \{(I,x) : x \in I \text{ and } I \cap E \text{ is finite}\}.$$

must be a full cover.

Take any bounded interval $[a,b]$ and choose a partition $\pi \subset \beta$ of the interval $[a,b]$. Then, evidently, $E \cap [a,b]$ is a finite set. We have proved:

> **[Bolzano-Weierstrass]** Every infinite bounded set of real numbers must have a point of accumulation.

Exercise 95, page 85

Let E be the set of points x where $F'(x) \geq 0$ and let $z_1, z_2, z_3, \ldots$ be a list of the remaining points. Let $\varepsilon > 0$.

Define

$$\beta_1 = \{([u,v],w) : F(v) - F(u) > -\varepsilon(v-u)\}.$$

This can be checked to be a full cover of E. Define

$$\beta_2 = \left\{([u,v],w) : w = z_n \text{ for some } n \text{ and } F(v) - F(u) > -\varepsilon 2^{-n}\right\}.$$

Check that $\beta = \beta_1 \cup \beta_2$ is a full cover. Accordingly, by the Cousin covering lemma, if $[a,b]$ is any interval then there are are points

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

and associated points $\{\xi_i\}$ so that

$$\pi = \{([x_{i-1},x_i],\xi_i) : i = 1,2,\ldots,n\} \subset \beta.$$

Hence

$$F(b) - F(a) = \sum_{i=1}^{n} [F(x_i) - F(x_{i-1})] > -\varepsilon \left[b - a + \sum_{n=1}^{\infty} 2^{-n} \right] = -\varepsilon [b - a + 1].$$

Since ε is an arbitrary positive number $F(b) \geq F(a)$.

Exercise 96, page 85

Define

$$\beta = \{([u,v],w) : F \text{ is bounded below on } [u,v]\}.$$

Use the fact that F is lower semicontinuous to check this to be a full cover. Apply the Cousin covering lemma.

Exercise 97, page 85

One direction is easy. If F is Lipschitz then, for some number M,

$$|f(x) - f(y)| \leq M|x-y|$$

for all x, y. In particular, for all $h \neq 0$,

$$\left|\frac{F(x+h)-F(x)}{h}\right| \leq \left|\frac{M|(x+h)-x|}{h}\right| = M.$$

The other direction uses a simple covering argument. Suppose that F satisfies the stated condition and define

$$\beta = \{([u,v],w) : \ w \in [u,v] \text{ and } \ |F(v)-F(u)| \leq M(v-u)\}.$$

This is evidently a full cover. Take any interval $[c,d]$. By the Cousin lemma there is a partition

$$\{([x_{i-1},x_i],\xi_i) : i = 1,2,3,\dots,n\}$$

of the whole interval $[c,d]$ contained in β. For this partition

$$|F(d)-F(c)| \leq \sum_{i=1}^{n} |F(x_i)-F(x_{i-1})| \leq \sum_{i=1}^{n} M|x_i - x_{i-1}| = M(d-c).$$

Thus F is Lipschitz.

Exercise 104, page 88

The former represents a sum taken over *all* elements in the partition π while the latter sum contains only those elements (if any) $([u,v],w) \in \pi$ for which $[u,v]$ is a subinterval of $[c,d]$. It is the usual convention to consider that an empty sum has value zero.

Exercise 105, page 89

The former represents a sum taken over *all* elements in the partition π while the latter sum contains only those elements (if any) $([u,v],w) \in \pi$ for which w belongs to the set E. It is the usual convention to consider that an empty sum has value zero.

Exercise 110, page 91

It satisfies the definition easily, with $G = \emptyset$ in fact.

Exercise 111, page 91

If

$$E = \{x_1, x_2, \ldots x_N\}$$

and $\varepsilon > 0$, then the sequence of intervals

$$\left(x_i - \frac{\varepsilon}{2N}, x_i + \frac{\varepsilon}{2N}\right) \quad i = 1, 2, 3, \ldots, N$$

covers the set E and the sum of all the lengths is ε. The union of these intervals is an open set G that contains E; by the subadditivity property the Lebesgue measure $\lambda(G)$ is smaller than ε.

Exercise 112, page 91

If

$$E = \{x_1, x_2, \ldots\}$$

and $\varepsilon > 0$, then the sequence of intervals

$$\left(x_i - \frac{\varepsilon}{2^{i+1}}, x_i + \frac{\varepsilon}{2^{i+1}}\right) \quad i = 1, 2, 3, \ldots$$

covers the set E. Let G be the union of these intervals. Since

$$\sum_{k=1}^{\infty} 2\left(\frac{\varepsilon}{2^{k+1}}\right) = \sum_{k=1} \varepsilon 2^{-k} = \varepsilon,$$

we conclude (from Lemma 2.9) that $\lambda(G) < \varepsilon$.

Exercise 113, page 91

Let $\varepsilon > 0$. Choose n so that $(2/3)^n < \varepsilon$. Then the nth stage intervals in the construction of the Cantor set give us 2^n closed intervals each of length $(1/3)^n$. This covers the Cantor set with 2^n closed intervals of total length $(2/3)^n$, which is less than ε. If the closed intervals trouble you (the definition requires open intervals), see Exercise 116 or argue as follows. Since $(2/3)^n < \varepsilon$ there is a positive number δ so that

$$(2/3)^n + \delta < \varepsilon.$$

Enlarge each of the closed intervals to form a slightly larger open interval, but change the length of each only enough so that the sum of the lengths of all the 2^n closed intervals does not increase by more than δ. The resulting collection of open intervals also covers the Cantor set, and the sum of the length of these intervals is less than ε. Thus the Cantor set has Lebesgue measure zero.

Exercise 115, page 92

Since E has Lebesgue measure zero, there is an open set G containing E for which $\lambda(G) < \varepsilon$. Let $\{(a_k, b_k)\}$ denote the component intervals of G. By the Heine-Borel theorem there is a finite N so that

$$\{(a_k, b_k) : k = 1, 2, \ldots, N\}$$

covers the set E. Since

$$\sum_{k=1}^{N} (b_k - a_k) \leq \lambda(G) < \varepsilon.$$

the proof is complete.

Exercise 118, page 92

You will have to decide on which of the integrals of Chapter 1 to use to answer this question. The general Newton integral or (equivalently) the Henstock-Kurzweil integral would be needed in general. The Riemann integral would not exist for all choices of sets E that have measure zero.

Exercise 141, page 95

You will have to decide on which of the integrals of Chapter 1 to use to answer this question. For example you might wish to assume that f is Riemann integrable and try this. cf. Exercise 118.

Exercise 155, page 103

We know from Lemma 2.20 that, if E is an interval and if F has zero variation on E, then F is constant on E. If E is open then F would be constant on each component of E, but need not be constant on E itself. In general, though, the fact that F has zero variation on a set E that contains no intervals would not say much about whether F is constant or not.

Exercise 157, page 103

Don't forget to include the statement that F must be defined on an open interval that contains the point x_0. You should verify that it means precisely that F is defined on an open set containing the point x_0 and is continuous at that point.

Exercise 166, page 108

You might develop the proof using these easy steps first:

(A) Suppose that $G : [c,d] \to \mathbb{R}$ satisfies
$$\frac{G(y)-G(x)}{y-x} \geq r > 0$$
for all $x \neq y$ in $[c,d]$ and that $E \subset [c,d]$ has $G(E)$ a set of Lebesgue measure zero. Then E also has Lebesgue measure zero.

(A) Suppose that $F : [c,d] \to \mathbb{R}$ and $E \subset [c,d]$ satisfies
$$\frac{F(y)-F(x)}{y-x} \geq r > 0$$
for all $x \in E$, $y \in [c,d]$ with $x \neq y$. Then there is a function $G : [c,d] \to \mathbb{R}$ that satisfies
$$\frac{G(y)-G(x)}{y-x} \geq r > 0$$
for all $x \neq y$ in $[c,d]$ and $F(x) = G(x)$ for all $x \in E$.

Now to prove the statement in the exercise. It is enough to assume that $F'(x) > r > 0$ for every $x \in E$. For each $u \in E$ choose $\delta(u) > 0$ so that
$$\frac{F(v)-F(u)}{v-u} > r. \tag{8.1}$$
for $|u-v| < \delta(u)$. Now, as in the proof of Theorem 2.26, we apply one of our standard decomposition methods to obtain an increasing sequence of sets
$$E_n = \{x \in E : \delta(x) > 1/n\}$$
whose union is E. Take any interval $[c,d]$ with length less than $1/n$ and observe that
$$\frac{F(y)-F(x)}{y-x} \geq r > 0$$
for all $x \in E_n \cap [c,d]$, $y \in [c,d]$ with $x \neq y$ in $[c,d]$. Then, by statement (B), there is a function $G : [c,d] \to \mathbb{R}$ that satisfies
$$\frac{G(y)-G(x)}{y-x} \geq r > 0$$
for all $x \neq y$ in $[c,d]$ and $F(x) = G(x)$ for all $x \in E_n \cap [c,d]$. Note that $F(E_n \cap [c,d]) = G(E_n \cap [c,d])$ is a set of measure zero. Then by statement (A) it follows that $E_n \cap [c,d]$ is also a set of measure zero.

Deduce now from this that E itself must also be a set of measure zero.

Exercise 173, page 112

Use Lemma 2.32.

Exercise 174, page 112

Let C the collection of points in (a,b) at which there is no derivative. This is countable and, since F is continuous, F has zero variation on C. Now take any Lebesgue measure zero set $N \subset (a,b)$. We know that F has zero variation on $C \cap N$ and, by Lemma 2.32, we know that F has zero variation on $N \setminus C$. It follows that F has zero variation on N.

Exercise 180, page 125

Do not attempt this exercise without gaining first some experience in applications of the Hölder and Minkowski inequalities. A proof can be found in Natanson [63, p. 257]. Note that the condition

$$\sum_{i=1}^{n} \frac{|F(x_i) - F(x_{i-1})|^p}{(x_i - x_{i-1})^{p-1}} \leq K$$

in Riesz's criterion is equivalent to the inequality

$$\sum_{i=1}^{n} \left| \frac{F(x_i) - F(x_{i-1})}{(x_i - x_{i-1})} \right|^p (x_i - x_{i-1}) \leq K,$$

which is closer in appearance to a typical Riemann sum.

A related theorem of F. Riesz is also interesting in that it expresses this same idea as kind of pth power absolute continuity condition.

Theorem 8.1 (F. Riesz) *Let $1 < p < \infty$. A necessary and sufficient condition in order for a function $F : [a,b] \to \mathbb{R}$ to be representable in the form*

$$F(x) = C + \int_a^x f(t)\,dt \quad (a \leq x \leq b)$$

for some constant C and for some function f on $[a,b]$ with the property that $|f|^p$ is integrable is that, for all $\varepsilon > 0$ a positive δ can be found so that

$$\sum_{i=1}^{n} \frac{|F(b_i) - F(a_i)|^p}{(b_i - a_i)^{p-1}} < \varepsilon$$

for any system $\{(a_i,b_i)\}$ of pairwise disjoint subintervals of (a,b) satisfying $\sum_{i=1}^{n}(b_i - a_i) < \delta$.

Exercise 182, page 128

The language of the upper and lower integrals alone suggests that upper integrals should exceed lower integrals. It is best not to take this for granted, but to prove this fact in this exercises. Make use in your proof of the fact that the intersection of two full covers, is again a full cover.

Exercise 185, page 128

cf. Exercise 52.

Exercise 186, page 129

Infinite values are allowed but we would have to avoid $\infty + (-\infty)$ or $-\infty + \infty$. This is simpler if you first check that a single value $f(b)$ is irrelevant to the computations so that you may assume that $f(b) = 0$. Then ensure that any partition π contained in your choice of β of the interval $[a,b]$, $[a,c]$ or $[b,c]$ would have to contain an element (I,b).

Exercise 187, page 129

Check, first, that full covers do in fact contain endpointed partitions (as well as ordinary partitions). Then note that, if a partition π contains a pair $([u,v],w)$ for which $u < w < v$ that element can be replaced by the two items $([u,w],w)$ and $([w,v],w)$. That does not change the Riemann sums here because, for example,

$$f(w)[v-u] = f(w)[w-u] + f(w)[v-w].$$

Finally check that if β is a full cover there must be a smaller full cover $\beta' \subset \beta$ so that $([u,v],w) \in \beta'$ with $u < w < v$ if and only if both $([u,w],w)$ and $([w,v],w)$ are in β'.

Exercise 188, page 129

Use β to find estimates for the upper and lower integrals,

$$\overline{\int_a^b} f(x)\,dx - \underline{\int_a^b} f(x)\,dx < 2\varepsilon.$$

(Later we will show that this condition is, in fact, both necessary and sufficient.

Exercise 194, page 143

First show that the sequence

$$\left\{ \int_a^b f_n(x)\,dx \right\}$$

is a Cauchy sequence and, hence, converges to a real number I. Then it is almost obvious how to exploit the inequality

$$\left| \int_a^b f_n(x)\,dx - \sum_{[u,v],w)\in\pi} f_n(w)(v-u) \right| < \varepsilon$$

that is available from the hypothesis that $\{f_n\}$ is a sequence of equi-integrable functions on $[a,b]$.

Exercise 195, page 143

This question was posed on-line in the MathOverFlow forum. Professor Gerry Edgar suggested this:

> "What if you try these functions $\{f_n\}$ on $[0,1]$ ... piecewise linear, with $f_n(0)=0$, $f_n(1/n)=1$, $f_n(2/n)=1$, $f_n(3/n)=4$, $f_n((1)=0$."

Exercise 196, page 143

Among other ideas, you might want to think about the following questions:

1. If a sequence of functions $\{f_n\}$ is equicontinuous must it be equi-integrable in either the Henstock-Kurzweil or Riemann senses?

2. Can a condition that uses an inequality of the form
$$\sum_{([u,v],w)\in\pi} \omega f_n([u,v])(v-u) < \varepsilon$$
characterize equi-integrability in the Riemann sense?

Exercise 197, page 147

Check first that you need only prove the case where $E \subset (a,b)$. Then it is just a matter of looking carefully at the definitions of the two concepts.

Exercise 198, page 147

Use the subadditive property of open sets expressed in Lemma 2.9.

Exercise 207, page 161

Just apply Exercise 206.

Exercise 208, page 162

Write $h = f \circ g$. Let $G \subset \mathbb{R}$ be an arbitrary open set. Note that $h^{-1}(G) = f^{-1}\left(g^{-1}(G)\right)$. Since G is open and g is continuous we know that $g^{-1}(G)$ is also open. Then by Exercise 204 we conclude that $f^{-1}\left(g^{-1}(G)\right)$ is measurable. Since $h^{-1}(G)$ is measurable for every choice of open set G it follows (from Exercise 204 yet again) that h is measurable.

It is clear that the argument would not apply in the different order $g \circ f$. To complete the exercise, however, requires finding a suitable counterexample. Begin by showing that there is a continuous increasing function ϕ and a measurable set E for which $\phi(E)$ is not measurable. Let g be the function inverse to ϕ and let $f = \chi_E$.

Exercise 209, page 164

Let A be the set of *all* points where a function f has a finite derivative. We know that each set of the form

$$\{x : \overline{D}f(x) > c\}$$

and

$$\{x : \underline{D}f(x) < c\}$$

is measurable. Thus the set A' of points at which f does not have a derivative (finite or infinite) can be expressed as the union of the family of sets

$$A_{pq} = \{x : \underline{D}f(x) < p < q < \overline{D}f(x)\}$$

for rational numbers p and q. It follows that A' is also measurable. Again the set A'' of points where $f'(x) = \pm\infty$ can be written as

$$A'' = \bigcap_{n=1}^{\infty} \{x : \overline{D}f(x) < -n\} \cup \bigcap_{n=1}^{\infty} \{x : \underline{D}f(x) > n\}.$$

Thus this set is measurable. But $A = A' \setminus A''$.

Exercise 213, page 173

Apply Fatou's lemma to the non-negative sequence given by $g - f_n$.

Exercise 218, page 182

Let $\varepsilon > 0$ and suppose that $F'(x) = f(x)$ at every point. Define

$$\beta = \{([u,v],w) : |F(v) - F(u) - f(w)(v-u)| \leq< \varepsilon(v-u).$$

Check that β is a full cover of $\mathbb{R}$ and that it satisfies (4.21).

Conversely suppose that β has been chosen to be a full cover of $\mathbb{R}$ that satisfies (4.21). Fix $x \in \mathbb{R}$ and determine $\delta > 0$ so that whenever (I,x) satisfies $x \in I$ and $\lambda(I) < \delta$ then necessarily $(I,x) \in \beta$. Note that if $([c,d],x)$ is any pair for which $c \leq x \leq d$ and $d - c < \delta$ then necessarily $([c,d],x)$ is in β and the set π containing only this one pair is itself also a partition of $[c,d]$. Consequently, using (4.21),

$$|F(d) - F(c) - f(x)(d-c)| < \varepsilon(d-c).$$

But this verifies that $F'(x) = f(x)$.

Exercise 219, page 182

Suppose that $F'(x) = f(x)$ everywhere. Apply Lemma 218 to find a full cover β for which for every compact interval $[a,b]$ and every partition $\pi \subset \beta$ of $[a,b]$,

$$\sum_{(I,x)\in\pi} |\Delta F(I) - f(x)\lambda(I)| < \varepsilon\lambda([a,b])/2. \tag{8.2}$$

Let π_1, $\pi_2 \subset \beta$ be partitions of $[a,b]$. Apply (8.2) to each of them and add to obtain that

$$\left|\sum_{(I,z)\in\pi} f(z)\lambda(I) - \sum_{(I',z')\in\pi'} f(z')\lambda(I')\right| < \varepsilon\lambda([a,b]). \tag{8.3}$$

Now simply rearrange (8.3) to obtain (4.22). Conversely suppose that the statement (4.22) in the theorem holds for ε and β. This is a stronger statement than the Cauchy criterion and so f is integrable on every compact interval. Thus there is a function F that will serve as the indefinite integral for f on any interval. From (4.22) we deduce that

$$\left|\Delta F([a,b]) - \sum_{(I,z)\in\pi} f(z)\lambda(I)\right| < 2\varepsilon\lambda([a,b]) \tag{8.4}$$

must be true for any partition π of $[a,b]$ from the cover β.

Fix $x\in\mathbb{R}$ and determine $\delta>0$ so that whenever (I,x) satisfies $x\in I$ and $\lambda(I)<\delta$ then necessarily $(I,x)\in\beta$. Note that if $([c,d],x)$ is any pair for which $c\le x\le d$ and $d-c<\delta$ then necessarily $([c,d],x)$ is in β and the set π containing only this one pair is itself also a partition of $[c,d]$. Consequently, using (8.4),

$$|F(d)-F(c)-f(x)(d-c)| < 2\varepsilon(d-c).$$

But this verifies that $F'(x)=f(x)$.

Exercise 225, page 184

Yes and no, uniform convergence is very strong, but the integrals are defined in very special ways as antiderivatives with exceptional sets.

Suppose that f_n are Newton integrable on $[a,b]$ in some sense with indefinite integrals F_n. If $f_n\to f$ uniformly then F_n converges uniformly to some continuous function $F:[a,b]\to\mathbb{R}$. (Check this.)

Let x_0 be a point at which $F_n'(x_0)=f_n(x_0)$ for each n. Let $\varepsilon>0$ and choose an integer N so that

$$|f_n(x)-f(x)|<\varepsilon$$

for all $x\in[a,b]$ and all $n\ge N$. Observe that f is integrable at least in the sense of the Henstock-Kurzweil integral. (Any uniform limit of integrable functions would be integrable.) Thus for any x, y in the interval

$$|F(y)-F(x)-(F_n(y)-F_n(x))| \le \int_x^y |f(t)-f_n(t)|\,dt < \varepsilon(y-x).$$

Use this to deduce that $F'(x_0)=f(x_0)$.

Thus the classical Newton has this property. that a uniform limit of a sequence of classically Newton integrable functions is also classically Newton integrable. The naive Newton has this property. that a uniform limit of naively

Newton integrable functions is also naively Newton integrable. The same can be verified for the utility and the general versions.

The elementary version, however, employs a finite set of exceptions. At each stage of the sequence above $F_n'(x) = f_n(x)$ for all x except for x in some finite set E_n. We can verify that $F'(x) = f(x)$ for all x except for x in the set $E = \bigcup_{n=1}^{\infty} E_n$. But that set E may not be finite itself.

For example, simply let $f_n(x) = 0$ for all x excepting that $f_n(1/i) = 1/i$ for each point $x = 1/i$ with $i = 1, 2, 3, \dots, n$. Then f_n converges uniformly to a function which is not integrable in the elementary Newton sense.

Exercise 226, page 185

The following are trivial:

$$\text{[unif]} \Rightarrow \text{[meas]}$$

and

$$\text{[unif]} \Rightarrow \text{[a.u.]}.$$

The following are easy:

$$\text{[unif]} \Rightarrow \text{[mean]},$$

$$\text{[a.u.]} \Rightarrow \text{[a.e.]},$$

and

$$\text{[a.u.]} \Rightarrow \text{[meas]}.$$

The only difficult one is

$$\text{[a.e.]} \Rightarrow \text{[a.u.]}$$

whose proof we supply in Section 4.10.6 as Egorov's theorem.

Exercise 227, page 186

This is just an observational question. Look at the displays. We need counterexamples to justify our views that these implication are invalid:

$$\text{[a.e.]} \not\Rightarrow \text{[unif]} \quad \text{and} \quad \text{[a.e.]} \not\Rightarrow \text{[mean]},$$

$$\text{[a.u.]} \not\Rightarrow \text{[unif]} \quad \text{and} \quad \text{[a.u.]} \not\Rightarrow \text{[mean]},$$

$$\text{[meas]} \not\Rightarrow \text{ any of the others.}$$

Note, however, that we do not need counterexamples for all possible failed implications. For example, a counterexample that shows that

$$\text{[meas]} \not\Rightarrow \text{[a.u.]}$$

would automatically allow us to conclude that

$$\text{[meas]} \not\Rightarrow \text{[unif]} \quad \text{and} \quad \text{[meas]} \not\Rightarrow \text{[a.e.]},$$

Exercise 228, page 186

The conclusion follows from the easy inequality

$$\lambda(\{x \in [a,b] : |f_n(x) - f(x)| \geq \eta\}) \leq \eta^{-1} \int_a^b |f_n(t) - f(t)|\, dt.$$

Exercise 229, page 188

First of all select a subsequence $\{f_{n_k}\}$ so that

$$\liminf_{n\to\infty} \int_a^b f_n(x)\, dx = \lim_{k\to\infty} \int_a^b f_{n_k}(x)\, dx.$$

This subsequence $\{f_{n_k}\}$ also converges in measure to f. By Exercise 4.48 there is a further subsequence (i.e. a subsequence of the sequence $\{f_{n_k}\}$) that converges almost everywhere on $[a,b]$. Now just apply the original version of Fatou's lemma to this subsubsequence.

Exercise 231, page 189

If $(-\infty,\infty)\setminus N$ contains even a single sequence of points $\{x_n\}$ with $|x_n| \to \infty$ then $\{f_n\}$ cannot converge uniformly on $(-\infty,\infty)\setminus N$.

Exercise 232, page 190

The new implications, available under the assumption that the sequence is dominated, are:

$$\text{[a.e.]} \Rightarrow \text{ [mean]},$$

$$\text{[meas]} \Rightarrow \text{ [mean]},$$

and

$$\text{[a.u.]} \Rightarrow \text{ [mean]}.$$

But a moment's thought shows that there is only one that needs to be proved:

$$\text{[meas]} \Rightarrow \text{ [mean]}.$$

The other two implications would follow from this.

Exercise 233, page 190

Compare with Exercise 4.35. Make sure to consult also Exercises 4.48 and 229 for some techniques useful in handling convergence in measure.

Exercise 235, page 194

It was A. Denjoy [Bull. Soc. Math. France 43 (1915), 161–248] who first introduced the notion of approximately continuous functions of a real variable. In that paper he proved that they belong to Baire class 1 and have the Darboux property (i.e., intermediate value property). More recent and simpler proofs can be found in several texts (e.g., [11], [83] and in the article of Chen [1].

Exercise 236, page 194

Construct a covering argument that is identical to that used in the proof of Theorem 95. Use the approximate version of the Cousin lemma (i.e., Lemma 4.8). See also the references in the answer for Exercise 235.

Exercise 237, page 195

Simply show that, if this is not the case, then the indefinite integral of f would have bounded variation on $[a,b]$. For a nonabsolutely integrable function, this cannot be the case.

Exercise 238, page 196

It is easy to check that [uniform convergence] does not imply [bounded convergence]. We show that [uniform convergence] does imply [dominated convergence]. Suppose that $f_n \to f$ uniformly on $[a,b]$ where each f_n is Lebesgue integrable. Choose an integer N so that $|f_n(x) - f(x)| < 1$ for all $x \in [a,b]$ and all $n \geq N$. Set, for each $x \in [a,b]$,

$$g(x) = \max\{|f_1(x)|, |f_2(x)|, |f_3(x)|, \dots |f_{N-1}(x)|, |f_N(x)| + 1\}.$$

The g is Lebesgue integrable on $[a,b]$ and g dominates the sequence $\{f_n\}$. Thus [uniform convergence] implies [dominated convergence] on a finite internal $[a,b]$. (This would not be the case on an infinite interval.)

Exercise 239, page 197

This is trivial on on a finite internal $[a,b]$. (This would not be the case on an infinite interval.)

[1] Jau D. Chen, A note on approximate continuity, Tamkang J. Math. 5 (1974), no. 1, 109–111.

Exercise 240, page 197

We assume that, for some Lebesgue integrable function g, $|f_n(x)| \leq g(x)$ for all integers n and a.e. points $x \in [a,b]$. Let $\varepsilon > 0$ and choose $\delta > 0$ so that if G is an open set for which $\lambda(G) < \delta$ then

$$\int_a^b g(x)\chi_G(x)\,dx < \varepsilon.$$

for all integers n. [This uses Theorem 4.55.) Then

$$\int_a^b |f_n(x)|\chi_G(x)\,dx \leq \int_a^b g(x)\chi_G(x)\,dx < \varepsilon$$

for all integers n.

Exercise 241, page 197

Arrange such a sequence $f_n : [0,1] \to \mathbb{R}$ converging pointwise to zero for which

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = 0$$

and for which

$$\sup_n \{f_n(x)\} = 1/x$$

for all $0 < x < 1$.

Exercise 242, page 197

If the sequence $\{g_n\}$ satisfies the Vitali equi-integrability condition, then Theorem 4.56 supplies the limit

$$\lim_{n\to\infty} \int_a^b g_n(x)\,dx = 0.$$

In the other direction suppose that this limit is valid. Let $\varepsilon > 0$. Choose an integer N so that

$$\int_0^1 g_n(x)\,dx < \varepsilon$$

if $n \geq N$. For each integer $i = 1, 2, \ldots, N-1$, choose $\delta_i > 0$ so that if G is an open set for which $\lambda(G) < \delta_i$ then

$$\int_a^b |g_i(x)|\chi_G(x)\,dx < \varepsilon.$$

Set

$$\delta = \min\{\delta_1, \delta_2, \delta_3, \ldots, \delta_{N-1}\}.$$

Then, if G is an open set for which $\lambda(G) < \delta$,

$$\int_a^b |g_n(x)|\chi_G(x)\,dx < \varepsilon$$

for all integers n. Thus the sequence $\{g_n\}$ satisfies the Vitali equi-integrability condition.

Exercise 243, page 201

This would be an essential exercise for beginning students. Although an interpretation of the integral as area makes this identity quite obvious, the connection with differentiation serves as a good reminder where this theory originates.

Exercise 251, page 211

Let E be a measurable subset of $[0,1]$. Then the function $f(x) = \chi_E(x)$ is a nonnegative integrable function. Does $f = f^+$ belong to $\mathcal{S}_\ell^\uparrow[0,1]$? That would mean that there is a function $g \in \mathcal{S}_\ell^\uparrow[0,1]$ with $f = g$ a.e. in that interval. Since g is l.s.c. the set

$$G = \{x \in (0,1) : g(x) > 1/2\}$$

is an open set and a.e. point of G belongs to E. This is not the case, however, for every measurable set. (For example take a Cantor set of positive measure.)

Exercise 252, page 211

An example is given in James Foran and Sandra Meinershagen [29], Some answers to a question of P. Bullen, Real Anal. Exchange 13 (1987/88), no. 1, 265–277.

Exercise 258, page 215

Define $F : \mathbb{R} \to \mathbb{R}$ by $F(0) = F(1/(2n-1)) = 0$ and $F(1/2n) = 1/n$ for all $n = 1,2,3,\ldots$. Extend F to be linear on each of the intervals contiguous to these points where it has so far been defined. Show that F is absolutely continuous but that Vitali's condition does not hold on the interval $[0,1]$.

Exercise 259, page 215

Perhaps hard to spot. Note that the condition does not specify that the intervals should be nonoverlapping. Show that every Lipschitz function satisfies this condition. Then show that a function that satisfies the condition of *must* be a Lipschitz function.

Exercise 261, page 217

The only property of ultrafilters that is needed for this proof is the following fundamental fact that the student can look up elsewhere: if $\mathcal{U}$ is an ultrafilter on Π and α is any nonempty subset of Π then either $\alpha \in \mathcal{U}$ or else $\Pi \setminus \alpha \in \mathcal{U}$.

To prove Theorem 4.71 note that the only part that is not completely immediate is the identity

$$(\mathcal{U})\underline{\int_a^b} f(x)\,dx = (\mathcal{U})\overline{\int_a^b} f(x)\,dx$$

that is required to hold for all functions $f:[a,b]\to\mathbb{R}$. Let t be any real number for which

$$t < (\mathcal{U})\overline{\int_a^b} f(x)\,dx.$$

Define the following subsets of Π:

$$\alpha = \left\{\pi\in\Pi : \sum_{([u,v],w)\in\pi} f(w)(v-u) > t\right\}$$

and its complement

$$\alpha' = \left\{\pi\in\Pi : \sum_{([u,v],w)\in\pi} f(w)(v-u) \leq t\right\}.$$

By the fundamental property of ultrafilters (just mentioned) one of these two sets must belong to $\mathcal{U}$. But if $\alpha' \in \mathcal{U}$ then that would imply that

$$(\mathcal{U})\overline{\int_a^b} f(x)\,dx \leq t$$

which is impossible. It follows that $\alpha \in \mathcal{U}$ and, hence, that

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

Since this is true for all t for which

$$t < (\mathcal{U})\overline{\int_a^b} f(x)\,dx$$

the identity

$$(\mathcal{U})\underline{\int_a^b} f(x)\,dx = (\mathcal{U})\overline{\int_a^b} f(x)\,dx$$

must hold.

Exercise 262, page 218

For any covering relation β it is clear that

$$\mathrm{Var}(r\lambda,\beta[E]) \leq \mathrm{Var}(f\lambda,\beta) \leq \mathrm{Var}(s\lambda,\beta)$$

and from this one can deduce that

$$r\lambda(E) = V^*(r\lambda, E) \le V^*(f\lambda, E) \le V^*(s\lambda, E) = s\lambda(E).$$

Note that it would also be true that

$$r\lambda(E) = V_*(r\lambda, E) \le V_*(f\lambda, E) \le V^*(s\lambda, E) = s\lambda(E).$$

Exercise 266, page 219

This exercise asserts that when $\lambda(E)$ is zero so too is $\int_E f(x)\,dx$. This is considered an absolute continuity condition. In Exercise 274 we considered a different version of absolute continuity asserting that if $\lambda(E)$ is "small" so too is $\int_E f(x)\,dx$.

Let $E_n = \{x \in E : 1/n < f(x)\}$. Check that

$$n\lambda(E_n) \le \int_{E_n} f(x)\,dx \le \int_E f(x)\,dx = 0.$$

Thus $\lambda(E_n) = 0$ for each n and so also if $E' = \{x \in E : f(x) \ne 0\}$ then

$$\lambda(E') \le \sum_{n=1}^{\infty} \lambda(E_n) = 0.$$

Exercise 273, page 220

For illustrative purposes only we begin the proof with the bounded case. Suppose that $f(x) < N$ for all $x \in E$. Choose $\delta = \varepsilon/N$ and observe that, if $\lambda(G) < \delta$ then the inequalities in the measure estimates of an earlier exercise in this section provide

$$\int_{E\cap G} f(x)\,dx \le N\lambda(G) < \varepsilon.$$

Thus the proof in the bounded case is trivial and does not require that f be measurable.

Exercise 274, page 220

This exercise asserts that when $\lambda(E)$ is small so too is $\int_E f(x)\,dx$. This is considered an absolute continuity condition. In Exercise 267 we consider a different version of absolute continuity asserting that if $\lambda(E)$ is zero so too is $\int_E f(x)\,dx$. Note that there are finiteness assumptions in this (stronger) version.

The argument in the preceding exercise suggests how to proceed. Let

$$A_n = \{x : n-1 \le f(x) < n\}.$$

From the fact that f is measurable we can deduce that A_n is measurable. Thus we can select an open set G_n for which $B_n = A_n \setminus G_n$ is closed and $\lambda(G_n) <$

$\varepsilon 2^{-n} n^{-1}$. That also requires

$$\int_{E\cap A_n} f(x)\,dx \leq \int_{E\cap B_n} f(x)\,dx + \int_{E\cap A_n\cap G_n} f(x)\,dx$$
$$\leq \int_{E\cap B_n} f(x)\,dx + \varepsilon 2^{-n}.$$

Note that $\{B_n\}$ is a disjointed sequence of closed sets whose union B can be handled by the usual additive properties of measures over such sets. Thus

$$\sum_{n=1}^{\infty} \int_{E\cap A_n} f(x)\,dx \leq \sum_{n=1}^{\infty} \int_{E\cap B_n} f(x)\,dx + \varepsilon$$
$$= \int_{E\cap B} f(x)\,dx + \varepsilon \leq \int_E f(x)\,dx + \varepsilon < \infty.$$

In particular there must be an integer N sufficiently large that

$$\sum_{n=N+1}^{\infty} \int_{E\cap A_n} f(x)\,dx < \varepsilon/2.$$

Choose $\delta = \varepsilon/(2N)$ and let G be any open set for which $\lambda(G) < \delta$. Since

$$E\cap G = \{x \in E\cap G : f(x) < N\} \cup \bigcup_{n=N+1}^{\infty} (G\cap E\cap A_n)$$

we have

$$\int_{E\cap G} f(x)\,dx \leq N\lambda(G) + \sum_{n=N+1}^{\infty} \int_{E\cap A_n} f(x)\,dx < \varepsilon.$$

Exercise 275, page 220

The proof repeats a number of techniques we have already seen in the proof of Theorem 274.

Each of the sets appearing in the statement of the theorem is measurable, because f is measurable. Select open sets G_{kr} so that $B_{rk} = A_{kr} \setminus G_{kr}$ is closed and so that

$$\lambda(G_{kr}) < \varepsilon 2^{-|k|-1} r^{-k}.$$

That also requires

$$\int_{E\cap A_{kr}} f(x)\,dx \leq \int_{E\cap B_{kr}} f(x)\,dx + \int_{E\cap A_{kr}\cap G_{kr}} f(x)\,dx$$
$$\leq \int_{E\cap B_{kr}} f(x)\,dx + \varepsilon 2^{-|k|-1}.$$

Note that $\{B_{kr}\}$ is a disjointed sequence of closed sets whose union B_r can be handled by the usual additive properties of measures over such sets.

Now we compute:

$$\int_E f(x)\,dx \le \sum_{k=-\infty}^{\infty} \int_{E\cap A_{kr}} f(x)\,dx$$

$$\le \sum_{k=-\infty}^{\infty} \int_{E\cap B_{kr}} f(x)\,dx + \varepsilon \le \sum_{k=-\infty}^{\infty} r^k \lambda(E\cap B_{kr}) + \varepsilon$$

$$\le r \sum_{k=-\infty}^{\infty} \int_{E\cap B_{kr}} f(x)\,dx + \varepsilon = r \int_{E\cap B_r} f(x)\,dx + \varepsilon \le r \int_E f(x)\,dx + \varepsilon.$$

Exercise 276, page 220

The first identity

$$\int_E f(x)\,dx = V^*(f\lambda, E)$$

is just our definition. Thus the intent of the exercise is to prove just that

$$V^*(f\lambda, E) = V_*(f\lambda, E).$$

Repeat Exercise 275 and this time deduce the related inequality

$$V_*(f\lambda, E) \le \sum_{k=-\infty}^{\infty} r^k \lambda(E\cap A_{kr}) \le rV_*(f\lambda, E).$$

Essentially this is accomplished because the Lebesgue measure can be estimated by either full covers or by fine covers (this is the Vitali covering theorem).

A comparison of the two inequalities shows that $V^*(f\lambda, E) = V_*(f\lambda, E)$.

Exercise 277, page 222

If one were to follow the outline in detail, then the second step requires careful justification. It is not clear, even after some reflection, that the value of the integral does not depend on the particular representation of f as a sum $\sum_{i=1}^{\infty} c_i \chi(E_i)$. Specifically one must prove the following statement:

Suppose that

$$f(x) = \sum_{i=1}^{\infty} c_i \chi(E_i)$$

for numbers $\{c_1, c_2, c_3, \dots\}$ *in* $\mathbb{R}^+ \cup \{+\infty\}$ *and sets* $\{E_1, E_2, E_3, \dots\}$ *belonging to* $\mathcal{M}$*. Suppose also that*

$$f(x) = \sum_{j=1}^{\infty} d_j \chi(A_j)$$

for numbers $\{d_1, d_2, d_3, \dots\}$ *in* $\mathbb{R}^+ \cup \{+\infty\}$ *and sets*

$\{A_1, A_2, A_3, \dots\}$ *belonging to $\mathcal{M}$. Then*

$$\sum_{i=1}^{\infty} c_i \mu(E \cap E_i) = \sum_{j=1}^{\infty} d_i \mu(E \cap A_i).$$

A better and more conventional approach would be to define the concepts in a manner that allows the clearest development and then verify later that these statements (for elementary and measurable functions) are correct. If, instead, we were to take the statements given as definitions, then the justification added in italics above must be made.

Exercise 278, page 223

This exercise and the two that follow it are deceptively easy. The hard work is justifying that there is no ambiguity in defining the integral of an elementary function. Once that is established then these properties flow from it. Here suppose that

$$f(x) = \sum_{i=1}^{\infty} a_i \chi_{A_i}(x)$$

and

$$g(x) = \sum_{i=1}^{\infty} b_i \chi B_i(x)$$

for numbers $\{a_1, b_1, a_2, b_2, \dots\}$ in $\mathbb{R}^+ \cup \{+\infty\}$ and sets $\{A_1, B_1, A_2, B_2, \dots\}$ belonging to $\mathcal{M}$. Then the function $F(x) = f(x) + g(x)$ has a representation as the sum

$$a_1 \chi_{A_1}(x) + b_1 \chi_{B_1}(x) + a_2 \chi_{A_2}(x) + b_2 \chi_{B_2}(x) + a_3 \chi_{A_3}(x) + \dots$$

and so (by definition)

$$\int_E [f(x) + g(x)]\, d\mu(x) = \int_E F(x)\, d\mu(x) =$$

$$a_1 \mu(A_1 \cap E) + b_1 \mu(B_1 \cap E) + a_2 \mu(A_2 \cap E) + \mu \chi(B_2 \cap E) + a_3 \mu(A_3 \cap E) + \dots$$

and this sum is evidently the same as

$$\int_E f(x)\, d\mu(x) + \int_E g(x)\, d\mu(x).$$

Thus the right perspective on this integration theory is not really that these properties (additivity and monotone convergence properties) are derivable from the definitions as that these properties have been designed to hold. The definition itself contains these properties as simple consequences.

Exercise 280, page 223

This exercise and the two that precede it are deceptively easy. The hard work is justifying that there is no ambiguity in defining the integral of an elementary function. Once that is established then these properties flow from it. Here suppose that

$$g_i(x) = \sum_{k=1}^{\infty} a_{ik}\chi_{A_{ik}}(x)$$

and

$$f(x) = \sum_{i=1}^{\infty} g_i(x)$$

for numbers $\{a_{ik}\}$ in $\mathbb{R}^+ \cup \{+\infty\}$ and sets $\{A_{ik}\}$ belonging to $\mathcal{M}$. Then the function $f(x)$ has a representation as the sum

$$f(x) = \sum_{i=1}^{\infty}\sum_{k=1}^{\infty} a_{ik}\chi_{A_{ik}}(x).$$

While this is formulated as a double sum it can be easily visualized as a sum of a single sequence so that f is a nonnegative elementary function. Thus by definition

$$\int_E f(x)\,d\mu(x) = \sum_{i=1}^{\infty}\sum_{k=1}^{\infty} a_{ik}\mu(E \cap A_{ik})$$

and this sum is evidently the same as

$$\sum_{i=1}^{\infty} \int_{E_i} g_i(x)\,d\mu(x).$$

Again the right perspective on this integration theory is not really that the monotone convergence property is derivable from the definitions as that this property holds by design. The definition itself contains these properties as simple consequences.

Exercise 281, page 227

This needs the Cousin lemma so that for any given full cover β there exists at least one subset $\pi \subset \beta$ that is a partition of $[a,b]$

Exercise 285, page 229

Take as a full cover β the collection of pairs $([u,v],w)$ for which $w \in [u,v]$ but $[u,v]$ never overlaps both of the intervals $[0,1/2]$ or $[1/2,1]$ unless $w = 1/2$. Then all partitions π of $[a,b]$ from β can be split neatly at the point $1/2$.

Exercise 286, page 229

Take as a full cover β the collection of pairs $([u,v],w)$ for which $w \in [u,v]$ but $[u,v]$ never overlaps two of the intervals $[\xi_{i-1},\xi_i]$ unless w is one of the points $\{\xi_i\}$. Then all partitions π of $[a,b]$ from β can be split neatly at the points ξ_i.

Exercise 287, page 229

Both integrals exist but have different values, which you can check. If you were schooled in the Riemann-Stieltjes integral then you might recall this example was used to illustrate *non-existence* of the Riemann-Stieltjes integral. These differences in the two theories are mostly irrelevant since most applications will assume that one function is continuous and the other has bounded variation.

Exercise 288, page 229

Warning: If you were schooled in the Riemann-Stieltjes integral before learning *this* Stieltjes integral you may think not. Otherwise just check that the existence of the integral (finitely that is) on $[a,b]$ and $[b,c]$ is equivalent to the existence of the integral on $[a,c]$.

Exercise 291, page 229

Hint: $|dG(x)|$ is subadditive whereas $dG(x)$ is additive.

Exercise 300, page 238

Prove just for monotonic saltus functions. Assume that sequences $\{x_n\}$, $\{u_n\}$, and $\{v_n\}$ are given with $u_n \geq 0$, $v_n \geq 0$ and the series $\sum_{n=1}^{\infty} u_n$ and $\sum_{n=1}^{\infty} v_n$ both assumed to be convergent. Define the corresponding monotonic jump functions $f_n : \mathbb{R} \to \mathbb{R}$ so that $f_n(t) = 0$ for $t < x_n$, $f_n(x_n) = u_n$ and $f_n(t) = u_n + v_n$ for $t > x_n$. Clearly $f_n'(t) = 0$ at every point except possibly at $t = x_n$. (If either u_n or v_n is not zero then, certainly, $f_n'(x_n)$ cannot exist.) The derivative of the saltus function

$$s(t) = \sum_{n=1}^{\infty} f_n(t)$$

can now be handled by the Fubini differentiation theorem (Theorem 2.42). We must have

$$s'(t) = \sum_{n=1}^{\infty} f_n'(t) = 0$$

at almost every point t.

Exercise 314, page 248

We can simplify the argument and assume that F is defined on the whole real line. We wish to show that

1. $\lambda_F(\emptyset) = 0$.

2. For any sequence of sets E, E_1, E_2, E_3, ... for which $E \subset \bigcup_{n=1}^{\infty} E_n$ the inequality
$$\lambda_F(E) \leq \sum_{n=1}^{\infty} \lambda_F(E_n)$$
must hold.

This result is often described by the following language that splits the property (2) in two parts:

Subadditivity: $\lambda_F\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \lambda_F(E_n).$

Monotonicity: $\lambda_F(A) \leq \lambda_F(B)$ if $A \subset B$.

The monotonicity is obvious. This allows us to prove the additivity assertion above just in the special case that the sets $\{E_n\}$ are pairwise disjoint so that $E = \bigcup_{n=1}^{\infty} E_n$ is now a disjoint union. If $\lambda_F(E_n) = \infty$ for any integer n there is nothing to prove so we may suppose all of these are finite.

Let $\varepsilon > 0$. For each integer n choose a full cover β_n of E_n so that
$$\sup_{\pi \subset \beta_n} \sum_{([u,v],w)\in\pi} |F(v) - F(u)| < \lambda_F(E_n) + \varepsilon 2^{-n}.$$
Then write
$$\beta = \bigcup_{n=1}^{\infty} \beta_n[E_n].$$
This is a full cover of E and consequently
$$\lambda_F(E) \leq \sup_{\pi \subset \beta} \sum_{([u,v],w)\in\pi} |F(v) - F(u)|.$$
Take any subpartition $\pi \subset \beta$ and observe that
$$\sum_{([u,v],w)\in\pi} |F(v) - F(u)| \leq \sum_{n=1}^{\infty} \sum_{([u,v],w)\in\pi[E_n]} |F(v) - F(u)| \leq \sum_{n=1}^{\infty} \left[\lambda_F(E_n) + \varepsilon 2^{-n}\right].$$
From this it follows that
$$\lambda_F(E) \leq \varepsilon + \sum_{n=1}^{\infty} [\lambda_F(E_n)]$$
and the subadditive property follows.

Exercise 318, page 249

One direction is easy. If β is full cover of $(a,b))$ and we prune out intervals not inside (a,b) by writing $\beta' = \beta((a,b))$, then it is clear that
$$\lambda_F((a,b)) \leq \sup_{\pi \subset \beta'} \sum_{([u,v],w)\in\pi} |F(v) - F(u)| \leq \mathrm{Var}(F, [a,b]).$$
In the other direction ETC

Exercise 319, page 249

Suppose that $E \subset (-L, L)$ and that $|F'(x)| < M$ for all $x \in E$. Then

$$\beta = \left\{ ([u,v],w) : (u,v) \subset (-L,L), \ \ \frac{|F(v)-F(u)|}{v-u} < M \right\}$$

is a full cover of E Thus, for any subpartition $\pi \subset \beta$,

$$\sum_{([u,v],w)\in\pi} |F(v)-F(u)| \leq 2ML.$$

It follows that $\lambda_F(E) \leq 2ML$ and so is finite.

Exercise 321, page 254

Develop the Henstock zero variation criterion for this integral and check that the usual zero derivative procedure will supply this fact.

Exercise 325, page 269

Note that there is no typo in the first inequality: the full variation is needed on the right-hand side. The second inequality, the easier to check, follows from the fact that the intersection of two full covers is again full. The first inequality follows from the fact that the intersection of two covering relations, one of which is full and the other fine, is again fine.

Exercise 330, page 270

If $f'(x) = 0$ for all $x \in E \setminus N$ show that $\lambda_f(E \setminus N) = 0$.

Exercise 332, page 270

It is enough to suppose that $\lambda_f(E) < \infty$. Let $C = \{x \in E : \lambda_f(\{x\}) > 0\}$; this set must include every point at which f fails to be continuous. Now let $C_n = \{x \in E : \lambda_f(\{x\}) > 1/n\}$. From measure-properties

$$\lambda_f(C_n) \leq \lambda_f(E) < \infty.$$

But if C_m contains k points then

$$\lambda_f(C_n) \geq k/n.$$

It follows that each set C_n is finite and hence the set C must be countable.

Exercise 333, page 270

Some authors would use the term weakly continuous at a point x_0 to mean that there is at least one sequence $c_n \to x_0$ and so that $|c_n - x_0| > 0$ and

$$f(c_n) - f(x_0) \to 0.$$

This condition is a little stronger than the definition in the text. For example the function $f(x) = 0$ if $x \neq 0$ and $f(0) = 1$ is weakly continuous at 0 in our sense but not in the stronger sense. The property in the exercise is dictated by the particular definition that we use for fine covers.

Here is a proof. Since f is weakly continuous at x_0 we know, by definition, that $\lambda_f^\star(\{x_0\}) = 0$. For each integer n we can select a fine cover β_n of the set $\{x_0\}$ so that $\mathrm{Var}(\Delta f, \beta_n) < 1/n$. From β_n we can select a pair $([c_n, d_n], x_0)$ for which $d_n - c_n < 1/n$. Note that $c_n \leq x_0 \leq d_n$ and

$$|f(d_n) - f(c_n)| \leq \mathrm{Var}(\Delta f, \beta_n) < 1/n.$$

This pair of sequences $\{c_n\}$ and $\{d_n\}$ has all the properties that we need except they need not be monotonic. But there is a monotonic subsequence of the $\{c_n\}$ so that we can consider that we have selected that subsequence. Take a further subsequence so that both sequences are monotonic. The new sequences have all the properties that we need.

Exercise 334, page 270

Let

$$E = \{x : \liminf_{(I,x) \Longrightarrow x} |\Delta f(I)| > 0\}$$

and

$$E = \{x : \liminf_{(I,x) \Longrightarrow x} |\Delta f(I)| > 1/n\}.$$

The set of points where f is not weakly continuous is exactly the set $E = \bigcup_n E_n$. Note that $\beta = \{(I,x) : |\Delta f(I)| > 1/n\}$ is a full cover of E_n and apply the decomposition lemma from Section 2.2.3.

Exercise 336, page 270

Recall that F has finite variation on (a,b) if there is a number M and a full cover β of (a,b) so that

$$\sum_{([u,v],w)\in\pi} |F(v) - F(u)| \leq M$$

whenever π is a subpartition, $\pi \subset \beta$. If F has bounded variation on $[a,b]$ then certainly $M = \mathrm{Var}(F,[a,b])$ will work.

In the converse direction we suppose that M and β have been chosen with this property. For every subinterval $[c,d] \subset [a,b]$ there is a partition π contained in β for which evidently

$$|F(d)-F(c)| = \left|\sum_{([u,v],w)\in\pi}[F(v)-F(u)]\right| \leq \sum_{([u,v],w)\in\pi}|F(v)-F(u)| \leq M.$$

Fix some point x_0 in (a,b) and then we have the bound $|F(x)| \leq M+|F(x_0)|$ for every point x in (a,b).

Now we estimate $\mathrm{Var}(F,[a,b])$. Take any choice of points

$$a = s_0 < s_1 < \cdots < s_{n-1} < s_n = b.$$

We note that

$$|F(s_1)-F(s_0)| \leq |F(a)|+M+|F(x_0)|$$

and that

$$|F(s_n)-F(s_{n-1})| \leq |F(b)|+M+|F(x_0)|.$$

We may choose a partition π from β so that π contains a partition of each of the remaining intervals $[s_1,s_2]$, $[s_2,s_3]$, ..., $[s_{n-2},s_{n-1}]$. This provides the inequality

$$\sum_{i=1}^{n}|F(s_i)-F(s_{i-1})| \leq |F(a)|+M+|F(x_0)|+|F(b)|+M+|F(x_0)|+M =$$

$$|F(a)|+|F(b)|+3M+2|F(x_0)|.$$

This offers us an upper bound for $\mathrm{Var}(F,[a,b])$ and we have proved that F has bounded variation on $[a,b]$.

Exercise 338, page 271

If f is locally recurrent at every point of a set E then

$$\beta = \{(I,x) : \Delta f(I) = 0\}$$

is a fine cover of E. Thus

$$\lambda_f^{\star}(E) \leq \mathrm{Var}(\Delta f,\beta) = 0.$$

Exercise 339, page 271

Define

$$\beta = \{(I,x) : \Delta f(I) \geq 0\}$$

and notice that this is a full cover of E. Apply the decomposition from Section 2.2.3 for β. There is an increasing sequence of sets $\{E_n\}$ with $E = \bigcup_{n=1}^{\infty} E_n$ and a sequence of compact intervals $\{I_{kn}\}$ covering E so that if x is any point in E_n and I is any subinterval of I_{kn} that contains x then (I,x) belongs to β.

We check that f is nondecreasing on each set $D_{nk} = E_n \cap I_{kn}$ in a certain strong way. For if either x or y belongs to the set D_{nk} and $[x,y] \subset I_{kn}$ then one of the pairs $([x,y],x)$ or $([x,y],y)$ belongs to β which requires that $f(x) \leq f(y)$.

Let $c = \inf D_{nk}$ and $d = \sup D_{nk}$. Suppose that $c = d$. Then D_{nk} contains a single point c and $\lambda_f(\{c\}) < \infty$, i.e., $\lambda_f(D_{nk}) < \infty$. Suppose instead that $c < d$. Let $D'_{nk} = D_{nk} \cap (c,d)$ so that D_{nk} contains, at most, two points c and d more than the set D'_{nk}. Let $\beta' = \beta[D_{nk}] \cap \beta((c,d))$. Then β' is a full cover of D'_{nk}. Let $\pi = \{\{[c_i,d_i],x_i)\}$ be any subpartition contained in β'. We see from the manner in which f increases relative to the set D_{nk} that

$$\sum_i |f(d_i) - f(c_i)| \leq 2[f(d) - f(c)].$$

It follows that

$$\lambda_f(D'_{nk}) \leq \mathrm{Var}(\Delta f, \beta') \leq 2[f(d) - f(c)] < \infty.$$

Consequently,

$$\lambda_f(D_{nk}) \leq \lambda_f(D'_{nk}) + \lambda_f(\{c\}) + \lambda_f(\{d\}) < \infty$$

too, so that in either case $\lambda_f(D_{nk})$ is finite. It follows that λ_f is σ-finite on the set E since that set has been expressed as a union of a sequence of sets on each of which λ_f is σ-finite.

Exercise 340, page 271

Define three sets E_1, E_2, and E_3. E_1 is the set of points at which f is locally nondecreasing. E_2 is the set of points at which $-f$ is locally nondecreasing. E_3 is the set of points at which f is locally recurrent. Since f is continuous it has the Darboux property. From that we see that $E_1 \cup E_2 \cup E_3 = \mathbb{R}$ since there are no other possibilities.

But $\lambda^\star_f(E_3) = 0$ and λ_f is σ-finite on E_1 and E_2 (Exercise 339). It follows that the smaller measure $\lambda^\star_f$ must be σ-finite.

Exercise 341, page 271

Since the hint suggests that we can use Theorem 6.20 let us do so. There must be a sequence of compact sets $\{E_n\}$ covering E and a sequence of continuous functions of bounded variation $\{g_n\}$ so that f is Kolmogorov equivalent to g_n on E_n. In particular, we know that $g'_n(x)$ exists at almost every point. Therefore the set of points in E_n at which $f'(x) = g'_n(x)$ fails is a set of measure zero, say N_n. It follows that f is differentiable at every point of E with the possible exception of points in the measure zero set $\bigcup_{n=1}^\infty N_n$.

Exercise 344, page 276

Let

$$E = \{x : D^- f(x) < D_+ f(x)\}$$

and, for each rational number r, let

$$E_r = \{x : D^- f(x) < r < D_+ f(x)\}.$$

Note that E is the union of the countable collection of sets E_r taken over all rationals r. For each x in E_r there is a $\delta(x) > 0$ so that, for all $0 < h < \delta(x)$,

$$\Delta f([x-h,x]) < r\lambda([x-h,x])$$

and

$$\Delta f([x,x+h) > r\lambda([x,x+h]))\}$$

because of the values of the Dini derivatives.

Let

$$E_{rn} = \{x \in E : \delta(x) > 1/n\}$$

and check that

$$E_r = \bigcup_{n=1}^{\infty} E_{rn}.$$

We claim that, for each n, the set E_{rn} is countable. Indeed there cannot be two points x and y with $x < y$ in E_{rn} closer together than $1/n$. For if so, let $h = y - x$, note that $0 < h < \delta(x) < 1/n$ and $0 < h < \delta(y) < 1/n$. That would mean that

$$\Delta f([x,y]) < r\lambda([x,y]) < \Delta f([x,y)$$

which is impossible. Accordingly each E_{rn} is countable and so too also is E. The other set of the theorem can be handled by an identical proof.

Exercise 346, page 277

Consider first the set

$$A = \{x : D^- f(x) < D^+ f(x)\}$$

and, for each rational number r, let

$$A_r = \{x : D^- f(x) < r < D^+ f(x)\}.$$

Note that A is the union of the countable collection of sets A_r taken over all rational numbers r.

For each x in A_r we have $D^- f(x) < r$. Thus there is a $\delta(x) > 0$ so that, for all $0 < h < \delta(x)$,

$$f(x) - f(x-h) < rh.$$

For each $n = 1, 2, 3, \dots$ and each $k = 0, \pm 1, \pm 2, \dots$ write

$$A_{rnk} = \left[\frac{k-1}{n}, \frac{k}{n}\right] \cap \left\{x \in A_r : \delta(x) > \frac{1}{n}\right\}.$$

Notice that

$$f(x) - f(y) < r(x-y)$$

for all $x < y$ with $x,\, y \in A_{rnk}$ and check that

$$A_r = \bigcup_{k=-\infty}^{\infty} \bigcup_{n=1}^{\infty} A_{rnk}.$$

Finally let E_{rnk} denote the closure of the set A_{rnk}. Each set E_{rnk} is compact and we claim that it contains no subinterval; in particular then it is a meager subset of $\mathbb{R}$.

Should such a set E_{rnk} contain an interval $[a,b]$ then, by the continuity of f we must conclude that the inequality stated above would require, for all $a < y < x < b$, that

$$f(x) - f(y) \leq r(x-y).$$

Consequently there would be no points y in (a,b) at which $r < D^+ f(y)$. But this is impossible since the set A_{rnk} is dense in the set E_{rnk}.

Thus we have displayed

$$A_r \subset \bigcup_{k=-\infty}^{\infty} \bigcup_{n=1}^{\infty} E_{rnk}$$

as a subset of a union of a sequence of meager subsets of $\mathbb{R}$.

It follows that the set A defined above is also a meager subset of $\mathbb{R}$. In a similar way we can conclude that each of the sets

$$\{x : D^- f(x) > D^+ f(x)\}$$
$$\{x : D_- f(x) > D_+ f(x)\}$$

and

$$\{x : D_- f(x) < D_+ f(x)\}$$

is a meager subset of $\mathbb{R}$. From this the theorem follows.

Exercise 348, page 277

Suppose that f is not nondecreasing on $[a,b]$. Then we can choose points $a \leq a' < b' \leq b$ with $f(b') < f(a')$. Thus $[f(b'), f(a')]$ is a nonempty compact subinterval of $[c,d]$. Take any y between $f(b')$ and $f(a')$. Let

$$M(y) = \sup\{x \in (a', b') : f(x) = y\}.$$

Check that $f(x) = y$ and that $D^+ f(x) \leq 0$ whenever $x = M(y)$. Thus, $y \in f(D)$. Consequently $f(D)$ contains $(f(b'), f(a'))$ and so, also, all compact subintervals contained in this open interval.

Exercise 349, page 277

We break the proof into a number of steps that follow Morse's original exposition.

Step 1. Suppose that f is strictly decreasing on a compact set $E \subset [a,b]$. If E

contains no subinterval, then we claim that $f(E)$ is a compact subset of $[c,d]$ that also contains no interval.

We can define a strictly decreasing, continuous function $g:\mathbb{R}\to\mathbb{R}$ so that $f(x)=g(x)$ for all x in E by making g continuous and linear on all the open intervals complementary to E. We know that $f(E)=g(E)$ would be compact. Suppose, contrary to what we want, that $g(E)$ contains a subinterval J of $[c,d]$. We consider the inverse function g^{-1} which maps that subinterval J back into E. Such a function would be continuous and therefore maps J to some interval. That would require E to contain an interval.

Step 2. Define, for each integer $n=1,2,3,\dots,$

$$E_n=\{x\in[a,b]: f(x+h)-f(x)\le -h/n \text{ whenever } 0\le h\le 1/n\}$$

Then we will prove that E_n is a compact subset of $[a,b]$ that contains no interval and that $f(E_n)$ is a compact subset of $[c,d]$ that contains no interval.

It is easy to check, using the continuity of f, that E_n is closed. Thus both E_n and $f(E_n)$ must be compact. We subdivide $[a,b]$ into a finite collection $\{J_k\}$ of compact, nonoverlapping subintervals of $[a,b]$, covering all of that interval and each of length less than $1/n$. It is easy to see that f is strictly decreasing on each set $J_k\cap E_n$. By our hypotheses the set A is dense in $[a,b]$ so that no one of these sets $J_k\cap E_n$ can contain an interval. In particular E_n itself can contain no interval. Moreover, by step 1, we conclude that $f(J_k\cap E_n)$ is a compact set that contains no subintervals of $[c,d]$. It follows that $f(E_n)$ is contained in the finite union of such sets and so must itself contain no subintervals of $[c,d]$.

Step 3. The set B is a meager subset of $[a,b]$ and the set $f(B)$ is a meager subset of $[c,d]$. This follows from step 2 since B is the union of the sequence of sets $\{E_n\}$ each of which is a meager subset of $[a,b]$, while $f(B)$ is the union of the sequence of sets $\{f(E_n)\}$, each of which is a meager subset of $[c,d]$.

Step 4. Suppose now that f is not nondecreasing on $[a,b]$. Then we can choose points $a\le a'<b'\le b$ with $f(b')<f(a')$. Thus $[f(b'),f(a')]$ is a nonempty compact subinterval of $[c,d]$. We know from the proof of the preliminary lemma that f maps the set

$$D=\{x\in[a,b]: D^+f(x)\le 0\}$$

onto a set containing the open interval $(f(b'),f(a'))$. But we already have established that the set $f(B)$ is a meager subset of $[c,d]$. Using the fact that $B\cup C=D$, we conclude that $f(B)\cup f(C)=f(D)\supset(f(b'),f(a'))$. Thus $f(C)$ must contain a Thus $f(C)$ must contain a residual subset of the interval $[f(b'),f(a')]$.

Exercise 351, page 277

For example, consider the set
$$E = \{x : D^+ f(x) < r\}$$
and write, for positive integers m and n,
$$E_{mn} = \{x : \ f(x+t) - f(x) - rt + t/m \leq 0 \text{ for all } 0 \leq t \leq 1/n\}.$$
Since f is continuous, we can check that each set E_{mn} is closed. But
$$E = \bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} E_{mn}$$
reveals that E must be Lebesgue measurable.

Exercise 354, page 278

Use the Darboux property of continuous functions. As a more challenging exercise the student may wish to prove this without the assumption of continuity.

Exercise 356, page 278

This follows immediately from Lemma 6.19 since we know (Exercise 340) that, for every continuous function f, the measure $\lambda_f^\star$ must be σ-finite.

Exercise 358, page 279

Let
$$\beta_1 = \{(I,x) : s\lambda(I) < |h(I,x))\}$$
and
$$\beta_2 = \{(I,x) : |h(I,x))| < r\lambda(I)\}\,.$$
Note that β_1 is a fine cover of E and that β_2 is a full cover of E. Let β be any full cover of E and note that $\beta_1 \cap \beta$ is a fine cover of E and that $\beta_2 \cap \beta$ is a full cover of E. Thus
$$V^*(h,E) \leq \mathrm{Var}(h, \beta \cap \beta_2) \leq r\,\mathrm{Var}(\lambda, \beta \cap \beta_2) \leq r\,\mathrm{Var}(\lambda, \beta).$$
From this it follows that
$$V^*(h,E) \leq r\lambda^*(E).$$
Similarly
$$sV_*(\lambda,E) \leq \mathrm{Var}(s\lambda, \beta \cap \beta_1) \leq \mathrm{Var}(h, \beta \cap \beta_1) \leq \mathrm{Var}(h,\beta).$$
From this it follows that
$$s\lambda_*(E) \leq V^*(h,E).$$

Exercise 359, page 279

Use the methods already seen for Exercise 358.

Exercise 361, page 279

Write
$$E_n = \{x \in E : \overline{\mathrm{lip}}_f(z) < n\}.$$
By Lemma 6.16
$$\lambda_f(E_n \cap [-n,n]) \leq n\lambda(E_n \cap [-n,n]) < \infty.$$
It follows that f has σ-finite variation in E. Note then, that if N is a null subset of E,
$$\lambda_f(N) \leq \sum_{n=1}^{\infty} \lambda_f(E_n \cap N) \leq n\lambda(E_n \cap N) = 0.$$
This proves the final assertion.

Exercise 362, page 282

If S is a null set then $Z = S$ solves the exercise. Otherwise construct such a set by first taking a countable dense subset Z_1 of S. [The endpoints of the complementary intervals will suffice, unless S contains an interval. If S does contain an interval then include all rational numbers in that interval.] Now Z_1 is a countable subset of S and so has measure zero. For each integer n choose an open set G_n containing Z_1 with $\lambda(G_n) < 1/n$. Finally check that $Z = S \cap \bigcap_{n=1}^{\infty} G_n$ is a $\mathcal{G}_\delta$-set and that $\lambda(Z) = \lambda(Z_1) = 0$.

Exercise 364, page 290

Take g' to denote the derivative of g where that exists and 0 otherwise; such a function is measurable and we will be able to apply Exercise 274.

Observe first that if $[c,d]$ is any compact interval then
$$|\Delta g([c,d])| \leq \int_{[c,d]} |g'(x)|\,dx.$$
This follows from the fact that g is continuous so that
$$|\Delta g([c,d])| \leq \lambda(g([c,d]) \leq \lambda(g([c,d] \cap N) + \lambda(g([c,d] \setminus N) =$$
$$\lambda(g([c,d] \cap N) \leq \lambda_g([c,d] \cap N) = \int_{[c,d]} |g'(x)|\,dx.$$
Use Theorem 6.7 and Theorem 6.24. Now apply Exercise 274 to obtain, for every $\varepsilon > 0$, a $\delta > 0$ so that if G is an open set with $\lambda(G) < \delta$ then
$$\int_G |g'(x)|\,dx < \varepsilon.$$

In particular, if we are given any sequence of nonoverlapping intervals $\{[c_n,d_n]\}$ for which $\sum_n \lambda([c_n,d_n]) < \delta$ then there is an open set G covering these intervals for which $\lambda(G) < \delta$; it follows that

$$\sum_n |\Delta g([c_n,d_n])| \leq \sum_n \int_{[c_n,d_n]} |g'(x)|\,dx \leq \int_G |g'(x)|\,dx < \varepsilon.$$

Exercise 365, page 294

Under these hypotheses there is an indefinite integral F of the function f on the open interval (a,b). If F is uniformly continuous on (a,b) then we know that f is integrable on all of $[a,b]$. Thus it is enough to establish that when f has continuous upper and lower integrals on $[a,b]$ it follows that F is uniformly continuous on (a,b).

Exercise 366, page 294

Define F appropriately, starting with

$$F(z) = \sum_{[a_i,b_i]\subset[a,z]} \int_{a_i}^{b_i} f(x)\,dx$$

for any $z \in E$ and, for $z \in (a_j,b_j)$, set

$$F(z) = F(a_j) + \int_{a_j}^{z} f(x)\,dx.$$

Obtain $V^*(\Delta F - f\lambda,[a,b]) = 0$ from $V^*(f\lambda,E) = 0$, $V^*(\Delta F,E) = 0$, and

$$V^*(\Delta F - f\lambda,[a,b]\setminus E) \leq \sum_{i=1}^{\infty} V^*(\Delta F - f\lambda,[a_i,b_i]).$$

Exercise 367, page 300

Let

$$\beta = \{(u,v],w) : \ f\chi_{[u,v]} \text{ belongs to } \mathcal{J}\}.$$

This is a full cover of the set of points at which f is not $\mathcal{J}$-singular. Consequently if $[a,b]$ is a compact interval that contains no $\mathcal{J}$-singular points of f there must be a partition π of $[a,b]$ from β. Since $f\chi_{[u,v]}$ belongs to $\mathcal{J}$ for each $(u,v],w) \in \pi$ it follows that $f\chi_{[a,b]}$ also belongs to $\mathcal{J}$ (by property (3) of Definition 6.43 describing integration methods).

Exercise 369, page 307

It follows from Theorem 6.52 which we prove a little later on that, in fact,

$$[L(b)-L(a)] \leq \underline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f(x)\,dx \leq [U(b)-U(a)].$$

That shows immediately that

$$[L(b)-L(a)] \leq [U(b)-U(a)].$$

The reader might, however, want to write out a simple proof using a Cousin covering argument, using Theorem 6.52 as a model.

Exercise 370, page 307

This is the definition that Saks [73, p. 191] uses for major and minor functions. This allows the function being integrated to assume infinite values.

Exercise 371, page 308

This is the definition that can often be found in the literature for the major and minor functions employed in defining the Perron integral. This allows some flexibility in choosing major and minor functions. It does not, however, change the scope of the integral itself.

Exercise 372, page 308

Saks [73, p. 1252] alludes to this as a further possible definition for major and minor functions that could be used to develop a Perron integral.

Exercise 379, page 317

Choose $z \in E$ but not in E_1. Consider the intervals

$$I_n = (z-1/n, z+1/n).$$

If $E_1 \cap I_n \neq \emptyset$ for all n then we can deduce that z would have to belong to the closed set E_1.

Exercise 380, page 317

If $n=1$ this is obvious. Use induction on n.

Exercise 381, page 318

Suppose not, i.e., suppose that none of the sets E_k contain a portion of E. Then, using Exercise 380, select a portion $E \cap (a_1, b_1)$ so that $E \cap (a_1, b_1) = \emptyset$ and pass to a closed subinterval $[c_1, d_1] \subset (a_1, b_1)$ for which $E \cap [c_1, d_1] \neq \emptyset$. Continue inductively, choosing portions

$$E \cap (a_n, b_n) \subset [c_{n-1}, d_{n-1}]$$

and closed subintervals $[c_n, d_n] \subset (a_n, b_n)$ for which $E \cap [c_n, d_n] \neq \emptyset$.

The nested sequence of intervals $\{[c_n, d_n]\}$ all must contain a point z of E in common. But this point z cannot belong to any of the sets E_k which is in contradiction to the hypothesis that $E \subset \bigcup_{k=1}^{\infty} E_k$.

Exercise 382, page 318

To adjust the proof, at the nth stage of the induction select the interval $(a_n, b_n) \subset G_n$. The point z you will find must belong to each of the G_n and, consequently, to E. Sets of the form $E = \bigcap_{j=1}^{\infty} G_j$ for some sequence $\{G_j\}$ of open sets are said to be sets of type $\mathcal{G}_\delta$.

Exercise 384, page 321

This is "intuitively obvious." Certainly in dimension one, length is additive, in dimension two area is additive, in dimension three volume is additive, etc.

Well no. While the truth of the statement is hardly surprising and it is indeed trivial in dimension one, a proof would still be needed. Not all textbooks might supply such a proof but if you search enough there should be a number of examples. McShane proves this as Lemma 2-1 (p. 255) of his text and includes the following comment:

> In higher-dimensional spaces the result is still true, but the proof of that fact is tedious. Some people may think that this additivity is "intuitively evident" and that it is a waste of time to prove it. But even in the plane there are far more complicated dissections of an interval into subintervals than simple checkerboard patterns. . . . Besides that, who can honestly say that he has any clear-cut "intuitions" about 19-dimensional space?

E.J. McShane, *Unified Integration*, Academic Press (1983).

Exercise 385, page 321

This is the higher dimensional version of the Cousin lemma that was used extensively in the elementary chapters. As is usual in mathematics the higher dimension version can be proved by a similar method provided one takes the time to modify the argument as needed. The key tool in dimension one was the nested interval property asserting that a shrinking sequence of closed bounded intervals converged to a point. The same is true in higher dimensions. Having established this fact the proof of the Cousin lemma is then straightforward.

If you need to see a formal proof see Theorem 3-1, p. 258 in E. J. McShane, *Unified Integration*, Academic Press (1983). Henstock also takes the trouble to prove this assertion in detail in Theorem 4.1, p. 43 of R. Henstock, Lectures on the Theory of Integration, World Scientific (1988).

Exercise 394, page 326

Do not use Theorem 7.8 since that is not the point of the exercises.

Exercise 395, page 326

You can use Lemma 7.2.

Exercise 399, page 330

This will require an application of the dominated convergence theorem. For details that can be used to prove this exercise as well as the preceding two exercises see the proof of Theorem 4–1, pp. 262–264 in E. J. McShane, *Unified Integration*, Academic Press (1983).

Exercise 401, page 331

This is given as Corollary 4–2, pp. 265-266 in E. J. McShane, *Unified Integration*, Academic Press (1983).

Exercise 402, page 331

You should be able to verify that

$$\int_{-1}^{1}\left(\int_{-1}^{1} f(x,y)\,dx\right)dy = \int_{-1}^{1}\left(\int_{-1}^{1} f(x,y)\,dy\right)dx = 0$$

while the double integral

$$\int\int_{[-1,1]\times[-1,1]} f(x,y)\,dxdy$$

fails to exist. For the double integral consider first the integrals over the squares $[n^{-1},1]\times[n^{-1},1]$ and $[-1,-n^{-1}]\times[-1,-n^{-1}]$ for $n=2,3,4,\dots$.

Exercise 403, page 331

You should be able to verify that
$$\int_0^1\left(\int_{-1}^1 f(x,y)\,dy\right)dx=0$$
but that both
$$\int_{-1}^1\left(\int_0^1 f(x,y)\,dx\right)dy$$
and the double integral
$$\int\int_{[-1,1]\times[-1,1]} f(x,y)\,dxdy$$
fails to exist. For the double integral consider first the integrals over the intervals $[n^{-1},1]\times[0,1]$ for $n=2,3,4,\dots$.

Exercise 404, page 332

You should be able to verify that
$$\int_0^1\left(\int_{-1}^1 f(x,y)\,dy\right)dx=1$$
but that
$$\int_{-1}^1\left(\int_0^1 f(x,y)\,dx\right)dy=-1$$
while the double integral
$$\int\int_{[0,1]\times[0,1]} f(x,y)\,dxdy$$
fail to exist. For the double integral consider that the function $f(x,y)>4/(27x^2)$ at every point in the set
$$\{(x,y):n^{-1}\le x\le 1,\ 0<y<x/2\}.$$

Exercise 405, page 332

Well you can indeed define anything you want but it needs to be consistent and useful. There are (see Exercise 403) situations in which only one of the expressions
$$\int_c^d\left(\int_a^b f(x,y)\,dx\right)dy$$

and

$$\int_a^b \left(\int_c^d f(x,y)\,dy \right) dx$$

exists. So which order do we take as our definition? There are also situations in which both exist (see Exercise 404) but have different values! This student's version of an integral would not even, then, allow us to rotate the axes by a right-angle without changing the integral radically. There are also situations in which both integrals exist and have the same value but the double integral does not exist in our sense (see Exercise 402) and shouldn't exist since it leads to unpleasant conclusions.

Exercise 406, page 334

Take any subdivision of $[a,b]$,

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

Let

$$E_i = \{x \in E : x_{i-1} < f(x) \leq x_i\}$$

and recall that

$$\mathcal{L}^n(E_i) = w(x_i) - w(x_{i-1}).$$

Check that

$$x_{i-1}[w(x_i) - w(x_{i-1})] \leq \int_{E_i} f(x)\,dx \leq x_i[w(x_i) - w(x_{i-1})].$$

This connects the Riemann sums

$$\sum_{i=1}^n x_i[w(x_i) - w(x_{i-1})] \text{ and } \sum_{i=1}^n x_i[w(x_i) - w(x_{i-1})]$$

with the integral

$$\int_{\{x \in E: a < f(x) \leq b\}} f(x)\,dx = \sum_{i=1}^n \int_{E_i} f(x)\,dx.$$

Exercise 407, page 334

Remember that, the infinite integral

$$\int_{-\infty}^{\infty} s dw(s)$$

would be the same as

$$\lim_{n to \infty} \int_{-n}^{n} s dw(s).$$

Thus we need to show that

$$\int_E f(x)\,dx = \lim_{n to \infty} with \int_{\{x \in E: -n < f(x) \leq n\}} f(x)\,dx$$

which is a simple measure-theoretic computation.

Bibliography

[1] A. Aleksandrov, ÜUber die Aquivalenz des Perronschen und des Denjoyschen Integralbegriffes, Math. Z. 20 (1924), 213–222.

[2] D. Austin, A geometric proof of the Lebesgue differentiation theorem. Proc. Amer. Math. Soc. 16 (1965) 220–221. 116

[3] Robert G. Bartle, *A modern theory of integration*, American Mathematical Society, Providence, 2001.

[4] H. Bauer, Der Perronschen Integralbegriff und seine Beziehung zum Lebesgueschen, Monatsch. Math. Phys. 26(1915), 153–198. 306

[5] Hana Bendová and Jan Malý, An elementary way to introduce a Perron-like integral, Annales Academiae Scientiarum Fennicae Mathematica, Volumen 36, 2011, 153–164. 36, 37

[6] Sterling K. Berberian, Classroom Notes: Regulated Functions: Bourbaki's Alternative to the Riemann Integral. Amer. Math. Monthly 86 (1979), no. 3, 208–211. 43

[7] M. W. Botsko, An elementary proof of Lebesgue's differentiation theorem, Amer. Math. Monthly 110 (2003), no. 9, 834–838. 70, 116

[8] David Bressoud, *Radical Approach to Real Analysis*, Washington, D.C.: Mathematical Association of America, 2007. 28

[9] David Bressoud, *A Radical Approach to Lebesgue's Theory of Integration.*, Cambridge, MA: Cambridge University Press, 2008. 350

[10] D. S. Bridges, Lecture notes on a Riesz-method development of the Lebesgue integral: http://www.math.canterbury.ac.nz/˜d.bridges/ 209

[11] Andrew M. Bruckner, Differentiation of real functions (Lecture Notes in Math. 659, Springer- Verlag, New York, 1978). 371

[12] Andrew M. Bruckner and Brian S. Thomson, Real variable contributions of G. C. Young and W. H. Young, Expo. Math. 19 (2001), no. 4, 337–358.

[13] Andrew M. Bruckner, Brian S. Thomson, and Judith B. Bruckner, Real Analysis, 2nd Ed., ClassicalRealAnalyis.com (2008). 153, 222

[14] Peter S. Bullen, Non-absolute integrals: a survey, Real Anal. Exchange 5 (1980), 195–259.

[15] Peter S. Bullen, Nonabsolute integration in the twentieth century, AMS Special Session on Nonabsolute Integration University of Toronto, Toronto, Canada, September 23-24, 2000. Available for download: http://www.emis.de/proceedings/Toronto2000/

[16] C. Carathédory, *Vorlesungen über reelle Funktionen*, Leipzig-Berlin (1918). 209

[17] V. G. Celidze and A. G. Dzvarseisvili, *The theory of the Denjoy integral and some applications* (trans. Peter S. Bullen), World Scientific, Singapore, 1989.

[18] K. Ciesielski, "How good is Lebesgue measure?" Math. Intelligencer 11(2), 1989, 54–58.

[19] K. Ciesielski, *Set Theory for the Working Mathematician*, Cambridge, England: Cambridge University Press, 1997. 305

[20] P. J. Daniell, A general form of integral, Ann. of Math. (2) 19 (1917/18), 279–294. 200

[21] Roy O. Davies, An elementary proof of the theorem on change of variable in Riemann integration, The Mathematical Gazette, Vol. 45, No. 351 (1961), 23–25. 260

[22] Arnaud Denjoy, Une extension de l'intégrale de M. Lebesgue, C. R. Acad. Sci. Paris 154 (1912), 859-862.

[23] Arnaud Denjoy, Sur les fonctions derivées sommables, Bull. Soc. Math. France 43 (1915), 161–248. 193

[24] Arnaud Denjoy, *Mémoire sur la dérivation et son calcul inverse*, Gauthier-Villars, Paris (1954).

[25] Jacques Dixmier, Cours de mathématiques du premier cycle. Cahiers Scientifiques, Fasc. 30. Gauthier-Villars, Paris, 1967. 43

[26] Randall Dougherty and Alexander S. Kechris, The complexity of antidifferentiation. Adv. Math. 88 (1991), no. 2, 145–169. 61, 298, 306

[27] Vasile Ene, Riesz type theorems for general integrals, Real Anal. Exchange 22, no. 2 (1997), 714–733. 124

[28] Claude-Alain Faure, The Lebesgue differentiation theorem via the rising sun lemma, Real Anal. Exchange 29 (2003/04), no. 2, 947–951. 70

[29] James Foran and Sandra Meinershagen, Some answers to a question of P. Bullen, Real Anal. Exchange 13 (1987/88), no. 1, 265–277. 373

[30] Chris Freiling, How to compute antiderivatives, Bull. Symbolic Logic 1 (1995), no. 3, 279–316. 6, 306

[31] Chris Freiling, On the problem of characterizing derivatives, Real Anal. Exchange 23 (1997/98), no. 2, 805–812.

[32] Casper Goffman, A bounded derivative which is not Riemann integrable. Amer. Math. Monthly 84 (1977), no. 3, 205–206. 52

[33] L. Gordon and S. Lasher, An elementary proof of integration by parts for the Perron integral, Proc. Amer. Math. Soc. 18 (1967), 394-398. [14]

[34] Russell A. Gordon, *The integrals of Lebesgue, Denjoy, Perron, and Henstock*, American Mathematical Society, Providence, 1994.

[35] D. C. Gillespie, The Cauchy Definition of a Definite Integral, Annals of Mathematics, Second Series, Vol. 17, No. 2 (1915) 61–63. 350

[36] Judith V. Grabiner, Who gave you the epsilon? Cauchy and the origins of rigorous calculus, American Mathematical Monthly 90 (3), 1983, 185–194. 11

[37] John W. Hagood, The Lebesgue differentiation theorem via nonoverlapping interval covers, Real Anal. Exchange 29 (2003/04), no. 2, 953–956. 311

[38] John W. Hagood and Brian S. Thomson, Recovering a function from a Dini derivative, Amer. Math. Monthly 113 (2006), no. 1, 34–46. 312

[39] H. Hake, Über de la Vallee Poussins Ober-und Unterfunktionen, Math. Ann. 83 (1921), 119–42.

[40] Ralph Henstock, *Lectures on the theory of integration*, World Scientific, Singapore, 1988.

[41] Ralph Henstock, *The general theory of integration*, Oxford University Press, Oxford, 1991.

[42] E. Kamke, Zur Definition der approximate stetigen Funktionen, Fund. Math. I0 (1927), 431–433. 193

[43] Hyman Kestelman, Change of variable in Riemann integration, The Mathematical Gazette, Vol. 45, No. 351 (1961), 17–23. 260

[44] Douglas S. Kurtz and Charles W. Swartz, *Theories of integration: the integrals of Riemann, Lebesgue, Henstock-Kurzweil, and Mcshane,* World Scientific, Singapore, 2004.

[45] J. Kurzweil, Generalized ordinary differential equations and continuous dependence on a parameter. Czechoslovak Math. J. 7(82), 1957, 418–449.

[46] J. Kurzweil, *Nichtabsolut konvergente Integrale*, Teubner-Texte zur Mathematik 26, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1980.

[47] Jaroslav Kurzweil, *Henstock-Kurzweil integration: its relation to topological vector spaces*, World Scientific, Singapore, 2000.

[48] Jaroslav Kurzweil, *Integration between the Lebesgue integral and the Henstock-Kurzweil integral: its relation to local convex vector spaces*, World Scientific, Singapore, 2002.

[49] Solomon Leader, *The Kurzweil-Henstock integral and its differentials*, Marcel Dekker, New York, 2001.

[50] Henri Lebesgue, Remarques sur la définition de l'intégrale, Bull. Sci. Math. **29** (1905) 272–275. 6

[51] Peng Yee Lee, *Lanzhou lectures on Henstock integration*, World Scientific, Singapore, 1989.

[52] Peng Yee Lee and R. Výborný, *The Integral: An Easy Approach after Kurzweil and Henstock*, Cambridge. Univ. Press, 2000, ISBN 0 521 77968 5.

[53] Beppo Levi, Ricerche sulle funzioni derivate, Rnd. dei Lincei, (5), Vol. XV, 1906 p. 437.

[54] H. Looman, Über die Perronsche integral definition, Math. Ann. 93 (1935), 153–156.

[55] Jaroslav Lukeš, A topological proof of Denjoy-Stepanoff theorem. Časopis Pěst. Mat. 103 (1978), no. 1, 95–96, 98. 194

[56] J. Marcinkiewicz, Sur les séries de Fourier, Fund. Math. 27 (1936), 38–69. 311

[57] Jean Mawhin, *Analyse–fondements, techniques, evolution*, De Boeck Universite, Brussels, 1992.

[58] Robert M. McLeod, *The generalized Riemann integral*, The Mathematical Association of America, Washington, 1980.

[59] Edward James McShane, On Perron integration, Bull. Amer. Math. Soc. 48 (1942), 718–726.

[60] Edward James McShane, *Integration*, Princeton University Press, 1944.

[61] F. Medvedev (trans. R. Cooke), *Scenes from the history of real functions*, Birkhauser Verlag, Basel, 1991. 7

[62] Patrick Muldowney, *A general theory of integration in function spaces*, Longman, Harlow, 1987.

[63] I. P. Natanson, *Theory of functions of a real variable*, 2nd ed., revised, New York, 1960. 364

[64] Jiří Navrátil, A note on the theorem on change of variable in a Riemann integral, (Czech. English summary) Časopis Pěst. Mat., Vol. 106 (1981), No. 1, 79–83. 261

[65] O. Perron, Über den Integralbelgriff, Sitzber', Heidelberg Akad. Wiss. Abt. A 16 (1914), 1–16. 306

[66] I. N. Pesin, *Classical and modern integration theory*, (New York, 1970).

[67] Washek Pfeffer, *The Riemann approach to integration*, Cambridge University Press, Cambridge, 1993.

[68] Washek Pfeffer, *Derivation and integration*, Cambridge University Press, Cambridge, 2001.

[69] David Preiss and Jaromír Uher, A remark on the substitution for the Riemann integral. (Czech. English summary) Časopis Pěst. Mat., Vol. 95 (1970), No. 4, 345–347. 260, 261

[70] Frédéric Riesz, Sur certain systèmes singulier d'equations intégrales, Annales de L'École Norm. Sup., Paris (3) 28 (1911), 33–62. 123

[71] Frédéric Riesz, "Sur un Théorème de Maximum de Mm. Hardy et Littlewood", Journal of the London Mathematical Society 7 (1) (1932), 10–13. 66

[72] Herbert E. Robbins, Note on the Riemann integral, American Math. Monthly, Vol. 50, No. 10 (Dec., 1943), 617–618. 255

[73] Stanisław Saks, *Theory of the Integral*, 2nd revised Ed., Dover, New York, 1964. 209, 259, 310, 311, 392

[74] D. N. Sarkhel, A criterion for Perron integrability, Proc. Amer. Math. Soc. 7 (1978), 109-112. MR 58 17006. 311

[75] Eric Schechter, *Handbook of analysis and its foundations*, Academic Press, San Diego, 1997.

[76] Stefan Schwabik, *Generalized ordinary differential equations*, World Scientific, Singapore, 1992.

[77] James Serrin and Dale E. Varberg, A general chain rule for derivatives and the change of variables formula for the Lebesgue integral, Amer. Math. Monthly, 76 (1969) 514–520. 258, 259

[78] W. Stepanoff, Sur une proprieté caractéristique des fonctions mesurables, Rec. Math. Soc. Math. Moscou 30 (1924), 487–489. 193

[79] Charles Swartz, *Introduction to gauge integrals*, World Scientific, Singapore, 2001.

[80] E. Talvila, Characterizing integrals of Riemann integrable functions, Real Anal. Exchange 33, no. 2 (2007), 487-488. 120

[81] G. P. Tolstov, Sur l'intégrale de Perron, Mat. Sb. 5 (1939), 647–659: 1. 208. 311

[82] G. P. Tolstov, Parametric differentiation and the narrow Denjoy integral. Mat. Sb. (N.S.) 53(95), 1961, 387–392.

[83] Brian S. Thomson, *Real functions*. Lecture Notes in Mathematics, 1170. Springer-Verlag, Berlin, 1985. 371

[84] Brian S. Thomson, Characterizations of an indefinite Riemann integral. Real Anal. Exchange 35 (2010), no. 2, 487–492. 120

[85] Brian S. Thomson, *The Calculus Integral*, ClassicalRealAnalyis.com (2010). 41

[86] Brian S. Thomson, On Riemann sums, Real Analysis Exchange 37, no. 1 (2011), 221-242.

[87] Brian S. Thomson, A strong kind of Riemann integrability, College Mathematics Journal 43:4, (2012) 313–321.

[88] Brian S. Thomson, Judith B. Bruckner, and Andrew M. Bruckner, *Elementary Real Analysis, 2nd Edition*, ClassicalRealAnalyis.com (2008). ISBN-13: 978-1434843678, ISBN-10: 143484367X 153

[89] Brian S. Thomson, Judith B. Bruckner, and Andrew M. Bruckner, *Elementary Real Analysis: The Dripped Version*, ClassicalRealAnalyis.com (2009). ISBN-13: 978-1438248509, ISBN-10: 1438248504

[90] G. Vitali, Una proprietà delle funzioni misurabili, Istit. Lombardo Rend. (2), 38 (1905) 599–603. 209

[91] W. H. Young, Upper and lower integration, Proc. London Math. Soc. (2) 2 (1904), 52–66. 200

[92] W. H. Young, A note on the property of being a differential coefficient, Proc. London Math. Soc. (2) **9** (1911) 360–368.

[93] Elias Zakon, *Mathematical Analysis I*, ISBN 1-931705-02-X, published by The Trillia Group, 2004. 42

Index