THE INTEGRAL CALCULUS

Alan Smithee

For information about this publication consult the web site
classicalrealanalysis.com

Direct all correspondence to Alan Smithee
calculusprogram **at** gmail.com

1991 *Mathematics Subject Classification.* Primary 26A06, 26A24, 26A39, 26A42; Secondary 26A45, 26A46,

Key words and phrases. Derivative, integral, Newton integral, Henstock-Kurzweil integral.

VERSION 0.22 [BETA]
OCTOBER 24, 2006.
Direct correspondence to Alan Smithee .

ABSTRACT. The D.R.I.P. program (DUMP THE RIEMANN INTEGRAL) proposes that the usual calculus treatment of the Riemann integral and the introduction of so-called improper integrals be dropped in favor of a simpler integral which includes those theories. These notes suggest a way to do this.

Preface

These informal calculus notes are offered to those instructors who may wish to design a calculus sequence that drops the Riemann integral in favor of the more natural integral on the real line that has been called, at diverse times, by many different names[1].

The program to drop the classical Riemann integral from the undergraduate curriculum has been called by some

D.R.I.P. [Dump the Riemann Integral Project]

and has a small and mostly ignored band of followers. Encouraged by the authors of the web site

www.classicalrealanalysis.com

I have made these notes freely available to anyone wishing to use them for these purposes. I am grateful to these three authors, but absolve them of any responsibility for the many flaws of this presentation.

I am not the originator of these ideas, merely a scribe who has undertaken to deliver and interpret some notes that fell into my possession under unusual circumstances. These will be related in a later publication

THE FUNDAMENTAL PROGRAM OF THE CALCULUS,
by Alan Smithee (2007) *[to appear]*.

The notes are not classroom tested. Indeed I am far too timid to try them out. My chairman would have my hide if I were to teach our calculus students anything other than the Riemann integral and the improper integral from the standard prescribed texts. Moreover, should I try covertly to expose the students in my sections to these ideas and these simpler methods, they might fail in the general examinations. For example, my students would arrive at the formula

$$\int_a^b x^p\,dx = \frac{b^{p+1}-a^{p+1}}{p+1} \quad (p>0)$$

by the simple justification of the differentiation formula

$$\frac{d}{dx}x^{p+1}/(p+1) = x^p.$$

But this would be graded as incomplete because traditional calculus students (using the dreadful Riemann integral) are obliged also to verify that the function $f(x)=x^p$ is in fact Riemann integrable. The very fact that it is a derivative of another

[1]The Denjoy integral, the Perron integral, the Denjoy-Perron integral, the restricted integral of Denjoy, the Denjoy total, the Henstock integral, the Kurzweil integral, the Henstock-Kurzweil integral, the Kurzweil-Henstock integral, the generalized Riemann integral, the Riemann-complete integral, the gage integral, the gauge integral, etc. We call it simply *the integral*.

function is adequate justification in the correct theory of integration, but not at all for the Riemann integral.

More distressingly would be their solution to this problem: Verify that

$$\int_0^1 x^p\,dx = \frac{1}{1+p} \qquad (-1 < p < 0).$$

My students would use the same differentiation formula (valid for $0 < x \leq 1$) and merely point out that the function

$$F(x) = \frac{x^{p+1}}{p+1}$$

is continuous on $[0,1]$ provided that $-1 < p < 0$. They would then claim that

$$\int_0^1 x^p\,dx = F(1) - F(0) = \frac{1}{1+p} \qquad (-1 < p < 0).$$

Again wrong? Few of my colleagues would understand that this is perfectly correct; they would demand instead some nonsense about unbounded functions, improper integrals, and limits on the right-hand side at the endpoint 0.

But if these worries do not alarm you then, certainly, try out the notes. But let me know of the successes and failures as well as improvements that can be made in the notes.

Alan Smithee

Direct correspondence to Alan Smithee

Course Design

The DRIP program is best viewed as a thought experiment for academic mathematicians. There is certainly no chance of the early calculus courses changing. Any textbook writer who decided to alter material for a subsequent edition would receive frantic and panicked e-mails from his editor. The rule of the market seems to be that all calculus books must look essentially like all previously successful calculus books. But this shouldn't stop us from thinking about it. Some brave souls may even try the ideas out.

0.1. Step 1: Use Newton's integral

The least ambitious course of action for the calculus reformer to try is to introduce only the Newton integral. Drop any attempt at starting integration theory off with Riemann sums and limits of these. Have such lectures ever been successful?

The fundamental theorem of the calculus becomes then a definition, not a theorem. In the narrowest version

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

means only that F is a continuous function on $[a, b]$ and that

$$F'(x) = f(x) \text{ for } a < x < b \text{ with finitely many exceptions.}$$

[See the Newton versions of the integral in Chapter 1.]

It is hard to imagine that this would handicap the students in any truly serious way since they are unlikely to have to integrate functions with more than a finite number of discontinuities. The notes clear the way for a countably infinite number of discontinuities if that is needed. When more is required the students should be ready, in any case, for the Lebesgue integral.

If you feel this little ambition, then you are obliged at least to mention (without proof) Theorem 5.12 which asserts that bounded functions with countably many discontinuities are integrable. For unbounded functions give the comparison test from Theorem 5.21. At that point any calculus treatment can be used since most of the theorems that are stated in such courses are for continuous functions.

The Newton integral offers the best introductory approach to integration. The final [advanced] version of the Newton integral can be expressed in this way:

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

means that F is a continuous function on $[a, b]$, that

$$F'(x) = f(x) \text{ for almost every } a < x < b\ ,$$

and the function F does not grow on the exceptional set. This integral is the natural integral; it includes the Riemann integral, the Lebesgue integral, and the usual improper integrals.

In the past this integral might have seemed quite distant from the calculus curriculum: there was always a firewall [measure theory] separating the freshman calculus integral from the graduate level Lebesgue integral. But the reader of these notes will see that, by Chapter 4, with nothing more than elementary techniques, all of this can be realized. (The firewall remains safely in place until Chapter 5.)

Zakon's Analysis text. There is currently at least one analysis textbook available[2] that follows exactly this program, replacing the Riemann integral by the Newton integral (with countably many exceptions):

> **Mathematical Analysis I**, by Elias Zakon, ISBN 1-931705-02-X, published by The Trillia Group, 2004. 355+xii pages, 554 exercises, 26 figures, hypertextual cross-references, hyperlinked index of terms. Download size: 2088 to 2298 KB, depending on format.

This can be downloaded freely from the web site

www.trillia.com/zakon-analysisI.html

Inexpensive site licences are available for instructors wishing to adopt the text.

Zakon's text offers a serious analysis course at the pre-measure theory level, and commits itself to the Newton integral. There are rigorous proofs and the presentation is carried far enough to establish that all regulated functions are integrable in this sense. Since regulated functions are uniform limits of step functions, it is easy to anticipate how this could be done.

One can argue, however, that any analysis course that demands such a serious treatment would be ill advised to stay too long at the level of this early stage Newton integral. Since the technical details for the treatment of the natural integral are no harder than the details in the rest of Zakon's text, instructors should at least consider the merits of Step 2.

0.2. Step 2: Replace the Riemann integral by the natural integral

For courses that would normally develop integration theory in a somewhat rigorous treatment [elementary real analysis courses in North America] introduce the natural integral, not the Riemann integral. The general ideas are identical, but the payoff of using the more general integral is huge.

There will be some reluctance. The ideas in Chapter 2, 3 and 4 are entirely elementary and completely consonant with the scope and intentions of any traditional elementary analysis course. But the instructor may never have seen them.

A full treatment of the integral will require the technical details in Chapter 2. The proof in Section 2.7 is a bit of a *tour de force* for calculus level students. It is not really harder than many other traditional proofs of theoretical calculus, but it requires a serious grasp of the definitions of continuity and differentiability as well as some facility with the covering relations used. But once the students rise to this level, the instructor will find that there are no new hurdles to place before them:

[2] I am indebted to Bradley Lucier, the founder of the Trillia Group, for this reference. The text has been used by him successfully for beginning graduate students at Purdue.

all of the remaining theorems in the text are at the same level and use very similar methods. Indeed it is this unity of methods that makes the subject so accessible.

0.3. Step 3: Prepare the way for the measure theory

Chapter 5 is largely for the instructor and students who will shortly go on to study the Lebesgue integral. It contains many familiar ingredients from the Lebesgue theory without invoking any of the machinery of measure theory.

Traditionally a graduate course in measure and integration develops much of the machinery of the Lebesgue integral and then, late in the game, shows that the Lebesgue integral includes the Riemann integral. That would then seem an end of it. What is neglected often is a discussion of how the improper integral of the calculus fits in (or more significantly does not fit in).

For those of us imagining a DRIP program, we would have to consider that the students have previously learned the natural integral on the real line but no measure theory. How would we reconcile the graduate measure course with the earlier elementary course that happens to develop an integral that is a proper extension of the Lebesgue integral?

In Chapter 5 the characterizations of the Lebesgue measure using notions of full and fine covers is given. (This is the Vitali covering theorem, albeit in an unusual expression.) The chapter is devoted to proving some of the fundamental (and deeper) results in the theory (the differentiation of the integral, Lebesgue's differentiation theorem for monotonic functions) and exhibiting where the measure-theoretic integral fits into the natural integration theory. The level of material is significantly higher and more background would be assumed of the student.

D.R.I.P. HISTORY

DUMP THE RIEMANN INTEGRAL PROJECT

> *The genuine history of ideas is not for amateurs: it requires skill, dedication, integrity, and certainly training in the discipline of the academic historian. There is another kind of history, wherein the details are mustered in order to tell a story. One decides in advance on the narrative one wishes to tell and selects the historical events and their interpretation to further that narrative. This history is in the latter tradition. There are, naturally, far more historians of this type than of the genuine type.*

The search for what should best be called *the natural integral on the real line* becomes rather heated in the late nineteenth century. By then the limitations of the popular Riemann integral became more fully appreciated. Most attention had been paid on the unsolved problem of how best to integrate unbounded functions. The discovery that there were bounded derivatives that failed to be Riemann integral meant to most that there was something seriously and fundamentally wrong with that theory.

Lebesgue's solution at the beginning of the twentieth century was to base a theory of integration on measure theory. There was enough of that to hand. Peano, Jordan, and Cantor had developed the theory of sets and their measure to a useful level. That was taken to a brilliant further step by Borel. For Lebesgue the path was clear and he developed it in a masterful way. He claimed as his motivation the fact that there existed differentiable functions F for which F' was not integrable in any of the known methods.

We know now that the premise itself is doomed to failure: measure theory alone cannot capture the integral of all functions that arise in analysis. In particular it cannot integrate all derivatives. Suppose that a function f on an interval is given and that the sets

$$\{x \in [a,b] : r \leq f(x) \leq s\}$$

together with their Lebesgue measures are given for all real r, s. Then the integral of f can be constructed from this information only for Lebesgue integrable functions. It can be done for all bounded derivatives; it cannot be done for *all* derivatives.

Fortunately for posterity Lebesgue did not know this in advance. His solution of the problem led to a development within a few decades of all of the tools of measure theory. His launch of this program was a greater contribution than finding the natural integral on the real line would have been.

Most teachers of the theory remark that Lebesgue abandoned the Riemann sums approach in favor of the measure-theoretic methods and that this is the source of the success of his project. In fact Lebesgue felt rather naturally compelled to show that his integral could be expressed as a limit of Riemann sums. He didn't, however, find the right filter to express this. If he had found it and thereby discovered the natural integral on the real line it might have delayed the far more important development of measure theory. But it is wrong to suggest that the success of the Lebesgue theory rests alone in the shift from Riemann sums to measures. Other techniques we now know would have generalized both integrals.

The correct and natural integral on the real line was discovered by Perron and Denjoy in the decade after Lebesgue's first work. Denjoy's idea was to extend the Lebesgue integral by a transfinite series of extensions. The series of extensions are motivated by the properties that derivatives possess. Perron's idea was to describe an integral that would have to include the Lebesgue integral and also integrate all derivatives.

Not much attention was paid to the theory of Denjoy and Perron except by some of the best analysts of the pre WWII period. The seminal book of the period,

> **The theory of the integral** by Stanislaw Saks, New York, (1937).

includes an account of the measure theory as well as the methods of Perron and Denjoy. Formidable mathematicians of the era (e.g., Banach, Steinhaus, Zygmund) were thoroughly grounded, not only in measure theory, but in these now mostly dead subjects. Later generation textbooks drop the subject. By the appearance of

> **Measure Theory** by Paul Halmos, Van Nostrand, New York, (1950).

(a standard text for many mid-twentieth century students) the subject is considered arcane and little discussed outside of certain circles[3].

In the late 1950s the simple definition of the natural integral was discovered. Henstock was working on nonabsolute integration and investigating ideas of Ward. Those ideas gelled into his formulation of a generalized Riemann integral, that he saw quickly was equivalent to the Denjoy-Perron integral. Kurzweil, at the same time, was investigating differential equations on the real line and observed the same simple formal definition of an integral that would express what he needed.

So by 1960 or so the stage was set for the introduction into the undergraduate curriculum of the natural integral on the real line. The definition was simple and natural enough for introduction at an early level. The connection with the fundamental theorem of the calculus was simpler and more compelling than in either the Riemann or Lebesgue integration theories. Measure theory could be introduced early on in simple ways or deferred until later. The students would be spared the confused mess caused by having numerous integration theories in place of one unifying theory: the Riemann integral, the improper Riemann integral, the Lebesgue integral, the Denjoy-Perron integral with their oddly overlapping domains would disappear.

So why didn't it happen? There were advocates.

> Henstock, Ralph "A Riemann-type integral of Lebesgue power." Canad. J. Math. 20 1968 79–87.

[3]A certain critic of the time preferred to refer to these "circles" by the more arrogant term "*cul de sac*."

offered a very readable and lucid treatment that might have had an influence if the curricula of the time had been less inflexible. At about this time, according to reliable reports, Henstock announced at an international meeting that "the Lebesgue integral is dead." It is curious that this statement was met largely with instant derision rather than any attempt to understand. Henstock meant by no means that measure-theoretic methods should be dropped; indeed he used them himself extensively. He meant that as a special integral on the real line, the Lebesgue integral should be replaced. Indeed the integral would disappear as a definition and reappear as a theorem: for all absolutely integrable functions f the integral

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

can be constructed from the sets

$$\{x \in [a,b] : r \leq f(x) \leq s\}$$

together with their Lebesgue measures for all real r, s. Later, in graduate school, the student could learn that this theorem serves as a useful definition of an integral in situations where no other structure than a measure space is available.

At about the same time appeared the memoir

> McShane, E. J. *A Riemann-type integral that includes Lebesgue-Stieltjes, Bochner and stochastic integrals.* Memoirs of the American Mathematical Society, No. 88 American Mathematical Society, Providence, R.I. 1969 54 pp.

which suggested just how flexible the methods were but attracted attention mostly from specialists. Much later

> Bartle, Robert G., "Return to the Riemann integral." Amer. Math. Monthly 103 (1996), no. 8, 625–632.

seemed at its appearance to suggest that there was some mainstream interest in the natural integral. But it is hard to detect any changes in the usual textbooks and the usual teaching of the subject.

The first formal attempt at influencing the undergraduate curriculum might be considered the beginning of the D.R.I.P. program. Robert Bartle, Ralph Henstock, Jaroslav Kurzweil, Eric Schechter, Stefan Schwabik, and Rudolf Výborný distributed a letter to publishers' representatives at the Joint Mathematics Meetings in San Diego, California, in January 1997. Their letter is available online at this website:

http://www.math.vanderbilt.edu/ schectex/ccc/gauge/letter/

More recently the web site

http://www.classicalrealanalysis.com

has been constructed and contains some information on the program without itself being an advocate.

Contents

CHAPTER 1

Newton's Original Integral

Integration as it was originally conceived of by Newton is the process inverse to differentiation. Such a process, he saw, would have numerous applications in geometry and physics.

In modern language we can describe the situation this way. His language was different but this is essentially how he viewed this process.

Definition 1.1 (Original Newton Integral). Suppose that f is a function defined on an interval (a, b) and that we can find a continuous function $F : [a, b] \to \mathbb{R}$ so that

$$F'(x) = f(x) \text{ for every } x \text{ with } a < x < b.$$

Then we will say that F is *an indefinite integral* of f on $[a, b]$ and we will write

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

and call the latter *the definite integral* of f on $[a, b]$.

1.0.1. Notes.

- If a function f has an indefinite integral then it has *many* indefinite integrals. If F is one such function then

$$F + \text{a constant}$$

  is also an indefinite integral.
- If both F and G are indefinite integrals of f on $[a, b]$ then

$$F = G + \text{a constant}$$

  .
- Using just this description, how will we know that f has an indefinite integral on an interval $[a, b]$? We have to *find* a function F, check that it is continuous on $[a, b]$, find the derivative of F, and then check that

$$F'(x) = f(x)$$

  at every point x inside the open interval (a, b).
- This is a *descriptive* definition. It says that we shall describe a certain situation using the terms "indefinite integral" and "definite integral." It does not say how we would ever arrive at such a situation; given a particular function f we might never be able to decide whether or not it is has an indefinite integral or a definite integral. Nor, if it does have a definite integral, would we be ever able to find its definite integral.

- How then can we generally evaluate $\int_a^b f(x)\,dx$? There will be other methods, but for now and for most problems in the calculus there is little option but to first find an indefinite integral F and then compute $F(b) - F(a)$.

1.0.2. Exercises.

EXERCISE 1. Check the first statement of the Notes (easy).

EXERCISE 2. Check the second statement of the Notes (needs the mean-value theorem).

EXERCISE 3. Find an indefinite integral for $f(x) = x^2$ that works on any interval.

EXERCISE 4. Evaluate

$$\int_0^1 x^2\,dx.$$

EXERCISE 5. Evaluate

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx.$$

(Note that the function being integrated does not have a value at $x = 0$; according to our definition it needs to be defined on the open interval $(0, 1)$ only.)

SOLUTION IN SECTION 8.1.1.

EXERCISE 6. Show the existence of and compute the value of the integral

$$\int_a^b x^p\,dx$$

for all $p > 0$ and all $a < b$.

EXERCISE 7. Show the existence of and compute the value of the integral

$$\int_0^b x^p\,dx$$

for all $-1 < p < 0$ and all $b > 0$.

EXERCISE 8. Show that the integral

$$\int_0^b x^p\,dx$$

does not exist in this sense for any $p \leq -1$ and $b > 0$.

1.1. Beyond the original Newton integral

The key technical fact that allows the descriptive definition just given for the Newton integral to work can be expressed as follows:

LEMMA 1.2. If F and G are both continuous functions on an interval $[a, b]$ and if $F'(x) = G'(x)$ for all $a < x < b$ then

$$F(b) - F(a) = G(b) - G(a).$$

PROOF IN SECTION 7.4.1.

This means that, while a function can have many different indefinite integrals, it has just one definite integral

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

that can be computed by taking *any* single indefinite integral F and finding the value of $F(b) - F(a)$.

We can generalize the integral by taking advantage of the following slight extension of Lemma 1.2 above.

LEMMA 1.3. If F and G are both continuous functions on an interval $[a, b]$ and if $F'(x) = G'(x)$ for all $a < x < b$ with at most finitely many exceptions then

$$F(b) - F(a) = G(b) - G(a).$$

PROOF IN SECTION 7.4.2.

This means an integration theory could be developed that would allow some exceptional points at which we do not have to check that the indefinite integral has the correct derivative. It is likely that Newton would have instantly accepted this version and would have recognized its utility.

1.1.1. An extension of the Newton integral.

DEFINITION 1.4 (Modified Newton Integral). Suppose that f is a function defined on an interval (a, b) except possibly at finitely many points. Suppose that we can find a continuous function $F : [a, b] \to \mathbb{R}$ so that $F'(x) = f(x)$ for every x with $a < x < b$ with perhaps finitely many exceptions. Then we will say that F is *an indefinite integral* of f on $[a, b]$ and we will write

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

and call the latter *the definite integral* of f on $[a, b]$.

Note that the definition does not require of the function being integrated that it be defined at every point of the interval (a, b). For example we could ask whether the following integral exists:

$$\int_a^b \frac{1}{\sqrt{|x-1|}\sqrt{|x-2|}\sqrt{|x-3|}\sqrt{|x-4|}}\,dx.$$

(Warning: don't try this one yet!) Most older calculus texts would have trouble with this and might insist that the interval (a, b) not include any of the points 1, 2, 3, or 4.

Sometimes they would allow this but require that some particular values be preassigned at the exceptional points (even though they would later turn out to be irrelevant).

EXERCISE 9. Compute

$$\int_{-1}^1 \frac{1}{\sqrt{|x|}}\,dx.$$

SOLUTION IN SECTION 8.1.2.

1.2. Larger exceptional sets

As integration theory developed over the centuries since Newton it became clear that the theory required quite large exceptional sets, certainly larger than just a few points. But this also requires us to characterize those sets that can be so neglected and also to describe what we must require of an indefinite integral so that these sets can be ignored.

At a calculus level we can easily go one step further, even if we cannot quite approach the full generality needed. The key is to push Lemma 1.2 and Lemma 1.3 much further. We cannot do this with help from the mean-value theorem as before: indeed the proof is deferred to the next chapter where we introduce a new technique needed for the integration theory.

LEMMA 1.5. Suppose that F and G are both continuous functions on an interval $[a, b]$. Suppose that there is a sequence of points e_1, e_2, e_3, ... of points from $[a, b]$ and that $F'(x) = G'(x)$ for all $a < x < b$ except possibly at the points e_1, e_2, e_3, ... Then

$$F(b) - F(a) = G(b) - G(a).$$

PROOF IN SECTION 7.4.3.

1.3. A version of the Newton integral for elementary calculus

Assuming Lemma 1.5 for the moment we can introduce a much improved version of the Newton integral.

DEFINITION 1.6 (Modified Newton Integral). Suppose that f is a function defined on an interval (a, b) except possibly at the points of a sequence e_1, e_2, e_3, ... from $[a, b]$. Suppose that we can find a continuous function $F : [a, b] \to \mathbb{R}$ so that $F'(x) = f(x)$ for every x with $a < x < b$ with perhaps the exception of the points e_1, e_2, e_3, ... Then we will say that F is *an indefinite integral* of f on $[a, b]$ and we will write

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

and call the latter *the definite integral* of f on $[a, b]$.

The only justification needed would be to use Lemma 1.5 to check that if F and G both qualify to be indefinite integrals of f on an interval $[a, b]$, then F and G differ by a constant so that $F(b) - F(a) = G(b) - G(a)$. Thus the definite integral is unambiguously defined.

There is one subtle point here that might be missed. Suppose that two functions F and G both qualify to be indefinite integrals of f. That means that there is some sequence of points e_1, e_2, e_3, ... from $[a, b]$ and that $F'(x) = f(x)$ provided x is not one of the points in this sequence. It also means that there is some sequence of points e'_1, e'_2, e'_3, ... from $[a, b]$ [not necessarily the same sequence as before] and that $G'(x) = f(x)$ provided x is not one of the points in this sequence.

Accordingly, we observe that $F'(x) = G'(x)$ except possibly at points belonging to the combined sequence

$$e_1, e'_1, e_2, e'_2, e_3, e'_3, e_4, e'_4, \dots$$

allowing us to apply Lemma 1.5. From that lemma we deduce that F and G differ by a constant and that $F(b) - F(a) = G(b) - G(a)$. Thus the definite integral is unambiguously defined no matter what indefinite integral we choose to use.

1.4. Integration formulas

The Newton integral inherits its formulas directly from the standard differentiation formulas. If we review the latter we will be able to deduce useful and attractive formulas for the integral.

For the integral of Chapter 3, which is much more general than the Newton versions here, these formulas remain true but will require some attention to hypotheses; they will not usually follow trivially from the differentiation formulas.

1.4.1. Sum formula. One of the first formulas we encounter in the calculus is that for the sum of two derivatives:

$$\frac{d}{dx}\left\{F(x)+G(x)\right\} = \frac{d}{dx}F(x) + \frac{d}{dx}G(x).$$

From that we obtain the sum formula for integrals:

$$\int_a^b \left\{f(x)+g(x)\right\}\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

The hypotheses allowing this are: f has an indefinite integral F and g has an indefinite integral G where both F and G are continuous on $[a,b]$ with

$$F'(x) = f(x) \quad\text{and}\quad G'(x) = g(x)$$

for all points in (a,b) excepting possibly some sequence of exceptional points. This sum formula will be available for the Chapter 3 integral under much weaker hypotheses.

1.4.2. Integration by parts. One of the most studied of the formulas we encounter in the calculus is that for the product of two derivatives:

$$\frac{d}{dx}\left\{F(x)G(x)\right\} = F'(x)G(x) + F(x)G'(x)$$

From that we obtain the formula for integrals known as *integration by parts:*

$$\int_a^b \left\{f(x)G(x)+F(x)g(x)\right\}\,dx = F(b)G(b) - F(a)G(b).$$

The hypotheses allowing this are: f has an indefinite integral F and g has an indefinite integral G where both F and G are continuous on $[a,b]$ with

$$F'(x) = f(x) \quad\text{and}\quad G'(x) = g(x)$$

for all points in (a,b) excepting possibly some sequence of exceptional points. There are versions of integration by parts formulas for the general integration theory, but they require very different proofs.

1.4.3. Integration by substitution. The formula for the derivative of the composition of two function is known sometimes as the chain rule:

$$\frac{d}{dx}F(G(x)) = F'(G(x))G'(x)$$

From that we obtain the formula for integrals known as *integration by substitution:*

$$\int_c^d f(s)\,ds = \int_a^b f(G(t))g(t)\,dt = F(G(b)) - F(G(a)).$$

The most transparent assumptions usually used to justify this are that G is a continuous, nondecreasing function on $[a,b]$ with $G(a) = c$ and $G(b) = d$ (so G maps the interval $[a,b]$ onto the interval $[c,d]$); $G'(x) = g(x)$ everywhere in (a,b); and F is a continuous function on $[c,d]$ with $F'(x) = f(x)$ at each point of (c,d).

Facility in using the formula computationally is a skill often drilled into students in an earlier course than this. For a theoretical course (as this is) we recognize the formula as only an application of the chain rule. There are more general versions available under weaker hypotheses for the integral of Chapter 3.

1.5. Preview

The Newton integral is a sufficient tool for most of elementary calculus needs; we should be informed of some of its theory. One of the defects in the presentation to this point is that we do not know what functions can be integrated by this method.

To be sure if a given function f is the derivative of some other function F then we know how the procedure works. But what sufficient conditions can be stated for a function f in order that we can be assured that such an indefinite integral exists? We cannot always be placed in the uncomfortable position of computing an indefinite integral in order to be assured that there is one.

We report here, by way of a preview, some of the theory that will clarify the situation. Proper statements of these facts appear in Chapter 3.

Theorem 1.7. In order that a function f possess an integral

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

on a compact interval $[a,b]$ in the sense of the Newton integral of this chapter the following are sufficient:

- f is continuous at every point of $[a,b]$.
- f is continuous at every point of (a,b) and is bounded.
- f is continuous at every point of (a,b) with the exception possibly of some sequence of points and f is bounded.
- f is continuous at every point of (a,b) with at most finitely many exceptions and dominated by another function g for which

$$0 \le f(x) \le g(x) \quad (a < x < b),$$

  where g is continuous on (a,b) (again allowing finitely many exceptions) and the integral $\int_a^b g(x)\,dx$ is assumed to exist.

In any of these cases we can be assured that f has an indefinite integral F, that

$$\int_c^d f(x)\,dx = F(d) - F(c) \quad (a \le c < d \le b)$$

and that

$$F'(x) = f(x)$$

at every point x at which f is continuous.

Example. One of the most important functions of elementary calculus with broad applications to probability, statistics, and physics is the function

$$F(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-x^2}\,dx.$$

(Interpret $\int_0^t$ as $-\int_t^0$ in the case where $t < 0$.)

Since the integrand $f(x) = e^{-x^2}$ is continuous at every point we know that this integral exists. We know too that $F'(x) = f(x)$ at every point. The student will be surprised to learn that there is no other formula for F that can be expressed in terms of the usual elementary functions of the calculus. In fact then this *is* the formula for F. We can use the Newton integral here because we have a theory that ensures us that there is an indefinite integral, even though we can not explicitly write it down.

Once we have this integral formula we can find other ways of using it. For example the series methods of the differential calculus and the series methods of Chapter 4 would allow us to show that

$$F(x) = \frac{2}{\sqrt{\pi}} \left(x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \frac{x^9}{216} - \dots \right).$$

1.6. The General Newton integral

The version of the Newton integral promoted in this chapter can be described this loose way:

We write

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

provided that we can determine a continuous function $F : [a, b] \to \mathbb{R}$ so that

$F'(x) = f(x)$ at all but a negligible set of points in (a, b).

By *negligible* here we mean the set of points X where the identity $F'(x) = f(x)$ might fail or simply be unknown can be written as a sequence. The justification, again in loose language is this:

a continuous function $F : [a, b] \to \mathbb{R}$ *cannot grow* on a negligible set of points e_1, e_2, e_3, ...

By the early twentieth century it was recognized that a larger class of negligible sets was needed for many problems. The class of sets that we might need to neglect are called the *null sets*. A null set is small in certain senses, but might be too large to be written out as a sequence of points. It turns out that while continuous functions do not grow on sequences of points, they might grow on null sets. Thus the definition of the Newton integral requires both a relaxation in the class of sets to be neglected and a tightening of the requirement on the function F. The definition of the modern version of the Newton integral is this:

> We write
> $$\int_a^b f(x)\,dx = F(b) - F(a)$$
> meaning that there is a continuous function $F : [a,b] \to \mathbb{R}$ so that:
> (a) There is a null set N,
> (b) $F'(x) = f(x)$ for all x in (a,b) except possibly for x in the null set N, and
> (c) this function F does not grow on the negligible set N.

This integral is properly defined as a Newton integral in Chapter 3. An equivalent constructive definition is given in Chapter 4. A measure-theoretic account is given in Chapter 5. This is the correct integral for the calculus.

CHAPTER 2

Covering Relations

The full definition of the integral will require the following notions:

- Covering relations.
- Full covers.
- Partitions of a compact interval $[a, b]$.
- Riemann sums.
- Cousin covering lemma.

We motivate these concepts in this chapter.

2.1. First mean-value theorem for integrals

The original Newton integral, the student will recall, requires of indefinite integrals that the derivative requirement holds at *every* point (no exceptional set is allowed). Let us return to that briefly.

How can we determine the value of a definite integral

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

for a function f? According to the definition we need to find an indefinite integral F first and then compute $F(b) - F(a)$. Finding an indefinite integral may be a much harder task than simply evaluating this single number $F(b) - F(a)$.

The mean value theorem for derivatives gives a hint. According to that theorem

$$F(b) - F(a) = f(\xi)(b - a)$$

for at least one point ξ in (a, b). That gives the identity

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

but we do not know precisely which point ξ to choose. This result is called the *first mean-value theorem* for the integral; we see it is available for the narrowest version of the Newton integral, the one where the indefinite integral F has the integrand f as its derivative at *every* point inside the interval.

This relation between an interval $[a, b]$ and some selected point ξ is called a *covering relation*. While the covering relation suggested by the first mean-value theorem for integrals is a useful one it cannot be made the basis for defining an integral.

2.2. Riemann sums

The identity

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

that we have just seen might be improved by subdividing the interval $[a, b]$ by intermediate points:

$$a = a_1 < b_1 = a_2 < b_2 = a_3 < \cdots < a_n < b_n = b.$$

This expresses the interval $[a, b]$ as the union of a collection of n nonoverlapping, compact subintervals:

$$[a_1, b_1], [a_2, b_2], [a_3, b_3], \ldots, [a_n, b_n].$$

The mean-value theorem of the differential calculus, as before, asserts that we can select a point ξ_i inside each interval $[a_i, b_i]$ so that

$$F(b_i) - F(a_i) = f(\xi_i)(b_i - a_i).$$

This leads to a new covering relation

$$\pi = \{([a_1, b_1], \xi_1), ([a_2, b_2], \xi_2), ([a_3, b_3], \xi_3), \ldots, ([a_n, b_n], \xi_n\}$$

which is called *a partition*. The partition is denoted as π (the letter is chosen so as to use the Greek letter corresponding to "P," not to have anything to do with areas of circles).

Using this partition π we have

$$F(b) - F(a) = \sum_{i=1}^{n} [F(b_i) - F(a_i)] = \sum_{i=1}^{n} f(\xi_i)(b_i - a_i)$$

and consequently

$$\int_a^b f(x)\, dx = \sum_{i=1}^{n} f(\xi_i)(b_i - a_i).$$

The sums here are called *Riemann sums for the function f* (or sometimes, more correctly, Cauchy sums) and there is a long history of using such sums and partitions of intervals to estimate integrals.

Again, however, this discussion does not move us much closer to finding a definition of the integral since we would be unable to choose the correct partition to use unless we already knew F and f.

2.3. Riemann sums constructed from the derivative

We have seen that if f is Newton integrable on an interval $[a, b]$ then for some partition of that interval

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

we have the identity

$$\int_a^b f(x)\, dx = \sum_{i=1}^{n} f(\xi_i)(b_i - a_i).$$

2.3.1. A full cover. The idea is a good one if we approach it differently. Let us assume that

$$F'(x) = f(x) \quad (a \le x \le b)$$

(which is a somewhat stronger assumption than we need for Newton's integral).

Let $\epsilon > 0$ and consider the collection β of all the pairs of intervals and points of the form

$$([c, d], \xi)$$

for which ξ is in $[c, d]$, with $[c, d] \subset [a, b]$, and for which

$$\left| \frac{F(d) - F(c)}{d - c} - f(\xi) \right| < \epsilon.$$

That means that every pair $([c, d], \xi)$ from β will satisfy

$$|F(d) - F(c) - f(\xi)(d - c)| < \epsilon(d - c).$$

This is a covering relation too (a collection of interval-point pairs) and we can check that it is quite a large collection. By the definition of the derivative we know that for every point ξ inside $[a, b]$ there is a $\delta > 0$ so that if

$$c \le \xi \le d$$

and

$$0 < d - c < \delta$$

then $([c, d], \xi)$ belongs to the collection β. This means that β is *full* in a sense we will make precise elsewhere: β contains all pairs $([c, d], \xi)$ for which the interval is sufficiently small.

2.3.2. Partitions and Riemann sums. Now suppose that β happens to contain a partition of the interval $[a, b]$. This would seem likely since β is a very large collection of pairs (I, x) with small I. Thus we assume there is a subset π of β,

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

so that the intervals $[a_i, b_i]$ are nonoverlapping and combine to form the whole interval $[a, b]$. In that case

$$\left| F(b) - F(a)) - \sum_{i=1}^{n} f(\xi_i)(b_i - a_i) \right| = \left| \sum_{i=1}^{n} [F(b_i) - F(a_i)] - \sum_{i=1}^{n} f(\xi_i)(b_i - a_i) \right|$$

$$= \left| \sum_{i=1}^{n} \{[F(b_i) - F(a_i)] - f(\xi_i)(b_i - a_i)\} \right| \le \sum_{i=1}^{n} |[F(b_i) - F(a_i)] - f(\xi_i)(b_i - a_i)|$$

$$< \sum_{i=1}^{n} \epsilon(b_i - a_i) = \epsilon(b - a).$$

2.3.3. Approximation to the integral by Riemann sums. That would mean that

$$\left| \int_a^b f(x)\, dx - \sum_{i=1}^{n} f(\xi_i)(b_i - a_i) \right| < \epsilon(b - a).$$

This computation shows that the integral can be *approximated* by Riemann sums, by a different method that does not rely on the mean-value theorem to select just the right points to use. To define an integral by this method will require that we utilize this notion of a *full* covering relation.

2.4. Full covers of a set

A covering relation is simply a collection β of interval-point pairs $([c,d],x)$ where $[c,d]$ is a compact interval and x some point belonging to $[c,d]$.

DEFINITION 2.1. A covering relation β is *full* at a point x_0 if there is $\delta > 0$ so that β contains all pairs $([c,d],x_0)$ for which $c \leq x_0 \leq d$ and $0 < d-c < \delta$.

DEFINITION 2.2. A covering relation β is *full cover* of a set E if β is full at each point x_0 belonging to the set E.

2.4.1. Mandatory Exercises. At the risk of sounding too severe, we instruct the reader to master each of the following exercises. All of the calculus applications of these ideas are based on them.

EXERCISE 10 (Continuity at a point). Let F be continuous at a point x_0, let $\epsilon > 0$, and write

$$\beta = \{([c,d],x_0) : c \leq x_0 \leq d \text{ and } |F(d) - F(c)| < \epsilon\}.$$

Show that β is full at x_0.

SOLUTION IN SECTION 8.2.1.

EXERCISE 11 (Continuity at a point). A smaller and more useful covering relation uses the notion of oscillation[1]. Let F be continuous at a point x_0, let $\epsilon > 0$, and write

$$\beta = \{([c,d],x_0) : c \leq x_0 \leq d \text{ and } \omega F([c,d]) < \epsilon\}.$$

Show that β is full at x_0.

SOLUTION IN SECTION 8.2.2.

EXERCISE 12 (Continuity at points in a set). Let F be continuous at each point of a set E, let $\epsilon > 0$, and write

$$\beta = \{([c,d],x) : c \leq x \leq d \text{ and } |F(d) - F(c)| < \epsilon\}.$$

Show that β is a full cover of E.

SOLUTION IN SECTION 8.2.3.

EXERCISE 13 (Derivative). Let F be differentiable at a point x_0 with $F'(x_0) = t$, let $\epsilon > 0$, and write

$$\beta = \{([c,d],x_0) : c \leq x_0 \leq d \text{ and } |F(d) - F(c) - t(d-c)| \leq \epsilon(d-c)\}.$$

Show that β is full at x_0.

SOLUTION IN SECTION 8.2.4.

EXERCISE 14 (Derivative at points in a set). Let F be differentiable at each point of a set E, let $\epsilon > 0$, and write

$$\beta = \{([c,d],x) : c \leq x \leq d,\ x \in E \text{ and } |F(d) - F(c) - F'(x)(d-c)| \leq \epsilon(d-c)\}.$$

Show that β is a full cover of E.

---

[1] Here $\omega F([c,d])$ denotes the *oscillation* of the function F on the interval $[c,d]$. We define it (loosely) as the greatest value obtained by $|F(s) - F(t)|$ for any choices of points s and t in the interval $[c,d]$.

SOLUTION IN SECTION 8.2.5.

EXERCISE 15 (Intersection of covers). Let β_1 and β_2 be full at a point x_0. Show that $\beta_1 \cap \beta_2$ is full at x_0.

EXERCISE 16 (Intersection of covers). Let β_1 and β_2 be full covers of a set E. Show that $\beta_1 \cap \beta_2$ is also a full cover of E.

EXERCISE 17 (Union of covers). Let β_1 be a full cover of a set E_1 and let β_2 be a full cover of a set E_2. Show that $\beta_1 \cup \beta_2$ is a full cover of $E_1 \cup E_2$.

2.5. Full covers and Cousin covers

We prefer to say merely that β is a *full cover* if β is a full cover of $\mathbb{R}$, i.e., if β is full at every real number. The entire theory of differentiation and integration of the calculus can be presented in a way that directly relates to full covers.

If a covering relation β is a full cover then we have expressed the opinion that it should contain a partition of any interval $[a,b]$, i.e., there should be a subset π of β,

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

so that the intervals

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

form a nonoverlapping collection of subintervals that make up all of $[a,b]$.

Note that, if our goal is to have partitions of $[a,b]$, we do not quite need β to be full at the endpoints a and b since we would use only subintervals of $[a,b]$ and not concern ourselves with what is happening on the left at a or what is happening on the right at b. This leads to the following definition, which slightly relaxes the condition defining full covers and also focusses on our need for partitions.

DEFINITION 2.3. A covering relation β is a *Cousin cover* of the compact interval $[a,b]$ provided that, at each point x in $[a,b]$, there is a $\delta > 0$ so that β contains all pairs $([c,d],x)$ for which $c \le x \le d$, $[c,d] \subset [a,b]$ and $0 < d - c < \delta$.

2.5.1. Exercises.

EXERCISE 18. Let β be a Cousin cover of a compact interval $[a,b]$. Show that β is a Cousin cover of any compact subinterval $[c,d] \subset [a,b]$.

EXERCISE 19. Let β be a full cover of $[a,b]$. Show that β is a Cousin cover of $[a,b]$. Is the converse true?

EXERCISE 20. Let β_1 and β_2 be Cousin covers of a compact interval $[a,b]$. Show that $\beta_1 \cap \beta_2$ is Cousin cover of $[a,b]$.

EXERCISE 21. Let $[a,b]$ be a compact interval and let β_1 be a full cover of the open interval (a,b). Write, for any $\delta > 0$,

$$\beta_2 = \{([a,d],a) : a < d < a + \delta\}$$

and

$$\beta_3 = \{([c,b],b) : b - \delta < c < b\}$$

Show that $\beta_1 \cup \beta_2 \cup \beta_3$ is a Cousin cover of $[a,b]$.

Convention: Any full cover is necessarily also a Cousin cover of any compact interval $[a,b]$. Conversely if the focus of attention is on a particular compact interval and β is a Cousin cover of $[a,b]$ then it is simple matter to enlarge β to be a full cover by adding in a covering relation that is irrelevant. Thus in all definitions using a Cousin cover it may be convenient to simply use full covers. The reader will then forgive the writer if, when defining a full cover, he neglects to add in enough irrelevant interval-point pairs to make the cover a true full cover.

Differentiation bases: The collection of all full covers is a filtering family that serves both as a basis for differentiation and a basis for integration. A formal presentation of these ideas at an advanced level may make the flow of the argument transparent. For elementary courses any reference to the "filter of full covers" would only obscure the discussion.

2.6. Cousin covering lemma

LEMMA 2.4 (Cousin). Let β be a full cover. Then β contains a partition of every compact interval $[a,b]$, i.e., there is a subset π of β,

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

so that the intervals $[a_i,b_i]$ are nonoverlapping and combine to form the whole interval $[a,b]$.

PROOF IN SECTION 7.5.1.

COROLLARY 2.5. Let β be a Cousin cover of a compact interval $[a,b]$. Then β contains a partition of every compact subinterval of $[a,b]$.

PROOF IN SECTION 7.5.2.

2.7. An application of the Cousin lemma

The Cousin lemma offers us a technique that can be used to prove Lemma 1.5 that we just skipped over. This lemma follows from the following statement which we now prove, using the Cousin covering lemma:

> *Suppose that F is a continuous function. Suppose that there is a sequence of points e_1, e_2, e_3, $\dots$ of points and that $F'(x) = 0$ for all x except possibly at the points e_1, e_2, e_3, $\dots$. Then F is constant.*

The next chapter is devoted to an elaborate study of this lemma, in a more general exposition. It is, nonetheless, worth working through the details here as a preliminary to the studying the collection of definitions and techniques to be used there.

Proof. Fix an interval $[a,b]$. The proof is obtained by showing, for any $\epsilon > 0$, that

$$|F(b) - F(a)| < \epsilon.$$

This can only be true for all $\epsilon > 0$ if $F(b) = F(a)$. This shows that F is in fact constant.

We use a covering argument. We need to construct a full cover β using a different construction at the points where $F'(x) = 0$ and at the points e_1, e_2, $\dots$ where the derivative need not be zero. There are infinitely many steps so this value of ϵ will need to be split into infinitely many small pieces by the following simple identity:

$$\epsilon = \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{8} + \frac{\epsilon}{16} + \dots .$$

The first step in the covering argument takes advantage of the fact that

$$F'(x) = 0$$

at points other than $e_1, e_2, \dots$. Write

$$\eta = \frac{\epsilon}{2(b-a)}$$

and

$$\beta_1 = \{([c,d],x) : \ a < c \le x \le d < b \text{ and } |F(d) - F(c)| < \eta(d-c)\}.$$

This takes care of all points where we know that the derivative is zero. To handle the remaining points, use the continuity of F and form a collection β_2 consisting of all of the following pairs:

$$\{([c,d],e_i) : \ c \le e_i \le d \text{ and } |F(d) - F(c)| < \epsilon 2^{-1-i} \text{ for } i = 1,2,3,\dots \ \}.$$

It is easy to check that $\beta = \beta_1 \cup \beta_2$ is a full cover. Now to complete the proof is merely a computation using partitions from β of $[a,b]$. We know by the Cousin covering lemma that there is at least one partition π of $[a,b]$ that is a subset of β. Write

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

Then

$$|F(b) - F(a)| = \left|\sum_{i=1}^{n}[F(b_i) - F(a_i)]\right| \le \sum_{i=1}^{n}|F(b_i) - F(a_i)| = \sum_1 + \sum_2 .$$

Here we have split the sum into two parts, the first part chosen over members of the partition from β_1 and the remainder chosen over members of the partition from β_2. For the first part of the sum we check that

$$\sum_1 = \sum_1 |F(b_i) - F(a_i)| < \sum_{i=1}^{n}\eta(b_i - a_i) \le \frac{\epsilon(b-a)}{2(b-a)} = \frac{\epsilon}{2}.$$

For the second part of the sum we check that

$$\sum_2 = \sum_2 |F(b_i) - F(a_i)| < \frac{\epsilon}{4} + \frac{\epsilon}{8} + \frac{\epsilon}{16} + \dots = \frac{\epsilon}{2}.$$

Putting these together gives us that

$$|F(b) - F(a)| < \epsilon$$

as we needed to prove.

CHAPTER 3

Growth of a Function on a Set

The calculus is concerned to a great extent with the growth of a function either *at a point* or *on a set*. The derivative measures the growth of a function at a point. A different concept is needed for the growth of a function on a set.

For a monotonic function F, growth on an interval $[a, b]$ can be simply measured as

$$|F(b) - F(a)|,$$

the absolute value of the difference of the end value and the beginning value. For a nonmonotonic function we should be interested in the all the ups and downs, not just the beginning and end values.

A measurement of the sums

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

taken over nonoverlapping subintervals would be appropriate. This notion appears in the early literature and was formalized by Jordan in the late 19th century under the terminology "variation of a function."

For a first theoretical course in calculus we do not need (at first) the actual measurement of growth. What we do need is the notion that a function has *zero growth* or *fails to grow on a set*.

This leads to the following notions, explored in this chapter:

- A function F *does not grow* on a set E.
- A set N is a *null set* if the identity function $F(x) = x$ does not grow on N.
- A function F is *continuous at a point* x_0 if F does not grow on the singleton set $\{x_0\}$.
- A function F is *continuous* if F does not grow on any singleton set $\{x_0\}$.
- A function is *absolutely continuous*[1] if it does not grow on any null set.
- A function F with a zero derivative $F'(x) = 0$ at every point of a set E does not grow on E.

The first five of these are definitions. The last one reveals the connection between zero derivatives and zero growth[2].

[1]This notion of absolute continuity is more general than that commonly used in elementary analysis courses: see the the stronger Vitali version below.

[2]There is a converse but it lies deeper than the material in this chapter: if F does not grow on a set E then $F'(x) = 0$ at every point x in E except possibly for some null set.

3.1. Growth of a function on a set

The "change" of a function F on an interval $[a,b]$ is given by the increment

$$\Delta F([a,b]) = F(b) - F(a).$$

What we wish to express is the variability of the function on a set E.

DEFINITION 3.1. A function $F : \mathbb{R} \to \mathbb{R}$ is said *not to grow* on a set E if for every $\epsilon > 0$ there can be found a full cover β of that set E so that

$$\sum_{i=1}^{n} |\Delta F([a_i, b_i])| < \epsilon$$

whenever the subpartition

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

is chosen from β.

Recall that in order for the subset γ to be a subpartition, we require merely that the intervals $\{[a_i, b_i]\}$ do not overlap. The collection γ here is not necessarily a partition (although it may be) and so we cannot use the letter "π." It is what we have called a *subpartition* since it could be (but won't be) expanded to be a partition.

3.1.1. Exercises.

EXERCISE 22. Suppose that $F : \mathbb{R} \to \mathbb{R}$ does not grow on a set E_1 and that $E_2 \subset E_1$. Show that then F does not grow on E_2.

EXERCISE 23. Suppose that $F : \mathbb{R} \to \mathbb{R}$ does not grow on the sets E_1 and E_2. Show that then F does not grow on the union $E_1 \cup E_2$.

EXERCISE 24. Suppose that $F : \mathbb{R} \to \mathbb{R}$ does not grow on each member of a sequence of sets E_1, E_2, E_3, ... Show that then F does not grow on the union $\bigcup_{n=1}^{\infty} E_n$.

SOLUTION IN SECTION 8.3.10.

3.2. Fundamental growth lemma on intervals

The fundamental growth theorem that we need shows that only constant functions can fail to grow on an interval.

THEOREM 3.2. *Suppose that a function $F : \mathbb{R} \to \mathbb{R}$ does not grow on an interval (open or closed). Then F is constant on that interval.*

3.3. Zero derivatives imply zero growth

There is an immediate connection between the derivative and growth on a set. In the simplest case we see that a function cannot grow on a set on which it has everywhere a zero derivative.

THEOREM 3.3. *Suppose that a function $F : \mathbb{R} \to \mathbb{R}$ has a zero derivative $F'(x)$ at every point x of a set E. Then F does not grow on E.*

3.4. Null or negligible sets

We define a null set to be those sets on which the simplest function $F(x) = x$ does not grow.[3] This can be given its own definition. If $F(x)$ then

$$|F(b_i) - F(a_i)| = (b_i - a_i)$$

so that the failure of F to grow on a set E can be described simply by using the sums

$$\sum_{i=1}^{n}(b_i - a_i)$$

taken over a subpartition.

DEFINITION 3.4. A set E is said to be a *negligible set* (or a *null set*) if for every $\epsilon > 0$ there can be found a full cover β of that set E so that

$$\sum_{i=1}^{n}(b_i - a_i) < \epsilon$$

whenever the collection

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

is a subpartition chosen from β.

LEMMA 3.5. Every finite set is null.

PROOF IN SECTION 7.6.1.

LEMMA 3.6. Every set E whose elements can be written out as a sequence of points e_1, e_2, e_3, ... is null.

PROOF IN SECTION 7.6.2.

LEMMA 3.7. No open interval (a, b) is null.

PROOF IN SECTION 7.6.3.

3.4.1. Exercises.

EXERCISE 25. Show that a set E is a null set if and only if the function $F(x) = x$ does not grow on E.

SOLUTION IN SECTION 8.3.4.

EXERCISE 26. Show that a Lipschitz function cannot grow on a null set.

SOLUTION IN SECTION 8.3.7.

EXERCISE 27. Suppose that $F : \mathbb{R} \to \mathbb{R}$ is differentiable at each point of a set E. Show that then F cannot grow on any null subset of E.

SOLUTION IN SECTION 8.3.11.

EXERCISE 28. Show that every subset of a null set is a null set.

SOLUTION IN SECTION 8.3.1.

[3] The fact that the function $F(x) = x$ does not grow on a set E will lead to the fact, in turn, that *most* functions of the calculus cannot grow on E.

EXERCISE 29. Show that the union of two sets null sets is again a null set.

SOLUTION IN SECTION 8.3.2.

EXERCISE 30. Show that the union of a sequence of null sets is again a null set.

SOLUTION IN SECTION 8.3.3.

3.5. Classification of functions by growth properties

3.5.1. Continuous at a point.

DEFINITION 3.8. A function $F : \mathbb{R} \to \mathbb{R}$ is continuous at a point x_0 if F does not grow on the set $E = \{x_0\}$.

EXERCISE 31. Show that a function $F : \mathbb{R} \to \mathbb{R}$ is continuous at a point x_0 if and only if for every $\epsilon > 0$ there is a $\delta > 0$ so that $|F(x_0) - F(x)| < \epsilon$ whenever $|x - x_0| < \delta$.

3.5.2. Continuous function. A function is continuous if it is continuous at every point.

DEFINITION 3.9. A function $F : \mathbb{R} \to \mathbb{R}$ is continuous if F does not grow on any singleton set $E = \{x_0\}$.

EXERCISE 32. Show that a function is continuous if and only if it does not grow on any finite set.

SOLUTION IN SECTION 8.3.5.

EXERCISE 33. Show that a function is continuous if and only if it does not grow on any set E whose elements can be written as a sequence $E = \{e_1, e_2, e_3, \dots\}$.

SOLUTION IN SECTION ??.

3.5.3. Continuous in Cauchy's uniform sense. According to Exercise 31 a function $F : \mathbb{R} \to \mathbb{R}$ is continuous at a point x_0 if and only if for every $\epsilon > 0$ there is a $\delta > 0$ so that $|F(x_0) - F(x)| < \epsilon$ whenever $|x - x_0| < \delta$. Cauchy introduced this notion without noticing that the choice of δ could well depend on the point x_0 and not merely on the ϵ. That mistaken notion is now used as a definition.

DEFINITION 3.10. A function $F : [a, b] \to \mathbb{R}$ is uniformly continuous [in Cauchy's sense] on $[a, b]$ provided that for every $\epsilon > 0$ there is a $\delta > 0$ so that $|F(x) - F(y)| < \epsilon$ whenever x and y are points in $[a, b]$ for which $|x - y| < \delta$.

It is transparent that should a function F satisfy this condition then F must be continuous at each point in (a, b). (We cannot say much about continuity at a and b since the points x and y in the definition are chosen from inside $[a, b]$.)

The converse is significant.

THEOREM 3.11. *Suppose that the function $F : \mathbb{R} \to \mathbb{R}$ is continuous. Then F is uniformly continuous [in Cauchy's sense] on any compact interval $[a, b]$.*

PROOF IN SECTION 7.6.4.

3.6. Absolutely continuous function

DEFINITION 3.12. A function $F : \mathbb{R} \to \mathbb{R}$ is absolutely continuous if F does not grow on any null set.

The following are either immediate or easy to prove:

- The function $F(x) = x$ is absolutely continuous.
- Every Lipschitz function is absolutely continuous.
- Every differentiable function is absolutely continuous.
- A linear combination of absolutely continuous functions is absolutely continuous.
- Every absolutely continuous function is continuous.

Thus there is an abundance of absolutely continuous functions (indeed nearly all the functions of the calculus).

LEMMA 3.13. Suppose that the function $F : \mathbb{R} \to \mathbb{R}$ has a finite derivative at every point. Then F is absolutely continuous.

3.6.1. Exercises.

EXERCISE 34. Show that every absolutely continuous function is continuous.

EXERCISE 35. Suppose that a continuous function $F : \mathbb{R} \to \mathbb{R}$ has a finite derivative at every point with the exception of some sequence of points e_1, e_2, e_3, ... Show that F is absolutely continuous.

EXERCISE 36. Show that every Lipschitz function is absolutely continuous.

EXERCISE 37. Show that a linear combination of absolutely continuous functions is absolutely continuous.

3.7. Absolutely continuous in Vitali's sense

The uniform continuity criterion of Cauchy, the ϵ–δ condition just given, has a parallel for absolute continuity.

DEFINITION 3.14. A function $F : [a,b] \to \mathbb{R}$ is absolutely continuous in Vitali's sense on $[a,b]$ provided that for every $\epsilon > 0$ there is a $\delta > 0$ so that

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

whenever $\{[x_i, y_i]\}$ are nonoverlapping subintervals of $[a,b]$ for which

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

It is transparent that absolutely continuity in Vitali's sense easily implies uniform continuity [in Cauchy's sense]. The lemma we now state shows that the Vitali condition is also stronger than absolute continuity in our sense. (It is in fact strictly stronger: there are absolutely continuous functions that are not absolutely continuous in Vitali's sense.)

LEMMA 3.15. Suppose that the function $F : [a,b] \to \mathbb{R}$ is absolutely continuous in Vitali's sense on the interval $[a,b]$. Then F cannot grow on a null subset of (a,b).

SOLUTION IN SECTION 8.3.9.

3.7.1. Exercises.

EXERCISE 38. Show that every Lipschitz function is absolutely continuous in Vitali's sense.

SOLUTION IN SECTION 8.3.8.

EXERCISE 39. Give an example of an absolutely continuous function that is not Lipschitz.

EXERCISE 40. Show that there is an absolutely continuous function F that is not absolutely continuous according to Vitali's definition on the interval $[0, 1]$.

CHAPTER 4

The General Newton Integral

We have seen a preview of the definition of the correct integral on the real line in Section 1.6. Now we can clarify that vague discussion by exploiting the notion of null sets and the concept of a function not growing on a set.

4.1. Modern theory

Here are the ingredients of the modern theory:

- *Negligible sets* (or *null sets* as they are also known) are the sets that can be ignored for defining the indefinite integral.
- The indefinite integral F must be chosen in such a way that the function F *does not grow* on the exceptional set.

The modern integral can be described now this way.

DEFINITION 4.1. Suppose that f is a function defined on an interval (a, b) excepting possibly at points belonging to some negligible set. Suppose that we can find a continuous function $F : \mathbb{R} \to \mathbb{R}$ and a negligible set N so that

(a) $F'(x) = f(x)$ for every $a < x < b$ with the exception possibly of points x in the set N.
(b) The function F does not grow on the set N.

Then we would say that F is *an indefinite integral* of f on $[a, b]$ and write

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

Clearly this integral includes all of the weaker versions of the Newton integral discussed in Chapter 1.

4.1.1. Exercises.

EXERCISE 41. Show that if a function F is an indefinite integral of a function f on $[a, b]$ with

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

then F cannot grow on any null subset of (a, b).

EXERCISE 42. Show that a function $F : \mathbb{R} \to \mathbb{R}$ is an indefinite integral of a function f on $[a, b]$ with

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

if the following two conditions hold:

(a) $F'(x) = f(x)$ for each x in (a, b) with the exception perhaps of points x in some null set N, and
(b) the function F is absolutely continuous.

The special case where item (b) is replaced by absolute continuity in Vitali's sense on the interval $[a, b]$ gives rise to the narrower integral known as the *Lebesgue integral*. A further restriction to the much smaller class of Lipschitz functions would reduce the scope further and introduce an integral that copes only with bounded functions (but which in all other respects is sufficiently general).

4.2. Key constancy lemma

The key lemma that allows the descriptive definition of the integral that extends Newton's original idea to allow negligible sets is this. The proof is easy, following from the material in the preceding chapter.

LEMMA 4.2. Let $F : \mathbb{R} \to \mathbb{R}$ be a continuous function and let N be a negligible set. Suppose that $F'(x) = 0$ for each x in the interval (a, b) except possibly at points in N and suppose that F does not grow on N. Then F is constant on the interval $[a, b]$.

PROOF IN SECTION 7.6.5.

COROLLARY 4.3. Let $f : [a, b] \to \mathbb{R}$. Let F, $G : \mathbb{R} \to \mathbb{R}$ be continuous functions and let N_1 and N_2 be a negligible sets. Suppose that $F'(x) = f(x)$ for each x in the interval (a, b) except possibly at points in N_1 and suppose that F does not grow on N_1. Suppose that $G'(x) = f(x)$ for each x in the interval (a, b) except possibly at points in N_2 and suppose that G does not grow on N_2. Then $F - G$ is constant on the interval $[a, b]$.

PROOF IN SECTION 7.6.6.

CHAPTER 5

The Integral

5.1. Towards a definition of the integral

A formal constructive definition of the Newton integral used in the calculus is possible, using the following simple concepts:

- Cousin cover of a compact interval $[a, b]$.
- Partition of a compact interval $[a, b]$.
- Riemann sum over a partition.

We introduce a new word, *integrable*, that applies to those functions for which a certain procedure applies. To date we have merely *described* integrals by using the relation

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

available when an indefinite integral F happens to be available.

5.2. Formal constructive definition of the integral

Note that the function f being integrated in the constructive[1] definition that we now give must be defined at *all* points of the interval $[a, b]$ in order for this definition to be interpreted. In our descriptive theory we could ignore a negligible set. Here it will be a theorem (not part of the definition) that a negligible set can be ignored.

DEFINITION 5.1. Suppose that f is a function defined everywhere on an interval $[a, b]$. Then f is said to be *integrable* on $[a, b]$ provided that there is a real number c with the following property: for every $\epsilon > 0$ it is possible to select a Cousin cover[2] β of the interval $[a, b]$ in such a way that

$$\left| \sum_{i=1}^{n} f(\xi_i)(b_i - a_i) - c \right| < \epsilon$$

for every partition

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

of the interval $[a, b]$ chosen from β. In that case we will write

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

[1]While we call our definition *the constructive definition*, that is really a misnomer since, in full generality, there is no general method of using properties of the function f to construct the Cousin covers β that appear in the definition.

[2]The definition is unchanged if "Cousin cover of $[a, b]$" is replaced merely by "full cover." That offers a formal simplification, although in practise all arguments would be focused on the particular interval $[a, b]$ and a Cousin cover effectively used.

and call this *the definite integral* of f on $[a,b]$.

5.3. Relation with Newton's integral

Our very first task must be to establish that this integral includes all of the methods we have been using so far. Note that the first three conditions are contained in the fourth condition which describes the most general Newton integral.

THEOREM 5.2. Let F and f be functions defined on a compact interval $[a,b]$. Suppose that F is continuous on $[a,b]$ and that one of the four following conditions holds:

(a) $F'(x) = f(x)$ for every $a < x < b$.
(b) $F'(x) = f(x)$ for every $a < x < b$ with finitely many exceptions.
(c) There is a sequence of points $e_1, e_2, e_3, \ldots$ from the interval $[a,b]$ and $F'(x) = f(x)$ at every point of (a,b) with possibly the exception of the points $e_1, e_2, e_3, \ldots$.
(d) $F'(x) = f(x)$ for every $a < x < b$ excepting possibly some null subset N of (a,b) and the function F does not grow on N.

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

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

PROOF IN SECTION 7.7.1.

The integral includes the Newton integral (as this theorem shows). In fact the two integrals are equivalent, although that will take much more advanced material to establish and will not appear until Chapter 5.

5.4. Ignoring a sequence of points

In our earlier definition of Newton's integral

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

of a function f we paid no attention to the values $f(a)$ and $f(b)$, and indeed in many cases the function f had no such values. For our modified version of Newton's integral we ignored a sequence of points at which the function need not be defined nor the derivative checked.

The definition of the constructive integral seems to demand that the function be defined for *every* value in $[a,b]$ in order that the integral

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

assume a value. Here we show that the values can be freely altered at any particular sequence of points.

LEMMA 5.3. Let $e_1, e_2, e_3, \ldots$ be a sequence of points from the interval $[a,b]$ and suppose that f is a function defined on $[a,b]$ and for which $f(x) = 0$ for every point x not belonging to this sequence. Then f is integrable on $[a,b]$ and

$$\int_a^b f(x)\,dx = 0.$$

PROOF IN SECTION 7.7.5.

LEMMA 5.4. Let e_1, e_2, e_3, ... be a sequence of points from the interval $[a, b]$ and suppose that f and g are functions defined on $[a, b]$ and for which $f(x) = g(x)$ for every point x not belonging to this sequence. Then f is integrable on $[a, b]$ if and only if g is integrable on $[a, b]$, and if one of the functions is integrable then

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

PROOF IN SECTION 7.7.6.

5.5. The convention for ignoring sequences of points

Let e_1, e_2, e_3, ... be a sequence of points from the interval $[a, b]$ and f a function defined at every point of $[a, b]$ except possibly at the points of the sequence. Then f is said to be integrable if the function g defined by $g(x) = f(x)$ wherever this is defined and $g(x) = 0$ elsewhere, is integrable on $[a, b]$. The integral of f is interpreted as

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

In effect, then, if some sequence of points is troublesome for the integration computation or argument that you wish to use, just change the values at those points to something more convenient. The most popular choice is to set $f(x) = 0$ at as many points in some sequence as you wish.

EXERCISE 43. Let f be the function defined to be 1 at every irrational number and as 0 at every rational number. Is f integrable on any interval and, if so, what is $\int_a^b f(x)\,dx$?

SOLUTION IN SECTION 8.4.1.

5.6. Ignoring null sets

In Section 5.4 we saw how a sequence of points can be ignored for the purposes of integration theory. Here we see that the situation extends to the class of negligible (null) sets.

LEMMA 5.5. Let N be a negligible subset of the interval $[a, b]$ and suppose that f is a function defined on $[a, b]$ and for which $f(x) = 0$ for every point x not belonging to N. Then f is integrable on $[a, b]$ and

$$\int_a^b f(x)\,dx = 0.$$

SOLUTION IN SECTION 7.7.7.

LEMMA 5.6. Let N be a negligible subset of the interval $[a, b]$ and suppose that f and g are functions defined on $[a, b]$ and for which $f(x) = g(x)$ for every point x not belonging to N. Then f is integrable on $[a, b]$ if and only if g is integrable on $[a, b]$, and if one of the functions is integrable then

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

SOLUTION IN SECTION 7.7.8.

This allows our earlier convention to be broadly generalized: change the value of any integrated function freely on a negligible set. No integration statement or integration formula will change.

5.7. Order

THEOREM 5.7. Let f and g be functions integrable on a compact interval $[a, b]$ and suppose that

$$f(x) \leq g(x)$$

for every $a \leq x \leq b$. Then

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

PROOF IN SECTION 7.7.9.

5.8. Linear combinations

THEOREM 5.8. Let f and g be functions integrable on a compact interval $[a, b]$ and let r and s be real numbers. Then the function $rf + sg$ is also integrable on $[a, b]$ and

$$\int_a^b [rf(x) + sg(x)]\,dx = r\int_a^b f(x)\,dx + s\int_a^b g(x)\,dx.$$

PROOF IN SECTION 7.7.10.

5.9. Additivity

THEOREM 5.9. Let f be a function that is integrable on two adjacent intervals $[a, c]$ and $[c, b]$ with $a < c < b$. Then f is integrable on $[a, b]$ and

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

PROOF IN SECTION 7.7.11.

5.10. Comparing Riemann sums

Let us compare two different Riemann sums

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

and

$$\sum_{j=1}^{m} f(\xi_i')(b_i' - a_i')$$

taken over two different partitions of a compact interval $[a, b]$:

$$\pi_1 = \{([a_i, b_i], \xi_i) : i = 1, 2, \dots, n\}$$

and

$$\pi_2 = \{([a_j', b_j'], \xi_j') : j = 1, 2, \dots, m\}.$$

At first sight these seem rather difficult to subtract, but look more closely.

Some rewriting will help in the comparison. First of all use

$$\mathcal{L}([a,b]) = b - a$$

to denote the *length* of the interval $[a,b]$. We can even allow $a = b$ so that $[a,a]$ is a degenerate interval of length $\mathcal{L}([a,a]) = 0$. The length of the empty set $\mathcal{L}(\emptyset)$ should also be zero.

Then the two Riemann sums assume the form

$$\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i])$$

and

$$\sum_{j=1}^{m} f(\xi'_j)\mathcal{L}([a'_j,b'_j]).$$

Note that if $[c,d]$ is any subinterval of $[a,b]$ the collection

$$\pi_3 = \{([a_i,b_i] \cap [c,d], \xi_i) : i = 1,2,\dots,n\}$$

is really a partition of $[c,d]$ provided that we can ignore the odd bits where

$$[a_i,b_i] \cap [c,d]$$

is degenerate or empty. In particular it is easy to compute that

$$\mathcal{L}([c,d]) = \sum_{i=1}^{n} \mathcal{L}([a_i,b_i] \cap [c,d]).$$

Thus we arrive at the odd-looking expression for the Riemann sums above:

$$\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) = \sum_{i=1}^{n}\sum_{j=1}^{m} f(\xi_i)\mathcal{L}([a_i,b_i] \cap [a'_j,b'_j])$$

and

$$\sum_{j=1}^{m} f(\xi'_j)\mathcal{L}([a'_j,b'_j]) = \sum_{i=1}^{n}\sum_{j=1}^{m} f(\xi'_j)\mathcal{L}([a_i,b_i] \cap [a'_j,b'_j]).$$

Subtracting these two sums is much easier. This gives us our little computational lemma which we will use as the basis for a characterization of the integral.

Lemma 5.10. Let f be a function defined on a compact interval $[a,b]$ and let π_1 and π_2 be two partitions of the interval $[a,b]$:

$$\pi_1 = \{([a_i,b_i], \xi_i) : i = 1,2,\dots,n\}$$

and

$$\pi_2 = \{([a'_j,b'_j], \xi'_j) : j = 1,2,\dots,m\}.$$

Then

$$\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) - \sum_{j=1}^{m} f(\xi'_j)\mathcal{L}([a'_j,b_j]') = \sum_{i=1}^{n}\sum_{j=1}^{m} [f(\xi_i) - f(\xi'_j)]\mathcal{L}([a_i,b_i] \cap [a'_j,b'_j]).$$

5.11. Necessary and sufficient condition for integrability

Theorem 5.11 (Cauchy criterion). Let f be a function defined on a compact interval $[a,b]$. Then f is integrable on $[a,b]$ if and only if for every $\epsilon > 0$ there exists a Cousin cover β of $[a,b]$ such that

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{m}[f(\xi_i)-f(\xi_j')]\mathcal{L}([a_i,b_i]\cap[a_j',b_j'])\right| < \epsilon$$

for every pair of partitions π_1 and π_2 of the interval $[a,b]$ chosen from β:

$$\pi_1 = \{([a_i,b_i],\xi_i) : i = 1,2,\dots,n\}$$

and

$$\pi_2 = \{([a_j',b_j'],\xi_j') : j = 1,2,\dots,m\}.$$

proof in section 7.7.12.

5.12. Integrating continuous functions

We can integrate most of the functions of the elementary calculus by recognizing them as derivatives. All polynomials are derivatives and so integrable on any interval. The trigonometric functions $\sin x$ and $\cos x$ also arise in differentiation formulas, so they too are integrable.

For theoretical reasons we need to know some sufficient properties a function might have that imply integrability. A famous and useful result is that continuous functions are integrable.

Theorem 5.12. Let f be a function defined on an interval $[a,b]$ and let e_1, e_2, e_3, … be a sequence of points of $[a,b]$. Suppose that f is bounded on $[a,b]$ and continuous at each point with the exception possibly of the points e_1, e_2, e_3, … in the sequence. Then f is integrable.

proof in section 7.7.15.

Corollary 5.13. Let f be a function continuous at each point of an open interval (a,b). Then if f is bounded on (a,b), f is integrable on $[a,b]$.

Corollary 5.14. Let f be a function continuous at each point of a compact interval $[a,b]$. Then f is integrable on $[a,b]$.

proof in section 7.7.14.

5.13. Functions continuous almost everywhere

A function is said to be *continuous almost everywhere* if it is continuous at every point with the exception of a null set. We now prove the following generalization of Theorem 5.12 which, we recall, asserts that functions that are bounded and continuous everywhere except at some sequence of points is integrable. Here we show that the exceptional set can be much larger.

Theorem 5.15. Let f be a bounded function defined on an interval $[a,b]$ that is continuous at each point with the exception possibly of a null set. Then f is integrable.

solution in section 7.7.16.

5.14. The indefinite integral

The essential ingredient in the Newton definition of the integral of a function f on a compact interval $[a,b]$ is that there must be an indefinite integral, i.e., a function F defined on $[a,b]$ for which we can compute

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

for all $a < t \leq b$ and

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

for all $a \leq c < d \leq b$.

The constructive definition of the integral does not seem to promise an indefinite integral, just simply a single number

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

which gives the value of the definite integral on $[a,b]$.

It is a highly important feature of the integral that we do indeed have an indefinite integral and that indefinite integral is continuous. (We cannot, however, be yet assured that $F'(x) = f(x)$ at most points as would be the case for Newton's integral.)

THEOREM 5.16. Let f be integrable on a compact interval $[a,b]$. Then f is integrable on every compact subinterval $[c,d]$ of $[a,b]$ and there is a continuous function F defined on $[a,b]$ for which

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

for all $a \leq c < d \leq b$.

PROOF IN SECTION 7.7.17.

In general we need not expect a "formula" for an indefinite integral unless we have a particularly simple formula already for f. But in writing about that indefinite integral we can always simply use

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

for $a < t \leq b$ and $F(a) = 0$. Some users of the calculus do not hesitate to use the formula

$$\int_a^a f(x)\,dx = 0$$

so that

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

will have meaning for all $a \leq t \leq b$.

5.14.1. Henstock's criterion.

THEOREM 5.17 (Henstock's Criterion). The following statement is a necessary and sufficient condition in order for a function F to be an indefinite integral for a function f on an interval $[a,b]$: for every $\epsilon > 0$ there is a Cousin cover β of $[a,b]$ such that

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

for all partitions

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

of $[a,b]$ contained in β.

PROOF IN SECTION 7.7.18.

5.14.2. Indefinite integrals. If a function f is integrable on an interval $[a,b]$ then it has an indefinite integral F (many indefinite integrals in fact) and the relation between f and F is expressed by the Henstock criterion which is merely equivalent to the relation

$$F(d) - F(c) = \int_c^d f(x)\,dx \qquad (a \le c < d \le b).$$

What other properties must F have?

THEOREM 5.18. If $F : [a,b] \to \mathbb{R}$ is an indefinite integral of a function f on an interval $[a,b]$ then the function F does not grow on null sets.

This theorem asserts that the indefinite integral F is always absolutely continuous in the general sense. In particular F is also continuous. It need not be Lipschitz (unless f is bounded) or absolutely continuous [in Vitali's sense] unless f is absolutely integrable.

PROOF IN SECTION 7.7.19.

COROLLARY 5.19. If $F : [a,b] \to \mathbb{R}$ is an indefinite integral of f on an interval $[a,b]$ then the function F is both continuous and absolutely continuous [general sense] on $[a,b]$.

5.15. Differentiating the integral

If f is integrable on an interval $[a,b]$ in Newton's sense then we know from that descriptive definition of the integral that it has an indefinite integral F. This function F is continuous on $[a,b]$,

$$F(t) = \int_a^t f(x)\,dx + \text{some constant}$$

for every $a < t \le b$, and

$$F'(x) = f(x)$$

at every point of (a,b) with the exception possibly of points in some given sequence of points.

In particular at any given point x_0 in (a,b) we cannot be sure that

$$F'(x_0) = f(x_0).$$

Most points have this property but the point x_0 chosen may not.

For the constructive definition of the integral we are not yet sure that $F'(x_0) = f(x_0)$ would have to be true at *any* point. The following theorem is useful in calculus courses for providing a simple situation in which we will know what the derivative is.

THEOREM 5.20. Let f be integrable on $[a, b]$ and let F be an indefinite integral of f. If $a < x_0 < b$ and x_0 is a point of continuity of the function f then

$$F'(x_0) = f(x_0).$$

PROOF IN SECTION 7.7.20.

5.16. A comparison test

THEOREM 5.21. Let f and g be functions, continuous at each point of an open interval (a, b) with finitely many exceptions. Suppose that

$$0 \leq f(x) \leq g(x)$$

for every x in (a, b). Then f is integrable on $[a, b]$ if g is integrable on $[a, b]$. If f is not integrable on $[a, b]$ then g is not integrable on $[a, b]$.

PROOF IN SECTION 7.7.21.

5.17. Summing inside the integral

For a great many calculus applications we need to be able to use the formula

$$\int_a^b \left(\sum_{n=1}^{\infty} f_n(x) \right) dx = \sum_{n=1}^{\infty} \left(\int_a^b f_n(x)\, dx \right).$$

The integral allows exactly such a computation under some simple hypotheses. This is particularly important in applications since, often, the best formula one can obtain for a function f is some summation formula

$$f(x) = \sum_{n=1}^{\infty} f_n(x).$$

The functions f_n may have a simple formula, but there might be nothing more specific that can be said about f other than this summation formula.

THEOREM 5.22. Suppose that f_1, f_2, f_3, ... is a sequence of nonnegative functions defined and integrable on a compact interval $[a, b]$. If, for each x in $[a, b]$

$$f(x) = \sum_{n=1}^{\infty} f_n(x),$$

and if

$$\sum_{n=1}^{\infty} \left(\int_a^b f_n(x)\, dx \right) < \infty$$

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

$$\int_a^b f(x)\, dx = \sum_{n=1}^{\infty} \left(\int_a^b f_n(x)\, dx \right). \tag{5.1}$$

PROOF IN SECTION 7.7.22.

Remark: As usual in the theory a sequence of points or even a null set where we do not need to verify that

$$f(x) = \sum_{n=1}^{\infty} f_n(x),$$

can be introduced. Simply set all the values at those points equal to zero and none of the integrals will be altered.

5.17.1. Monotone convergence theorem.

THEOREM 5.23 (Monotone convergence theorem). Let $\{f_n\}$ be a nondecreasing sequence of integrable functions on an interval $[a, b]$ and suppose that

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for every x in $[a, b]$. Then f is integrable on $[a, b]$ and

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

provided only that this limit is finite.

PROOF IN SECTION 7.7.23.

5.18. Infinite integrals

We often require an extension of the integral so as to allow infinite values so that a statement such as

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

has meaning. We simply alter Definition 5.1 in a way that is familiar for most infinite limits.

DEFINITION 5.24. Suppose that f is a function defined everywhere on an interval $[a, b]$. Then

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

provided that for every $M > 0$ it is possible to select a Cousin cover β of the interval $[a, b]$ in such a way that

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

for every partition

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

of the interval $[a, b]$ chosen from β.

Such a function would not be integrable by Definition 5.1 and we do not alter that language. Instead we would say that *the integral exists* and is infinite.

5.18.1. Summing inside the integral. We now can extend the formula we have used before:

$$\int_a^b \left(\sum_{n=1}^{\infty} f_n(x) \right) dx = \sum_{n=1}^{\infty} \left(\int_a^b f_n(x)\, dx \right).$$

Our assumptions previously (Theorem 5.22) required the integrals to exist as finite values and the sum of the integrals to converge. Now infinite values are allowed.

THEOREM 5.25. Suppose that $f_1,\ f_2,\ f_3,\ \ldots$ is a sequence of nonnegative functions defined on a compact interval $[a, b]$ and suppose that

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

exists (possibly infinite) for each n. Suppose that, for each x in $[a, b]$,

$$f(x) = \sum_{n=1}^{\infty} f_n(x).$$

Then

$$\int_a^b f(x)\, dx = \sum_{n=1}^{\infty} \left(\int_a^b f_n(x)\, dx \right). \tag{5.2}$$

This theorem is contained in the original proof in Section 7.7.22; the reader can easily extract an argument from there.

CHAPTER 6

Measure Theory

Warning: This is not an elementary chapter!

The measure theory in this chapter should not be considered in the same pedagogical light as the earlier four chapters. Up to now no techniques beyond simple compactness arguments are invoked. If we are willing to define "elementary real analysis" by that definition [nothing past compactness] then those chapters are elementary. Granted all our compactness arguments appeal to the Cousin covering lemma, but at bottom we are not using any idea deeper than the nested interval property.

For those readers who would like to see how the theory so far can be meshed with the usual theory of Lebesgue's measure we offer this chapter. Most students and readers might prefer to use a standard treatment of measure and integration and simply recognize [or accept] that the theory of the first four chapters can be placed within the setting of modern measure and integration as outlined here.

The reader will follow some new ideas in this chapter and will be required to review some old ones. The notion of supremum and infimum of a set of real numbers

$$\sup S \quad \text{and} \quad \inf S$$

are needed. Without experience in sup/inf manipulations, open/closed sets, and indeed much of an elementary analysis course this chapter may prove inaccessible.

6.1. Duality for full covering relations

We add a dual notion to that of full cover. This should naturally belong in Chapter 3 but, to keep the presentation entirely elementary, this more subtle idea and its exploitation was delayed. The definitions are close parallels to Definitions 2.1 and 2.2.

Definition 6.1. A covering relation β is *fine* at a point x_0 if for every $\delta > 0$, the relation β contains at least one pair $([c,d], x_0)$ for which $c \leq x_0 \leq d$ and $0 < d - c < \delta$.

Definition 6.2. A covering relation β is *fine cover* of a set E if β is fine at each point x belonging to the set E.

No discussion of full covers can proceed for long without fine covers making their appearance. This arises simply from the process of negation. The following elementary exercise should clarify.

Exercise 44. Let $f : \mathbb{R} \to \mathbb{R}$ be a function, let $\epsilon > 0$ and define the covering relations

$$\beta_1 = \{([c,d], x) : \ x \in [c,d] \text{ and } |f(d) - f(c)| < \epsilon\}$$

and

$$\beta_2 = \{([c,d],x) :\ x \in [c,d] \text{ and } |f(d) - f(c)| \geq \epsilon\}.$$

Show that

(a) β_1 is full at a point x_0 if and only if β_2 is not fine at that point.
(b) f is continuous at a point x_0 if and only if β_1 is full at x_0 for every choice of $\epsilon > 0$.
(c) f is discontinuous at a point x_0 if and only if β_2 is fine at x_0 for at least one choice of $\epsilon > 0$.
(d) f is continuous at each point in a set E if and only if β_1 is a full cover of E for every choice of $\epsilon > 0$.
(e) f is discontinuous at each point in a set E if and only if there is a positive function ϵ on E so that

$$\beta_3 = \{([c,d],x) :\ x \in [c,d] \text{ and } |f(d) - f(c)| \geq \epsilon(x)\}$$

is a fine cover of E.

6.2. Measures

Our goal is to provide three different definitions of Lebesgue's measure. The fact that all three are equivalent is known as *the Vitali covering theorem*. The three measures will be denoted[1] as

$$\mathcal{L}, \quad \mathcal{L}_*, \quad \text{and} \quad \mathcal{L}^*.$$

By a measure we mean a function defined for all sets and assuming real values (including $+\infty$) and satisfying the following two properties:

Definition 6.3. A set function $\mathcal{M}$ defined for all sets of reals numbers is a *measure*[2] on $\mathbb{R}$ if it has the following properties:

(a) $\mathcal{M}(\emptyset) = 0$.
(b) For any sequence of sets E, E_1, E_2, E_3, ... for which

$$E \subset \bigcup_{n=1}^{\infty} E_n$$

the inequality

$$\mathcal{M}(E) \leq \sum_{n=1}^{\infty} \mathcal{M}(E_n)$$

must hold.

[1]The intention is to annoy and irritate traditionalists who might insist that these symbols would indicate Lebesgue measure, Lebesgue outer measure, and Lebesgue inner measure. They do not! They indicate here the Lebesgue [outer] measure, the full measure, and the fine measure.

[2]Some (most?) authors call this an outer measure.

6.3. Lebesgue's measure

The measure $\mathcal{L}$ is known as Lebesgue's measure and assigns a natural length to every set. There are numerous presentations of Lebesgue's measure and any one of these can be consulted. We offer no proofs of the statements in this section. We simply list what we need.

One traditional construction take place in four steps. For the first step define

$$\mathcal{L}(\emptyset) = 0$$

as required by Definition 6.3. For the second step take an arbitrary open interval (a, b) where $-\infty \leq a < b \leq \infty$ and assign

$$\mathcal{L}((a, b)) = b - a.$$

6.3.1. Measure of open sets. For the third step we take an arbitrary open set G and display G as a sequence of component intervals

$$G = \bigcup_{k=1}^{\infty} (a_k, b_k).$$

(There may be only finitely many components but the argument is the same.)

Then we simply define the measure of G to be the sum of the lengths of the component intervals:

$$\mathcal{L}(G) = \bigcup_{k=1}^{\infty} \mathcal{L}((a_k, b_k)).$$

This is consistent with step 2 since an open interval (a, b) would be considered an open set with a single component. It is traditional to consider the empty set as an open set having *no components*; to be consistent with step 1 then merely requires us to consider an empty sum to be zero.

6.3.2. Measure of arbitrary sets. The final step in the construction of the Lebesgue measure is to approximate the measure of an arbitrary set by the measure of open sets that contain it.

Definition 6.4. For an arbitrary subset E of $\mathbb{R}$ define

$$\mathcal{L}(E) = \inf\{\mathcal{L}(G) : G \text{ an open set containing } E\}.$$

Theorem 6.5. $\mathcal{L}$ is a measure on $\mathbb{R}$.

6.3.3. Increasing unions of open sets. We do not list all of the properties of Lebesgue's measure, but there is one we shall need for our computations in the proof of a crucial covering theorem later. This theorem (or more general versions of it) is proved in all presentations of Lebesgue measure.

Theorem 6.6. Let G_1, G_2, G_3, ... be an increasing sequence of open subsets $\mathbb{R}$. Then

$$G = \bigcup_{k=1}^{\infty} G_k$$

is also an open set and

$$\mathcal{L}(G) = \lim_{n\to\infty} \mathcal{L}(G_n).$$

6.4. The full measure

Let

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

be a partition or subpartition. By its *length* we shall mean

$$V(\mathcal{L}, \gamma) = \sum_{i=1}^{n} \mathcal{L}([a_i, b_i]) = \sum_{i=1}^{n} (b_i - a_i).$$

This is extended to an arbitrary covering relation β by writing

$$V(\mathcal{L}, \beta) = \sup\{V(\mathcal{L}, \gamma) : \gamma \text{ a subpartition contained in } \beta\}.$$

DEFINITION 6.7. For an arbitrary subset E of $\mathbb{R}$ define

$$\mathcal{L}^*(E) = \inf\{V(\mathcal{L}, \beta) : \beta \text{ a full cover of } E\}.$$

THEOREM 6.8. $\mathcal{L}^*$ is a measure on $\mathbb{R}$.

PROOF IN SECTION 7.8.1.

6.5. The fine measure

DEFINITION 6.9. For an arbitrary subset E of $\mathbb{R}$ define

$$\mathcal{L}_*(E) = \inf\{V(\mathcal{L}, \beta) : \beta \text{ a fine cover of } E\}.$$

THEOREM 6.10. $\mathcal{L}_*$ is a measure on $\mathbb{R}$.

PROOF IN SECTION 7.8.2.

6.6. Vitali Covering Theorem

The Vitali covering theorem asserts that the Lebesgue measure $\mathcal{L}$ is identical to the two measures arising from full and fine covers.

THEOREM 6.6.1 (Vitali Covering Theorem). *For any set of real numbers E,*

$$\mathcal{L}(E) = \mathcal{L}_*(E) = \mathcal{L}^*(E).$$

PROOF IN SECTION 7.8.3.

6.7. Null sets; almost everywhere

We now have some additional characterizations for a set to be a null set. The first is the definition and it is easily seen to be equivalent to the statement that $\mathcal{L}^*(E) = 0$. The rest then follow from the Vitali covering theorem.

- [Definition 3.4] A set E is a null set if for every $\epsilon > 0$ there is a full cover β of E so that

$$\sum_{i=1}^{n} (b_i - a_i) < \epsilon$$

whenever $\gamma = \{([a_i, b_i], x_i) : i = 1, 2, \ldots, n\}$ is a subpartition chosen from β.

- [Fine version] A set E is a null set if for every $\epsilon > 0$ there is a fine cover β of E so that
$$\sum_{i=1}^{n}(b_i - a_i) < \epsilon$$
whenever $\gamma = \{([a_i, b_i], x_i) : i = 1, 2, \dots, n\}$ is a subpartition chosen from β.
- $\mathcal{L}(E) = 0$.
- $\mathcal{L}^*(E) = 0$.
- $\mathcal{L}_*(E) = 0$.

We also introduce some language that is in common use. If some property holds everywhere outside of a null set (i.e., everywhere outside of a set of Lebesgue measure zero) then that property is said to hold *almost everywhere*.

6.8. Differentiation of the integral

The best result we have so far as regards differentiation of the integral is Theorem 5.20 that allows us to assert this at points of continuity. The correct result requires an application of the Vitali covering theorem.

THEOREM 6.11. Let f be an integrable function on an interval $[a, b]$ with indefinite integral F. Then $F'(x) = f(x)$ almost everywhere in the interval $[a, b]$.

PROOF IN SECTION 7.8.6.

6.9. Descriptive characterization of the integral

We are now in a position to complete the Newton characterization of the integral. Up to this point we had everything here except for the statement that $F'(x) = f(x)$ almost everywhere. Now that we know this we have justified this theorem:

THEOREM 6.12. Let f be a function defined on a compact interval $[a, b]$. Then f is integrable on $[a, b]$ if and only if there exists a function F and a set N such that

(a) F is continuous on $[a, b]$,
(b) N is negligible,
(c) F does not grow on N,
(d) $F'(x) = f(x)$ at every point in (a, b) with the exception possibly of points in N.

In that case F is an indefinite integral of f on $[a, b]$ and
$$\int_a^b f(x)\,dx = F(b) - F(a).$$

COROLLARY 6.13. Let f be a function defined on a compact interval $[a, b]$. Then f is integrable on $[a, b]$ if and only if there exists an absolutely continuous [general sense] function $F : [a, b] \to \mathbb{R}$ so that $F'(x) = f(x)$ at almost every point in (a, b). In that case F is an indefinite integral of f on $[a, b]$ and
$$\int_a^b f(x)\,dx = F(b) - F(a).$$

6.9.1. Characterization of indefinite integrals. What functions are the indefinite integrals of integrable functions? The exact characterization requires only the notion that we already seen, functions that do not grow on null sets. Such functions we have described as absolutely continuous [general sense].

THEOREM 6.14. A necessary and sufficient condition that a continuous function $F : [a, b] \to \mathbb{R}$ is the indefinite integral of some integrable function f on $[a, b]$ is that F is absolutely continuous [general sense].

This proof is beyond the scope of a calculus course. We do know, from the material presented here, that the following is true, but we do not know without much more work that the underlined words can be removed:

> A necessary and sufficient condition that a continuous function $F : [a, b] \to \mathbb{R}$ is the indefinite integral of some integrable function f on $[a, b]$ is that F is differentiable at almost every point of (a, b) and absolutely continuous.

6.10. Lebesgue Differentiation Theorem

THEOREM 6.15. Let F be a continuous, nondecreasing function on an interval $[a, b]$. Then F is differentiable almost everywhere in the interval $[a, b]$.

PROOF IN SECTION 7.8.7.

6.11. Measurable sets

The class of sets that play a fundamental role in the measure theory is defined to be those sets that are "approximately closed" in the following sense.

DEFINITION 6.16. An arbitrary subset E of $\mathbb{R}$ is *measurable*[3] if for every $\epsilon > 0$ there is an open set G with $\mathcal{L}(G) < \epsilon$ and so that $E \setminus G$ is closed.

Immediately we see that closed sets are measurable and null sets are measurable. The full range of properties of measurable sets are discussed in any measure theory course; we are making do with as little development as possible consistent with the goal of displaying the highlights of the theory. We need the following:

THEOREM 6.17. The class of all measurable subsets of $\mathbb{R}$ forms a Borel family[4] that contains all closed sets and all null sets.

PROOF IN SECTION 7.8.8.

6.12. Measurable functions

DEFINITION 6.18. An arbitrary function $f : \mathbb{R} \to \mathbb{R}$ is *measurable* if for any real number r

$$E = \{x \in \mathbb{R} : f(x) < r\}$$

is a measurable set.

[3]Most courses will give a different definition of measurable and later on show that this property used here is equivalent.

[4]A collection of sets is a Borel family if it is closed under the formation of unions and intersections of sequences of its members, and contains the complement of each of its members. See the proof for details.

A function $f : [a, b] \to \mathbb{R}$ would be measurable if the corresponding set

$$E = \{x \in [a, b] : f(x) < r\}$$

is a measurable set. An equally simple way to think of this is that a function f defined on an interval is measurable if there is a measurable function $g : \mathbb{R} \to \mathbb{R}$ and $f(x) = g(x)$ for all $x \in [a, b]$.

EXERCISE 45. Let f be a measurable function. Show that each of $|f|$, $[f]^+$, and $[f]^-$ must also be measurable.

6.12.1. 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, ..., E_n and real numbers r_1, r_2, r_3, ..., r_n so that

$$f(x) = \sum_{k=1}^{n} r_k \chi_{E_k}(x)$$

for all real x.

LEMMA 6.19. Any simple function is measurable.

PROOF IN SECTION 7.8.9.

THEOREM 6.20. Every nonnegative, measurable function $f : \mathbb{R} \to \mathbb{R}$ can be written as the sum of a series of nonnegative simple functions:

$$f(x) = \sum_{k=1}^{\infty} f_n(x).$$

PROOF IN SECTION 7.8.10.

6.12.2. Limits of continuous functions.

THEOREM 6.21. Let $\{f_n\}$ be a sequence of continuous functions defined on the real line. Suppose that f is a function on $\mathbb{R}$ for which

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for almost every x. Then f is measurable.

PROOF IN SECTION 7.8.11.

6.12.3. Integrable functions are measurable.

THEOREM 6.22. Every function $f : [a, b] \to \mathbb{R}$ that is integrable on $[a, b]$ is measurable.

PROOF IN SECTION 7.8.12.

6.13. Lebesgue's program: construction of the 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 integrand. Our formal definition of the integral *appears* to do this. Since Cousin 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 then 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 it.

6.13.1. The measure step. The first step in Lebesgue's program is to establish for a broad class of sets E that

$$\int_a^b \chi_E(x)\,dx = \mathcal{L}(E \cap [a,b]).$$

We start by proving this statement for compact sets.

LEMMA 6.23. Let K be a compact subset of an interval $[a,b]$. Then χ_K is integrable on $[a,b]$ and

$$\mathcal{L}(K) = \int_a^b \chi_K(x)\,dx.$$

PROOF IN SECTION 7.8.13.

LEMMA 6.24. Let E be a measurable subset of a compact interval $[a,b]$. Then χ_E is integrable on $[a,b]$ and

$$\mathcal{L}(E) = \int_a^b \chi_E(x)\,dx.$$

PROOF IN SECTION 7.8.14.

6.13.2. 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, …, E_n and real numbers r_1, r_2, r_3, …, r_n so that

$$f(x) = \sum_{k=1}^{n} r_k \chi_{E_k}(x)$$

for all real x. It follows from the integration theory (Theorem 5.8) and the integration of characteristic functions (Lemma 6.24) that such a function is necessarily integrable on any compact interval $[a,b]$ and that

$$\int_a^b f(x)\,dx = \sum_{k=1}^{n} \left(\int_a^b r_k \chi_{E_k}(x)\,dx \right) = \sum_{k=1}^{n} r_k \mathcal{L}(E_k \cap [a,b]).$$

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.

6.13.3. Nonnegative measurable functions. We have seen (Theorem 6.20) that every nonnegative measurable function can be represented by simple functions. Consequently the integral of such a function can be constructed.

THEOREM 6.25. 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

$$f(x) = \sum_{k=1}^{\infty} f_n(x) \quad (a \le x \le b)$$

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 integrable on $[a,b]$ if and only if this series of integrals converges to a finite value.

COROLLARY 6.26. 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.

6.13.4. Absolutely integrable functions. A function f is *absolutely integrable* on an interval $[a,b]$ if both f and $|f|$ are integrable on that interval.

THEOREM 6.27. Let f be a measurable function on an interval $[a,b]$. Then f is absolutely integrable if and only if

$$\int_a^b |f(x)|\,dx < \infty.$$

PROOF IN SECTION 7.8.15.

Our final theorem for Lebesgue's program shows that the integral is constructible by his methods for all absolutely integrable functions. We see in the next section that this is as far as one can go.

THEOREM 6.28. If f is absolutely 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 IN SECTION 7.8.16.

6.13.5. 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 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 6.29. 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 IN SECTION 7.8.17.

CHAPTER 7

PROOFS

The proofs assume the usual calculus background for definitions of continuity and derivatives. We need the mean-value theorem of the calculus but little more than that. It would be useful if the student knew that continuous functions on compact intervals are bounded, but even that is avoided in the proofs by relying on covering arguments instead.

7.1. The nested interval property

There is some unapologetic use of compact intervals in a way that might not satisfy the most rigorous requirements. We construct them when we need them (without using sups and infs) and we assume that we can use this important property:

> **Nested interval property**: If $[a_1, b_1] \supset [a_2, b_2] \supset [a_3, b_3] \dots$ is a shrinking sequence of compact intervals with lengths decreasing to zero,
>
> $$\lim_{n\to\infty}(b_n - a_n) = 0$$
>
> then there is a unique point z that belongs to each of the intervals.

There are only two instances where the property is used but they are crucial. The first is for establishing the Cousin covering lemma; the second for giving the Cauchy criterion for integrability of a function.

7.2. Managing epsilons

The student should have had some familiarity with ϵ, δ proofs. Certainly the notions of continuity, limits, and derivatives in an earlier course will have prepared the rudiments.

In such arguments here we frequently have several steps or many steps. With two steps the student is by now accustomed to splitting ϵ into two pieces

$$\epsilon = \frac{\epsilon}{2} + \frac{\epsilon}{2}.$$

Then the argument needed to show something is smaller than ϵ breaks into showing the two separate pieces are smaller than $\frac{\epsilon}{2}$.

It is only moderately more difficulty to handle infinitely many pieces in a proof. In the calculus as presented in these notes we frequently have to handle some condition described by an infinite sequence of steps. For that a very simple device is available, similar to splitting the ϵ into two or three or more pieces:

$$\epsilon = \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{8} + \frac{\epsilon}{16} + \dots + \frac{\epsilon}{2^n} + \dots.$$

It would be reasonable to use this computation even prior to any study of sequences and series. In fact, however, an infinite series can be avoided in all the proofs anyway since all sums are finite. Thus the student never needs anything beyond the inequality

$$\frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{8} + \frac{\epsilon}{16} + \cdots + \frac{\epsilon}{2^n} < \epsilon.$$

This can be proved with elementary algebra.

7.3. Full covering arguments

While there may appear to be a lot of material presented in the notes (Chapters 1-4), by far the bulk of it uses one basic kind of argument, relying only on the manipulation of full covers. This one technique, once mastered, will carry the student through nearly all of the proofs and exercises.

Indeed the course itself could be defined as going only to the extremes that can be handled by full covering arguments. In Chapter 4 we are just on the edge of introducing measure theory but hold back by relying only on full coverings. Thus sets of measure zero (null sets) are defined and handled only by full covers. Similarly the growth of a continuous function on a set could easily lead to the study of variational measures, but reduces instead to a simple full cover statement.

There is a dual notion: a fine covering argument, i.e., what most would call a Vitali covering argument. That is the line that the course does not cross until Chapter 5, and that line defines what we can prove and what we cannot prove until the advanced material.

7.4. Chapter 1

7.4.1. Proof of Lemma 1.2. If F and G are continuous functions on an interval $[a,b]$ and if

$$F'(x) = G'(x)$$

for all $a < x < b$ then

$$F(b) - F(a) = G(b) - G(a).$$

Proof. To see this define a new function $H(x) = F(x) - G(x)$. This function H is continuous and

$$H'(x) = F'(x) - G'(x) = 0$$

for every $a < x < b$. By the mean-value theorem of the differential calculus this means that

$$H(b) - H(a) = H'(\xi) = 0$$

for at least one point $a < \xi < b$. It follows that

$$F(b) - F(a) = G(b) - G(a).$$

■

7.4.2. Proof of Lemma 1.3. If F and G are continuous functions on an interval $[a,b]$ and if

$$F'(x) = G'(x)$$

for all $a < x < b$ with at most finitely many exceptions then

$$F(b) - F(a) = G(b) - G(a).$$

Proof. Let us suppose that the exceptions are at the points

$$a < x_1 < x_2 < x_3 < \cdots < x_n < b.$$

By Lemma 1.2 we know that

$$F(x_1) - F(a) = G(x_1) - G(a),$$
$$F(x_2) - F(x_1) = G(x_2) - G(x_1),$$
$$F(x_3) - F(x_2) = G(x_3) - G(x_2),$$

etc., finishing with

$$F(b) - F(x_n) = G(b) - G(x_n).$$

Add these identities together (enjoying the many cancelations) and obtain

$$F(b) - F(a) = G(b) - G(a).$$

■

7.4.3. Proof of Lemma 1.5. This is the generalization of Lemma 1.3 and Lemma 1.5. While the mean-value theorem was adequate for the proof of those two lemmas, a new technique is needed here. The proof appears in Section 2.7 using the covering argument techniques that are central to this course. Work through the first few sections of Chapter 2 before attempting this proof.

7.5. Chapter 2

7.5.1. Proof of the Cousin Covering Lemma (Lemma 2.4). The central technical tool needed for the discussion of integrals and handling the proofs is the Cousin covering lemma, asserting that Cousin covers must contain partitions. The proof here depends on the nested interval property of compact intervals.

Proof. Note, first, that if β fails to contain a partition of $[a,b]$ 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 or equal to 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 Cousin 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 $[a,b]$. ■

7.5.2. Proof of Corollary 2.5. According to Exercise 18, if β is a Cousin cover of $[a,b]$ then β is a Cousin cover of every compact subinterval. It follows from the Cousin covering lemma that β thus contains a partition of every compact subinterval of $[a,b]$.

7.6. Chapter 3

7.6.1. Proof of Lemma 3.5. We show that every finite set is null.

Proof. Let $\epsilon > 0$. Let $S = \{s_1, s_2, s_3, \dots, s_m\}$ be a finite set. Define the covering relation

$$\beta = \{([c,d], s_j) : c \le s_j \le d,\ 0 < d - c < \epsilon/(2m) \quad (j = 1, 2, 3, \dots, m)\}.$$

This is a full cover of S. Take any collection

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

from β chosen in such a way that the intervals $\{[a_i, b_i]\}$ do not overlap. Observe that each $x_i = s_j$ for some j and that there can be no more than two different pairs $([a_i, b_i], x_i)$ for which x_i is the same s_j.

Thus each interval $[a_i, b_i]$ is shorter in length than $\epsilon/(2m)$ and there can be at most $2m$ of them. It follows that

$$\sum_{i=1}^{n} (b_i - a_i) < \epsilon.$$

By definition then S is a null set. ■

7.6.2. Proof of Lemma 3.6. Every set E whose elements can be written out as a sequence of points is null.

Proof. Let $\epsilon > 0$. Let $S = \{s_1, s_2, s_3, \dots\}$ be the set. Define the covering relation

$$\beta = \{([c,d], s_j) : c \le s_j \le d,\ 0 < d - c < \epsilon 2^{-j-1} \quad (j = 1, 2, 3, \dots)\}.$$

This is a full cover of S. Take any collection

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

from β chosen in such a way that the intervals $\{[a_i, b_i]\}$ do not overlap. Then each interval $[a_i, b_i]$ corresponds to some $x_i = s_j$ and is shorter in length than $\epsilon 2^{-j}$. Again there are at most two such correspondences. Thus

$$\sum_{i=1}^{n} (b_i - a_i) < 2 \sum_{j=1}^{\infty} \epsilon 2^{-j-1} = \epsilon.$$

By definition then S is a null set. ■

7.6.3. Proof of Lemma 3.7. No open interval is null.

Proof. Suppose that (a,b) is null. Let $[c,d]$ be a compact subinterval of (a,b) and let $\epsilon = d - c > 0$.

Choose a full cover β of (a,b) so that for any collection

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

from β chosen in such a way that the intervals $\{[a_i, b_i]\}$ do not overlap, we must have
$$\sum_{i=1}^{n}(b_i - a_i) < \epsilon/2.$$
It is easy to check that β is a Cousin cover of the interval $[c, d]$. Thus there is a partition
$$\{([a_i, b_i], x_i) : i = 1, 2, \dots, n\}$$
of the interval $[c, d]$ contained in β. From this partition we find that
$$d - c = \sum_{i=1}^{n}(b_i - a_i) < \epsilon/2 = (d - c)/2.$$
Since this is impossible it must be false that (a, b) is null. ■

7.6.4. Proof of Lemma 3.11.

Suppose that the function $F : \mathbb{R} \to \mathbb{R}$ is continuous. Then F is uniformly continuous [in Cauchy's sense] on any compact interval $[a, b]$.

Proof. Let $\epsilon > 0$. Define the covering relation β consisting of all $([y, z], x)$ with $y \le x \le z$ and $|F(x') - F(x)| < \epsilon/4$ whenever $y \le x' \le z$. This β is a full cover, because F is continuous at each point. Thus there is a partition
$$\{([a_i, b_i], x_i) : i = 1, 2, \dots, n\}$$
of the interval $[a, b]$ contained in β. Define δ to be the minimum of the lengths $(b_i - a_i)/2$ $(i = 1, 2, \dots, n)$. Then if c and d are points in $[a, b]$ with $d - c < \delta$ just check that
$$|F(d) - F(c)| < \epsilon.$$
■

7.6.5. Proof of Lemma 4.2.

Let F be a continuous function on an interval $[a, b]$ and let N be a negligible set. Suppose that $F'(x) = 0$ for each x in the interval (a, b) except possibly at points in N and suppose that F does not grow on N. Then F is constant.

Proof. Let $\epsilon > 0$. Let $E = [a, b] \setminus N$ so that $F'(x) = 0$ at each point in E. Define the covering relation
$$\beta_1 = \{([c, d], x) : a < c \le x \le d < b,\ x \in E,\ \text{ and } |F(d) - F(c)| < \eta(d - c)\}$$
where
$$\eta = \frac{\epsilon}{2(b - a)}.$$
This is a full cover of E.

We use the fact that F does not grow on the set N to choose a covering relation β_2 that is a full cover of N and so that F will not grow on that cover by more than $\epsilon/2$. Specifically we choose a full cover β_2 of N consisting of pairs $([c, d], x)$ with $x \in N$ in such a way that
$$\sum_{i=1}^{n} |\Delta F([a_i, b_i])| < \epsilon/2$$

whenever the collection

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

is chosen from β_2 in such a way that the intervals $\{[a_i,b_i]\}$ do not overlap.

Now $\beta = \beta_1 \cup \beta_2$ is evidently a Cousin cover of $[a,b]$. Let

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

be a partition of $[a,b]$ chosen from β. Then

$$|F(b) - F(a)| = \left|\sum_{i=1}^{n} F(b_i) - F(a_i)\right| \leq \sum_{i=1}^{n} |F(b_i) - F(a_i)|\,.$$

But this sum easily splits into two parts (corresponding to pairs from β_1 and β_2), each part being smaller than $\epsilon/2$. Thus

$$|F(b) - F(a)| < \epsilon$$

for all $\epsilon > 0$. From this it follows that $F(b) = F(a)$. The same argument can be applied to any subinterval of $[a,b]$ so F must be constant. ∎

7.6.6. Proof of Corollary 4.3. The corollary requires a couple of steps to establish. Write $E_1 = N_1 \setminus N_2$ and $E_2 = N_2 \setminus N_1$. Note that $E_1 \cup E_2 = N_1 \cup N_2$ is a null set. The function F does not grow on E_1 (because of the assumption). But it also does not grow on E_2 because it has a finite derivative there and E_2 is a null set. (Use Exercise 27 for this.) The same applies to G and E_2. Now apply Lemma 4.2 to $H = F - G$ and the null set $N_1 \cup N_2$.

7.7. Chapter 4

7.7.1. Proof of Theorem 5.2(a). In Section 5.3 we have stated that the integral includes each of the three variants of the Newton integral from Chapter 1. We need really only one proof, since the final condition in the statement of the theorem includes the other three. For teaching purposes let us present proofs for each of the three statements.

> Assume that F is continuous on $[a,b]$ and $F'(x) = f(x)$ for every $a < x < b$. Then f is integrable on $[a,b]$ with
>
> $$\int_a^b f(x)\,dx = F(b) - F(a).$$

We need to start with a positive ϵ, find an appropriate Cousin cover and then show that all the Riemann sums are within ϵ of $F(b) - F(a)$. Since the proof has three steps, one at the point a, one at the point b, and a final step handling all the points in (a,b) we split ϵ into three parts:

$$\epsilon = \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4}.$$

Proof. Let $\epsilon > 0$. Taking advantage of the fact that F is continuous at a and b, we define the covering relations

$$\beta_1 = \{([a,c],a) : |F(c) - F(a)| + |f(a)|(c-a) < \epsilon/4\}$$

and

$$\beta_2 = \{([c,b],c) : |F(b) - F(c)| + |f(b)|(b-c) < \epsilon/4\}.$$

At the remaining points we take advantage of the derivative formula $F'(x) = f(x)$ to define

$$\beta_3 = \{([c,d],x) : c \leq x \leq d \;\text{ and }\; |F(d) - F(c) - f(x)(d-c)| < \eta(d-c)\}$$

where

$$\eta = \frac{\epsilon}{2(b-a)}.$$

Combine the covering relations

$$\beta = \beta_1 \cup \beta_2 \cup \beta_3.$$

Simply check (point-by-point) that β is a Cousin cover of $[a,b]$. At the points a and b this uses the continuity of F. At the points $a < x < b$ this uses the definition of the derivative.

Finally let π be any partition of $[a,b]$ chosen from β,

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

Define

$$\gamma_1 = \pi \cap \beta_1, \quad \gamma_2 = \pi \cap \beta_2, \quad \text{and} \quad \gamma_3 = \pi \cap \beta_3,$$

so that γ_k $(k = 1,2,3)$ just contains elements $([a_i,b_i],\xi_i)$ that happen also to belong to β_k. Here the γ_k are not themselves partitions, but subsets (possibly empty) of the partition π. We might call them *subpartitions*. Together they combine[1] to form π.

There remain only some computations now:

$$\left|\sum_{i=1}^{n} f(\xi_i)(b_i - a_i) - [F(b) - F(a)]\right| = \left|\sum_{i=1}^{n} \{f(\xi_i)(b_i - a_i) - [F(b_i) - F(a_i)]\}\right|$$

$$\leq \sum_{i=1}^{n} |f(\xi_i)(b_i - a_i) - [F(b_i) - F(a_i)]| \leq \sum_{\gamma_1} + \sum_{\gamma_2} + \sum_{\gamma_3}$$

where the three sums are taken over elements in the corresponding subpartition γ_k, i.e., that correspond to the situation in which the pairs $([a_i,b_i],\xi_i)$ belong to β_1, β_2, or β_3. (An empty sum is always interpreted as zero.)

It is easy to check that

$$\sum_{([a_i,b_i],\xi_i)\gamma_1} |f(\xi_i)(b_i - a_i) - [F(b_i) - F(a_i)]| < \epsilon/4,$$

$$\sum_{([a_i,b_i],\xi_i)\gamma_2} |f(\xi_i)(b_i - a_i) - [F(b_i) - F(a_i)]| < \epsilon/4,$$

since the sums are either empty or contain just an element corresponding to the endpoint a or b. The remaining sum

$$\sum_{([a_i,b_i],\xi_i)\gamma_3} |f(\xi_i)(b_i - a_i) - [F(b_i) - F(a_i)]| < \eta \sum_{i=1}^{n} (b_i - a_i) \leq \eta(b-a) = \epsilon/2.$$

Putting this together, we have proved that

$$\left|\sum_{i=1}^{n} f(\xi_i)(b_i - a_i) - [F(b) - F(a)]\right| < \epsilon$$

[1] We haven't assumed that the γ_k are disjoint; they need not be as defined, but we could arrange them to be so in a similar argument if that were needed.

for any choice of partition from β. Thus, directly from the definition, f is integrable on $[a,b]$ and

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

■

7.7.2. Proof of Theorem 5.2(b). We prove the second version is a similar manner. Now we are assuming that

$$F'(x) = f(x)$$

for every $a < x < b$ with possibly p exceptions, say at x_1, x_2, x_3, $\dots x_p$. Again we need to establish that f is integrable on $[a,b]$ with

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

We need to start with a positive ϵ, find an appropriate Cousin cover and then show that all the Riemann sums are within ϵ of $F(b) - F(a)$. The proof now has extra steps, to handle the p extra points where the derivative might fail. This suggests that we split ϵ into more parts:

$$\epsilon = \frac{\epsilon}{2} + \frac{\epsilon}{2p+4} + \frac{\epsilon}{2p+4} + \cdots + \frac{\epsilon}{2p+4}.$$

Proof. Let $\epsilon > 0$. Construct β_1 and β_2 as before but use the smaller choice $\epsilon/(2p+4)$ rather than $\epsilon/4$:

$$\beta_1 = \{([a,c],a) : |F(c) - F(a)| + |f(a)|(c-a) < \epsilon/(2p+4)\}$$

and

$$\beta_2 = \{([c,b],c) : |F(b) - F(c)| + |f(b)|(b-c) < \epsilon/(2p+4)\}.$$

Add a new covering relation to take care of the points x_1, x_2, $\dots$, x_p. We know that F is continuous at these points so each of these can be handled in the same way we handled a and b:

$$\beta_4 = \{([c,d],x_i) : |F(d) - F(c)| + |f(x_i)|(d-c) < \epsilon/(2p+4),\ i = 1,2,3,\dots,p\}.$$

The rest of the proof is identical except that there will be four pieces of the sum to handle rather than the three pieces from the first proof. ■

7.7.3. Proof of Theorem 5.2(c). We generalize the proof by one more step, now assuming that

$$F'(x) = f(x)$$

for every $a < x < b$ with possibly infinitely many exceptions, say at a sequence of points e_1, e_2, e_3, $\dots$ from the interval $[a,b]$. Again we need to establish that f is integrable on $[a,b]$ with

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

We need to start with a positive ϵ, find an appropriate Cousin cover and then show that all the Riemann sums are within ϵ of $F(b)-F(a)$. The proof now has infinitely many steps, to handle all the extra points where the derivative might fail. This suggests that we split ϵ into infinitely many parts:

$$\epsilon = \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{8} + \cdots + \frac{\epsilon}{2^n}.$$

Proof. Let $\epsilon > 0$. Construct β_1 and β_2 as before but using the choices $\epsilon/4)$ and $\epsilon/8$:

$$\beta_1 = \{([a,c],a) : |F(c) - F(a)| + |f(a)|(c-a) < \epsilon/4\}$$

and

$$\beta_2 = \{([c,b],c) : |F(b) - F(c)| + |f(b)|(b-c) < \epsilon/8\}.$$

Add a new covering relation to take care of all the points e_1, e_2, ..., e_p, ... in the sequence. We know that F is continuous at these points so each of these can be handled in the same way we handled a and b:

$$\beta_4 = \{([c,d],e_i) : |F(d) - F(c)| + |f(e_i)|(d-c) < \epsilon 2^{-p-2},\ i = 1,2,3,\ldots,\}.$$

The rest of the proof is identical except that you will be able to use the identity

$$\frac{\epsilon}{8} + \frac{\epsilon}{16} + \frac{\epsilon}{32} + \ldots \frac{\epsilon}{2^n} + \cdots = \frac{\epsilon}{4}$$

to handle the Riemann sum contribution from β_4. ■

7.7.4. Proof of Theorem 5.2(d). We generalize the proof by a final step, now assuming that

$$F'(x) = f(x)$$

for every $a < x < b$ with the exception of some null set N and that F does not grow on N. Again we need to establish that f is integrable on $[a,b]$ with

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

Proof. We can assume that N contains the endpoints a and b. Let us also use the function $f_1(x) = f(x)$ for $x \in [a,b] \setminus N$ and $f_1(x) = 0$.

Let $\epsilon > 0$. Construct

$$\beta_1 = \{([c,d],x) : c \le x \le d \text{ and } |F(d) - F(c) - f(x)(d-c)| < \eta(d-c)\}$$

where

$$\eta = \frac{\epsilon}{2(b-a)}.$$

This is a full cover of $[a,b] \setminus N$. Since F does not grow on N there is a full cover β_2 of the set N with the property that

$$\sum_{(I,x)\in\gamma} |\Delta F(I)| < \epsilon/2$$

for any subpartition γ from β_2.

The union $\beta = \beta_1 \cup \beta_2$ is a Cousin cover of $[a,b]$. A simple argument [similar to our other three variants, but in fact a bit easier] will show that

$$\left| F(b) - F(a) - \sum_{(I,x)\in\pi} f_1(x)\mathcal{L}(I) \right| < \epsilon$$

for any partition π from β_2. This proves that f_1 [not f] is integrable on $[a,b]$ and that

$$\int_a^b f_1(x)\,dx = F(b) - F(a).$$

The final argument, replacing f_1 by f, is given in the next sections. When two functions agree everywhere except on a null set they are essentially the same function as the next sections show.

■

7.7.5. Proof of Lemma 5.3. We show that a function that is zero everywhere but at the points of some sequence must be integrable and have a zero integral.

Proof. Let $\epsilon > 0$. Define the covering relations

$$\beta_1 = \{([c,d],x) : a \le c \le x \le d \le b,\ x \ne e_p,\ p = 1,2,3,\dots,\}.$$

and

$$\beta_2 = \{([c,d],e_p) : a \le c \le e_p \le d \le b,\ \ p = 1,2,3,\dots,\ \ |f(e_p)|(d-c) < \epsilon 2^{-p-1}\}.$$

It is easy to check that $\beta = \beta_1 \cup \beta_2$ is a Cousin cover of $[a,b]$.

Let

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

be any partition of $[a,b]$ from β. Note that if $\xi_i = x_p$ for some p then

$$|f(\xi_i)(b_i - a_i)| < \epsilon 2^{-p-1}$$

and there can be at most two such i corresponding to one p. At all other points $f(\xi) = 0$ so

$$|f(\xi_i)(b_i - a_i)| = 0.$$

Thus

$$\left|\sum_{i=1}^n f(\xi_i)\mathcal{L}([a_i,b_i]) - 0\right| < 2\sum_{p=1}^\infty \epsilon 2^{-p-1} = \epsilon.$$

By definition, then, we have verified that f is integrable on $[a,b]$ and that

$$\int_a^b f(x)\,dx = 0.$$

■

7.7.6. Proof of Lemma 5.4. We show that two functions f and g that agree everywhere but at the points of some sequence must behave identically as regards integrability and have the same integral. The proof is closely related to that for Lemma 5.3.

Proof. Let $\epsilon > 0$. Define the covering relations

$$\beta_1 = \{([c,d],x) : a \le c \le x \le d \le b,\ x \ne e_p,\ p = 1,2,3,\dots,\}.$$

and

$$\beta_2 = \{([c,d],e_p) : c \le e_p \le d,\ p = 1,2,3,\dots,\ \ [|f(e_p)| + |g(e_p)|](d-c) < \epsilon 2^{-p-1}\}.$$

It is easy to check that $\beta' = \beta_1 \cup \beta_2$ is a Cousin cover of $[a,b]$.

Suppose that g is integrable on $[a,b]$. If so, then there is a Cousin cover β'' with the property that for any partition

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

of $[a,b]$ from β'',

$$\left|\sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i,b_i]) - \int_a^b g(x)\,dx\right| < \epsilon/2.$$

Define $\beta = \beta' \cap \beta''$. This is a Cousin cover of $[a,b]$ (because of Exercise 20). Let

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

be any partition of $[a,b]$ from β. A simple computation, using the ideas we just saw in the proof of Lemma 5.3, shows that, since π is a subset of β',

$$\left|\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) - \sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i,b_i])\right| < \epsilon/2.$$

But π is also a subset of β'', so we know that

$$\left|\sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i,b_i]) - \int_a^b g(x)\,dx\right| < \epsilon/2.$$

Putting these two inequalities together we deduce that

$$\left|\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) - \int_a^b g(x)\,dx\right|$$

$$\leq \left|\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) - \sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i,b_i])\right| + \left|\sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i,b_i]) - \int_a^b g(x)\,dx\right|$$

$$< \epsilon/2 + \epsilon/2 = \epsilon.$$

By definition then we have verified that f is integrable on $[a,b]$ and that it has an integral equal to $\int_a^b g(x)\,dx$. ■

7.7.7. Proof of Lemma 5.5. We show that a functions $f : [a,b] \to \mathbb{R}$ that is zero except on a null set N must have a zero integral.

Proof. Let $\epsilon > 0$. Let N be the null set where f may not vanish. For each integer $n = 1,2,3,\dots$ define the sets

$$N_n = \{x \in N : n-1 \leq |f(x)| < n\}.$$

These too are null sets and their union is all of N. Choose full covers β_n of the set N_n so that

$$\sum_{([u,v],w)\in\gamma} v-u < \epsilon 2^{-n-2}/n$$

for any subpartition γ contained in β_n. There must be a Cousin cover β of $[a,b]$ with the property that $\beta[E_n] \subset \beta'[E_n]$ for each n.

Let

$$\pi = \{([u_i,v_i],w_i) : w = 1,2,\dots,n\}$$

be any partition of $[a, b]$ from β. Mostly $f(w_i) = 0$, but when this is not zero, w_i belongs to some N_n and so the pair $([u_i, v_i], w_i)$ belongs to some β_n. In that case we use

$$|f(w_i)| < n$$

and the estimates for the covers β_n.

We deduce that

$$\left|\sum_{i=1}^{n} f(w_i)(v_i - u_i)\right| < \epsilon.$$

Just drop the part of the sum with $f(w_i) = 0$ and handle the separate pieces for which $w_i \in N_n$. By definition then we have verified that f is integrable on $[a, b]$ and that it has an integral equal to zero. ■

7.7.8. Proof of Lemma 5.6. We show that two functions f and g that agree everywhere but at the points of some null set must behave identically as regards integrability and have the same integral. The proof is closely related to that for Lemma 5.3 and Lemma 5.4.

Proof. Let $\epsilon > 0$. Let N be the null set where f and g may differ. For each integer $n = 1, 2, 3, \ldots$ define the sets

$$N_n = \{x \in N : n - 1 \le |f(x)| + |g(x)| < n\}.$$

These too are null sets and their union is all of N. Choose full covers β_n of the set N_n so that

$$\sum_{([u,v],\xi)\in\gamma} v - u < \epsilon 2^{-n-2}/n$$

for any subpartition γ contained in β_n. There must be a Cousin cover β' of $[a, b]$ with the property that $\beta'[E_n] \subset \beta'[E_n]$ for each n.

Suppose that g is integrable on $[a, b]$. If so, then there is a Cousin cover β'' with the property that for any partition

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

of $[a, b]$ from β'',

$$\left|\sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i, b_i]) - \int_a^b g(x)\,dx\right| < \epsilon/2.$$

Define $\beta = \beta' \cap \beta''$. This is a Cousin cover of $[a, b]$ (because of Exercise 20). Let

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

be any partition of $[a, b]$ from β. A simple computation, using the ideas we have already used, shows that

$$\left|\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i, b_i]) - \sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i, b_i])\right| < \epsilon/2.$$

This is because, mostly $f(\xi_i) = g(\xi_i)$, but when they are not equal, ξ_i belongs to some N_n and so the pair $([a_i, b_i], \xi_i)$ belongs to some β_n. In that case we use

$$|f(\xi_i) = g(\xi_i)| \le |f(\xi_i)| + |g(\xi_i)| < n$$

and the estimates for the covers β_n.

But π is also a subset of β'', so we know that

$$\left|\sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i,b_i]) - \int_a^b g(x)\,dx\right| < \epsilon/2.$$

Putting these two inequalities together we deduce that

$$\left|\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) - \int_a^b g(x)\,dx\right|$$

$$\leq \left|\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) - \sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i,b_i])\right| + \left|\sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i,b_i]) - \int_a^b g(x)\,dx\right|$$

$$< \epsilon/2 + \epsilon/2 = \epsilon.$$

By definition then we have verified that f is integrable on $[a,b]$ and that it has an integral equal to $\int_a^b g(x)\,dx$. ■

7.7.9. Proof of Theorem 5.7.

Proof. Since f is integrable there is a Cousin cover β_1 of $[a,b]$ with the property that for any partition

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

from β'' of the interval $[a,b]$,

$$\left|\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) - \int_a^b f(x)\,dx\right| < \epsilon/2.$$

Similarly, since g is integrable there is a Cousin cover β_2 with the property that for any partition

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

from β'',

$$\left|\sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i,b_i]) - \int_a^b g(x)\,dx\right| < \epsilon/2.$$

The intersection $\beta = \beta_1 \cap \beta_2$ is again a Cousin cover of $[a,b]$. Let

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

be any partition of $[a,b]$ chosen from β. We are assuming that

$$f(x) \leq g(x)$$

for every $a \leq x \leq b$. Consequently

$$\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) \leq \sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i,b_i]).$$

Combining these three inequalities for this partition π, we find that

$$\int_a^b f(x)\,dx - \epsilon/2 \leq \int_a^b g(x)\,dx + \epsilon/2.$$

Since ϵ is any positive number we conclude that

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

■

7.7.10. Proof of Theorem 5.8.

Our goal here is to establish the useful formula

$$\int_a^b [rf(x) + sg(x)]\,dx = r\int_a^b f(x)\,dx + s\int_a^b g(x)\,dx.$$

The proof is hardly more than an exercise on the simple theme: the intersection of two Cousin covers is again a Cousin cover.

Proof. It is enough to illustrate the method for the case

$$h(x) = f(x) + g(x)$$

with both f and g integrable on $[a, b]$. Thus we are proving the formula

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

Let $\epsilon > 0$. Since f is integrable on $[a, b]$ there is a Cousin cover β_1 of that interval with the property that for any partition

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

of the interval $[a, b]$ from β'',

$$\left|\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i, b_i]) - \int_a^b f(x)\,dx\right| < \epsilon/2.$$

Similarly, since g is integrable on $[a, b]$ there is a Cousin cover β_2 of that interval with the property that for any partition

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

of $[a, b]$ from β'',

$$\left|\sum_{i=1}^{n} g(\xi_i)\mathcal{L}([a_i, b_i]) - \int_a^b g(x)\,dx\right| < \epsilon/2.$$

The intersection $\beta = \beta_1 \cap \beta_2$ is again a Cousin cover of $[a, b]$. Let

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

be any partition of $[a, b]$ chosen from β. Then we easily check that

$$\left|\sum_{i=1}^{n} [f(\xi_i) + g(\xi_i)\mathcal{L}([a_i, b_i]) - \int_a^b f(x)\,dx - \int_a^b g(x)\,dx\right| < \epsilon.$$

By definition this means that $h(x) = f(x) + g(x)$ is integrable on $[a, b]$ and that

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

■

7.7.11. Proof of Theorem 5.9. Our goal is to establish the useful formula

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

for integrating over two adjacent intervals $[a, c]$ and $[c, b]$.

Proof. We begin by invoking our right to change the value of the function at points that are annoying. We decide now that $f(c) = 0$.

Since f is integrable on $[a, c]$ there is a Cousin cover β_1 of $[a, c]$ with the property that for any partition π of the interval $[a, c]$ from β_1,

$$\left| \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) - \int_a^b f(x)\,dx \right| < \epsilon/2.$$

Similarly, since f is also integrable on $[c, b]$ there is a Cousin cover β_2 of $[c, b]$ with the property that for any partition π of the interval $[c, b]$ from β_2,

$$\left| \sum_{(I,x)\in\pi} f(x)\mathcal{L}(I) - \int_b^c f(x)\,dx \right| < \epsilon/2.$$

We construct a Cousin cover β of the full interval $[a, b]$ by extracting what we need from β_1 and β_2 and adding in some bits to take care of the point c. Define β to consist of

(a) all pairs $([s, t], x)$ with $[s, t] \subset [a, c]$ chosen from β_1,
(b) together with all pairs $([s, t], x)$ with $[s, t] \subset [c, b]$ chosen from β_2,
(c) and, finally, also those pairs $([s, t], c)$ for which $([s, c], c)$ belongs to β_1 and $([c, t], c)$ belongs to β_2.

It is easy to check that β is a Cousin cover of the full interval $[a, b]$. Let

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

be any partition of $[a, b]$ chosen from β. We can determine two partitions π_1 and π_2 so that $\pi_1 \subset \beta_1$, $\pi_2 \subset \beta_2$ and the Riemann sums match up thusly:

$$\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i, b_i]) = \sum_{(I,x)\in\pi_1} f(x)\mathcal{L}(I) + \sum_{(I,x)\in\pi_2} f(x)\mathcal{L}(I).$$

The only adjustment needed to find these two partitions is to check whether π happens to contain an element $([s, t], c)$ with $s < c < t$. If so split that element into $([s, c], c)$ and $([c, t], c)$. That will have no effect on the Riemann sums. But having made that split the partition π divides neatly into parts in β_1 and parts in β_2.

Then we easily check that

$$\left| \sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i, b_i]) - \int_a^c f(x)\,dx + \int_c^b f(x)\,dx \right| < \epsilon.$$

By definition this means that $f(x))$ is integrable on $[a, b]$ and that

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

■

7.7.12. Proof of Theorem 5.11 [Necessity]. Let us prove the theorem in two separate steps, the first for sufficiency and the second for necessity.

> **Necessary condition:** If a function f defined on a compact interval $[a,b]$ is integrable on $[a,b]$ then for every $\epsilon > 0$ there exists a Cousin cover β of $[a,b]$ such that
> $$\left|\sum_{i=1}^{n}\sum_{j=1}^{m}[f(\xi_i) - f(\xi_j')]\mathcal{L}([a_i,b_i]\cap[a_j',b_j'])\right| < \epsilon$$
> for every pair of partitions π_1 and π_2 of the interval $[a,b]$ chosen from β:
> $$\pi_1 = \{([a_i,b_i],\xi_i) : i = 1,2,\dots,n\}$$
> and
> $$\pi_2 = \{([a_j',b_j'],\xi_j') : j = 1,2,\dots,m\}.$$

Proof. Let $\epsilon >$ and choose a Cousin cover β of $[a,b]$ so that every Riemann sum from β is within $\epsilon/2$ of the value
$$c = \int_a^b f(x)\,dx.$$
Then if π_1 and π_2 are partitions of the interval $[a,b]$ chosen from β:
$$\pi_1 = \{([a_i,b_i],\xi_i) : i = 1,2,\dots,n\}$$
and
$$\pi_2 = \{([a_j',b_j'],\xi_j') : j = 1,2,\dots,m\}$$
we know that
$$\left|\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) - \sum_{j=1}^{m} f(\xi_j')\mathcal{L}([a_j',b_j'])\right|$$
$$\leq \left|\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i,b_i]) - c\right| + \left|\sum_{j=1}^{m} f(\xi_j')\mathcal{L}([a_j',b_j']) - c\right| < \epsilon/2 + \epsilon/2 = \epsilon.$$
But Lemma 5.10 allows us to write this difference of the two Riemann sums in precisely the way described by the theorem. ■

7.7.13. Proof of Theorem 5.11 [Sufficiency]. The other direction in the proof of this theorem lies deeper and will require an appeal to the nested interval property.

> **Sufficient condition:** Suppose that f is a function defined on a compact interval $[a,b]$ so that for every $\epsilon > 0$ there exists a Cousin cover β of $[a,b]$ such that
> $$\left|\sum_{i=1}^{n}\sum_{j=1}^{m}[f(\xi_i) - f(\xi_j')]\mathcal{L}([a_i,b_i]\cap[a_j',b_j'])\right| < \epsilon$$
> for every pair of partitions π_1 and π_2 of the interval $[a,b]$ chosen from β:
> $$\pi_1 = \{([a_i,b_i],\xi_i) : i = 1,2,\dots,n\}$$

and

$$\pi_2 = \{([a'_j, b'_j], \xi'_j) : j = 1, 2, \dots, m\}.$$

Then f is integrable on $[a, b]$.

Proof. Start off by selecting a Cousin cover β_n of $[a, b]$ in such a way that β_1 and so that

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{m} [f(\xi_i) - f(\xi'_j)] \mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) \right| < 2^{-1}$$

for every pair of partitions π_1 and π_2 of the interval $[a, b]$ chosen from β_1.

For each positive integer $n = 2, 3, 4, \dots$ then select (inductively) a Cousin cover β_n of $[a, b]$ in such a way that $\beta_n \subset \beta_{n-1}$ so that

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{m} [f(\xi_i) - f(\xi'_j)] \mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) \right| < 2^{-n}$$

for every pair of partitions π_1 and π_2 of the interval $[a, b]$ chosen from β_n.

In particular we know from Lemma 5.10 that this means any two such Riemann sums are this close together:

$$\left| \sum_{i=1}^{n} f(\xi_i) \mathcal{L}([a_i, b_i]) - \sum_{j=1}^{m} f(\xi'_j) \mathcal{L}([a'_j, b'_j]) \right| < 2^{-n}.$$

Let S_n denote the set of all values arising from Riemann sums using the Cousin cover β_n. We now know that the diameter of the set S_n no greater than 2^{-n}. We know too, that since $\beta_n \subset \beta_{n-1}$ $(n = 2, 3, 4, \dots)$ the sets $\{S_n\}$ form a shrinking sequence:

$$S_1 \supset S_2 \supset S_3 \supset \dots .$$

Let us choose for each n a compact interval $[a_n, b_n]$ in such a way that $S_n \subset [a_n, b_n]$, $b_n - a_n \leq 2^{-n}$ and

$$[a_n, b_n] \subset [a_{n-1}, b_{n-1}] \quad (n = 2, 3, 4, \dots).$$

By invoking the nested interval property, we know that there is a unique point that belongs to each of the intervals $[a_n, b_n]$. Call this point c and observe that c is within 2^{-n} of every point in the set S_n.

Let $\epsilon > 0$, choose an integer m so that $2^{-m} < \epsilon$, and set $\beta = \beta_m$. Let

$$\pi_1 = \{([a_i, b_i], \xi_i) : i = 1, 2, \dots, n\}$$

be any partition from β. Then the Riemann sum for π is within 2^{-m} of c, i.e.,

$$\left| \sum_{i=1}^{n} f(\xi_i) \mathcal{L}([a_i, b_i]) - c \right| \leq 2^{-m} < \epsilon.$$

This verifies exactly the condition in the definition for f to be integrable on $[a, b]$.
■

7.7.14. Proof of Theorem 5.12 [Simple case]. Let us prove the simplest version of this theorem to illustrate the methods; this is the version expressed in the second Corollary. We prove the theorem without any assumptions[2] about the function f being bounded, but in fact whenever a function is continuous at *every* point of a compact interval $[a, b]$ it is necessarily bounded.

> Let f be a continuous function on an interval $[a, b]$. Then f is integrable.

Proof. Let $\epsilon > 0$ and set

$$\eta = \frac{\epsilon}{2(b-a)}.$$

Define the covering relation

$$\beta = \{([c,d],x) : a \le c \le x \le d \le b, \quad \omega f([c,d]) < \eta\}.$$

(Recall that $\omega f([c, d])$ is the oscillation of the function f on the interval $[c, d]$; for continuous functions this is small when the interval is small.)

Using the continuity of f at each point we easily check that β is a Cousin cover of $[a, b]$. Consider a pair of partitions π_1 and π_2 of the interval $[a, b]$ chosen from β:

$$\pi_1 = \{([a_i, b_i], \xi_i) : i = 1, 2, \dots, n\}$$

and

$$\pi_2 = \{([a'_j, b'_j], \xi'_j) : j = 1, 2, \dots, m\}.$$

We shall estimate

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{m}[f(\xi_i) - f(\xi'_j)]\mathcal{L}([a_i, b_i] \cap [a'_j, b'_j])\right|$$

Note that if

$$[a_i, b_i] \text{ and } [a'_j, b'_j]$$

do not overlap then there is no contribution to the sum. If $[a_i, b_i]$ and $[a'_j, b'_j]$ happen to overlap then we notice that this forces

$$|f(\xi_i) - f(\xi'_j)| < 2\eta.$$

This observation allows us to estimate the sum that we need in order to apply the Theorem 5.11:

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{m}[f(\xi_i) - f(\xi'_j)]\mathcal{L}([a_i, b_i] \cap [a'_j, b'_j])\right|$$

$$< 2\eta\sum_{i=1}^{n}\sum_{j=1}^{m}\mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) \le 2\eta(b-a) = \epsilon$$

Consequently we have verified the condition in Theorem 5.11 and it follows that f must be integrable on $[a, b]$. ■

[2] The reader may find it curious that most elementary classes [using the Riemann integral] would be unable to prove this theorem until after proving that continuous functions on compact intervals are uniformly continuous.

7.7.15. Proof of Theorem 5.12 [General case]. Let us prove the general version of this theorem by nearly repeating the same proof but also using familiar methods for handling a sequence of exceptional points.

> Let f be a function defined on an interval $[a, b]$ and continuous at each point with the exception possibly of a sequence of points e_1, e_2, e_3, ... Then, provided f is bounded, f must be integrable on $[a, b]$.

Proof. Let $\epsilon > 0$ and suppose that $|f(x)| < M$ for all x in the interval $[a, b]$. Set

$$\eta = \frac{\epsilon}{4(b-a)}.$$

Define the covering relation

$$\beta_1 = \{([c,d],x) : a \leq c \leq x \leq d \leq b, \quad \omega f([c,d]) < \eta\}.$$

The relation β_1 fails to be a Cousin cover of $[a, b]$ only at the points where the function may be discontinuous.

We remedy that by adding in a further covering relation:

$$\beta_2 = \{([c,d],e_p) : a \leq c \leq e_p \leq d \leq b, \quad p = 1,2,3,\dots, \quad (d-c) < \epsilon 2^{-p-2}/M\}.$$

It is easy to check that $\beta = \beta_1 \cup \beta_2$ is a Cousin cover of $[a, b]$.

Consider a pair of partitions π_1 and π_2 of the interval $[a, b]$ chosen from β:

$$\pi_1 = \{([a_i, b_i], \xi_i) : i = 1, 2, \dots, n\}$$

and

$$\pi_2 = \{([a'_j, b'_j], \xi'_j) : j = 1, 2, \dots, m\}.$$

We shall estimate

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{m} [f(\xi_i) - f(\xi'_j)] \mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) \right|$$

Note that if

$$[a_i, b_i] \text{ and } [a'_j, b'_j]$$

do not overlap then there is no contribution to the sum. If $[a_i, b_i]$ and $[a'_j, b'_j]$ happen to overlap then we notice that either (i) one of the points ξ_i or ξ'_j is from the sequence of points e_1, e_2, e_3, ..., or else (ii) f is continuous at both points and we have arranged to have the inequality

$$|f(\xi_i) - f(\xi'_j)| < 2\eta.$$

In case (i) we would use merely that

$$|f(\xi_i) - f(\xi'_j)| \leq |f(\xi_i)| + |f(\xi'_j)| < 2M.$$

This observation allows us to estimate the sum that we need in order to apply the Theorem 5.11:

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{m} [f(\xi_i) - f(\xi'_j)] \mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) \right|$$

$$< 2\eta \sum_{i=1}^{n} \sum_{j=1}^{m} \mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) + 2M \sum_{p=1}^{\infty} \left[\epsilon 2^{-p-1}/M\right] \leq 2\eta(b-a) + \epsilon/2 = \epsilon.$$

Consequently we have verified the condition in Theorem 5.11 and it follows that f must be integrable on $[a, b]$.

■

7.7.16. Proof of Theorem 5.15. We prove this generalization of Theorem 5.12 by mimicking the proof given for that theorem. It is essentially just a matter of replacing a negligible set (a sequence of points) with a possibly larger one (any null set).

Proof. Let $\epsilon > 0$ and suppose that $|f(x) < M$ for all x in the interval $[a, b]$. Set

$$\eta = \frac{\epsilon}{4(b-a)}.$$

Define the covering relation

$$\beta_1 = \{([c,d],x) : a \leq c \leq x \leq d \leq b, \quad \omega f([c,d]) < \eta\}.$$

The relation β_1 fails to be a Cousin cover of $[a, b]$ only at the points where the function may be discontinuous.

Let N be the null set consisting of points where we do not know if f is continuous. By the definition of a null set we may select a full cover β_2 of N in such a way that

$$\sum_{i=1}^{n}(b_i - a_i) < \epsilon/(8M)$$

whenever a collection

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

is chosen from β_2 in such a way that the intervals $\{[a_i, b_i]\}$ do not overlap.

It is easy to check that $\beta = \beta_1 \cup \beta_2$ is a Cousin cover of $[a, b]$. Consider a pair of partitions π_1 and π_2 of the interval $[a, b]$ chosen from β:

$$\pi_1 = \{([a_i, b_i], \xi_i) : i = 1, 2, \dots, n\}$$

and

$$\pi_2 = \{([a'_j, b'_j], \xi'_j) : j = 1, 2, \dots, m\}.$$

Suppose that $[a_i, b_i]$ and $[a'_j, b'_j]$ overlap. Then either (i) one of the points ξ_i or ξ'_j is from the set N or else (ii) we must have the inequality

$$|f(\xi_i) - f(\xi'_j)| < 2\eta.$$

In case (i) we would use merely that

$$|f(\xi_i) - f(\xi'_j)| \leq |f(\xi_i)| + |f(\xi'_j)| < 2M.$$

This observation allows us to estimate the sum that we need in order to apply Theorem 5.11:

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{m} [f(\xi_i) - f(\xi'_j)] \mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) \right|$$

.

We split this sum into two parts: the first part consists of summing only over i and j for which case (i) holds. For that either ξ_i belongs to N in which case $([a_i, b_i], \xi_i)$ is from β_2 or ξ'_j belongs to N and so the corresponding element $([a'_j, b'_j], \xi'_j)$ is from β_2. That sum is smaller than

$$\sum_1 \left|f(\xi_i) - f(\xi'_j)\right| \mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) << 2M[\epsilon/8M] + 2M[\epsilon/8M] = \epsilon/2.$$

The second part consists of summing only over i and j for which case (ii) holds, allowing us to use the inequality

$$|f(\xi_i) - f(\xi'_j)| < 2\eta.$$

That sum is smaller than

$$\sum_2 \left|f(\xi_i) - f(\xi'_j)\right| \mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) < \eta \sum_{i=1}^{n} \sum_{j=1}^{m} \mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) = \eta(b-a) = \epsilon/2.$$

.

Putting these together we have

$$\left|\sum_{i=1}^{n} \sum_{j=1}^{m} [f(\xi_i) - f(\xi'_j)]\mathcal{L}([a_i, b_i] \cap [a'_j, b'_j])\right| < \epsilon.$$

Consequently we have verified the condition in Theorem 5.11 and it follows that f must be integrable on $[a, b]$.

■

7.7.17. Proof of Theorem 5.16. We now show that, whenever a function f is integrable on a compact interval, it is also integrable on every compact subinterval $[c, d]$ of $[a, b]$ and there is an indefinite integral, i.e., a continuous function F defined on $[a, b]$ for which

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

for all $a \leq c < d \leq b$.

Proof. Let $\epsilon > 0$ and choose a Cousin cover β of $[a, b]$ so that whenever we are given any two partitions π_1 and π_2 of $[a, b]$ chosen from β we must have both

$$\left|\sum_{(I,x)\in\pi_1} f(x)\mathcal{L}(I) - \int_a^b f(x)\, dx\right| < \epsilon/2.$$

and

$$\left|\sum_{(I,x)\in\pi_2} f(x)\mathcal{L}(I) - \int_a^b f(x)\, dx\right| < \epsilon/2.$$

Adding these we see that

$$\left|\sum_{(I,x)\in\pi_1} f(x)\mathcal{L}(I) - \sum_{(I,x)\in\pi_2} f(x)\mathcal{L}(I)\right| \leq$$

$$\left|\sum_{(I,x)\in\pi_1} f(x)\mathcal{L}(I) - \int_a^b f(x)\, dx\right| + \left|\sum_{(I,x)\in\pi_2} f(x)\mathcal{L}(I) - \int_a^b f(x)\, dx\right| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Now let $[c, d]$ be any compact subinterval of $[a, b]$. Certainly β is a Cousin cover also of $[c, d]$. Let us take any two partitions π_3 and π_4 of $[c, d]$ chosen from β. Since

β contains a partition of every subinterval of $[a, b]$ we can find a subset π of β so that

$$\pi_3 \cup \pi \text{ and } \pi_4 \cup \pi$$

are partitions of $[a, b]$ with π itself partitions of $[a, c]$ and $[d, b]$. (If $c = a$ or $d = b$ then we drop one of these intervals.)

Now we simply observe that

$$\left| \sum_{(I,x)\in\pi_3} f(x)\mathcal{L}(I) - \sum_{(I,x)\in\pi_4} f(x)\mathcal{L}(I) \right|$$

$$= \left| \sum_{(I,x)\in\pi_3\cup\pi} f(x)\mathcal{L}(I) - \sum_{(I,x)\in\pi_4\cup\pi} f(x)\mathcal{L}(I) \right| < \epsilon.$$

According to Theorem 5.11 this condition is enough to assure us that f is integrable on $[c, d]$.

Finally then, if we set

$$F(s) = \int_a^s f(x)\,dx,$$

then we have, for $a < s < t \le b$, by an application of Theorem 5.9, that

$$\int_a^s f(x)\,dx + \int_s^t f(x)\,dx = \int_a^t f(x)\,dx.$$

This shows that

$$\int_s^t f(x)\,dx = F(t) - F(s).$$

To complete the proof we must show that this function F is continuous. It is enough for us to show that F is continuous on the right at a. The same arguments would succeed to show that F is continuous on the left at b. Finally then, by arguing on the interval $[a, c]$ and $[c, b]$ we would know that F is continuous on the right and left at c.

Let us make our simplifying assumption that $f(a) = 0$ (which will change nothing). Now we show that F is continuous on the right at a by showing that

$$|F(t) - F(a)| < \epsilon$$

whenever $a < t < a + \delta$, where we select δ simply by requiring that it be chosen so that

$$([a, t], a) \in \beta$$

whenever $a < t < a + \delta$.

Fix t for which $a < t < a + \delta$. We know that f is integrable on $[t, b]$ and that β is a Cousin cover of $[t, b]$. Thus we may choose π_1, a partition of the interval $[t, b]$ chosen from β_1:

$$\pi_1 = \{([a_i, b_i], \xi_i) : i = 2, \dots, n\}$$

chosen so that

$$\left| F(b) - F(t) - \sum_{(I,x)\in\pi_1} f(x)\mathcal{L}(I) \right| < \epsilon/2.$$

Note that $a_2 = t$. Define $([a_1, b_1], \xi_1) = ([a, t], a)$ and consider the enlarged partition

$$\pi_2 = \{([a_i, b_i], \xi_i) : i = 1, 2, \dots, n\},$$

now a partition of the larger interval $[a, b]$. We note that $t = a_1$ and $\xi_1 = a$ so that $f(\xi_1) = f(a) = 0$. In particular then

$$\sum_{(I,x)\in\pi_1} f(x)\mathcal{L}(I) = \sum_{(I,x)\in\pi_2} f(x)\mathcal{L}(I).$$

Now we recall, since π_2 is a partition of $[a, b]$ from β, that

$$\left| F(b) - F(a) - \sum_{(I,x)\in\pi_2} f(x)\mathcal{L}(I) \right| < \epsilon/2.$$

Adding these inequalities we have

$$|F(t) - F(a)| < \epsilon.$$

■

7.7.18. Proof of Theorem 5.17 [Henstock Criterion].

Proof. Suppose first that this condition holds: for every $\epsilon > 0$ there is a Cousin cover β of $[a, b]$ such that

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

for all partitions

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

of $[a, b]$ contained in β. Notice that the same statement would be true if π is a partition of a subinterval $[c, d]$ (since we could add to it to form a partition of all of $[a, b]$. That means that for such a partition π,

$$\left| \sum_{i=1}^{n} f(\xi_i)(b_i - a_i) - [F(d) - F(c)] \right| \leq$$

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

This proves that

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

and so F is an indefinite integral of f.

For the other direction we suppose that F is an indefinite integral of f on $[a, b]$. Then it is an (almost) immediate consequence of the fact that

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

that we can choose a Cousin cover β so that for any partition π of $[a, b]$ contained in β

$$\left| \sum_{([c,d],x)\in\pi} \{f(x)(d - c) - [F(d) - F(c)]\} \right| < \epsilon/3.$$

(We can insist that β contains only pairs (I, x) with $I \subset [a, b]$.)

Let now γ be an arbitrary subpartition from β. It is an easy matter to add to γ to produce a partition π of $[a, b]$ with $\gamma \subset \pi \subset \beta$.

Suppose that $\eta > 0$ and that there are m gaps needing to be filled to make up the partition, say the intervals $[c_1, d_1]$, $[c_2, d_2]$, $\dots [c_m, d_m]$. Because we know that

$$\int_{c_k}^{d_k} f(x)\,dx = F(d_k) - F(c_k) \quad (k = 1 = 2, \dots, m)$$

we can select additional partitions π_1, π_2, $\dots \pi_m$ of the gap intervals $[c_1, d_1]$, $[c_2, d_2]$, $\dots [c_m, d_m]$ in such a way that each π_k is a subset of β and

$$\left| \sum_{([c,d],x)\in\pi_k} \{f(x)(d-c) - [F(d) - F(c)]\} \right| < \eta/m.$$

Define

$$\pi = \gamma \cup \pi_1 \cup \pi_2 \cup \dots \pi_k.$$

This is a partition of the full interval $[a, b]$ and is a subset of β. Consequently assembling all the pieces in an appropriate way we deduce that

$$\left| \sum_{([c,d],x)\in\gamma} f(x)(d-c) - [F(d) - F(c)] \right| < \epsilon/3 + m(\eta/m) = \epsilon/3 + \eta.$$

As this would hold for any $\eta > 0$ we must have in fact that

$$\left| \sum_{([c,d],x)\in\gamma} f(x)(d-c) - [F(d) - F(c)] \right| \leq \epsilon/3 < \epsilon/2.$$

This inequality is not exactly what we need, but it is only one step away from the inequality that we do need. Take any partition π of $[a, b]$ contained in β. Write

$$\gamma_1 = \{([c,d],x) \in \pi : f(x)(d-c) - [F(d) - F(c)] \geq 0\}$$

and

$$\gamma_2 = \{([c,d],x) \in \pi : f(x)(d-c) - [F(d) - F(c)] < 0\}.$$

Consequently

$$\sum_{([c,d],x)\in\gamma} |f(x)(d-c) - [F(d) - F(c)]| = \sum_{([c,d],x)\in\gamma_1} [f(x)(d-c) - [F(d) - F(c)]]$$
$$+ \sum_{([c,d],x)\in\gamma_2} [-f(x)(d-c) + [F(d) - F(c)]] < \epsilon/2 + \epsilon/2 = \epsilon.$$

■

7.7.19. Proof of Theorem 5.18. We show that the indefinite integral of an integrable function cannot grow on null sets. This is the same as saying that such a function is absolutely continuous (in the general sense).

Proof. This follows with not too much trouble from the Henstock criterion. Let N be an arbitrary null set and write for $n = 1, 2, 3, \dots$

$$N_n = \{x \in N : |f(x)| < n\}.$$

We wish to show that F does not grow on N. It is enough to show that F does not grow on each set N_n since it then follows that F does not grow on the set N which is the union of the sequence $\{N_n\}$.

Fix an integer n and let $\epsilon > 0$. There exists a Cousin cover β_1 of $[a,b]$ such that

$$\sum_{(I,x)\in\pi} |\Delta F(I) - f(x)\mathcal{L}(I)| < \epsilon/2, \tag{7.1}$$

for every partition π of $[a,b]$ contained in β_1. Since N_n is a null set there is a Cousin cover β_2 so that

$$\sum_{(I,x)\in\pi} \mathcal{L}(I) < \epsilon/(2n), \tag{7.2}$$

for every subpartition π contained in $\beta_2[N_n]$. Let $\beta = \beta_1 \cap \beta_2$. This too is a Cousin cover of $[a,b]$.

Suppose now that π is a subpartition contained in $\beta[N_n]$. Then from (7.1) and (7.2) we deduce that

$$\sum_{(I,x)\in\pi} |\Delta F(I)| \leq \sum_{(I,x)\in\pi} |f(x)\mathcal{L}(I)| + \epsilon/2$$

$$\leq \sum_{(I,x)\in\pi} n\mathcal{L}(I) + \epsilon/2 < \epsilon.$$

By definition this shows that F cannot grow on the set N_n. ■

7.7.20. Proof of Theorem 5.20. We now show that the derivative of the indefinite integral is the function integrated, at least at those points where the function is continuous. In symbols we are verifying the formula

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

to be valid at any point t in between a and b for which f is continuous.

Proof. Let $\epsilon > 0$, let $a < x_0 < b$, suppose that F is an indefinite integral of f on $[a,b]$ and that f is continuous at x_0. Then there is a $\delta > 0$ so that

$$|f(x) - f(x_0)| < \epsilon$$

for all x in the interval $[x_0 - \delta, x_0 + \delta]$. Take any $0 < s < \delta$ and observe that

$$F(x_0 + s) - F(x_0) = \int_{x_0}^{x_0+s} f(x)\,dx.$$

Thus

$$|F(x_0 + s) - F(x_0) - f(x_0)s| = \left|\int_{x_0}^{x_0+s} [f(x) - f(x_0)]\,dx\right|$$

$$\leq \int_{x_0}^{x_0+s} |f(x) - f(x_0)|\,dx \leq \int_{x_0}^{x_0+s} \epsilon\,dx = \epsilon s.$$

Rewriting this we see that

$$\left|\frac{F(x_0 + s) - F(x_0)}{s} - f(x_0)\right| \leq \epsilon$$

for all $0 < s < \delta$. A similar argument would show that precisely the same inequality is true for all $0 > s > -\delta$. By definition then $F'(x_0) = f(x_0)$. ■

7.7.21. Proof of Theorem 5.21.

> Let f and g be continuous functions at each point of an open interval (a, b) with finitely many exceptions. Suppose that $0 \leq f(x) \leq g(x)$ for every x in (a, b). Then f is integrable on $[a, b]$ if g is integrable on $[a, b]$.

The student can add a sequence of exceptional set of points into this assertion, but make sure, however, to assume that f is bounded on each subinterval $[c, d]$ with $a < c < d < b$.

Proof. We can split the interval $[a, b]$ into finitely many pieces for which a discontinuity can occur only at an endpoint. Prove that f is integrable on each piece and then add up the pieces.

Thus the argument reduces to handling just a simple situation. Assume f and g are continuous for all $a < x \leq b$. Let G be an indefinite integral for g,

$$G(t) = \int_a^t g(x)\,dx \quad (a < t \leq b).$$

The function f is integrable on any subinterval $[t, b]$ for $a < t < b$ merely because it is continuous there. Define

$$F(t) = -\int_t^c f(x)\,dx \quad (a < t \leq b).$$

Note that

$$\int_t^s f(x)\,dx = F(s) - F(t) \quad (a < s < t \leq b).$$

Note that $F'(x) = f(x)$ at every point in (a, b) by the continuity of f and F is continuous on $(a, b]$, but not yet defined at a.

From the inequality $0 \leq f(x) \leq g(x)$ we easily conclude that

$$0 \leq F(b) - F(t) \leq G(b) - G(t)$$

and so

$$F(b) - G(b) + G(t) \leq F(t)$$

for all $a < t \leq b$. Note that F and G are increasing on (a, b) and F is bounded below by

$$F(b) - G(b) + G(a)$$

. Thus

$$\lim_{t \to 0+} F(t) = p$$

exists. Take $F(a) = p$ and note that F is now a continuous function on $[a, b]$ and qualifies to be an indefinite integral of f on $[a, b]$ in the original Newton sense. It follows that f is integrable on $[a, b]$. ∎

7.7.22. Proof of Theorem 5.22. The proof here will likely be skipped over for most calculus courses as it is quite detailed. The argument requires nothing more than a covering theorem argument, but it is rather delicate.

We wish to obtain the formula

$$\int_a^b \left(\sum_{n=1}^{\infty} f_n(x) \right) dx = \sum_{n=1}^{\infty} \left(\int_a^b f_n(x)\,dx \right).$$

To prove this formula is not sophisticated. It merely uses the definition of the integral. Given an $\epsilon > 0$ can we find a Cousin cover so that the Riemann sums for the cover are within ϵ of the value we require for the integral? The computations are not so easy.

Proof. Let $\epsilon > 0$. We prove first that there is a Cousin cover β' of $[a,b]$ with the property that for any partition π of $[a,b]$ chosen from β we shall have for a lower bound of the Riemann sums of f over π the number

$$\sum_{n=1}^{\infty}\left(\int_a^b f_n(x)\,dx\right) - \epsilon.$$

Then we construct a Cousin cover β'' of $[a,b]$ with the property that for any partition π of $[a,b]$ chosen from β we shall have for an upper bound of the Riemann sums of f over π the number

$$\sum_{n=1}^{\infty}\left(\int_a^b f_n(x)\,dx\right) + \epsilon.$$

That would then mean that, with $\beta''' = \beta' \cap \beta''$, for any partition π of $[a,b]$ chosen from β''' we shall have all the Riemann sums of f over π within ϵ of the number

$$\sum_{n=1}^{\infty}\left(\int_a^b f_n(x)\,dx\right).$$

Here is how to construct β': Choose first an integer N large enough so that

$$\sum_{n=1}^{N}\left(\int_a^b f_n(x)\,dx\right) \geq \sum_{n=1}^{\infty}\left(\int_a^b f_n(x)\,dx\right) - \epsilon/2.$$

This is possible since we have assumed that the series of integrals converges.

Then choose Cousin covers β_n of the interval $[a,b]$ $(n = 1, 2, \dots, N)$ so that all sums

$$\left|\sum_{(I,x)\in\pi} f_n(x)\mathcal{L}(I) - \int_a^b f_n(x)\,dx\right| < \epsilon 2^{-n-1}$$

whenever $\pi \subset \beta_n$ is a partition of $[a,b]$. Let

$$\beta' = \bigcap_{n=1}^{N} \beta_n.$$

By Exercise 20 we can see that this too is a Cousin cover of $[a,b]$, one that is contained in all of the others.

Take any partition of $[a,b]$ with $\pi \subset \beta'$, and compute

$$\sum_{\pi} f(x)\mathcal{L}(I) \geq \sum_{\pi}\left(\sum_{n=1}^{N} f_n(x)\mathcal{L}(I)\right) = \sum_{n=1}^{N}\left(\sum_{\pi} f_n(x)\mathcal{L}(I)\right) \geq$$

$$\sum_{n=1}^{N}\left(\int_a^b f_n(x)\mathcal{L}(I) - \epsilon 2^{-n-1}\right)$$

$$\geq \sum_{n=1}^{N}\left(\int_a^b f_n(x)\,dx\right) \geq \sum_{n=1}^{\infty}\left(\int_a^b f_n(x)\,dx\right) - \epsilon/2$$

$$\geq \sum_{n=1}^{\infty} \left(\int_a^b f_n(x)\,dx \right) \geq \sum_{n=1}^{\infty} \left(\int_a^b f_n(x)\,dx \right) - \epsilon.$$

This gives the lower bound for all Riemann sums from β' that we wanted.

The proof for the other direction is similar, 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

$$tf(x) \leq \sum_{n=1}^{N(x)} f_n(x).$$

Choose, again and using the same ideas as before, Cousin covers β_n of $[a, b]$ $(n = 1, 2, \dots)$ so that $\beta_1 \supset \beta_2 \supset \beta_3 \dots$ and all sums

$$\left| \sum_{(I,x)\in\pi} f_n(x)\mathcal{L}(I) - \int_a^b f_n(x)\mathcal{L}(I) \right| < \epsilon 2^{-n}$$

whenever $\pi \subset \beta_n$ is a partition of $[a, b]$.

Let

$$E_n = \{x \in [a, b] : N(x) = n\}.$$

We use these sets to carve up the covering relations. Write

$$\beta_n[E_n] = \{(I, x) \in \beta_n : x \in E_n\}$$

and we form β'' to be the combined collection of all these smaller covers $\beta_n[E_n]$. In the usual mathematical notation this means that

$$\beta'' = \bigcup_{n=1}^{\infty} \beta_n[E_n].$$

It is easy to check that β'' is now a Cousin cover of $[a, b]$.

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 $j = 1, 2, 3, \dots, N$,

$$\gamma_j = \{(I, x) \in \pi : x \in E_j\}$$

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

$$\pi = \gamma_1 \cup \gamma_2 \cup \dots \cup \gamma_N.$$

Check the following computations, making sure to use the fact that for $x \in E_i$,

$$tf(x) \leq f_1(x) + f_2(x) + \dots + f_i(x).$$

$$\sum_{\pi} tf(x)\mathcal{L}(I) = \sum_{i=1}^{N} \sum_{\gamma_i} tf(x)\mathcal{L}(I) \leq \sum_{i=1}^{N} \sum_{\gamma_i} \left(f_1(x) + f_2(x) + \dots + f_i(x)\right)\mathcal{L}(I)$$

$$= \sum_{j=1}^{N} \left(\sum_{\sigma_j} f_j(x)\mathcal{L}(I) \right) \leq$$

$$\sum_{j=1}^{N}\left(\int_a^b f_j(x)\,dx+\epsilon 2^{-j}\right)\leq\sum_{j=1}^{\infty}\left(\int_a^b f_j(x)\,dx\right)+\epsilon.$$

We have established this inequality for all $t<1$:

$$t\left(\sum_{\pi} tf(x)\mathcal{L}(I)\right)\leq\sum_{j=1}^{\infty}\left(\int_a^b f_j(x)\,dx\right)+\epsilon.$$

This gives the upper bound for all Riemann sums from β'' that we wanted. ■

7.7.23. Proof of Theorem 5.23. This theorem follows directly from Theorem 5.22 and the identity

$$f(x)=f_1(x)+\sum_{n=1}^{\infty}\left(f_n(x)-f_{n-1}(x)\right).$$

7.8. Chapter 5

7.8.1. Proof of Theorem 6.8. We show that the full version of the measure $\mathcal{L}^*$ is a measure on $\mathbb{R}$ in the sense of Definition 6.3.

Proof. It is just a matter of interpreting the definition to conclude that $\mathcal{L}^*(\emptyset)=0$. Note that if $A\subset B$ then it is easy to check that

$$\mathcal{L}^*(A)\leq\mathcal{L}^*(B).$$

This is merely because any full cover of B would necessarily be also a full cover of A.

Suppose now that we have a sequence of sets E, E_1, E_2, E_3, … for which

$$E\subset\bigcup_{n=1}^{\infty}E_n.$$

We can suppose that each $\mathcal{L}^*(E_n)<\infty$ otherwise the identity that we require

$$\mathcal{L}^*(E)\leq\sum_{n=1}^{\infty}\mathcal{L}^*(E_n)$$

must hold trivially.

Let $\epsilon>0$ and select full covers β_i of E_i so that

$$V(\mathcal{L},\beta_i)\leq\mathcal{L}^*(E_i)+\epsilon 2^{-i}\quad(i=1,2,3,\dots).$$

Note that $\beta=\bigcup_{i=1}^{\infty}\beta_i$ is a full cover of E (merely because it is full at each point in E). Thus

$$\mathcal{L}^*(E)\leq V(\mathcal{L},\beta)\leq\sum_{i=1}^{\infty}V(\mathcal{L},\beta_i)\leq\sum_{i=1}^{\infty}\left(\mathcal{L}^*(E_i)+\epsilon 2^{-i}\right)=\left(\sum_{i=1}^{\infty}\mathcal{L}^*(E_i)\right)+\epsilon.$$

From that we deduce that

$$\mathcal{L}^*(E)\leq\sum_{n=1}^{\infty}\mathcal{L}^*(E_n).$$

Here is a more careful estimate leading to the inequality

$$V(\mathcal{L},\beta)\leq\sum_{i=1}^{\infty}V(\mathcal{L},\beta_i)$$

that the reader may not have seen as immediate. Take any subpartition γ from β and write $\gamma_i = \gamma \cap \beta_i$. Since γ is a finite collection there must be an integer N so that

$$\gamma = \bigcup_{i=1}^{N} \gamma_i.$$

Then the inequality

$$V(\mathcal{L}, \gamma) \leq \sum_{i=1}^{N} V(\mathcal{L}, \gamma_i)$$

is immediate and can be used to obtain the inequalities used above. ■

7.8.2. Proof of Theorem 6.10. We show that the fine version of the measure $\mathcal{L}_*$ is a measure on $\mathbb{R}$ in the sense of Definition 6.3.

Proof. The proof repeats the proof for Theorem 6.8. It is a matter of interpreting the definition to conclude that $\mathcal{L}_*(\emptyset) = 0$. Note that if $A \subset B$ then it is easy to check that

$$\mathcal{L}_*(A) \leq \mathcal{L}_*(B).$$

This is merely because any fine cover of B would necessarily be also a fine cover of A.

Suppose now that we have a sequence of sets E, E_1, E_2, E_3, ... for which

$$E \subset \bigcup_{n=1}^{\infty} E_n$$

We can suppose that each $\mathcal{L}^*(E_n) < \infty$ otherwise the identity that we require

$$\mathcal{L}_*(E) \leq \sum_{n=1}^{\infty} \mathcal{L}_*(E_n)$$

must hold trivially.

Let $\epsilon > 0$ and select fine covers β_i of E_i so that

$$V(\mathcal{L}, \beta_i) \leq \mathcal{L}_*(E_i) + \epsilon 2^{-i}.$$

Note that $\beta = \bigcup_{i=1}^{\infty} \beta_i$ is a fine cover of E (merely because it is fine at each point in E). Thus

$$\mathcal{L}_*(E) \leq V(\mathcal{L}, \beta) \leq \sum_{i=1}^{\infty} V(\mathcal{L}, \beta_i) \leq \sum_{i=1}^{\infty} \left(\mathcal{L}_*(E_i) + \epsilon 2^{-i}\right).$$

From that we deduce easily that

$$\mathcal{L}_*(E) \leq \sum_{n=1}^{\infty} \mathcal{L}_*(E_n).$$

■

7.8.3. Step 1: Vitali covering theorem. For the first step in the proof of the Vitali covering theorem we prove a preliminary covering theorem that is of interest on its own.

LEMMA 7.1 (Radó Covering Theorem). Let E be a set of finite Lebesgue measure. Suppose that β is a covering relation with the property that for each point x in E there is at least one pair (I,x) in β. Then there is a subpartition $\gamma \subset \beta$ so that

$$\mathcal{L}(E) \leq 3 \left\{ \sum_{(I,x)\in\gamma} \mathcal{L}(I) \right\}.$$

Proof. In the language introduced above we are showing, that

$$\mathcal{L}(E) \leq 3V(\mathcal{L},\beta).$$

Let

$$G_0 = \bigcup \{(c,d) : ([c,d],x) \in \beta\}$$

We leave it as an exercise for the reader to check that G_0 is open and includes every point of E with at most countably many exceptions. Thus

$$\mathcal{L}(E) \leq \mathcal{L}(G_0).$$

[Let C be the collection of points in E that do not belong to G_0. Each such point t must correspond to an element $([t,s],x)$ or $([s,t],x)$ from β. Let C_n be the elements of C that correspond to such an interval with length greater than $1/n$.]

Since G_0 is open, we can select a list of open intervals $\{(c_j,d_j)\}$ with rational endpoints so that

$$G_0 = \bigcup_{j=1}^{\infty} (c_j,d_j).$$

Note that for every $t \in G_0$ there is at least one pair $(I,x) \in \beta$ and an index j for which $t \in (c_j,d_j) \subset I$. Thus we can use the sequence $\{(c_j,d_j)\}$ to construct a sequence $\{([a_i,b_i],x_i)\}$ from β with the property that

$$G_0 = \bigcup_{j=1}^{\infty} (a_i,b_i).$$

Simply determine, for each $j = 1,2,3,\dots$, whether or not it is possible to find $(I,x) \in \beta$ for which $(c_j,d_j) \subset I$. If so add (I,x) as the next element of the sequence; if not pass to the next j.

Thus, since $\mathcal{L}(E) \leq \mathcal{L}(G_0)$ and $\mathcal{L}(E)$ is finite, there must exist an integer N large enough so that

$$\mathcal{L}(E) \leq \frac{3}{2}\mathcal{L}\left(\bigcup_{i=1}^{N} (a_i,b_i) \right).$$

This uses Theorem 6.6 applied to this increasing sequence of open sets.

From the finite sequence,

$$([a_i,b_i],x_i) \quad (i = 1,2,3,\dots,N)$$

discard all redundant elements, i.e., any element for which the union above would not change if it were deleted. Relabel the (now nonredundant) sequence as

$$([a_1,b_1],x_1),\ ([a_2,b_2],x_2),\ \dots,\ ([a_N,b_N],x_N)$$

where the order is adjusted so that $a_1 < a_2 < \cdots < a_N$. (We need not worry about $a_i = a_j$ since that cannot occur if any redundancy has already been eliminated.)

Set

$$\gamma_1 = \{([a_i, b_i], x_i) : \ i \text{ is odd}\}$$

and

$$\gamma_2 = \{([a_i, b_i], x_i) : \ i \text{ is even}\}.$$

By the way that this has been constructed both of these are subpartitions and

$$\frac{2}{3}\mathcal{L}(E) \leq \mathcal{L}\left(\bigcup_{i=1}^{N}(a_i, b_i)\right) \leq \sum_{(I,x)\in\gamma_1} \mathcal{L}(I) + \sum_{(I,x)\in\gamma_2} \mathcal{L}(I).$$

We choose $\gamma = \gamma_1$ or γ_2 depending on which yields the larger of these two sums on the right-hand side of the inequality. This finishes the proof. ∎

7.8.4. Step 2: Vitali Covering Theorem. Towards the goal of proving that

$$\mathcal{L}(E) = \mathcal{L}_*(E) = \mathcal{L}^*(E)$$

we prove the following.

STEP 2: $\mathcal{L}_*(E) \leq \mathcal{L}^*(E) \leq \mathcal{L}(E)$.

Proof. This first inequality $\mathcal{L}_*(E) \leq \mathcal{L}^*(E)$ is a trivial consequence of the fact that any full cover of a set E is necessarily also a fine cover of E. For the second inequality let G be any open set containing E. Define

$$\beta = \{(I, x) : \ (I, x) \text{ with } x \in I \subset G\}.$$

Then β is a full cover of E. Note that

$$\mathcal{L}^*(E) \leq V(\mathcal{L}, \beta) \leq \mathcal{L}(G).$$

This follows simply because any subpartition $\gamma \subset \beta$ would have

$$\sum_{(I,x)\in\gamma} \mathcal{L}(I) \leq \mathcal{L}\left(\bigcup_{(I,x)\in\gamma} I\right) \leq \mathcal{L}(G).$$

Since the inequality

$$\mathcal{L}^*(E) \leq \mathcal{L}(G)$$

holds for all open sets containing E it follows from the definition of $\mathcal{L}(E)$ that

$$\mathcal{L}^*(E) \leq \mathcal{L}(E).$$

∎

Because of Step 2 the proof of the Vitali covering theorem can be obtained by verifying just the inequality

$$\mathcal{L}(E) \leq \mathcal{L}_*(E).$$

We shall first show that the theorem follows from a more familiar classical form of that theorem, namely the following statement. This should be considered the geometric form, that is somewhat obscured by considering only the identity of the three measures.

7.8.5. STEP 3: Geometrical form of the Vitali theorem.

LEMMA 7.2 (Geometrical form of the Vitali theorem). Let E be a set of finite Lebesgue measure, $\epsilon > 0$ and β any fine cover of E. Then there exists a subpartition $\gamma \subset \beta$ so that

$$\mathcal{L}\left(E \setminus \bigcup_{(I,x)\in\gamma} I\right) < \epsilon. \tag{7.3}$$

Proof. Our proof uses the Radó covering lemma. Use the notation, for any subpartition γ,

$$V(\mathcal{L},\gamma) = \sum_{(I,x)\in\gamma} \mathcal{L}(I)$$

and

$$\sigma(\gamma) = \bigcup_{(I,x)\in\gamma} I$$

to simplify the writing. Note that $\sigma(\gamma)$ is a finite union of compact intervals, and so also a compact set.

Start with an open set $G_0 \supset E$ for which $\mathcal{L}(G_0) < \infty$ and prune β by considering

$$\beta_0 = \{(I,x) \in \beta : I \subset G_0\}.$$

This β_0 is a fine cover of E and so, by Radó's lemma, we can choose a subpartition $\gamma_1 \subset \beta_0$ for which

$$\mathcal{L}(E) \le 3V(\mathcal{L},\gamma_1).$$

Let

$$G_1 = G_0 \setminus \sigma(\gamma_1)$$

which is evidently an open set that contains $E_1 = E \setminus \sigma(\gamma_1)$. If E_1 is empty or has zero measure we are done.

In general we proceed inductively. Choose a bounded open set G_0 containing E. Set $E_0 = E$,

$$\beta_0 = \{(I,x) \in \beta : I \subset G_0\}.$$

For any $n = 1, 2, \dots$, unless E_{n-1} is empty or has zero measure we select, by Radó's lemma, a subpartition $\gamma_n \subset \beta_{n-1}$ for which

$$\mathcal{L}(E_{n-1}) \le 3V(\mathcal{L},\gamma_n).$$

Let

$$G_n = G_{n-1} \setminus \sigma(\gamma_n)$$

which is evidently an open set that contains $E_n = E_{n-1} \setminus \sigma(\gamma_1)$. Set

$$\beta_n = \{(I,x) \in \beta : I \subset G_n\}.$$

None of the subpartitions γ_n overlap and so, in particular,

$$\gamma'_N = \bigcup_{n=1}^{N} \gamma_n$$

is itself a subpartition contained in β. If the process stops then it is easy to verify that

$$\mathcal{L}\left(E \setminus \sigma(\gamma'_N)\right) = 0.$$

If the process does not stop then for some large N we must have

$$\mathcal{L}\left(E \setminus \sigma(\gamma'_N)\right) < \epsilon. \tag{7.4}$$

To see this note that all the subpartitions have been pruned to lie in the open set G_0. From this it follows that

$$\sum_{i=1}^{\infty} V(\mathcal{L}, \gamma_n) \leq \mathcal{L}(G_0) < \infty.$$

Choose N so large that $V(\mathcal{L}, \gamma_N) < \epsilon/3$ and it will follow that

$$\mathcal{L}(E_{N-1}) \leq 3V(\mathcal{L}, \gamma_N) < \epsilon.$$

But this is exactly (7.4). ■

STEP 4. For any set E of finite Lebesgue measure,

$$\mathcal{L}(E) \leq \mathcal{L}_*(E) \leq \mathcal{L}^*(E) \leq \mathcal{L}(E).$$

Proof. For any such set E and any fine cover β of E select $\pi \subset \beta$ so that (7.3) holds. Then, writing $\sigma(\pi) = \bigcup_{(I,x)\in\pi} I$, we have

$$\mathcal{L}(E) \leq \mathcal{L}(E - \sigma(\pi)) + \mathcal{L}(\sigma(\pi)) < \epsilon + V(\mathcal{L}, \pi).$$

Consequently $\mathcal{L}(E) \leq V(\mathcal{L}, \beta)$ for all fine covers β of E. Thus

$$\mathcal{L}(E) \leq \mathcal{L}_*(E) \leq \mathcal{L}^*(E) \leq \mathcal{L}(E).$$

■

STEP 5. The statement

$$\mathcal{L}(E) = \mathcal{L}_*(E) = \mathcal{L}^*(E)$$

is valid for all sets.

This final step is simple. We know that the statement holds for all sets of finite Lebesgue measure. Any set of infinite measure contains subsets of arbitrarily large measure.

7.8.6. Proof of Theorem 6.11. Let f be an integrable function on an interval $[a, b]$ with indefinite integral F. Then $F'(x) = f(x)$ almost everywhere in the interval $[a, b]$.

Proof. We show for the upper and lower derivates[3] of F show that

$$\overline{D}F(x) = \underline{D}F(x) = f(x)$$

outside of a null set. The set of points where $\underline{D}F(x) < f(x)$ can be analyzed by considering the sets

$$E_{pq} = \{x \in [a, b] : \underline{D}F(x) < p < q < f(x)\}$$

for rational numbers $p < q$. If each of these is a null set, then

$$E = \{x \in [a, b] : \underline{D}F(x) < f(x)\} \tag{7.5}$$

is also a null set since it is the union of a sequence of null sets.

[3]The reader is assumed to know what these mean.

To this end, let $\epsilon > 0$ and select β_0 to be a Cousin cover of $[a,b]$ chosen so that the Henstock criterion (Theorem 5.17) is satisfied. Thus for any subpartition γ contained in β_0

$$\sum_{([c,d],x)\in\gamma} |F(d) - F(c) - f(x)(d-c)| < \epsilon.$$

Define the covering relation

$$\beta_1 = \{([c,d],x) : x \in [c,d],\ \ F(d) - F(c) < p(d-c) < q(d-c) < f(x)(d-c)\}.$$

This is a fine cover of E_{pq}: check that it is fine at each point x in that set, by using the inequalities

$$\underline{D}F(x) < p < q < f(x).$$

The intersection $\beta_0 \cap \beta_1$ is also a fine cover of E_{pq}. We claim that

$$(q-p)\mathcal{L}_*(E_{pq}) \leq (q-p)V(\mathcal{L}, \beta_0 \cap \beta_1) \leq \epsilon.$$

To see this select any subpartition γ from $\beta_0 \cap \beta_1$ and check this computation:

$$(q-p)\sum_{(I,x)\in\gamma} \ell(I) \leq \sum_{(I,x)\in\gamma} [q\ell(I) - p\ell(I)] \leq$$

$$\sum_{([c,d],x)\in\gamma} |f(x)(d-c) - [F(d) - F(c)]| < \epsilon.$$

As $\epsilon > 0$ is arbitrary and $q - p > 0$ it follows that $\mathcal{L}_*(E_{pq}) = 0$. As announced above, this shows that the set E in (7.5) is a null set.

Exactly the same argument, with suitable modifications, shows that the set of points where $\overline{D}F(x) > f(x)$ is a null set. Consequently, outside of a null set,

$$\underline{D}F(x) = \overline{D}F(x) = f(x).$$

The theorem follows. ∎

7.8.7. Proof of Theorem 6.15 [Lebesgue Differentiation Theorem]. Let F be a continuous, nondecreasing function on an interval $[a,b]$. Then F is differentiable almost everywhere in the interval $[a,b]$.

We prove first a narrower version (note the emphasis):

> Let F be a continuous, *strictly increasing* function on an interval $[a,b]$. Then F has a finite derivative at almost every point in the interval $[a,b]$.

The keys to the proof are the following two growth lemmas,

Lemma 7.3. Let f be a continuous, strictly increasing function and let E be a set with the property that $\underline{D}f(x) < p$ for each $x \in E$. Then

$$\mathcal{L}(f(E)) \leq p\mathcal{L}(E). \tag{7.6}$$

Proof. Let G be any open set containing E. Define the covering relation

$$\beta = \left\{[c,d],x) : \frac{f(d)-f(c)}{d-c} < p \text{ and } x \in [c,d] \subset G\right\}.$$

This is a fine cover of E because of the fact that $\underline{D}f(x) < p$ for each x in E.

The function f maps this fine cover of E to a fine cover of $f(E)$ as follows:

$$\beta' = \{(f(I), f(x)) : (I,x) \in \beta\}.$$

This is because the image under f of a compact interval I containing a point x is necessarily a compact interval $f(I)$ that contains the point $f(x)$. Since f is continuous there must be for each $x \in E$ pairs $(f(I), f(x))$ in β' with small intervals $f(I)$ merely because there must pairs (I, x) in β with arbitrary small intervals I.

Consequently we can use β' to estimate the Lebesgue measure of $f(E)$. We claim that

$$\mathcal{L}_*(f(E)) \leq V(\mathcal{L}, \beta') \leq p\mathcal{L}(G).$$

From this claim, true for all such open sets G containing E, we can deduce that

$$\mathcal{L}_*(f(E)) \leq p\mathcal{L}(E). \tag{7.7}$$

The Vitali covering theorem assures us that (7.6) and (7.7) are identical.

To check the claim take any subpartition $\gamma \subset \beta$

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

corresponding to a subpartition from β':

$$\gamma' = \{(f([a_i, b_i]), f(x_i)) : i = 1, 2, \dots, n\}$$

Compute

$$\sum_{i=1}^{n} \mathcal{L}(f([a_i, b_i]) = \sum_{i=1}^{n} [f(b_i) - f(a_i)] \leq \sum_{i=1}^{n} p[b_i - a_i] \leq p\mathcal{L}(G).$$

■

Lemma 7.4. Let f be a continuous, nondecreasing function and let E be a set with the property that $\overline{D}f(x) > q$ for each $x \in E$. Then

$$\mathcal{L}(f(E)) \geq q\mathcal{L}(E). \tag{7.8}$$

Proof. Let G be any open set containing $f(E)$. Define the covering relation

$$\beta = \left\{[c, d], x) : \frac{f(d) - f(c)}{d - c} > q \text{ and } x \in [c, d] \subset f^{-1}(G)\right\}.$$

This is a fine cover of E because of the fact that $\overline{D}f(x) > p$ for each x in E.

Consequently we can use β to estimate the Lebesgue measure of E. We claim that

$$q\mathcal{L}_*(E) \leq V(\mathcal{L}, \beta) \leq p\mathcal{L}(G).$$

From this claim, true for all such open sets G containing $f(E)$, we can deduce that

$$q\mathcal{L}_*(E) \leq \mathcal{L}(f(E)). \tag{7.9}$$

The Vitali covering theorem assures us that (7.8) and (7.9) are identical.

To check the claim take any subpartition $\gamma \subset \beta$

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

Compute

$$q\sum_{i=1}^{n} (b_i - a_i) \leq \sum_{i=1}^{n} [f(b_i) - f(a_i)] \leq \sum_{i=1}^{n} \mathcal{L}(f([a_i, b_i]) \leq \mathcal{L}(G).$$

■

The two growth lemmas are used to check that each of the following sets is a null set:

$$A_{pq} = \{x \in [a,b] : \underline{D}f(x) < p < q < \overline{D}F(x)\} \tag{7.10}$$

and

$$B = \{x \in [a,b] : \overline{D}f(x) = \infty\}. \tag{7.11}$$

The first set (7.10) must satisfy

$$\mathcal{L}(f(E_{pq})) = \mathcal{L}(E_{pq}) = 0$$

because of properties (7.6) and (7.8). That means that the set where the upper derivative exceeds the lower derivative:

$$A = \{x \in [a,b] : \underline{D}f(x) < \overline{D}F(x)\}$$

must also be a null set since this set is the countable union of the collection of sets

$$E_{pq} = \{x \in [a,b] : \underline{D}f(x) < p < q < \overline{D}F(x)\}$$

taken over all rational numbers $p < q$. If each of these is a null set then so too is A.

Also the second set (7.11) must satisfy

$$q\mathcal{L}(B) \leq \mathcal{L}(f(B)) \leq f(b) - f(a)$$

for all positive values of q, because of property (7.8). But this can hold only if $\mathcal{L}(B) = 0$.

It follows so far that a continuous, strictly increasing function f on a compact interval $[a,b]$ must have a finite derivative almost everywhere. If f is merely nondecreasing then consider the function

$$g(x) = f(x) + x.$$

This function is a continuous, strictly increasing function that must have a finite derivative almost everywhere, hence so too must the original function f.

7.8.8. Proof of Theorem 6.17. 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.

Proof. Here are the details of the proof. Items (c), (d), and (e) are exactly 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:

(a) Every null set is measurable.
(b) Every closed set is measurable.
(c) 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.
(d) 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.
(e) If E is measurable then the complement $\mathbb{R} \setminus E$ is also measurable.

Items (a) and (b) are easy. Let us prove (e) first. Let E be measurable and E' is its complement. Let $\epsilon > 0$ and choose an open set G_1 so that $E \setminus G_1$ is closed and $\mathcal{L}(G_1) < \epsilon/2$. Let O be the complement of $E \setminus G_1$; evidently O is open.

First find an open set G_2 with $\mathcal{L}(G_2) < \epsilon/2$ 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 $\epsilon > 0$ and choose open sets G_n so that $\mathcal{L}(G_n) < \epsilon 2^{-n}$ 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

$$\mathcal{L}(G) \leq \sum_{n=1}^{\infty} \mathcal{L}(G_n) \leq \sum_{n=1}^{\infty} \epsilon 2^{-n} = \epsilon.$$

Finally

$$E' = E \setminus G = \bigcap_{n=1}^{\infty} (E_n \setminus G_n)$$

is closed.

For (d), 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 (e) and hence B is measurable by (d). The complement of B, namely the set A, is measurable by (e) again. ■

7.8.9. Proof of Theorem 6.19. 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

$$E_s = \{x : f(x) < s\}$$

must be in terms of the sets $\{E_k\}$. In each case we see that E_s is some simple combination of measurable sets and so is itself measurable. ■

7.8.10. Proof of Theorem 6.20. Every nonnegative, measurable function $f : \mathbb{R} \to \mathbb{R}$ can be written as the sum of a series of nonnegative simple functions:

$$f(x) = \sum_{k=1}^{\infty} f_n(x).$$

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be a nonnegative, measurable function. We express f as

$$f(x) = \sum_{i=1}^{\infty} r_i \chi A_i(x)$$

for a sequence of positive real numbers $\{r_i\}$ and a sequence of measurable sets $\{A_i\}$. 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

$$A_k = \left\{ x : f(x) \geq r_k + \sum_{j<k} r_j \chi A_j(x) \right\}$$

inductively, starting with $A_0 = \emptyset$. Then check that

$$f(x) = \sum_{k=1}^{\infty} r_k \chi A_k(x)$$

at every x. ■

7.8.11. Proof of Theorem 6.21. Let $\{f_n\}$ be a sequence of continuous functions defined on the real line. Suppose that f is a function on $\mathbb{R}$ 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 open, and hence measurable. This follows from the continuity 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. ■

7.8.12. Proof of Theorem 6.22. Every function $f : [a, b] \to \mathbb{R}$ that is integrable on $[a, b]$ is measurable.

Proof. Let F be an indefinite integral for f. We can suppose that $F(x) = F(a)$ for all $x \le a$ and that $F(x) = F(b)$ for all $x \ge b$. Note that the sequence of continuous functions

$$\frac{F(x + 1/n) - F(x)}{1/n}$$

converges for almost every x on the real line. For almost every x in $[a, b]$ its limit is $F'(x) = f(x)$ (using Theorem 6.11) while for *all* points outside $[a, b]$ the limit is zero. The limit function of a sequence of continuous functions is always measurable (Theorem 6.21) and it follows from this that f is a measurable function. ■

7.8.13. Proof of Theorem 6.23. Let K be a compact subset of an interval $[a, b]$. Then χ_K is integrable on $[a, b]$ and

$$\mathcal{L}(K) = \int_a^b \chi_K(x)\,dx.$$

Proof. Let G_1 be the open set complementary to K and consider the covering relation

$$\beta_1 = \{(I, x) : x \in K \text{ and } I \subset [a, b] \text{ or } x \notin K \text{ and } I \subset G_1\}.$$

This is a Cousin cover of $[a, b]$. Let $\pi \subset \beta_1$ be a partition of $[a, b]$. Note that for any $(I, x) \in \pi$ either $x \in K$ and $\chi_K(x) = 1$ or $x \notin K$, $\chi_K(x) = 0$ and $I \cap K = \emptyset$. Set $\pi[K] = \{(I, x) \in \pi : x \in K\}$. Thus

$$\bigcup_{(I,x)\in\pi[K]} I \supset K$$

and hence

$$\sum_{(I,x)\in\pi} \chi_K(x)\mathcal{L}(I) = \sum_{(I,x)\in\pi[K]} \mathcal{L}(I) \ge \mathcal{L}(K).$$

This gives a lower bound for all Riemann sums from any partition $\pi \subset \beta_1$ of $[a, b]$:

$$\mathcal{L}(K) \le \sum_{(I,x)\in\pi} \chi_K(x)\mathcal{L}(I). \tag{7.12}$$

In the other direction let $\epsilon > 0$ and take any open set $G_2 \supset K$ for which $\mathcal{L}(G_2) \le \mathcal{L}(K) + \epsilon$. Consider the covering relation

$$\beta_2 = \{(I, x) : x \in K \text{ and } I \subset G_2 \text{ or } x \notin K \text{ and } I \subset [a, b]\}.$$

This, too, is a Cousin cover of $[a, b]$. Let $\pi \subset \beta_2$ be a partition of $[a, b]$. Note that for any $(I, x) \in \pi$ either $x \in K$, $\chi_K(x) = 1$ and $I \subset G_2$ or $x \notin K$ and $\chi_K(x) = 0$. Thus

$$\sum_{(I,x)\in\pi} \chi_K(x)\mathcal{L}(I) = \sum_{(I,x)\in\pi[K]} \mathcal{L}(I) \le \mathcal{L}(G).$$

This gives an upper bound for all Riemann sums from any partition $\pi \subset \beta_2$ of $[a, b]$:

$$\sum_{(I,x)\in\pi} \chi_K(x)\mathcal{L}(I) \le \mathcal{L}(G_2) \le \mathcal{L}(K) + \epsilon \tag{7.13}$$

Set $\beta = \beta_1 \cap \beta_2$; this is a Cousin cover of $[a,b]$. We deduce from the inequalities (7.12) and (7.13) that, for all Riemann sums from any partition $\pi \subset \beta$ of $[a,b]$,

$$\mathcal{L}(K) \leq \sum_{(I,x)\in\pi} \chi_K(x)\mathcal{L}(I) < \mathcal{L}(K) + \epsilon. \tag{7.14}$$

The identity of the theorem is now clear. ■

7.8.14. Proof of Theorem 6.24. Let E be a measurable subset of a compact interval $[a,b]$. Then χ_E is integrable on $[a,b]$ and

$$\mathcal{L}(E) = \int_a^b \chi_E(x)\,dx.$$

Proof. Using the definition of a measurable set find a decreasing sequence of open sets $\{G_n\}$ with

$$\lim_{n\to\infty} \mathcal{L}(G_n) = 0$$

so that

$$K_n = E \setminus G_n$$

is closed. Note that this expresses χ_E as the limit of an increasing sequence of functions

$$\lim_{n\to\infty} \chi_{K_n}(x) = \chi_E(x)$$

at every point x except in the null set $N = \bigcap_{n=1}^{\infty} G_n$. It follows from the monotone convergence theorem that

$$\int_a^b \chi_E(x)\,dx = \lim_{n\to\infty} \int_a^b \chi_{K_n}\,dx = \lim_{n\to\infty} \mathcal{L}(K_n) = \mathcal{L}(E).$$

■

7.8.15. Proof of Theorem 6.27. Let f be a measurable function on an interval $[a,b]$. Then f is absolutely integrable if and only if

$$\int_a^b |f(x)|\,dx < \infty.$$

Proof. We know, from Exercise 45, that the functions $|f|$, $[f]^+$, and $[f]^-$ are also measurable. The finiteness of this integral implies (by Corollary 6.26) 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. ■

7.8.16. Proof of Theorem 6.28. If f is absolutely 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 absolutely 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. ■

7.8.17. Proof of Theorem 6.29. 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 45 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 6.27) 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. ■

CHAPTER 8

SOLUTIONS FOR SELECTED EXERCISES

8.1. Chapter 1

8.1.1. Solution for Exercise 5. To evaluate

$$\int_0^1 \frac{1}{\sqrt{|x|}}\,dx$$

it is enough to find explicitly an indefinite integral. (Not always possible but here it is.)

The function $F(x) = 2\sqrt{x} = 2x^{\frac{1}{2}}$ has $f(x) = \frac{1}{\sqrt{x}}$ as its derivative at every point $0 < x < 1$. Check that F is continuous on $[0,1]$. Then we can conclude that F is an indefinite integral for f on $[0,1]$ so that

$$\int_0^1 \frac{1}{\sqrt{|x|}}\,dx = F(1) - F(0) = 2.$$

8.1.2. Solution for Exercise 9. (cf. Exercise 5.)
To evaluate

$$\int_{-1}^1 \frac{1}{\sqrt{|x|}}\,dx$$

it is enough to find explicitly an indefinite integral. (Not always possible but here it is with some thinking.)

The function $F_1(x) = 2\sqrt{x} = 2x^{\frac{1}{2}}$ certainly has $f(x) = \frac{1}{\sqrt{x}}$ as its derivative at every point $0 < x < 1$. Similarly the function $F_2(x) = -2\sqrt{-x} = 2[-x]^{\frac{1}{2}}$ has $f(x) = \frac{1}{\sqrt{-x}}$ as its derivative at every point $-1 < x < 0$.

Now define the function $F(x)$ to be $F_1(x)$ on the interval $[0,1]$ and to be $F_2(x)$ on the interval $[-1,0]$ (noting that there is a match at $x = 0$). Check that this function is continuous on $[-1,1]$ and that

$$\frac{d}{dx}F(x) = \frac{1}{\sqrt{|x|}}$$

at all points with the solitary exception $x = 0$. Thus

$$\int_{-1}^1 \frac{1}{\sqrt{|x|}}\,dx = F(1) - F(-1) = 4.$$

8.2. Chapter 2

8.2.1. Solution for Exercise 10. Let F be continuous at a point x_0, let $\epsilon > 0$. Then we can choose $\delta > 0$ so that

$$|F(x) - F(x_0)| < \epsilon/2$$

for all $|x - x_0| < \delta$.

Note that if an interval $[c, d]$ contains the point x_0 and if $d - c < \delta$, then $x_0 - c < \delta$ and $d - x_0 < \delta$. But

$$|F(d) - F(c)| = |[F(d) - F(x_0)] + [F(x_0) - F(c)]|$$
$$\leq |F(d) - F(x_0)| + |F(x_0) - F(c)| < \epsilon/2 + \epsilon/2 = \epsilon.$$

This shows that the covering relation

$$\beta = \{([c, d], x_0) : c \leq x_0 \leq d \text{ and } |F(d) - F(c)| < \epsilon\}$$

contains all these pairs for which $[c, d]$ contains x_0 and $|c - d| < \delta$. Thus β is full at x_0.

8.2.2. Solution for Exercise 11. A smaller and more useful covering relation for continuity uses the following version of Exercise 10: Let $\epsilon > 0$, x_0 a real number and write for a function F

$$\beta = \{([c, d], x_0) : c \leq x_0 \leq d \text{ and } \omega F([c, d]) < \epsilon\}.$$

Suppose that F is continuous at the point x_0. Then β is full at x_0.

Here $\omega F([c, d])$ denotes the *oscillation* of the function F on the interval $[c, d]$. We define it (loosely[1]) as the greatest value obtained by

$$|F(s) - F(t)|$$

for any choices of points s and t in the interval $[c, d]$.

For this exercise repeat the arguments for Exercise 10: Since F is continuous at a point x_0, let $\epsilon > 0$. Then we can choose $\delta > 0$ so that

$$|F(x) - F(x_0)| < \epsilon/3$$

for all $|x - x_0| < \delta$.

Suppose that the interval $[c, d]$ contains the point x_0 and $d - c < \delta$. Take any two points s and t from $[c, d]$. Then $x_0 - s < \delta$ and $t - x_0 < \delta$. Thus

$$|F(s) - F(t)| = |[F(s) - F(x_0)] + [F(x_0) - F(t)]|$$
$$\leq |F(s) - F(x_0)| + |F(x_0) - F(t)|$$
$$< \epsilon/3 + \epsilon/3 = 2\epsilon/3.$$

Consequently

$$\omega F([c, d]) \leq 2\epsilon/3 < \epsilon.$$

This shows that β is full at x_0.

8.2.3. Solution for Exercise 12. Let F be continuous at each point of a set E. Then, by Exercise 10, the covering relation

$$\beta = \{([c, d], x) : c \leq x \leq d \text{ and } |F(d) - F(c)| < \epsilon\}.$$

is a full at each point in E. By definition, then, β is a full cover of E.

---

[1]The correct definition is as

$$\omega F([c, d]) = \sup\{|F(s) - F(t)| : c \leq s \leq t \leq d\}.$$

but as we are not assuming that the student is fully briefed on sups and infs [suprema and infima] so this can be taken in a more relaxed sense.

8.2.4. Solution for Exercise 13. Let F be differentiable at a point x_0, let $\epsilon > 0$. Then we can choose $\delta > 0$ so that

$$|F(x) - F(x_0) - F'(x_0)(x - x_0)| \leq \epsilon|x - x_0|/2$$

for all $|x - x_0| < \delta$. Suppose that $[c, d]$ is an interval containing x_0 and with length less than δ. It must be true that both

$$|F(x_0) - F(c) - F'(x_0)(x_0 - c)| \leq \epsilon(x_0 - c)/2$$

and

$$|F(d) - F(x_0) - F'(x_0)(d - x_0)| \leq \epsilon(d - x_0)/2.$$

But

$$F(d) - F(c) - F'(x_0)(d - c) = [F(d) - F(x_0) - F'(x_0)(d - x_0)]$$

$$+[F(x_0) - F(c) - F'(x_0)(x_0 - c)].$$

Thus simply add these two inequalities to obtain that

$$|F(d) - F(c) - F'(x_0)(d - c)| \leq \epsilon(d - c)|.$$

This shows that the covering relation

$$\beta = \{([c, d], x_0) : c \leq x_0 \leq d \text{ and } |F(d) - F(c) - F'(x_0)(d - c)| \leq \epsilon(d - c)\}.$$

is full at x_0.

8.2.5. Solution for Exercise 14. This follows directly from Exercise 13.

8.2.6. Solution for Exercise 15. Let β_1 and β_2 be full at a point x_0. If β_1 is full at a point x_0 then there is a $\delta_1 > 0$ so that β_1 contains all pairs $([c, d], x_0)$ for which it is true that both $d - c < \delta_1$ and $c \leq x_0 \leq d$. But if β_2 is also full at x_0 then there is a $\delta_2 > 0$ so that β_2 contains all pairs $([c, d], x_0)$ for which it is true that both $d - c < \delta_2$ and $C \leq x_0 \leq d$. Set

$$\delta = \min\{\delta_1, \delta_2\}.$$

Then both β_1 and β_2 (and hence also $\beta_1 \cap \beta_2$) contains any pair $([c, d], x_0)$ for which it is true that both $d - c < \delta$ and $c \leq x_0 \leq d$. Thus $\beta_1 \cap \beta_2$ is full at x_0.

8.2.7. Solution for Exercise 16. This follows directly from Exercise 15.

8.2.8. Solution for Exercise 17. Let β_1 be a full cover of a set E_1 and let β_2 be a full cover of a set E_2. Then $\beta_1 \cup \beta_2$ is full at any point x in E_1 (merely because $\beta_1 \cup \beta_2$ contains β_1). Similarly $\beta_1 \cup \beta_2$ is full at any point x in E_2 (merely because $\beta_1 \cup \beta_2$ contains β_2). Thus $\beta_1 \cup \beta_2$ is a full cover of $E_1 \cup E_2$.

8.3. Chapter 3

8.3.1. Solution for Exercise 28. That every subset of a null set is also a null set follows directly from the definition.

8.3.2. Solution for Exercise 29. Suppose that E_1 and E_2 are null sets and let $E = E_1 \cup E_2$ be their union. Let $\epsilon > 0$. Choose a full cover β of E_1 so that

$$\sum_{i=1}^{n}(b_i - a_i) < \epsilon/2$$

whenever the collection

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

is chosen from β_1 in such a way that the intervals $\{[a_i, b_i]\}$ do not overlap. Similarly choose a full cover β_2 of E_2 with the same property.

Let $\beta = \beta_1 \cup \beta_2$. Then, by Exercise 17, β is a full cover of E. Take any subpartition

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

from β, i.e., any subcollection with the property that the intervals $\{[a_i, b_i]\}$ do not overlap. Define

$$\gamma_1 = \gamma \cap \beta_1 \quad \text{and} \quad \gamma_2 = \gamma \cap \beta_2$$

so that γ_1 and γ_2 contain only elements $([a_i, b_i], \xi_i)$ that happen also to belong to β_1 or β_2. Here the γ_k are also subpartitions (perhaps empty)[2].

This allows us to estimate the sum

$$\sum_{i=1}^{n}(b_i - a_i) \leq \sum_{([a_i,b_i],\xi_i)\gamma_1}(b_i - a_i) + \sum_{([a_i,b_i],\xi_i)\gamma_2}(b_i - a_i) =$$

$$< \epsilon/2 + \epsilon/2 = \epsilon.$$

We have verified the statement that E is null.

8.3.3. Solution for Exercise 30. Suppose that E_1, E_2, E_3, ... is a sequence of null sets and let

$$E = E_1 \cup E_2 \cup E_3 \cup \dots$$

be their union. Let $\epsilon > 0$.

For each integer $k = 1, 2, 3, \dots$ choose a full cover β_k of E_k so that

$$\sum_{i=1}^{n}(b_i - a_i) < \epsilon 2^{-k}$$

whenever the collection

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

is chosen from β_k in such a way that the intervals $\{[a_i, b_i]\}$ do not overlap.

Let β be the union of the sequence of covering relations $\{\beta_k\}$. Then, it is easy to check (point-by-point) that β is a full cover of E. Take any subpartition

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

from β, i.e., any subcollection with the property that the intervals $\{[a_i, b_i]\}$ do not overlap. Define

$$\gamma_k = \gamma \cap \beta_k\}.$$

[2]We haven't assumed that the γ_1 and γ_2 are disjoint; they need not be as defined, but we could arrange them to be so in a similar argument if that were needed. Since they may have a pair in common we have had to use an inequality in the proof where an equal sign might have seemed appropriate.

so that γ_k $(k = 1, 2, 3, \dots)$ just contains elements $([a_i, b_i], \xi_i)$ from γ that happen also to belong to β_k. Most of these new subpartitions are empty: in fact since γ contains only n elements, there are at most n of the γ_k that are nonempty. Certainly we can find a large enough integer K so that $\gamma_k = \emptyset$ for $k > K$

Now, since γ_k is a subset of β_k, we have

$$\sum_{([a_i,b_i],x_i)\in\gamma_k} (b_i - a_i) < \epsilon 2^{-k}.$$

(Interpret an empty sum as zero.)

It follows that

$$\sum_{i=1}^{n}(b_i - a_i) \leq \sum_{k=1}^{K}\left(\sum_{([a_i,b_i],x_i)\in\gamma_k} (b_i - a_i)\right) < \sum_{k=1}^{K}\epsilon 2^{-k} < \epsilon.$$

We have verified the statement that E is null.

8.3.4. Solution for Exercise 25. A set E is a null set if and only if the function $F(x) = x$ does not grow on E. This follows merely from writing out the definitions for the two statements and checking that they are indeed identical. Make sure to use the fact that

$$\omega F([c, d]) = d - c$$

for the function $F(x) = x$.

8.3.5. Solution for Exercise 32. We show that a continuous function F does not grow on a finite set. Let $\epsilon > 0$ and let $E = \{e_1, e_2, e_3, \dots, e_k\}$. Define

$$\beta = \{([c, d], e_j) : \ j = 1, 2, \dots, k \text{ and } \omega F([c, d]) < \epsilon/(2k)\}.$$

By the fact that F is continuous at each point in E we see that β is a full cover of E. Take any collection

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

from β so that the intervals $\{[a_i, b_i]\}$ do not overlap. Then, by the way in which β was constructed, for any $i = 1, 2, \dots, n$ we must have $x_i = e_j$ for some $j = 1, 2, \dots, k$ and $\omega F([c, d]) < \epsilon/k$. Note that there are, at most, two choices of i with the same j. It follows that

$$\sum_{i=1}^{n} \omega F([a_i, b_i]) < 2k(\epsilon/(2k) = \epsilon.$$

By definition, then, F does not grow on E.

8.3.6. Solution for Exercise ??. We show that a continuous function F does not grow on a set E whose elements can be written out as a sequence of points e_1, e_2, ...

Let $\epsilon > 0$. Define

$$\beta = \{([c, d], e_j) : \ j = 1, 2, \dots, \text{ and } \omega F([c, d]) < \epsilon 2^{-j-1}\}.$$

By the fact that F is continuous at each point in E we see that β is a full cover of E. Take any collection

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

from β so that the intervals $\{[a_i, b_i]\}$ do not overlap. Then, by the way in which β was constructed, for any $i = 1, 2, \dots, n$ we must have $x_i = e_j$ for some $j = 1, 2, \dots,$

and $\omega F([c,d]) < \epsilon 2^{-j-1}$. Note that there are, at most, two choices of i with the same j. It follows that

$$\sum_{i=1}^{n} \omega F([a_i,b_i]) < 2\sum_{j=1}^{\infty} \epsilon 2^{-j-1} = \epsilon.$$

By definition, then, F does not grow on E.

8.3.7. Solution for Exercise 26. We show that a Lipschitz function F cannot grow on a negligible set N. Since F is Lipschitz there is a positive constant M so that

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

for all real x and y.

Let $\epsilon > 0$. Use the fact that N is a negligible set to construct a full cover β of N with the property that

$$\sum_{i=1}^{n}(b_i - a_i) < \epsilon/M$$

whenever the collection

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

is chosen from β in such a way that the intervals $\{[a_i,b_i]\}$ do not overlap. For this same collection observe that

$$|\Delta F([a_i,b_i])| \leq M(b_i - a_i).$$

Consequently

$$\sum_{i=1}^{n} |\Delta F([a_i,b_i])| \leq \sum_{i=1}^{n} M(b_i - a_i) < \epsilon.$$

By definition, then, F does not grow on N.

8.3.8. Solution for Exercise 38. Since F is Lipschitz there is a positive constant M so that

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

for all real x and y. Let $\epsilon > 0$. Set $\delta = \epsilon/M$. Take any collection of intervals $\{[a_i,b_i]\}$ that do not overlap and for which

$$\sum_{i=1}^{n}(b_i - a_i) < \delta.$$

Observe that

$$|\Delta F([a_i,b_i])| \leq M(b_i - a_i).$$

Thus

$$\sum_{i=1}^{n} |\Delta F([a_i,b_i])| \leq \sum_{i=1}^{n} M(b_i - a_i) < M\delta = \epsilon.$$

Thus, by definition, F is absolutely continuous.

8.3.9. Solution for Exercise 3.15. We show that an absolutely continuous function F on a compact interval $[a,b]$ cannot grow on a negligible set N contained in $[a,b]$.

Let $\epsilon > 0$. By the definition of absolute continuity there is a $\delta > 0$ so that whenever a collection of subintervals

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

of $[a,b]$ is given that do not overlap and for which

$$\sum_{i=1}^{n} (b_i - a_i) < \delta$$

it must be the case that

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

Use the fact that N is a negligible set to construct a full cover β of N with the property that

$$\sum_{i=1}^{n} (b_i - a_i) < \delta$$

whenever the collection

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

is chosen from β in such a way that the intervals $\{[a_i, b_i]\}$ do not overlap. For this same collection observe immediately that

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

By definition, then, F does not grow on N.

8.3.10. Solution for Exercise 24. Suppose that F is a continuous function F that does not grow on each member of a sequence of sets E_1, E_2, E_3, ... Let

$$E = E_1 \cup E_2 \cup E_3 \cup \dots$$

be their union. Let $\epsilon > 0$.

Since F does not grow on a E_k there can be found a full cover β_k of that set E_k so that

$$\sum_{i=1}^{n} \omega F([a_i, b_i]) < \epsilon 2^{-k}$$

whenever the collection

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

is chosen from β_k in such a way that the intervals $\{[a_i, b_i]\}$ do not overlap.

Let β be the union of all the covering relations β_k. Then, β is a full cover of E. Take any subpartition

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

from β, i.e., any subcollection with the property that the intervals $\{[a_i, b_i]\}$ do not overlap. Define

$$\gamma_k = \gamma \cap \beta_k.$$

Most of these are empty: in fact since γ contains only n elements, there are at most n of the γ_k that are nonempty. Choose a large enough integer K so that $\gamma_k = \emptyset$ for $k \geq K$

Now, since γ_k is a subset of β_k, we have

$$\sum_{([a_i,b_i],x_i)\in\gamma_k} \omega F([a_i, b_i]) < \epsilon 2^{-k}.$$

(Interpret an empty sum as zero.)

It follows that

$$\sum_{i=1}^{n} \omega F([a_i, b_i]) \leq \sum_{k=1}^{K} \left(\sum_{([a_i,b_i],x_i)\in\gamma_k} \omega F([a_i, b_i]) \right) < \sum_{k=1}^{K} \epsilon 2^{-k} < \epsilon.$$

We have verified the statement that F does not grow on E.

8.3.11. Solution for Exercise 27. Suppose that F is a continuous function that is differentiable at each point of a set E. For each integer $k = 1, 2, 3, \dots$ let E_k be the set of points in E at which $|F'(x)| < k$. Let N be any null subset of E. We will show that F does not grow on any set $N \cap E_k$. (The demonstration is very close to the same argument used for Lipschitz functions.)

Define the covering relation

$$\beta_k = \{([c, d], x) : \ x \in E_k \text{ and } \omega F([c, d]) \leq k(d - c)\}.$$

Because of our assumption that $F'(x)$ exists at each point $x \in E_k$ and that $|F'(x)| < k$ it is easy to check that this is a full cover of E_k.

Let $\epsilon > 0$ and fix an integer k. Since N is null, there is a full cover β'_k of N so that

$$\sum_{i=1}^{n} (b_i - a_i) < \epsilon/k$$

whenever the subpartition

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

is chosen from β'_k, i.e., a finite subcollection with the property merely that the intervals $\{[a_i, b_i]\}$ do not overlap.

Now consider the covering relation

$$\beta = \beta'_k \cap \beta_k.$$

This is evidently a full cover of the set $N \cap E_k$. Take any subpartition

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

from β. Then, by the way in which β was constructed, each

$$\omega F([a_i, b_i]) \leq k(b_i - a_i)$$

and, moreover,

$$\sum_{i=1}^{n} (b_i - a_i) < \epsilon/k$$

It follows that

$$\sum_{i=1}^{n} \omega F([a_i, b_i]) \leq \sum_{i=1}^{n} k(b_i - a_i) < \epsilon.$$

By definition, then, F cannot grow on the negligible set $N \cap E_k$.

By Exercise 24 it follows that F cannot grow on the union of the sequence of sets $\{N \cap E_k\}$. But this union is all of N.

8.4. Chapter 4

8.4.1. Solution for Exercise 43. Let f be the function defined to be 1 at every irrational number and as 0 at every rational number. Define $g(x) = 1$ everywhere and observe that f and g agree except at a sequence of points, namely the sequence of all rational points. Consequently f is integrable on any interval since g is integrable on any interval and

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

where $G(x) = x$.

Additional Reading

The texts listed here, all of them except for the first, use the so-called "gauge" version of the natural integral. In this text we have chosen to express this idea using covering relations. I see no advantages in the gauge version except one that has been promoted occasionally. For readers familiar (or too familiar) with the usual Riemann integral one can promote the natural integral by pointing out that it is formally only a very small change in the classical definition of Riemann. Where before was a small positive number δ measuring the fineness of partitions, now there is a small positive function δ performing that measurement. But having committed oneself to this unfortunate definition we will now be required to manipulate this gauge and refer constantly in proofs and definitions to the gauge. For the majority of the proofs in the theory this adds a level of messiness and detail that has irritated many potential users.

- Alan Smithee, *The Fundamental Program of the Calculus*, [to appear] 2007.
- Douglas S. Kurtz and Charles W. Swartz, *Theories Of Integration: The Integrals Of Riemann, Lebesgue, Henstock-Kurzweil, and Mcshane*, (Series in Real Analysis) World Scientific Publishing Company (2004) [ISBN: 9812388435].
- Robert G. Bartle, *A modern theory of integration*, Providence, American Mathematical Society, 2001.
- V.G. Celidze and A.G. Dzvarseisvili, *The theory of the Denjoy integral and some applications* (trans. P.S. Bullen), Singapore, World Scientific, 1989.
- Russell A. Gordon, *The integrals of Lebesgue, Denjoy, Perron, and Henstock*, Providence, American Mathematical Society, 1994.
- Ralph Henstock, *Lectures on the theory of integration*, World Scientific, Singapore, 1988.
- Ralph Henstock, *The General Theory of Integration*, (Oxford Mathematical Monographs) Oxford University Press, (1991) [ISBN: 019853566X]
- Solomon Leader, *The Kurzweil-Henstock integral and its differentials*, Marcel Dekker, New York, 2001.
- Lee Peng-Yee, *Lanzhou lectures on Henstock integration*, Singapore, World Scientific, 1989.
- Robert M. McLeod, *The generalized Riemann integral*, Washington, The Mathematical Association of America, 1980.
- Patrick Muldowney, *A general theory of integration in function spaces*, Harlow, Longman, 1987.

- Washek Pfeffer, *The Riemann approach to integration*, Cambridge University Press, Cambridge, 1993.
- Washek Pfeffer, *Derivation and integration*, Cambridge University Press, Cambridge, 2001.
- Charles Swartz, *Introduction to gauge integrals*, Singapore, World Scientific, 2001.
- Stefan Schwabik, *Generalized ordinary differential equations*, Singapore, World Scientific, 1992.
- Lee Peng Yee and Rudolf Vyborny, *The integral: an easy approach after Kurzweil and Henstock*, Cambridge University Press, Cambridge, 2000.

The following are in languages other than English:

- Jean Mawhin, *Analyse–fondements, techniques, evolution*, De Boeck Universite, Brussels, 1992.
- Štefan Schwabik: *Integrace v R (Kurzweilova teorie)* [= Integration in R (Kurzweil's theory)] Karolinum, Praha (1999) 326 pages, ISBN 80-7184-967-7.

Essay: Why the Riemann integral is more difficult

Most instructors of the calculus are likely to hold the mistaken belief that the natural integral on the real line (presented in these notes) is a more difficult choice for a theoretical calculus course than the Riemann integral. After all, they might reason, the natural integral is more general, more general even than the Lebesgue integral. Therefore does it not follow that the theory must be harder?

They have heard that the natural integral is easier than the Lebesgue integral since it does not require the development of all of the tools of measure theory at the outset. This they would regard as a curiosity; by the time the student reaches the stage of learning advanced integration theory why not throw measure theory at them anyway? They will need it.

But it can hardly seem possible that this integral is actually a competitor in simplicity and ease of application to the Riemann integral. Isn't the Riemann integral the universal choice for an introductory course?

This brief essay attempts to show that the Riemann integral presents barriers to an adequate theory of integration that the natural integral does not. Even if the ambition of the course is at the level of a first real analysis course, the latter integral is easier to present.

8.5. The fundamental theorem of the calculus

In all presentations of the Riemann integral the following theorem is stated and perhaps proved:

THEOREM 8.5.1. *Let F be a differentiable function on an interval $[a,b]$ and suppose that <u>the function $f(x) = F'(x)$ is Riemann integrable</u> on $[a,b]$. Then*

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

The underlined words are the offensive ones here. Without them the theorem is invalid. Their presence makes the proof fairly sophisticated. Also what have we proved? Until we know what functions are integrable the theorem is useless!

Essentially the proof [which need not be given formally here] follows this sketch. Use an ϵ, δ definition of the integrability of f to obtain a subdivision

$$a = a_0 < a_1 < \cdots < a_{n-1} < a_n = b$$

and collect these three facts:

$$F(b) - F(a) = \sum_{i=1}^{n}[F(a_i) - F(a_{i-1})], \tag{8.1}$$

$$\sum_{i=1}^{n}[F(a_i) - F(a_{i-1})] = \sum_{i=1}^{n} f(\xi_i)[a_i - a_{i-1}], \tag{8.2}$$

and

$$\sum_{i=1}^{n} f(\xi_i)[a_i - a_{i-1}] \approx \int_a^b f(x)\,dx. \tag{8.3}$$

The first identity (8.1) is just some algebra. The second identity (8.2) calls on the mean-value theorem of the calculus to select the points ξ_i from the intervals $[a_{i-1}, a_i]$ and the final approximation (8.3) uses the integrability hypothesis on f.

There are two very nontrivial ideas here: the integrability is used to select the δ and the mean-value theorem to make the selection of the points that provide the connection between the values of the Riemann sum and the value $F(b) - F(a)$.

In contrast the natural integral uses no sophisticated theory, simply the definition of the derivative and the definition of the integral. The additional hypothesis is removed: removing an hypothesis can sometimes make a proof easier.

THEOREM 8.5.2. *Let F be a differentiable function on an interval $[a, b]$. Then the function $f(x) = F'(x)$ is integrable on $[a, b]$ and*

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

The three computations reduce to two computations that can now be sketched as

$$F(b) - F(a) = \sum_{i=1}^{n}[F(a_i) - F(a_{i-1})] \tag{8.4}$$

and

$$\sum_{i=1}^{n}[F(a_i) - F(a_{i-1})] \approx \sum_{i=1}^{n} f(\xi_i)[a_i - a_{i-1}]. \tag{8.5}$$

The first identity (8.4), as before, is just some algebra. The second approximation (8.5) is merely the definition of the derivative. The partitions allowed are exactly the ones that allow such an approximation.

8.6. Improving the Riemann version

We can make the Riemann version just as easy provided we recognize the real structure of what that theory is. Let us define

DEFINITION 8.1. A function F is said to be *uniformly differentiable* on an interval $[a, b]$ if there is a function $F'(x)$ with the property that for every $\epsilon > 0$ there is a $\delta > 0$ so that

$$\left|\frac{F(d) - F(c)}{d - c} - F'(x)\right| < \epsilon$$

whenever $x \in [c, d] \subset [a, b]$ and $0 < d - c < \delta$.

Now the correct elementary version of the fundamental theorem of the calculus for the Riemann integral is this:

THEOREM 8.6.1. *Let F be a uniformly differentiable function on an interval $[a,b]$. Then the function $f(x) = F'(x)$ is Riemann integrable on $[a,b]$ and*

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

The proof is obvious, also elementary. It follows using only the definition of the integral and the definition of the uniform derivative.

But how might it be applied in a first course? Checking that a derivative exists uniformly would be a bit overwhelming for most students. The following lemma gives an elementary way to apply this.

LEMMA 8.2. Let F be a differentiable function on an interval $[a,b]$ and suppose that the function $f(x) = F'(x)$ is continuous at each point of $[a,b]$. Then F is uniformly differentiable on $[a,b]$.

Proof. The proof uses only the Bolzano-Weierstrass version of a compactness argument; it does not require that the student know that f is uniformly continuous. If F is not uniformly differentiable on $[a,b]$ then there must be an $\epsilon_0 > 0$ so that for every $\delta > 0$

$$\left| \frac{F(d) - F(c)}{d - c} - f(x) \right| \geq \epsilon_0$$

for at least one choice of $x \in [c,d] \subset [a,b]$ for which $0 < d - c < \delta$.

That allows the selection of a sequence of points x_n and containing intervals $[c_n, d_n]$ for which

$$\left| \frac{F(d_n) - F(c_n)}{d_n - c_n} - f(x_n) \right| \geq \epsilon_0$$

and for which $0 < d_n - c_n < 2^{-n}$. There must be a convergent subsequence $\{x_{n_k}\}$ converging to a point z in $[a,b]$. The continuity of f at z and the fact that $F'(z) = f(z)$ will contradict this construction. ■

With the lemma to hand we now have a version of the fundamental theorem for the Riemann integral that is useful and elementary [assuming that one accepts this compactness argument as elementary].

COROLLARY 8.3. Let F be a differentiable function on an interval $[a,b]$ and suppose that the function $f(x) = F'(x)$ is continuous at each point of $[a,b]$. Then f is Riemann integrable on $[a,b]$ and

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

This might be considered acceptable for a first course. The theorem that continuous functions are always Riemann integrable is deeper still. It requires proving that continuous functions are uniformly continuous and some harder arguments to establish that uniformly continuous functions are integrable.

The natural integral trumps all of this as regards the fundamental theorem. All derivatives are integrable! The proof is a complete triviality.

8.7. Continuous functions are integrable

Any modestly ambitious course on the Riemann integral will prove that continuous functions are integrable. More accurately such a course will prove that *uniformly* continuous functions are integrable and likely show that functions everywhere continuous in a compact interval are uniformly continuous there.

Let us review the methods. Roughly speaking both the natural integral and the Riemann integral can be described this way: provided a partition

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

of an interval $[a, b]$ is "small enough" [or sufficiently fine?] then all the Riemann sums

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

will be close to the value of the integral.

The technique used to characterize integrability [Cauchy criterion] can be similarly described as asserting that whenever two partitions

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

and

$$\pi' = \{([a'_j, b'_j], \xi'_j) : j = 1, 2, \dots, m\}$$

are both "small enough" then the corresponding Riemann sums must be close together, i.e.,

$$\sum_{i=1}^{n} f(\xi_i)(b_i - a_i) \approx \sum_{j=1}^{m} f(\xi'_j)(b'_j - a'_j).$$

For both the natural integral and the Riemann integral, then, an existence theorem will naturally require an estimate of the difference between two different Riemann sums. In the notes we have exploited the trivial identity

$$\sum_{i=1}^{n} f(\xi_i)\mathcal{L}([a_i, b_i]) - \sum_{j=1}^{m} f(\xi'_j)\mathcal{L}([a'_j, b_j]') =$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} [f(\xi_i) - f(\xi'_j)]\mathcal{L}([a_i, b_i] \cap [a'_j, b'_j])$$

for this purpose. [Here the length of an interval I is described by using the length function $\mathcal{L}(I)$.]

Now we can discuss our existence theorems.

THEOREM 8.4. If f is uniformly continuous on $[a, b]$ then f is Riemann integrable on $[a, b]$.

Given $\epsilon > 0$ one selects $\delta > 0$ so that if $\mathcal{L}([c, d]) < 2\delta$ the function f cannot oscillate too much on $[c, d]$, specifically so that

$$\omega f([c, d]) < \epsilon/(b - a).$$

Then if the two partitions π and π' above are both "smaller than this δ" it is easy to estimate the difference between the Riemann sums. The only contributions to the sum occur when

$$\mathcal{L}([a_i, b_i] \cap [a'_j, b'_j]) \neq 0$$

and when that happens ξ_i and ξ'_j are inside an interval of length smaller than 2δ. Consequently

$$\left|f(\xi_i) - f(\xi'_j)\right| < \epsilon/(b-a).$$

The proof is easy enough then to write up. Compare it with the version for the natural integral.

THEOREM 8.5. If f is continuous at each point of $[a,b]$ then f is integrable on $[a,b]$.

We collect all the pairs

$$\beta = \{([c,d],x) : \omega f([c,d]) < \epsilon/[2(b-a)]\}$$

we easily check that β satisfies the necessary requirements at each point merely because f is continuous at each point. Then if the two partitions π and π' above are both contained in this β it is, again, easy to estimate the difference between the Riemann sums. The ideas are identical and the arithmetic is almost the same.

But there is a curious difference. The natural integral is sufficiently flexible that the proof could be carried through without knowing in advance that f is uniformly continuous; we merely used continuity at each point. For the Riemann integral this is not the case; uniform continuity was an essential assumption.

We might argue [we will argue!] that this is misleading. It is merely an artifact of the proof that uniform continuity is required. The student is easily led to believe that nonuniformity of the continuity would destroy integrability. Is the function $f(x) = \sin x^{-1}$ integrable on $[0,1]$? Certainly. But the student of the Riemann integral would likely not believe it. The student of the natural integral can prove that bounded functions with only a finite number of discontinuities are integrable; this same method of proof works with some mild modifications.

The classical Riemann integral, as it is usually taught, little prepares the student for an understanding even of its own properties, let alone an understanding of the true nature of integration theory on the real line.

8.8. Limits and limitations

If $\{f_n\}$ is a convergent sequence of integrable functions one often needs to have access to the formula

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b \left[\lim_{n\to\infty} f_n(x)\right]\,dx.$$

The greatest limitation for this formula when it is the Riemann integral that is being taught is that the limit function may be integrable in more general senses, but does not happen to be integrable in the Riemann sense.

As the student doubtless learns, one has to view this problem as involving a change in the order of limits. Normally some kind of assumption of uniformity can be used to allow this. For this problem there are two choices: Either the nature of the convergence of $\{f_n(x)\}$ to its limit is uniform in x or the nature of the integrability of the functions $\{f_n\}$ is uniform.

For the Riemann integral we can focus on these two using these conventional definitions:

DEFINITION 8.6. A sequence of functions $\{f_n\}$ converges uniformly to a function f on an interval $[a, b]$ if for every $\epsilon > 0$ there is an integer N so that

$$|f_n(x) - f(x)| < \epsilon$$

for all $n \geq N$ and all x in $[a, b]$.

DEFINITION 8.7. A sequence of functions $\{f_n\}$ is equi-Riemann-integrable on an interval $[a, b]$ if for every $\epsilon > 0$ there is a $\delta > 0$ so that

$$\left| \int_a^b f_n(x)\, dx - \sum_{i=1}^{n} f_n(\xi_i)(b_i - a_i) \right| < \epsilon$$

for all partitions $\pi = \{([a_i, b_i], \xi_i) : i = 1, 2, \dots, n\}$ of the interval $[a, b]$ finer[3] than δ.

These definitions lead to the following entirely elementary theorem, which is the best that the Riemann integral has to offer without developing new tools.

THEOREM 8.8. Let $\{f_n\}$ be an everywhere convergent sequence of Riemann integrable functions on an interval $[a, b]$. Then

$$\lim_{n\to\infty} \int_a^b f_n(x)\, dx = \int_a^b \left[\lim_{n\to\infty} f_n(x) \right] dx.$$

provided either that

(a) The sequence $\{f_n\}$ converges uniformly on the interval $[a, b]$, or
(b) The sequence $\{f_n\}$ is equi-Riemann-integrable on the interval $[a, b]$.

The proof is trivial. Most elementary courses, committed though they are to the Riemann integral, prove only the first version, the version in which the convergence is assumed to be uniform. The other version, in which the integrability is assumed to be uniform, appears to be rare. Presumably this is because the conditions might seem hard to verify and so the student will find few occasions to apply the condition. These course miss an opportunity to explore a different kind of uniformity.

Note that the first condition [uniform convergence] implies the second condition [equi-integrability]. This condition might seem to be worth exploring, but leads nowhere that the student should be taken. It is equivalent to requiring uniform boundedness [a reasonable requirement for the Riemann integral] in addition to equicontinuity almost everywhere. But exploring that is troublesome with no better tools to hand. Worse yet it is misleading. The equicontinuity plays a role only in ensuring that the limit function is Riemann integrable, not in establishing the formula. Again the student could well be encouraged in the illusion that equicontinuity plays some kind of role in the theory of integral limits.

For the natural integral it is the equi-integrable assumption that is the strongest and most compelling. Again, though, the proof remains trivial. Thus the theory is as easy as for the Riemann integral, but the theory is not misleading. In time the student will learn that this version includes formally the most common assumptions of the advanced theory, the bounded convergence theorem, the dominated convergence theorem, the monotone convergence theorem etc.

[3]Finer than δ here means that each interval in the partition has length smaller than δ.

8.9. Arzelà's bounded convergence theorem

As a test of the Riemann integral consider the following problem which is a slight, but critical, variant of the problem that we discussed in the preceding section. While it might seem that this problem shows again a limitation of the Riemann integral the purpose of this section is to defend the Riemann integral against this charge; this does not advance our thesis but we need to be fair.

> **Problem.** Suppose that $\{f_n\}$ is a convergent sequence of functions with
> $$f(x) = \lim_{n\to\infty} f_n(x)$$
> at each point of a compact interval $[a, b]$. Suppose that all of these functions, *including* f, are Riemann integrable on $[a, b]$. Under what conditions can we conclude that
> $$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx?$$

The advanced reader knows immediately that the most obvious choice of conditions for this problem is that the functions are uniformly bounded, i.e., that $|f_n(x)| \leq M$ for some real number M and all x in $[a, b]$ and integers n. The elementary reader would have assumed uniform convergence which is unfortunate. But how would the more ambitious hypothesis be handled without the tools of measure theory and Lebesgue's version of the bounded convergence theorem?

There are a number of discussions of this topic at the level of an elementary real analysis course that can be found in the Monthly. Luxemburg's article is particularly worth reading as it contains a wealth of historical observations from a master mathematician.

(a) W. A. J. Luxemburg, Arzelà's Dominated Convergence Theorem for the Riemann Integral, *The American Mathematical Monthly*, Vol. 78, No. 9. (Nov., 1971), pp. 970–979.
(b) Jonathan W. Lewin, A Truly Elementary Approach to the Bounded Convergence Theorem, *The American Mathematical Monthly*, Vol. 93, No. 5 (May, 1986), pp. 395–397.
(c) Jonathan W. Lewin, Some Applications of the Bounded Convergence Theorem for an Introductory Course in Analysis, The *American Mathematical Monthly*, Vol. 94, No. 10 (Dec., 1987), pp. 988–993.

Interestingly enough this problem is not as advanced as it might appear and can be easily handled by the Riemann theory of integration with only a few simple measure-theoretic tools. In the problem show that

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

by establishing this simpler version that substitutes $g_n(x) = |f_n(x) - f(x)|$:

> *Lemma 1: Suppose that* $0 \leq g_n(x) \leq 2M$ *for all* $x \in [a, b]$ *and that*
> $$\lim_{n\to\infty} g_n(x) = 0 \quad (a \leq x \leq b).$$
> *Then*
> $$\lim_{n\to\infty} \int_a^b g_n(x)\,dx = 0.$$

By using just the lower integral the problem gets reduced in complexity. Indeed once we have expressed the problem in this manner, it should be transparent that it is rather less ambitious an undertaking than might be imagined.

To tackle this problem we assemble the following tools. These are close to an introduction to Lebesgue's theory of measure, but can clearly be considered as belonging naturally to an account of the Riemann integral.

Define an elementary set to be a finite union of closed intervals. Define the Lebesgue measure of such an elementary set E to be $\mathcal{L}(E)$, merely the sum of the lengths of the intervals. Assemble and prove some of the simpler properties of Lebesgue measure, but just for elementary sets.

Prove the following not-too-difficult lemma:

> *Lemma 2: Let $\{A_n\}$ be a contracting sequence of bounded sets with empty intersection. Define*
>
> $$t_n = \sup\{\mathcal{L}(E) : \ E \textit{ an elementary subset of } A_n\}.$$
>
> *Then* $\lim_{n\to\infty} t_n = 0$.

This will require that the student knows that an intersection of a sequence of compact sets that has the finite intersection property must have a nonempty intersection. Nothing deeper than that is needed. Then introduce step functions and make sure that the student is aware that the lower integral is approximated from below by step functions.

Now the proof of Lemma 1, and hence of the Arzelà bounded convergence theorem, is attainable by a very simple argument.

Proof. Let $\epsilon > 0$ and define the set of points

$$A_n = \{x \in [a,b] : g_k(x) \geq \epsilon/[2(b-a)] \text{ for at least one integer } k \geq n\}.$$

The sequence $\{A_n\}$ forms a contracting sequence of bounded sets with empty intersection. Consequently, by Lemma 2, there must be an integer N so that every elementary subset $E \subset A_N$ must have measure smaller than $\epsilon/(4M)$.

Let $n \geq N$ and suppose that s is a step function for which

$$0 \leq s(x) \leq g_n(x)$$

for all x in $[a,b]$. Let E be the elementary set of points x in $[a,b]$ for which

$$s(x) \geq \epsilon/[2(b-a)].$$

Certainly $E \subset A_N$ so that

$$\mathcal{L}(E) \leq \epsilon/(4M).$$

On the points x in E we can use that $s(x) \leq 2M$ while on remaining points x in $[a,b]$ outside of E we know that $s(x) < \epsilon/[2(b-a)]$.

Consequently the integral of the step function s is easily estimated to be

$$\int_a^b s(x)\,dx \leq 2M\mathcal{L}(E) + \epsilon(b-a)/[2(b-a)] < \epsilon.$$

All such step functions satisfy this inequality and so it follows that, for all $n \geq N$,

$$\underline{\int_a^b} g_n(x)\,dx \leq \epsilon.$$

■

Although this simple argument can be used to defend the Riemann integral we still remain concerned that the requirement that the limit function be assumed to be Riemann integrable, which clearly reduces the level of difficulty, also reduces the applicability of the Arzelà theorem.

8.10. The bounded convergence theorem for the natural integral

It is useful to compare the correct version of the bounded convergence theorem with the Arzelà version. This version is deeper because the integrability of the limit function is proved, not assumed.

> **Bounded Convergence Theorem.** Suppose that $\{f_n\}$ is a uniformly bounded and convergent sequence of functions with
> $$f(x) = \lim_{n\to\infty} f_n(x)$$
> at each point of a compact interval $[a,b]$. Suppose that all of the functions, $\{f_n\}$ are integrable on $[a,b]$. Then f too must be integrable on $[a,b]$ and
> $$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx.$$

We sketch a proof of this theorem based on some of the rudiments of Lebesgue's theory of measure. In Lebesgue's version the measure plays a key role in the definition of the integral and in the development of its properties. For the natural integral the definition is obtained much more simply, but the measure theory remains as one of the most important tools in understanding the integral. The first lemma is a harder version of Lemma 2.

> *Lemma 3: Let $\{A_n\}$ be a contracting sequence of bounded sets with empty intersection. Then for every $\epsilon > 0$ there is an open set G and an integer N so that*
> $$\mathcal{L}(G) < \epsilon \text{ and } A_N \subset G.$$

We present the proof of the bounded convergence theorem using some of the same language and steps as for the proof of the Arzelà version. The key is to use the measure lemma (Lemma 3) to show that the sequence of functions $\{f_n\}$ must be equi-integrable.

Proof. We suppose that $|f_n(x)| < M$ for all n and x, that each f_n is integrable on $[a,b]$ and that $f(x) = \lim_{n\to\infty} f_n(x)$. Set
$$g_k(x) = |f_k(x) - f(x)|.$$
Let $\epsilon > 0$ and define the set of points
$$A_n = \{x \in [a,b] : g_k(x) \geq \epsilon/[12(b-a)] \text{ for at least one integer } k \geq n\}.$$
The sequence $\{A_n\}$ forms a contracting sequence of bounded sets with empty intersection. Consequently, by Lemma 3, there must be an integer N and an open set G so that
$$\mathcal{L}(G) < \epsilon/(12M)$$
and so that
$$A_N \subset G.$$

Choose a Cousin cover β of $[a,b]$ so that for all $n = 1,2,3,\dots,N$ and any partition π from β,

$$\left| \int_a^b f_n(x)\,dx - \sum_{(I,x)\in\pi} f_n(x)\mathcal{L}(I) \right| < \epsilon/6.$$

This is possible since each function is integrable and there are only finitely many functions to handle.

Define

$$\beta_1 = \{(I,x) \in \beta : \text{ if } x \in A_N \text{ then } I \subset G\}.$$

This removes some elements from β but the result is still a Cousin cover of $[a,b]$.

Let n be any integer and consider a pair of partitions π_1 and π_2 of $[a,b]$ chosen from β_1. From the way in which we chose β we see that

$$\left| \sum_{(I,x)\in\pi_1} f_n(x)\mathcal{L}(I) - \sum_{(I,x)\in\pi_2} f_n(x)\mathcal{L}(I) \right| < \epsilon/3 \tag{8.6}$$

for $n = 1,2,3,\dots,N$. Our goal is to obtain a similar inequality for all remaining integers n.

Consider now any integer n greater than N. In the computation below we make use of the fact that if x in A_N and $(I,x) \in \beta_1$ then

$$|f_n(x) - f_N(x)| \le 2M$$

and

$$I \subset G$$

while on remaining points x in $[a,b]$ outside of A_N we know that

$$g_N(x) < \epsilon/[12(b-a)] \ \text{ and } \ g_n(x) < \epsilon/[12(b-a)]$$

so that, in particular,

$$|f_n(x) - f_N(x)| \le |f_n(x) - f(x)| + |f(x) - f_N(x)| < 2\,\{\epsilon/[12(b-a)]\} = \epsilon/[6(b-a)].$$

Thus for any partition π of $[a,b]$ from β_1 we must have

$$\left| \sum_{(I,x)\in\pi} [f_n(x) - f_N(x)]\mathcal{L}(I) \right| < (2M)\mathcal{L}(G) + (b-a)\,\{\epsilon/[6(b-a)]\} < \epsilon/3.$$

From that we deduce that the inequality (8.6) can be improved to assert that

$$\left| \sum_{(I,x)\in\pi_1} f_n(x)\mathcal{L}(I) - \sum_{(I,x)\in\pi_2} f_n(x)\mathcal{L}(I) \right| < \epsilon$$

must in fact hold for *all* integers n.

This proves that the sequence $\{f_n\}$ is equi-integrable. As we have seen in Section 8.8 we can conclude quickly and simply now that the limit function f is integrable and that its integral is the limit of the integrals of the f_n. ■

8.11. Measure theory

Some of the rudiments of Lebesgue's theory of measure can easily be introduced in a first real analysis course. But when the only integration theory available is the Riemann integral there hardly seems any point.

Consider that, as part of a discussion of open sets, closed sets, and compact sets one also introduces the following concepts:

(a) The measure $\mathcal{L}(G)$ of an open set G is the sum of the lengths of the component intervals.

(b) The measure $\mathcal{L}(K)$ of a compact set K is the infimum of the measures $\mathcal{L}(G)$ for open sets G containing K.

(c) The null sets are the sets contained in open sets of arbitrarily small measure.

(d) Null functions are functions that vanish outside of a null set.

This is not an ambitious set of topics but it is a worthwhile collection of ideas to introduce at an elementary level. If the integration theory used in the course is the natural integral, not the Riemann integral, then the following are easy enough to state and prove:

(a) If f is a null function then it is integrable on every interval $[a, b]$ and $\int_a^b f(x)\,dx = 0$.

(b) If K is a closed set and $[a, b]$ any compact interval then χ_K is integrable on $[a, b]$ and

$$\int_a^b \chi_K(x)\,dx = \mathcal{L}(K \cap [a, b]).$$

These are false for the Riemann integral. Even bounded null functions need not be Riemann integrable. For most closed sets K the characteristic function χ_K would not be Riemann integrable. Doubtless one could find some artificial applications of measure theory to Riemann integration but they would serve only a weak pedagogic purpose.

The Quiz

This quiz is designed for graduate students who have just completed their study of the Lebesgue integral and are expected to remember their freshman calculus study of the Riemann integral. They will be unable to use the natural integral, of course, but should be able to answer using the techniques and theory of the Riemann and Lebesgue integrals.

8.12. The Quiz

Solve each of the following questions using

- THE INTEGRAL[4].
- The Riemann integral.
- The Lebesgue integral.

#1. What is

$$\int_0^1 x^2\,dx?$$

SOLUTION ON PAGE 114.

#2. What is

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx?$$

SOLUTION ON PAGE 115.

#3. What is

$$\int_{-1}^1 \frac{1}{\sqrt{|x|}}\,dx?$$

SOLUTION ON PAGE 116.

#4. What is your point?

SOLUTION ON PAGE 117.

[4]i.e., the natural integral on the real line, otherwise known as the Denjoy integral, the Perron integral, the Denjoy-Perron integral, the restricted integral of Denjoy, the Denjoy total, the Henstock integral, the Kurzweil integral, the Henstock-Kurzweil integral, the Kurzweil-Henstock integral, the generalized Riemann integral, the Riemann-complete integral, the gage integral, the gauge integral, etc. We call it here simply THE INTEGRAL.

8.13. Problem #1

$$\int_0^1 x^2\,dx?$$

8.13.1. The integral: Observe that, with $f(x) = x^2$ and $F(x) = x^3/3$, we have $F'(x) = f(x)$ at every point of the interval $[0, 1]$.

Consequently we can use the following theorem of integration theory (known sometimes as the fundamental theorem of the calculus).

> *Theorem A. If $F'(x) = f(x)$ at every point of $[a, b]$ then f is integrable on $[a, b]$ and the value of the integral is exactly*
>
> $$\int_a^b f(x)\,dx = F(b) - F(a).$$

We conclude that f is integrable on [0,1] and the value of the integral is

$$\int_0^1 x^2\,dx = F(1) - F(0) = 1/3.$$

8.13.2. The Riemann integral: Observe that, with $F(x) = x^3/3$, we have $F'(x) = f(x)$ at every point of the interval $[0, 1]$. We cannot use Theorem A since that is *false* for the Riemann integral. But there is a rather pathetic variant we can use:

> *Theorem A^R. If $F'(x) = f(x)$ at every point of $[a, b]$ and assume in addition that <u>f is Riemann integrable on $[a, b]$</u>. Then the value of the integral is exactly $F(b) - F(a)$.*

Thus we check first that f is continuous on $[0, 1]$, we apply a theorem asserting that continuous functions are Riemann integrable and finally we conclude that f is integrable on $[0, 1]$ and the value of the integral is $F(1) - F(0) = 1/3$.

8.13.3. The Lebesgue integral: [No, you are not allowed to say the Riemann integral is included in the Lebesgue integral, thereby reducing to the solution above!]

Observe that, with $F(x) = x^3/3$, we have $F'(x) = f(x)$ at every point of the interval $[0, 1]$. We cannot use Theorem A since that is false for both the Riemann integral and the Lebesgue integral. But the following happens to be true:

> *Theorem A^L. If $F'(x) = f(x)$ at every point of $[a, b]$ and assume in addition that <u>f is Lebesgue integrable on $[a, b]$</u> then the value of the integral is exactly $F(b) - F(a)$.*

This too is a pathetic variant. But, even so, it is very seldom proved in analysis courses and the average graduate student would be unlikely to know the statement or be able to prove it. (Most would believe it but be amazed at not being able to prove it.) So we would not be able to appeal to this theorem.

But there is a weaker variant we can use:

> *Theorem B^L. If $F'(x) = f(x)$ at [almost] every point of $[a, b]$ and <u>if F is absolutely continuous on $[a, b]$</u> then f is integrable on $[a, b]$ and the value of the integral is exactly $F(b) - F(a)$.*

Thus we check first that F is Lipschitz on $[0,1]$ with Lipschitz constant 2 (just use the mean-value theorem) and then invoke a theorem asserting that Lipschitz functions are absolutely continuous. Again we conclude that f is integrable on $[0,1]$ and the value of the integral is $F(1) - F(0) = 1/3$.

8.14. Problem #2

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx?$$

Note: The student will likely be disturbed by the fact that the integrand is undefined at the point $x = 0$. This is a mystery. All Riemann type theories are uninfluenced by the value of the function at a single point. So for both the integral and the Riemann integral simply ignore the point or assign some other convenient value there. For the Lebesgue integral (and indeed for THE INTEGRAL) a set of measure zero may be ignored, so the single point $x = 0$ shouldn't distress graduate students.

8.14.1. The integral. Observe that, with $F(x) = 2\sqrt{x}$, we have $F'(x) = f(x)$ at every point of the interval $[0,1]$ with one exception. Consequently we cannot use Theorem A above, since that requires a derivative at *every* point. Instead we have the following theorem of integration theory, a modification of Theorem A.

> *Theorem C. If F is a continuous function on $[a,b]$ and if $F'(x) = f(x)$ at every point of $[a,b]$ with a number of exceptions [at most countably many exceptions], then f is integrable on $[a,b]$ and the value of the integral is exactly $F(b) - F(a)$.*

We conclude that f is integrable on $[0,1]$ and the value of the integral is

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx = F(1) - F(0) = 2.$$

8.14.2. The Riemann integral. Trick question! You cannot use the Riemann integral for unbounded functions. A properly disciplined calculus student will use the *improper* version of the Riemann integral.

Observe that, with

$$F(x) = 2\sqrt{x}$$

we have $F'(x) = f(x)$ at every point of the interval $[0,1]$ with one exception, at $x = 0$. But this function f is unbounded on $[0,1]$ and so f is not Riemann integrable. But most students have learned an extension of the Riemann integral, known as "the improper integral." Theorem C is false for both the Riemann and improper Riemann integrals.

Instead the student must fall back on the following canonical ritual of the elementary calculus. This function f is continuous on $[t,1]$ for all $t > 0$ so is Riemann integrable. By applying Theorem A^R determine that this integral has value

$$\int_t^1 \frac{1}{\sqrt{x}}\,dx = F(1) - F(t) = 2 - 2\sqrt{t}.$$

Take the limit as $t \to 0+$ from the right and, using the continuity of the function F, obtain the finite value $F(1) - F(0) = 2$.

Finally we conclude that f is integrable [in the sense of the improper Riemann integral] on $[0, 1]$ and the value of the integral is $F(1) - F(0) = 2$.

8.14.3. The Lebesgue integral. Observe that, with $F(x) = 2\sqrt{x}$, we have $F'(x) = f(x)$ at every point of the interval $[0, 1]$ with one exception. We cannot use Theorem C since that is false for the Lebesgue integral. We cannot (surprisingly for some students) use the "improper" prescribed procedure just used either, for that is not available for the Lebesgue integral. [The ritual which was nearly a religious obligation in elementary calculus is forbidden to acolytes of the Lebesgue integral.]

We could try for this theorem:

> *Theorem C^L. If F is a continuous function on $[a, b]$, if $F'(x) = f(x)$ at every point of $[a, b]$ with countably many exceptions and if f is Lebesgue integrable on $[a, b]$, then the value of the integral is exactly $F(b) - F(a)$.*

This is true, but it is not at all taught in typical graduate courses. So the student has little recourse but to return to Theorem A^L and verify that the function F is absolutely continuous.

This function is not Lipschitz so that fact will need verification, likely from the definition. Again we conclude (after some considerable labor) that f is integrable on $[0, 1]$ and the value of the integral is $F(1) - F(0) = 2$.

An alternate method would be to use the monotone convergence theorem and truncate the function f by $f_n(x) = \min\{n, f(x)\}$, then letting $n \to \infty$ and seeing that f_n increases to f. But in either the case the student is embarrassed by the considerable machinery of Lebesgue theory that needs to be brought to bear on a problem which caused much less grief as a freshman student.

8.15. Problem #3

$$\int_{-1}^{1} \frac{1}{\sqrt{|x|}}\, dx?$$

Note: Once again he student will likely be alarmed by the fact that the integrand is undefined at the center of the interval. Indeed most calculus texts would avoid this; they are less bothered if the undefined points occur at an endpoint.

8.15.1. The integral. It is easy enough to tailor an indefinite integral: take $F(x) = 2\sqrt{x}$ for $x \geq 0$ and as $-2\sqrt{-x}$ for $x < 0$. Then F is continuous and $F'(x) = f(x)$ with one exception. Apply the usual fundamental theorem of the calculus (Theorem C) to obtain

$$\int_{-1}^{1} \frac{1}{\sqrt{|x|}}\, dx? = F(1) - F(-1) = 4.$$

8.15.2. The Riemann integral. Split this into the two integrals

$$\int_{-1}^{1} \frac{1}{\sqrt{|x|}} = \int_{-1}^{0} \frac{1}{\sqrt{|x|}} + \int_{0}^{1} \frac{1}{\sqrt{|x|}}.$$

Then worry. There don't seem to be any theorems about the improper Riemann integral in our calculus book that allow this, but it seems likely!

8.15.3. The Lebesgue integral. See the solution for the Riemann integral. That would be ok for the Lebesgue integral.

8.16. Question #4

What is your point?

Well if your students have no troubles with problems #1, #2, or #3 then maybe there is no point. There are anecdotes about graduate students given problem #1 on an oral exam and freezing in a panic.

Index